UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Lexical category biases in children's early vocabulary: Decomposing linguistic and cultural predictors across 32 languages and dialects

Permalink

https://escholarship.org/uc/item/9vz8m1nd

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Hagihara, Hiromichi
Tan, Alvin Wei Ming
Frank, Michael C.

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Lexical category biases in children's early vocabulary: Decomposing linguistic and cultural predictors across 32 languages and dialects

Hiromichi Hagihara (hiromichi.h.hus@osaka-u.ac.jp)[1,2], Alvin W. M. Tan[3] & Michael C. Frank[3]
[1]Graduate School of Human Sciences, The University of Osaka
[2]International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo
[3]Department of Psychology, Stanford University

Abstract

Children's early vocabulary is often skewed towards nouns and away from verbs, but these biases are thought to vary cross-linguistically. Prior work attempting to explore this variation has often been limited in language coverage, predictors, or bias estimation methods. Drawing on large-scale cross-linguistic resources (Wordbank, WALS, GramBank, and CHILDES), we provide a systematic decomposition of early lexical category biases into typological, frequency-based, and cultural components. We first derive noun, verb, and adjective bias estimates from early productive vocabularies across 32 languages and dialects. We then test whether cross-linguistic variation in these biases can be predicted by grammatical features and part-of-speech frequency in child-directed speech, as well as broader cultural factors. Several of these predictors reliably explain variation in noun and verb bias, whereas adjective bias is comparatively poorly predicted. These findings highlight multiple classes of factors shaping variation in early lexical biases.

Keywords: language development; word learning; vocabulary; syntactic bias; cross-linguistic comparison

Introduction

Children do not acquire words at random. Children's early vocabulary shows systematic regularities, such that concrete, frequent words, and words embedded in specific contexts are learned earlier than others (e.g., Braginsky et al., 2019; Roy et al., 2015; Tan et al., 2024). At the same time, the composition of early vocabularies also shows systematic biases. One well-documented bias concerns syntactic categories. Many languages exhibit a robust noun bias (Bornstein et al., 2004; Frank et al., 2021; Gentner, 1982; Waxman et al., 2013), with object words dominating early vocabularies compared with relation words. In contrast, some languages, such as Mandarin Chinese (Hung & Chang, 2023; Tardif, 1996), Korean (Choi & Gopnik, 1995), and Tseltal (Casillas et al., 2024), show a weaker noun bias or even a verb bias. Cross-linguistic comparisons of lexical biases not only elucidate multiple developmental pathways for language acquisition (Bates et al., 1996; Nelson, 1973; Rowe & Weisleder, 2020) but also provide insight into the ontogenetic origins of linguistic influences on perception and cognition (Hagihara et al., 2026; Imai et al., 2016; Whorf, 1956).

A range of accounts have been proposed to explain how early lexical biases arise (Tan & Frank, 2025). From a cognitive perspective that is a prominent approach in developmental science, nouns are referentially transparent because objects are more easily individuated from their environments than actions or attributes (Fitch et al., 2025; Gentner, 1982), whereas verbs involve more variable ways of lexicalizing relations among objects, and these vary across languages (Gentner & Boroditsky, 2001). Linguistic accounts suggest that morphological complexity or saliency in utterance positions (Caselli et al., 1995; Goldfield, 1993; Kim et al., 2000; Kim et al., 2025; Tardif et al., 1997, 2008), or differences in information required to grasp syntactic frames (Gillette et al., 1999) affect lexical biases. More simply, overall part-of-speech frequencies vary across languages (e.g., Camaioni & Longobardi, 2001), biasing learning toward more frequent categories (Verhagen et al., 2022). Cultural contexts may also influence how children incorporate words into their lexicons. Cultures differ in the priority given to object labeling vs. describing relations among objects (Choi & Gopnik, 1995; Fernald & Morikawa, 1993; Tardif et al., 1999) and in attention to objects or actions (Kuwabara & Smith, 2022; Lavin et al., 2006; Senzaki & Shimizu, 2022; Waxman et al., 2013). Together, these three accounts (i.e., cognitive, linguistic, and cultural) suggest multiple factors that may shape early lexical category biases.

Although a substantial body of research has examined lexical category biases in early vocabulary, the literature remains fragmented, making it difficult to obtain a comprehensive view of how multiple factors affect these biases. Many studies either focus on a restricted set of languages (often biased toward Indo-European languages), with few exceptions (e.g., Bornstein et al., 2004; Braginsky et al., 2019), or examine only a small set of explanatory factors (e.g., Tan & Frank, 2025).

Furthermore, earlier work has often operationalized lexical biases using measures such as raw word counts (e.g., Tardif et al., 1997), category proportions (e.g., Bornstein et al., 2004), or ratios of nouns to the sum of nouns and verbs (e.g., Tardif et al., 1999), all susceptible to an availability problem. When more nouns than verbs are available, a noun-heavy vocabulary may arise even under random learning. Recent approaches address this issue by normalizing biases relative to expected vocabulary size, estimating over- or under-representation relative to chance (Frank et al., 2021; Tan & Frank, 2025). However, studies using such methods often collapse verbs and adjectives into a single predicate category (e.g., Frank et al., 2021), despite cross-linguistic differences in their grammatical status and distribution (e.g., Dixon, 2004). Given that noun learning is traditionally contrasted with verb learning specifically (e.g., Gentner, 1982), separating verbs and adjectives is crucial.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Here, the present study aims to provide a comprehensive, cross-linguistic decomposition of early lexical category biases into typological, frequency-based, and cultural components. The recent availability of large-scale resources, such as Wordbank (Frank et al., 2017), WALS (Dryer & Haspelmath, 2013), GramBank (Skirgård et al., 2023), and CHILDES (MacWhinney, 2000), enables broader language coverage with more principled bias estimation and a richer set of predictors. We first derive noun, verb, and adjective bias estimates from early productive vocabularies across languages and dialects. We then ask whether variation in these biases can be predicted by typological features, frequency in child-directed speech, and broader cultural factors. Together, this approach enables a systematic assessment of the sources of early lexical category biases across diverse languages. The data and code are available at: https://github.com/hagi-hara/lexical-category-biases.

Analysis 1: Estimating lexical category biases

We calculate bias estimates for nouns, verbs, and adjectives and treat them as dependent variables in subsequent analyses.

Early vocabulary data

Children's vocabulary data were obtained from the MacArthur–Bates or Oxford Communicative Development Inventories (CDIs; Hamilton et al., 2000; Marchman et al., 2023), parent-report vocabulary checklists. CDI data were extracted from Wordbank, an open repository of CDIs (Frank et al., 2017, accessed June 18, 2025). We primarily focused on the Words & Sentences form, which measures productive vocabulary of several hundred words. To ensure compatibility with typological data from WALS and GramBank (i.e., open databases that compile cross-linguistic structural features), we included only languages with sufficient typological data coverage, yielding 32 languages and dialects (12−36 months, N = 58−9,986 per language or dialect).

Bias estimation

Following previous work (Braginsky et al., 2019; Frank et al., 2021; Tan & Frank, 2025), CDI items were classified into five categories: nouns, verbs, adjectives, function words, and other, among which we focused on nouns, verbs, and adjectives. For each CDI administration, we computed the proportion of produced items within each category, as well as total vocabulary size as the proportion of all CDI items. At the group level, category proportions were modeled as a function of total vocabulary size using a generalized linear model. When plotted with total vocabulary proportion on the x-axis and category proportion on the y-axis (Figure 1), the diagonal represents chance-level acquisition, and deviations indicate lexical category bias. Biases were then quantified as the signed area between a third-order polynomial fit—constrained to pass through (0, 0) and (1, 1)—and the diagonal, with positive values indicating over-representation and negative values indicating under-representation. A third-order polynomial allows for a more flexible fit than a quadratic model.

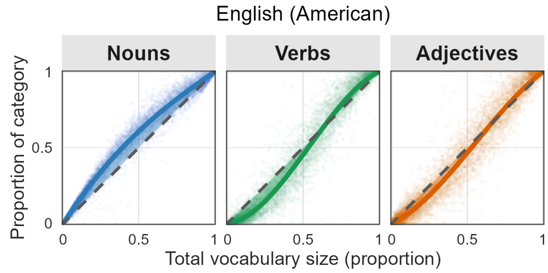

Figure 1: Example of lexical category bias.

Results and discussion

Figure 2 shows the estimated lexical category biases for each language and dialect. Nouns were over-represented in all languages, consistent with previous findings (Frank et al., 2021), except in Kiswahili. Noun bias was generally stronger in Indo-European languages (e.g., German, Danish, and American English), whereas Mandarin, Cantonese, Japanese, and Turkish showed weaker noun biases. Korean exhibited a strong noun bias, diverging from Choi and Gopnik (1995) but aligning with Bornstein et al. (2004). Regional varieties of English, French, and Spanish varied in noun bias strength. This suggests that factors beyond language structure (e.g., culture) may modulate lexical biases.

Verb biases showed a roughly inverse pattern to noun biases, with both over- and under-representation across languages. Separate analyses of each lexical bias revealed that, although noun and verb bias estimates were negatively correlated (r = −.67, p < .001), they are not mutually exclusive: a language with a verb bias can still exhibit a noun bias.

Adjectives showed little over-representation and were often under-represented, consistent with their later acquisition (e.g., Sandhofer & Smith, 2007; Waxman & Booth, 2001). Adjective bias estimates did not significantly correlate with either noun bias (r = −.21, p = .25) or verb bias (r = .17, p = .36). This pattern suggests that verbs and adjectives should be analyzed separately in cross-linguistic research, rather than collapsed into a single category.

Analysis 2: Typological predictors

Using bias estimates for each lexical category, we examine whether cross-linguistic variation in the bias strength can be predicted from morphosyntactic features from WALS (Dryer & Haspelmath, 2013) and GramBank (Skirgård et al., 2023).

Predictor selection and description

Typological predictors were selected based on their potential relevance to early lexical category biases. Candidate features were filtered using three criteria: (a) minimal missing data, (b) sufficient cross-linguistic variability (excluding features where only a single language was exceptional), and (c) simple morphosyntactic properties rather than complex constructions (e.g., *wh*-dependencies). For a small subset of features (4%), values were unavailable in both WALS and GramBank (e.g.,

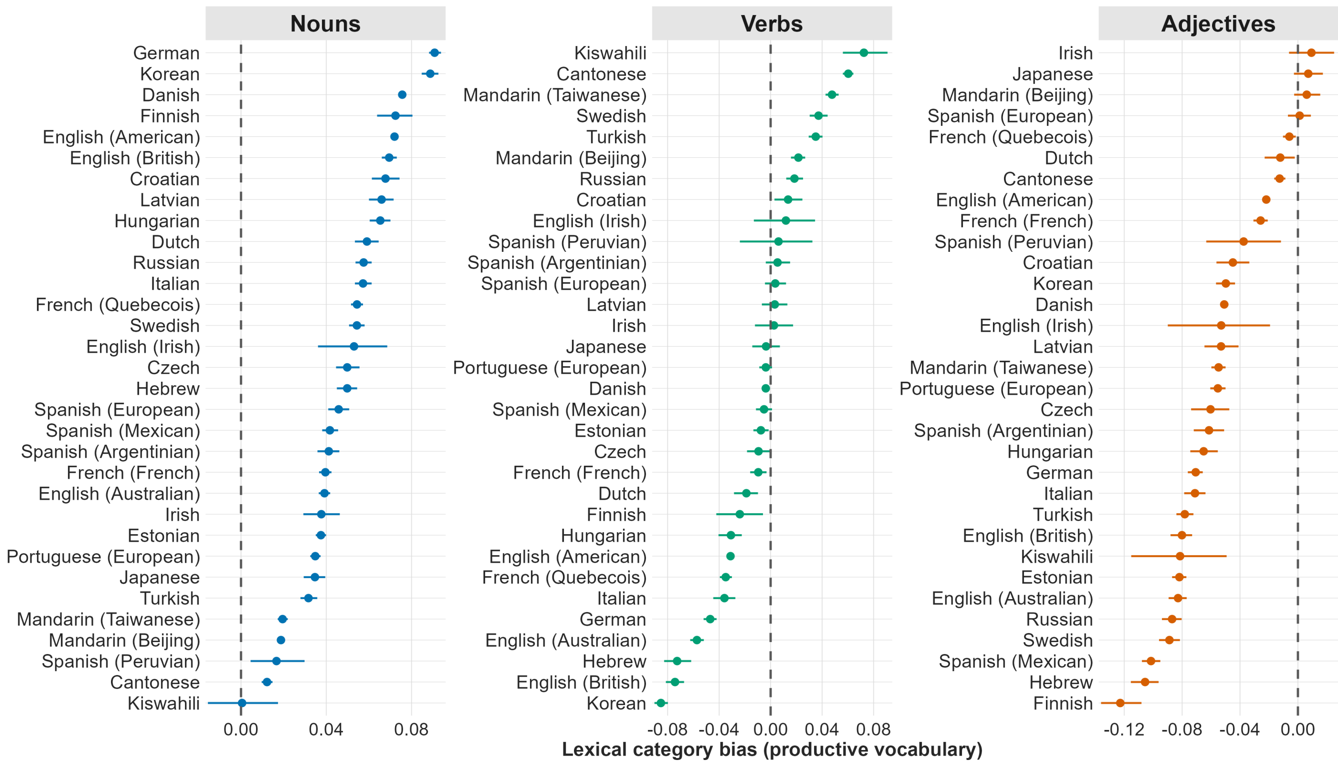

Figure 2: Lexical category biases for each language and dialect. Points show bias estimates, and horizontal lines indicate bootstrapped 95% CIs. Dashed vertical lines represent chance-level. The same conventions apply to subsequent figures.

whether Cantonese permits subject omission) and were filled in from descriptive dictionaries. The final predictor set included 16 features, broadly concerning inflectional morphology, word order, and phrase structure constraints.

Nominal predictors. *Articles* captured the presence of definite and/or indefinite articles (absent; definite only; indefinite only; both). *Case number* described nominal inflection (none; 2–4; ≥5 cases). *Case position* captured how case is realized on noun phrases (absent; suffixal; prepositional clitics; postpositional clitics; mixed). *Plural marking* indexed how plurality is encoded on nouns (absent; prefixal; suffixal; mixed). *Nominal plurality* indicated whether grammatical plural marking is available (absent; present). *Numeral classifiers* indexed whether numerals require classifiers when modifying nouns (absent; present). *Adposition–noun phrase order* captured basic nominal word order (no dominant order; prepositional; postpositional).

Verbal predictors. *Tense marking* indexed whether verbs morphologically distinguish past and/or future tense (none; past-only; both). *Verbal person marking* captured whether verbs index grammatical arguments (none; agent only; agent and patient). *Tense–aspect position* described where tense and aspect are realized on the verb (absent; prefixal; suffixal; mixed). *Verbal suppletion* described whether a language uses distinct verb forms to express tense or aspect contrasts, such as different roots for past vs. non-past (absent; present).

Adjectival predictors. *Comparative marking* indexed whether adjectives have a dedicated morphological comparative marker (absent; present). *Verbal adjectives* captured whether adjectives behave like verbs in predicative position, sharing verbal morphology rather than requiring a separate construction (absent; present).

Cross-category predictors. *Word order* captured dominant subject-verb-object order (SVO; SOV; VSO; no dominant order). *Pro-drop* indexed whether pronominal subjects can be omitted in pragmatically unmarked clauses (absent; present). *Adjective–noun order* captured relative order within noun phrases (adjective–noun; noun–adjective).

Analysis

To test whether typological features predict bias strength, we fit language-level elastic net linear models, separately for nouns, verbs, and adjectives. Given the limited number of languages and potentially correlated predictors, elastic net was used to shrink coefficients and reduce overfitting; tuning selected ridge regression ($\alpha = 0$). We used leave-one-language-out cross-validation for model evaluation, repeatedly fitting the model on all but one language and predicting the held-out language. This procedure provides an out-of-sample check of model stability while using nearly all available data for fitting in each iteration. Uncertainty was quantified with 95% bootstrap confidence intervals (CIs) obtained by resampling languages.

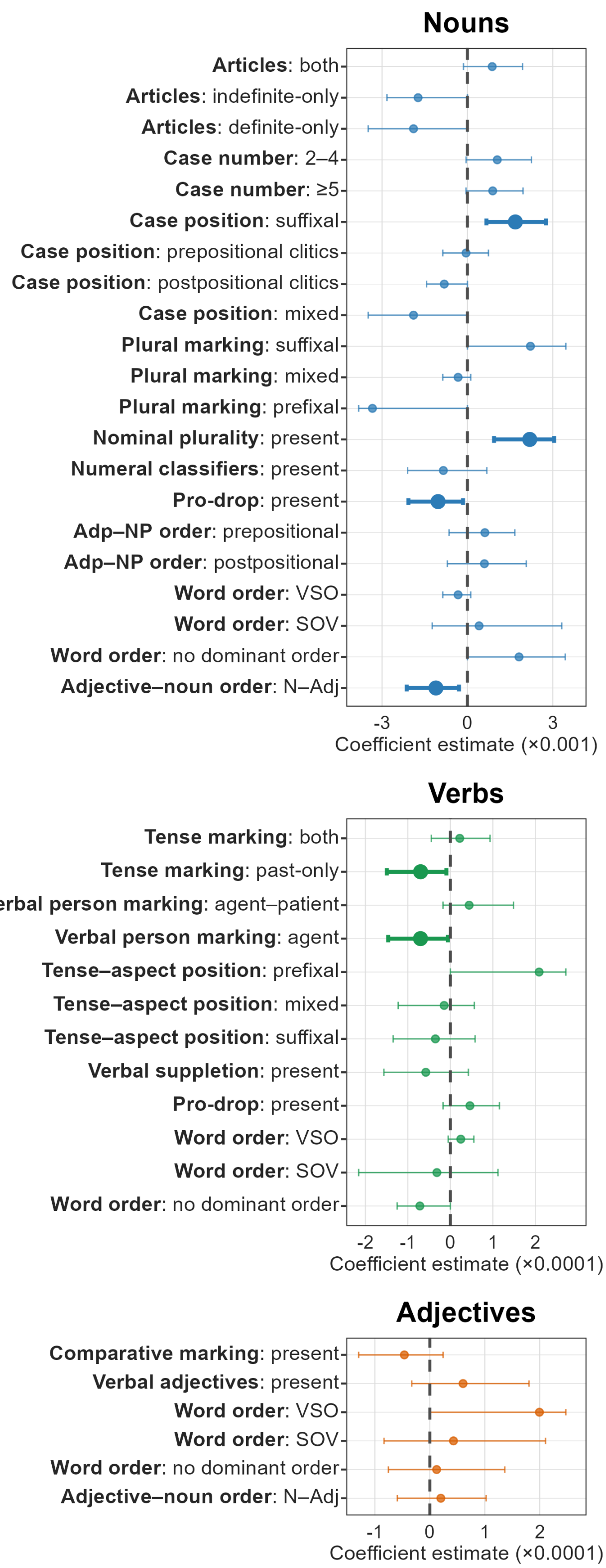

Figure 3: Ridge regression coefficients predicting lexical category biases from typological features. Larger points and thicker bars indicate effects whose CIs exclude zero. Coefficients are shown relative to baseline reference categories (e.g., absent, none, SVO, Adj–N).

Results and discussion

Several typological features predicted cross-linguistic variation in noun or verb bias, but not adjective bias (Figure 3). Suffixal case marking (e.g., Danish, Finnish, Korean) positively predicted noun bias ($b = 0.002$). Relatively stable noun stems with suffixal case marking may enhance the structural cohesion of noun phrases and thereby increase noun salience. Partial variability in word forms in the input has been linked to earlier acquisition (Moore & Bergelson, 2024). This aligns with the negative effect of pro-drop ($b = -0.001$), which can reduce noun salience (Camaioni & Longobardi, 2001; Tardif et al., 1997) (but see Bornstein et al., 2004). However, some pro-drop languages (e.g., Korean, Turkish) also have noun inflection via suffixal case marking, leaving the link between case marking and noun bias unresolved.

Nominal plurality positively predicted noun bias ($b = 0.002$). This may reflect a language's bias toward objects, consistent with previous work on noun learning (Imai & Mazuka, 2007; Yoshida & Smith, 2003). However, languages without obligatory plurality marking were only Japanese, Mandarin, and Cantonese, so this effect should be interpreted cautiously, as it may reflect correlations with other properties.

Other noun-related function words (e.g., articles) did not predict noun bias, suggesting that noun bias may depend on specific morphology bound to nouns, beyond simple co-occurrence with function words that can aid noun learning (Barbir et al., 2023). Word order showed no robust effect, although noun–adjective order (e.g., French, Hebrew, Spanish) predicted weaker noun bias ($b = -0.001$).

Among verbal predictors, past tense marking and verbal agent marking (bs $= -0.00007$) were negatively associated with verb bias. Given the very small effect sizes, these should be interpreted cautiously. Notably, only Mandarin and Cantonese lack past tense marking. A non-significant but positive effect of prefixal tense–aspect marking was driven solely by Kiswahili, again suggesting limited generalizability.

No typological features significantly predicted adjective bias. This result suggests that adjective acquisition may be more strongly shaped by other influences, such as cultural or pragmatic factors.

Analysis 3: Frequency-based predictors

This analysis examines whether overall part-of-speech frequency predicts lexical category bias across languages.

Predictor processing and analysis

Following Tan et al. (2024), we used child-directed speech corpora from CHILDES (MacWhinney, 2000), lemmatized using UDPipe (Straka, 2018), to compute item-level lemma frequencies. Lemma frequencies were Laplace-smoothed and log-transformed, then averaged within each lexical category to obtain category-level predictors. Analyses were restricted to 30 languages and dialects, excluding Kiswahili and Latvian due to data unavailability. For each category, we fit a linear model predicting lexical category bias from standardized

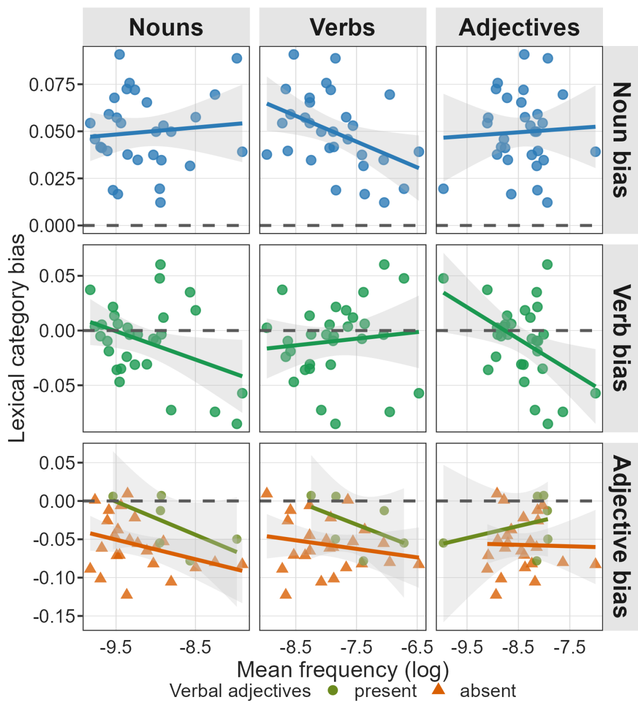

Figure 4: Relationships between mean category frequency in child-directed speech and lexical category bias. Adjective trends are shown separately by verbal adjective status.

overall frequency measures of nouns, verbs, and adjectives[1].

Results and discussion

As reported in prior work (e.g., Tardif et al., 1997), lexical category bias was related to overall part-of-speech frequency in child-directed speech (Figure 4). Noun bias was positively predicted by noun frequency (b = 0.56, p = .021) and more strongly by a negative association with verb frequency (b = −0.72, p < .001). This suggests that noun bias reflects not only noun frequency itself but also relative frequency differences between nouns and verbs. For verb bias, only verb frequency showed a positive effect (b = 0.45, p = .030). However, noun frequency did not reliably predict verb bias (b = −0.46, p = .057), suggesting an asymmetric relationship between noun and verb biases[2].

For adjective bias, neither noun frequency (b = −0.51, p = .061) nor verb frequency (b = 0.068, p = .76) was predictive. However, given that verb bias showed clear frequency effects, we explored whether adjective bias might similarly depend on whether adjectives behave in a verb-like manner. Adding the verbal adjective term as a fixed effect improved model fit (ΔAIC = 6.81; ΔR^2_{adj} = .21). Languages with verbal adjectives (Cantonese, Japanese, Korean, Mandarin, Turkish) showed overall less negative adjective bias than those without (b = 1.23, p = .007). Adjective bias was positively associated with adjective frequency (b = 0.45, p = .031) and negatively associated with noun frequency (b = −0.69, p = .007). These results suggest that adjective bias may reflect both the syntactic availability of adjectives as independent predicates and relative frequency contrasts with nouns.

Analysis 4: Cultural predictors

Our final analysis explored whether cultural orientations predict variation in lexical category bias. Possible mechanisms include cross-cultural differences in attention to focal objects vs. relations/background (Masuda & Nisbett, 2001), children's attention (Heise et al., 2025), and caregiver–child interaction styles (Casillas et al., 2020). More analytic, object-focused cognition may favor nouns, whereas more holistic, relation-focused cognition may favor verbs. However, large-scale cross-national measures of cognitive style remain limited. We thus used Individualism–Collectivism (IDV–COL) (Minkov & Kaasa, 2022) as a proxy, following work linking individualistic (collectivistic) cultures to relatively analytic (holistic) cognition (Nisbett et al., 2001), although these claims remain debated (e.g., Wong et al., 2021). We also included Flexibility–Monumentalism (FLX–MON) exploratorily from the same source given its broad data coverage.

Predictor processing and analysis

IDV–COL contrasts individual freedom and autonomy with collective norms and in-group loyalty. This is among the most widely used dimensions in cross-cultural research (Komisarof & Akaliyski, 2025). Higher values indicate stronger individualism. FLX–MON contrasts adaptive, context-sensitive selves with more stable, consistency-oriented selves, with higher values indicating greater flexibility. Scores were available for 31 countries/regions, with Latvian unavailable. For each lexical category, we fitted linear models predicting bias, including standardized fixed effects for IDV–COL and FLX–MON[3].

Results and discussion

Figure 5 shows the relationships between lexical category bias and cultural measures. IDV–COL positively predicted noun bias (b = 0.01, p < .001) and negatively predicted verb bias (b = −0.02, p = .008). Adjective bias was not explained by this factor (b = −0.00, p = .70). FLX–MON did not significantly predict bias for any lexical category (ps > .080). The pattern for IDV–COL is consistent with the idea that more object-focused cognition, correlated with individualism, favors nouns, whereas more relation-focused cognition favors

[1] We also computed the mean length of utterance in words (MLU-w) from CHILDES. For each item, MLU-w was calculated over utterances in which the word appeared, serving as a proxy for syntactic complexity. However, category-level MLU-w did not predict lexical category bias (ps > .24). Prior work has linked caregiver utterance length to children's word learning (Elmlinger et al., 2025; Roy et al., 2009; Tan et al., 2024), but this effect may relate to overall vocabulary growth rather than the formation of lexical category biases.

[2] Noun frequency was positively correlated with both verb (r = .58, p < .001) and adjective frequency (r = .58, p < .001), whereas verb and adjective frequencies were not (r = .25, p = .18).

[3] IDV–COL was positively correlated with FLX–MON (r = .42, p = .017).

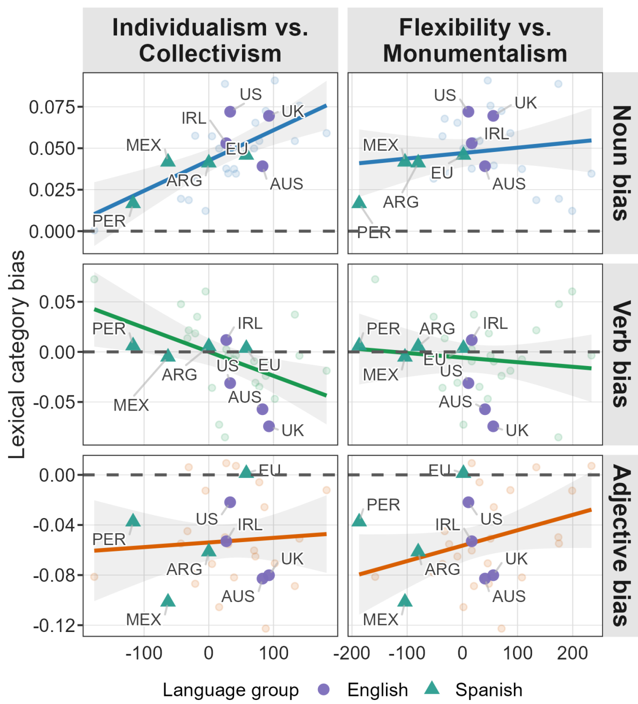

Figure 5: Relationships between cultural predictors and lexical category bias. English and Spanish varieties are highlighted as illustrative cases. English: US (American), AUS (Australian), UK (British), IRL (Irish); Spanish: EU (European), ARG (Argentinian), MEX (Mexican), PER (Peruvian).

verbs. Cultural values influence caregiving practices and developmental pathways (Keller, 2016), and this may shape linguistic input and attentional guidance.

Regional varieties within the same language further suggest that bias variation is partly aligned with cultural measures. For example, Spanish varieties cluster along the regression line linking noun bias to IDV–COL. This supports the view that lexical biases reflect not only linguistic structure but also the cultural contexts in which languages are learned, consistent with recent analyses of bilingualism (Tan & Frank, 2025).

General discussion

Young children often begin by learning more nouns than verbs, but this finding is not universal across languages (e.g., Frank et al., 2021). Both the presence of this lexical category bias and its variation across languages offer the possibility of shedding light on the underlying mechanisms of early vocabulary growth. This opportunity has been blocked, however, due to the lack of systematic measurement of these biases. We took advantage of new coordinated data resources to quantify lexical category biases and use them as a target for predictive modeling based on typological, frequency, and cultural predictors. All three types of predictors showed strong relations with category biases, particularly noun and/or verb bias, in the subset of languages for which they were available.

The frequency effects we observed suggest that some differences in lexical bias across languages may simply be due to differences in patterns of usage across languages. Given that children's acquisition is sensitive to word frequency (Frank et al., 2021; Goodman et al., 2008), if one lexical category is overall more frequent in a particular language compared with another, then it may be learned faster for that reason alone.

On the other hand, our typological and cultural findings suggest that frequency may not be the full explanation. Across both nouns and verbs, the typological features that related to lexical category bias strength appeared to be those that would lead to greater prominence for that category. For example, for nouns both plurality and case marking (which likely increase prominence) were positively associated, whereas pro-drop (which decreases both frequency and prominence) was negatively associated. Similarly, greater collectivism, which has often been linked to greater salience of (social) relations over individual entities, was associated with smaller noun biases and more positive verb biases. Overall, these findings suggest that beyond the distribution of words, linguistic structure and cultural context also play a role in which words are acquired earlier—and by extension, which lexical categories are over- or under-represented[4].

Although this study substantially expands the number of languages and predictors examined, several limitations remain. First, it is observational in nature and considers many interrelated predictors across a relatively small language sample, with typological and cultural predictors often correlated. Future work should broaden data coverage (e.g., CDI, CHILDES) across diverse language families (Kidd & Garcia, 2022; Scaff et al., 2025). Second, we were unable to quantitatively operationalize cognitive predictors (e.g., Gentner, 1982). Given that verbs, like nouns, vary in how easily they evoke mental images (Ma et al., 2009) and that semantic granularity differs across languages (Hagihara et al., 2026) such factors may also influence verb bias. Finally, tests of cultural hypotheses were limited by the availability of cross-national measures; access to a broader range of indicators, such as those capturing tendencies to verbalize emotions, may help constrain accounts of adjective bias in future work.

Language learning has both striking commonalities across cultures and languages as well as areas of substantial variation. Lexical category bias is an area in which variation has been posited to relate to fundamental aspects of children's learning abilities. Yet to date quantitative evidence on the nature of this variation has been limited. Our hope is that, by making precise measurements of cross-linguistic biases, they can be a target for future explanatory modeling. More broadly, we hope that our work contributes to efforts to characterize and explain both variability and consistency in early language development.

[4] A follow-up analysis explored whether detected typological/cultural predictors were associated with lexical-category frequency. Most were not, except that adjective–noun order predicted noun frequency (b = 0.46, p = .035), and individualism was negatively correlated with verb frequency ($r = -.36$, $p < .001$). For these, simple frequency may offer a proximal explanation. Future work should compare the relative contributions of these predictor classes.

Acknowledgements

We are grateful to the researchers who made their data publicly available, and to those who have advocated for open science, without whom this work would not have been made. We also thank the research participants and their families for providing the data. We appreciate the assistance of Nanako Kimura, Haruna Yamagishi, and Weitung Lee for their support in data handling, including data organization and coding. Finally, we thank Sho Tsuji for connecting the authors and facilitating this collaboration.

References

Barbir, M., Babineau, M. J., Fiévet, A.-C., & Christophe, A. (2023). Rapid infant learning of syntactic–semantic links. *Proceedings of the National Academy of Sciences*, *120*(1). https://doi.org/10.1073/pnas.2209153119

Bates, E., Dale, P. S., & Thal, D. (1996). Individual differences and their implications for theories of language development. https://doi.org/10.1111/b.9780631203124.1996.00005.x

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., Pêcheux, M.-G., Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, *75*(4), 1115–1139. https://doi.org/10.1111/j.1467-8624.2004.00729.x

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Camaioni, L., & Longobardi, E. (2001). Noun versus verb emphasis in Italian mother-to-child speech. *Journal of Child Language*, *28*(3), 773–785. https://doi.org/10.1017/s0305000901004846

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10*(2), 159–199. https://doi.org/10.1016/0885-2014(95)90008-x

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a tseltal mayan village. *Child Development*, *91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

Casillas, M., Foushee, R., Méndez Girón, J., Polian, G., & Brown, P. (2024). Little evidence for a noun bias in Tseltal spontaneous speech. *First Language*, *44*(6), 600–628. https://doi.org/10.1177/01427237231216571

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language*, *22*(3), 497–529. https://doi.org/10.1017/s0305000900009934

Dixon, R. M. W. (2004). Adjective classes in typological perspective. In R. M. W. Dixon & A. Y. Aikhenvald (Eds.), *Adjective classes* (pp. 1–49). Oxford University Press. https://doi.org/10.1093/oso/9780199270934.003.0001

Dryer, M., & Haspelmath, M. (2013). The World Atlas of Language Structures. https://doi.org/10.5281/zenodo.13950591

Elmlinger, S. L., Levy, J. A., & Goldstein, M. H. (2025). Immature vocalizations elicit simplified adult speech across multiple languages. *Current Biology*, *35*(4), 871–881.e3. https://doi.org/10.1016/j.cub.2024.12.052

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, *64*(3), 637. https://doi.org/10.2307/1131208

Fitch, A., Lieberman, A. M., Frank, M. C., Brough, J., Valleau, M., & Arunachalam, S. (2025). Referential transparency of verbs in child-directed input by Japanese and American caregivers. *Journal of Child Language*, *52*(6), 1367–1382. https://doi.org/10.1017/s0305000924000382

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/s0305000916000209

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning the wordbank project*. MIT Press.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), *Language development, Vol. 2: Language, thought, and culture* (pp. 301–334). Lawrence Erlbaum.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge University Press.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176. https://doi.org/10.1016/s0010-0277(99)00036-0

Goldfield, B. A. (1993). Noun bias in maternal speech to one-year-olds. *Journal of Child Language*, *20*(1), 85–99. https://doi.org/10.1017/s0305000900009132

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

Hagihara, H., Barbir, M., Yoshida, H., & Smith, L. B. (2026). Rethinking the origins of cross-language effects: How heard verbs influence Japanese- and English-speaking children's attention to the details of actions. *Developmental Science*, *29*(2). https://doi.org/10.1111/desc.70138

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a british communicative development inventory. *Journal of Child Language*, *27*(3), 689–705. https://doi.org/10.1017/s0305000900004414

Heise, M. J., Meristo, M., Ueno, M., Itakura, S., & Carlson, S. M. (2025). Cultural differences in visual attention emerge in infancy. *Infancy*, *30*(1), e12651. https://doi.org/10.1111/infa.12651

Hung, Y.-F., & Chang, C.-J. (2023). Developing early lexical composition in mandarin-speaking children: A longitudinal study. *Journal of Child Language*, *51*(4), 903–924. https://doi.org/10.1017/s0305000922000654

Imai, M., Kanero, J., & Masuda, T. (2016). The relation between language, culture, and thought. *Current Opinion in Psychology*, *8*, 70–77. https://doi.org/10.1016/j.copsyc.2015.10.011

Imai, M., & Mazuka, R. (2007). Language-relative construal of individuation constrained by universal ontology: Revisiting language universals and linguistic relativity. *Cognitive Science*, *31*(3), 385–413. https://doi.org/10.1080/15326900701326436

Keller, H. (2016). Psychological autonomy and hierarchical relatedness as organizers of developmental pathways. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150070. https://doi.org/10.1098/rstb.2015.0070

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, *42*(6), 703–735. https://doi.org/10.1177/01427237211066405

Kim, M., McGregor, K. K., & Thompson, C. K. (2000). Early lexical development in English- and Korean-speaking children: Language-general and language-specific patterns. *Journal of Child Language*, *27*(2), 225–254. https://doi.org/10.1017/s0305000900004104

Kim, M., Mun, J., Kim, G., Jeon, S., & Choi, Y. (2025). Systematic reexamination of early verb dominance: Verbal and nonverbal characteristics of caregiver input and their contribution to long-term language outcomes. *Developmental Psychology*. https://doi.org/10.1037/dev0002089

Komisarof, A., & Akaliyski, P. (2025). New developments in hofstede's individualism-collectivism: A guide for scholars, educators, trainers, and other practitioners. *International Journal of Intercultural Relations*, *107*, 102200. https://doi.org/10.1016/j.ijintrel.2025.102200

Kuwabara, M., & Smith, L. B. (2022). Focus on one or more? cultural similarities and differences in how parents talk about social events to preschool children. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.778960

Lavin, T. A., Hall, D. G., & Waxman, S. R. (2006). East and west: A role for culture in the acquisition of nouns and verbs. In K. A. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word* (pp. 525–543). Oxford University PressNew York. https://doi.org/10.1093/acprof:oso/9780195170009.003.0021

Ma, W., Golinkoff, R. M., Hirsh-Pasek, K., McDonough, C., & Tardif, T. (2009). Imageability predicts the age of acquisition of verbs in chinese children. *Journal of Child Language*, *36*(2), 405–423. https://doi.org/10.1017/s0305000908009008

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk.* (3rd). Lawrence Erlbaum Associates.

Marchman, V. A., Dale, P., & Fenson, L. (2023). *MacArthur-Bates Communicative Development Inventories User's Guide and Technical Manual* (3rd edition). Brookes Publishing Co.

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of japanese and americans. *FJournal of Personality and Social Psychology*, *81*(5). https://doi.org/10.1037/0022-3514.81.5.922

Minkov, M., & Kaasa, A. (2022). Do dimensions of culture exist objectively? a validation of the revised minkov-hofstede model of culture with world values survey items and scores for 102 countries. *Journal of International Management*, *28*(4), 100971. https://doi.org/10.1016/j.intman.2022.100971

Moore, C., & Bergelson, E. (2024). Wordform variability in infants' language environment and its effects on early word learning. *Cognition*, *245*, 105694. https://doi.org/10.1016/j.cognition.2023.105694

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38*(1/2), 1. https://doi.org/10.2307/1165788

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological review*, *108*(2), 291. https://psycnet.apa.org/doi/10.1037/0033-295X.108.2.291

Rowe, M. L., & Weisleder, A. (2020). Language development in context. *Annual Review of Developmental Psychology*, *2*(1), 201–223. https://doi.org/10.1146/annurev-devpsych-042220-121816

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668. https://doi.org/10.1073/pnas.1419773112

Roy, B. C., Frank, M. C., & Roy, D. (2009). Exploring word learning in a high-density longitudinal corpus. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

Sandhofer, C., & Smith, L. B. (2007). Learning adjectives in the real world: How learning nouns impedes learning adjectives. *Language Learning and Development*, *3*(3), 233–267. https://doi.org/10.1080/15475440701360465

Scaff, C., Loukatou, G., Cristia, A., & Havron, N. (2025). Demographic biases in naturalistic language recordings in the childes database. *Developmental Science*, *28*(3), e70011. https://doi.org/10.1111/desc.70011

Senzaki, S., & Shimizu, Y. (2022). Different types of focus: Caregiver–child interaction and changes in preschool children's attention in two cultures. *Child Development*, *93*(3), e348–e356. https://doi.org/10.1111/cdev.13731

Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latarche, J. J., Lesage, J., Weber, T., Witzlack-Makarevich, A., Passmore, S., Chira, A., Maurits, L., Dinnage, R., Dunn, M., Reesink, G., Singer, R., Bowern, C., Epps, P., Hill, J., . . . Gray, R. D. (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances*, *9*(16). https://doi.org/10.1126/sciadv.adg6175

Straka, M. (2018). UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task. In D. Zeman & J. Hajič (Eds.), *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies* (pp. 197–207). Association for Computational Linguistics. https://doi.org/10.18653/v1/K18-2020

Tan, A. W. M., & Frank, M. C. (2025). The origins of syntactic category biases: Evidence from early vocabularies of bilingual children. *Proceedings of the 47th Annual Conference of the Cognitive Science Society*.

Tan, A. W. M., Loukatou, G., Braginsky, M., Mankewitz, J., & Frank, M. C. (2024). Predicting ages of acquisition for children's early vocabulary across 27 languages and dialects. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society*.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, *32*(3), 492–504. https://doi.org/10.1037/0012-1649.32.3.492

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929–938. https://doi.org/10.1037/0012-1649.44.4.929

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the "noun bias" in context: A comparison of English and Mandarin. *Child Development*, *70*(3), 620–635. https://doi.org/10.1111/1467-8624.00045

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, *24*(3), 535–565. https://doi.org/10.1017/s030500099700319x

Verhagen, J., van Stiphout, M., & Blom, E. (2022). Determinants of early lexical acquisition: Effects of word- and child-level factors on Dutch children's acquisition of words. *Journal of Child Language*, *49*(6), 1193–1213. https://doi.org/10.1017/s0305000921000635

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*(3), 217–242. https://doi.org/10.1006/cogp.2001.0764

Waxman, S. R., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H.-j. (2013). Are nouns learned before verbs? infants provide insight into a long-standing debate. *Child Development Perspectives*, *7*(3), 155–159. https://doi.org/10.1111/cdep.12032

Whorf, B. L. (1956). Science and linguistics. In J. B. Caroll (Ed.), *Language, thought and reality* (pp. 207–219). MIT Press.

Wong, V. C., Wyer Jr, R. S., Wyer, N. A., & Adaval, R. (2021). Dimensions of holistic thinking: Implications for nonsocial information processing across cultures. *Journal of Experimental Psychology: General*, *150*(12), 2636–2658. https://psycnet.apa.org/doi/10.1037/xge0001060

Yoshida, H., & Smith, L. B. (2003). Shifting ontological boundaries: How Japanese- and English-speaking children generalize names for animals and artifacts. *Developmental Science*, *6*(1), 1–17. https://doi.org/10.1111/1467-7687.00247_1