The Clustering Approach
An input-driven approach to parameter setting

Alan Hezao Ke[1], Jingying Xu[1], & Lijun Ding[2]

[1]Michigan State University, [2]University of Wisconsin-Madison

Introduction

Two approaches on the computational modeling of child language acquisition:

Grammar selection vs. Direct parameter setting

- **The grammar selection approach**
  e.g., Gibson & Wexler's Triggering Learning Algorithm, Yang's Variational Learner (VL):
  - Language learner identifies the target grammar with the correct set of parameters by navigating through the hypothesis space with all possible human grammars defined by UG.

  Main challenge:
  - The requirement to track a large hypothesis space and slow convergence.
- **The direct parameter setting approach**
  e.g., Fodor & Sakass Structural Triggering Learner models (Fodor 1998a, Sakas & Fodor 2012)
  - Language learner tracks a single grammar.
  - The parameters are learned directly and individually by parsing of linguistic input.

  Main challenge:
  - The need of extensive search during the online parsing of each sentence to identify (un)ambiguous triggers.

The hybrid approach

Critical observation. These two approaches complement each other:

- The former effectively learns the target grammar in the presence of ambiguous triggers, thereby avoiding extensive searching during online parsing.
- The latter tracks the probability of a single grammar, thus free of tracking the whole grammar pool.

→ We should look for an integrative, hybrid approach.

A precursor of the hybrid approach: Yang's (2002) Naive Parameter Learner (NPL)

Parameters are learned individually, but grammar selection is still used for parsing the input.

- **Challenge**: When the parameter number increases, the grammar pool expands exponentially: this increases the difficulty of finding the target grammar (Nazarov et al. 2021).

The clustering approach (CA)

CA adopts the hybrid approach from NPL and uses the same probability updating rules for parameter learning.

Important differences between CA and NPL:

- CA assumes a hierarchically structured pool of possible human grammars, clustered by parameters and their values (see Fig. 1 and Fig. 2; cf. Roberts 2019).

→ This achieves the structured organization of the parameters as proposed by Fodor (1998b) without the need to assume default parameter values.

- The parameter with the most extreme probability (closer to either 1 or 0) becomes a (dynamically updated) *sampling parameter that clusters the grammar pool.*
- A grammar is selected from the clustered grammar pool to parse a given input sentence.
- If parsing succeeds, the probabilities of all parameters *used in the parsing* will increase.
- If parsing fails, the probability *of the sampling parameter for clustering* will decrease.
- Once a sampling parameter is set, the grammar pool is reduced by half, significantly shrinking the hypothesis space for future grammar selection.
- *Input driven*: the sampling parameter is ultimately determined by the input: it is the parameter that is most consistently detected in the input.

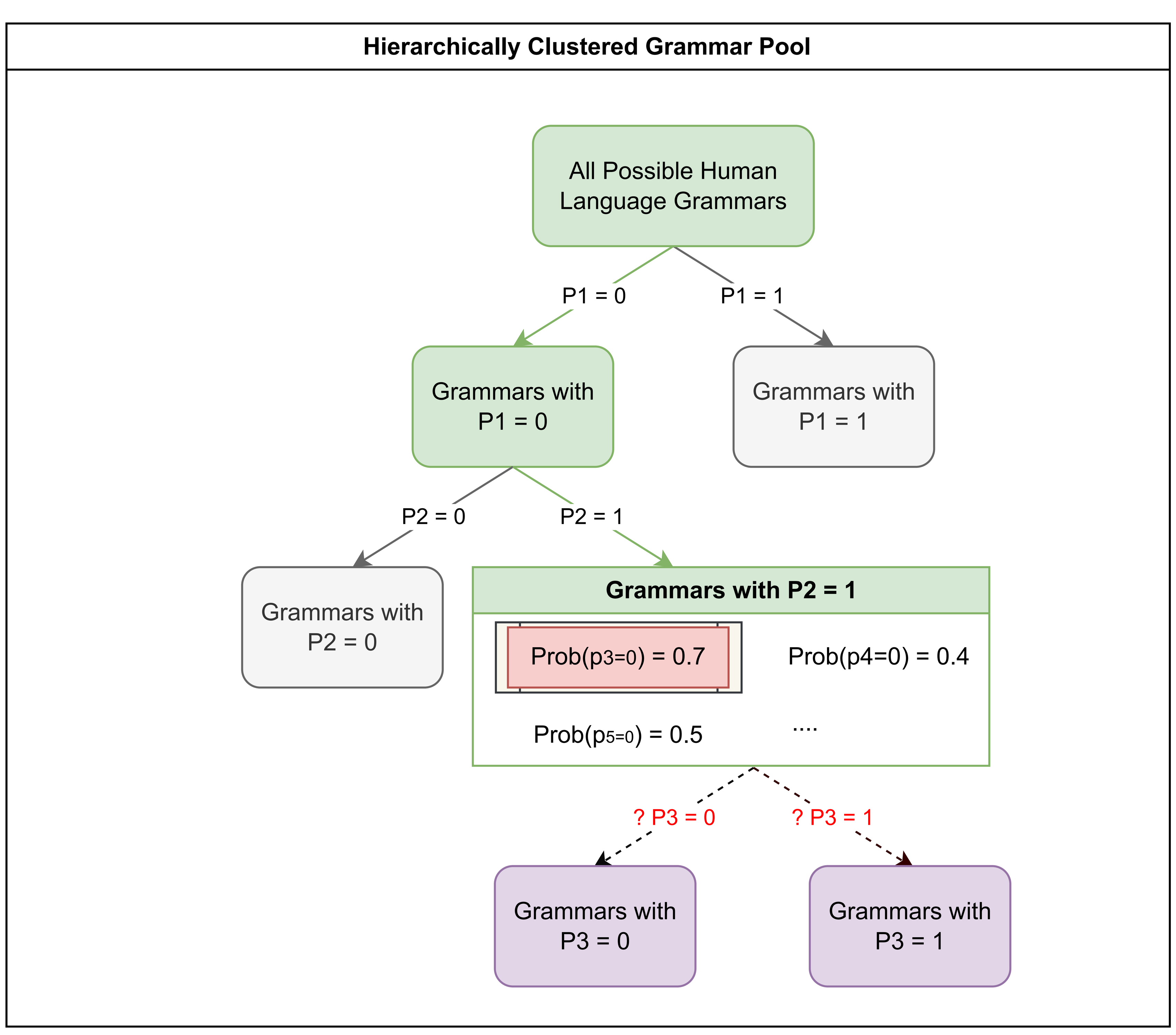

Figure 1. Grammar pool clustered by parameters that are hierarchically ordered

Selected References

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

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

Fodor, Janet Dean. 1998b. Parsing to learn. *Journal of Psycholinguistic research* 27(3). 339–374.

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

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

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

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

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

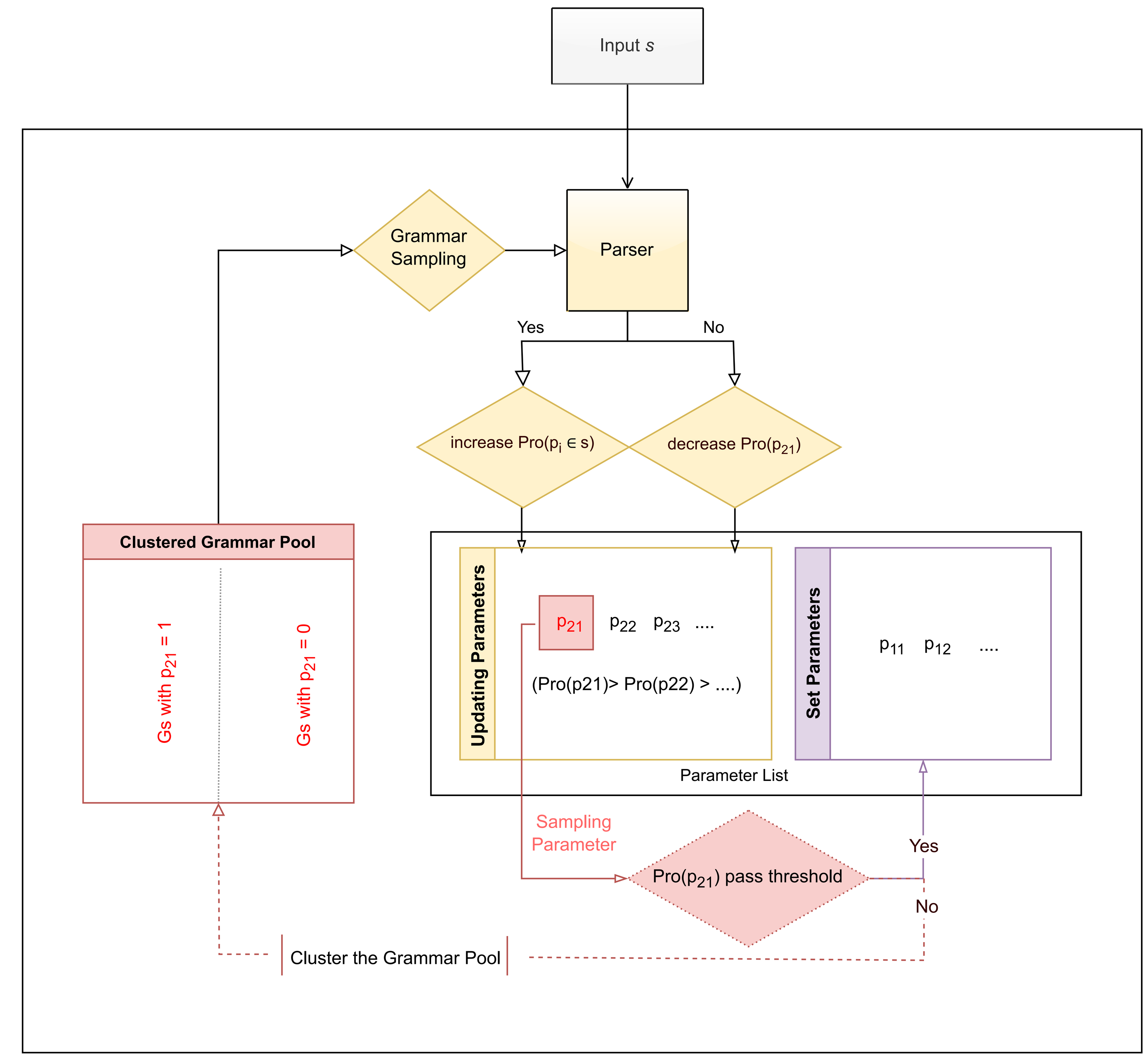

Figure 2. Parameter setting in the clustering approach

Simulations

We conducted simulations for CA, in comparison with VL and NPL with synthesized data as input (see the setup of the computation modeling in Table 1).

Input sentences	Synthesized sentences (from languages with 2, 4, 8, or 12 parameters respectively) with different lengths, depending on the amount of parameters (out of 2, 4, 8, and 12 parameters) that are used to generate these sentences. The frequency of the parameters in the corpora follows a Zipfian distribution.
Settings	Learning rate = 0.05, threshold = 0.9, Iteration = 400

Table 1. Critical specifications of the computational models

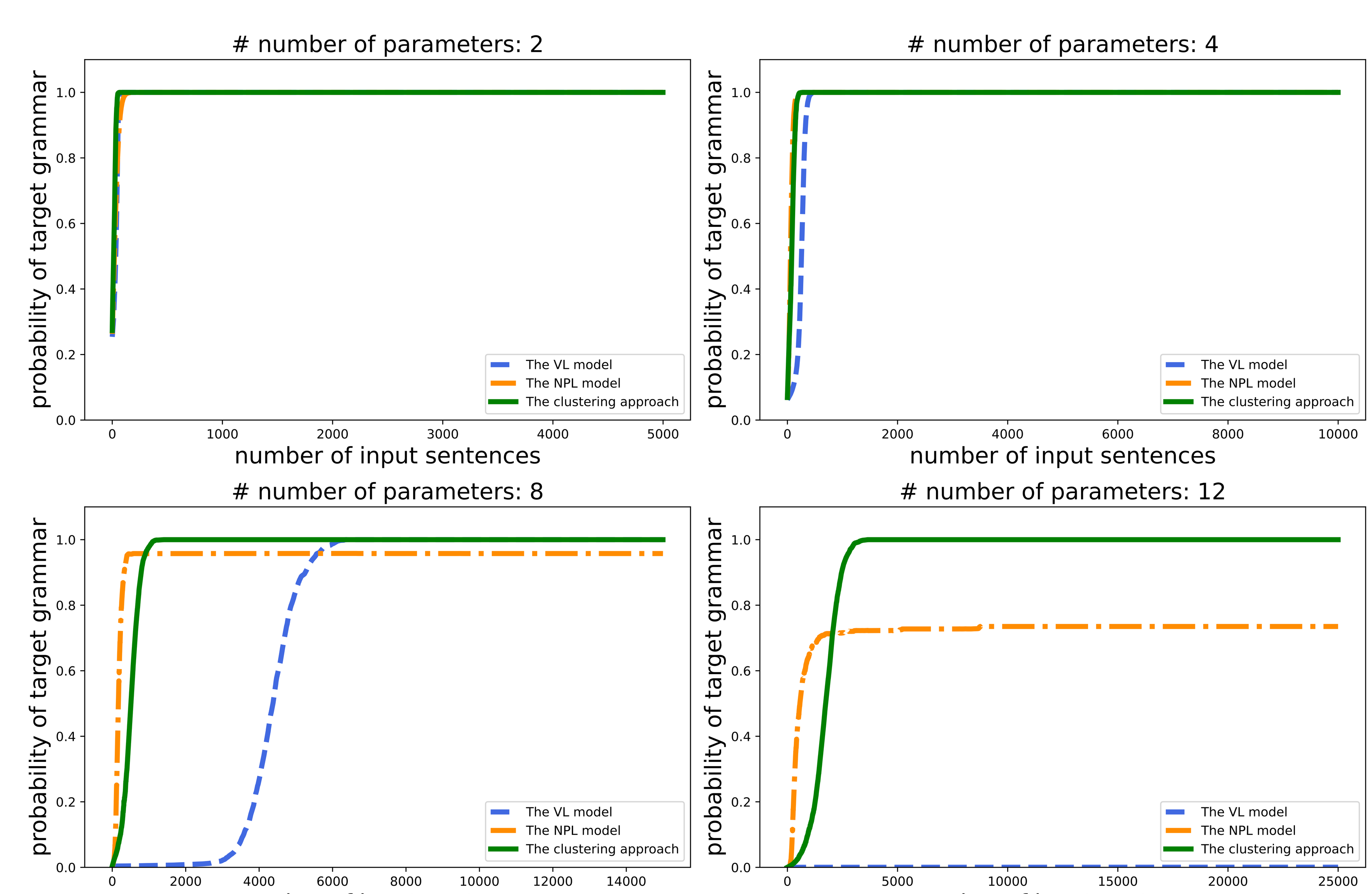

Figure 3. Comparison of VL, NPL, and CA on synthesized data

Results

- The three approaches do not differ when learning simple grammars with 4 or fewer parameters.
- VL predicts a lack of learning for a significant portion of the overall learning process when dealing with 12 parameters.
- NPL encounters challenges when learning the target grammar with 12 parameters. As the number of parameters further increases, it will face similar problems as VL.
- In contrast, CA demonstrates gradual or successful parameter learning in these cases.

Conclusions

- CA not only addresses the challenges of the grammar selection approach and the direct parameter setting approach, but also offers computational evidence that parameter setting remains a viable framework worthy of further exploration.
- This study supports an input-driven approach that aligns with an emergentist perspective on the acquisition of syntactic parameters (Biberauer & Roberts 2015, Roberts 2019).
- This study provides a computationally feasible method for hierarchically structuring the parameters based on input information.