

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Verbs or Nouns? A cross-linguistic study examining the effect of morphological complexity and input on children's early lexical development

Permalink

<https://escholarship.org/uc/item/7254q3jb>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Hardegger, Selma
Mazara, Jekaterina
KuÅàntay, Aylin C
[et al.](#)

Publication Date

2024

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

Verbs or Nouns? A cross-linguistic study examining the effect of morphological complexity and input on children's early lexical development

Selma Hardegger (selma.hardegger@uzh.ch)

Department of Comparative Language Science, University of Zurich, Switzerland

Jekaterina Mažara (jekaterina.mazara@uzh.ch)

Department of Comparative Language Science & Center for the Interdisciplinary Study of Language Evolution,
University of Zurich, Switzerland

Birgit Hellwig (bhellwig@uni-koeln.de)

Department of Linguistics, University of Cologne, Germany

Aylin Küntay (akuntay@ku.edu.tr)

Department of Psychology, Koç University, Istanbul, Türkiye

Barbara Pfeiler (bpfeiler@cephcis.unam.mx)

Peninsular Center in Humanities and Social Sciences, National Autonomous University of Mexico, Mérida, Mexico

Sabine Stoll (sabine.stoll@uzh.ch)

Department of Comparative Language Science & Center for the Interdisciplinary Study of Language Evolution,
University of Zurich, Switzerland

Abstract

Despite considerable differences in the structures of the world's languages as well as child-rearing practices, children across the world show remarkable similarities in their time frame of early lexical comprehension and production. This congruent development is related to the maturation of cognitive skills like individuation, generalization, symbol recognition, object and event representation. A long-standing assumption, based on several studies and languages (e.g., Gentner 1982, Bornstein et al. 2004), is that these cognitive skills make children universally favor nouns over predicates since the conceptual salience of nouns facilitates concept-label-mapping. In the past four decades numerous studies have addressed this question from a variety of methodological approaches and have taken into account several possible factors responsible for this development, both linguistic and extra-linguistic. A large-scale comparative study investigating the use of nouns and verbs in naturalistic contexts has been missing so far. To test for potentially universal tendencies we examine the use of noun and verb forms in child-surrounding adult language as well as children's production in longitudinal naturalistic data of 10 grammatically maximally diverse languages. We assess the effect of morphological complexity on the children's distribution of verbs and nouns in comparison to that found in surrounding speakers. We show that in languages with complex verb morphology, children exhibit a higher deviation of their noun-to-verb ratio from what they encounter in their input. This deviation gradually diminishes as they become more productive in the use of their target language. Despite language-specific differences in morphological complexity, proficiency in verb morphology serves as a cross-linguistic explanation for the development of children's noun-to-verb ratios. This finding contributes to the understanding of the cognitive mechanisms underlying children's early use of their language system.

Keywords: language development; lexical acquisition; morphological structure; cross-linguistic; longitudinal corpus study

Introduction

Across languages, typically developing children exhibit considerable similarities when they use their first words. This

is related to general developmental milestones that occur towards the end of their first year. However, learning individual words is not enough and children need to generalize parts of speech to become proficient speakers. A long-standing assumption has been that children learn the meanings of nouns more easily than verb meanings (e.g., Macnamara, 1972; Nelson, 1973; Gentner, 1978). This preference is attributed to easier processing assuming that nouns form more "natural" categories and facilitate object-label mapping (e.g., Gentner, 2006; Gentner & Boroditsky, 2001). A wide range of studies including experiments, questionnaires (Fenson et al., 1991), and naturalistic data strengthen the claim of a universal early "noun bias" (e.g., Gentner, 1982; Clark, 1993; Bornstein et al., 2004; Imai et al., 2008; Waxman et al., 2013). However, there are exceptions to this trend. In languages like Korean, Mandarin, and some Mayan languages, a verb preference has been observed in the earliest acquisition phase (e.g. Choi & Gopnik, 1993; Tardif, 1996; Tardif, Shatz, & Naigles, 1997; Brown, 1998; de León, 1999; Casillas, Foushee, Méndez Girón, Polian, & Brown, 2024). So far, it is unresolved whether these different preferences are an artifact of the methods used or whether findings of noun or verb preference in certain languages reflect language-specific features that influence children's early production, as proposed in Stoll, Bickel, Lieven, Banjade, et al. (2012). It is difficult to compare experimental results with findings from naturalistic studies or questionnaires as they tap into different behaviors. Further, even though we know that languages differ with respect to their proportion of nouns and verbs, we do not know how these distributions play out in the speech children encounter and how their own production relates to this ratio. To make cross-linguistic generalizations, we therefore need comparable data sets and the same methods of measuring the

production of the youngest speakers and their surrounding adults.

Here we ask: (i) Do we find a general noun bias in children’s early production (e.g., Imai et al., 2008; Waxman et al., 2013; Childers, Porter, Dolan, Whitehead, & McIntyre, 2020) and if yes, (ii) are there language-specific differences in how strong this bias is as a function of linguistic properties.

Children’s early lexical inventories and use of word forms are shaped by a variety of factors. Careful examination of language-specific effects is therefore crucial to be able to tease apart language-related properties from potentially universal more general cognitive mechanisms. Here we focus on the role of morphology. Research on morphologically more complex languages suggests that the morphology of a language and its acquisition might impact the learning of nouns and verbs. In a study of noun-to-verb ratios in Chintang (Sino-Tibetan, Nepal) which has highly complex, polysynthetic verb form structure and heavily makes use of noun omissions, Stoll, Bickel, Lieven, Banjade, et al. (2012) found that children between 2;0 and 4;0 exhibit a high noun-to-verb ratio, i.e., noun prevalence, even though the language they hear from surrounding adults is heavily verb focused (low noun-to-verb ratio). As the children become more proficient in their use of verb morphology, their relative proportion of verbs gradually increases towards the level of surrounding adults. Stoll, Bickel, Lieven, Banjade, et al. (2012) suggested a general mechanism whereby children’s morphological development in verb forms drives the decrease in their noun-to-verb ratios. Similar results were found by (Taverna & Waxman, 2020) in a study on Wichi. These two studies suggest that morphological complexity overrules factors such as frequency or positional salience in adult language in predicting the development of children’s noun-to-verb ratios.

To assess whether these tendencies are universal, it is important to test this hypothesis in as varied a sample of languages as possible (Stoll & Bickel, 2013). In this study, we use naturalistic longitudinal data of 44 target children and their surrounding adults from 10 typologically maximally diverse linguistic communities from the ACQDIV database (Moran, Schikowski, Pajović, Hysi, & Stoll, 2016). Using longitudinal corpora of everyday interactions allows us to assess children’s unstructured, naturalistic noun and verb production and to compare it to their surrounding speakers in the same context.

We compare noun-to-verb distributions of individual target children to the surrounding speakers of their own language rather than comparing them across languages. This allows us to assess which factors influence children’s development in direct comparison to the production of adult speakers in the same context. To do so, we measure the noun-to-verb ratio in children and model the deviation of their rate from that of surrounding speakers as a function of their age, morphological development, the frequency of nouns and verbs in the input, and language. This approach makes it possible to see whether there are similar trends and regularities across all

languages irrespective of their structural differences (Stoll & Bickel, 2013).

Data

The data are part of *ACQDIV*, a large database of naturalistic longitudinal recordings of children and their ambient language (Moran et al., 2016) in languages selected for maximal typological diversity based on an algorithm described in Stoll & Bickel, 2013. Some features related to verb form complexity are summarized in Table 2. For this study, we use 10 corpora: Chintang (Sino-Tibetan, Stoll et al., 2015), English (Indo-European, Theakston, Lieven, Pine, & Rowland, 2001), Japanese (Japonic, Miyata & Nisisawa, 2012), KuWaru (Finisterre-Huon, Rumsey et al., 2019), Nungon (Chimbu-Wahgi, Sarvasy, 2017), Qaqet (Baining, Hellwig, Dawuda, Frye, & Reetz, 2014), Russian (Indo-European, Stoll & Meyer, 2008), Sesotho (Bantu, Demuth, 2015), Turkish (Turkic, Küntay, Koşbac, & unpublished, n.d.), and Yucatec (Mayan, Pfeiler, n.d.).

All recordings are fully transcribed and morphologically glossed. For the purposes of comparison, speakers above the age of 12 are considered adult speakers and are counted as part of the child-surrounding adult input. The recordings vary in the diversity of contexts they capture due to the different lifestyles of the communities. Some communities, such as Chintang, spend most of their day outdoors and children are surrounded by many different speakers of various age groups, including several other children. Others, like English and Russian, consist mostly of dyadic interactions indoors. While the recording context can influence the use of language, within each recording, adults and children are within the same context and, therefore, their production is contextually as comparable as possible. The number of target children for each language as well as the age range are given in table 1.

Table 1: Number of target children (44 in total) whose production was analyzed for this study and their age range.

Language	No. of children	Age range
Chintang	6	0;7 - 4;6
English	12	1;8 - 3;0
Japanese	7	1;4 - 5;1
KuWaru	1	2;2 - 3;0
Nungon	2	2;3 - 3;2
Qaqet	3	1;11 - 3;6
Russian	5	1;4 - 6;9
Sesotho	3	2;1 - 4;7
Turkish	2	0;11 - 3;0
Yucatec	3	1;5 - 4;8

Table 2: Typological parameters and features related to verb form complexity of the languages in the sample, adapted from Stoll and Bickel (2013).

Language	Canonical V position	Verb synthesis	V agreement	Inflectional compactness
Chintang	V=3	high	some	distributive
English	V=2	low	some	cumulative
Japanese	V=3	medium	none	cumulative
KuWaru	V=3	low	some	separative
Nungon	V=3	low available	some	separative
Qaqet	V=2	low	some	cumulative
Russian	V=2	low	some	cumulative
Sesotho	V=2	low	some	distributive
Turkish	V=3	medium	some	separative
Yucatec	V=1/V=2	low	some	separative

For the purposes of this study, we rely on the annotations of the individual language experts and their native speaker transcribers and informants. Analogous to previous research on the frequency of nouns and verbs in early lexical learning, this means that forms identified in children’s production are annotated in relation to adult grammar and noun and verb morphology. The diversity in morphological complexity can be illustrated by Chintang and English verb morphology. English is highly analytic with very few synthetic verb forms whereas Chintang is polysynthetic with more than 4000 verb forms which are the result of 140 unique grammatical elements that can be combined with a verb stem (Stoll, Mazara, & Bickel, 2017).

Distributions of nouns and verbs

With the exception of English, all examined corpora had more distinct full verb forms than noun forms (see Figure 1). The reason for this diverse distribution in English might be the small number of inflectional forms as well as the comparatively limited argument dropping leading to a richer noun inventory.

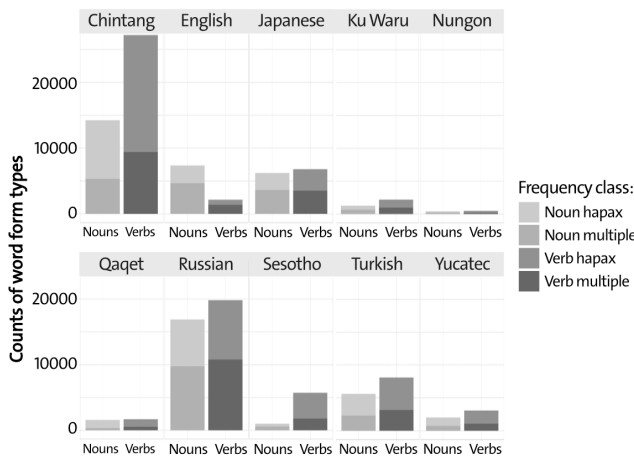


Figure 1: The distribution of noun and verb full form types in adults’ speech. The forms are divided into forms occurring multiple times and single-occurrence forms (“hapax”)

The noun-to-verb ratios observed in adults of the individual

corpora showed a considerable degree of variation. Chintang adults’ showed the lowest rate of nouns as measured in tokens (between 0.2 and 0.4), although their token noun-to-verb rates were somewhat higher. The average trends observed in children are summarized in Table 3 and explored in the regression analyses below.

Table 3: Noun-to-verb ratios in adult and children’s production as measured in tokens.

Language	Adults’ N/(N+V)	Children
Chintang	0.2 - 0.4	0.8 and gradual decrease between 3;0 and 4;0
English	0.6	0.8 to 1 at start, gradual decrease
Japanese	0.4 - 0.6	0.8 and gradual decrease
KuWaru	0.4	0.6 - 0.75 throughout
Nungon	0.4	0.6
Qaqet	0.25	fluctuating
Russian	0.4	varied; 0.8 - 0.9, then decrease and 0.5 for others
Sesotho	0.35	0.4, lower for type counts
Turkish	0.3 - 0.4	0.8 and decrease
Yucatec	0.4	0.8, fluctuating

The semantic content of the most frequently used words in children’s early inventories as well as adults’ production across all corpora show substantial similarities. The most commonly used nouns are kinship terms and proper names, reflecting the attention to people in the interactions recorded in these corpora. The most frequently used verb stems across all age groups were also shared between adults and children. This indicates that changes in the distributions should be attributed to changes in morphological and syntactic abilities rather than simply semantic development.

Modeling effect of morphological development

Model assumptions and variable coding

The outcome variable of our models is the noun-to-verb ratio produced by an individual target child during a particular recording session. For the present analysis, only synthetic noun and verb forms were included. To cover all possible frequency effects, we compute the counts of all synthetic verb and noun forms for tokens, full form (stem+affixes) types, and V- and N-stem types. We distinguish between stem and full form types to capture possible growth in flexibility in using lexical stems and fully inflected forms over time. The noun-to-verb ratios are computed as $N(nouns) / (N(nouns) + N(verbs))$. This avoids division by 0 and constrains the space of possible outcomes to values between 0 (no nouns) and 1 (all nouns). We fit separate models for type (stem and full form) and token frequencies.

Adults’ noun-to-verb ratios are computed based on all adults present during the corresponding recording session.

If children do, in fact, show a general tendency to use more nouns than surrounding adult speakers, this would be an effect observed across all languages for both type and token counts. However, the amount by which children deviate from the distributions found in adults will vary depending on the morphological complexity of their language. If the children

have to learn a more complex system, we expect their proficiency to approach adult levels at a more gradual pace and their deviation from adults' noun-to-verb distributions to both be stronger and decrease more gradually. This reflects the added challenge of processing complex morphology.

To model morphological proficiency, we separately compute the entropy of verb and noun forms used by adults and target children and take the ratio of child-to-adult entropy as a quantification of proficiency. Entropy is an information theoretical measure that has been used in linguistic studies to assess how uniform or skewed a distribution over some units is (Shannon, 1948; Moscoso del Prado Martín, Kostic, & Baayen, 2004). Stoll, Bickel, Lieven, Banjade, et al. (2012) adapted the method for the description of morphological productivity in child language corpora to reflect how evenly forms are distributed across the space of possible stems of verbs and nouns. A skewed distribution generally indicates the tendency to use a limited set of stem+affixes combinations, while a more uniform distribution points to a more flexible use of affixes across multiple stems. Higher entropy indicates lower predictability due to a higher number of possible combinations, while younger children tend to exhibit low entropy because their production consists of fewer distinct stem+affixes combinations. We do not expect adult language to show an even distribution across all possible stems, since linguistic production tends to be characterized by skewed distributions. Therefore maximal entropy is not a good indicator of adult-like performance. Instead, we adopt the ratio of children's entropy divided by adults' entropy as a measure of adult performance approximation per individual recording session. A ratio close to 1 indicates that children's use of verb or noun morphology within a session is similar to that of adults, while a value below 1 indicates lower productivity and proficiency in children.

The comparison to each child's surrounding speech ensures that we are not comparing entropy rates across languages but instead evaluating their development within each language and each individual child. Random effects are introduced for individual children, allowing us to take all data into account, even if some languages include smaller numbers of focal children than others. Additionally, entropy measures can be affected by small sample sizes and might incorrectly reflect the actual underlying frequency distributions. To correct for this, we use the Chao-Shen entropy estimator, which adjusts the computation by increasing the estimated number of unseen types Chao and Shen (2003). This corrects for the underestimation of encountered types due to undersampling through an estimation of the true underlying inventory of types based on the number of low-frequency types in the observed distribution.

Following previous work (Stoll, Bickel, Lieven, Banjade, et al., 2012; Stoll et al., 2017), we treat all affixes, both inflectional and derivational, as part of the affix-bundle, since we cannot reliably claim that children have an understanding of the distinction between the two classes of affixes, nor do

we treat compound verb stems as separate stems but rather as a derivational form of the main stem. This treatment of affixes makes the languages more comparable and is adequate at this level of linguistic development, since it does not assume any dependence on children's knowledge of the parts that make up the verb or noun forms. Therefore, we operate with full verb and noun forms for computing relative entropy measures and modeling proficiency in verb and noun morphology across languages.

We test two hypotheses with regard to the reported noun prevalence in early language development: i) Children learning a language with more complex verb morphology will have a greater deviation from adult entropy levels and show a more gradual approach towards adult-like performance during the early phases of language acquisition. ii) There is a correlation between the change in noun-to-verb ratios and the increase in relative child-to-adult entropy measures. This is based on the assumption that children who have to learn a more complex verb morphology only gradually approach the level of verb production seen in adults, so their noun-to-verb ratios will remain higher for longer. Conversely, in languages with less complex verb morphology, children's development will be less steep or even non-existent because their noun-to-verb ratios are expected to be more similar to those produced by adults. As a matter of control, we also test the relevance of acquiring competence in noun morphology for predicting the development of children's noun-to-verb ratios across languages.

Models

To assess the effect of age, noun-to-verb ratios in input, and morphological proficiency, we fitted a number of generalized linear (beta-binomial) mixed effects models with noun-to-verb ratios as the outcome variable. The models were grouped by the different units of comparison, i.e., for the ratios of full form types, full form tokens, and stem types. The fixed effects are: i) *entropy ratio of noun forms child/adults*, ii) *entropy ratio of verb forms child/adults*, iii) *noun-to-verb ratio in input*, iv) *language*, and v) *age*. We did not find substantial collinearity between *age of target child* and *relative entropy*. To account for individual variation and the different numbers of children in each corpus, we model individual children as random effects.

Models were evaluated in terms of goodness of fit and suitable models were evaluated against each other using the Akaike Information Criterion (AIC) to select the best parsimonious model (Faraway, 2016) starting from the most complex model (including all additive and interactive terms and random effects) and proceeding downwards to simpler ones. Model comparisons were conducted separately for each group of models, i.e., individual models were evaluated for each of the possible outcome variables independently: ratios of full form types, full form tokens, and stem types. The different outcome measures represent different levels of analysis and, therefore, these models cannot, and should not, be compared against each other.

The following four models have the most explanatory power at each of the different levels of analysis:

Model 1: token N-to-V full forms \sim rel. entropy full **verb** forms + age + language + input-N-to-V

Model 2: type N-to-V types \sim rel. entropy full **verb** forms + age + language + input-N-to-V

Model 3: token N-to-V full forms \sim rel. entropy full **noun** forms + age + language + input-N-to-V

Model 4: type N-to-V types \sim rel. entropy full **noun** forms + age + language + input-N-to-V.

All models also include random intercepts and slopes for individual target children to allow for individual variation.

Results: Effect of morphological proficiency

Model 1: Figure 2 shows the relationship between children's token noun-to-verb ratio and their relative entropy of verb forms for individual languages. Overall, child-to-adult entropy ratios negatively correlate with noun-to-verb ratios in children ($\beta = -0.6941[-1.03, -0.36], p = 4.14^{-5}$). This is in line with the main hypothesis which states that with an increase in morphological proficiency of verb forms, children's noun-to-verb ratios decrease towards the level of adults. Age also has a significant negative effect ($\beta = -0.0015$ per day of age $[-0.0022, -0.0006], p = 0.0003$), reflecting the decrease in N-to-V ratio over time.

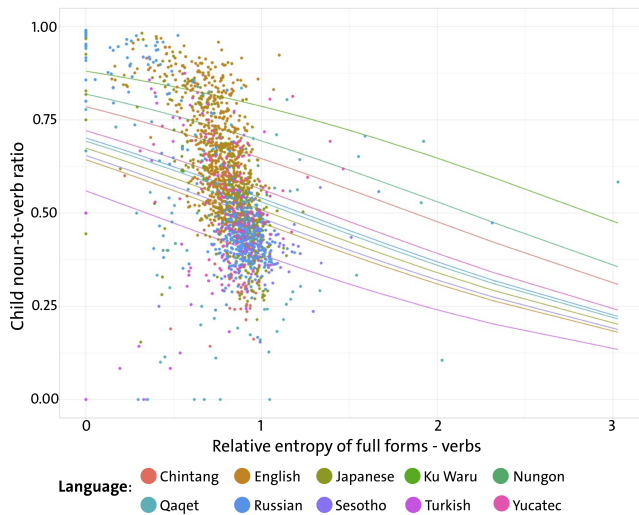


Figure 2: Language-level models of the effect of relative verb form entropy on children's noun-to-verb ratios.

We also evaluated the effect of input noun-to-verb ratios on the noun-to-verb ratios found in children's production, which was significant (but had various strengths) for all languages, with the exception of Japanese. The strongest effect was found for Chintang ($\beta = 5.7040[3.93, 7.48], p = 2.8^{-10}$), see Figure 3. This suggests that for many languages, contextual differences might play a significant role in moderating the proportion of nouns and verbs used by both adults and children.

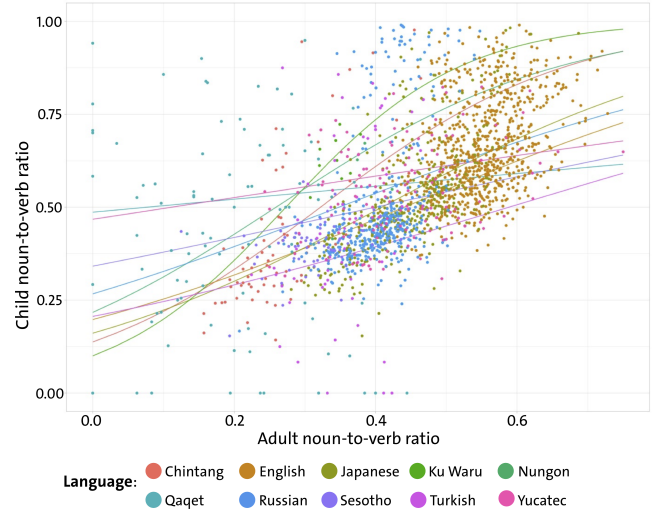


Figure 3: Language-level models showing the effect of input noun-to-verb distributions on children's production in the same recording.

Model 2: This model assesses the effect of verb morphology development on noun-to-verb type ratios. Here only age ($\beta = -0.0005$ per day $[-7.78, -0.00], p = 9.8^{-5}$) and session-specific noun-to-verb ratios in input had a significant effect across all languages. The only language for which relative verb form entropy showed a strong significant effect on noun-to-verb ratios was Chintang ($\beta = -1.7226[-2.54, -0.90], p = 4.0^{-5}$).

Model 3: This model examines the effect of noun morphology proficiency on token N-to-V distributions. Here both age ($\beta = -0.0016$ per day of age $[-0.002, -0.001], p = 8.46^{-10}$) and relative noun entropy ($\beta = -0.4900[-0.84, -0.14], p = 0.0061$) showed a significant effect. Input distributions had a strong effect in Chintang and Yucatec but only a weak effect or only a trend in other languages.

Model 4: This model examines the effect of noun morphology proficiency on type N-to-V distributions. Here we find a positive effect of the relative entropy of nominal morphology. This means the more similar children are in their proficiency in noun morphology to adults, the higher their noun-to-verb ratio ($slope = 0.5575[3.23, 0.79], p = 3.24^{-6}$ across all languages). In model 4, as in the other models, language-specific effects of input noun-to-verb ratios are stronger in some languages, the strongest effect was observed in Chintang ($\beta = 4.8791[3.81, 5.95], p = 2^{-16}$) and Yucatec ($\beta = 1.7146[0.46, 4.08], p = 1.55^{-06}$).

Discussion

Previous studies have not found a conclusive answer regarding the strength of a potential universal "noun bias" in early language acquisition. Rather than concentrating on children's abstract ability to learn a noun or verb word, we examined the

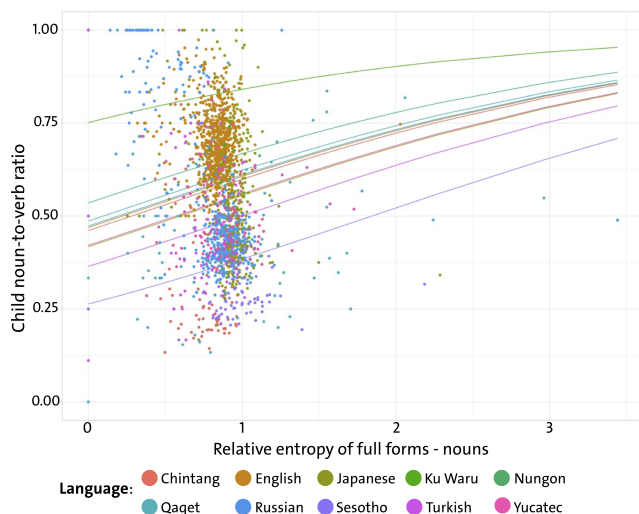


Figure 4: Language-level models showing the effect of relative entropy of noun morphology on children's noun-to-verb ratios.

use of these forms in everyday interactions in 10 maximally diverse languages. Given the strong link between input patterns and children's learning trajectories, a higher proportion of nouns in children's production is especially surprising in languages with a considerable under-representation of nouns. Analyses that focus on either experimental approaches to retention or comprehension of verb and noun labels as well as studies that concentrate on assessing the proportion of nouns and verbs represented in children's early lexical inventories, provide only part of the picture. To complement these studies, we assessed the potential effect of multiple independent variables on the use of noun and verb forms in naturalistic contexts. To ensure that all levels of comparison are taken into account, we evaluated distributions of nouns and verbs in full form types, full form tokens, and stem types.

In our sample of maximally diverse languages, all children showed a higher noun-to-verb rate and thus noun preference compared to surrounding adults in the same recordings. With the exception of Sesotho children, this ratio was well above 0.5 for children, whereas adult speakers' ratio was below 0.5, with the exception of English adults who use more noun than verb forms. The models described here provide additional evidence for a previously proposed developmental pattern: children's gradual decrease in noun-to-verb ratios towards adult levels is correlated with an increase in their proficiency in verb morphology (Stoll, Bickel, Lieven, Paudyal, et al., 2012; Taverna & Waxman, 2020). Since this observation holds for a sample of typologically highly diverse languages, our results suggest that this might be a universal tendency during the early years of language acquisition. This is an important addition to the literature, since a majority of studies, besides the two mentioned above, focus on semantic or lexical properties of nouns and verbs as the preferred explanation for

these patterns. Our results, however, highlight an important correlation between morphological development and the proportion of nouns and verbs in children's production.

Model 1 showed that relative verb form entropy has an effect on the noun-to-verb distribution measured in tokens, while it does not have the same explanatory power for explaining noun-to-verb distributions in types (model 2). This means that, as children's productivity in verb morphology increases, their use of nouns and verbs becomes more like that of adults. Model 3 showed that children's relative noun form entropy had a similar relationship with their noun-to-verb ratios: as proficiency increases, noun-to-verb ratios measured in tokens decrease toward adult levels. However, in a reversal of this trend, model 4, which captured the relationship between relative entropy of nominal morphology and noun-to-verb ratios measured in types, shows a positive correlation between the increase in proficiency and noun-to-verb rates. This result was unexpected and merits further investigation in future studies.

Unlike relative entropy, which had a similar effect across languages, the effect of noun and verb distributions in input was more dependent on the individual languages and varied considerably. We found the strongest effect of input distributions on children's production in Chintang, which is surprising given that Chintang children show one of the greater noun-to-verb ratio deviations compared to their input.

Given the fairly distinct developmental trajectories this comparison uncovered, future work must include a more detailed exploration of the forms that we examine in the use of adults and children as well as the interaction between various properties of input distributions and children's production. The fact that adults' production seems to show an effect on children's production in some languages but less so or not at all in others, opens up interesting questions of linguistic properties of the forms, information structure, conversational conventions as well as the role of child-directed as opposed to child-surrounding speech for future studies. Given the variety of linguistic structures and child-rearing conventions, comparisons should include careful examination of both fine-grained, highly language-specific aspects of word form structures and conversational practices but also try to develop an understanding of the interplay of more general effects and cross-linguistically valid cognitive mechanisms influencing language development.

Acknowledgments

The creation of the database used in this paper was supported by funding from the European Research Council (ERC) under the European Union's Seventh Framework Program (FP7-2007-2013) (Grant agreement No. 615988; PI Sabine Stoll) as part of the project 'Acquisition processes in maximally diverse languages: Min(d)ing the ambient languages (ACQ-DIV)'.

References

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park,

- S.-Y., Pascual, L., ... 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.
- Brown, P. (1998). Children's first verbs in Tzeltal: evidence for an early verb category. *Linguistics*, 36(4).
- Casillas, M., Foushee, R., Méndez Girón, J., Polian, G., & Brown, P. (2024, January). Little evidence for a noun bias in Tzeltal spontaneous speech. *First Language*, 01427237231216571.
- Chao, A., & Shen, T.-J. (2003). Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environmental and ecological statistics*, 10(4), 429–443.
- Childers, J. B., Porter, B., Dolan, M., Whitehead, C. B., & McIntyre, K. P. (2020, February). Does children's visual attention to specific objects affect their verb learning? *First Language*, 40(1), 21–40.
- Choi, S., & Gopnik, A. (1993). Nouns are not always learned before verbs: An early verb spurt in Korean. In *The Proceedings of the 25th Annual Child Language Research Forum* (Vol. 25, p. 96).
- Clark, E. V. (1993). *The Lexicon in Acquisition* (1st ed.). Cambridge University Press.
- de León, L. (1999). Verbs in Tzotzil (Mayan) early syntactic development. *International Journal of Bilingualism*, 3(2-3), 219–239.
- Demuth, K. (2015). *Demuth Sesotho Corpus*.
- Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Boca Raton, FL: CRC Press.
- Fenson, L., Dale, P. S., Reznick, J., Thal, D., Bates, E., Hartung, J. P., ... Reilly, J. S. (1991). *Technical manual for the MacArthur communicative development inventories*. San Diego, CA: San Diego State University.
- Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child development*, 988–998.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *BBN report; no. 4854*. (Publisher: Champaign, Ill.: University of Illinois at Urbana-Champaign, Center for the ...)
- Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets words: how children learn verbs* (p. 544-564). Oxford: Oxford University Press.
- Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (p. 215-256). Cambridge: Cambridge University Press.
- Hellwig, B., Dawuda, C., Frye, H., & Reetz, S. (2014). *The Qaqet Corpus at the Language Archive Cologne*.
- Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in chinese-, english-, and japanese-speaking children. *Child Development*, 79, 979–1000.
- Küntay, A. C., Koşbac, D., & unpublished, S. S. T. (n.d.). *Koç University Longitudinal Language Development Database on language acquisition of 8 children from 8 to 36 months of age*.
- Macnamara, J. (1972). Cognitive basis of language learning in infants. *Psychological review*, 79(1).
- Miyata, S., & Nisisawa, H. Y. (2012). *Japanese CHILDES: The 2012 CHILDES manual for Japanese*. Pittsburgh, PA: TalkBank.
- Moran, S., Schikowski, R., Pajović, D., Hysi, C., & Stoll, S. (2016). The ACQDIV Database: Min(d)ing the Ambient Language. In N. Calzolari et al. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Paris, France: European Language Resources Association (ELRA).
- Moscoso del Prado Martín, F., Kostic, A., & Baayen, R. H. (2004). Putting the bits together: an information theoretical perspective on morphological processing. *Cognition*, 94, 1–18.
- Nelson, K. (1973, February). Structure and Strategy in Learning to Talk. *Monographs of the Society for Research in Child Development*, 38(1/2), 1.
- Pfeiler, B. u. (n.d.). *Pfeiler Yucatec Child Language Corpus*.
- Rumsey, A., Noma, A., Reed, L., Peck, N., Tongeren, C., & Yam, S. (2019). *ACQDIV portion of the Ku Waru Child Language Socialization Study (KWCLSS)*.
- Sarvasy, H. (2017). *Sarvasy Nungon Corpus*.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.
- Stoll, S., & Bickel, B. (2013). Capturing diversity in language acquisition research. *Language typology and historical*, 195–216.
- Stoll, S., Bickel, B., Lieven, E., Banjade, G., Bhatta, T. N., Gaenszle, M., ... Rai, N. K. (2012). Nouns and verbs in chintang: children's usage and surrounding adult speech. *Journal of Child Language*, 39, 284–321.
- Stoll, S., Bickel, B., Lieven, E., Paudyal, N. P., Banjade, G., Bhatta, T. N., ... Rai, N. K. (2012, March). Nouns and verbs in Chintang: children's usage and surrounding adult speech. *Journal of Child Language*, 39(2), 284–321.
- Stoll, S., Lieven, E., Banjade, G., Bhatta, T. N., Gaenszle, M., Paudyal, N. P., ... Bickel, B. (2015). *Audiovisual corpus on the acquisition of Chintang by six children*.
- Stoll, S., Mazara, J., & Bickel, B. (2017). The acquisition of polysynthetic verb forms in chintang. In M. Fortescue, M. Mithun, & N. Evans (Eds.), *Handbook of polysynthesis* (p. 495-517). Oxford University Press.
- Stoll, S., & Meyer, R. (2008). *Audio-visual longitudinal corpus on the acquisition of Russian by 5 children*.
- Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental psychology*, 32(3), 492.
- 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.
- Taverna, A. S., & Waxman, S. R. (2020). Early lexical acquisition in the Wichi language. *Journal of Child Language*, 47(5), 1052–1072.
- Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language*, 28(1), 127–152.
- Waxman, S., 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.