Define a CFG with \( N = \{ S, NP, VP, Det, N, PRP, V \} \) which can parse the sentences below:

- I like the cat
- He wants this job
- The cat ate her

\[
S \rightarrow NP \_VP \\
VP \rightarrow V \_NP \\
NP \rightarrow PRP \_| \_DET \_N
\]

New grammar:

\[
S \rightarrow NP^S \_VP^S \\
VP^S \rightarrow V \_NP^V \_VP \\
NP^S \rightarrow PRP \\
NP^S \rightarrow DET \_N \\
NP^V \_VP \rightarrow PRP \\
NP^V \_VP \rightarrow DET \_N
\]

Consider what happens to the size of the grammar.
2. Can we use terminal splitting to help us resolve PP attachment as in:

1. saw (the boy in the park) vs. 1. saw (the boy) in the park.

Terminal splitting can't help us...
The important information is the relationship between the lexical items 'saw', 'boy' and 'park'.
1. Write an equation for the probability of:

\[ VP(\text{dumped}, \text{VBD}) \rightarrow \text{VBD}(\text{dumped}, \text{VBD}) \] \[ \text{NP}(\text{sacks}, \text{NNS}) \] \[ \text{PP(} \text{into}, \text{P}) \]

i.e. the probability of this tree piece:

- \[ VP(\text{dumped}, \text{VBD}) \]
- \[ \text{VBD}(\text{dumped}, \text{VBD}) \]
- \[ \text{NP}(\text{sacks}, \text{NNS}) \]
- \[ \text{PP(} \text{into}, \text{P}) \]

First generate the head:

\[ P_h(\text{H}|\text{LHS}) = P(\text{VBD}(\text{dumped}, \text{VBD}) | \text{VP(} \text{dumped}, \text{VBD})) \]

Next generate the left elements of the head:

\[ P_{l1}(\text{STOP} | \text{VP(} \text{dumped}, \text{VBD})), \text{VBD}(\text{dumped}, \text{VBD})) \]

Then generate the right elements of the head:

\[ P_{r1}(\text{NP(} \text{sacks}, \text{NNS}) | \text{VP(} \text{dumped}, \text{VBD})), \text{VBD}(\text{dumped}, \text{VBD})) \]

\[ P_{r2}(\text{PP(} \text{into}, \text{P}) | \text{VP(} \text{dumped}, \text{VBD})), \text{VBD}(\text{dumped}, \text{VBD})) \]

\[ P_{r3}(\text{STOP} | \text{VP(} \text{dumped}, \text{VBD})), \text{VBD}(\text{dumped}, \text{VBD})) \]

Multiply all these together for the probability of the whole rule.

Note that these probabilities are easier to estimate from small amounts of data. E.g. \( P_{r2} \) is estimated from all time ‘dumped into’ occurs and is independent of what was dumped.