The Acquisition of Case Systems in Typologically Diverse Languages: Children Gradually Generalize Grammatical Rules

Olivier Rüst, Marco Baroni, and Sabine Stoll

Language as we know it, be it spoken or written, is always presented in linear form; words are strung together like beads on a thread. This superficially linear ordering of information, however, hides the latent information on paradigmatic categories and hierarchical dependencies that constitutes grammar. In many languages, such relations are expressed by changing the form or the lexical environment of words. When, where and which marker is used and what these markers express is a great challenge to first language learners. One such potent marker is case, which indicates the roles in a sentence and how constituents relate to each other in a hierarchical structure. How are the paradigmatic relations of a case system learned? How do children learn to adapt different forms to a desired context, i.e., how do they learn to choose the appropriate case? How do they realize how to code roles with a specific case and use the hierarchical dependencies among constituents?

Children have been shown to extract information from the input they hear and initially develop a language that is *item-specific*, where grammatical phenomena are centered around specific lexical items (e.g., Lieven & Tomasello 2008; Goldberg 2006; Abbot-Smith & Tomasello 2006; Tomasello 2000; MacWhinney 1999). Starting from item-specific learning, children must, however, be able to eventually abstract grammatical categories, i.e., they must be able to use these grammatical categories independent of the lexical items. A natural hypothesis about how this is accomplished is that of a *gradually emerging productivity* of, for example, verbal inflections or case, starting from presumably rote-learned item combinations found in the input (Ambridge & Lieven 2011).

As an instance, at an early age host-case combinations are learned item- specifically, e.g., a Turkish speaking child may learn *radyoda* as a chunk, although it can theoretically be segmented into the host (the lexical element that case is attached to) *radyo* and the locative case marker *da*. Only as the child grows to have more language experience is the combination of the lexical item and case marker abstracted into "slots", allowing for a potentially gradual increase in combinations. At this stage, the child realizes that *radyoda* is made up of *radyo* and *da* and starts applying the locative marker *da* to other hosts.

* Olivier Rüst, University of Zürich, ruest.olivier@uzh.ch. Marco Baroni, ICREA, mbaroni@gmail.com. Sabine Stoll, University of Zürich, sabine.stoll@uzh.ch.

 Proceedings of the 46th annual Boston University Conference on Language Development, ed. Ying Gong and Felix Kpogo, 672-685. Somerville, MA: Cascadilla Press.

Previous studies have focused on the lexically restricted use of constructions in the earliest stages of acquisition, and only touched upon the consequent emergence of continuously more abstract structures during the later stages of acquisition. These studies have, indeed, found compelling evidence for lexically restricted use of transitivity (e.g., Theakston et al. 2001; Akhtar 1999; Tomasello & Brooks 1998), determiners (Mariscal 2009; Kirjavainen et al. 2009; Pine & Lieven 1997), verbal paradigms (Mueller Gathercole et al. 1999; Rubino & Pine 1998; Pizzuto & Caselli 1992; Berman 1982) and complex sentences (Diessel & Tomasello 2005) in young children. There are fewer studies characterizing the phase in which item-specific elements are generalized to rules. Dąbrowska (2000) qualitatively investigated the development of interrogative structures in one English speaking child and found gradually more generalized structural schemata of question formation. Mazara & Stoll (2019), in a quantitative study, found an increase in the variability of Russian past tense verb forms that are marked for aspect as children grow older. Dąbrowska (2000)'s study cannot, however, be generalized due to its qualitative character and Mazara & Stoll (2019)'s results could also be explained by vocabulary learning, i.e., the increase in variability in past verb forms marked for aspect might simply derive from the acquisition of novel (and thus item-specific) hosts.

We present here the first thorough empirical characterization of generalization from item-specific constructions to abstract schemata, in the context of the acquisition of host-case paradigms.

On the one hand, studying this phenomenon in its general form requires relatively dense longitudinal samples of naturalistic data from languages with typologically different kinds of case systems. For these purposes, we rely on the ACQDIV database (Moran et al. 2016). This database contains samples of child-directed and child-produced speech from maximally typologically diverse languages (Stoll & Bickel 2013b). We chose Chintang (Sino-Tibetan, Nepal), Russian (Indo-European), Japanese (Japanese), and Turkish (Turkic) (Chintang: Stoll et al. 2015; Russian: Stoll & Meyer 2008; Japanese: Miyata 2012; Nisisawa & Miyata 2010; Miyata & Nisisawa 2010; Nisisawa & Miyata 2009; Miyata & Nisisawa 2009; Miyata 2004a,b,c; Turkish: Küntay et al. Unpublished) due to case prevalence and corpus size.

On the other hand, in order to test the hypothesis of a gradual generalization of host-case combinations towards abstract schemata, we need a mathematical model of how a case paradigm would evolve under this hypothesis and possible alternatives.

We start by approximating such a model through simulations, by using conditional entropy as a measure of host-case variability (Study 1). The simulations model the acquisition of a case paradigm according to different hypotheses (item-specific or instant generalization, see Section 2) and serve to derive precise predictions from these.

In a second step (Study 2), we measure host-case variability in actual naturalistic acquisition data via conditional entropy in our 4 ACQDIV languages, and show that host-case combinations are gradually generalized towards abstract schemata by comparing these patterns to those resulting from the simulations.

In the following sections we present conditional entropy, its relation to our research question and we formulate our hypotheses. This is then followed by the simulation study. Subsequently, we analyze host-case variability in naturalistic data and conclude with a discussion on typological variables that may influence case acquisition.

1. Measuring Case Productivity with Conditional Entropy

The information-theoretic *entropy* measure (Shannon 1948) has been used in language acquisition studies to quantify the amount of predictability/variability of probabilistic variables (e.g., Mazara & Stoll 2019; Lester & del Prado Martín 2016, 2015; Stoll & Bickel 2013a). Shannon (1948) defined entropy as

$$H(X) = \sum_{i=1}^{n} P(x_i) log P(x_i)$$

where n is the number of distinct types and $P(x)$ the probability of occurrence of a type. If the probability of types is relatively uniform (i.e., all types are more or less equally likely to occur), then this results in a relatively high entropy value. Consequently if, for instance, the entropy of the hosts of a case is comparatively high, this indicates a more variable use of hosts, or, equivalently, their low predictability. In terms of acquisition, an increase in entropy then implies that the hosts of this case no longer consist of few predictable elements, but are used with greater flexibility.

Following this intuition, we could, for example, track entropy of the hosts of a specific case - say dative - in the development of a child. We may observe that the entropy of dative hosts increases over time. However, such an increase in entropy could cue two different phenomena. First, it might cue the extension of known forms to the dative case, which would indicate gradual generalization (the phenomenon we are after). However, it could also simply be due to vocabulary learning, where a new noun is first acquired in the dative and used in the dative only (the case-specific acquisition of hosts). In this case, entropy increases because a larger number of hosts can now occur in the dative, but there is no true paradigmatic generalization and only item-specific acquisition of hosts.

Since our interest lies in characterizing the flexibility of a case system as a whole, that is, how the interaction between the relevant hosts of a language and the possible case markers develop, it is more appropriate to use *conditional entropy*, which quantifies the degree of variability in relation to *two* variables – in our case, hosts and case.

More precisely, conditional entropy allows to quantify the predictability or variability of cases when we know the hosts or, conversely, the variability of hosts when we know the cases. Conditional entropy then allows us to capture the variability of the two slots in host-case combinations conditional on each other, and thus lends itself to the evaluation of item-specificity of such combinations (for other uses of conditional entropy see Lester et al. 2018; Lester 2018).

Conditional entropy is calculated as follows. First, a case system is defined by the two variables case and host and has the *joint entropy*:

$$H(\text{case}, \text{host}) = \sum_{x \in \text{hosts}} \sum_{y \in \text{case}} P(x, y) log_2[P(x, y)]$$

Joint entropy quantifies the overall variability of host-case combinations, and is affected by factors such as vocabulary learning and the fact that certain hosts are more likely to appear with certain cases for independent reasons (e.g., geographical names are more likely to appear in a locative). Hence, we do not rely on joint entropy directly as a measure of generalization.

To calculate the conditional entropy of cases given hosts, $H(\text{case}|\text{host})$, the entropy of the hosts H(host) is subtracted from the joint entropy $H(\text{case}, \text{host})$:

$$H(\text{case}|\text{host}) = H(\text{case}, \text{host}) - H(\text{host})$$

Intuitively, conditional entropy measures how much uncertainty is left about the case-host distribution once the uncertainty in hosts is taken into account. If this distribution is entirely accounted for by hosts, conditional entropy would be 0, indicating that the case distribution is entirely predictable by the hosts.

Due to the fact that conditional entropy is not symmetrical, both $H(\text{case}|\text{host})$ and $H(\text{host}|\text{case})$ can be calculated and have different interpretations. $H(\text{host}|\text{case})$ measures the entropy of hosts if the case is known. In other words, it measures the average variability of hosts given the cases. Conversely, $H(\text{case}|\text{host})$ measures the entropy of cases if the host is known, that is, it measures the average variability of case marking across hosts, thus indicating case productivity.

We use $H(\text{case}|\text{host})$, that quantifies the variability of case marking for the specific hosts. This is most informative, as the item-specificity hypothesis predicts that hosts only appear in one (or very few) cases in an early stage of acquisition. Only in the subsequent development hosts are marked more variably with different cases. Using $H(\text{host}|\text{case})$ is less informative, as it quantifies the variability of hosts for the given cases. This measure is, again, not only affected by the generalization process that allows a specific host to be marked more variably for case as the child grows older, but also by vocabulary learning.

Special attention was paid to the calculation of entropy, as entropy estimations according to Shannon's original formula correlate with sample size (Paninski 2003; Carlton 1969; Basharin 1959). We use two recent entropy estimators, PYM and NSB (Archer et al. 2014; Nemenman et al. 2004), that are accurate even when applied to very small sample sizes.[1]

2. Hypotheses

The null hypothesis, H_0, assumes that as soon as a host is acquired, it can be used in all cases (Legate & Yang 2007; Yang 2004). According to this view, a parameter [±CASE] is probabilistically set depending on the evidence in the input. From a child's point of view, case acquisition immediately amounts to learning abstract rules that are not tied to specific lexical items. In effect, this (almost) full combinatoriality of hosts and cases results in a constant value of $H(\text{case}|\text{host})$. This number must be above zero, because case cannot be predicted by the hosts. The null hypothesis thus is: $H(\text{case}|\text{host})$ *shows a constant value above zero*.

If, however, case acquisition proceeds by item-specific memorization followed by progressive generalization towards abstract schemata, we predict the following behaviour: In an early stage of acquisition, where host-case co-occurrences are perfectly item specific, $H(\text{case}|\text{host})$ will be zero. This indicates that the predictability of case given a host is perfect. This is so, because each host type is uniquely associated with only one case at this stage. When the child starts using the same host in several cases, the predictability of case given hosts diminishes, leading to an increase in $H(\text{case}|\text{host})$.

The alternative hypothesis, H_1, thus predicts: $H(\text{case}|\text{host})$ *starts at zero and is followed by a gradual increase*.

3. Study 1: Simulating case acquisition

In order to ensure that our methodology can reliably distinguish between H_0 and H_1 and derive precise predictions from these hypotheses, we computationally simulate case acquisition. These simulations assume different settings according to the hypotheses: (1) any case may be used as soon as a host is learned (full-productivity from start) and (2) case is first acquired item-specifically and then progressively generalised. The exact procedure is illustrated in the schema.

---

[1]The NSB estimator is a Bayesian estimator that takes an infinite mixture of Dirichlet priors, with a parameter to be specified by the experimenter corresponding to the number of types that can theoretically be observed. This estimator is accordingly well-suited to estimate H(case), because we (generally) know how many cases a specific language has. However, it is typically difficult to decide *a priori* the number of host types that a language could possibly have. We thus use the PYM estimator to compute H(host). This latter method relies on a family of priors in countably infinite space with an agnostic cardinality via the Pitman-Yor process. The choice to use both estimators is due to the fact that in our simulations NSB more reliably corrects for sample size on samples with a low cardinality (number of unique observable types) and entropy values as we see in the CASE slot, whereas PYM is more reliable on distributions with higher cardinality and entropy values.

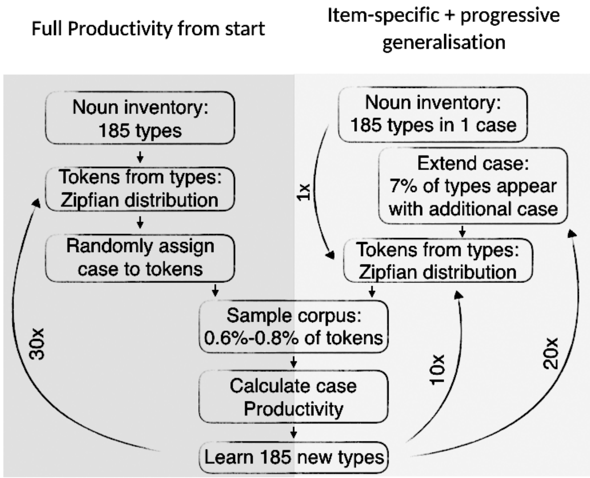

Figure 1: Schematic representation of simulations

Both simulations start out with an assumed noun inventory of 185 types, which is roughly expected from a three year old (e.g., Pérez-Pereira & Cruz 2018). In order to create tokens from these types we multiply each with a number sampled from a Zipfian distribution, which was chosen because it most closely reflects word frequencies in natural language. This results in a sample of 185 nouns, as they may be used in natural language.

After generating nominal tokens, the full-productivity hypothesis randomly assigns one out of five cases to each nominal token (number of cases was chosen arbitrarily). To recreate a corpus sampling procedure in the following step, between 0.6% and 0.8% of these tokens are sampled (this roughly corresponds to the "real" samples of a corpus with 4 hours of recording per month). Case productivity is then calculated on this reduced set of host-case co-occurrences. In a final step, the simulation learns 185 novel host types, which roughly corresponds to the number of new hosts a child learns in a month (e.g., Pérez-Pereira & Cruz 2018). At this point, the second cycle starts: The simulation now "knows" 370 types, and iterates over the steps described starting by creating tokens from the 370 types and then randomly assigning case and so on. This is repeated 30 times.

The item specific simulation on the other hand, creates the initial 185 hosts case-specific, i.e., each host type is uniquely associated with one of five cases. From these types of host-case combinations, tokens are created using the Zipfian distribution. As before, a corpus-recording regime is then simulated by sampling between 0.6% and 0.8% of the tokens, and case productivity is calculated on this reduced sample. This procedure of strictly learning host-case combinations is repeated for 10 cycles to simulate the item-specific phase. Starting in cycle 11, an additional step of case extension is added: After learning 185 novel host types, 7% of the overall known types at this point are extended to another case, i.e., 7% of host types can appear in one more case than initially learned[2]. This cycle of learning novel host types and expanding 7% of all learned hosts to a novel case is repeated 20 times.

These simulations result in two patterns that visualize the exact prediction of case productivity for each hypothesis (see top of figure 2).

4. Study 2: Cross-linguistic acquisition of case
4.1. Case Systems

In order to compare the simulated patterns to naturalistic data, we investigate case acquisition in Chintang, Russian, Turkish and Japanese.

Chintang has the cases nominative, genitive, ergative, locative, directionalis, commitative, modus, method, finalis and perlative, with various sub-cases especially in the locative. These cases are marked with bound morphemes, whose shape is dependent on the phonological environment. Additionally, the nouns in this language show polysynthetic morphology, which allows a host to be marked by several cases at once. These constituents are, however, not as frequent as Chintang allows for subject and object dropping if they can be inferred from context (the environment where they most frequently would occur) (Schikowski et al. 2015).

unambataŋbamubaŋa

unambataŋ	-bamu	-ba	-ŋa
house.of.parents.in.law	-loc.down	-loc	-erg.s

’from the in-laws’

Japanese has 8 different cases: nominative, topical marker, genitive, dative, accusative, lative, ablative and instrumental. In some cases a single noun may be marked by a combination of two case markers. The cases are, however, regularly marked via a post-nominal particle (Martin 2003).

densha	de
train	loc

On the train.

In contrast, Russian has the six cases nominative, accusative, genitive, dative, locative and instrumental. There are three main declension classes that roughly correspond to syntactic gender, with various sub-classes. Cases are marked with a bound morpheme that marks case and number together (polyexponence), which allows for 12 possible derivations of a single noun (Timberlake 2004). While the

[2] As we are entirely agnostic about the rate of extension, several settings were tested. We only show the setting with 7% of types being extended in each cycle.

form of the host is (mostly) stable, the case markers vary greatly from one (sub-) paradigm to another and forms are often subject to complete suppletion.

	singular	plural
nominative	chelovek	ljud**i**
genitive	chelovek**a**	ljud**ej**

Lastly, Turkish features the 5 cases accusative, dative, locative, ablative and genitive that are derived via suffixation. This suffixation is defined by vowel harmony, i.e. the exact shape of the suffix is conditional on the vowel in the stem (Göksel & Kerslake 2004), as shown on the locative.

radyo**da**	sepet**te**
radio-loc	basket-loc
’on the radio’	’in the basket’

To summarize, these case systems differ in a number of variables that may influence acquisition: (1) number of cases, (2) pervasiveness in language (input), (3) number of paradigmatic rules per case, (4) presence of phonological form adjustments, (5) number of case marker a host can take, (6) presence of polyexponence.

4.2. Corpora

Our language sample comes from the ACQDIV database (Moran et al. 2016) that contains typologically diverse longitudinal child-language data. A summary of relevant statistics on the used corpora can be found in the table.

	Children	Age range	Recording rhythm	Environment
Chintang	4	0;7 - 4;3.14	4h per month	outside
Russian	5	1;3.26 - 6;8.12	1h per week	home
Japanese (1)	4	2;11.27 - 5;4.18	70min per week	home
Japanese (2)	3	1;5.7 - 3;1.29	40-60min per week	home
Turkish	8	1;0.2 - 2;9.13	1h every 2 weeks	home

All our corpora contain three or more children that have been recorded for approximately 4h per month (except Turkish, 2h) and cover a total age range from 0;7 to 6;8.12. For Russian, Japanese and Turkish, the children were recorded at home. Chintang children, however, spend most of their time outside in the presence of other children and were thus recorded in this environment.

4.3. Method

In order to gauge case productivity of speakers (both adults and children), conditional entropy $H(\text{case}|\text{host})$ is calculated as described for each recording session. This value is cross-indexed with the child name and child age (of each specific recording session), the language spoken and the type of speech (child or child-directed speech).

We then use a linear Bayesian model (Bürkner 2018) to predict $H(\text{case}|\text{stem})$ with a first order interaction between child age and type of speech (child or child-directed) and a main effect for type of speech. Additional random intercepts are added for speaker, type of speech and language.

5. Results

We created simulations of a child's use of case in order to derive precise predictions of the hypotheses. H_0 predicts that children use case fully productively from early on (Legate & Yang 2007; Yang 2004), whereas H_1 suggests that case-use is first constrained to lexically specific items, and is only subsequently gradually generalized (Lieven & Tomasello 2008).

The simulations (see figure 2) result in clearly distinguishable predictions: H_0 results in constant case productivity, while H_1 shows a constant increase in case productivity after the item-specific phase.

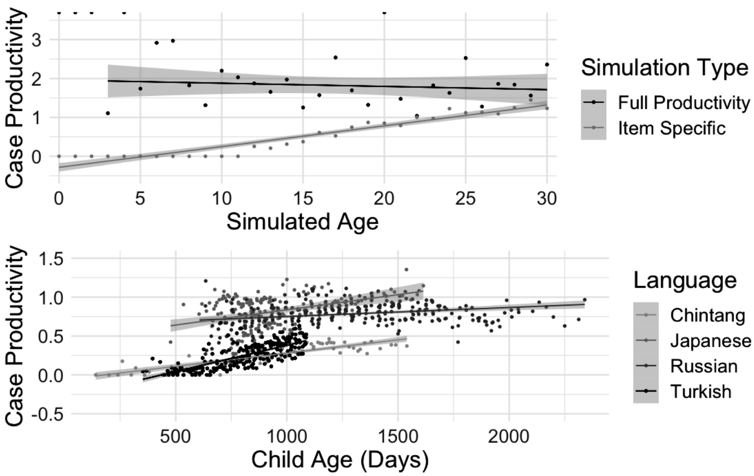

Figure 2: Results of simulation (top) and cross-linguistic study (bottom).

The comparison between the learning trajectories in naturalistic longitudinal data and those from the simulations reveals that the item-specific hypothesis with subsequent gradual generalization best explains our results. The specific trajectories of the four typologically diverse languages all show a gradual increase of case productivity, albeit following language specific trajectories: the Chintang and Turkish case systems are acquired slower whereas Russian and Japanese ones are acquired more rapidly.

The results of the cross-linguistic regression analysis are summarized in Table 1. We predict case productivity, operationalized as $H(\text{case}|\text{host})$, with a first order interaction between child age and type of speech (child or child-directed) and a main effect for type of speech with additional random intercepts for speaker, type of speech and language.

Table 1: Cross-linguistic Regression Analysis

Effects	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept	-0.05	1.75	-3.46	3.66
Child Speech	-0.49	2.52	-5.59	6.04
Child-directed Speech : Age	0.10	0.03	0.04	0.16
Child Speech : Age	**0.26**	**0.03**	**0.20**	**0.31**

The point estimate for child speech shows that their case productivity is lower than that for child-directed speech, although the 95% compatibility interval is extremely large. This point estimate serves as a sanity check, where the large variance is expected as there are huge cross-linguistic and inter-child differences. Interestingly, child-directed speech shows an increase in case productivity as children grow up. This may indicate that adults adjust their case use to the needs of children, and use case more variably as children grow up. Most importantly, however, there is a statistically credible increase of case productivity in child speech as children grow older. Tellingly, this increase is steeper in child than in child-directed speech, suggesting a convergence of child and child-directed speech in terms of their case productivity.

6. Discussion

The use of item-specific structures in early child speech is well attested in the literature. Children have to, however, become productive at some point. Here, we show how a grammatical category like case evolves from item-specific constructions and slowly becomes productive.

In a first step, we derive precise predictions from the item-specific hypothesis with subsequent generalization versus the full-productivity hypothesis, using simulations. Second, we investigate child case usage in four typologically diverse languages: Chintang, Japanese, Russian and Turkish. Third, and finally, by comparing the results of the simulation with those of the cross-linguistic study, we can show that the item-specific hypothesis followed by gradual generalization best explains our data.

While we do find cross-linguistic support for this hypothesis, children also show language specific learning trajectories. These may be explained by a number variables that influence the difficulty of abstracting a specific case system.

Chintang, as an instance, features 14 different cases, allows subject and object dropping, requires phonological form adjustments and can mark a single host with several cases. In effect, these children have to learn 14 different syntactic roles and their function, in a language where the environment that most frequently gives direct evidence, is rather scarce due to the subject and object dropping. Moreover, recognizing patterns that allow the recognition of case markers is difficult, as there are phonological form adjustments that may hinder a case marker being recognized as such, especially when they occur in concatenation with other cases. These could be factors that may explain why Chintang case is learned comparatively late.

Similarly, Turkish case is also acquired relatively slowly. Contrary to Chintang that features many possible variables that may complicate the acquisition of case, Turkish "only" features 5 cases that are marked with regular case markers. However, the vowel (and in some cases the consonant) in these morphemes is conditional on the phonological form of the host. The fact that Turkish case is acquired comparatively late, may indicate that children are extremely reliant on stable surface forms and struggle to recognize patterns if these are adjusted to the surrounding phonological environment.

Compared to this, the 6 Russian cases are acquired relatively quickly. This is interesting, as the Russian case system is rather complex: There are three main declension classes that imperfectly correlate with syntactic gender, and contain various sub-classes that contain many idiosyncrasies (e.g., indeclinable common nouns, compounds, appositives, names) and irregularities (e.g., suppletion). However, the surface form of most morphemes is stable, as they do not show any phonological adjustments. This may indicate that the ability to recognize (stable) patterns is much more detrimental than pure number of case and/or paradigms and their complexity.

Lastly, the Japanese system that contains 8 cases is also acquired rather quickly. Although it contains more cases than Russian, this system is extremely regular and would thus allow for easy recognition of surface patterns.

Overall then, the gradual abstraction of grammatical categories and specifically case, may be mitigated by a number of variables. A main detrimental factor, however, that influences speed/ease of acquisition seems to be the ability to recognize surface patterns via the phonological forms of morphemes and words. To further investigate typological effects, we plan a multi-variate regression analysis with a number of typological variables. In doing this, we hope to sharpen our understanding of the mechanisms that children rely on to detect patterns in the input and infer complex grammatical rules.

References

Abbot-Smith, Kirsten & Michael Tomasello. 2006. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review* 23(3). 275–290.

Akhtar, Nameera. 1999. Acquiring basic word order: Evidence for data-driven learning of syntactic structure. *Journal of Child Language* 26(2). 339–356.

Ambridge, Ben & Elena Lieven. 2011. *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press.

Archer, Evan, Il Memming Park & Jonathan W Pillow. 2014. Bayesian entropy estimation for countable discrete distributions. *The Journal of Machine Learning Research* 15(1). 2833–2868.

Basharin, Georgij. 1959. On a statistical estimate for the entropy of a sequence of independent random variables. *Theory of Probability & Its Applications* 4(3). 333–336.

Berman, Ruth A. 1982. Verb-pattern alternation: the interface of morphology, syntax, and semantics in Hebrew child language. *Journal of Child Language* 9(1). 169–191.

Bürkner, Paul-Christian. 2018. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal* 10(1). 395–411.

Carlton, A.G. 1969. On the bias of information estimates. *Psychological Bulletin* 71(2). 108.

Dąbrowska, Ewa. 2000. From formula to schema: The acquisition of English questions. *Cognitive Linguistics* 11(1-2). 83–102.

Diessel, Holger & Michael Tomasello. 2005. A new look at the acquisition of relative clauses. *Language* 882–906.

Göksel, Aslı & Celia Kerslake. 2004. *Turkish: A comprehensive grammar*. Routledge.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford University Press on Demand.

Kirjavainen, Minna, Anna Theakston & Elena Lieven. 2009. Can input explain children's me-for-I errors. *Journal of Child Language* 36(5). 1091–1114.

Küntay, Aylin Copty, Dilara Koçbaş & Süleyman Sabri Taşçı. Unpublished. Koç University Longitudinal Language Development Database on language acquisition of 8 children from 8 to 36 months of age .

Legate, Julie Anne & Charles Yang. 2007. Morphosyntactic learning and the development of tense. *Language Acquisition* 14(3). 315–344.

Lester, Nicholas. 2018. *The syntactic bits of nouns: How prior syntactic distributions affect comprehension, production, and acquisition*: UC Santa Barbara dissertation.

Lester, Nicholas, Daniel Baum & Tirza Biron. 2018. Phonetic duration of nouns depends on de-lexicalized syntactic distributions: Evidence from naturally occurring conversation. In *Cognitive Science*, 2035-2040.

Lester, Nicholas & Fermín Moscoso del Prado Martín. 2015. Constructional paradigms affect visual lexical decision latencies in English. In *Cognitive Science*, 1320-1325.

Lester, Nicholas & Fermín Moscoso del Prado Martín. 2016. Syntactic Flexibility in the Noun: Evidence from Picture Naming. In *Cognitive Science*, 2585-2590.

Lieven, Elena & Michael Tomasello. 2008. *Children's first language acquistion from a usage-based perspective.* Routledge/Taylor & Francis Group.

MacWhinney, Brian. 1999. *The emergence of language*. Taylor & Francis.

Mariscal, Sonia. 2009. Early acquisition of gender agreement in the Spanish noun phrase: starting small. *Journal of Child Language* 36(1). 143.

Martin, Samuel E. 2003. *A reference grammar of Japanese*. University of Hawaii Press.

Mazara, Jekaterina & Sabine Stoll. 2019. How Do Children Become Flexible in Their Use of Grammatical Categories? The Aspect Hypothesis Revisited. In *Proceedings of the Boston University Conference on Language Development* 44, 363–375. Cascadilla Press.

Miyata, Susanne. 2004a. Aki Corpus. *Pittsburgh, PA: TalkBank* .

Miyata, Susanne. 2004b. Ryo Corpus. *Pittsburgh, PA: TalkBank* .

Miyata, Susanne. 2004c. Tai Corpus. *Pittsburgh, PA: TalkBank* .
Miyata, Susanne. 2012. Japanese CHILDES: The 2012 CHILDES manual for Japanese .
Miyata, Susanne & Hiro Yuki Nisisawa. 2009. *MiiPro – Asato Corpus*. Pittsburgh, PA: TalkBank.
Miyata, Susanne & Hiro Yuki Nisisawa. 2010. MiiPro – Tomito Corpus. *Pittsburgh, PA: TalkBank* .
Moran, Steven, Robert Schikowski, Danica Pajović, Cazim Hysi & Sabine Stoll. 2016. The ACQDIV database: Min (d) ing the ambient language .
Mueller Gathercole, Virginia C, Eugenia Sebastián & Pilar Soto. 1999. The early acquisition of Spanish verbal morphology: Across-the-board or piecemeal knowledge? *International Journal of Bilingualism* 3(2-3). 133–182.
Nemenman, Ilya, William Bialek & Rob De Ruyter Van Steveninck. 2004. Entropy and information in neural spike trains: Progress on the sampling problem. *Physical Review E* 69(5). 056111.
Nisisawa, Hiro Yuki & Susanne Miyata. 2009. MiiPro – Nanami Corpus. *Pittsburgh, PA: TalkBank* .
Nisisawa, Hiro Yuki & Susanne Miyata. 2010. MiiPro – ArikaM Corpus. *Pittsburgh, PA: TalkBank* .
Paninski, Liam. 2003. Estimation of entropy and mutual information. *Neural computation* 15(6). 1191–1253.
Pérez-Pereira, Miguel & Raquel Cruz. 2018. A longitudinal study of vocabulary size and composition in low risk preterm children. *First Language* 38(1). 72–94.
Pine, Julian M & Elena VM Lieven. 1997. Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics* 18(2). 123–138.
Pizzuto, Elena & Maria Cristina Caselli. 1992. The acquisition of Italian morphology: Implications for models of language development. *Journal of Child Language* 19(3). 491–557.
Rubino, Rejane B & Julian M Pine. 1998. Subject–verb agreement in Brazilian Portuguese: what low error rates hide. *Journal of Child Language* 25(1). 35–59.
Schikowski, Robert, Netra Paudyal & Balthasar Bickel. 2015. 18. Flexible valency in Chintang. In *Volume 1 Introducing the Framework, and Case Studies from Africa and Eurasia*, 669–708. De Gruyter Mouton.
Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423.
Stoll, Sabine & Balthasar Bickel. 2013a. The acquisition of ergative case in Chintang. *The Acquisition of Ergativity* 9. 183–208.
Stoll, Sabine & Balthasar Bickel. 2013b. Capturing diversity in language acquisition research. *Language Typology and Historical Contingency* 195–216.
Stoll, Sabine, Elena Lieven, Goma Banjade, Toya Nath Bhatta, Martin Gaenszle, Netra P. Paudyal, Manoj Rai, Novel Kishor Rai, Ichchha P. Rai, Taras Zakharko, Robert Schikowski & Balthasar Bickel. 2015. *Audiovisual corpus on the acquisition of Chintang by six children*.
Stoll, Sabine & Roland Meyer. 2008. Audio-visual longitudinal corpus on the acquisition of Russian by 5 children .
Theakston, Anna L, Elena VM Lieven, Julian M Pine & Caroline F Rowland. 2001. The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language* 28(1). 127–152.
Timberlake, Alan. 2004. *A reference grammar of Russian*. Cambridge University Press.

Tomasello, Michael. 2000. The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences* 4(4). 156–163.

Tomasello, Michael & Patricia J Brooks. 1998. Young children's earliest transitive and intransitive constructions.

Yang, Charles D. 2004. Universal Grammar, statistics or both? *Trends in Cognitive Sciences* 8(10). 451–456.

Proceedings of the 46th annual Boston University Conference on Language Development

edited by Ying Gong
and Felix Kpogo

Cascadilla Press   Somerville, MA   2022

Copyright information

Proceedings of the 46th annual Boston University Conference on Language Development

ISSN 1080-692X
ISBN 978-1-57473-077-7 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com