Modeling the Acquisition of Question Variants in English

An Nguyen and Colin Wilson

Young children's production typically does not perfectly align with their input, especially when there is variation in the input. For example, Pozzan and Valian (2016) find that although children are exposed to both inverted polar questions (e.g., *Is Sally here?*) and non-inverted polar questions (e.g., *Sally is here?*), 3-to-5-year-olds uniformly apply inversion in their production. Similarly, Nguyen and Legendre (2021) find that English parental input of wh-questions includes both the wh-fronted (e.g., *Where is Sally?*) and wh-in-situ (e.g., *Sally is where?*) variant; however, children almost exclusively produce only the wh-fronted variant in both spontaneous speech and elicited task. Such regularizing behavior by children is observed not only in natural language studies but also in artificial language learning experiments. In their experiment, Hudson Kam and Newport (2005) expose adults and children to an artificial language with unpredictable variation in the determiner-noun system. Adults tend to match the frequency in the input even if that results in inconsistency in the language, while children tend to regularize to the most frequent variant and produce a more systematic language compared to their input. However, regularization is not a mere result of frequency effect. In another experiment, learners regularize to low-frequency but high-consistency variants while initially ignoring low-frequency and low-consistency ones (Hudson Kam and Newport, 2009). These studies suggest that when there are multiple variants of the same grammatical item in the input, children tend to only produce the dominant variant (or a subset of variants). Moreover, their regularizing pattern is not random but is conditioned by a number of factors, including but not limited to frequency and consistency.

Interestingly, the regularizing behavior is only shown in production while comprehension remains intact. Schwab, Casey, and Goldberg's (2018) artificial language learning experiment shows that although children regularize to the more frequent variant in the elicited production task, they demonstrate good comprehension for both the frequent and infrequent variants in the two-alternative-forced-choice task. Nguyen and Legendre (2021) also find that children have good comprehension of wh-in-situ and can even differentiate information-seeking wh-in-situ questions from repetition-seeking wh-in-situ (i.e., echo questions). Why, then, do children regularize their production? One potential answer is due to limited cognitive resources. Learning multiple variants of the same grammatical item is cognitive taxing, and learners can reduce the load by regularizing to fewer variants. This idea is supported by studies suggesting that increasing cognitive load can induce regularizing. Hudson Kam and Newport (2009) and Fedzechkina, Newport, and Jaeger (2017) show that by adding more

 Proceedings of the 46th annual Boston University Conference on Language Development, ed. Ying Gong and Felix Kpogo, 563-573. Somerville, MA: Cascadilla Press.

complexity to the learning task (e.g., increasing the number of variants), adult learners start to regularize their production as well. The asymmetry between comprehension and production could arise due to the difference in the amount of memory and planning required for comprehension versus production (Hendriks and Koster, 2010; Humphreys, 2012).

In this project, we use a computational model to explore what mechanism helps make children's production more systematic and regularized than their input. While we do not and cannot claim that this is how children actually acquire multiple variants, the goal of the model is to demonstrate a possible mechanism by which children regularize their production that captures many of the proposals other researchers have put forward, including the use of frequency and consistency as selecting factors. We focus on the acquisition of multiple English wh-question variants here, however, the model can be modified to work with other regularizing cases as well. In the next section (section 1), we will first present a brief overview of the English wh-questions and report current experimental findings on children's acquisition of such variants. We then describe the model in section 2, and discuss the theoretical contribution of the model in section 3. Section 4 discusses future work and concludes the paper.

1. Overview of wh-questions in English

Nguyen and Legendre's (2021) corpus study of 10 children finds that 3-5-year-olds are exposed to at least three different types of wh-questions in their input: fronted information-seeking questions (example 1a), in-situ information-seeking questions (i.e., probe questions) (example 1b), and in-situ repetition-seeking questions (i.e., echo questions) (example 1c).

(1) a. Where are you going?
b. A: Mary gave (unintelligible) to John. B: Mary gave WHAT to John?
c. On Thursday he ate through 4 strawberries, but he was still hungry. On Friday he ate through 5 oranges, but he was still what?

Fronted information-seeking questions are the most frequently used questions. This type of question has a flat/falling pitch accent and can be used in any contexts in which a question is called for. In-situ information-seeking questions similarly have a flat/falling pitch accent. However, they are typically only used in a discourse-given context where the speaker has some authority over the listener, hence they are most frequently found in quiz shows, courtroom and legal questioning, classroom, and child-directed speech. Finally, in-situ repetition-seeking questions, or echo questions, have a high rising pitch accent and heavy stress on the wh-word. This type of question is used to request for a repetition or clarification of a previous (unintelligible) utterance instead of new information. Hence, their usage is more restricted as they can only occur as a reaction to another utterance. In brief, these three types of questions are similar in some aspects and dissimilar in others, making it difficult to classify them based on just one linguistic dimension. Fronted questions and probe questions are semantically and prosodically similar, but they have different surface structures. Probe questions and echo questions share the same in-situ structure, but have different meanings and intonations

Wh-in-situ takes up about 15% of all main-clause questions in child-directed speech, although the exact percentage varies among children, with some being exposed to over 20% of wh-in-situ while some are exposed to less than 10%. In general, probe questions are more frequent than echo questions. Interestingly, corpus analysis shows that children produce less than 1% of wh-in-situ out of all wh-question production. One potential explanation is that children refrain from this type of question because of its context restrictions. It could be that in natural settings, it is rare for children to be in a situation where it is felicitous for them to use probe questions. However, Nguyen and Legendre (2021) conduct a study on the comprehension and production of wh-in-situ and find that 1) children demonstrated good comprehension of wh-in-situ, they could differentiate between probe questions and echo questions and gave appropriate new-information answer or repetition-answer depending on the type of questions they heard; 2) even with the appropriate pragmatic setting (discourse givenness and speaker authority), children showed a strong preference for fronted wh-questions by producing almost exclusively fronted questions (98.7%) in the elicited task. While this may be unexpected to traditional structural economy accounts, which predict that children would prefer the structurally simpler wh-in-situ questions than fronted questions (as they do not require wh-fronting or subject-auxiliary inversion), the result is in line with other regularizing studies, showing that children tend to regularize to the dominant variant when there are multiple variants in the input. In the next section, we describe a non-parametric Bayesian model of how a child could regularize to fronted wh-questions in the input initially yet learn all three types of questions eventually.

2. A non-parametric Bayesian model of wh-question learning

The defining property of non-parametric Bayesian models is not an absence of parameters (as the name might suggest), but rather the ability to infer both the number of parameters as well as their values from data. In the case of clustering, non-parametric models can make inferences about how many clusters underlie a given data set and the internal properties of each cluster. Due to this flexibility, non-parametric models have been successfully applied to many clustering problems in linguistics and cognitive science, including syllable, morpheme, and word segmentation (Goldwater, 2007; Johnson, 2008; O’Donnell, 2015; Seshadri, Remes, and Räsänen, 2017), phonetic category learning (Lee, O’Donnell, and Glass, 2015), syntactic/semantic rule learning (Abend et al., 2017), and psychological category induction (Sanborn, Griffiths, and Navarro, 2010).

While language-internal variation in wh-questions does exist, the number of types within a given language is likely to be quite small and certainly not unbounded (compare the number of lexical items that must be identified in word segmentation). We adopt the non-parametric approach because, unlike classical parametric finite mixture models, it does not force the learner to commit to the existence of a particular number of clusters (question types) in advance of analyzing the input data. For purposes of implementation, we place an upper bound of $K = 10$ as the maximum number of wh-question types that the model can learn. This bound is much larger than the subset of questions that we analyze here; it can easily be raised if languages with more than ten wh-question types are found to be attested.

2.1. Data

We collected English wh-questions in main clauses from four CHILDES audio corpora: HSLLD (Dickinson and Tabors, 2001), Snow (MacWhinney and Snow, 1990), Van Houten (Van Houten, 1986), and Weist (Weist and Zevenbergen, 2008). Questions were annotated by hand for wh-fronting and subject-auxiliary inversion. To extract the prosodic properties, the questions were forced-aligned with the Montreal Forced Aligner (McAuliffe et al., 2017) and subsequently analyzed using the PRAAT software (Boersma and Weenink, 2019) to extract the duration (measured in ms) and the f0 contour (final Hz - initial Hz) of the vowel in the wh-word.

Because the audio data available on CHILDES was limited, the 88 utterances coded as just described were reserved for testing only. The model was trained on 2000 simulated instances randomly generated according to the values in Table 1a. (As further spoken examples of wh-questions in CDS become available, we anticipate being able to train and test the model entirely on natural utterances.)

Each question type inferred by the model consists of a probability distribution over several properties. Instead of examining an exhaustive list of linguistic properties, we limit the present study to the two morphosyntactic properties and two prosodic properties that are most relevant for our English case studies. A more comprehensive typology of wh-questions across languages would require more properties, and the model can be straightforwardly modified to include additional parameters. The morphosyntactic properties are discrete variables that can take on two values (1 corresponding to the presence of a property and 0 corresponding to its absence). They include the position of the wh-word and the inversion status of the auxiliary. The prosodic properties consist of two continuous variables: the duration (milliseconds) and F0 contour (ΔHz) of the wh-word. One potential issue is that it may be difficult to determine whether the differences between wh-in-situ and fronted wh-questions arise from the accent on the wh-word or from the position of the wh-word (utterance-initial versus utterance-final prosody). However, as most studies on echo questions have analyzed their unique intonation in terms of their stressed wh-word (Artstein, 2012; Cheng and Rooryck, 2002), we suggest that the differences between probing and echo questions could emerge from the wh-word itself instead of utterance-final position.

Finally, the frequency distributions of question types in the simulated data matched distribution in CDS: 84% fronted questions, 9% probe questions, and 7%

echo questions. The probability distributions of the morphosyntactic and prosodic properties within each question type were also fit to the CDS utterances. All fronted wh-questions require auxiliary inversion while in-situ questions do not, thus the values for wh-fronting and inversion are identical. As for the continuous variables, echo questions typically have longer durations of the wh-word and a rising intonation which is expressed as positive values of ΔF0. Probe questions and fronted questions have a shorter duration on the wh-word, and negative or close to zero values of ΔF0 to indicate falling or flat intonation.

2.2. Model specification

The non-parametric model proposed here is technically a Dirichlet Process Mixture Model (e.g., Gershman and Blei, 2012), as specified below.

$$
\begin{aligned}
&\textbf{Cluster probabilities}\\
\alpha &\sim Gamma(1,1)\\
v_\ell \mid \alpha &\sim Beta(1,\alpha) \qquad \text{for } \ell = 1,\ldots,K-1\\
\alpha &\sim Gamma(1,1)\\
w_1 &= v_1\\
w_k &= v_k \prod_{\ell=2}^{k-1} 1 - v_\ell \qquad \text{for } k = 2,\ldots,K-1\\
w_K &= \prod_{\ell=1}^{K-1} 1 - v_\ell\\
&\textbf{Parameters of each cluster}\\
p_{kj} &\sim Beta(1,1) \qquad \text{for } j \in \{WhFront, Inver\}\\
\mu_{k\ell} &\sim Normal(M_\ell, S_\ell)\text{for } l \in \{WhDur, Wh_{\Delta F0}\}\\
\log \sigma_{k\ell} &\sim Normal(M_\sigma, S_\sigma)\\
&\textbf{Distribution of observations}\\
&p(\mathbf{y}_i|\mathbf{w},\mathbf{p},\mu,\sigma) \qquad \text{for } n = 1,\ldots,N\\
&= \sum_{k=1}^{K} w_k \prod_j Bernoulli(y_{ij}|p_{kj}) \cdot \prod_\ell Normal(y_{i\ell}|\mu_{k\ell},\sigma_{k\ell})
\end{aligned}
$$

The cluster mixture weights w_k are given by a stick-breaking procedure (Sethuraman, 1994). Starting with a unit length stick, in each step a portion of the stick is broken off according to v_l and assigned to w_k. The independent random variables v_l have the distribution *Beta(1, α)*. Higher values of α will yield less concentrated distributions, allowing the weights to decay more gradually. As *α* decreases, less of the unit-length stick will be left for subsequent values, yielding a smaller number of clusters. The concentrated parameter *α* can be regarded as the learner's belief about whether there are many or few wh-questions in a given language, a point we will return to in section 2.3

We represent each morphosyntactic property of the i^{th} question utterance as a binary value $y_{ij} \in \{0, 1\}$. In our data sets, we have two morphosyntactic properties (wh-fronting and subject-auxiliary inversion), hence $j \in \{1, 2\}$. Each question type k assigns a probability $p_{kj} \in [0,1]$ that the j^{th} property will be present (=1) in a question of that type. The prior probability distribution over p_{kj} is a *Beta(1,1)* distribution, which assigns equal prior probability to all values in $[0, 1]$. Similarly, each prosodic property of the i^{th} question utterance is a continuous variable y_{il}. There are two prosodic properties in our data (duration and F0 contour on the wh-word), therefore $l \in \{1, 2\}$. Each question type k places a $Normal(\mu_{kl}), \sigma_{kl}$ distribution on the l^{th} prosodic property. The prior on the mean μ_{kl} is a *Normal(100, 50)* distribution for duration (which is necessarily positive) and a *Normal(0, 50)* distribution for F0 contour (which can be rising or falling). The prior distribution on σ_{kl} was a broad log-normal distribution, allowing for substantial variation within each question type.

We implemented the model in the probabilistic programming language Stan (Carpenter et al., 2017) and assessed its ability to infer accurate wh-question types for English and to correctly categorize question utterances drawn from CDS corpora. The probability that the i^{th} question utterance, represented as two binary morpho-syntactic variables and two continuous prosodic variables, belongs to question type k is given by Bayes' Rule:

$$p(k|\mathbf{y}_i) = \frac{p(\mathbf{y}_i|k)\ w_k}{\sum_{k=1}^{K} p(\mathbf{y}_i|k')\ w_{k'}}$$

where each $p(y_i \mid k)$ is a product of two Bernoulli probabilities and two Normal densities.

2.3. Regularizing behavior

We attempt to model the regularizing pattern by manipulating two factors: a filtering rate and a parsimony bias α value.

The idea behind the filtering rate is that learners cannot utilize everything presented to them in the input when cognitive resources are limited. This ties back to the distinction between *input* vs *intake (e.g.,* Gass, 1997; Gagliardi and Lidz, 2014; Omaki and Lidz, 2015)*: input* is the data available in the environment, while *intake* is the data from the input that learners actually utilize to make inferences about the target grammar. Early on, with limited cognitive resources, children’s intake is smaller, in other words, the filtering rate to get to the intake from the input is higher. We used weighted random sampling for the intake instead of pure random sampling, in which utterances that are consistent and frequent are more likely to be selected. To reflect a developmental trajectory, the filtering rate was slowly reduced over time until the intake matches the input.

The parsimony bias α value represents the learner's initial bias about the number of wh-questions in their language. A larger value of α would allow the learner to be more flexible in learning more categories of wh-questions. We hypothesize that early on, learners would have a stronger preference to learn as few variants as possible to reduce the cognitive burden. Such bias is weakened over time and can eventually be overridden after sufficient exposure to the variable pattern. We thus compared three conditions:

1. Data-Alpha: both the amount of intake data and α value were manipulated to increase over time, in other words, the filtering rate and the parsimony bias decrease over time.
2. Data-only: the amount of intake data increases over time, but the parsimony bias stays constant at 1 - the highest value we have tested in this project.
3. Alpha-only: only the α value increases over time. The amount of intake data stays constant at 2000, which matches with the full input dataset we have.

2.4. Results

Inference proceeded by Markov-Chain Monte Carlo (MCMC) sampling, as implemented in Stan, for 5000 iterations with the initial 2500 samples discarded as burn-in. Trace plots indicated that all parameters settled on stable values within the burn-in period, therefore without loss of detail we present only average values over the remaining 2500 samples. The sampling run shown in Table 1b converged on three clusters, ordered in descending probability, that closely approximate the actual wh-question types in the training data in Table 1a (the other clusters inferred by the model had a total probability of 0.05 and are ignored here as noise). This run accurately classified 97.7% of the simulated question utterances on which it was trained, and 86.0% of the natural CDS test utterances. The main confusion in the test utterances was misclassification of echo questions as probe questions, which is also a mistake that children made in Nguyen and Legendre (2021) comprehension experiment.

Table 1. English questions based on CDS (a) and inferred by the model (b)

a)

Type	**%**	**WhFront**	**Inversion**	**WhDur (sd)**	**WhΔF0 (sd)**
Fronted	.84	1	1	150 (49)	-6 (36)
Probe	.09	0	0	208 (85)	-21 (61)
Echo	.07	0	0	254 (60)	108 (64)

b)

Cluster	**w**	$p_{WhFront}$	$p_{Inversion}$	μ_{WhDur} (σ)	$\mu_{Wh\ \Delta F0}$ (σ)
1 ≈ Fronted	.827	1.0	1.0	119 (48)	4 (35)
2 ≈ Probe	.097	0.03	0.03	203 (78)	-8 (68)
3 ≈ Echo	.076	0.01	0.01	270 (49)	120 (58)

Figure 1 reports the changes in the learning over time. In general, by slowly increasing the intake data and α to reflect developmental changes (Data:Alpha condition), the model displayed a regularizing pattern: initially the model only learned one cluster that showed the characteristics of fronted questions, then eventually expanded to two (fronted and in-situ). At the last two runs, with almost the full dataset, it was able to separate the two in-situ question types and learned all three clusters (fronted, in-situ echo, and in-situ information-seeking). The learned probability of fronted-questions was also initially boosted to almost 100% before stabilizing around the intake rate. The same pattern was observed in the two smaller in-situ clusters: when the model first learned to separate the two in-situ questions into cluster 2 and 3, initially the value of the more dominant PQs was boosted to ~0.12 before getting closer to the intake value at 0.097. This confirms a frequency-boosting pattern reported in previous papers. In the Data-only condition, the same trend emerged. However, compared to the Data:Alpha condition, fewer data points were required for cluster 1 to reach its target value and all three clusters to be learned. Finally, when only manipulating α in the Alpha-only condition, no effect was observed: all three clusters were learned at the same rate.

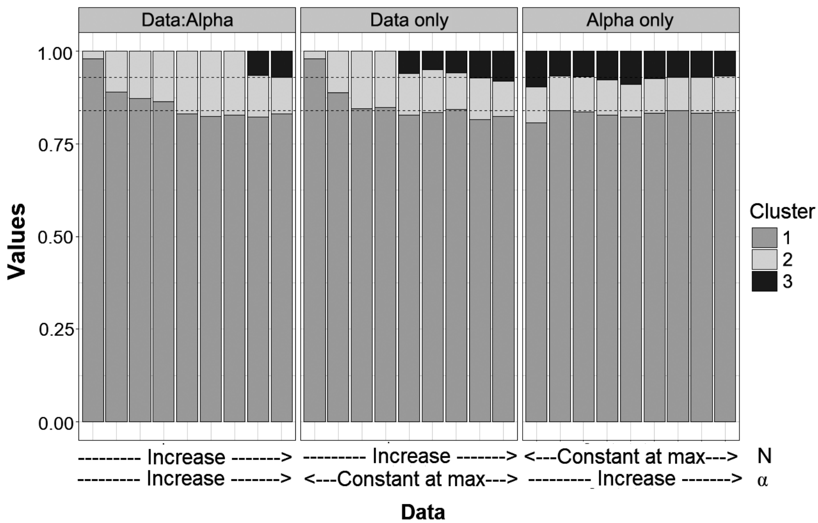

Figure 1. Learning over time in the three conditions

Note that the values of α as well as the amount of training data in this simulation were only used to represent abstract developmental changes and are not meant to be interpreted as precise values. The result in the Alpha-only condition certainly does not mean that a learner just needs to hear 2000 data points to learn different wh-question categories. Instead, the model is showing that different α values would require different (minimum) amounts of training data for the model to achieve learning. While parsimony bias alone may not be the primary motivation for regularizing, its interaction with intake quantity can capture the regularizing pattern in children.

3. Discussion

We have described a Bayesian model that learns to infer the number and characteristics of different wh-variants with fairly good accuracy. We also rely on the idea of an input filter to model the acquisition of wh-question variants. The model successfully captures the regularizing patterns often found when there is variation in the input using frequency and consistency as selecting factors, in line with previous studies (e.g., Hudson Kam and Newport, 2005, 2009; Schwab et al., 2018).

The interaction between the amount of intake data (N) and the learner's bias (α) shows the importance of intake quantity in regularizing. This is in line with Hendricks, Miller, and Jackson (2018)'s work on regularizing in Fering, a language with inconsistency in their gender-marking system. Specifically, Hendricks et al. find that bilingual children who are exposed to more Fering would have an adult-like pattern that preserves the inconsistency in their production, while children who are exposed to less Fering end up regularizing the inconsistent feature. In other words, the amount of intake can determine whether the learner would regularize their production or not. While the learner's bias alone may not induce a regularizing pattern, it can play a role during the learning process, as shown in the differences between the Data-only condition versus the Data-Alpha condition. The learner's bias can be used to account for cross-learner and cross-linguistic variation. A learner with a stronger parsimony bias may take longer to learn multiple variants of the same grammatical item compared to a learner with a weaker parsimony bias.

Finally, assuming that both comprehension and production utilize the same underlying learning mechanism but comprehension requires more cognitive load to master (Hendriks and Koster, 2010; Humphreys, 2012), the asymmetry between the two processes mentioned in the introduction could simply be because production requires more intake data and learning time than comprehension. Production thus 'lags' behind – when comprehension is in the three-cluster stage, production may still be in the one- or two-cluster stage.

4. Conclusion

The morphosyntactic properties considered in this project were manually coded and the prosodic properties were manually extracted. This makes the idealization that, first, these properties are always available to the learner in every instance of wh-questions and, second, they are always perceived correctly. In reality, both of these assumptions are likely to be violated. The availability and perception of these properties depend on many factors, including but not limited to the environment (e.g., noisy versus not noisy) and/or the attention of the listener. Some properties, such as duration, may be more prone to errors than others (Gussenhoven and Zhou, 2013). It is also possible that children misperceive some of their input, for example, mistakenly encoding a question with inversion as non-inversion.

Moreover, while the properties used in this project are representative, they are not comprehensive. We have only looked only at the prosody of the wh-word itself, but the intonation of the question at a sentence level can also bear important information for adults (Déprez, Syrett, and Kawahara, 2013). There are also pragmatic differences among the wh-variants. For instance, echo questions have a strong presupposition that the addressee knows the answer, and probe questions are typically used in more limited settings compared to fronted questions (see Biezma, 2020 for a more detailed discussion). We have limited our scope to morphosyntactic and prosodic properties because there has been no study that directly tests children's sensitivity to the pragmatic differences among questions types. However, to capture the full acquisition picture, the model should be expanded to additional linguistic dimensions, such as information structure and common ground.

While much work remains to be done to expand the range of variation that the model can accommodate, as well as to integrate the model with other aspects of syntax and to synergistically combine the learning of surface properties with that of semantics and pragmatics, the present results show the promise of adapting the idea of an input-filter mechanism and applying non-parametric Bayesian methods to model the acquisition of multiple wh-question variants.

References

Abend, Omri, Tom Kwiatkowski, Nathaniel Smith, Sharon Goldwater, and Mark Steedman. (2017). Bootstrapping language acquisition. *Cognition*,164:116–143.

Artstein, Ron. (2012). A focus semantics for echo questions. In ́Agnes Bende-Farkas and Arndt Riester, editors, *Workshop on Information Structure in Context*, 98–107. IMS, University of Stuttgart.

Biezma, Maria. (2020). Non-informative assertions: The case of non-optional wh-in-situ. *Semantics and Pragmatics, 13*, 18.

Boersma, Paul and David Weenink. (2019). Praat: doing phonetics by computer. [Computer program].

Cheng, Lisa and Johan Rooryck. (2002). Licensing wh-in-situ. *Syntax*, 3:1–19.

Déprez, Viviane, Kristen Syrett, and Shigeto Kawahara. (2013). The interaction of syntax, prosody, and discourse in licensing French wh-in-situ questions. *Lingua*, *124*, 4-19.

Dickinson, David and Patton Tabors. (2001). *Beginning literacy with language: Young children learning at home and school*. Paul Brookes Publishing, Baltimore, MD.

Fedzechkina, Maryia, Elissa Newport, & T. Florian Jaeger. (2017). Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive science, 41*(2), 416-446.

Gagliardi, Annie and Jeffrey Lidz. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language*, 90(1), 58–89.

Gass, Susan. 1997. *Input, Interaction, and the Second Language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gershman, Samuel and David Blei. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1):1–12.

Goldwater, Sharon. (2007). *Nonparametric Bayesian models of lexical acquisition*. Ph.D. thesis, Brown University.

Gussenhoven, Carlos, and Wencui Zhou. (2013). Revisiting pitch slope and height effects on perceived duration. In *INTERSPEECH 2013: 14th Annual Conference of the International Speech Communication Association*, 1365-1369.

Hendricks, Alison Eisel, Karen Miller, and Carrie Jackson. (2018). Regularizing unpredictable variation: Evidence from a natural language setting. *Language Learning and Development*, *14*(1), 42-60.

Hendriks, P, and Charlotte Koster. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, *120*(8), 1887-1897.

Hudson Kam, Carla, and Elissa Newport. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language learning and development*, *1*(2), 151-195.

Hudson Kam, Carla, and Elissa Newport. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, *59*(1), 30-66.

Humphreys, Gina. (2012). *Linking sentence production and comprehension: The neural mechanisms underlying production and comprehension control processes*. Doctoral dissertation, University of York.

Johnson, Mark. (2008). Unsupervised word segmentation for Sesotho using adaptor grammars. In *Proceedings of the Tenth Meeting of ACL Special Interest Group on Computational Morphology and Phonology*, 20–27.

Lee, Chia-ying, Timothy O'Donnell, and James Glass. (2015). Unsupervised lexicon discovery from acoustic input. *Transactions of the Association for Computational Linguistics*, 3:389–403.

MacWhinney, Brian and Catherine Snow. (1990). The child language data exchange system: An update. *Journal of Child Language*, 17:457–472.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. (2017). Montreal forced aligner: trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Conference of the International Speech Communication Association*, 498-502.

Nguyen, An, and Geraldine Legendre. (2021). The acquisition of wh-questions: Beyond structural economy and input frequency. *Language Acquisition*, 1-26.

O'Donnell, Timothy. (2015). *Productivity and reuse in language: A theory of linguistic computation and storage.* MIT Press.

Omaki, Akira, and Jeffrey Lidz. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition*, *22*(2), 158-192.

Sanborn, Adam, Thomas Griffiths, and Daniel Navarro. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, 117(4):1144–1167.

Schwab, Jessica, Lew-Williams Casey, and Adele Goldberg. (2018). When regularization gets it wrong: Children over-simplify language input only in production. *Journal of child language*, *45*(5), 1054-1072.

Seshadri, Shreyas , Ulpu Remes, and Okko Räsänen. (2017). Comparison of non-parametric Bayesian mixture models for syllable clustering and zero-resource speech processing. In *INTERSPEECH 2017*.

Sethuraman, Jayaram. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650.

Van Houten, Lori. 1986. Role of maternal input in the acquisition process: The communicative strategies of adolescent and older mothers with their language learning children. Paper presented at the *11th Annual Boston University Conference on Language Development*.

Weist, Richard and Andrea Zevenbergen. (2008). Auto-biographical memory and past time reference. *Language Learning and Development*, 4(4):291–308.

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

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