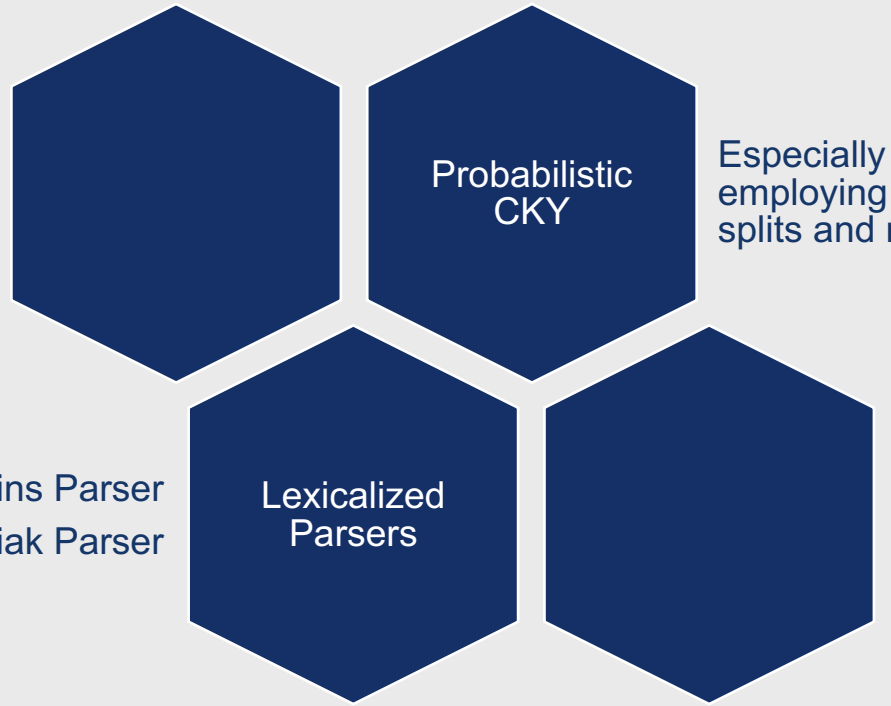
Probabilistic Lexicalized CFGs

Natalie Parde
UIC CS 421

Parsing Methods

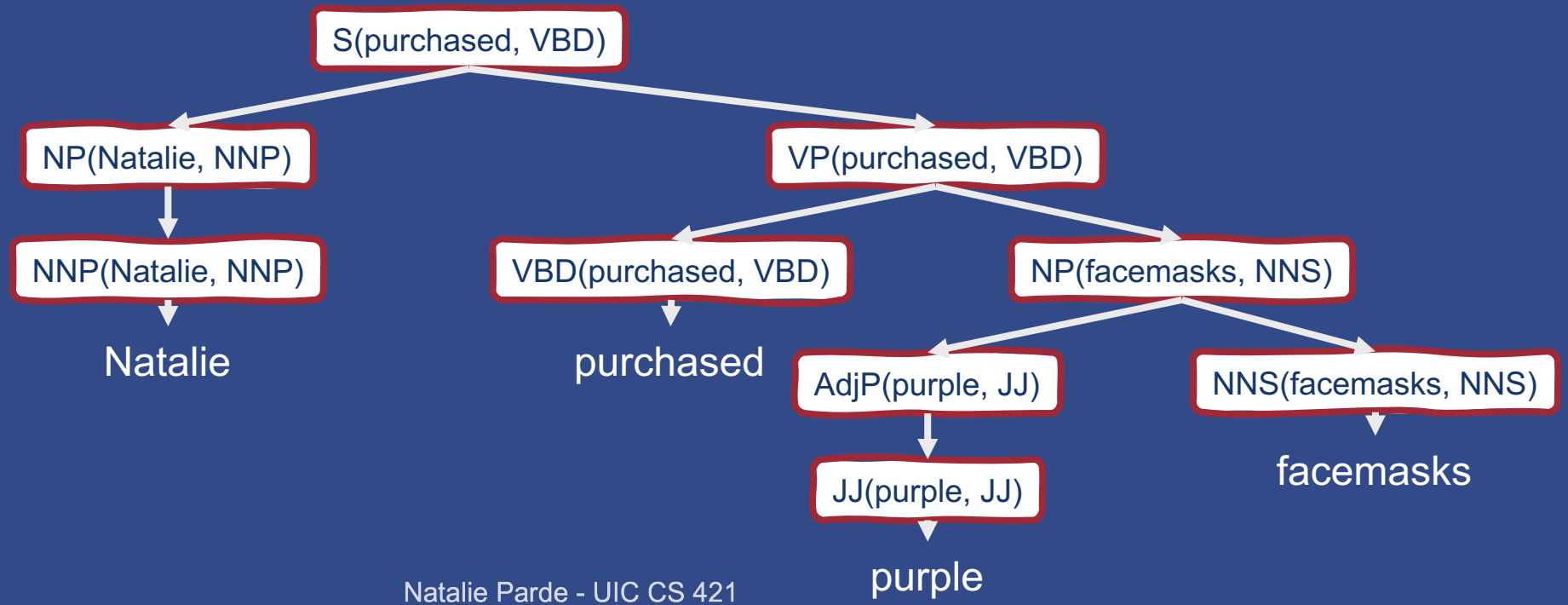


Especially when employing automated splits and merges!

Collins Parser
Charniak Parser

Lexicalized Parsers

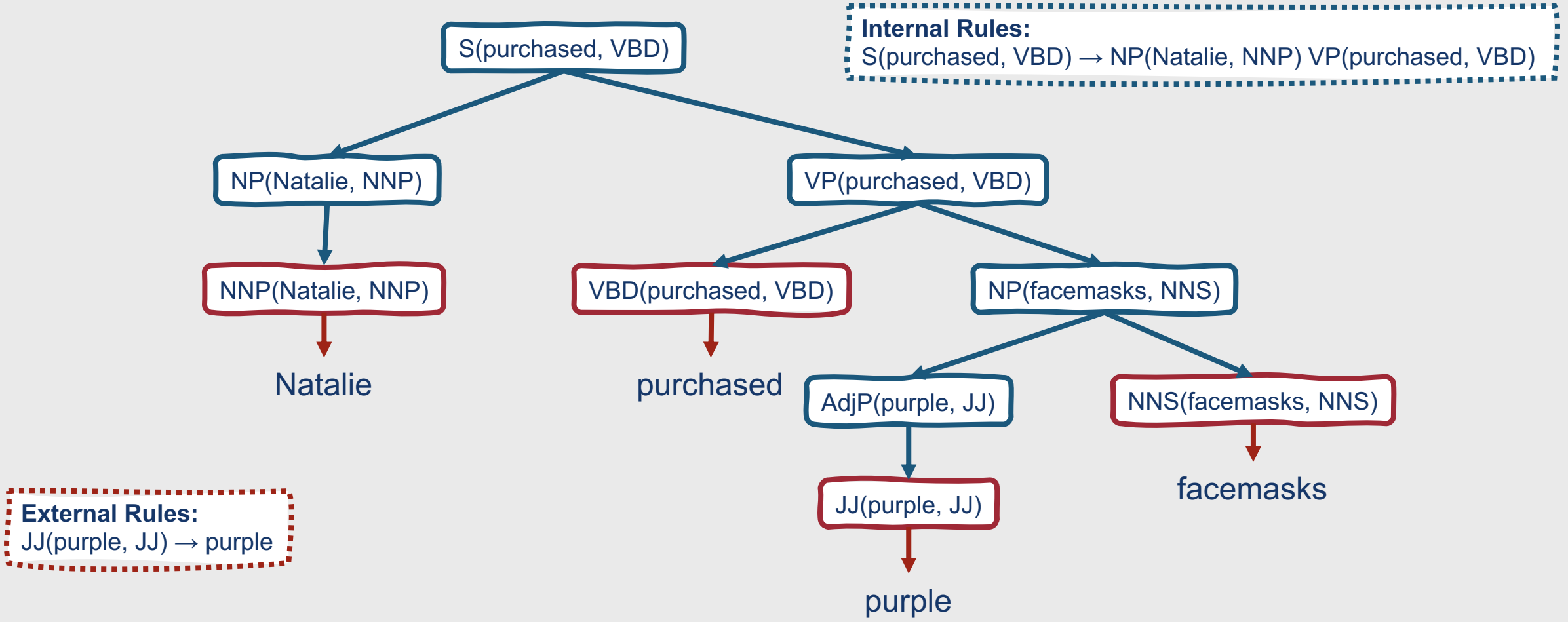
- Allow lexicalized rules
  - Non-terminals specify lexical heads and associated POS tags
  - NP(facemasks, NNS) → AdjP(purple, JJ) NNS(facemasks, NNS)



Lexicalized Grammars

- Intuitively, much like having many copies of the same production rule
  - NP(facemasks, NNS) → AdjP(purple, JJ) NNS(facemasks, NNS)
  - NP(facemasks, NNS) → AdjP(green, JJ) NNS(facemasks, NNS)
  - NP(sanitizers, NNS) → AdjP(purple, JJ) NNS(sanitizers, NNS)
- Two types of rules:
  - **Lexical Rules:** Generate a terminal word
  - **Internal Rules:** Generate a non-terminal constituent

Lexical vs. Internal Rules



Lexical vs. Internal Rules

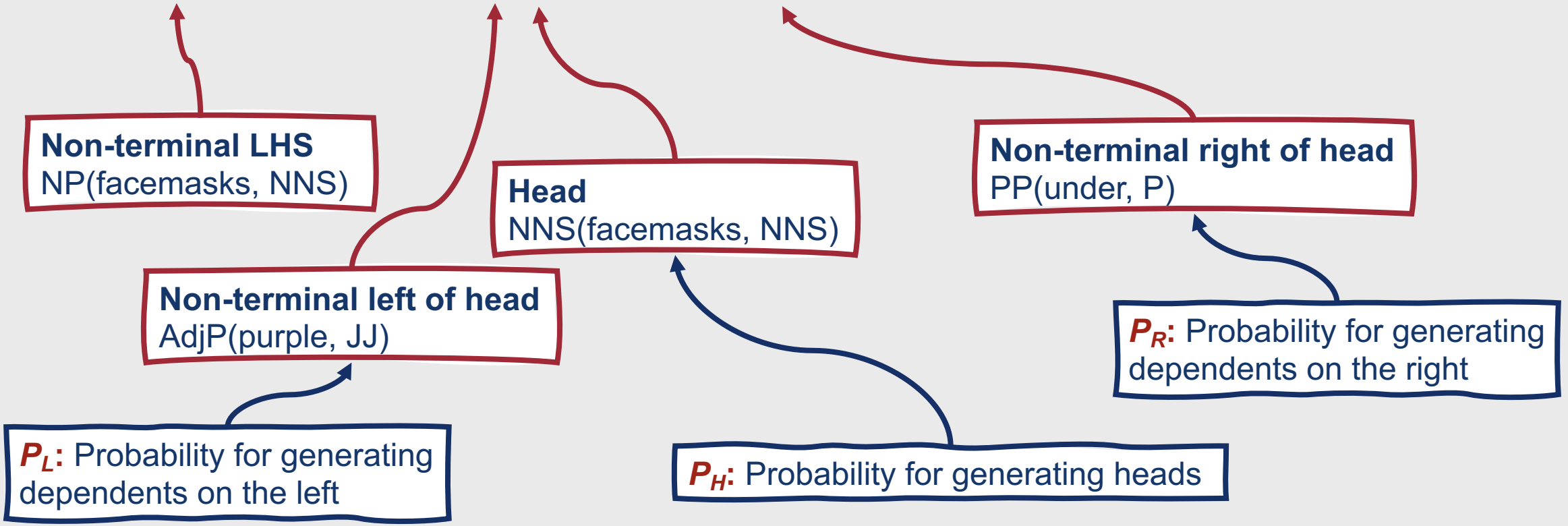
- **Lexical Rules**
  - Deterministic
    - JJ(purple, JJ) → purple
- **Internal Rules**
  - Require estimated probabilities
    - Normal maximum likelihood estimation won't work well because the counts will be too sparse
    - Instead, estimate the probability of an internal rule based on the product of the smaller, more reliable probability estimates comprising it

The Collins Parser

- Consider the following generic production rule:
  - $LHS \rightarrow L_n\ L_{n-1} \ldots L_1\ H\ R_1 \ldots R_{n-1}\ R_n$

The Collins Parser

- Consider the following generic production rule:
  - $LHS \rightarrow L_n\ L_{n-1} \ldots L_1\ H\ R_1 \ldots R_{n-1}\ R_n$



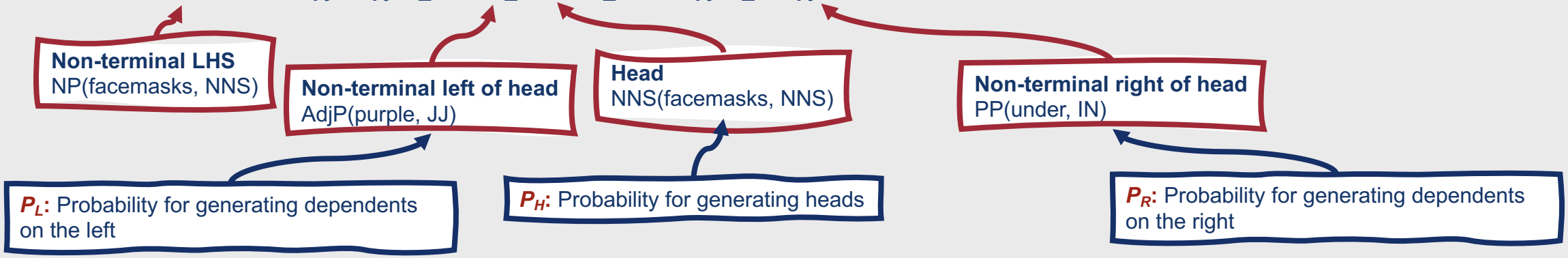
The Collins Parser

- Goal: Use P_H, P_L, and P_R to estimate the overall probability for the production rule
- Method:
  - Surround the righthand side of the rule with STOP non-terminals
    - NP(facemasks, NNS) → STOP AdjP(purple, JJ) NNS(facemasks, NNS) PP(under, IN) STOP
  - Compute the individual P_H, P_L, and P_R values for the head and the non-terminals to its left and right (including STOP non-terminals)
  - Multiply these together

Grab the **purple facemasks under the disinfectant**.

The Collins Parser

- Consider the following generic production rule:
  - $LHS \rightarrow L_n\ L_{n-1} \dots L_1\ H\ R_1 \dots R_{n-1}\ R_n$

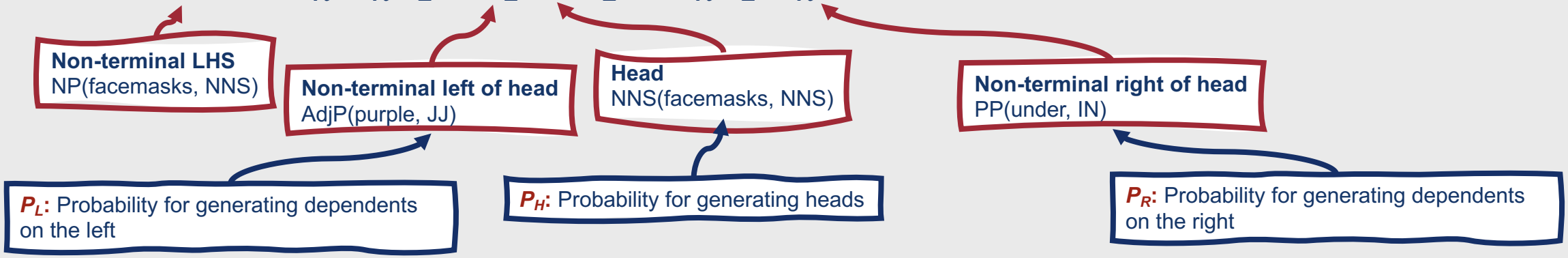


Grab the **purple facemasks under the disinfectant**.

NP(facemasks, NNS) → STOP AdjP(purple, JJ) NNS(facemasks, NNS) PP(under, IN) STOP

The Collins Parser

- Consider the following generic production rule:
  - $LHS \rightarrow L_n\ L_{n-1} \dots L_1\ H\ R_1 \dots R_{n-1}\ R_n$



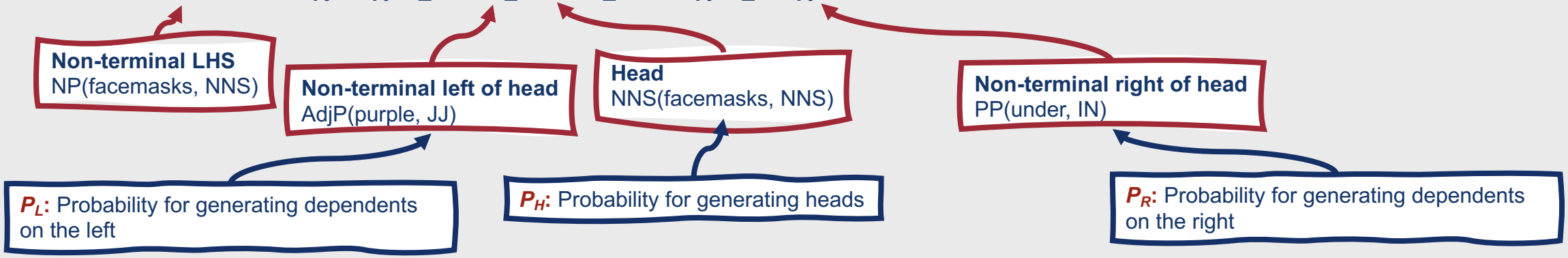
$P_H(H|LHS) = P(NNS(facemasks, NNS) \mid NP(facemasks, NNS))$

Grab the **purple facemasks under the disinfectant**.

NP(facemasks, NNS) → STOP AdjP(purple, JJ) **NNS(facemasks, NNS)** PP(under, IN) STOP

The Collins Parser

- Consider the following generic production rule:
  - $LHS \rightarrow L_n\ L_{n-1} \dots L_1\ H\ R_1 \dots R_{n-1}\ R_n$



Grab the **purple facemasks under the disinfectant**.

NP(facemasks, NNS) → **STOP AdjP(purple, JJ)** NNS(facemasks, NNS) PP(under, IN) STOP

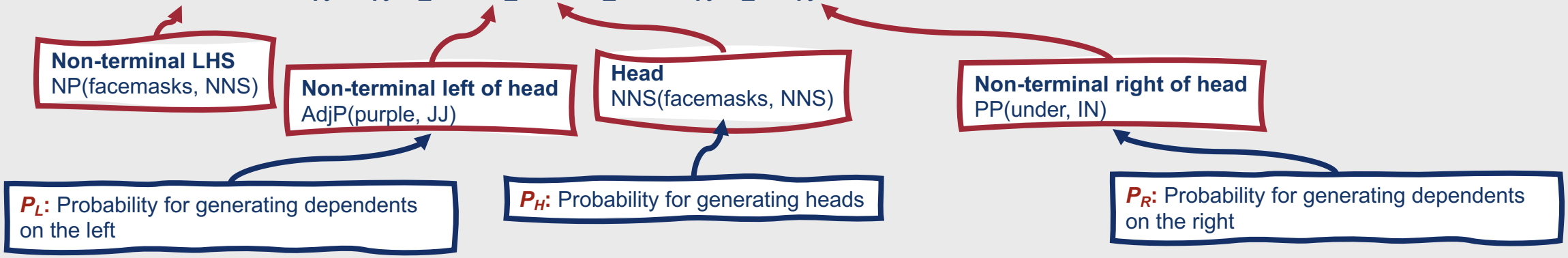
P_H(H|LHS) = P(NNS(facemasks, NNS) | NP(facemasks, NNS))

P_L(STOP|LHS H) = P(STOP | NP(facemasks, NNS) NNS(facemasks, NNS))

P_L(L_1|LHS H) = P(AdjP(purple, JJ) | NP(facemasks, NNS) NNS(facemasks, NNS))

The Collins Parser

- Consider the following generic production rule:
  - $LHS \rightarrow L_n\ L_{n-1} \dots L_1\ H\ R_1 \dots R_{n-1}\ R_n$



Grab the **purple facemasks under the disinfectant**.

NP(facemasks, NNS) → STOP AdjP(purple, JJ) NNS(facemasks, NNS) **PP(under, IN) STOP**

P_H(H|LHS) = P(NNS(facemasks, NNS) | NP(facemasks, NNS))

P_L(STOP|LHS H) = P(STOP | NP(facemasks, NNS) NNS(facemasks, NNS))

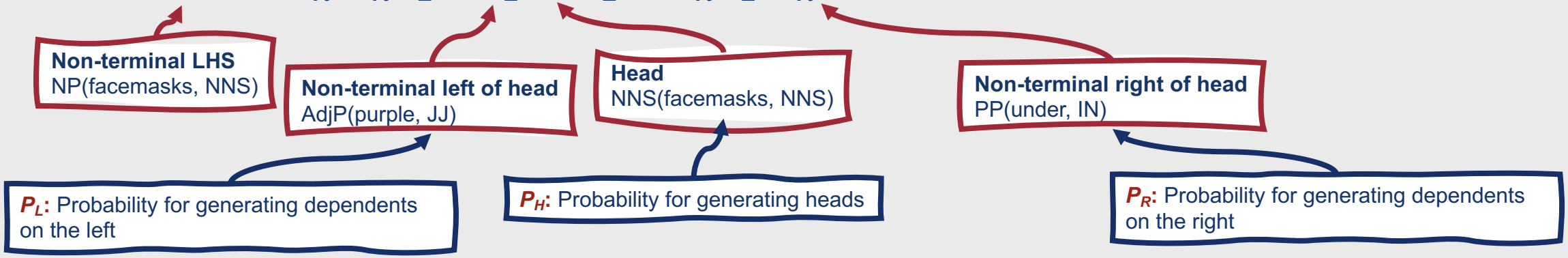
P_L(L_1|LHS H) = P(AdjP(purple, JJ) | NP(facemasks, NNS) NNS(facemasks, NNS))

P_R(R_1|LHS H) = P(PP(under, IN) | NP(facemasks, NNS) NNS(facemasks, NNS))

P_R(STOP|LHS H) = P(STOP | NP(facemasks, NNS) NNS(facemasks, NNS))

The Collins Parser

- Consider the following generic production rule:
  - $LHS \rightarrow L_n\ L_{n-1} \dots L_1\ H\ R_1 \dots R_{n-1}\ R_n$



Grab the **purple facemasks under the disinfectant**.

NP(facemasks, NNS) → STOP AdjP(purple, JJ) NNS(facemasks, NNS) **PP(under, IN) STOP**

$= P_H(H|LHS) * P_L(STOP|LHS\ H) * P_L(L_1|LHS\ H) * P_R(R_1|LHS\ H) * P_R(STOP|LHS\ H)$

$P_H(H|LHS) = P(NNS(facemasks, NNS) \mid NP(facemasks, NNS))$

$P_L(STOP|LHS\ H) = P(STOP \mid NP(facemasks, NNS)\ NNS(facemasks, NNS))$

$P_L(L_1|LHS\ H) = P(AdjP(purple, JJ) \mid NP(facemasks, NNS)\ NNS(facemasks, NNS))$

$P_R(R_1|LHS\ H) = P(PP(under, IN) \mid NP(facemasks, NNS)\ NNS(facemasks, NNS))$

$P_R(STOP|LHS\ H) = P(STOP \mid NP(facemasks, NNS)\ NNS(facemasks, NNS))$

Then, it's relatively easy to estimate the individual probabilities.

- Maximum likelihood estimate
- Much less subject to sparsity problems!

$P_R(R_1|\text{LHS H}) = P(\text{PP(under, IN)} \mid \text{NP(facemasks, NNS) NNS(facemasks, NNS)})$

$$\frac{\text{Count}(\text{NP}(\text{facemasks}, \text{NNS}) \textit{ with } \text{PP}(\text{under}, \text{IN}) \textit{ as a child to the right})}{\text{Count}(\text{NP}(\text{facemasks}, \text{NNS}))}$$