BUCLD 48 Proceedings
To be published in 2024 by Cascadilla Press
Rights forms signed by all authors

The Clustering Approach: An Input-Driven Approach to Parameter Setting

Alan Hezao Ke, Jingying Xu, and Lijun Ding

1. Introduction

Various models for parameter setting have been proposed within the Principles and Parameters (P & P) framework, among which we identify two main lines of research and call them the "grammar selection approach" and the "direct parameter setting approach". The former, exemplified by models like the Triggering Learning Algorithm (TLA, Gibson & Wexler 1994) model and the classical Variational Learner (VL, Yang 2002) model, views language acquisition as the selection of the target grammar, with the correct value setting of parameters, from all possible human grammars defined by UG. In contrast, the direct parameter setting approach, implemented in Fodor, Sakas, and colleagues' Structural Triggering Learner (STL, e.g., Fodor 1998, Sakas & Fodor 2012, Sakas et al. 2017), works with one single grammar and sets the values of parameters directly in the grammar. Showing that the two approaches complement each other, we introduce a new hybrid approach: the "Clustering Approach", built upon Yang's (2002) Naive Parameter Learner (NPL). Our goal is to demonstrate that an effective approach to modeling parametric variation in child language acquisition draws from three sources: language acquisition as grammar selection, language acquisition as direct parameter setting, and the understanding that parameters (or the pool of possible human grammars) are hierarchically structured by parametric clusters.

This paper is organized as follows. Section 2 reviews the direct parameter setting approach and the grammar selection approach, and makes a direct comparison of the two. Section 3 introduces our hybrid approach, the Clustering Approach. Section 4 presents the simulation results of the Clustering Approach in comparison with VL and NPL. Section 5 concludes the paper.

2. Grammar selection and direct parameter setting

2.1. The grammar selection approach

The grammar selection approach was implemented in various computational learning models, such as Gibson & Wexler's (1994) Triggering Learning Algorithm (TLA) and Yang's (2002) general Variational Learner (VL). Due to space

* Alan Hezao Ke, Michigan State University, `kehezao@msu.edu`. Jingying Xu, Michigan State University, `xujing21@msu.edu`. Lijun Ding, Texas A&M University, `lding@tamu.edu`.

constraints, we will only review Yang's VL here. The VL assumes a hypothesis space which consists of all possible human grammars. The learner's task is to navigate through this hypothesis space to identify the target grammar. The navigation is guided by the Linear Reward-Penalty Scheme (Bush & Mosteller 1951) as follows: A grammar G_i is sampled from the hypothesis space according to the probability distribution of the hypothesis space. Given an input sentence s, if G_i can analyze s, then the probability of G_i is increased, and meanwhile the probabilities of all other possible grammars in the hypothesis space are decreased. Otherwise, if G_i cannot parse s, decrease the probability of G_i and meanwhile increase the probabilities of all other possible grammars in the hypothesis space. When another input sentence is presented, a grammar that could be the same as or different from G_i is sampled to parse this new sentence. Again, the probabilities of all grammars in the hypothesis space are updated based on whether this grammar can parse s. This process continues until the probability of the target surpasses all competing grammars, signaling that the learner has learned the target grammar.

VL is a highly successful model for parameter setting, and it possesses several advantages. One is that it advances the line of research of integrating UG and statistical learning: UG provides a hypothesis space, while statistical learning guides the learner towards the target grammar. This learning algorithm has been demonstrated to consistently converge, given sufficient input (Straus 2008, Sakas et al. 2017). However, VL requires a substantial number of input sentences for the algorithm to converge on the target grammar, particularly when the number of parameters realistically reflects systematic cross-linguistic variations. Also, the model must keep track of the probabilities of all possible grammars in the hypothesis space and their updates. The number of competing grammars is determined by the number of parameters. The standard assumption is that all parameters are binary and thus each has two values. The number of possible grammar is equal to 2 raised to the power of k, where k is the number of parameters. As the number of parameters increases, the total number of possible grammar increases exponentially. This dramatic growth also makes tracking their probabilities challenging due to constraints on memory resources.

2.2. The direct parameter setting approach

The direct parameter setting approach has been implemented in Fodor and Sakas's STL models (e.g., Fodor 1998, Fodor & Sakas 2004, Sakas & Fodor 2012, Fodor 2017, Sakas et al. 2017; also see Howitt et al. 2021 for a model that incorporates many of STL's features). STL characterizes parameters and their values as UG-specified treelets, where a treelet is a sub-structure of a larger sentential tree. It assumes a direct mapping from the knowledge of parameters to parts of the structure in the parsing of actual input sentences. Language learners search a pool of all possible parametric treelets determined by UG. Once a treelet is used in online parsing, it joins the set of treelets employed for the target language, making it accessible for subsequent parsing. Consequently, whenever a new sentence is introduced for parsing, language learners need to search this set of treelets in use and find treelets that offer a proper structural analysis of that sentence. However, if the search into the set of treelets in use cannot find such treelets, then the parser must further search the pool of all possible parametric treelets to find new suitable ones for parsing the sentence. Another important property of STL is that the parser distinguishes between unambiguous and ambiguous triggers (cf. Roeper & Weissenborn 1990). Only unambiguous triggers are reliable cues for a specific parameter setting.[1]

The direct parameter setting approach enjoys important advantages. It significantly reduces the size of the hypothesis space compared to the grammar selection approach. This method no longer requires a search space encompassing all possible human grammars. Instead, the size of the hypothesis space is linearly proportional to the number of parameters. In the case of STL in which each parameter has two potential values, and each value corresponds to a specific treelet, the number of possible treelets in the hypothesis space is $2 \cdot n$, where n is equal to the number of parameters. For instance, when 13 parameters are considered and each has 2 value options, there are 26 treelets in the parametric treelet pool to select from.

However, the direct parameter setting approach also faces some significant challenges. First, to analyze a given input sentence, STL must search both the set of parametric treelets currently in use and potentially the entire pool of parametric treelets permitted by UG. Second, to determine whether a specific trigger is unambiguous, the parser must apply each of the treelets (including treelets in use and all possible treelets in the pool) in the analysis of every given input sentence. This ensures that no more than one set of treelets can successfully analyze a given sentence. If otherwise more than one set of treelets are applicable, the trigger will be instead identified as ambiguous. Furthermore, since there may be instances where individual treelets cannot fully analyze a sentence on their own, various combinations of treelets must be tried to achieve a successful analysis. Therefore, the STL approach generally necessitates an extensive search through *all* parametric treelets. It's crucial to emphasize that such a search is seen as overly taxing for the parser, especially during online parsing, where cognitive resources are expected to be limited.

It is important to note that the challenges faced by the direct parameter setting approach, particularly the requirement for extensive searching during online parsing, can be addressed by the grammar selection approach. The grammar selection approach avoids the extensive search challenge during online parsing noted in the direct parameter approach because the learner selects the grammar with a full set of parameters with their hypothetical values. Thus, any possible grammar selected from the grammar pool can be directly used as the grammar that guides

1 Different variants of the STL models deal with ambiguous input differently. Strong STL can learn from both unambiguous and ambiguous triggers. Weak STL implements serial parsing and it learns only from unambiguous triggers, ignoring all ambiguous triggers.

the parser. There is no need to learn the parameter values during parsing since all the parameter values have been presupposed or hypothesized in possible grammars. In addition, the grammar selection approach does not distinguish ambiguous triggers from unambiguous triggers and make use of both for learning. As we will explore further, computational models based on grammar selection can effectively set parameters without needing this differentiation, given sufficient input sentences (Sakas et al. 2017).[2]

In sum, both the grammar selection and the direct parameter setting approach face significant challenges, making neither of them an optimal solution within the parameter setting framework. However, they seem to counterbalance each other's shortcomings. The issues of the grammar selection approach are resolved by the direct parameter setting approach, and vice versa. This observation gives hope for a more ideal model. By effectively merging these two approaches, we might harness the strengths of both while bypassing their limitations. We will delve into such hybrid approaches in the subsequent section.

3. The hybrid approach

As pointed out in the last section, the direct parameter-setting approach and the grammar selection approach actually complement each other, suggesting the potential for an integrated, hybrid approach. In this section, we will introduce a precursor of our hybrid approach, Yang's (2002) Naive Parameter Learner (NPL), and our Clustering Approach.

Yang's NPL reserves grammar sampling for parsing from grammar selection but meanwhile adopts the update of the probabilities of parameters instead of possible grammars. Specifically, like STL, NPL sets parameters directly and individually. Each time an input sentence is encountered, a grammar comprising a list of parameter-value pairs is sampled to parse it. This is how the NPL approach addresses the challenges associated with the grammar selection approach: it reduces the search space and does not need to update and track the probabilities of possible grammars but only the probabilities of individual parameters.

However, an important concern with NPL is that as the amount of the possible grammars increases, (e.g., with more than 10 independent parameters), NPL, just like VL, also faces an extended period of exploration: identifying the target grammar that always correctly parses the input sentences is challenging with a random search into a vast hypothesis space. For example, the probability of finding the right grammar out of a hypothesis space generated by 20 parameters is $\frac{1}{2^{20}}$. This issue is inherent to all models that assumes grammar selection for parsing but updates all parameters in the selected grammar upon successful parsing. The issue is not merely the amount of input sentences the model requires to converge on the target grammar; more importantly, it implies that the learner does not set

2 However, Pearl (2011) and Nazarov et al. (2021) show that VL (and NPL) encounters difficulties learning all stress patterns from ambiguous input given in Dresher's (1999) corpus.

any parameters for a considerable amount of input sentences while exploring the hypothesis space. This prolonged exploration period is a consequence of the difficult random sampling of the target grammar from the grammar pool, irrespective of whether the learner has the competence to process language. When a non-target grammar is selected, it may either successfully parse the given input sentence or fail to do so. If successful, a list of parameters with incorrect values, differing from those in the target grammar, will be rewarded. If it fails, a list of parameters that may be part of the target grammar are penalized. The introduction of more parameters into the target grammar will simply result in greater confusion.

Such an issue is significant as we assume in the learning model that children are equipped with a parser capable of parsing the linguistic input, given a hypothesized grammar. This predicts that even if children already have the competence for parsing, they cannot learn parameters from the input simply because it is almost impossible to select the target grammar out of a big grammar pool. To compound this issue, a lack of parameter setting could persist even when the learner has successfully parsed a significant number of sentences. Such a prediction seems counter-intuitive and empirically unsupported, as "most (but not all) parameters are acquired fairly early" (Yang 2002: 46, also see Thornber & Ke to appear). The Clustering Approach attempts to address this issue. It posits that only the parameters used in the input sentence will be updated upon a successful parse, and only a limited number of parameters (=1 in this paper) that function as sampling parameters will be penalized upon a parsing failure. In what follows, we will detail the main mechanisms of the Clustering Approach.

First, the Clustering Approach adopts the idea that parameters are hierarchically structured (Baker 2008, Biberauer & Roberts 2015, Roberts 2019), sharing significant properties with those models that rely on parameter learning orders. The learning orders can be due to constraints from UG that limit access to certain types of input (Lightfoot 1989) or learning order biases, stemming from either innate factors or input-related distributional information, such as those applied in models for acquiring phonological parameters (Dresher 1999, Pearl 2011). In the Clustering Approach, these parametric clusters are learned in order, specifically, from hierarchically higher parameters to lower ones, with the hierarchy being determined completely by the input. Figure 1 depicts this hierarchical organization, representing an acquisition stage where two parameters have been set: P1 = 0 and P2 = 1. With these two parameters set, the grammar pool has shrunk to a quarter of its original size. That is, the learner at this point can disregard all possible grammars where P1 = 1 and P2 = 0. The learner's subsequent goal is to identify which next parameter can serve as a clustering criterion. This decision is entirely driven by the input: the parameter that has a probability *Prob*(P=0) whose *absolute value* of *Prob*(P=0) - 0.5 is the highest (P3 in Figure 1) will be prioritized and moved to the top of the list as the focus of learning. As will be detailed later, grammar selection from the current grammar pool is guided by the probabilities of P3=0 and P3=1. This is because grammar selection involves choosing a possible grammar from either the cluster of grammars with P3=0 or the cluster with P3=1. The probabilities of P3 (and other parameters detected in the input if the parsing is successful) will then be updated in the learning process. If a specific value of P3 consistently receives predominant and consistent support from the input, leading to a probability nearing 0 or 1 (or surpassing a threshold), P3 will become the subsequent parameter serving as a criterion to divide the grammar pool into two clusters: one with the probability set at 0, and the other at 1. The parameters that have been set were set in the same way. If a parameter is not used in a specific language, its probability will remain close to its initial value of 0.5, causing it to be pushed to the bottom of the hierarchy. While these parameters are present, they act as if they are irrelevant to the target grammar. In practice, they remain inactive, seeming as though they are not part of the learner's mental representation of the target grammar, which is the desired outcome. Such parameters do not increase the number of possible grammars for learning. In this sense, parameter setting in the Clustering Approach is driven by the input.

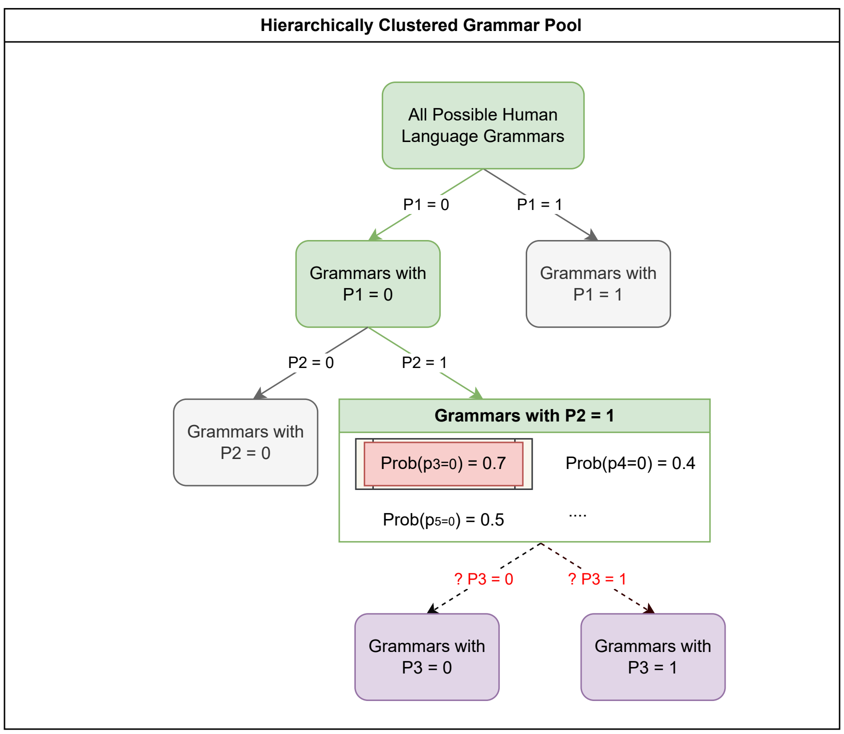

Figure 1: Grammar pool clustered by parameters that are hierarchically ordered

Before we proceed, it would be helpful to clarify the use of several terms

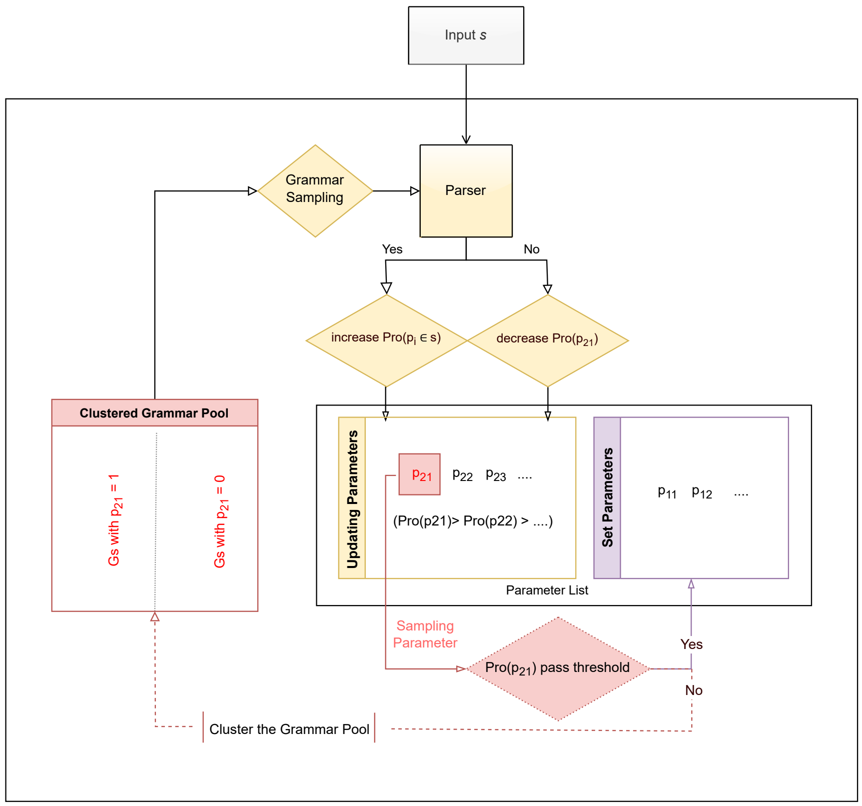

Figure 2: Parameter setting in the Clustering Approach

within the Clustering Approach. All parameters are embedded in the parser. Only the parameters that have been utilized in successful parsing will be collected into the set of *updating parameters*. We refer to the parameters that have been used to parse a specific sentence as the *parameters in use*. Only the parameter with the most extreme probability among the updating parameters will be employed for clustering the current grammar pool and sampling a grammar from the grammar pool to parse the input. This parameter is termed the *sampling parameter*. Once a parameter reaches thresholds, it will be classified into the *set parameters*, and its probability will no longer be updated. The set parameters reduce the size of the grammar pool as the possible grammars with different values than the set values are excluded from the future sampling process.

As a hybrid approach, the Clustering Approach incorporates many mechanisms from the NPL. Figure 2 presents a flowchart illustrating the Clustering Approach. We assume a grammar pool that consists of all possible grammars determined by or derived from UG or other innate knowledge. The size of this grammar pool corresponds to the number of all possible combinations of parameter values. Therefore, the initial size of the grammar pool is equal to what is assumed in both the VL and NPL models. Under the Clustering Approach, the grammar pool is hypothetical, designed solely to generate a possible grammar for the parser. It serves no other function beyond this. As such, it is "light-weighted."

Like the NPL, the learner samples a grammar from the grammar pool and uses that grammar to parse a given input sentence. This sampling process is guided by the probability of the sampling parameter and the set parameters. For example, as shown in Figure 2, if set parameters include two parameters P11=0 and P12=1, all the possible grammars of P11=1 or P12=0 will be excluded from the grammar pool. The updating parameters include P21, P22, and P23, with P21 having the most extreme probability value (closest either to 0 or 1). Therefore, P21 will be the sampling parameter at this point. The sampling parameter will be dynamically updated after each round of update of the parameters in use. A grammar will be sampled from the cluster of possible grammars with P21=0 and another cluster of possible grammars with P21=1. The probability of sampling a particular cluster is determined by the probability of P21=0 or P21=1. We also avoid tracking the probabilities of all possible grammars. The Clustering Approach, similar to the STL but different from NPL, tracks only the probabilities of updating parameters.

If the parser successfully constructs a syntactic tree for an input sentence by applying the parametric treelets associated with the selected grammar, we consider the parsing successful. In such instances, the learner identifies which parameters and their values have been utilized. identification is feasible due to STL's assumption that the parametric values can be mapped to treelets in a structure and vice versa.[3] These parameters are added to the set of updating parameters, if they are not already included. The probabilities of the updating parameters that are used in the parsed sentence are updated based on the Linear Reward-Penalty Scheme, which is also implemented in NPL.

It is important to note that the Clustering Approach updates the probabilities of only those parameters that are used in a parsed sentence, a key aspect of its input-driven nature. Via this process, a series of sampling parameters will be identified based on their consistent and frequent usage in the input sentences, leading to a learning order that is input-informed. This selective updating is a distinctive feature of the Clustering Approach, setting it apart from NPL, which updates all parameters in the selected grammar upon a successful parse.

If parsing fails, we assume that the parser discredits only the parameter that is used for sampling. The intuition behind the assumption is that if a part of an input sentence cannot be successfully parsed, then the parser should be able to know which part of the sentence causes a problem. The parser can then guess whether changing certain parameter values would render the sentence parsable.

3 This mapping may involve analytical ambiguity, which is outside the current scope of consideration.

For example, if the sequence <object, V> cannot be parsed by a "head-initial" parameter, then the parser should know if the parameter value were changed to head-initial, this sequence could be parsed successfully. Under this assumption, if the parsing failure is due to some of the current parameters in use, then the parser should decrease these parameters' probabilities only.

After a certain number of iterations, the probability of some specific parameter might reach the upper threshold (e.g., 0.9), causing the parameter to be set to 1, or it might fall under the lower threshold (e.g., 0.1), leading the parameter to be set to 0.[4] Once a parameter, P_i, has its value set to either 1 or 0, it will be utilized to cluster the grammar pool. Consequently, the grammar pool is grouped into two clusters: one where P_i is set to 1 and the other where it is set to 0. In future grammar sampling, only the cluster that has a P_i value consistent with the one set in the learning process will be considered. This effectively reduces the size of the grammar pool for grammar sampling by half, eliminating from consideration all grammars with the incorrect p_i value. As more parameters are set in their values, the grammar pool quickly shrinks. The target grammar emerges if the parameters are set according to their value in the target grammar.

4. Simulations with synthesized data

A number of simulations are conducted to explore the properties of the Clustering Approach in the learning of artificial languages, in comparison to two closely related benchmark models, namely, the VL and NPL. Recall that the primary goal is to verify whether the Clustering Approach can address a crucial issue in VL and NPL: namely, that with a larger number of parameters, the learner does not begin learning the target grammar or setting its parameters until after an extended period of exploration (even in the case where the learner has successfully parsed a significant number of input sentences). On the contrary, what we would like to see is that, even with a large hypothesis space, at least some parameters are learned earlier and overall the parameters are gradually learned in sequence. For example, the parameters that are frequently observed in the input should be learned earlier than the parameters that occur less frequently.

We construct a few corpora of synthesized sentences as part of the input to the model. The corpora are of m sentences, where $m = cn/2$, with the integers c representing a constant number and n the number of parameter. In the simulations with 12 parameters or less, we set c to be 5,000. That is, 8 parameters will have $cn/2 = 5000 \times 8/2 = 20000$ sentences. c is increased to 8000 for 16 and more parameters.

The sentences in the corpora are partitioned to groups with 4 + 2 + 2 + 2...+ 2, if more than 4 parameters are used. This is to create sentences with a variety of lengths. Each sentence is of the form $s_i = (p_j)_{j \in I_i}$, where $p_j \in \{0, 1\}$ is the value of the parameter j and $I_i \subset \{1, \ldots, n\}$ is the set of all the parameters in

4 The threshold can be adjusted to gauge the probability of setting the parameter to an incorrect value.

the sentence s_i. We construct eight artificial corpora of sentences which includes 2, 4, 8, 12, 16, 20, 24, and 28 parameters. The distribution of the parameters in sentences approximates a Zipf distribution (Zipf 1949), with some parameters occurring more frequently than others. The length of a sentence is defined by the number of parameters it consists. The sentences are created by randomly combining the parameters in the range of parameters.[5]

Figure 3 contrasts the Clustering Approach (represented by the solid line) with the VL model (dashed line) and the NPL model (dash-dotted line). The latter two act as benchmarks in the computational simulations. Figure 3 results from 400 iterations, with a learning rate set to 0.05 for all three approaches. The target grammar probability results are averages over the results obtained from all iterations. We set the high-bound and low-bound threshold to 0.9 and 0.1. For the Clustering Approach, this implies that if a parameter with a certain value has a probability higher than 0.9 or lower than 0.1, it is considered settled, and it will be used to cluster the grammar pool. Consequently, only the cluster that has the parameter value set accordingly will be used for grammar sampling for future parsing of the input. This reduces the size of the grammar pool by half, and thus increases the chance of identifying the target grammar from the grammar pool in the future.

These results suggest that the Clustering Approach method is indeed distinctive in terms of its beginning to learn from the input early and steadily converges to the target grammar. In addition, the number of input sentences the model requires for convergence does not increase too much with the increase in parameter number. The Clustering Approach can learn the parameters with significant fewer input sentences, especially when the parameter number is increased, for example, to 12 and more. This is due to a crucial property of the Clustering Approach: it learns the parameters that are used in the input and the parameters are learned in an order informed by the input.

By contrast, although the VL and NPL learners perform as well as the Clustering Approach when the parameter number equals 2 and 4, the number of sentences the learners require for them to even begin learning the target grammar increases exponentially as the number of parameters increases. Imagine that when the parameter number is increased to 40 (see, for example, Longobardi 2018), VL will require too many input sentences for the algorithm to actually start setting parameters. NPL will face a similar challenge. This is because when the parsing of an input sentence is successful with a selected grammar, the algorithm rewards all parameters in that grammar, including those not used in the input. This ultimately adds noise to the probabilities of the parameters and thus hinders learning. Again, the primary challenge for VL and NPL is not necessarily the quantity of sentences required to successfully learn the target grammar. By contrast, an important challenge for VL lies in the vast hypothesis space; with many possible grammars, learners may struggle to identify the target grammar. In addition, updates to the

5 See the shared scripts for detailed specifications of the corpora: `https://osf.io/qwef3/?view_only=461e1d47c7ca4980a1a3fc3460b6f20c`.

parameters are not accurately targeted at those that require updating; instead, the algorithm updates all parameters indiscriminately. This approach can lead to prolonged exploration periods during which parameters remain essentially unchanged, even though the learner has successfully processed a considerable number of input sentences. Note that the current simulations have simplified the learning process as the input corpora do not include noises and ambiguous sentences. This is reflected in the short actual learning periods, evident after the probability of the target grammar begins to increase. Such observation further highlights the significance of the problematic "flat" learning period. For example, VL does not learn the target grammar at all from input derived from a corpus generated by 12 parameters, whereas NPL stops setting any additional parameters of the target grammar after learning most of the parameters (between 8-10) under the 12 parameter condition.

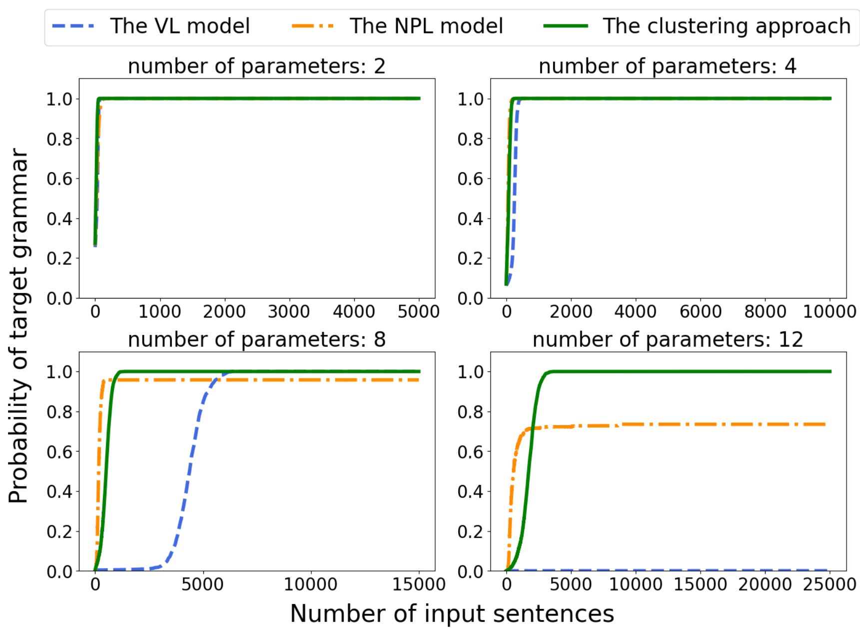

Figure 3: Comparing the Clustering Approach and Yang's (2002) VL and NPL modeled on synthesized data with 2, 4, 8, and 12 parameters.

To compare further the performance of the Clustering Approach and the NPL at a synthesized language with more parameters, we simulated their learning of languages with 16, 20, 24, and 28 parameters. Figure 4 is based on the same setting as the previous simulations. The results confirm that a hypothesis space generated by 16 to 28 parameters is challenging for the NPL model, but much less so for the Clustering Approach.

Another important aspect of the Clustering Approach is its emergent prop-

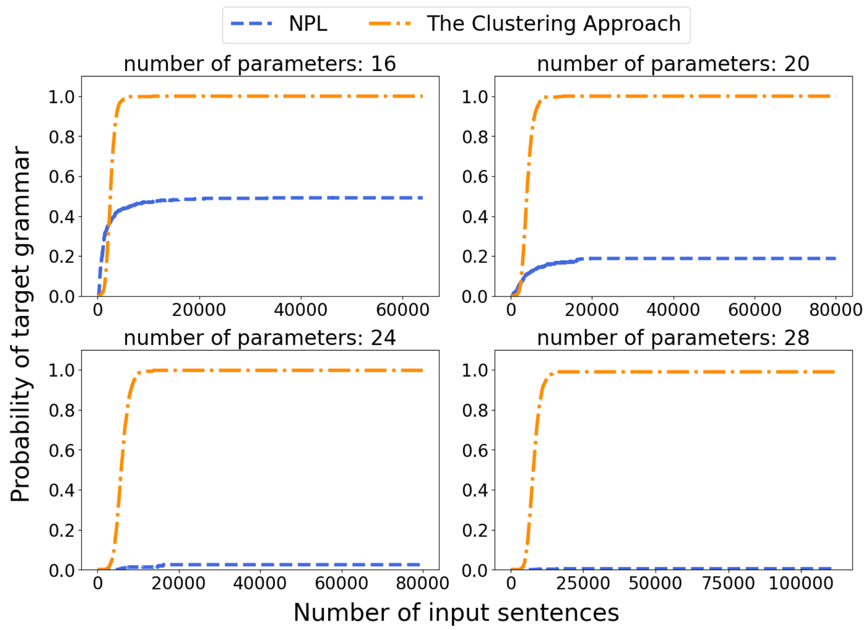

Figure 4: Comparing the Clustering Approach and Yang's (2002) NPL modeled on synthesized data with 16, 20, 24, and 28 parameters.

erty. This is related to a potential concern that for some parameters, their triggers might be so rare in the input that they might never reach either the upper or lower thresholds. The Clustering Approach allows parameters that are most frequent and consistent in the input to be set first, and the parameters which rarely occur or associate with significant variation will be set later. This is confirmed in empirical studies (cf. Legate & Yang 2007). Could some of the parameters which are extremely rare in the input be left unset in the end? The model's answer is yes. This suggests that children are uncertain about some parameters even though they have learned many other parameters. This predicts that in experimental studies, we should be able to observe some uncertainty with regard to those parameters (see e.g., Ke & Gao 2020). Which parameters are they? One of the central goals of our ongoing project is to identify the learning trajectories of different parameters, which has the potential to further test the empirical predictions of the Clustering Approach in experimental studies.

Finally, while the Clustering Approach consistently outperformed the NPL model in our simulations, it may not always successfully learn the target grammar under the following circumstances: (i) some parameters are too infrequent (e.g., 0.010 to 0.015 of the input) when the parameter number is increased to 20 and more; (ii) the corpus includes a fair amount of long sentences that consist of more than 20 parameters; (iii) the number of parameters is increased but the number of input sentences stays the same. Since these restrictions are reasonable constraints for a computational model of language acquisition, they do not raise immediate concerns but warrant further research.

5. Conclusions

In this paper, we have briefly reviewed some representative approaches to parameter setting for child language acquisition and made a contrast between the grammar selection approach and the direct parameter setting approach. A significant merit of the direct parameter setting approach is that it does not assume a vast hypothesis space. Instead, it tracks a single grammar, comprising a set of parameters that are set directly based on relevant structural information extracted from the primary linguistic data. Yet, its primary drawback is the necessity for identify relevant parametric treelets as well as ambiguous and unambiguous triggers during online parsing, presupposing an exhaustive search during online parsing.

By contrast, the grammar selection approach assumes the learner searches through a hypothesis space of possible grammars to identify the target grammar based on the learner's analysis of input sentences. Since the grammar selection approach assumes that the parser adopts the grammar that is sampled from the hypothesis space and applies it in the analysis of input sentences, it no longer needs to initiate immense searches to find appropriate parameters/parametric treelets for the analysis of the given input sentences. However, the grammar selection approach requires a significant number of input sentences to successfully navigate a vast hypothesis space. Interestingly, it is observed that the direct parameter setting approach addresses these issues directly: navigating through all possible human grammars becomes unnecessary when parameters are set directly.

A significant observation of this paper is that these approaches seem to complement each other in their major aspects. Consequently, we propose a hybrid method, the Clustering Approach, that integrates the grammar selection and direct parameter setting approaches, following the NPL model. The Clustering Approach envisions a pool of possible grammars without necessitating the tracking and evaluation of each; on the other hand, its algorithm sets parameter probabilities directly and individually, eliminating the need for extensive online searches, thanks to the grammar sampling from the pool during parsing. The Clustering Approach allows some parameters that are frequently and consistently observed in the input be set first and thus serve as clustering criteria. This dynamic nature of the Clustering Approach distinguishes itself from the NPL model. Based on the simulation results for the Clustering Approach and NPL, we contend that the NPL instead predicts the learner would gain minimal knowledge about the target grammar during the initial phase of exploration unless the target grammar is accidentally identified. Hence, the Clustering Approach offers a more realistic, input-driven model for the learning of parametric variation, paving the way for modeling child language acquisition with real linguistic data.

References

Baker, Mark C. 2008. *The atoms of language: The mind's hidden rules of grammar*. Basic books.

Biberauer, Theresa & Ian Roberts. 2015. Rethinking formal hierarchies: A proposed unification. *Cambridge occasional papers in linguistics* 7(1). 1–31.

Bush, Robert R & Frederick Mosteller. 1951. A mathematical model for simple learning. *Psychological Review* 58(5). 313.

Dresher, Bezalel Elan. 1999. Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30(1). 27–67.

Fodor, Janet Dean. 1998. Learning to parse? *Journal of Psycholinguistic Research* 27(2). 285–319.

Fodor, Janet Dean. 2017. Ambiguity, parsing, and the evaluation measure. *Language Acquisition* 24(2). 85–99.

Fodor, Janet Dean & William Gregory Sakas. 2004. Evaluating models of parameter setting. In *the proceedings of the 28th Annual Boston University Conference on Language Development*, vol. 1, 1–27. Cascadilla Press Somerville, MA.

Gibson, Edward & Kenneth Wexler. 1994. Triggers. *Linguistic Inquiry* 25(3). 407–454.

Howitt, Katherine, Soumik Dey & William Gregory Sakas. 2021. Gradual syntactic triggering: The gradient parameter hypothesis. *Language Acquisition* 28(1). 65–96.

Ke, Alan Hezao & Liqun Gao. 2020. Domain restriction in child mandarin: Implications for quantifier spreading. *Linguistics* 58(6). 1839–1875.

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

Lightfoot, David. 1989. The child's trigger experience: Degree-0 learnability. *Behavioral and Brain Sciences* 12(2). 321–334.

Longobardi, Giuseppe. 2018. Principles, parameters, and schemata: a radically underspecified UG. *Linguistic Analysis* 517–558.

Nazarov, Aleksei, Gaja Jarosz et al. 2021. The credit problem in parametric stress: A probabilistic approach. *Glossa* 6(1). 1–26.

Pearl, Lisa S. 2011. When unbiased probabilistic learning is not enough: Acquiring a parametric system of metrical phonology. *Language Acquisition* 18(2). 87–120.

Roberts, Ian. 2019. *Parameter hierarchies and universal grammar*. Oxford University Press.

Roeper, Thomas & Jürgen Weissenborn. 1990. How to make parameters work: Comments on Valian. In *Language processing and language acquisition*, 147–162. Springer.

Sakas, William Gregory & Janet Dean Fodor. 2012. Disambiguating syntactic triggers. *Language Acquisition* 19(2). 83–143.

Sakas, William Gregory, Charles Yang & Robert Berwick. 2017. Parameter setting is feasible. *Linguistic Analysis* 41. 391–408.

Straus, Kenneth Jerold. 2008. *Validation of a probabilistic model of language acquisition in children*: Northeastern University Boston Ph.D. dissertation.

Thornber, Molly & Alan Hezao Ke. to appear. Modeling the learning of syntactic parameters from parsed data. In *the proceedings of BUCLD 48*, MA: Cascadilla Press Somerville.

Yang, Charles D. 2002. *Knowledge and learning in natural language*. New York: Oxford University Press.

Zipf, George Kingsley. 1949. *Human behavior and the principle of least effort*. Addison-Wesley Press.