Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Shared structure and variation across populations, individuals and time in early vocabulary development: A data-driven latent-space approach

Hiromichi Hagihara

hiromichi.h@gmail.com

The University of Osaka https://orcid.org/0000-0003-3316-600X

Ye Li

The Hong Kong Polytechnic University

Chie Fukada

Kyoto Institute of Technology

Yasuhiro Kanakogi

The University of Osaka,

Masaaki Sakagami

The University of Osaka

Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11744224/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
There is **NO** Competing Interest.

We have already and explicitly mentioned the research ethics statement on the title page (page 3), which you might have looked over:Ethics Statement This study analyzed existing large-scale datasets, primarily from publicly available sources, and did not involve the collection of new data from human participants. Access to and reuse of data obtained through Databrary were approved by the Research Ethics Committee of the Graduate School of Human Sciences, The University of Osaka (HB-R8-073).Because

this study reused existing datasets and we did not collect new data from human participants, the "participant consent" issue should not be applicable.

Shared structure and variation across populations, individuals and time in early vocabulary development: A data-driven latent-space approach

Author Names and Affiliations

Hiromichi Hagihara [a,b,c,*] (hiromichi.h.hus@osaka-u.ac.jp)

Ye Li [d,*] (ye007.li@polyu.edu.hk)

Chie Fukada [e] (chieft@kit.ac.jp)

Yasuhiro Kanakogi [a] (kanakogi.hus@osaka-u.ac.jp)

Masaaki Sakagami [a,b,f] (masaakisakagami@gmail.com)

[a)] Graduate School of Human Sciences, The University of Osaka, Suita, JAPAN.

[b)] Faculty of Interdisciplinary Science and Engineering in Health Systems, Okayama University, Okayama, JAPAN.

[c)] International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo Institutes for Advanced Study, Tokyo, JAPAN.

[d)] Department of Language Science and Technology, The Hong Kong Polytechnic University, Kowloon, Hong Kong

[e)] Faculty of Arts and Sciences, Kyoto Institute of Technology, Kyoto, Japan.

[f)] Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan.

* HH and YL contributed equally to this work and share first authorship.

Corresponding Author

Hiromichi Hagihara, Graduate School of Human Sciences, The University of Osaka, 1-2 Yamadaoka, Suita-shi, Osaka 565-0871, JAPAN. / Faculty of Interdisciplinary Science and

Engineering in Health Systems, Okayama University, 3-1-1 Tsushima-naka, Kita-ku, Okayama 700-8530, JAPAN.

Email: hiromichi.h.hus@osaka-u.ac.jp

ORCID ID: https://orcid.org/0000-0003-3316-600X

Acknowledgements

We thank the researchers who generously shared their data and the broader open-science community whose efforts made this study possible. We are also grateful to the participants and their families for their contributions to the data. We thank Tenchi Mizutani and Hiroki Yamamoto for assistance with coding and for valuable feedback.

Data and Code Availability

The data and code used in this work are available at: https://anonymous.4open.science/r/vocab-dev-latent-space-2C5F/

Funding Sources

This research was supported by JST BOOST Program Japan Grant Number JPMJBY25D9 (to HH) and JSPS KAKENHI Grant Numbers JP23K11786 (to HH, CF, and MS) and JP20H05002 (to MS).

Conflicts of Interest

The authors have no conflict of interest to declare in the submission of this manuscript.

Ethics Statement

This study analyzed existing large-scale datasets, primarily from publicly available sources, and did not involve the collection of new data from human participants. Access to and reuse of data obtained through Databrary were approved by the Research Ethics Committee of the Graduate School of Human Sciences, The University of Osaka (HB-R8-073).

Author contributions

HH: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration, Funding acquisition.

YL: Conceptualization, Methodology, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization.

CF: Conceptualization, Writing - Review & Editing, Funding acquisition.

YK: Conceptualization, Writing - Review & Editing.

MS: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing - Review & Editing, Visualization, Funding acquisition.

Generative AI statement

Generative AI tools were used to assist with English-language editing, manuscript organization, and code organization and documentation. All AI-assisted content was reviewed and verified by the authors. All scientific decisions, analyses, interpretations, and final text and code remain the full responsibility of the authors.

Shared structure and variation across populations, individuals and time in early vocabulary development: A data-driven latent-space approach

Abstract

Young children's vocabularies exhibit shared developmental regularities alongside substantial variation in developmental pathways. Study 1 applied a variational autoencoder to caregiver-reported productive vocabularies from 9,093 English-learning children (16–30 months) to construct a developmental latent space. Study 2 tested generalizability in 1,056 Mandarin-learning children. Without age or word-category labels, the models organized vocabularies primarily by overall size, while two further dimensions of lexical composition emerged among similarly sized vocabularies: Nominality (prominence of nouns versus verbs/function words) and Commonality (widely shared versus idiosyncratic words). Although English was modestly noun-biased and Mandarin modestly verb-biased overall, within-language variation crossed this population-typical contrast. Study 3 mapped dense longitudinal records from four English-learning children (11–30 months) into the latent space and revealed within-child changes in lexical learning tendencies along these dimensions as vocabularies grew. Our framework integrates shared developmental regularities with variation across three scales: across populations, among individuals, and within individuals over development.

Keywords: language development, early word learning, infancy, cross-cultural research, data-driven analysis, individual variation, variation across time

Introduction

How young children learn words in their first few years of life is a striking feat. Research on early language development has revealed both shared developmental regularities and

substantial variation in the pathways through which children achieve them. Among the established population-level regularities across linguistic and cultural contexts are developmental milestones. Infants babble around 7 months, gesture communicatively between 6 and 8 months, produce first words around 12 months, and show a vocabulary spurt during their second year.[1–5] Across languages, early vocabularies are dominated by concrete nouns and people words.[6–10]

Variation is equally prominent within populations. Classic work has long documented individual differences in early word learning.[6,7,11,12] Longitudinal observations of children's first 50 words revealed individual differences in lexical composition: some children learned predominantly object names, whereas others learned more social and functional words for communicative purposes.[7] Large-scale caregiver-report data from 8-30-month-old children, collected using the MacArthur–Bates Communicative Development Inventories (CDI), a vocabulary checklist for young children, likewise revealed substantial differences among same-aged children in vocabulary composition, alongside differences in learning rate (learning words rapidly or at a gradual rate) and the onset of vocabulary spurt.[2,6,13] Such variation raises the question of how differences in vocabulary development arise and how they can be characterized.

Recent empirical studies have attempted to link variation in children's everyday auditory and visual environments to differences in word learning.[14–17] Extensive home recordings of a single child showed that the emergence of first productive words was predicted by the distinctiveness of their spatial (where words were heard), temporal (when words were heard), and linguistic (the conversational topics in which words appeared) contexts, beyond word frequency alone.[17] Similarly, home recordings of parent–child interactions with 6-month-old infants, followed by laboratory tests of common-noun knowledge, showed that better word comprehension was associated with word input quality – the co-occurrence of heard words and infants' visual

attention to their referents – rather than word frequency alone.[16] More broadly and critically, children's word learning takes place within their everyday experiences with surrounding people, home activities, and linguistic and cultural environments.[14,15,18,19] Such experiences can vary from child to child and may therefore constitute a major source of variation in vocabulary development that cannot be fully explained by shared population-level regularities alone.

Variation in early word learning occurs not only across children but also within children across different time points. Such temporal variation can be closely associated with children's sensorimotor development, which shapes caregivers' language input.[20–22] For instance, caregivers produce more locomotion verbs (e.g., "come") to their walking children than to their crawling children.[22] Vocabulary growth trajectories themselves may also differ across children.[23–26] An intensive longitudinal study of 45 infants aged 16‑30 months, whose expressive vocabulary was measured on 617 occasions, revealed that the rate of vocabulary growth changed over time (a slow early phase followed by a steep acceleration) and that the timing and steepness of these changes varied widely across individuals.[24] Beyond developmental changes in caregiver input and the pace of vocabulary growth, little is known about whether children's word learning tendencies – or the kinds of words they tend to learn – changes systematically within individuals over development, and whether such changes follow shared or child-specific patterns. Characterizing both regularities and variation in these within-child trajectories is therefore essential for understanding vocabulary development beyond differences between children.

Taken together, these findings suggest that developmental diversity coexists with regularities shared across children. Despite long-standing theories and recent advances in understanding early word learning, these shared regularities and diverse pathways have largely been characterized at different scales and through different methodological approaches. Large-

scale caregiver-report questionnaires, such as the CDI,[13,27] capture population-level developmental structure and between-child variation, whereas small-sample case studies track individual developmental trajectories densely over time.[17] However, few approaches characterize both macro-scale population structure and micro-scale within-child trajectories within a common framework. One way to bridge these macro- and micro-scale perspectives is to first extract and visualize shared developmental structure from large-scale population data, and then situate individual longitudinal trajectories within that developmental landscape. This would allow us to ask what is preserved and what varies across language samples, among children, and within children over time. By extracting shared developmental structure and quantifying variation relative to it across these scales, such a framework could provide a basis for identifying the processes that generate both developmental regularity and variation. Achieving this integration requires tools that can quantify and track variation in early word learning across individuals and over time, while accommodating a large sample, a large set of vocabulary items, and intensive longitudinal measurements.

To this end, we introduced a data-driven machine-learning approach based on a variational autoencoder (VAE).[28,29] A VAE is an unsupervised generative neural network trained to encode high-dimensional input data into a low-dimensional latent space and then reconstruct the original input from its latent representation. This approach offers several advantages for better characterizing word-learning variation across individuals and time points, compared with traditional approaches. First, rather than pre-classifying words into their distinctive categories (e.g., nouns, verbs, or adjectives[9,10,30]), a VAE has the potential to reveal bottom-up, data-driven groupings of words that share similar developmental patterns. Second, as a nonlinear latent-variable model, it can compress high-dimensional vocabulary profiles into a low-dimensional

space that represents both population-level vocabulary growth and differences in vocabulary composition among children at comparable developmental levels. Third, the same approach can be conveniently applied across languages to test whether analogous patterns of latent organization generalize across distinct linguistic contexts. Fourth, unlike methods that only project data into a lower-dimensional space, a VAE's decoder maps locations in the latent space, including intermediate locations not occupied by observed data, back onto estimated production probabilities for individual words. This allows us to characterize different vocabulary profiles across many individuals at a given benchmark (e.g., children with comparable vocabulary sizes) and to trace changes within a single child over time by estimating which kinds of words are more likely to be learned at a given location in the latent structure.

Building on this framework, the present study examined shared regularities and variation in early vocabulary development across three studies spanning three nested scales from *macro* to *micro*: population, within-population, and within-individual levels. Study 1 constructed a developmental latent space from large-scale, mostly cross-sectional productive vocabulary data from American English-learning children (9,093 children aged 16–30 months), identified major dimensions organizing vocabulary development, including overall vocabulary growth and variation in vocabulary composition, and characterized the words that distinguish positions along these dimensions. Study 2 applied the same approach to Beijing Mandarin Chinese-learning children (1,056 children aged 16–30 months) to test whether the latent organization identified in English generalized to a distinct linguistic context, including whether analogous dimensions emerged, and to examine how their population-level expression differed. For Studies 1 and 2, we used CDI data from Wordbank, an open-access database containing caregiver-reported vocabulary data from over 100,000 children across 50 languages and dialects.[31,32] Finally, Study 3 situated

densely sampled longitudinal vocabulary records from four English-learning children aged approximately 11–30 months – three from Sullivan et al.[33] and one from Roy et al.[17] – within the English latent space to trace how individual positions along the identified dimensions changed as their vocabularies grew. Together, these studies characterize developmental variation at three scales – across populations, among children within populations, and within children over time – while distinguishing overall vocabulary growth from variation in which words children learn.

Results

Study 1: Mapping population-level structure and between-child variation in American English

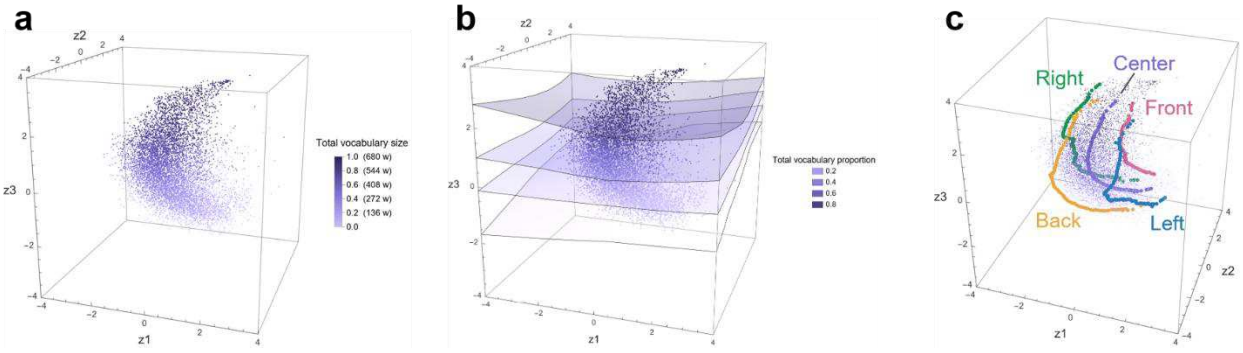
The developmental landscape of early vocabulary

We trained a VAE model on the American English CDI dataset (N = 9,093, 680 words) and visualized the resulting three-dimensional latent space (**Fig. 1a**). The CDI profiles formed a curved, arc-like structure with noticeable thickness, resembling a gently bent cucumber-like shape. Along this curved structure, total vocabulary size increased systematically from 0 to 680, progressing from bottom to top. The model successfully captured an overall developmental axis.

To further investigate how total vocabulary size was organized within the latent space, we used the VAE decoder to estimate total vocabulary size across the space and visualized locations with equivalent vocabulary sizes as isosurfaces (**Fig. 1b**). The isosurfaces were arranged approximately orthogonal to the curved developmental axis. This pattern indicates that movement along the arc corresponds to the overall progression of lexical development, whereas each isosurface represents profiles with comparable total vocabulary sizes. Variation within the same isosurface therefore reflects differences in which words children with the same vocabulary size

produce. These analyses, using the VAE encoder–decoder framework, provide an interpretable and quantitatively nuanced representation of early vocabulary development.

Fig. 1: Latent-space geometry of vocabulary development in American English.



(a) Each point represents one CDI administration (N = 9,093). Data formed a smooth, elongated manifold resembling a cucumber-like shape, along which total vocabulary size increased. The color indicates a gradual increase in total vocabulary size from the bottom to the top. **(b)** Isosurfaces of total vocabulary size, plotted at 20% intervals (with 4 slices). These are arranged approximately orthogonal to the curved developmental axis. Cross-sections of the manifold capture individual variation in lexical composition among children with equivalent vocabulary sizes. **(c)** Five representative developmental trajectories on the manifold. The central trajectory corresponds to the typical developmental path. Taking each point on this typical trajectory as a local origin, four additional trajectories were sampled at ±2 *SD* along directions of maximal variance within the isosurface. For clarity, these trajectories are referred to as the Left, Right, Back, and Front paths.

Nominality and Commonality as major dimensions of within-population variation

To identify the major dimensions of individual variation across levels of total vocabulary size, we analyzed the cross-sectional structure of the latent space. We defined a series of isosurfaces at 10% increments of total vocabulary size (0.1 to 0.9) and identified the central point of each. These points together trace the most typical developmental trajectory through the latent space. Individual variation was then characterized as deviations from this trajectory along two orthogonal directions that captured the largest variance within each isosurface. For clarity, we refer to these directions as the Left–Right and Back–Front axes.

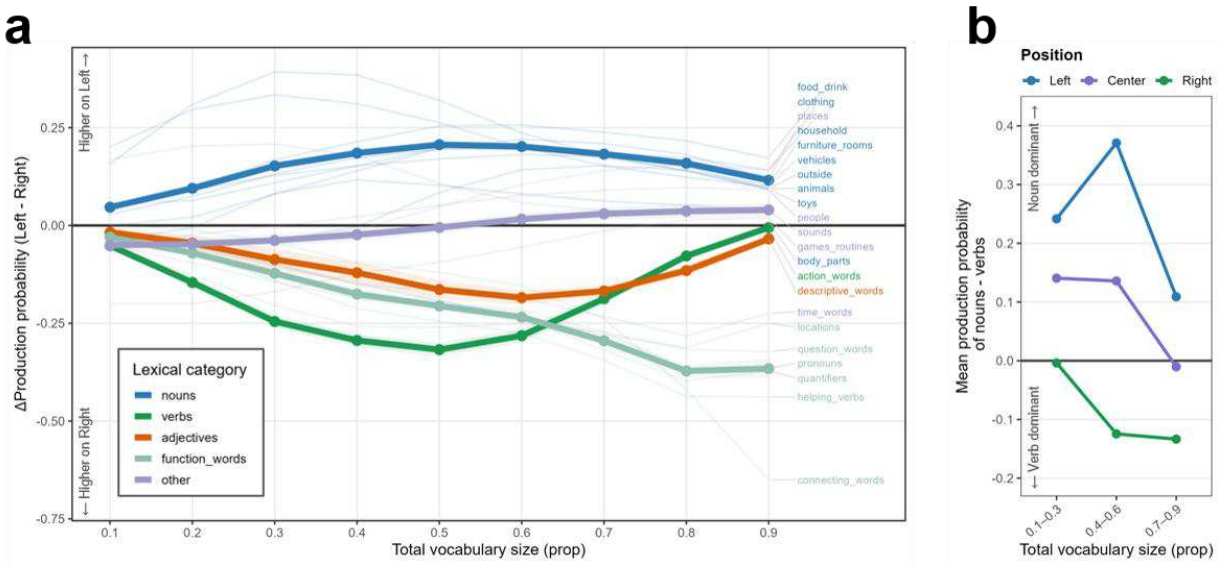
We generated synthetic CDI profiles to probe variation along these axes. At each isosurface, we sampled 100 latent-space points around the central point and locations at ±2 *SD* along each axis, and decoded them using the VAE. The decoder outputs continuous production probabilities for each word rather than binary "produced or not" CDI responses. This procedure yielded simulated vocabulary profiles spanning five representative developmental paths (**Fig. 1c**). We next used these profiles to interpret variation along the Left–Right and Back–Front axes.

We first examined the Left–Right axis. At each total vocabulary level, we calculated, for each CDI item, the difference in estimated production probability between the Left and Right samples (Left minus Right), and then averaged these differences within lexical categories. A clear and consistent pattern emerged across development (**Fig. 2a**): words referring to concrete objects (i.e., nouns) had higher probabilities on the Left, whereas predicates (i.e., verbs and adjectives) and function words (e.g., pronouns and helping verbs) had higher probabilities on the Right. The dominant contribution on the Right shifted across development, with verbs playing a larger role at early to intermediate stages and function words becoming more prominent later. Thus, the Left–

Right axis captures a continuum from noun-dominant to predicate- and function-word-dominant vocabularies. We therefore interpret this axis as reflecting *Nominality*.

We further confirmed this interpretation by computing the mean difference in production probability between nouns and verbs for each synthetic profile generated at a sampled location in the latent space. Profiles sampled from the Left were strongly noun-dominant, whereas those from the Right were verb-dominant (**Fig. 2b**). Profiles along the central trajectory also showed a modest noun bias, suggesting that the typical developmental pathway is systematically skewed toward nouns. This pattern is consistent with previous findings of noun bias in early English vocabularies.[9,10,34]

Fig. 2: The Left–Right axis captures Nominality in vocabulary development



(a) Differences in estimated production probability (Left minus Right) for each word, averaged within lexical categories across levels of total vocabulary size. Nouns show consistently higher probabilities on the Left side, whereas predicates (verbs and adjectives) and function words show

higher probabilities on the Right side. Despite no prior information about lexical categories being provided to the model, a noun–predicate distinction emerges in a data-driven manner along the Left–Right axis. **(b)** Mean differences in production probability between nouns and verbs for profiles sampled from the Left, Center, and Right positions. Estimates are averaged across three ranges of total vocabulary size: 0.1–0.3, 0.4–0.6, and 0.7–0.9. Points represent means, and error bars indicate standard errors. Profiles sampled from the Left are consistently noun-dominant, whereas those from the Right are verb-dominant. The Center profiles show a slight noun bias, suggesting that the typical developmental trajectory is not neutral but moderately noun-oriented. Estimates at each 10% increment of total vocabulary size are shown in **Supplementary Fig. S1**.

We next examined the Back–Front axis, the major within-isosurface direction orthogonal to the Left–Right (i.e., Nominality) direction. To characterize this dimension, we first defined a word-level index of acquisition timing. For each item, we estimated the total vocabulary size at which its production probability exceeded 0.5 using a logistic fit, and termed this measure the vocabulary of acquisition (VoA), analogous to age of acquisition (AoA) commonly used in developmental research.

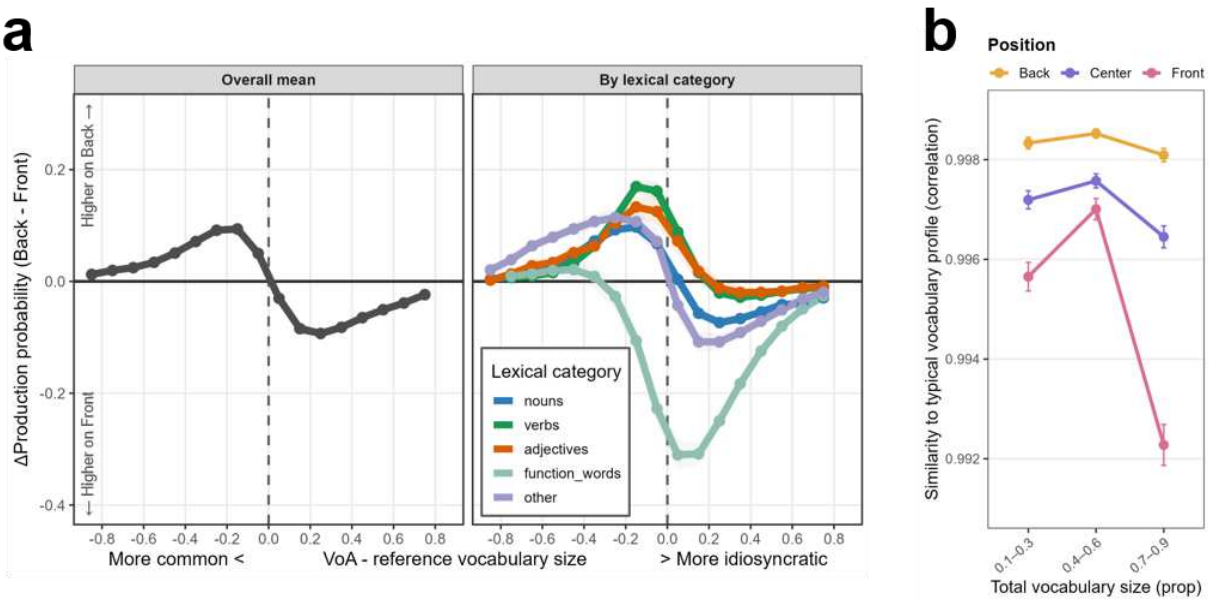
We then related each item's VoA to the reference total vocabulary size, defined as the developmental level from which each sample was drawn (i.e., the total vocabulary size corresponding to each isosurface). We computed their difference (VoA minus reference vocabulary size), and examined how this measure correlated with the mean difference in production probability between the Back and Front samples (Back minus Front). Items with VoA lower than the reference level (i.e., more commonly acquired words given the total vocabulary

size) showed higher probabilities on the Back side, whereas items with higher VoA (i.e., less commonly acquired words) showed higher probabilities on the Front side (**Fig. 3a**). This relationship was generally consistent both in the overall mean and across lexical categories. Thus, the Back-Front axis captures variation in whether children's vocabularies are composed of widely shared, early-acquired words or more idiosyncratic, typically later-acquired words. We therefore interpret this axis as reflecting *Commonality*.

We further evaluated this interpretation at the level of full CDI profiles. If the Back–Front axis indeed reflects Commonality, profiles sampled from the Back should be more similar to one another, whereas those from the Front should be more heterogeneous. To test this, we constructed, for each level of total vocabulary size and each position (Back, Center, Front), a mean CDI profile (a 680-dimensional vector of production probabilities), and computed the correlation between this mean profile and each sampled profile (**Fig. 3b**). Across total vocabulary levels, similarity was consistently highest for profiles sampled from the Back, intermediate for those from the Center, and lowest for those from the Front. Thus, Back profiles were more tightly clustered around their corresponding mean profile, whereas Front profiles were more heterogeneous. This pattern provides converging evidence that the Back–Front axis captures a continuum from widely shared to more idiosyncratic vocabularies.

Together, these analyses show that the latent space captures not only a dimension of overall developmental progression but also at least two distinct sources of individual diversity: Nominality and Commonality. The latent space provides an interpretable framework for characterizing diverse pathways of early vocabulary development.

Fig. 3: The Back-Front axis captures Commonality in vocabulary development



(a) Differences in estimated production probability between the Back and Front samples (Back minus Front) plotted against the difference between each word's vocabulary of acquisition (VoA) and the reference total vocabulary size. The left panel shows the overall pattern across all items, and the right panel shows the pattern across lexical categories. Words with lower VoA relative to the reference level (i.e., more commonly acquired words) show higher probabilities on the Back side, whereas words with higher VoA (i.e., less commonly acquired words) show higher probabilities on the Front side. **(b)** Similarity between sampled CDI profiles and the corresponding mean profile, computed as the correlation between each profile and the mean profile within each total vocabulary size level and position (Back, Center, Front). Estimates are averaged across three ranges of total vocabulary size: 0.1–0.3, 0.4–0.6, and 0.7–0.9. Points represent means, and error bars indicate standard errors. Across all three ranges, profiles sampled from the Back show higher similarity, indicating more homogeneous vocabulary composition, whereas profiles from the Front

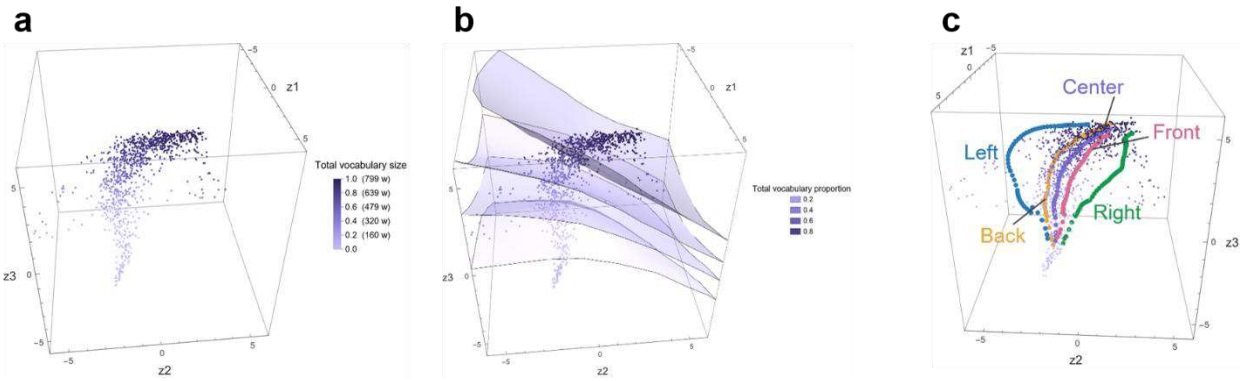
show lower similarity, reflecting greater variability across samples. Estimates at each 10% increment of total vocabulary size are shown in **Supplementary Fig. S2**.

Study 2: Testing the generalizability of the developmental structure in Mandarin Chinese

Is the structure uncovered in American English generalizable to other languages? To this end, we applied the same VAE-based framework to Mandarin Chinese, a linguistically distinct context frequently contrasted with English in developmental research and well represented in Wordbank. Mandarin differs from English in its early lexical profile, particularly in the relative prominence of verbs,[8–10,35–37] making it a stringent test case for generalizability. Using Mandarin CDI data (N = 1,056, 799 words), we trained a separate VAE model and analyzed the resulting latent space following the same procedure as for English. This allowed us to assess whether similar latent structures and dimensions of individual variation would emerge in a typologically distinct system.

The geometry of the three-dimensional latent space in Mandarin Chinese differed visibly from that in American English. Whereas the English data formed a thick, elongated, cucumber-like manifold, the Mandarin data exhibited a flatter, laterally extended, sea-ray-like structure (**Fig. 4a**). Despite this difference in global geometry, a clear developmental structure emerged. When data points were color-coded by total vocabulary size, they showed a systematic progression along a curved trajectory, similar to the developmental axis in English, and isosurfaces of total vocabulary size could be defined (**Fig. 4b**). These results demonstrate that a coherent axis of developmental progression from bottom to top is preserved in Mandarin, suggesting shared developmental organization despite differences in overall geometry across languages.

Fig. 4: Latent-space geometry of vocabulary development in Mandarin Chinese



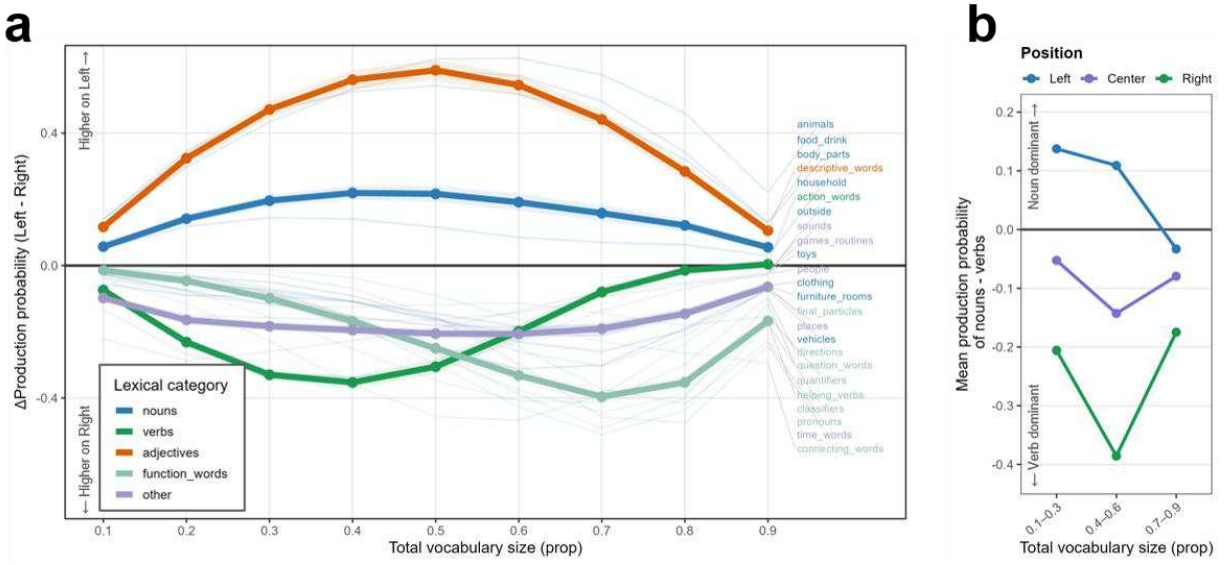
(a) Each point represents one CDI administration (N = 1,056). The data form a flattened, laterally extended manifold resembling a sea-ray-like shape, in contrast to the thicker structure observed in American English. Color indicates total vocabulary size, which increases systematically along a curved trajectory. **(b)** Isosurfaces of total vocabulary size, plotted at 20% intervals (with 4 slices). These are arranged approximately orthogonal to the curved developmental axis, indicating that a coherent axis of developmental progression is likewise preserved in Mandarin. **(c)** Five representative developmental trajectories extracted from the manifold, following the same procedure as in American English. The central trajectory corresponds to the typical developmental path, with four additional trajectories (Left, Right, Back, and Front) sampled at ±2 *SD* along directions of maximal variance.

Using the same procedure as in American English, we extracted two orthogonal axes of maximal variance in the Mandarin latent space (i.e., Left–Right and Back–Front orientations). At each level of total vocabulary size, we sampled the centroid (Center) and positions at ±2 *SD* along

each axis (Left, Right, Back, and Front), yielding five representative developmental trajectories (**Fig. 4c**). We then examined whether the Mandarin latent space could be interpreted analogously to the English latent space. For the Left–Right contrast, we computed the mean difference in production probability (Left minus Right) within each lexical category across total vocabulary levels. As in English, nouns showed higher probabilities on the Left, whereas verbs and function words were more prominent on the Right (**Fig. 5a**). One notable difference concerned adjectives. Whereas adjectives patterned with verbs on the Right in English, they were aligned with nouns on the Left in Mandarin. The contribution of adjectives to the Nominality dimension differed across the two languages.

We confirmed the broader Left–Right contrast at the level of full CDI profiles by examining differences in mean production probability between nouns and verbs across total vocabulary levels, as well as across the Left, Center, and Right positions. Consistent with the item-level analysis, profiles from the Left were noun-dominant, whereas those from the Right were verb-dominant. Profiles from the Center showed a slight Rightward bias, suggesting a verb-biased typical trajectory, in contrast to the noun-biased trajectory in English. This population-level verb bias is consistent with previous reports of a weaker noun bias (or a stronger verb bias) in Mandarin.[8–10,34–37]

Fig. 5: The Left–Right axis captures Nominality in Mandarin vocabulary development



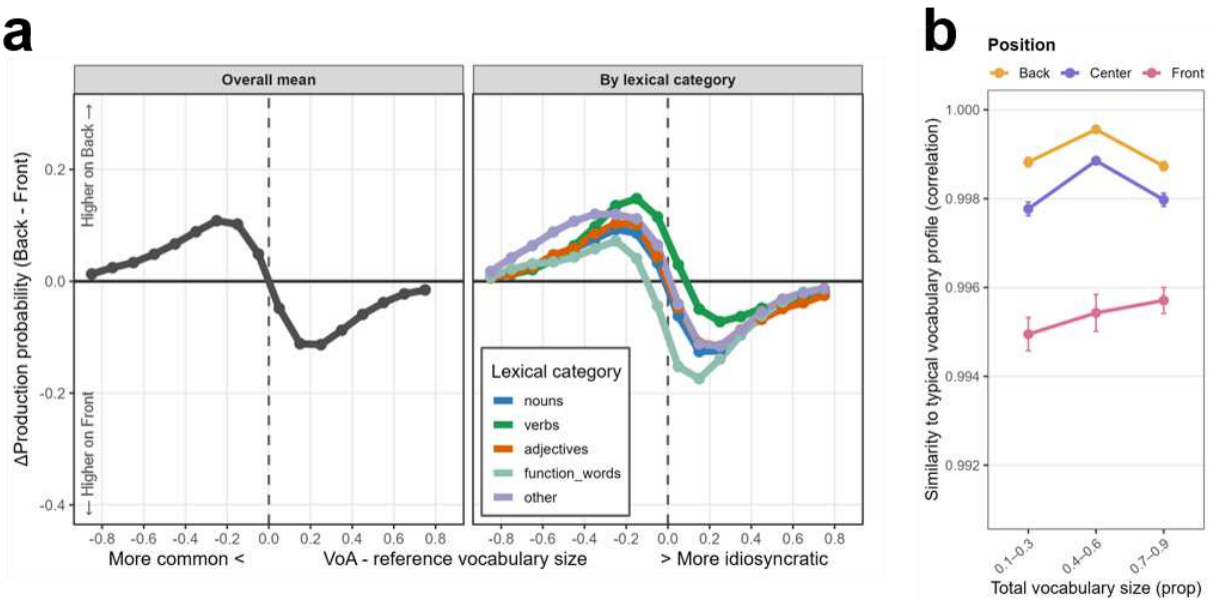
(a) Differences in estimated production probability (Left minus Right), averaged within lexical categories across levels of total vocabulary size. As in English, nouns show higher probabilities on the Left side, whereas verbs and function words are more prominent on the Right side. Notably, adjectives align with nouns on the Left side in Mandarin. **(b)** Mean differences in production probability between nouns and verbs across sampled profiles, summarized for the Left, Center, and Right positions. Profiles from the Left are noun-dominant, whereas those from the Right are verb-dominant. The Center profiles show a slight bias toward the Right, indicating a relatively stronger verb bias compared to English.

We next examined the Back-Front axis using the same procedure as in English. Consistent with the English results, items with lower VoA relative to the reference total vocabulary size showed higher production probabilities on the Back side, whereas items with higher VoA showed higher probabilities on the Front side (**Fig. 6a**). This relationship was generally consistent across

lexical categories, indicating that the Back-Front axis in Mandarin also captures variation from widely shared, early-acquired words to more idiosyncratic, typically later-acquired words. At the level of full CDI profiles, profiles sampled from the Back orientation were more similar to one another, whereas those from the Front orientation were more heterogeneous, with the Center intermediate, as observed in English (**Fig. 6b**).

Despite clear differences in the global latent-space geometry, American English and Mandarin Chinese showed striking similarities in their underlying organization. Three key dimensions were consistently recovered: an overall developmental axis corresponding to vocabulary growth, and individual variation interpretable as Nominality and Commonality. Importantly, within-language variation spanned the population-level contrast traditionally drawn between the two languages. Although English is often characterized as noun-friendly and Mandarin as verb-friendly,[38,39] some English vocabulary profiles extended toward the verb-biased end, whereas some Mandarin profiles extended toward the noun-biased end more characteristic of English than Mandarin. Thus, within-language variation crossed conventional cross-linguistic boundaries in early lexical development.

Fig. 6: The Back-Front axis captures Commonality in Mandarin vocabulary development



(a) Differences in estimated production probability (Back minus Front) plotted against VoA relative to the reference total vocabulary size. The left panel shows the overall pattern across all items, and the right panel shows the pattern across lexical categories. As in English, more commonly acquired words are biased toward the Back, whereas less commonly acquired words are biased toward the Front. **(b)** Similarity between sampled CDI profiles and the corresponding mean profile. As in English, profiles from the Back show higher similarity, whereas those from the Front show lower similarity, with the Center occupying an intermediate position.

Study 3: Tracing within-child trajectories through the developmental landscape

Our final analysis focuses on how development unfolds within individuals over time. Beyond cross-sectional differences across children, we asked whether the same child may follow different developmental tendencies at different stages. To address this, we leverage the

interpretable axes of the latent space to trace longitudinal trajectories of individual children. This approach allows us to characterize within-individual diversity in lexical composition, such as transitions from noun-dominant to verb- or function-word-dominant vocabularies, within the same latent structure used to assess between-child variation.

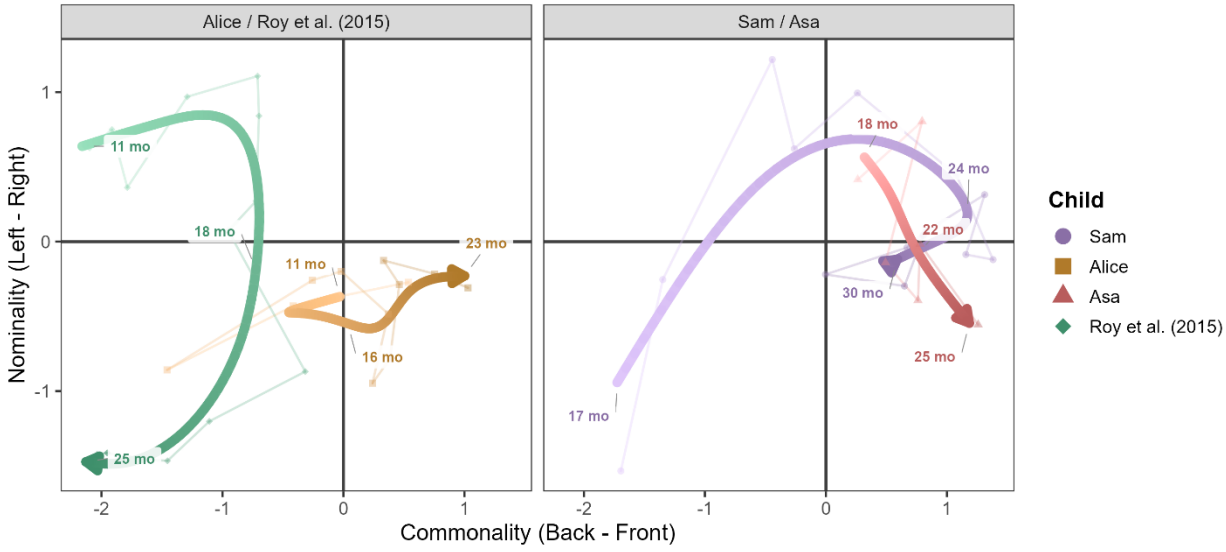
We examined vocabulary development in four English-learning children for whom dense, regularly sampled longitudinal data are publicly available. One child was drawn from the pioneering longitudinal study by Roy et al. (2015)[17] (hereafter, Child R), yielding 15 analyzable time points spanning 11 to 25 months. The remaining three children were drawn from the SAYCam corpus (Sullivan et al., 2021),[33,40] yielding 14 analyzable time points for Sam (17–30 months), 12 for Alice (11–23 months), and 6 for Asa (18–25 months). Unlike the SAYCam data, which included repeated CDI assessments, Child R's records were converted into CDI-like profiles from longitudinal records of each word's first production (see Methods).

We projected these longitudinal data into the latent space using the VAE encoder and quantified each observation's deviation along the Left–Right and Back–Front axes relative to the isosurface corresponding to its total vocabulary size (**Fig. 7**). This allowed us to track how biases in lexical composition, relative to the population-typical profile at a comparable vocabulary size, evolved over time within each individual. We observed clear heterogeneity in developmental trajectories. Sam showed an early bias toward less common verbs, followed by a rapid increase in nouns, accompanied by a shift toward more common vocabulary and a gradual convergence toward the population-typical profile. In contrast, Asa consistently favored more common words, while transitioning from a noun-dominant to a more predicate- and function-word-dominant profile. Alice similarly shifted toward more common vocabulary but maintained a relatively stable verb bias across development.

Child R showed a markedly distinct trajectory, characterized by a persistent tendency toward less common words and a substantial shift from a noun-dominant to a more predicate- and function-word-dominant profile. This pattern may reflect genuine individual differences, such as relatively advanced linguistic structure in early speech, but it may also be influenced by methodological differences in Child R's data (i.e., absence of directly measured CDI data) and potential sampling biases in the recording environment (e.g., recordings limited to the home environment, potentially lacking utterances produced in outdoor contexts). These possibilities warrant cautious interpretation.

Together, these results demonstrate that vocabulary development varies not only across individuals but also within individuals over time. While some children exhibit stable biases in lexical composition, others undergo substantial shifts in their developmental trajectories. This approach thus highlights developmental diversity as a dynamic process within a shared latent structure.

Fig. 7: Within-individual dynamics of vocabulary development



Longitudinal changes in vocabulary composition for four children, expressed in the coordinate system derived from the VAE latent space. CDI profiles for each child and time point were first encoded into the latent space, and their deviations along the Left–Right (*Nominality*) and Back–Front (*Commonality*) axes were quantified within the isosurface corresponding to their total vocabulary size. Deviations were normalized by the standard deviations within each isosurface, and each child's trajectory was smoothed using spline interpolation. The left panel shows Alice and Child R from Roy et al. (2015),[17] and the right panel shows Sam and Asa[33], with trajectories split across two panels for visual clarity. Labels indicate ages in months at the first, middle, and final time points. The trajectories reveal substantial within-individual variation in lexical biases over developmental time.

Discussion

Our findings showed that young children follow shared patterns of language development while taking distinct pathways across individuals and over time. Using a VAE, Study 1 mapped vocabulary development in a large American English sample, Study 2 tested whether the resulting organization generalized to Mandarin Chinese, and Study 3 traced within-child change by situating four densely documented English-learning children within the English latent space. Across Studies 1 and 2, the VAEs self-organized vocabulary profiles along a dominant gradient of overall vocabulary size in both languages, despite being trained without age or lexical category labels. At comparable vocabulary sizes, lexical composition varied along at least two dimensions: Nominality (noun-oriented to verb- and function-word-oriented profiles) and Commonality (widely shared to idiosyncratic words). Study 3 showed that children's lexical composition was not fixed but changed over time as vocabulary grew. Together, these findings offer a comprehensive account of shared developmental regularities and distinctive individual pathways within the same framework, moving beyond how many words children produce to address which words constitute their vocabularies and at which developmental stage.

The Nominality dimension mapped onto well-established grammatical categories in both languages, while the population-typical trajectories revealed cross-linguistic differences. The central trajectory tilted towards nouns (noun bias) in English but towards verbs (verb bias) in Mandarin Chinese, consistent with cross-linguistic work on noun–verb biases.[8–10,35–37] These lexical category biases are likely shaped by linguistic, input-related, and cultural factors,[10,34,41,42] whereas children's everyday experiences with their caregivers and learning environments may further invite variation across individuals within languages.[14,15,18,19] Our findings conveniently

display both cross-linguistic differences and individual-level variation in vocabulary development in the same representational framework.

Adjectives further suggest that Nominality may not be reducible to a simple noun–verb contrast. Whereas English adjectives patterned more closely with verbs, Mandarin adjectives aligned more closely with nouns. Adjectives vary cross-linguistically in their grammatical functions[43-45] and ease of learning during early vocabulary development.[10] Unlike English, Mandarin contains many adjectives functioning as predicatives without a copula[46], potentially facilitating their use, which results in Mandarin-learning children's positive bias towards acquiring predicatives, compared to English-learning counterparts.[9,10,30] Furthermore, cross-linguistic differences in caregiver input may influence children's adjective acquisition,[47] and contrasting adjective use (e.g., *big* versus *small*) has been associated with earlier adjective production.[48] Thus, the observed adjective pattern may result from language-specific grammatical structures and inputs, which go hand in hand with the broader caregiving and cultural practices.

Commonality captures a different form of individual-level variation. At the same level of vocabulary size, some profiles showed higher probabilities of producing words widely shared by most children, whereas others were more likely to produce less commonly shared words. Commonality therefore captures the extent to which children converge on a shared lexical core versus develop more idiosyncratic vocabulary compositions. What may drive variation in Commonality? One possibility lies in differences in children's everyday learning environments. Children's word learning is shaped by language-related factors, such as word frequency, concreteness, and contextual diversity.[30,49–51] Most importantly, the efficacy of word learning also depends on *how* words co-occur with other multimodal inputs, *when* words occur over time, and *where* they are encountered.[14,17,19,52–54] Although many such experiences are likely shared across

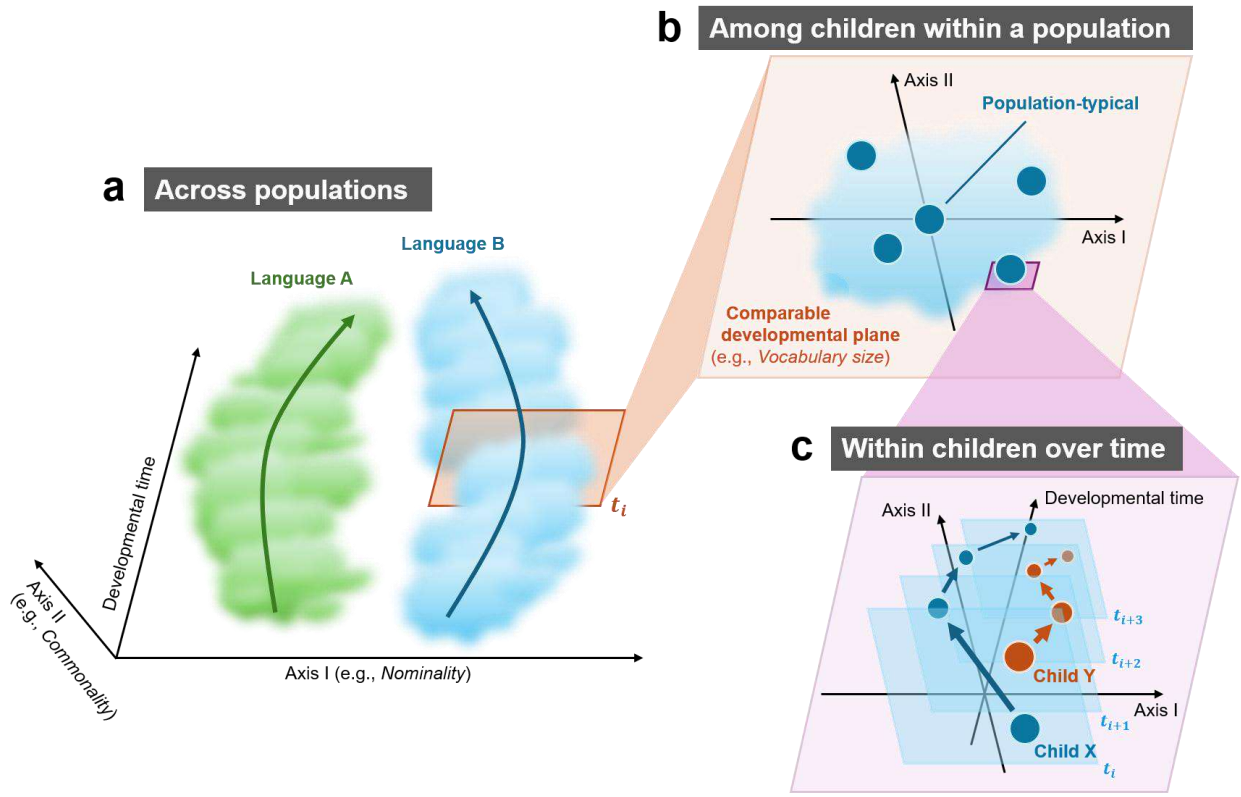
children and may support convergence on a common vocabulary profile, children who experience less widely shared objects, activities, or social interactions may be more likely to learn words that are uncommon in their peers' vocabularies. Children's own interests and intrinsic motivation may further shape which words they preferentially learn[55–57] (but see ref.[58]), as may perceptual and cognitive processes, such as the extent to which visually different exemplars are categorized the same.[59] This study did not aim to tease apart these conjectures, but future work should identify factors that give rise to variation in Commonality.

Whereas the large-scale data mapped the macro-scale developmental landscape, the four longitudinal cases revealed how micro-scale, within-child trajectories unfolded within this landscape. Relative to population-typical profiles at similar vocabulary sizes, some traveled along the Nominality and/or Commonality dimensions over time, whereas others remained stable. Sam's vocabulary profile, for example, shifted from containing more verbs to more nouns and from being idiosyncratic to more common, whereas Asa became more verb- and function-word-oriented while acquiring common words throughout. These patterns caution against categorizing young children into fixed learner types. Canonical accounts have identified distinctive learners' learning styles,[7,60] e.g., referential versus expressive learners, or distinctive pathways,[24] e.g., different steepness of vocabulary growth. Our findings go beyond these by suggesting that learning styles dynamically change over time: the same child switches their learning profile as vocabulary expands. Such within-child variation across time may reveal what's changed in learning environments and everyday experiences.[14,19] Together, the within-child variation conveys key developmental information rather than simple noise around a group average, consistent with dynamic systems approaches.[61,62] Despite the limited number of cases, these findings motivate future work to

examine individual-level variation across time points to determine when and how developmental shifts occur and what factors shape differences across individuals.

Efforts to broaden the generalizability of developmental science have emphasized comparisons across countries, cultures, and linguistic communities.[63–65] Our findings point to a complementary route to probing generalizability. Even within American English, one of the most studied populations in language development, population-level regularities coexist with substantial and structured variation across children and within children across time, much of which remains underexplored. Broadening the range of populations studied is necessary but cross-population comparison is not the only way to probe the generalizability of developmental theories. If each population is compressed to an average, important sources of developmental variation may remain obscured. In this conceptual framework (**Fig. 8**), we propose that early lexical development requires researchers to characterize variation at three nested scales: across populations, among individuals within a population, and across time within an individual. Our latent-space framework offers a convenient way to capture these nested scales, combining both overall developmental progress and variation in what constitutes children's vocabularies.

Fig. 8: Conceptual framework for capturing shared structure and variation in early vocabulary across populations, individuals, and time via nested scales.



Capturing shared structure and variation in early vocabulary can be characterized at three nested scales: across populations, among individuals within a population, and across development time points within an individual. (a) Across populations, overall developmental trajectories can differ across linguistic and/or sociocultural contexts (e.g., English, Mandarin Chinese). (b) Within a population, children with comparable vocabulary sizes can differ in their vocabulary compositions deviating from the population-typical mean (e.g., along the axes of Nominality and Commonality dimensions). (c) Within an individual, their vocabulary compositions can change over developmental time points along the dimensions. These scales capture shared structures (i.e., population-typical trajectory), as well as between-population differences, between-child variation, and within-child variation in early vocabulary development.

The present findings provide a quantitative description of how shared developmental structure coexists with variation across linguistic contexts, children, and time. What drives this pattern? This question echoes a longstanding debate in developmental science more broadly, concerning the interplay of biological constraints, experience, and social interaction, a debate in which language development has played a central role. The finding that Nominality and Commonality recurred across the two languages studied may reflect shared constraints or learning mechanisms across populations, possibly including innate predispositions.[66–68] Simultaneously, variation in how these dimensions are expressed across languages, children, and time may reflect differences in environmental and statistical regularities[19,69,70] and in social interaction and communicative experience.[71–73] Recent child-specific modeling likewise suggests that developmental trajectories may arise from the dynamic interplay between internal learning mechanisms and everyday experiences.[74] We do not adjudicate among these accounts, but provide a foundational framework for examining how these different sources may shape developmental regularity and variability. From this perspective, variability is not merely noise around a shared pattern, but part of the developmental phenomenon to be explained.

Several limitations constrain the generalizability of the conclusions. First, the American English and Mandarin Chinese datasets differed in sample size (9,093 versus 1,056 children) and items (680 versus 799 words). Because separate VAEs were trained on different datasets, differences in latent-space geometry may reflect sampling and measurement differences, alongside population differences. We therefore described each geometry separately rather than directly comparing them. Second, because we used three-dimensional latent spaces, Nominality and

Commonality capture two major dimensions of variation beyond overall vocabulary growth. Additional dimensions may emerge in higher-dimensional representations. Third, most observations used to construct the latent spaces were cross-sectional, so their organization reflects population-level structure rather than directly capturing within-child developmental dynamics. The four cases illustrate how individual children can travel through this interpretable structure along distinct trajectories, but are too few to establish systematic patterns of within-child change. Finally, Nominality and Commonality might not extend to other populations with different grammatical, lexical, or cultural characteristics.

Future work should move beyond describing variation to identifying the processes that generate it. First, large-scale longitudinal samples could track when changes in vocabulary composition emerge, which patterns remain stable, and whether they predict later learning outcomes. Second, linking repeated vocabulary measures with children's naturalistic linguistic, visual, and social environments could test whether shifts along Nominality or Commonality are preceded or accompanied by changes in caregiver inputs, object exposures, or social interactions; dense naturalistic recordings such as BabyView[75] may enable such analyses. Third, latent spaces can be extended to atypically developing populations, such as late talkers,[76,77] to examine whether and when developmental shifts occur similarly to or differently from those in typically developing children.

Together, we offer a novel latent-space framework for examining early vocabulary development that integrates both shared regularities and distinctive individual pathways. Across American English and Mandarin Chinese, the VAE recovered structured dimensions of overall vocabulary progress, Nominality, and Commonality, while revealing systematic differences in lexical composition among children with comparable vocabulary sizes and changes within

individuals over development. By comprehensively capturing variation across populations, among individuals, and within individuals across time within a shared developmental structure, our framework spanning three nested scales provides a foundation for identifying the processes that generate developmental regularity and variation.

Methods

Overview

We used a common VAE-based analytic framework across the three studies. Studies 1 and 2 applied the same modeling and latent-space characterization procedures to large-scale American English (Study 1) and Mandarin Chinese (Study 2) CDI datasets. Study 3 projected dense longitudinal vocabulary records onto the English latent space to trace within-child developmental trajectories.

Datasets

We analyzed vocabulary data from the MacArthur–Bates Communicative Development Inventories (CDIs), a widely used caregiver-report measure of early language development.[13,27] Caregivers report whether their child produces each word on a predefined vocabulary checklist. For Studies 1 and 2, CDI data were obtained from Wordbank,[31,32] an open repository of cross-linguistic CDI data (downloaded April 11, 2024). We focused on the Words and Sentences form, which assesses productive vocabulary.

We analyzed two languages frequently contrasted in the language development literature: American English and Mandarin Chinese.[8–10,35–37] For Study 1, the American English dataset comprised 9,093 administrations from children aged 16–30 months ($M = 23.03$ months, $SD = 4.50$). The vocabulary checklist contained 680 words grouped into 22 semantic categories (e.g., animal

names, action words). For Study 2, the Mandarin Chinese dataset comprised 1,056 administrations from children aged 16–30 months (M = 22.98 months, SD = 4.32) and 799 words grouped into 24 semantic categories. Each vocabulary profile was represented as a binary vector indicating whether each word was reported as produced. These high-dimensional word-production vectors served as the input to the VAE analyses.

For Study 3, to complement these large-scale data with dense longitudinal trajectories within individual children, we analyzed data from one child reported by Roy et al. (2015)[17] (i.e., Child R; 16 time points, 10–25 months) and three children from the SAYCam corpus[33,40] (Sam: 15 time points, 16–30 months; Alice: 12 time points, 11–23 months; Asa: 6 time points, 18–25 months). The earliest observations for Child R (10 months) and Sam (16 months) were excluded from the analysis because their total vocabulary sizes were too small (0–1 words), leaving 15 and 14 analyzable time points, respectively. These datasets were selected because they are accessible for research reuse and provide unusually dense, systematic longitudinal records, allowing us to trace changes in individual vocabulary profiles over time. The SAYCam data included longitudinal CDI assessments alongside day-long naturalistic recordings. Roy et al. (2015)[17], by contrast, derived high-resolution vocabulary records from continuous audio–visual recordings in the child's home environment. Over the first two years of life, 679 unique words were recorded as produced by the child, with the "birth" of each word (i.e., its first production) identified, although CDI data were not directly collected. We therefore constructed CDI-like longitudinal profiles by coding a word as produced at a given age if its first production had occurred by that age. Of the words recorded for this child, 364 could be matched to items on the American English CDI and were used for analysis.

Modeling using a variational autoencoder (VAE)

For Studies 1 and 2, we trained separate variational autoencoders (VAEs)[28,29] for American English and Mandarin Chinese, respectively. A VAE is a generative latent-variable model that learns a low-dimensional probabilistic representation of high-dimensional data through an encoder–decoder architecture. For an input vocabulary profile $\boldsymbol{x}$, the encoder parameterizes an approximate Gaussian posterior distribution $q_{\phi}(\boldsymbol{z}|\boldsymbol{x})$, where $\boldsymbol{z}$ is the latent vector and ϕ denotes the encoder parameters. A latent vector $\boldsymbol{z}$ sampled from this distribution is then passed through the decoder to reconstruct the original vocabulary profile. The model was trained by maximizing the evidence lower bound, which consists of an expected reconstruction log-likelihood minus the Kullback–Leibler divergence between the approximate posterior and a standard normal prior, $p(\boldsymbol{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$.

For each language, the input and output dimensionality corresponded to the number of CDI items: 680 for American English and 799 for Mandarin Chinese. Input vectors comprised binary indicators of word production (produced = 1, not produced = 0). The encoder contained a fully connected hidden layer of 512 units with a ReLU activation, followed by output layers parameterizing the mean and log-variance of a three-dimensional Gaussian latent distribution with diagonal covariance. The decoder contained a 512-unit hidden layer with a ReLU activation and a sigmoid output layer, yielding an estimated production probability for each CDI word. Accordingly, the reconstruction term was defined as the Bernoulli log-likelihood across CDI items. We restricted the latent representation to three dimensions to facilitate direct visualization and geometric characterization of developmental variation, rather than optimizing predictive performance. The nonlinear encoder allows complex structure in the high-dimensional vocabulary data to be represented in a low-dimensional space, while regularization toward a common prior

encourages latent representations to occupy a continuous, structured region. Together with the continuous decoder mapping from latent coordinates to word-production probabilities, this allowed us to probe intermediate locations in the latent space and examine how vocabulary composition changes along and across developmental trajectories. Importantly, the VAE was trained without information about predefined CDI semantic categories, grammatical classes, other linguistic annotations, or children's age (see **Supplementary Fig. S3** for model architecture).

Models were optimized using Adam with a learning rate of .001, a batch size of 128, and 30 training epochs, without early stopping. To reduce sensitivity to random initialization, we trained 1,000 candidate models with different random seeds. For each candidate model, seven-fold cross-validation was performed using a randomly generated data partition that differed across models. Model performance was evaluated using the validation evidence lower bound (ELBO) loss, and the model with the lowest validation loss was selected. Models with similarly low validation errors showed qualitatively similar latent-space structures. To locate each observed vocabulary profile in the latent space, we used the posterior mean $\boldsymbol{\mu}(\boldsymbol{x})$ as its representative coordinate rather than a random draw from the posterior distribution $q_{\phi}(\boldsymbol{z}|\boldsymbol{x})$.

For Study 3, no additional VAE was trained. Instead, each longitudinal English vocabulary profile was encoded using the American English VAE trained in Study 1, and the posterior mean $\boldsymbol{\mu}(\boldsymbol{x})$ was used as its latent coordinate.

Characterizing the latent developmental space (Studies 1 and 2)

To characterize how vocabulary development was organized within the latent space, we evaluated the decoder on a regular grid of 30 x 30 x 30 latent-space coordinates, spanning a region chosen to encompass the observed latent structure in each language. At each coordinate, the

decoder generated a vector of estimated production probabilities for all CDI words. We calculated the expected total vocabulary proportion as the mean of these probabilities and interpolated between grid points to construct isosurfaces corresponding to total vocabulary proportions from 10% to 90% in increments of 10%.

To further characterize variation among vocabulary profiles at comparable levels of overall vocabulary development, we examined observed profiles within a local window around each target isosurface. For each vocabulary-size level, profiles within a sample-size-adjusted range of the target total vocabulary size were included in the analysis (±150 words for American English and ±50 words for Mandarin Chinese). We calculated the centroid of these profiles at each vocabulary-size level and connected the centroids with a smooth curve to define the central developmental trajectory, conceptually representing the most typical developmental path, referred to as the Center trajectory in the Results. We then applied principal component analysis (PCA) to their three-dimensional latent coordinates. The first two principal components captured the major orthogonal directions of variation at each isosurface and defined the two local dimensions, referred to as the Left–Right and Back–Front axes in the Results. The orientations of these axes were aligned across vocabulary levels using the vector from the latent-space origin to the local centroid and the tangent to the central developmental trajectory at that centroid as geometric references. Thus, these directions were derived from the local distribution of vocabulary profiles at each developmental level.

For each isosurface, we defined five representative latent-space locations: the centroid (Center), locations at ±2 *SD* along the Left–Right principal-component axis (Left and Right), and locations at ±2 *SD* along the Back–Front principal-component axis (Back and Front). These

locations served as standardized reference points for sampling vocabulary profiles while holding overall vocabulary size approximately constant.

Sampling from the latent space (Studies 1 and 2)

We generated 100 random samples at each representative location in the latent space. Each representative location was treated as the mean of a Gaussian distribution, with the corresponding local variance estimated by averaging the encoder-derived variances of nearby observed vocabulary profiles in the latent space. Samples were drawn from this distribution to capture local variability around each representative location.

Because random sampling generally shifted the points slightly away from the target total-vocabulary isosurface, each sampled point was projected back onto the target isosurface before decoding. Specifically, for each sampled point, we first identified the isosurface corresponding to its estimated total vocabulary proportion and defined two orthogonal tangent vectors at that location. The direction perpendicular to that surface at the sampled point (i.e., the local surface normal) was then defined as the direction orthogonal to both tangent vectors. We moved the sampled point along this perpendicular direction until it intersected the target isosurface. This local normal projection approximates the nearest-point projection because deviations from the target isosurface were typically small (e.g., approximately 6–11 words on average at a total vocabulary size of 340 words, corresponding to the 50% total-vocabulary level in American English). The projected latent coordinates were then passed through the decoder to obtain estimated production probabilities for all CDI words.

Projecting longitudinal profiles into the latent space (Study 3)

For Study 3, each longitudinal vocabulary profile was encoded into the latent space learned from the American English CDI data in Study 1. Each observation was characterized relative to the local coordinate system at the isosurface corresponding to its total vocabulary size. We quantified its displacement from the Center trajectory along the local Left–Right and Back–Front axes and standardized these deviations by the corresponding within-isosurface standard deviations. This yielded time-varying scores for Nominality and Commonality that were comparable across developmental levels. Successive observations from each child were ordered chronologically to trace within-child trajectories through the developmental landscape. For visualization, trajectories were smoothed using spline interpolation.

References

1. Crais, E., Douglas, D. D. & Campbell, C. C. The intersection of the development of gestures and intentionality. *J. Speech Lang. Hear. Res.* **47**, 678–694 (2004).
2. Fenson, L. et al. Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* **59**, 1–185 (1994).
3. Goldfield, B. A. & Reznick, J. S. Early lexical acquisition: rate, content, and the vocabulary spurt. *J. Child Lang.* **17**, 171–183 (1990).
4. Iverson, J. M. Developing language in a developing body: the relationship between motor development and language development. *J. Child Lang.* **37**, 229–261 (2010).
5. Schneider, R. M., Yurovsky, D. & Frank, M. C. Large-scale investigations of variability in children's first words. in *Proc. 37th Annual Meeting of the Cognitive Science Society* (eds Noelle, D. C. et al.) 2110–2115 (Cognitive Science Society, 2015).

6. Bates, E., Dale, P. S. & Thal, D. J. Individual differences and their implications for theories of language development. In *Handbook of Child Language* (eds Fletcher, P. & MacWhinney, B.) 96–151 (Basil Blackwell, 1995).
7. Nelson, K. Structure and strategy in learning to talk. *Monogr. Soc. Res. Child Dev.* **38**, 1–135 (1973).
8. Tardif, T., Shatz, M. & Naigles, L. Caregiver speech and children's use of nouns versus verbs: a comparison of English, Italian, and Mandarin. *J. Child Lang.* **24**, 535–565 (1997).
9. Frank, M. C., Braginsky, M., Yurovsky, D. & Marchman, V. A. *Variability and Consistency in Early Language Learning: The Wordbank Project* (MIT Press, 2021).
10. Hagihara, H., Tan, A. W. M. & Frank, M. C. Lexical category biases in children's early vocabulary: decomposing linguistic and cultural predictors across 32 languages and dialects. in *Proc. Annual Meeting of the Cognitive Science Society* Vol. 48 (Cognitive Science Society, 2026).
11. Nelson, K. Individual differences in language development: implications for development and language. *Dev. Psychol.* **17**, 170–187 (1981).
12. Lieven, E. V. M., Pine, J. M. & Barnes, H. D. Individual differences in early vocabulary development: redefining the referential–expressive distinction. *J. Child Lang.* **19**, 287–310 (1992).
13. Fenson, L. et al. *MacArthur Communicative Development Inventories: User's Guide and Technical Manual* (Singular Publishing Group, 1993).
14. Tamis-LeMonda, C. S. & Masek, L. R. Embodied and embedded learning: child, caregiver, and context. *Curr. Dir. Psychol. Sci.* **32**, 369–378 (2023).

15. Tamis-LeMonda, C. S. Ecologically grounded observations offer unique insights into the embodied and embedded nature of infant learning. *Hum. Dev.* **69**, 282–301 (2025).
16. Bergelson, E. & Aslin, R. N. Nature and origins of the lexicon in 6-mo-olds. *Proc. Natl Acad. Sci. USA* **114**, 12916–12921 (2017).
17. Roy, B. C., Frank, M. C., DeCamp, P., Miller, M. & Roy, D. Predicting the birth of a spoken word. *Proc. Natl Acad. Sci. USA* **112**, 12663–12668 (2015).
18. Li, Y. & Benitez, V. L. Concurrences across time and sensorimotor capacities promote infant learning. *Child Dev. Perspect.* **19**, 99–107 (2025).
19. Smith, L. B., Jayaraman, S., Clerkin, E. & Yu, C. The developing infant creates a curriculum for statistical learning. *Trends Cogn. Sci.* **22**, 325–336 (2018).
20. Iverson, J. M. Developing language in a developing body, revisited: the cascading effects of motor development on the acquisition of language. *Wiley Interdiscip. Rev. Cogn. Sci.* **13**, e1626 (2022).
21. West, K. L., Fletcher, K. K., Adolph, K. E. & Tamis-LeMonda, C. S. Mothers talk about infants' actions: how verbs correspond to infants' real-time behavior. *Dev. Psychol.* **58**, 405–416 (2022).
22. West, K. L., Saleh, A. N., Adolph, K. E. & Tamis-LeMonda, C. S. "Go, go, go!" Mothers' verbs align with infants' locomotion. *Dev. Sci.* **26**, e13397 (2023).
23. Fernald, A. & Marchman, V. A. Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Dev.* **83**, 203–222 (2012).
24. Gómez Díaz, M., Fibla, L., Tsui, R. K.-Y. & Byers-Heinlein, K. Testing theories of the vocabulary spurt with monolingual and bilingual infants. *Dev. Psychol.* **60**, 1357–1371 (2024).

25. Pinstock, E. & Schramm, S. A. Vocabulary trajectories in German-speaking children from 18 months to three years: a growth mixture model. *J. Child Lang.* **53**, 62–79 (2026).
26. Rowe, M. L. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Dev.* **83**, 1762–1774 (2012).
27. Marchman, V. A., Dale, P. S. & Fenson, L. *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual* 3rd edn (Brookes Publishing, 2023).
28. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. Preprint at https://arxiv.org/abs/1312.6114 (2013).
29. Rezende, D. J., Mohamed, S. & Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. in *Proc. 31st International Conference on Machine Learning* (eds Xing, E. P. & Jebara, T.) Vol. 32, 1278–1286 (PMLR, 2014).
30. Braginsky, M., Yurovsky, D., Marchman, V. A. & Frank, M. C. Consistency and variability in children's word learning across languages. *Open Mind* **3**, 52–67 (2019).
31. Frank, M. C., Braginsky, M., Yurovsky, D. & Marchman, V. A. Wordbank: an open repository for developmental vocabulary data. *J. Child Lang.* **44**, 677–694 (2017).
32. Wordbank: an open database of children's vocabulary development. *Stanford University* https://wordbank.stanford.edu (accessed 1 November 2026).
33. Sullivan, J., Mei, M., Perfors, A., Wojcik, E. & Frank, M. C. SAYCam: a large, longitudinal audiovisual dataset recorded from the infant's perspective. *Open Mind* **5**, 20–29 (2021).
34. Zhang, Y., Kim, S. & Casillas, M. Cross-cultural variation in the noun bias in early vocabulary development: a systematic review. *Cogn. Sci*. **50**, e70241 (2026).
35. Tardif, T. Nouns are not always learned before verbs: evidence from Mandarin speakers' early vocabularies. *Dev. Psychol.* **32**, 492–504 (1996).

36. Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N. & Marchman, V. A. Baby's first 10 words. *Dev. Psychol.* **44**, 929–938 (2008).
37. Hung, Y.-F. & Chang, C. J. Developing early lexical composition in Mandarin-speaking children: a longitudinal study. *J. Child Lang.* **51**, 903–924 (2024).
38. Imai, M., Haryu, E. & Okada, H. Mapping novel nouns and verbs onto dynamic action events: are verb meanings easier to learn than noun meanings for Japanese children? *Child Dev.* **76**, 340–355 (2005).
39. Waxman, S., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K. & Song, H. J. Are nouns learned before verbs? Infants provide insight into a long-standing debate. *Child Dev. Perspect.* **7**, 155–159 (2013).
40. Sullivan, J., Mei, M., Perfors, A., Wojcik, E. & Frank, M. Head cameras on children aged 6 months through 31 months (SAYCam). *Databrary* https://doi.org/10.17910/b7.564 (2019).
41. Gentner, D. Why nouns are learned before verbs: linguistic relativity versus natural partitioning. In *Language Development, Vol. 2: Language, Thought, and Culture* (ed. Kuczaj, S. A. II) 301–334 (Lawrence Erlbaum, 1982).
42. Bornstein, M. H. *et al.* Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Dev.* **75**, 1115–1139 (2004).
43. Dixon, R. M. W. Adjective classes in typological perspective. In *Adjective Classes: A Cross-Linguistic Typology* (eds Dixon, R. M. W. & Aikhenvald, A. Y.) 1–49 (Oxford Univ. Press, 2004).
44. Dryer, M. S. & Haspelmath, M. (eds) *The World Atlas of Language Structures Online*. Zenodo https://doi.org/10.5281/zenodo.13950591 (2013).

45. Skirgård, H. et al. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Sci. Adv.* **9**, eadg6175 (2023).
46. Arcodia, G. F. The Chinese adjective as a word class. In *Word Classes: Nature, Typology and Representations* (eds Simone, R. & Masini, F.) 95–118 (John Benjamins Publishing Company, 2014).
47. Liu, S. X. *Early Vocabulary Development in English, Mandarin, and Cantonese: A Cross-Linguistic Study Based on CHILDES* (Master's thesis, University of Richmond, 2007).
48. Tribushinina, E. et al. The role of explicit contrast in adjective acquisition: a cross-linguistic longitudinal study of adjective production in spontaneous child speech and parental input. *First Lang.* **33**, 594–616 (2013).
49. Hills, T. T. The company that words keep: comparing the statistical structure of child- versus adult-directed language. *J. Child Lang.* **40**, 586–604 (2013).
50. Tan, A. W. M., Loukatou, G., Braginsky, M., Mankewitz, J. & Frank, M. C. Predicting ages of acquisition for children's early vocabulary across 27 languages and dialects. In *Proceedings of the 46th Annual Meeting of the Cognitive Science Society* 564–571 (2024).
51. Flores, A. Z., Montag, J. L. & Willits, J. A. Using known words to learn more words: a distributional model of child vocabulary acquisition. *J. Mem. Lang.* **132**, 104446 (2023).
52. Karmazyn-Raz, H. & Smith, L. B. Sampling statistics are like story creation: a network analysis of parent–toddler exploratory play. *Phil. Trans. R. Soc. B* **378**, 20210351 (2023).
53. Ramirez, F. M., Clerkin, E. M., Crandall, D. J. & Smith, L. B. A solution to generalized learning from small training sets found in infants' repeated visual experiences of individual objects. *Proc. Natl Acad. Sci. USA* **123**, e2536143123 (2026).

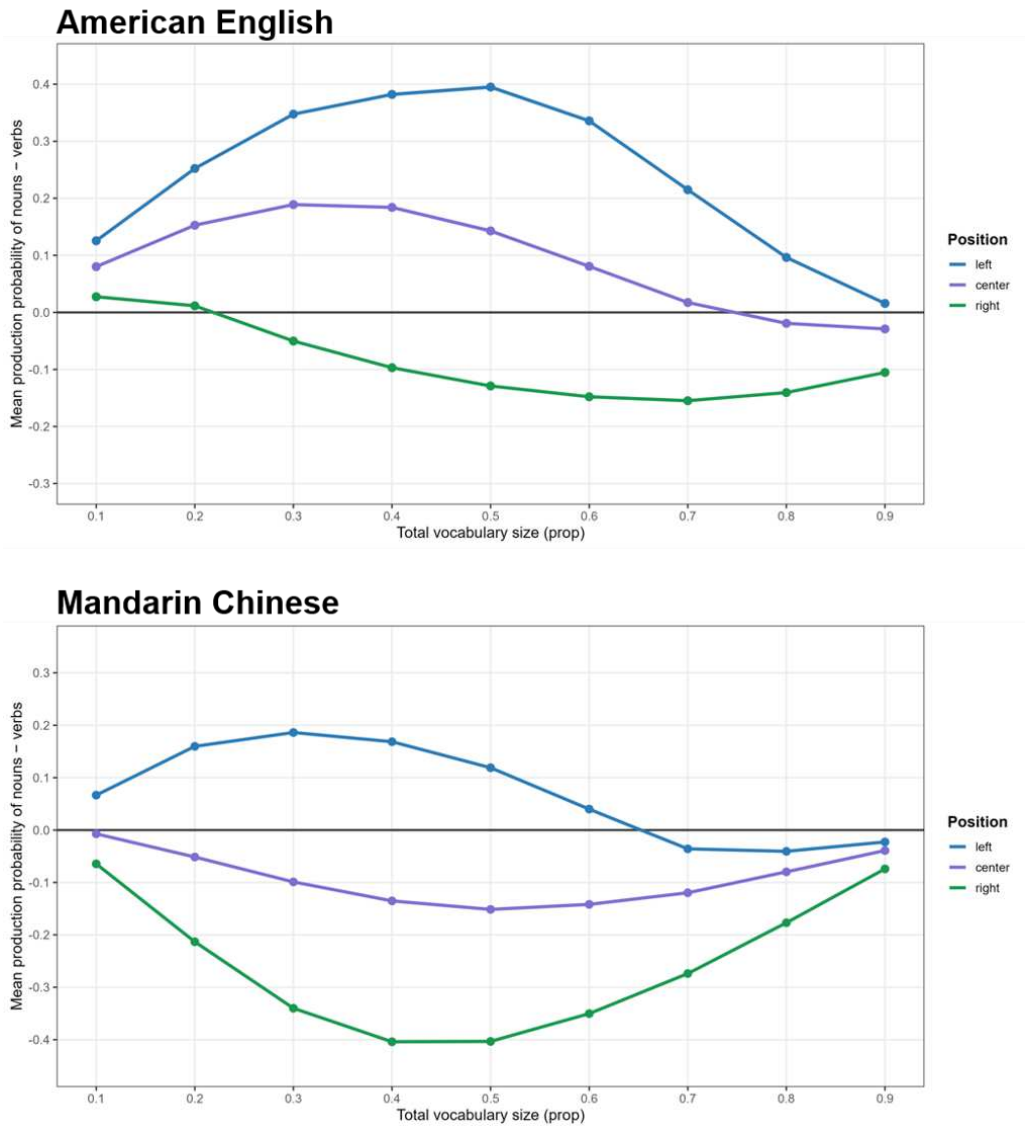
54. Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C. & Smith, L. B. Real-world visual statistics and infants' first-learned object names. *Phil. Trans. R. Soc. B* **372**, 20160055 (2017).
55. Ackermann, L., Hepach, R. & Mani, N. Children learn words easier when they are interested in the category to which the word belongs. *Dev. Sci.* **23**, e12915 (2020).
56. Lucca, K. & Wilbourn, M. P. Communicating to learn: infants' pointing gestures result in optimal learning. *Child Dev.* **89**, 941–960 (2018).
57. Madhavan, R. & Mani, N. Children's individual interests are sustained across development and predict later vocabulary development. *J. Child Lang.* 1–25 (2025).
58. Bazhydai, M., Wong, M. K., Altmann, E. C., Jones, S. D. & Westermann, G. No evidence for curiosity-driven information selection advantage in infants' novel word learning. *Dev. Sci.* **29**, e70101 (2026).
59. D'Souza, D., D'Souza, H., Mayor, J. & Tovar, Á. E. Explaining the comprehension–production vocabulary gap through neural networks and cross-syndrome evidence: insights from Williams syndrome. *Dev. Sci.* **29**, e70115 (2026).
60. Peters, A. M. Language learning strategies: does the whole equal the sum of the parts? *Language* **53**, 560–573 (1977).
61. Thelen, E. & Smith, L. B. *A Dynamic Systems Approach to the Development of Cognition and Action* (MIT Press, 1994).
62. Smith, L. B. & Thelen, E. Development as a dynamic system. *Trends Cogn. Sci.* **7**, 343–348 (2003).
63. Singh, L. A vision for a diverse, inclusive, equitable, and representative developmental science. *Dev. Sci.* **27**, e13548 (2024).

64. Singh, L. et al. Ethical and epistemic costs of a lack of geographical and cultural diversity in developmental science. *Dev. Psychol.* **61**, 1–18 (2025).
65. McMillan, B. T. M. et al. What 5000 babies can tell us about developing minds and how to study them. *Commun. Psychol.* **4**, 92 (2026).
66. Chomsky, N. *Aspects of the Theory of Syntax* (MIT Press, 1965).
67. Pinker, S. *The Language Instinct: How the Mind Creates Language* (William Morrow, 1994).
68. Markman, E. M. Constraints children place on word meanings. *Cogn. Sci.* **14**, 57–77 (1990).
69. Piaget, J. *The Origins of Intelligence in Children* (International Universities Press, 1952).
70. Yu, C. & Smith, L. B. Embodied attention and word learning by toddlers. *Cognition* **125**, 244–262 (2012).
71. Vygotsky, L. S. *Mind in Society: The Development of Higher Psychological Processes* (Harvard Univ. Press, 1978).
72. Tomasello, M. *Constructing a Language: A Usage-Based Theory of Language Acquisition* (Harvard Univ. Press, 2003).
73. Csibra, G. & Gergely, G. Natural pedagogy. *Trends Cogn. Sci.* **13**, 148–153 (2009).
74. Raviv, H. *et al.* The first 1000 days: modeling language learning from individual children's everyday speech. *Science* **394**, eaeh9311 (2026).
75. Long, B. L. *et al.* The BabyView dataset: high-resolution egocentric videos of infants' and young children's everyday experiences. In *Proceedings of the Cognitive Computational Neuroscience Conference* (8-page track) (2025).
76. Jiménez, E. & Hills, T. T. Differences in the semantic structure of the speech experienced by late talkers, late bloomers, and typical talkers. *Dev. Psychol.* **59**, 141–160 (2023).

77. Jiménez, E. & Hills, T. T. Semantic maturation during the comprehension–expression gap in late and typical talkers. *Child Dev.* **93**, 1727–1743 (2022).

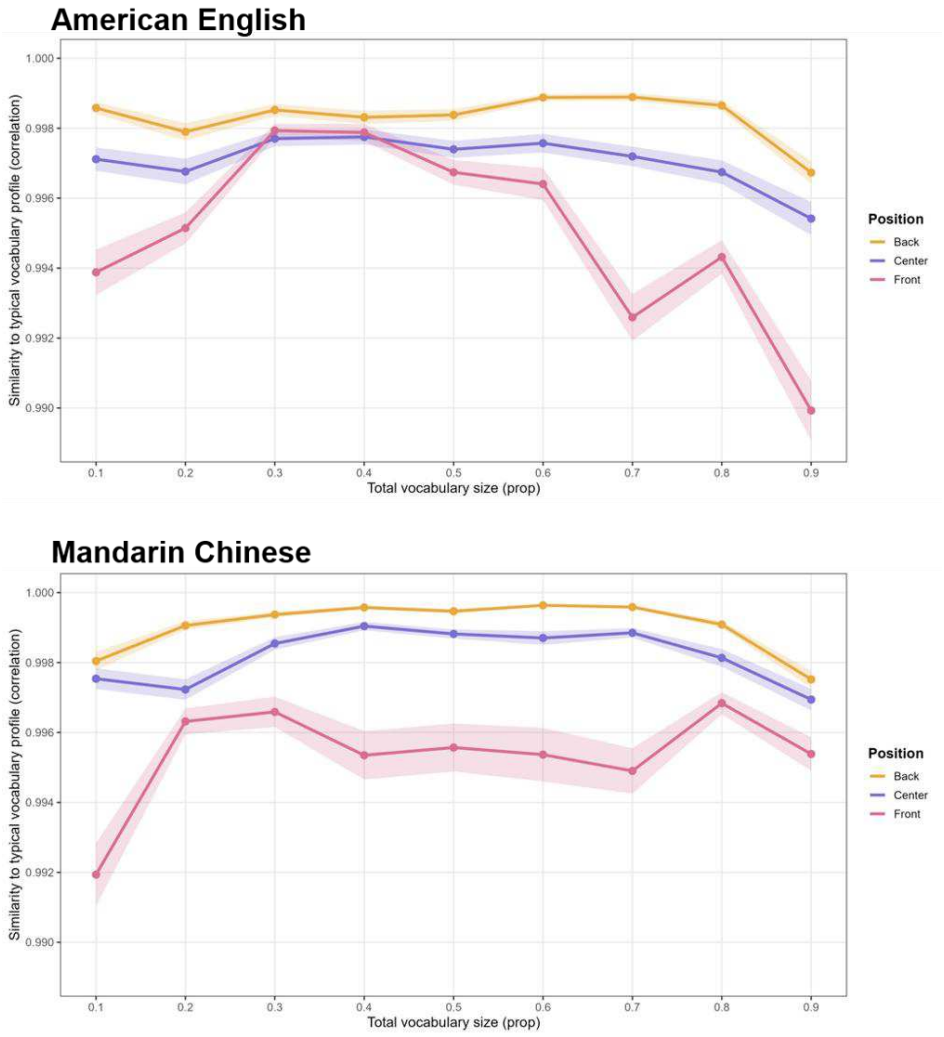
Supplementary Materials

Fig. S1: The Left–Right axis captures Nominality in vocabulary development



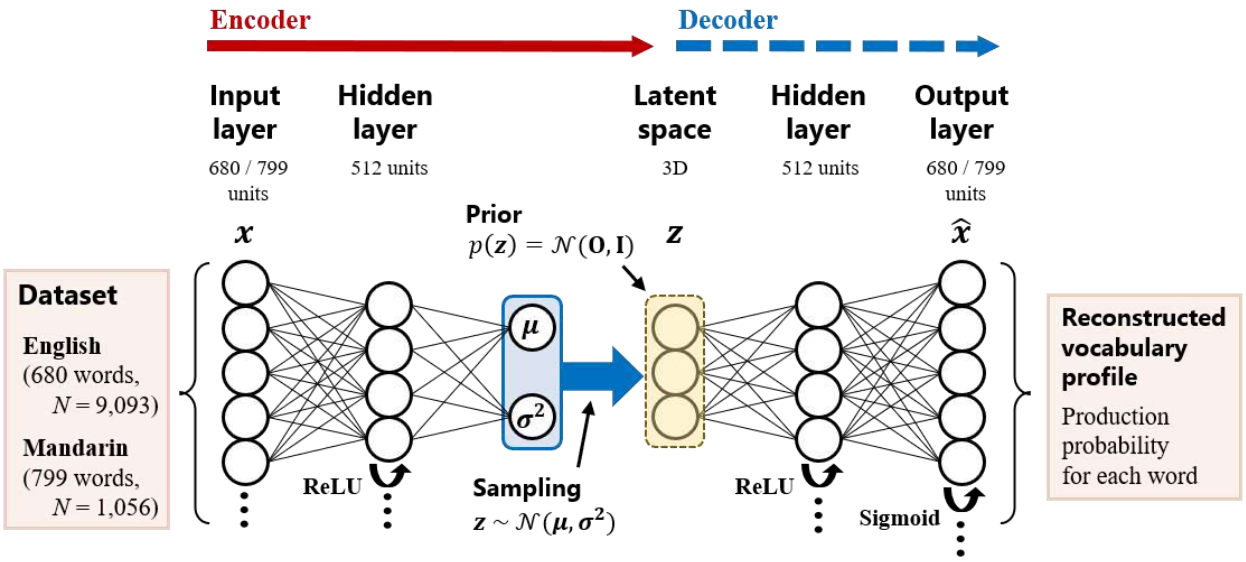
Mean differences in production probability between nouns and verbs across sampled profiles, summarized for the Left, Center, and Right positions. Profiles sampled from the Left are noun-dominant, whereas those from the Right are verb-dominant. The Center profiles show a slight noun bias for American English and a slight verb bias for Mandarin Chinese, suggesting that the typical developmental trajectory is not neutral but moderately noun- or verb-oriented.

Fig. S2: The Back-Front axis captures Commonality in vocabulary development



Similarity between sampled CDI profiles and the corresponding mean profile, computed as the correlation between each profile and the mean profile within each total vocabulary size level and position (Back, Center, Front). Profiles sampled from the Back show higher similarity, indicating more homogeneous vocabulary composition, whereas profiles from the Front show lower similarity, reflecting greater variability across samples.

Fig. S3: Architecture of the variational autoencoder (VAE).



Separate VAEs were trained for the American English (680 words; N = 9,093) and Mandarin Chinese (799 words; N = 1,056) datasets using the same architecture. Each binary vocabulary profile, with one input dimension per word, was encoded through a 512-unit hidden layer into a three-dimensional latent distribution. A latent representation was then sampled and decoded through a 512-unit hidden layer. The output dimensionality matched the input, yielding a reconstructed production probability for each word. ReLU activation was used for the hidden layers and sigmoid activation for the output layer. No information about children's age, predefined semantic categories, grammatical classes, or other linguistic annotations was provided to the models.