

Weighted CKY Algorithm

Miguel Ballesteros

Algorithms for NLP Course.
7-11

CarnegieMellon

CKY

- Kasami, 1965
- Younger, 1967
- Cocke and Schwartz, 1970

CKY Algorithm

- Cocke-Kasami-Younger algorithm.
- Recognition vs. Parsing:
 - Recognition - deciding the membership in the language
 - Parsing – Recognition+ producing a parse tree for it
- Parsing is more “difficult” than recognition (time complexity)
- **CKY**: bottom-up dynamic programming.

Probabilistic Context Free Grammars

- PCFG

- $G = (V, T, P, S, Q)$

- V : a finite set of variables, non-terminal symbols.

- T : a finite set of terminal symbols (equiv. To Σ in FSAs)

- P : a set of context free production rules, each of the form

- $A \rightarrow \alpha$, where $A \in V$, $\alpha \in (V \cup T)^*$

- S : a start non-terminal $S \in V$

- $Q: V \rightarrow [0, 1]$

Remembering ϵ -productions

- Formally, context-free grammars are allowed to have empty productions (ϵ = the empty string):

VP \rightarrow V NP

NP \rightarrow DT Noun

NP $\rightarrow \epsilon$

- These can always be eliminated without changing the language generated by the grammar:

- The grammar above becomes

– VP \rightarrow V NP

VP \rightarrow V ϵ

NP \rightarrow DT Noun

- The second production rule does not make a lot of sense, then ...

VP \rightarrow V NP

VP \rightarrow V

NP \rightarrow DT Noun

Remembering Chomsky Normal Form

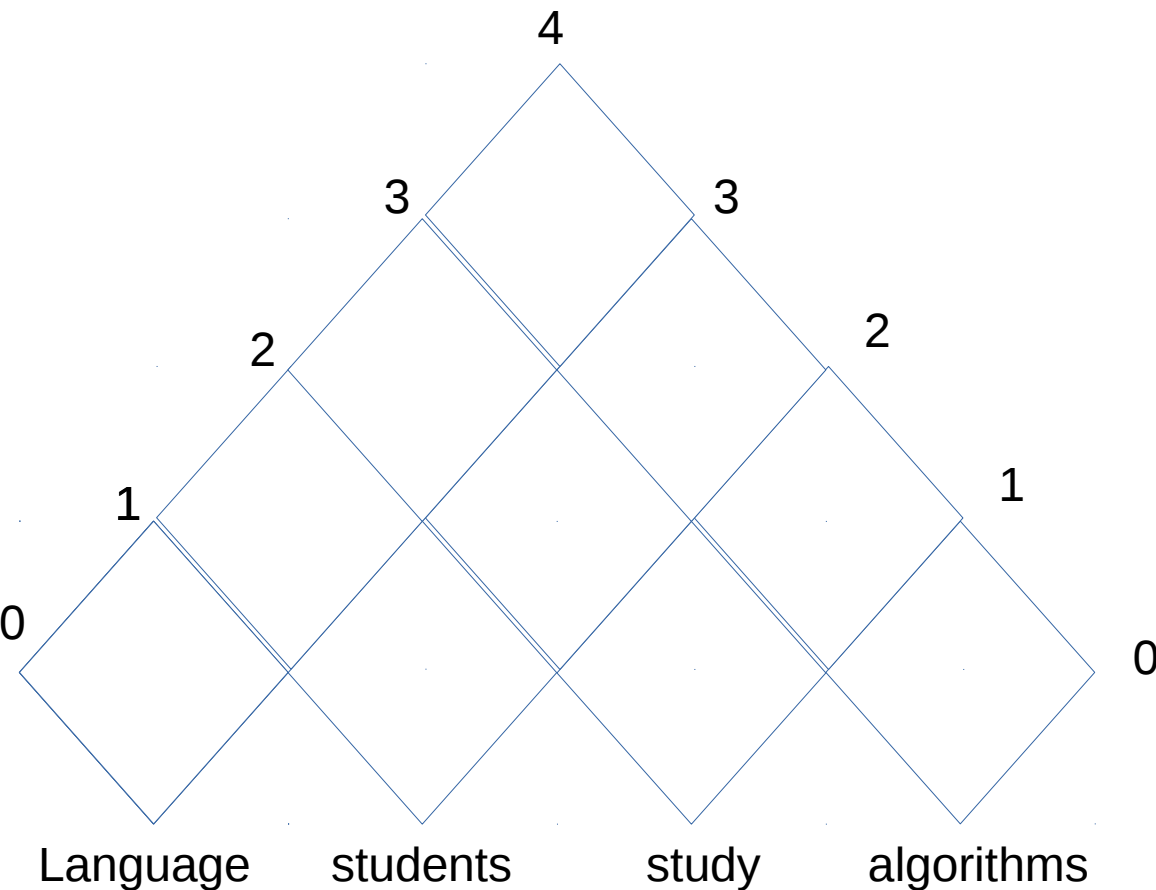
- Each production is of the form:
 - (i) $A \rightarrow BC$
 - (ii) $A \rightarrow a$

There you have your “binary branching”

- We saw that Any context-free language is generated by a context-free grammar in CNF.

Phrase structure parsing CKY algorithm

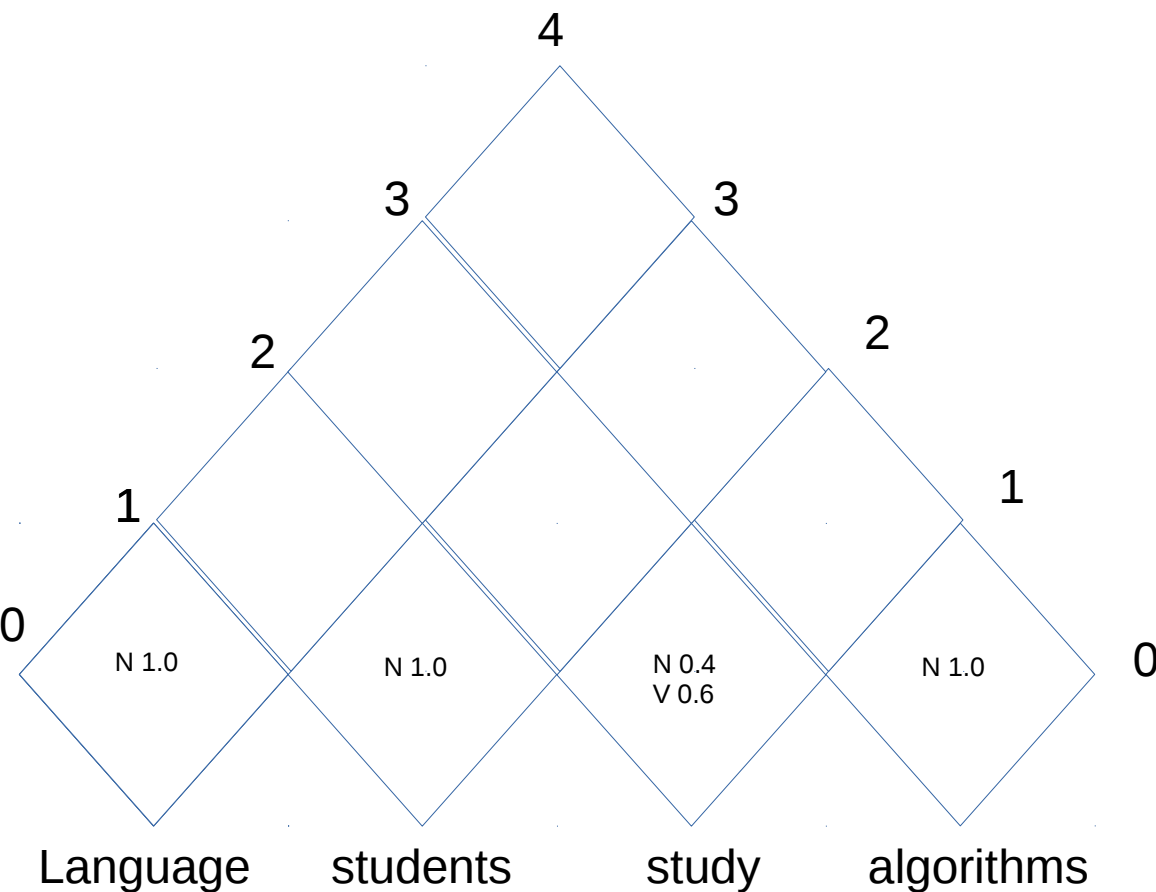
- Bottom-up dynamic programming by building a chart/triangle.



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow \text{students}$ 1.0
- $N \rightarrow \text{study}$ 0.4
- $V \rightarrow \text{study}$ 0.6
- $N \rightarrow \text{Algorithms}$ 1.0
- $N \rightarrow \text{Language}$ 1.0

Phrase structure parsing CKY algorithm

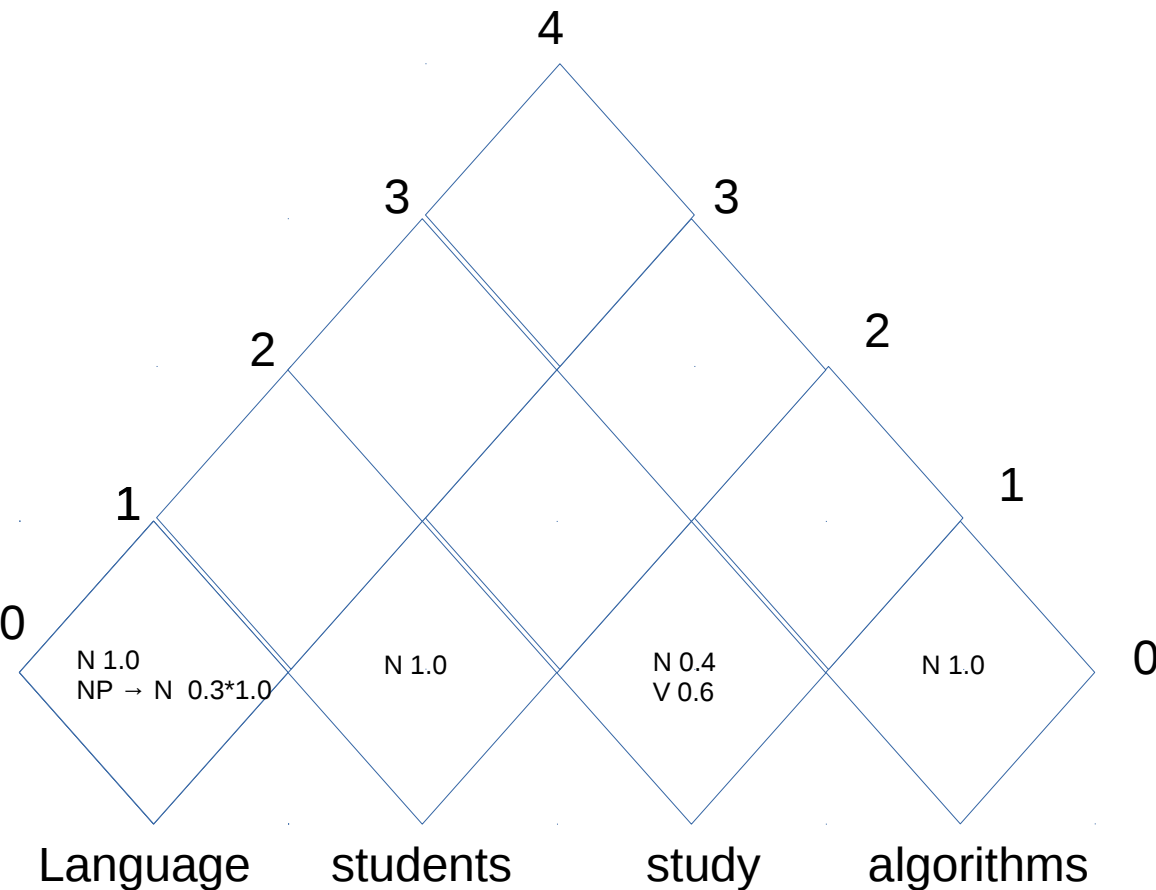
- First, lexicon rules.



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

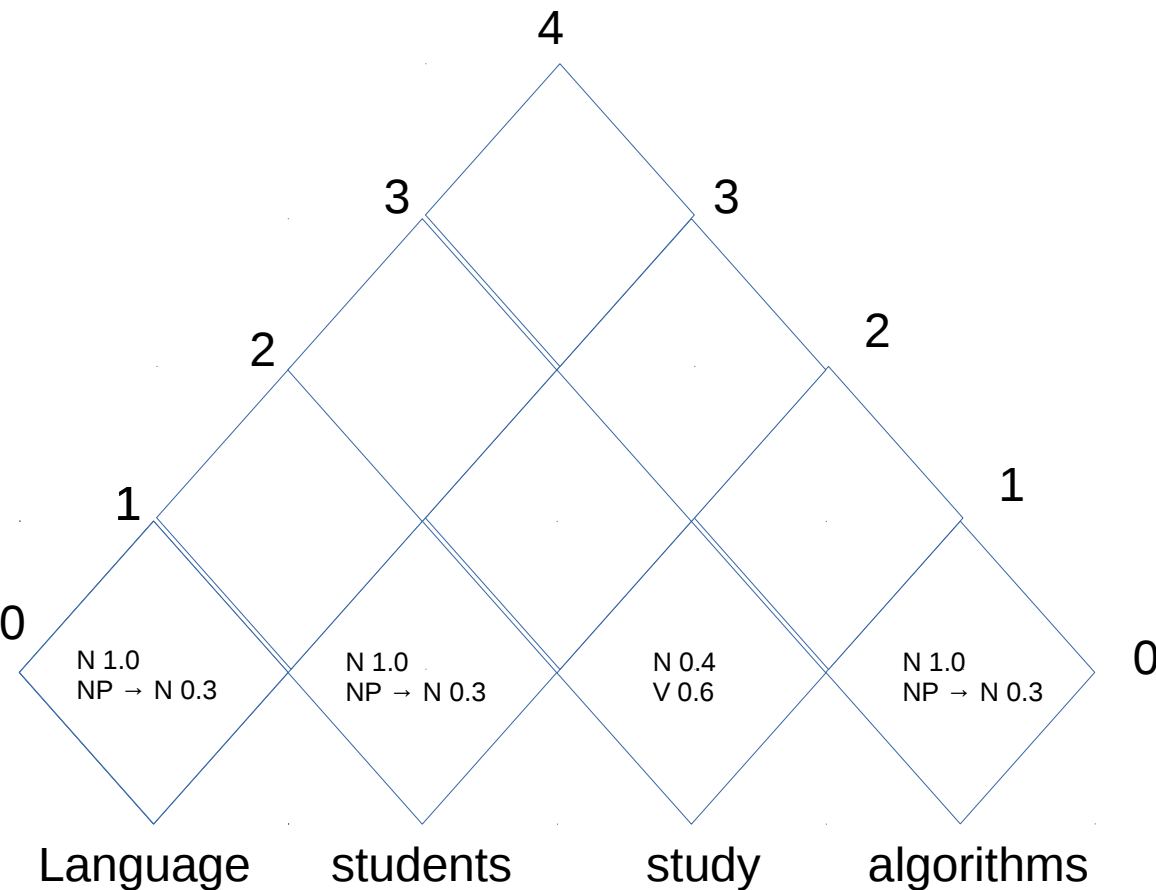
- Unary rules.



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow \text{students}$ 1.0
- $N \rightarrow \text{study}$ 0.4
- $V \rightarrow \text{study}$ 0.6
- $N \rightarrow \text{Algorithms}$ 1.0
- $N \rightarrow \text{Language}$ 1.0

Phrase structure parsing CKY algorithm

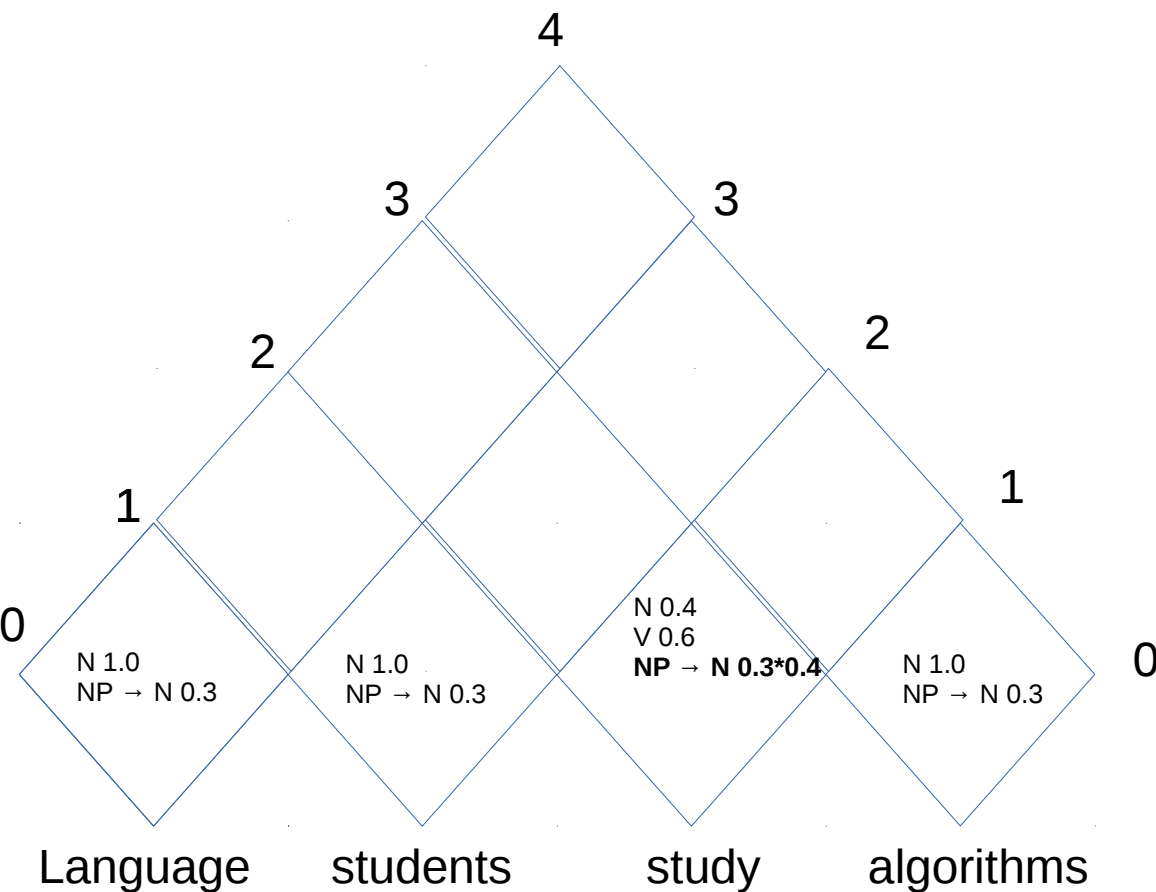
- Unary rules.



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

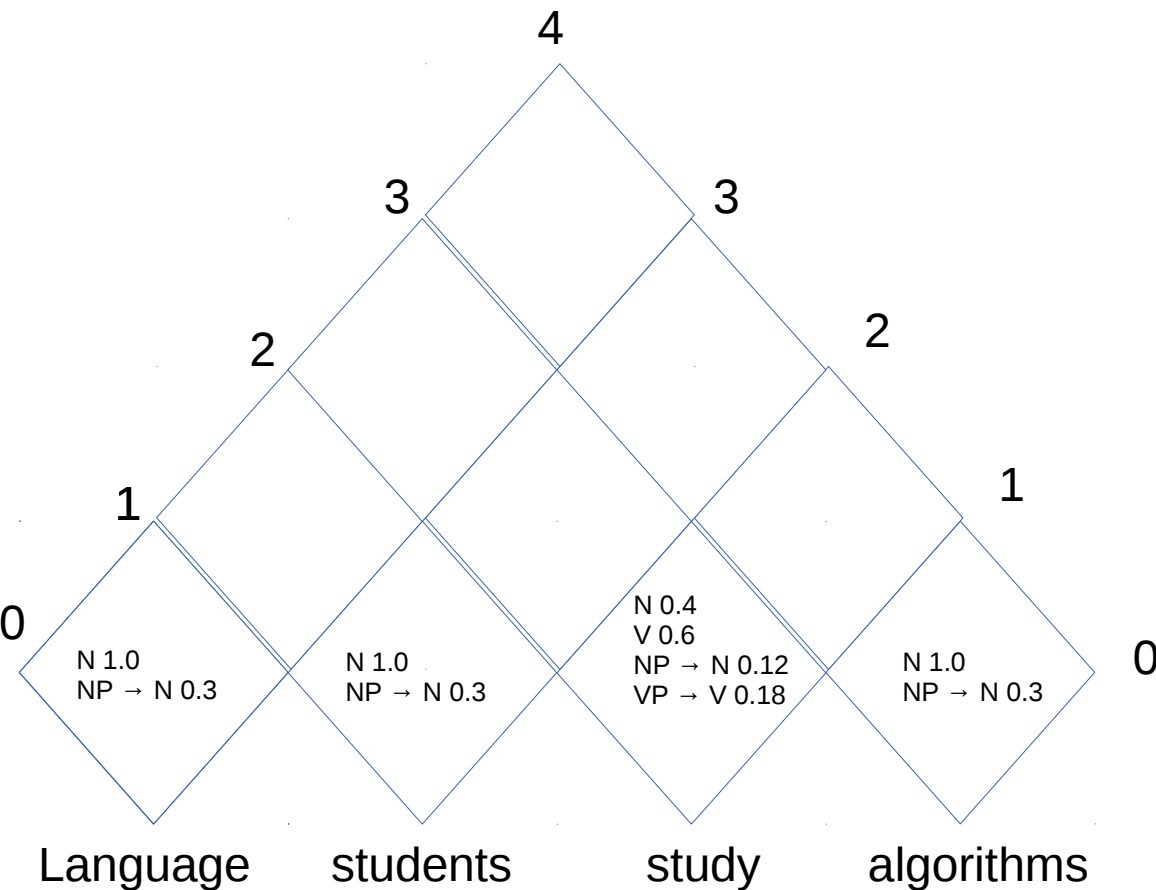
- Unary rules.



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow \text{students}$ 1.0
- $N \rightarrow \text{study}$ 0.4
- $V \rightarrow \text{study}$ 0.6
- $N \rightarrow \text{Algorithms}$ 1.0
- $N \rightarrow \text{Language}$ 1.0

Phrase structure parsing CKY algorithm

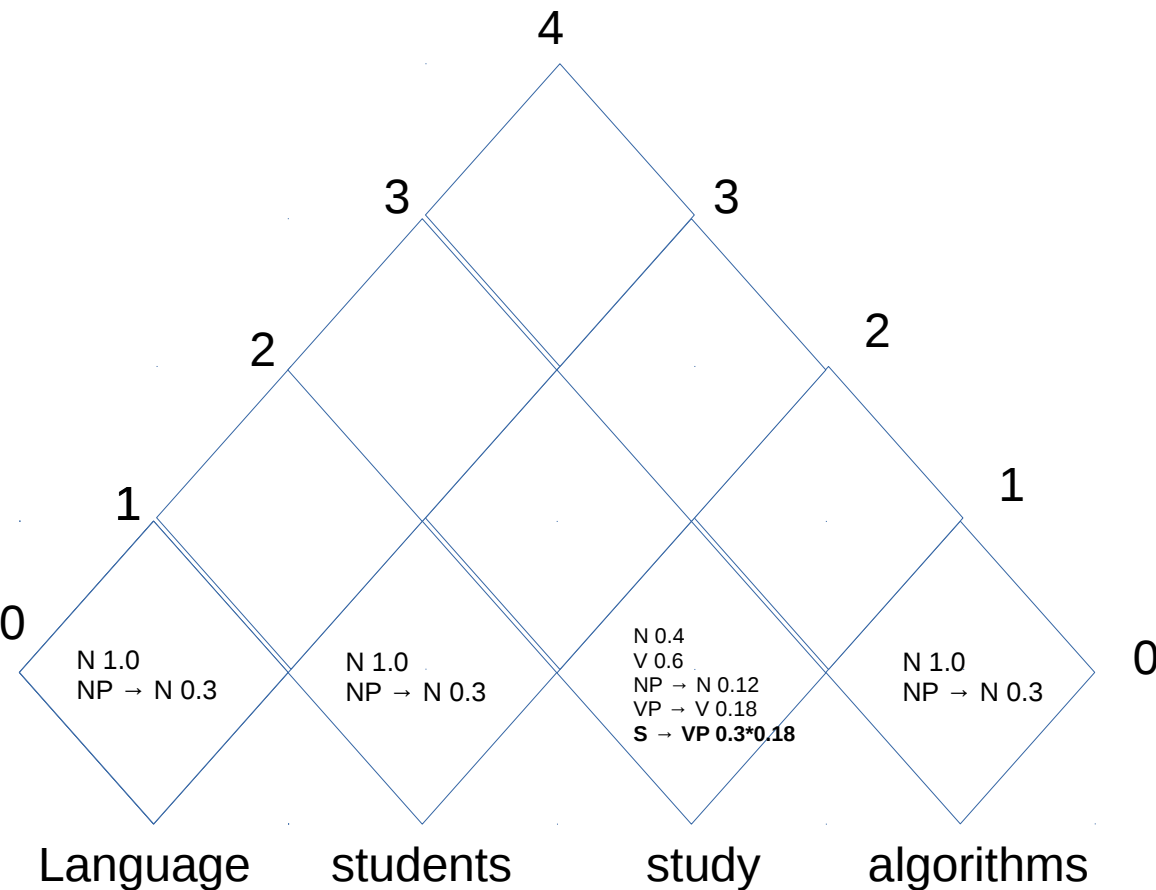
- Unary rules.



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

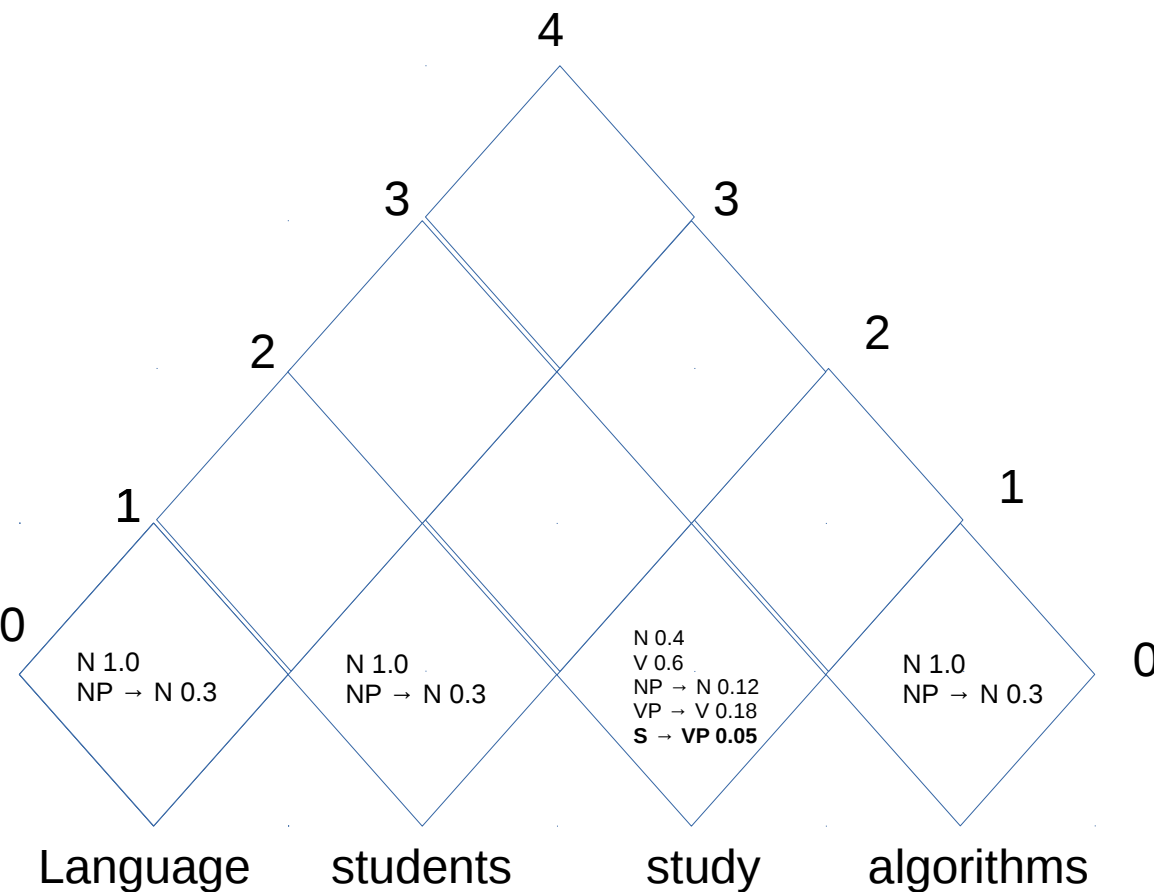
- More unary rules ?



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

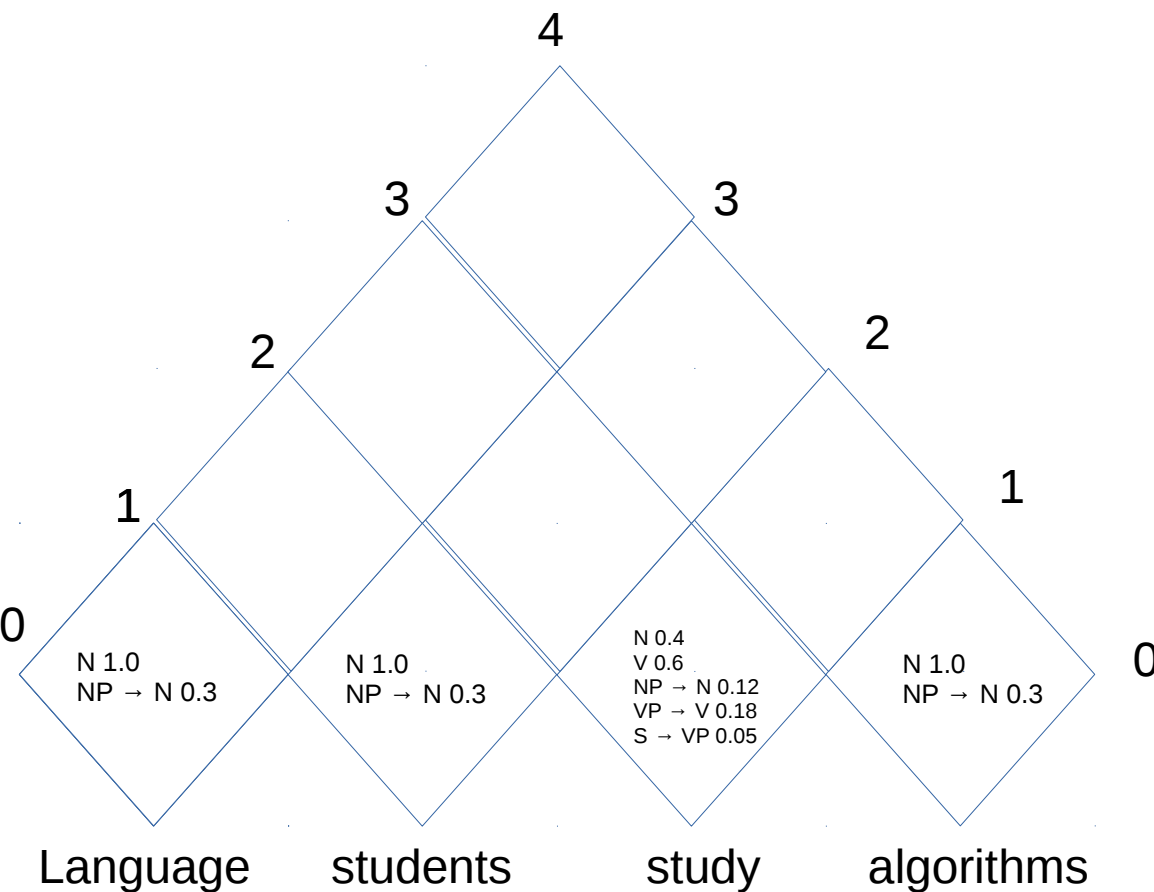
- More unary rules ?



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

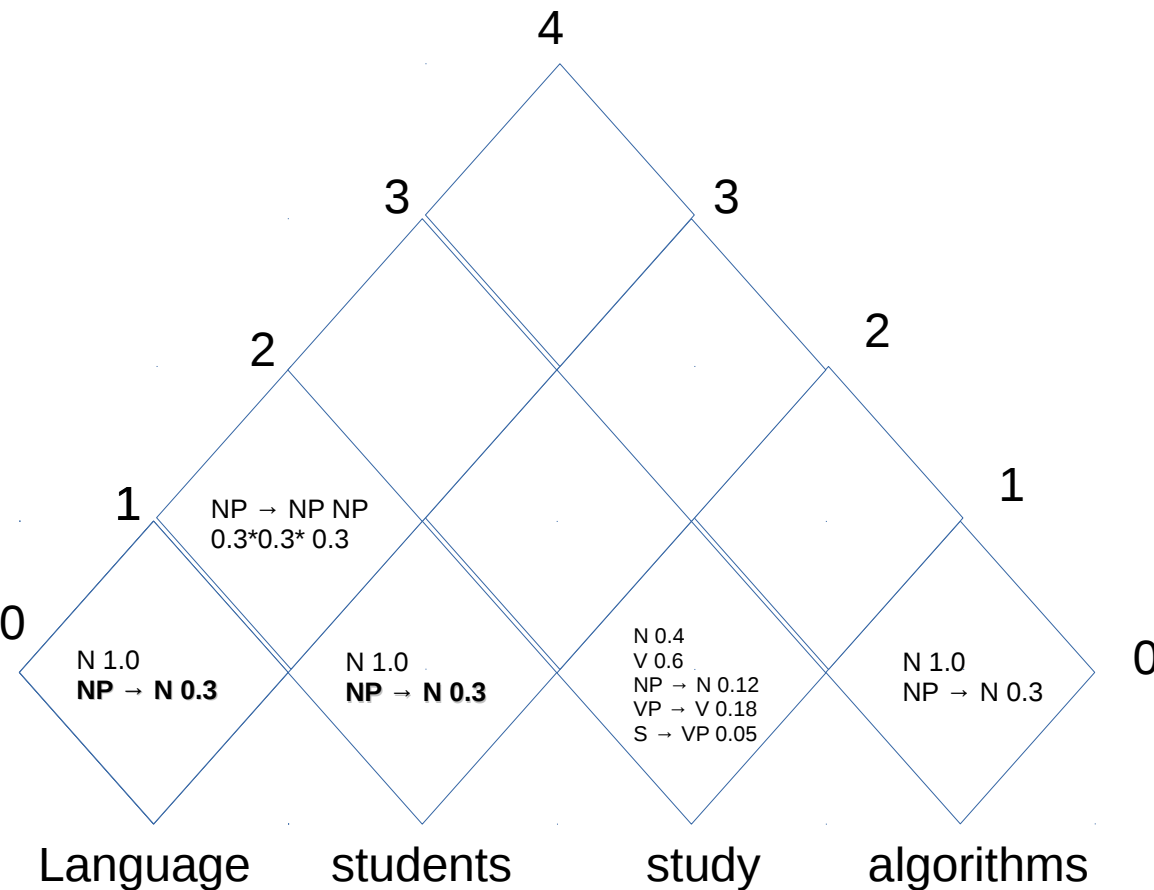
- Now, binary rules



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

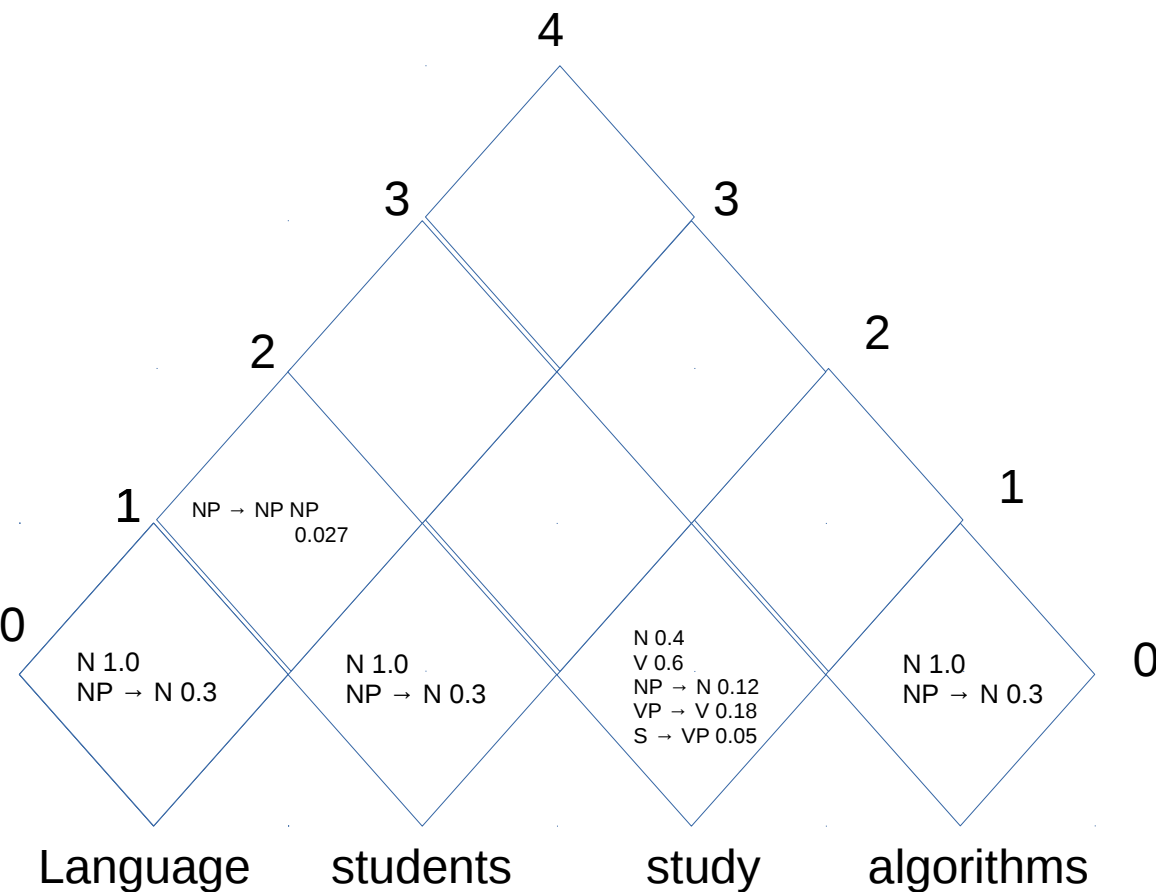
- Now, binary rules



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- **$NP \rightarrow NP NP$ 0.3**
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow$ students 1.0
- $N \rightarrow$ study 0.4
- $V \rightarrow$ study 0.6
- $N \rightarrow$ Algorithms 1.0
- $N \rightarrow$ Language 1.0

Phrase structure parsing CKY algorithm

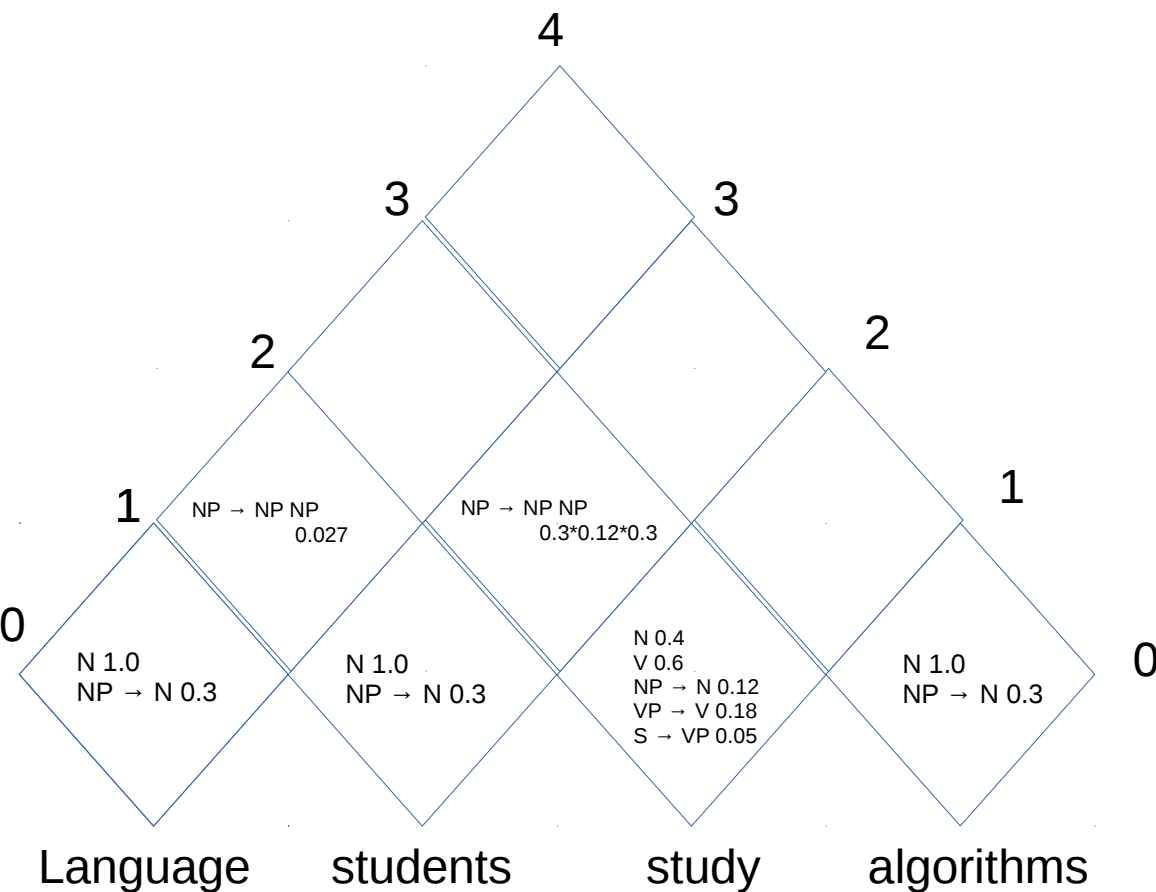
- Now, binary rules



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

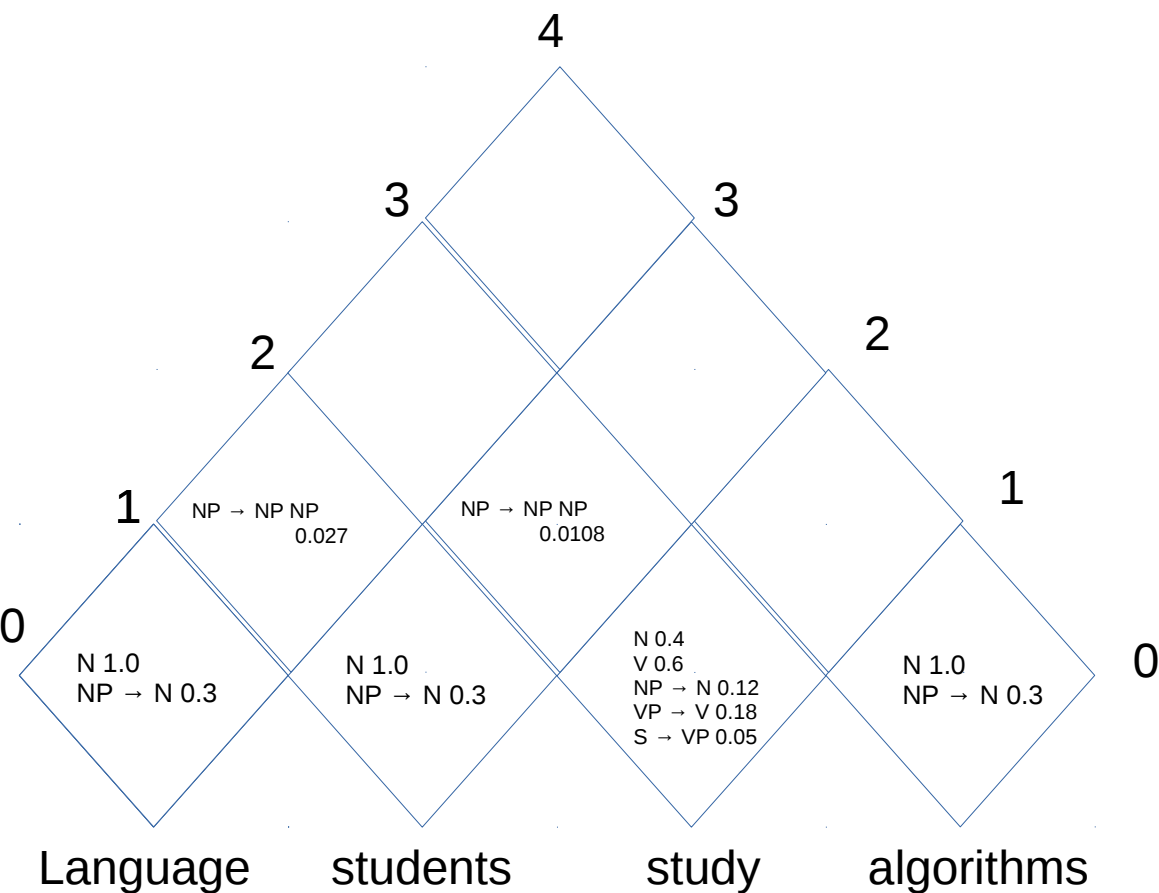
- Now, binary rules



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

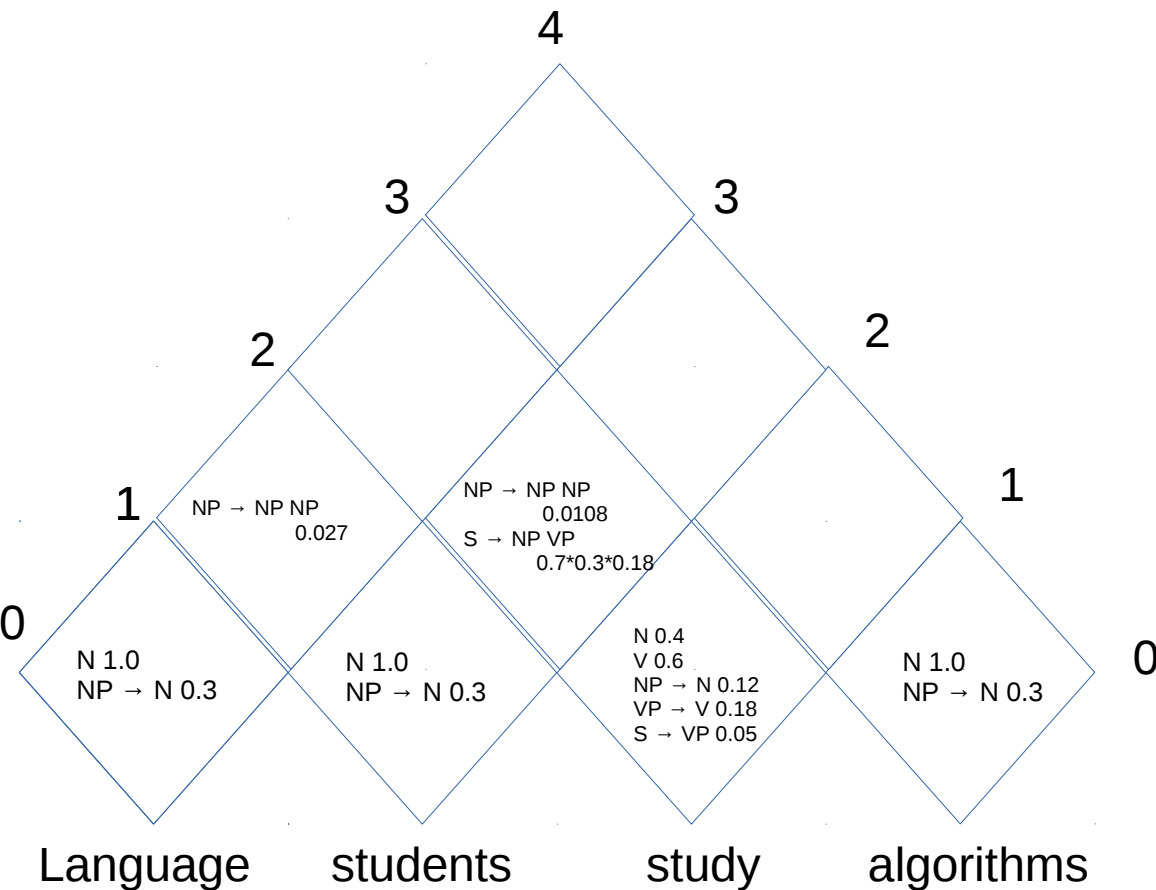
- Now, binary rules



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

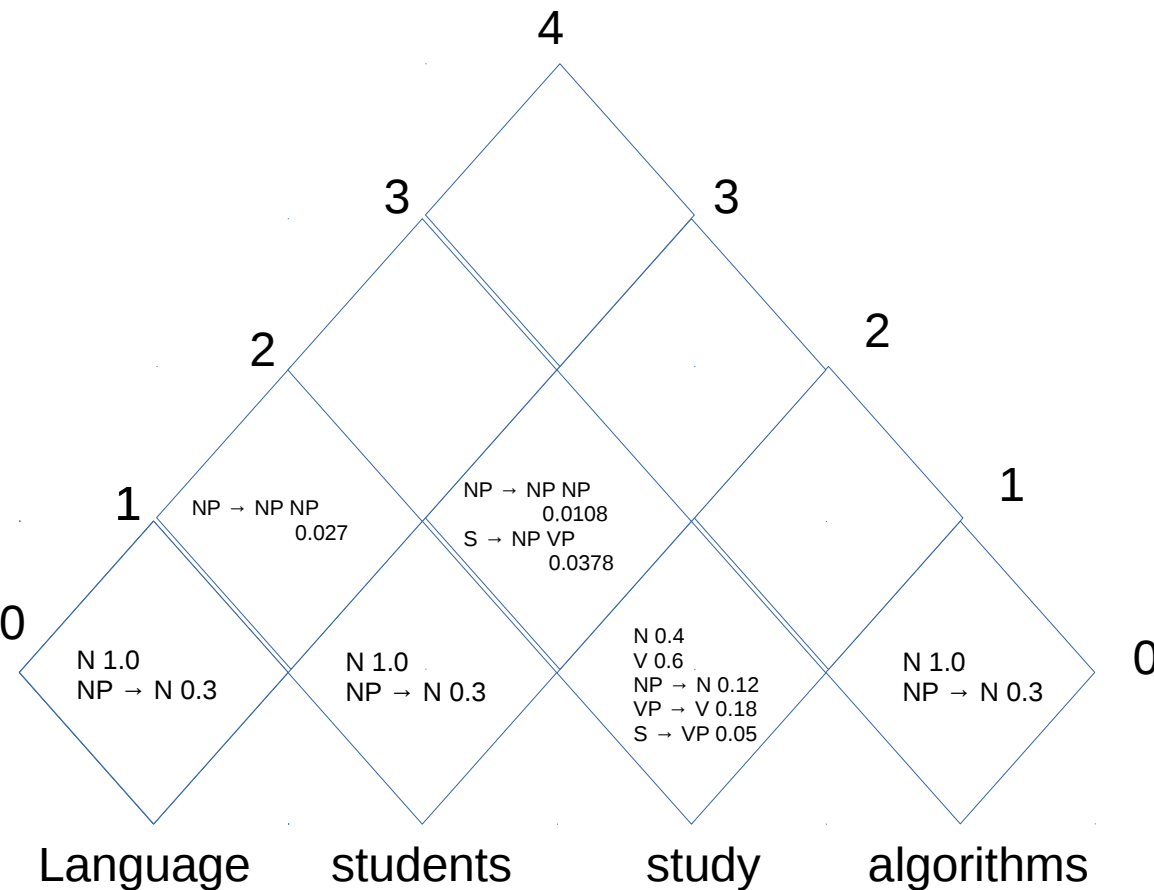
- Now, binary rules



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow \text{students}$ 1.0
- $N \rightarrow \text{study}$ 0.4
- $V \rightarrow \text{study}$ 0.6
- $N \rightarrow \text{Algorithms}$ 1.0
- $N \rightarrow \text{Language}$ 1.0

Phrase structure parsing CKY algorithm

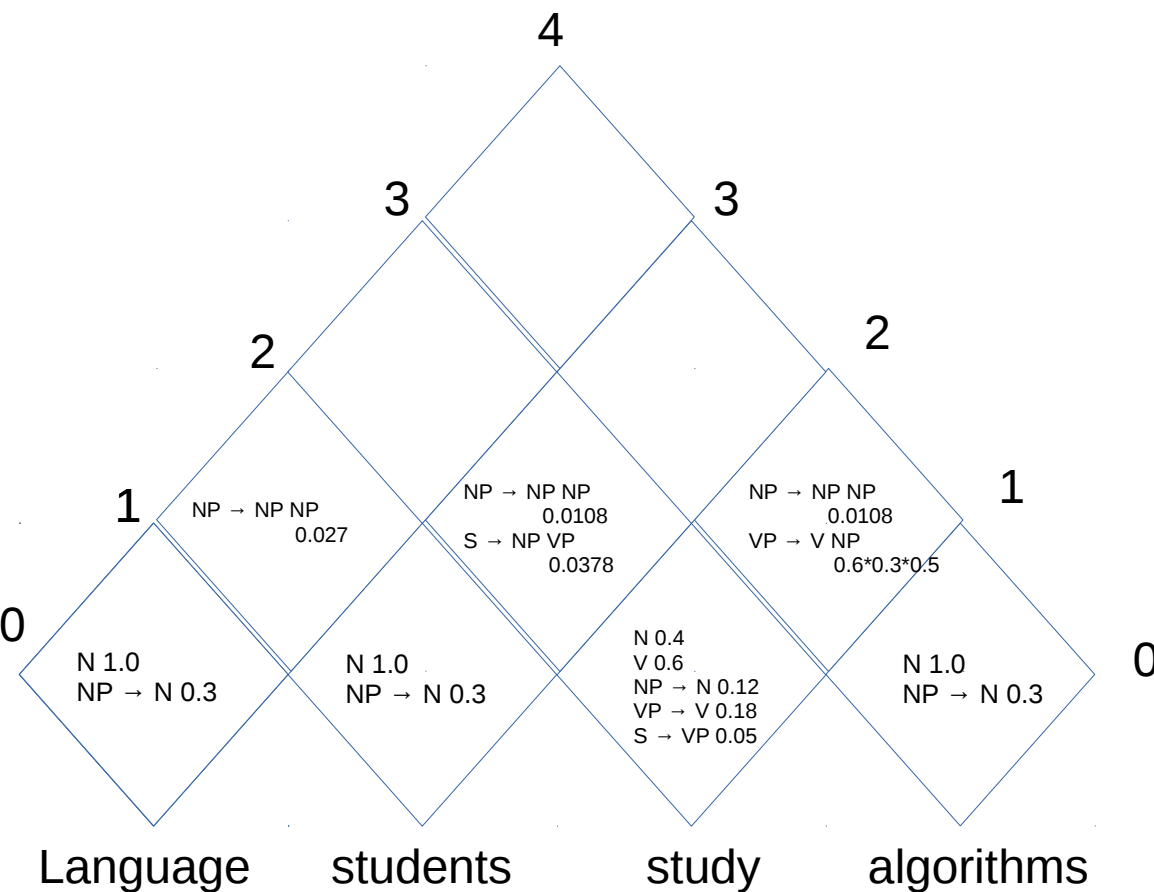
- Now, binary rules



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

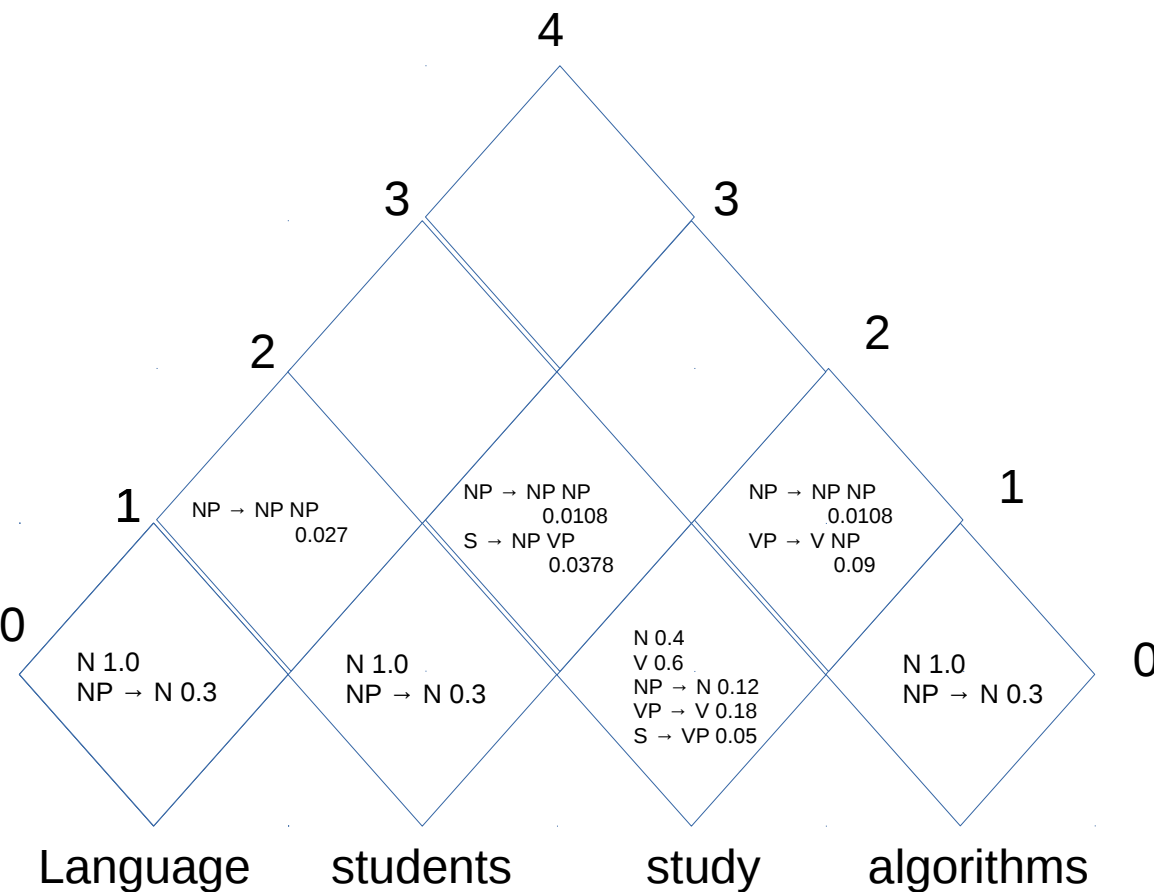
- Now, binary rules



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

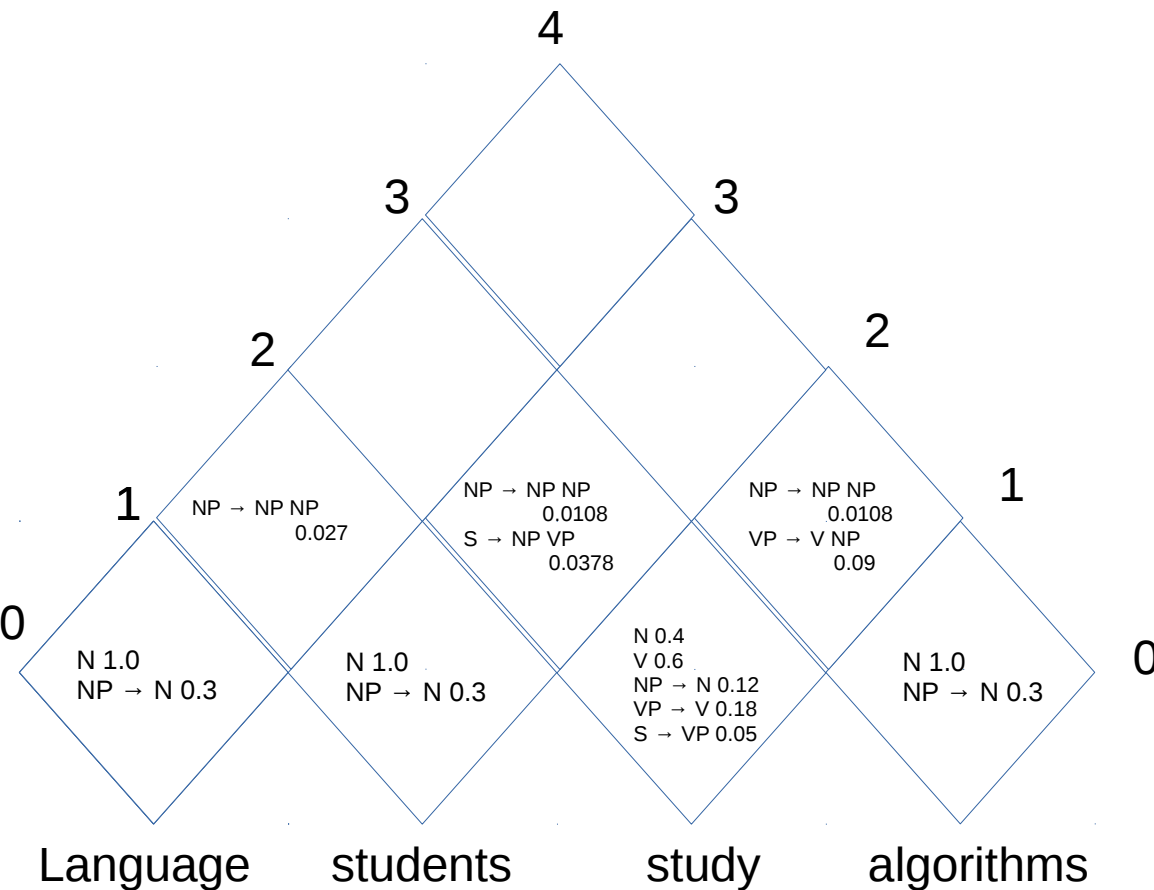
- Now, binary rules



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

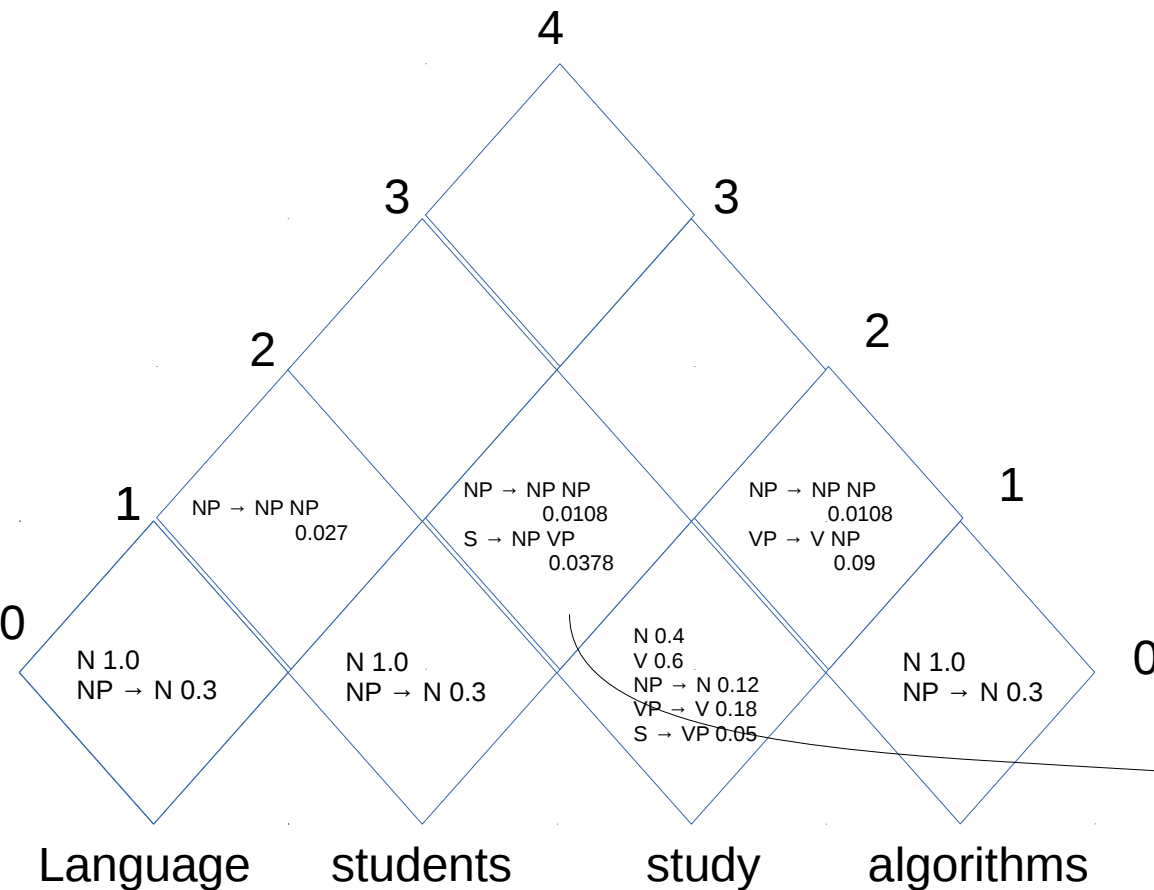
- Are there unary rules that are better?



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

- Are there unary rules that are better?

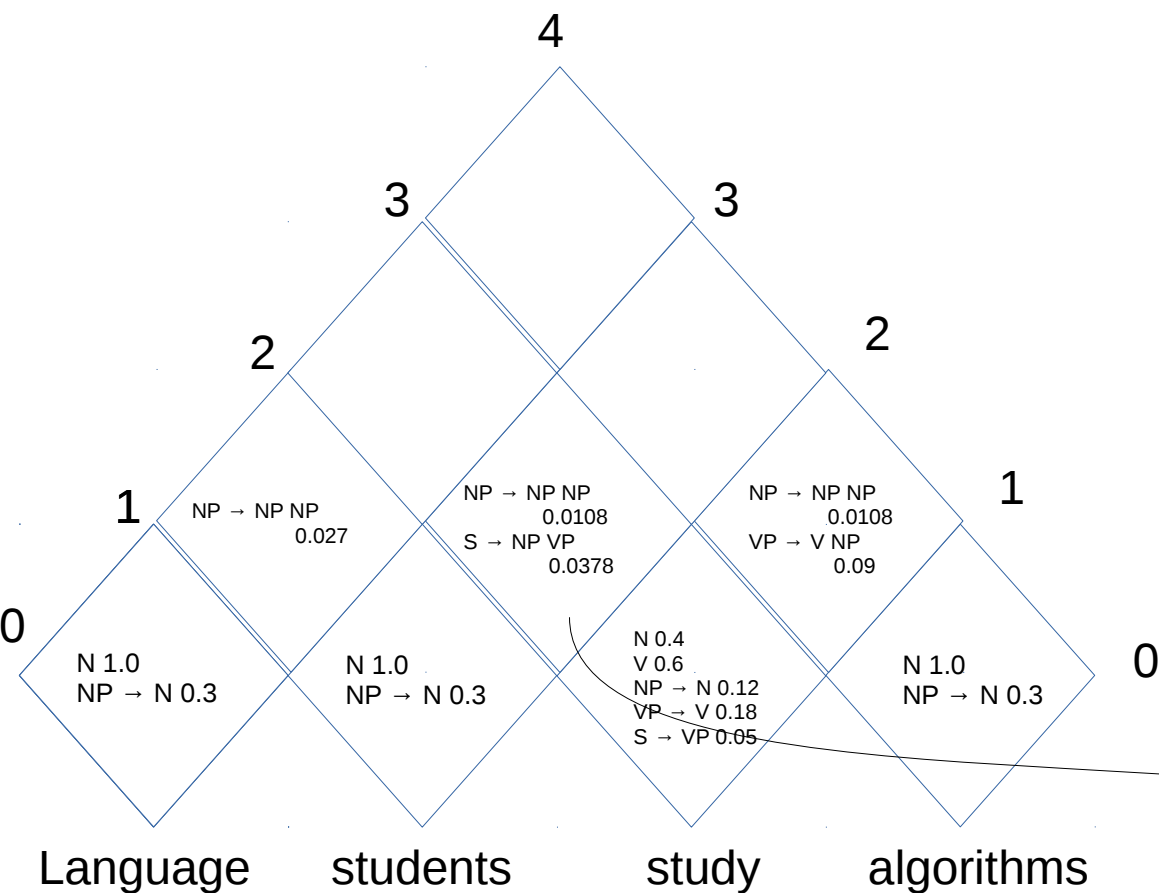


- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

S → VP 0.3*0.18

Phrase structure parsing CKY algorithm

- Are there unary rules that are better?

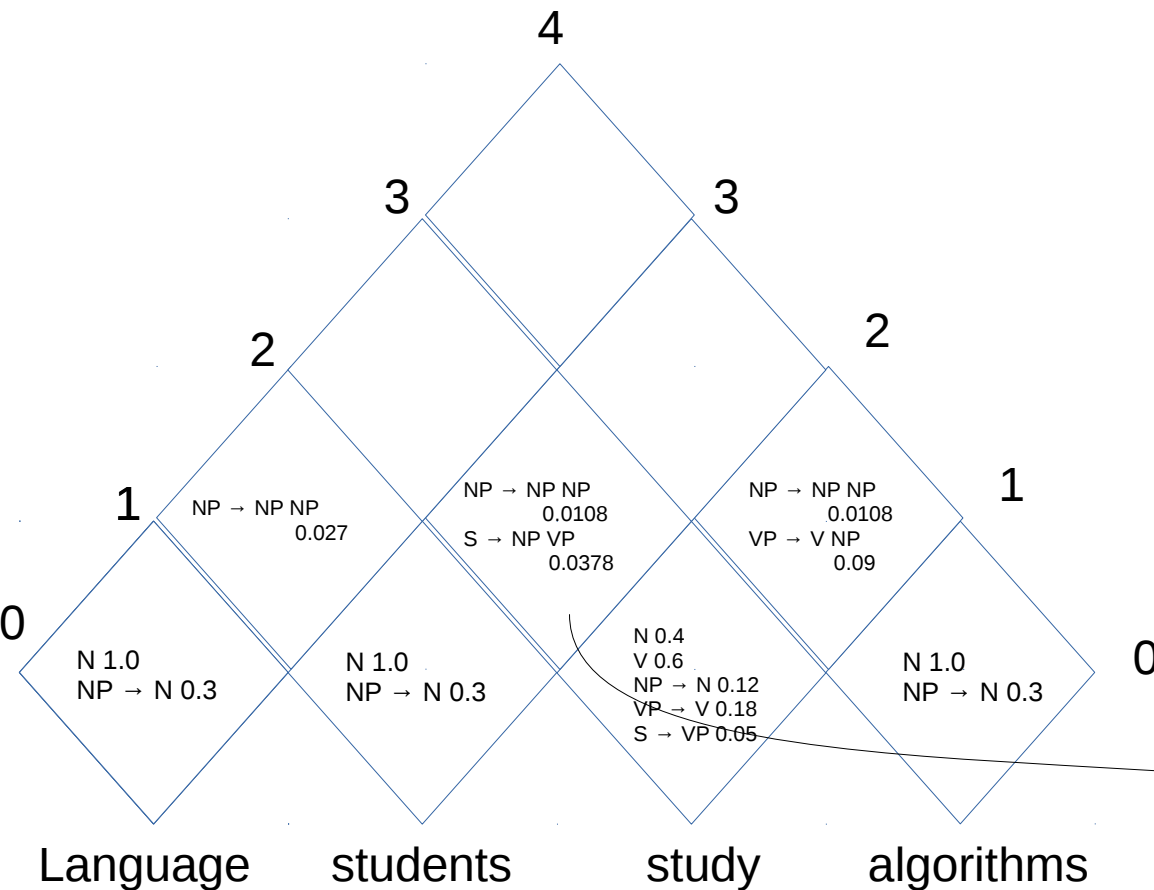


- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

$S \rightarrow VP \quad 0.3 \cdot 0.18 = 0.054$

Phrase structure parsing CKY algorithm

- Are there unary rules that are better?



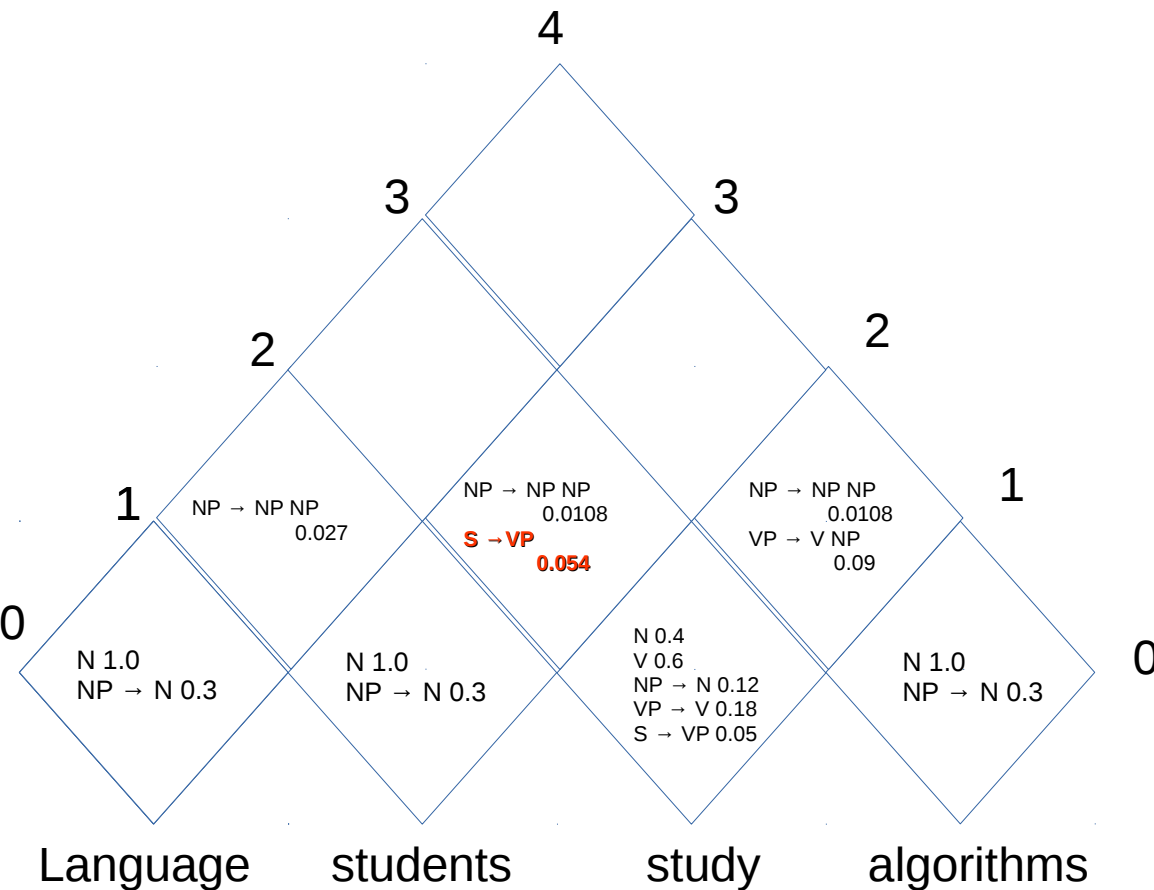
- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

$$S \rightarrow VP \quad 0.3 * 0.18 = 0.054$$

Is $0.054 > 0.0378$! yes, we replace the rule

Phrase structure parsing CKY algorithm

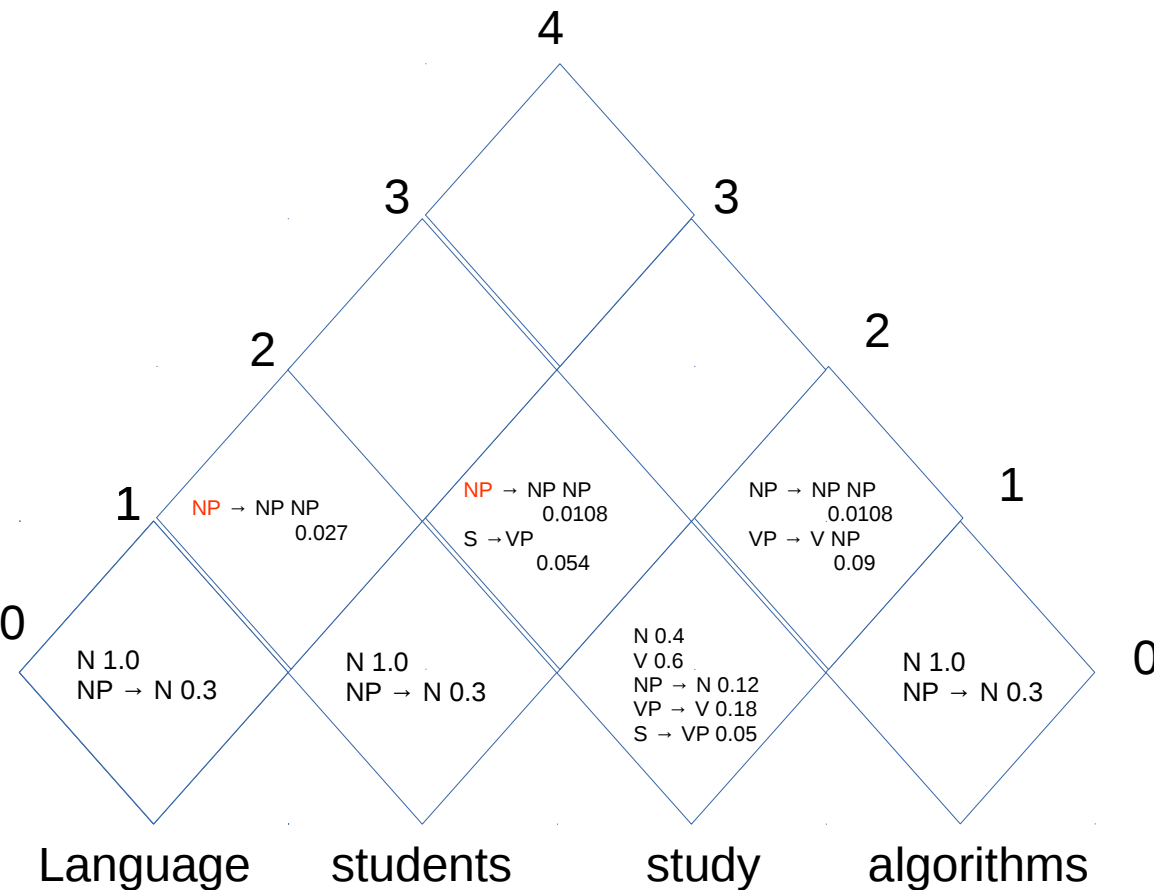
- Are there unary rules that are better?



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

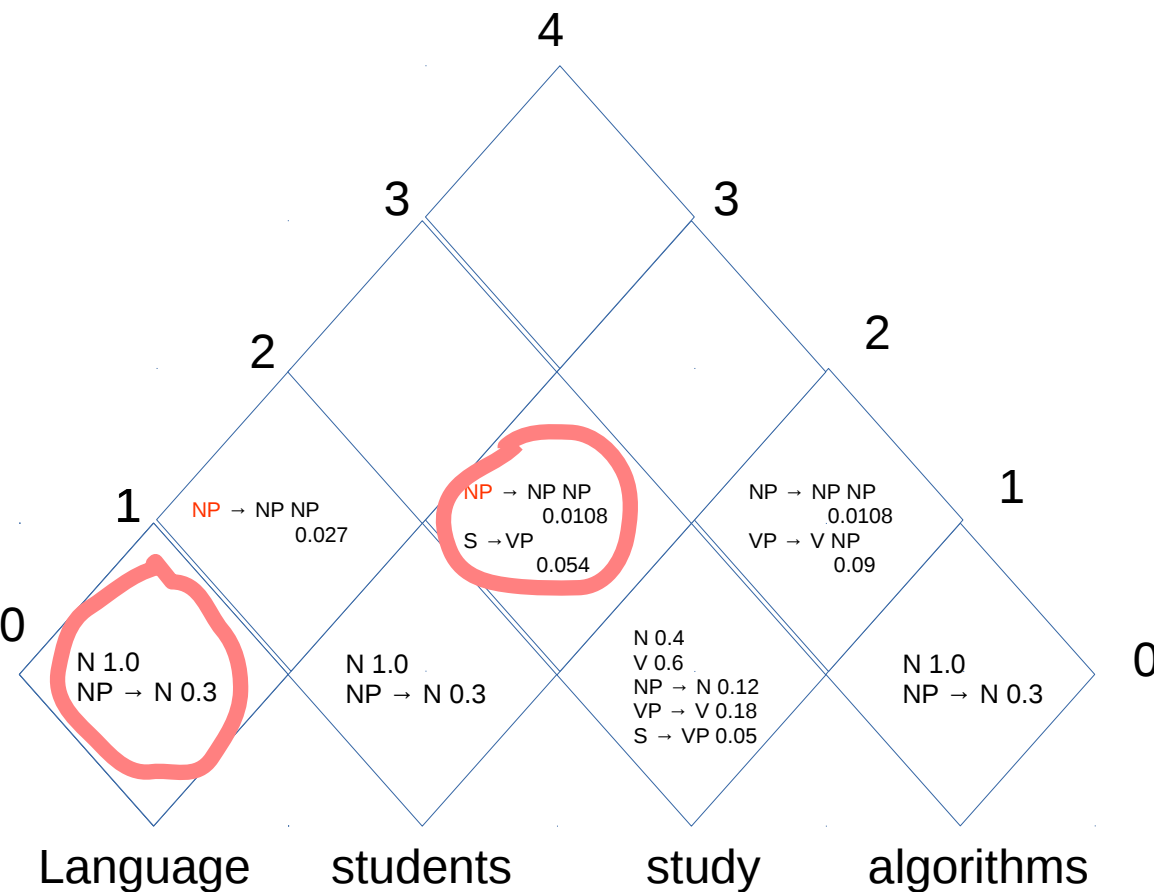
- Now, binary rules again



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

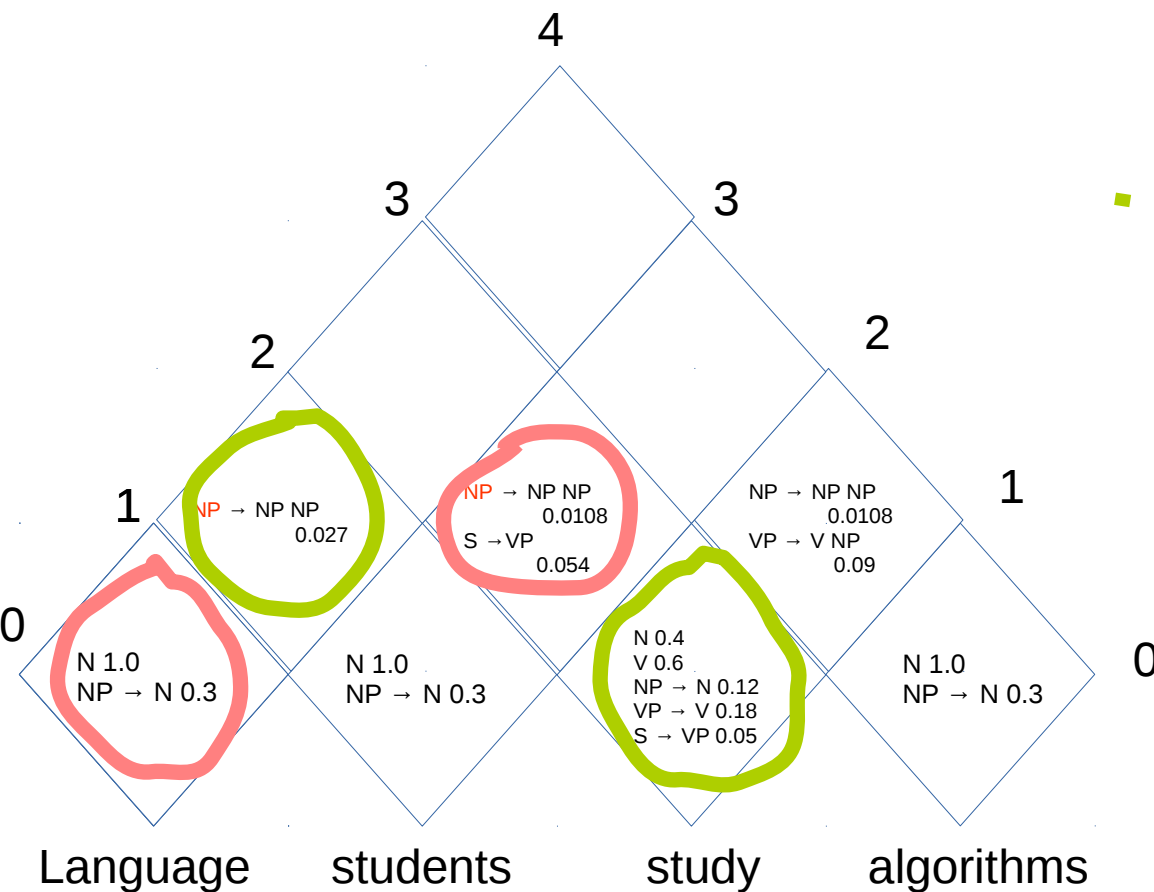
- Now, binary rules again



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow \text{students}$ 1.0
- $N \rightarrow \text{study}$ 0.4
- $V \rightarrow \text{study}$ 0.6
- $N \rightarrow \text{Algorithms}$ 1.0
- $N \rightarrow \text{Language}$ 1.0

Phrase structure parsing CKY algorithm

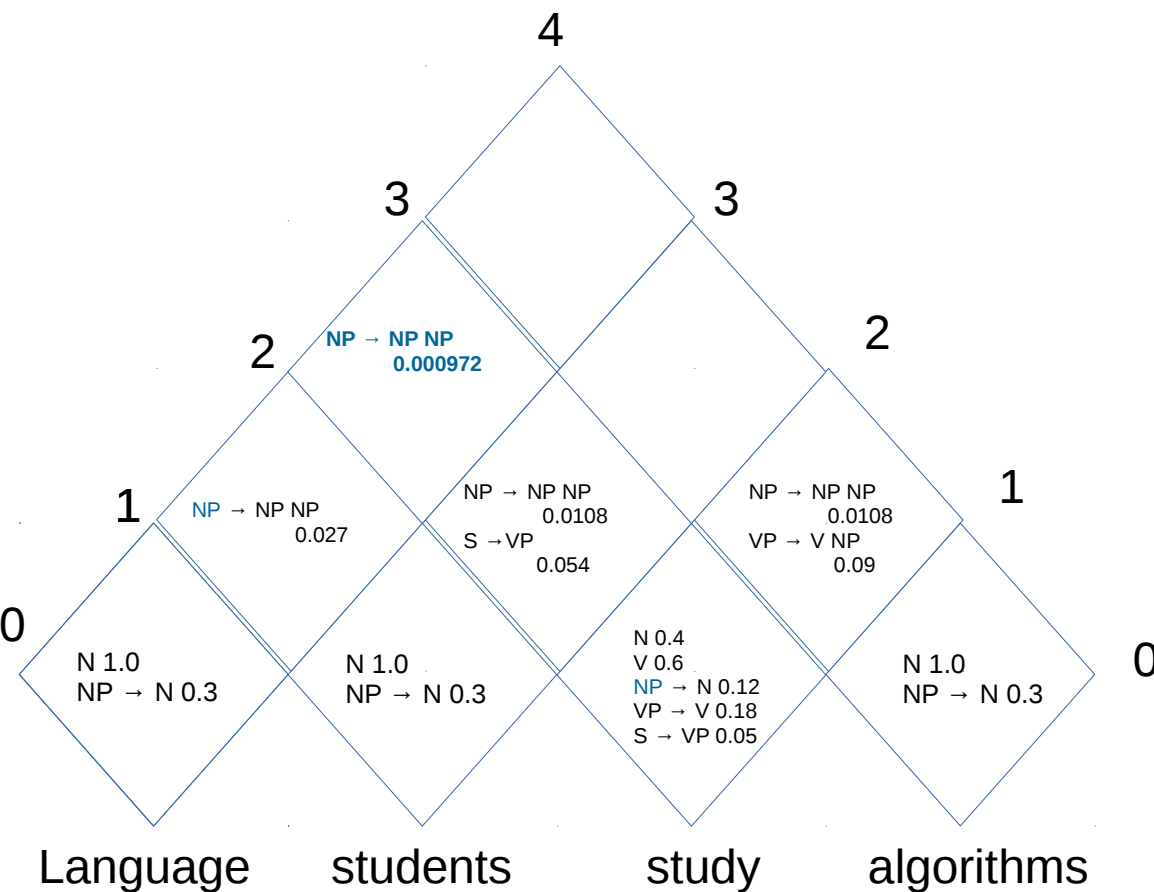
- Now, binary rules again



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

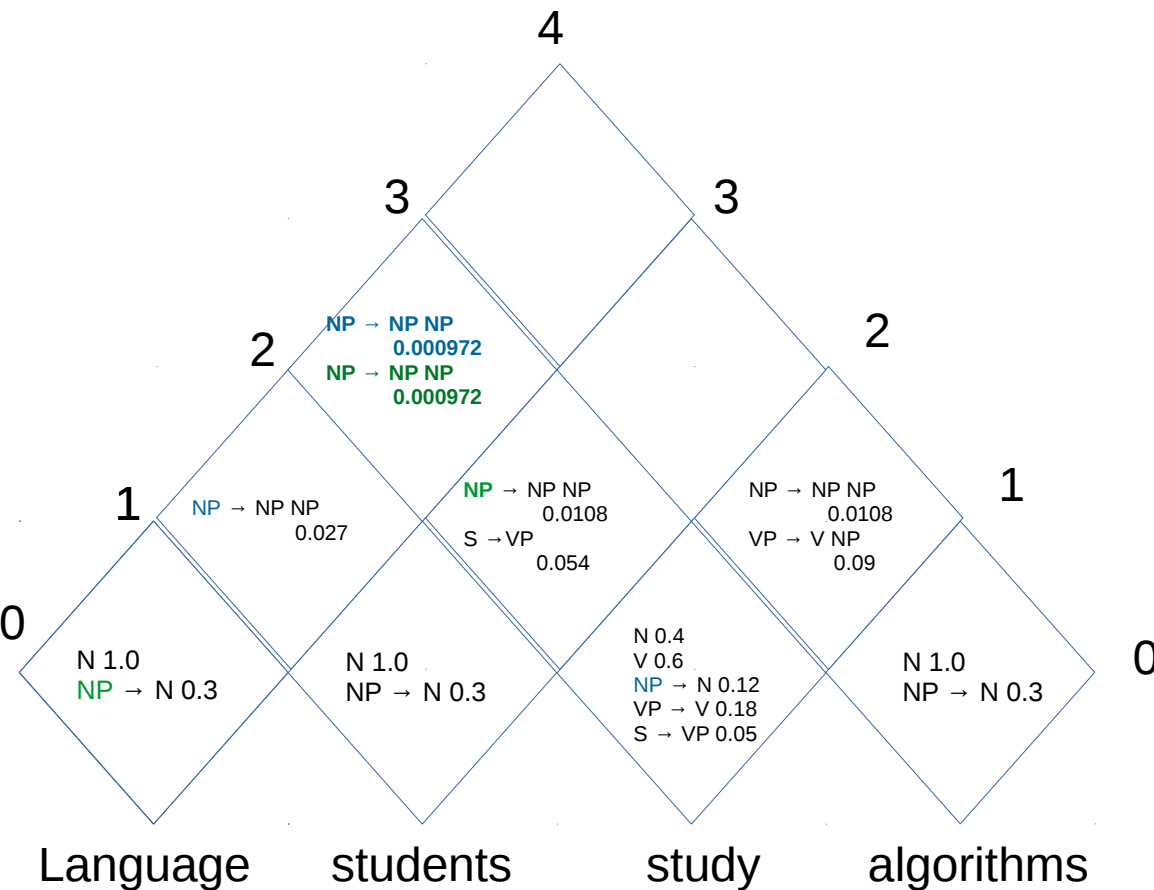
- Now, binary rules again



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

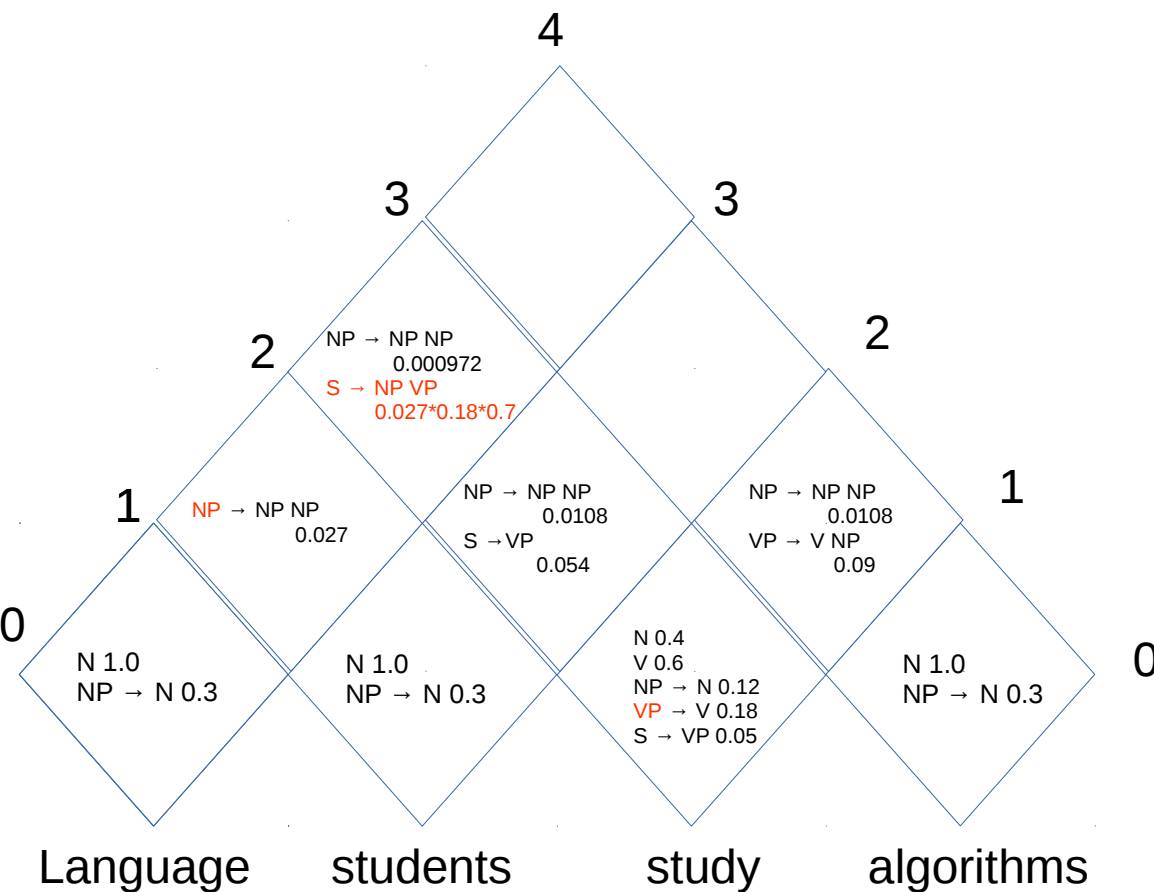
- Now, binary rules again



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

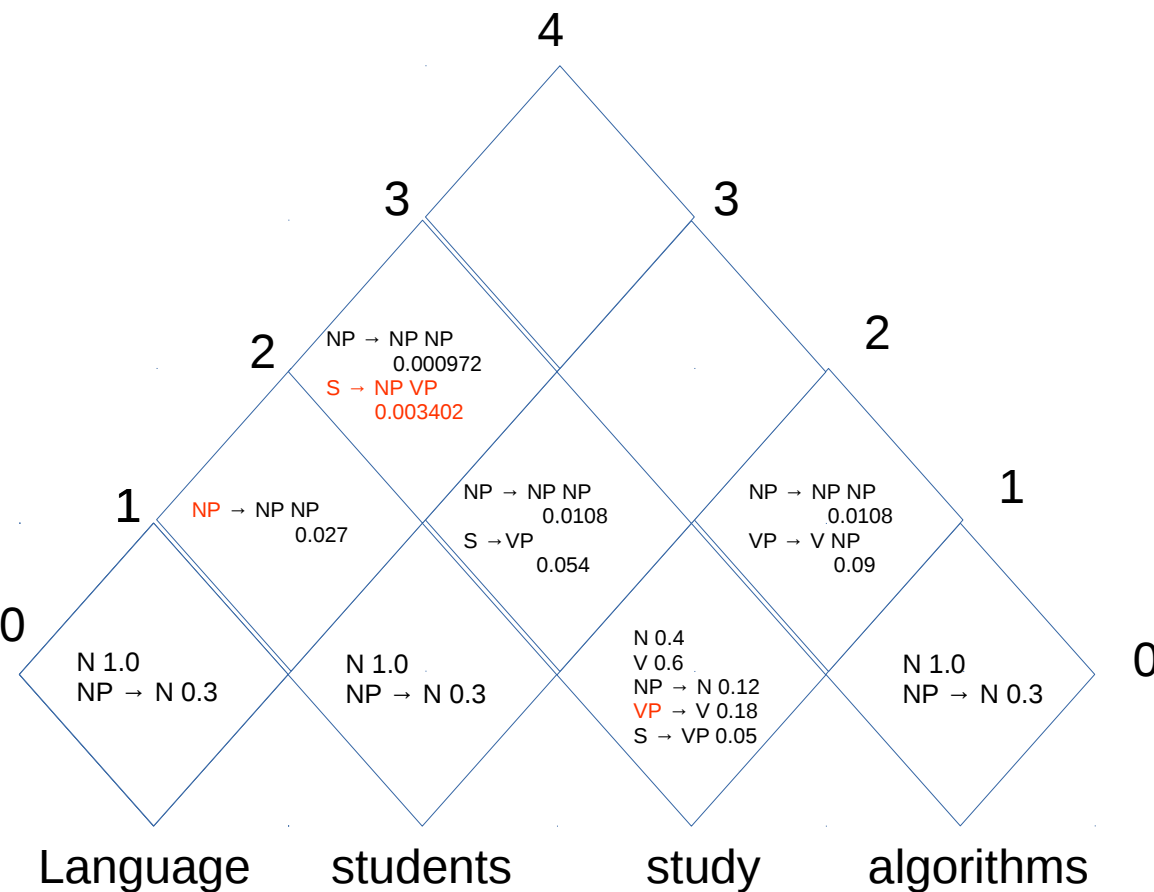
- Now, binary rules again



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

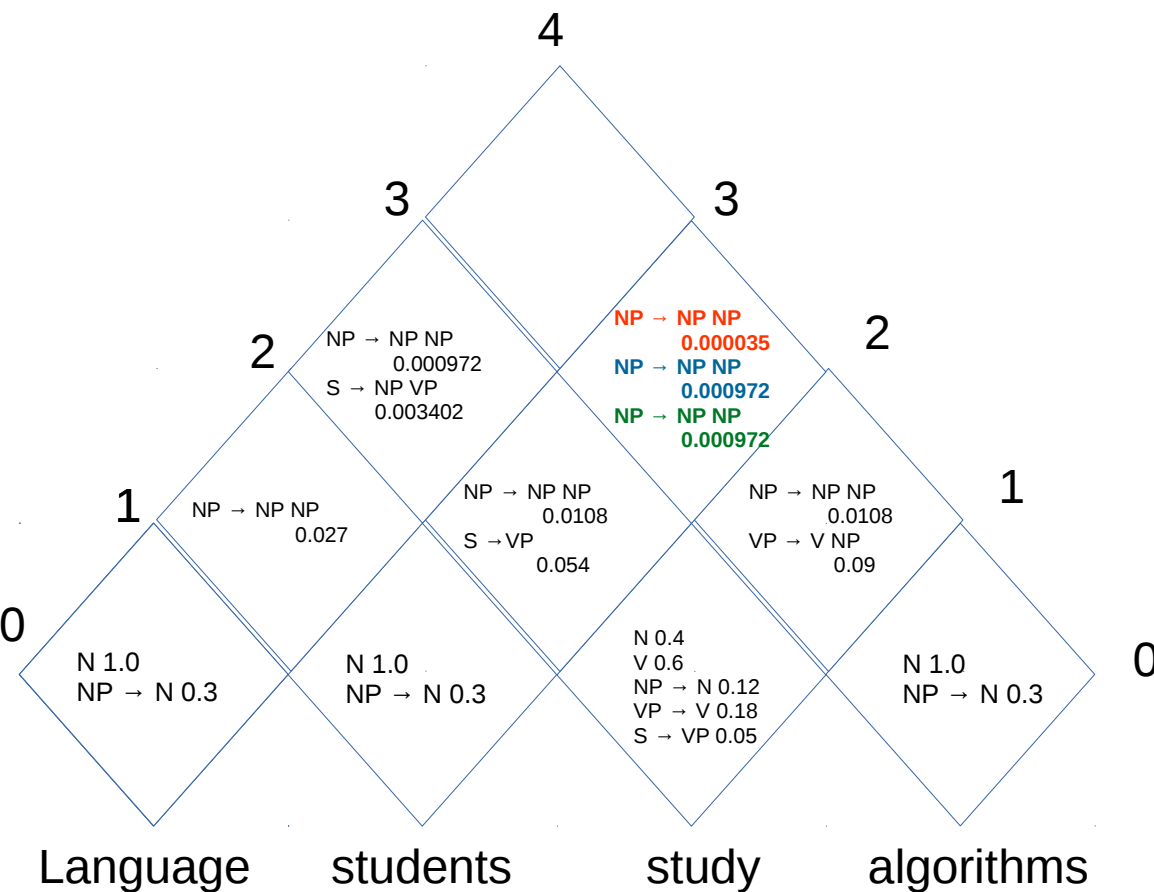
- Now, binary rules again



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

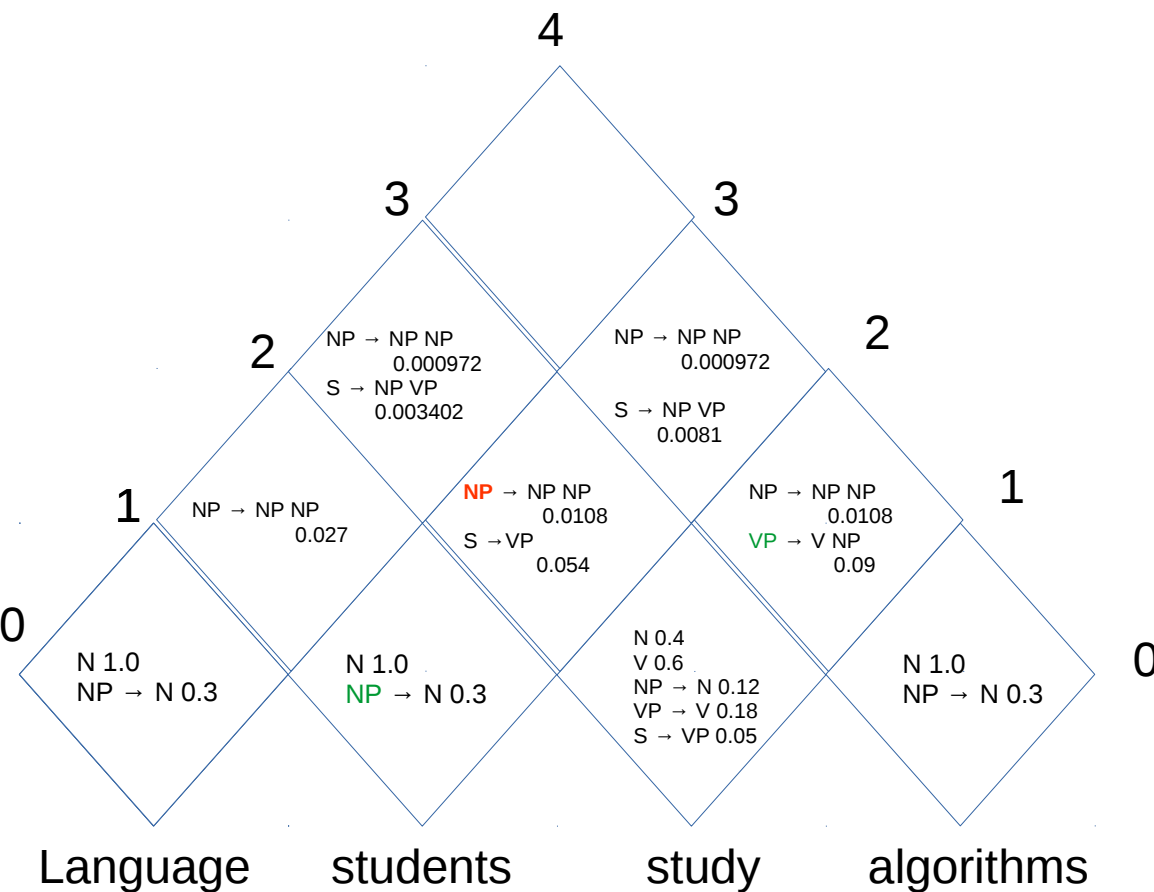
- Now, binary rules again



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

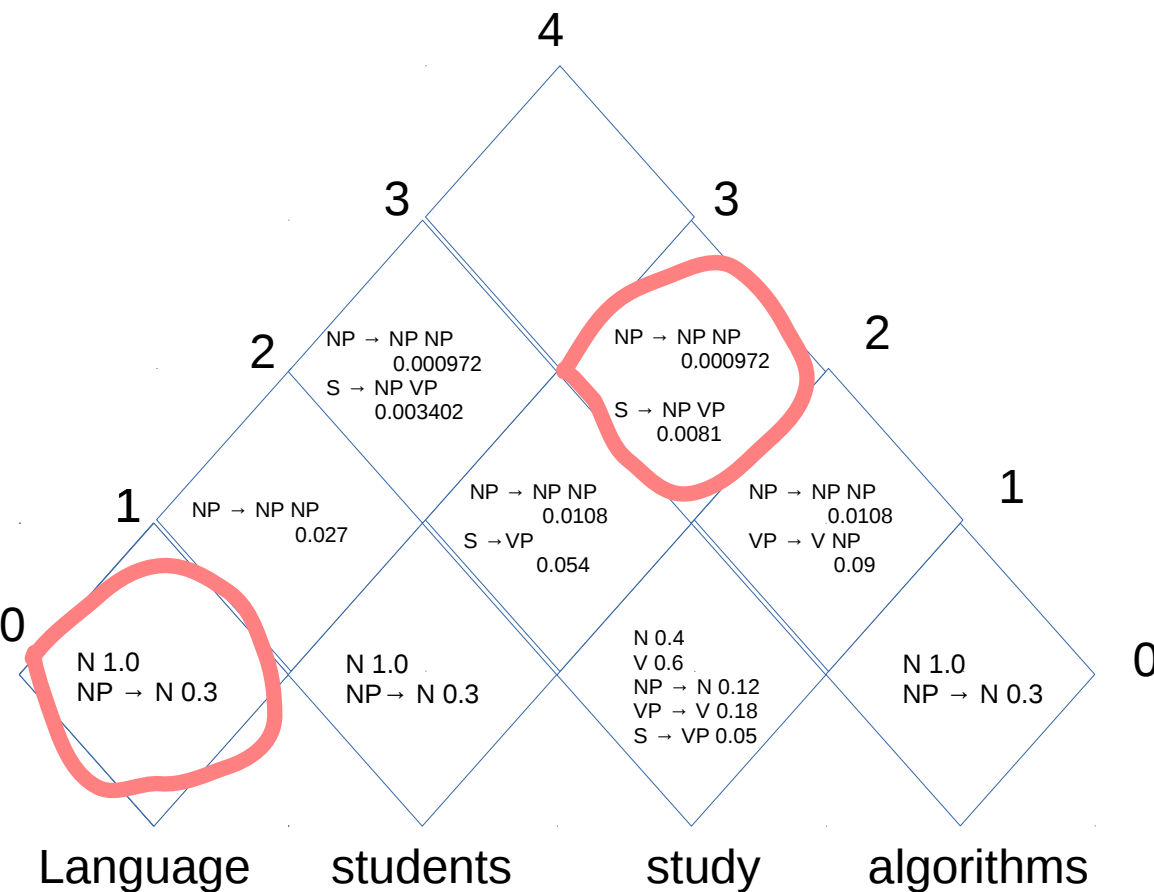
- Now, binary rules again



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

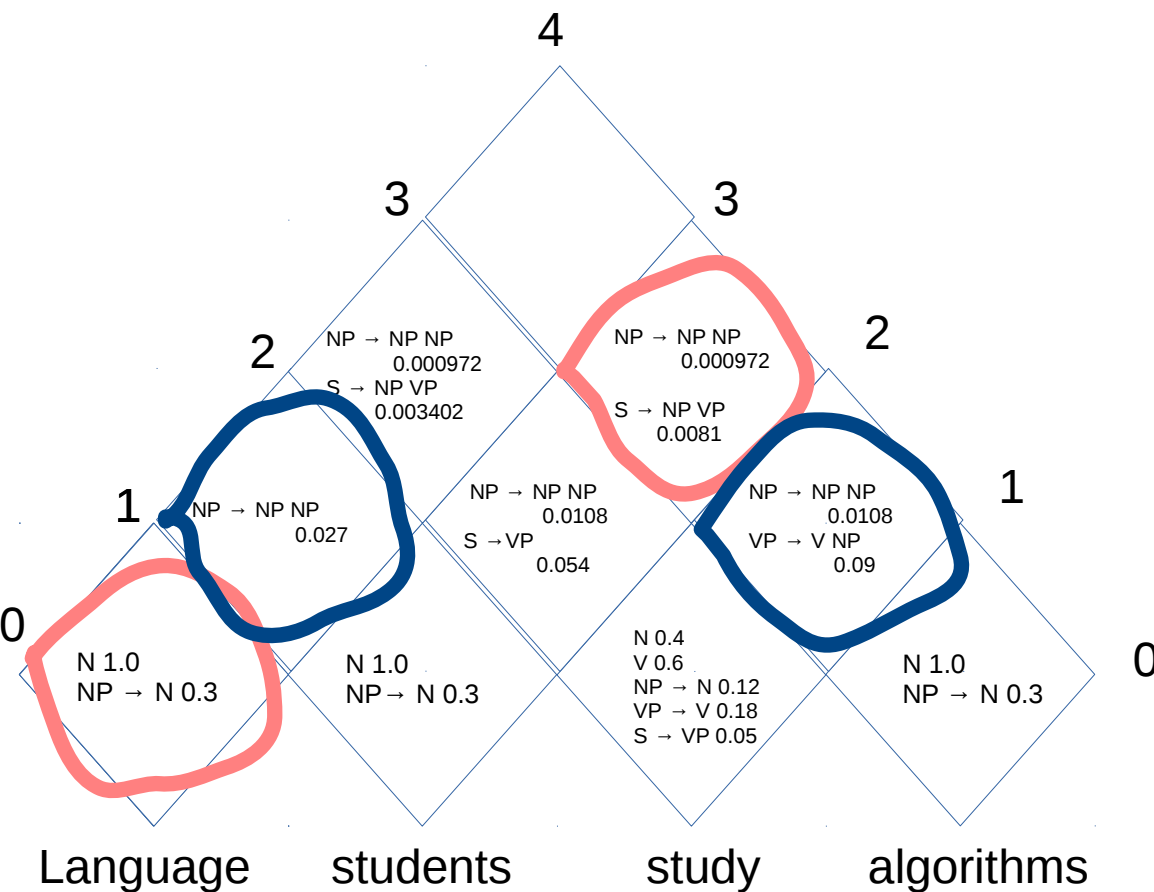
- Now, binary rules again.



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

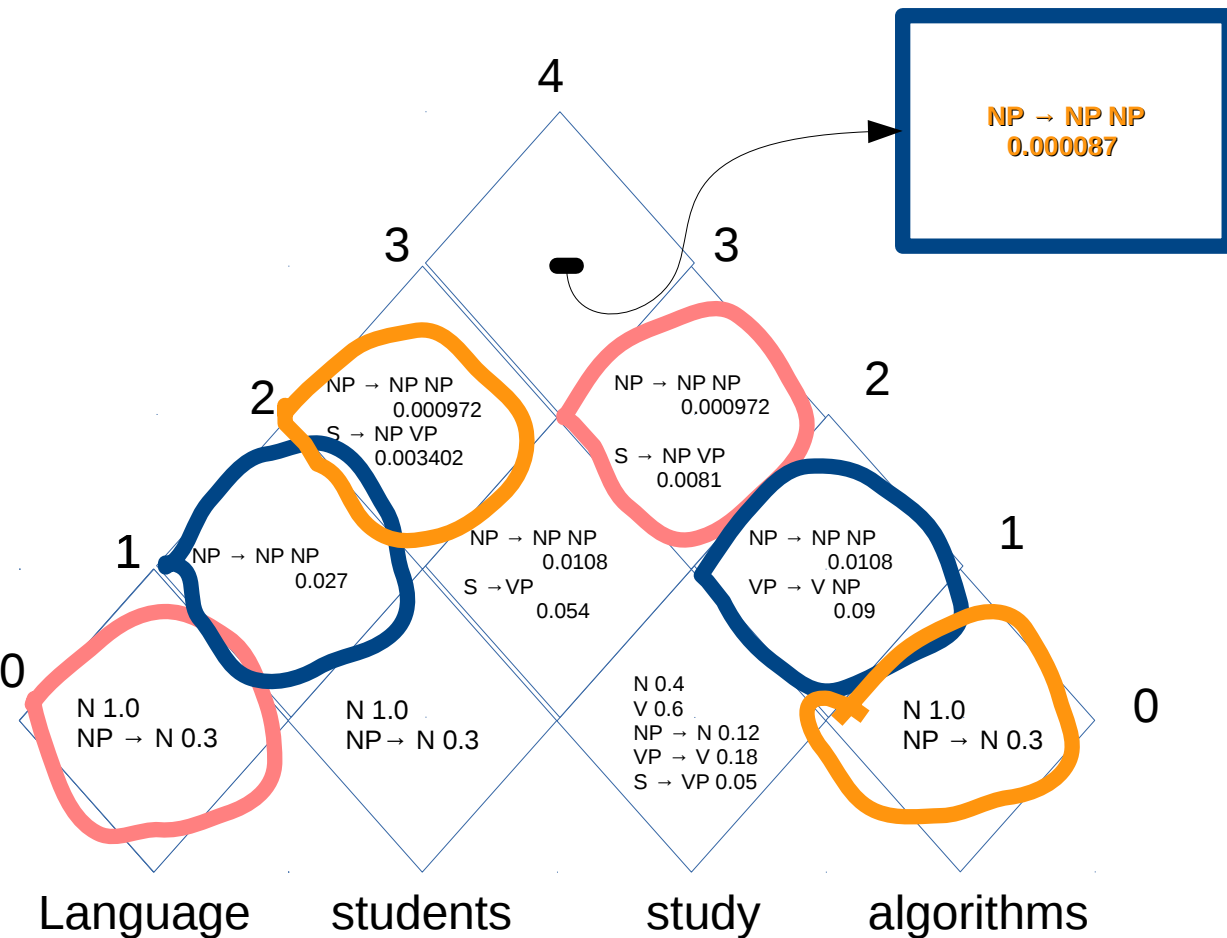
- Now, binary rules again.



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

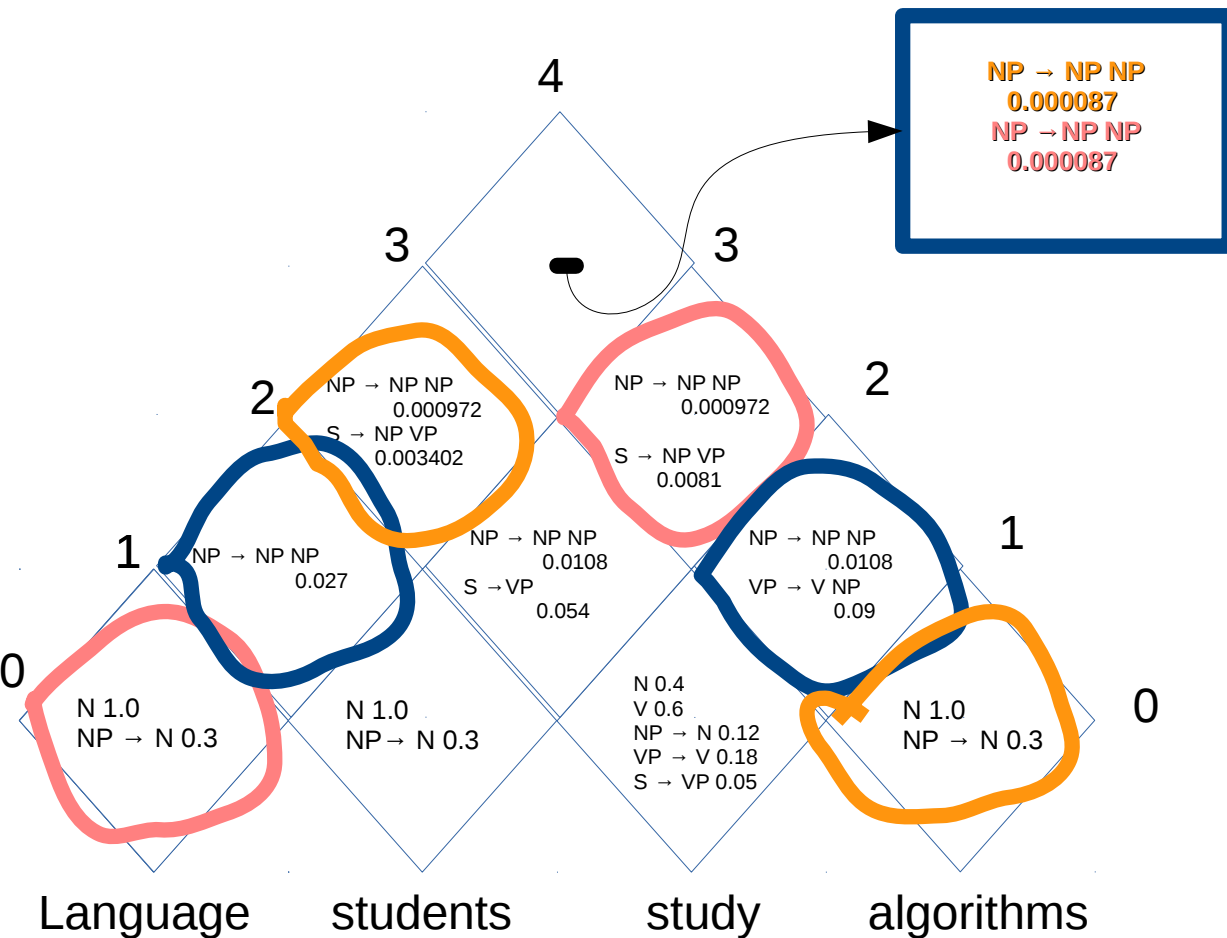
- Now, binary rules again.



- $S \rightarrow NP VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP NP$ 0.3
- $NP \rightarrow NP PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P NP$ 1.0
- $N \rightarrow students$ 1.0
- $N \rightarrow study$ 0.4
- $V \rightarrow study$ 0.6
- $N \rightarrow Algorithms$ 1.0
- $N \rightarrow Language$ 1.0

Phrase structure parsing CKY algorithm

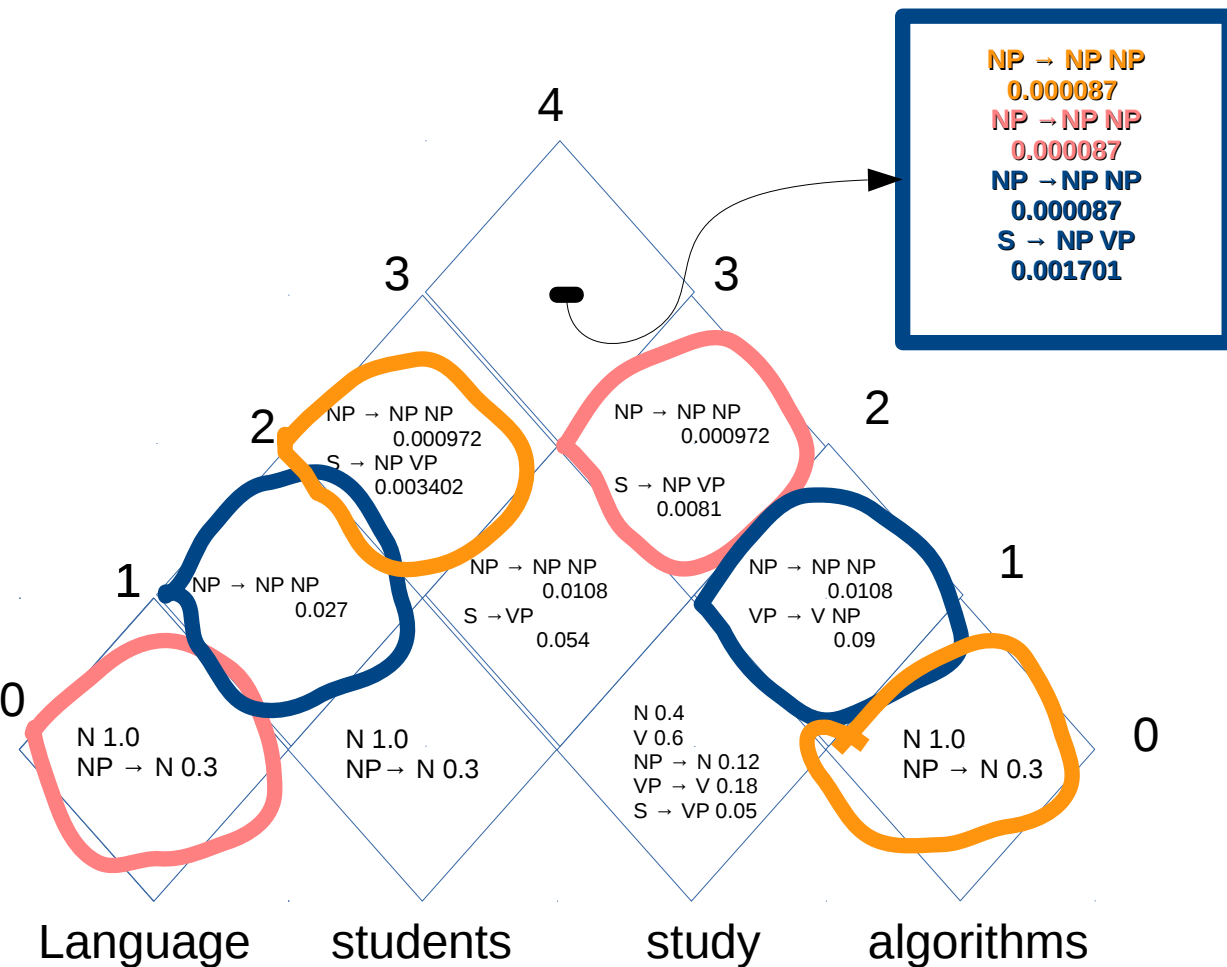
- Now, binary rules again.



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

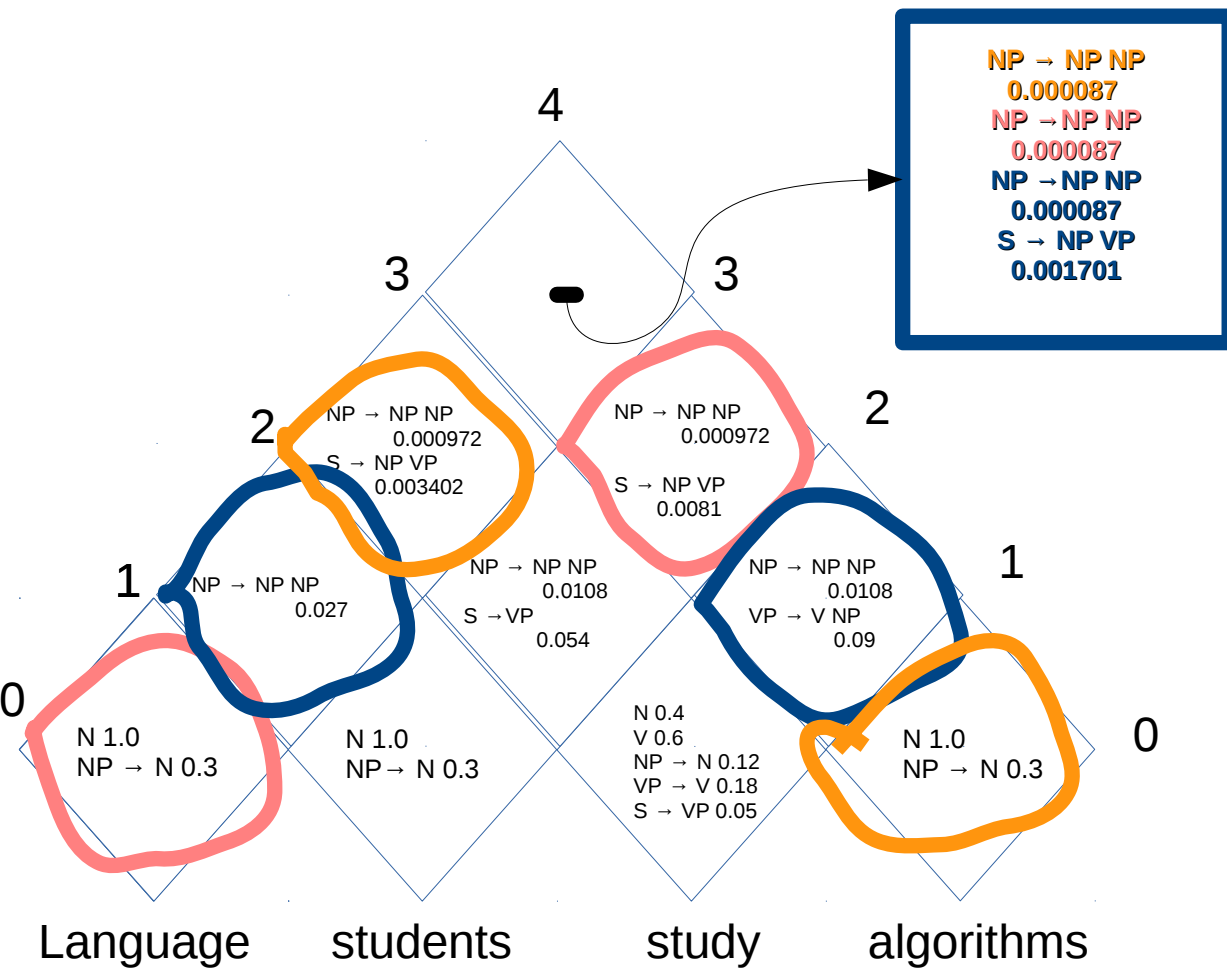
- Now, binary rules again.



- S → NP VP 0.7
- S → VP 0.3
- VP → V NP 0.5
- VP → V 0.3
- NP → NP NP 0.3
- NP → NP PP 0.4
- NP → N 0.3
- PP → P NP 1.0
- N → students 1.0
- N → study 0.4
- V → study 0.6
- N → Algorithms 1.0
- N → Language 1.0

Phrase structure parsing CKY algorithm

- Now, binary rules again.



- $S \rightarrow NP \ VP$ 0.7
- $S \rightarrow VP$ 0.3
- $VP \rightarrow V \ NP$ 0.5
- $VP \rightarrow V$ 0.3
- $NP \rightarrow NP \ NP$ 0.3
- $NP \rightarrow NP \ PP$ 0.4
- $NP \rightarrow N$ 0.3
- $PP \rightarrow P \ NP$ 1.0
- $N \rightarrow \text{students}$ 1.0
- $N \rightarrow \text{study}$ 0.4
- $V \rightarrow \text{study}$ 0.6
- $N \rightarrow \text{Algorithms}$ 1.0
- $N \rightarrow \text{Language}$ 1.0

