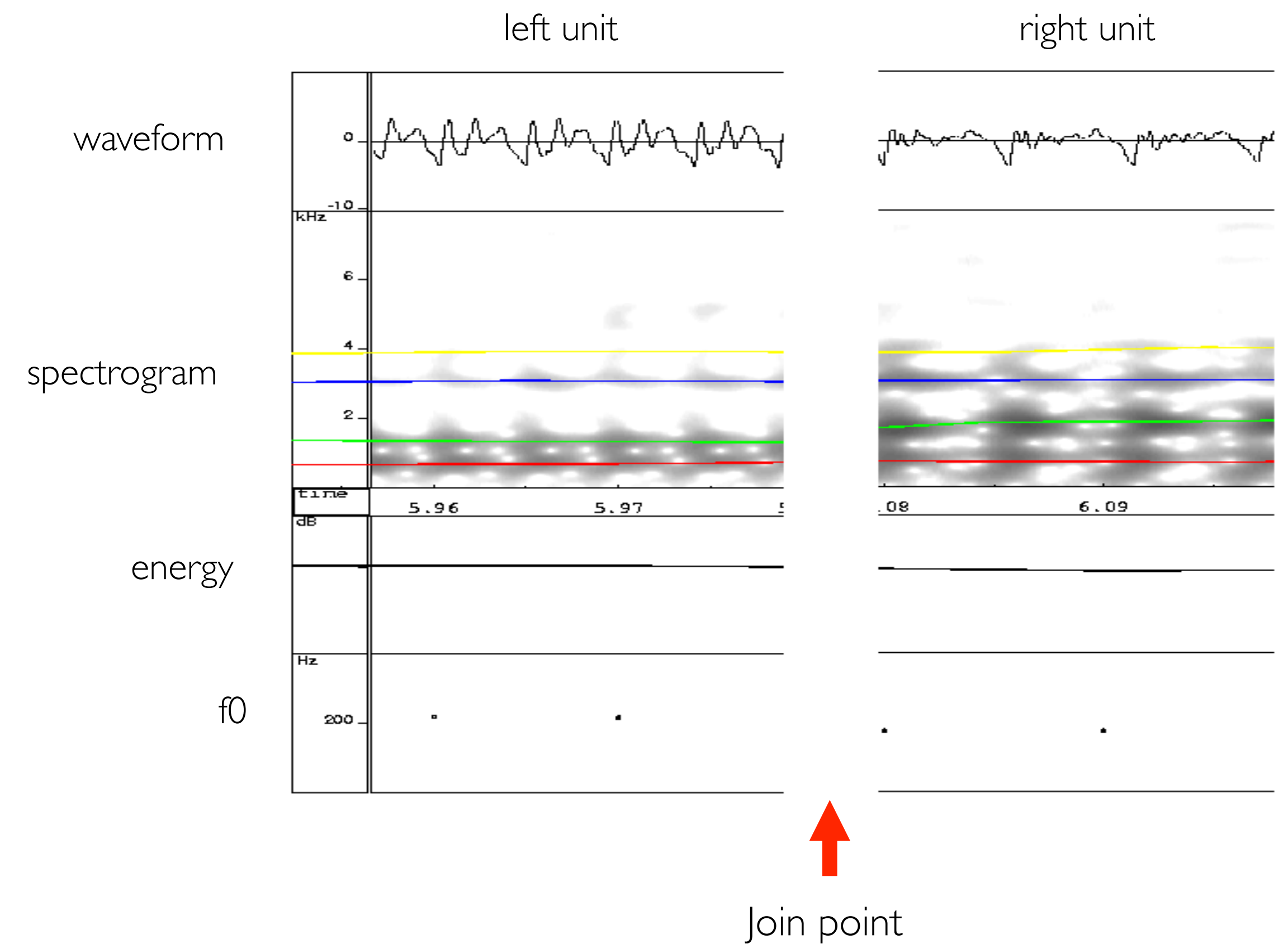


Unit selection

- discussion points

Join cost

- What acoustic features should we use?
- Any other improvements you can think of?
- What knowledge could you inject into the cost function?
 - e.g., phonetic

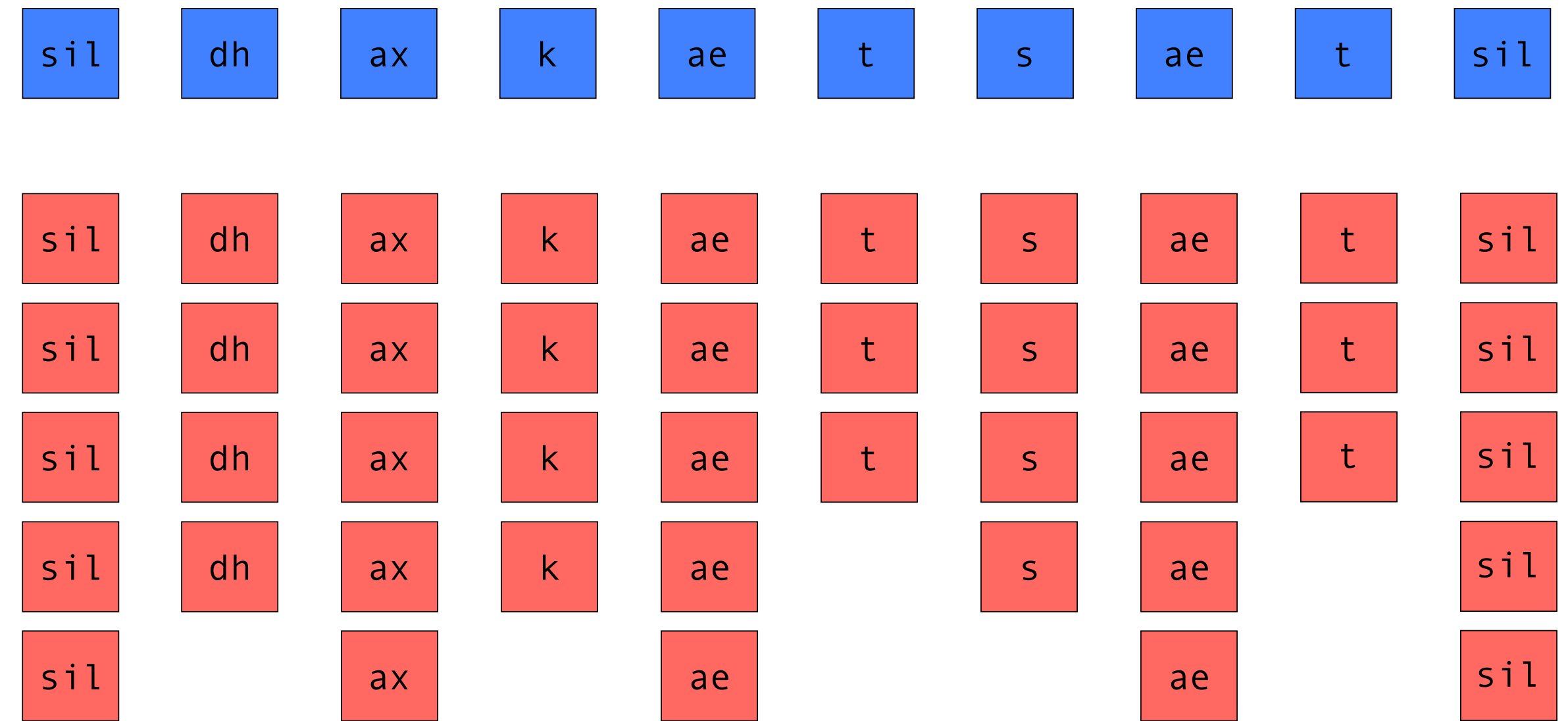


Join cost

- Which of these would be a better place to make a join, and why?
 - vowel
 - diphthong
 - stop
 - approximant
 - unvoiced fricative

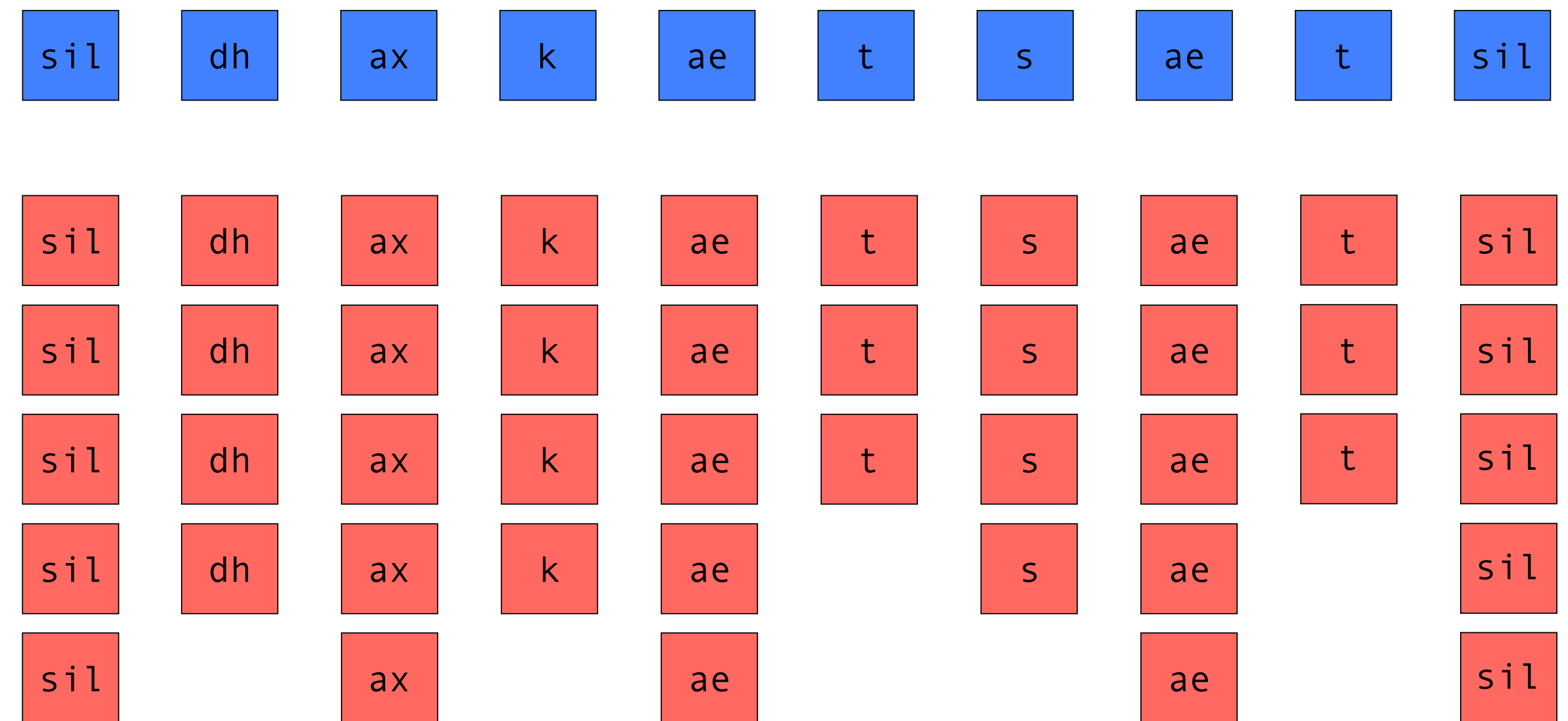
Can you make the search *faster*?

- Main computations are
 - target cost
 - join cost
- How many of each are computed per sentence being synthesised?



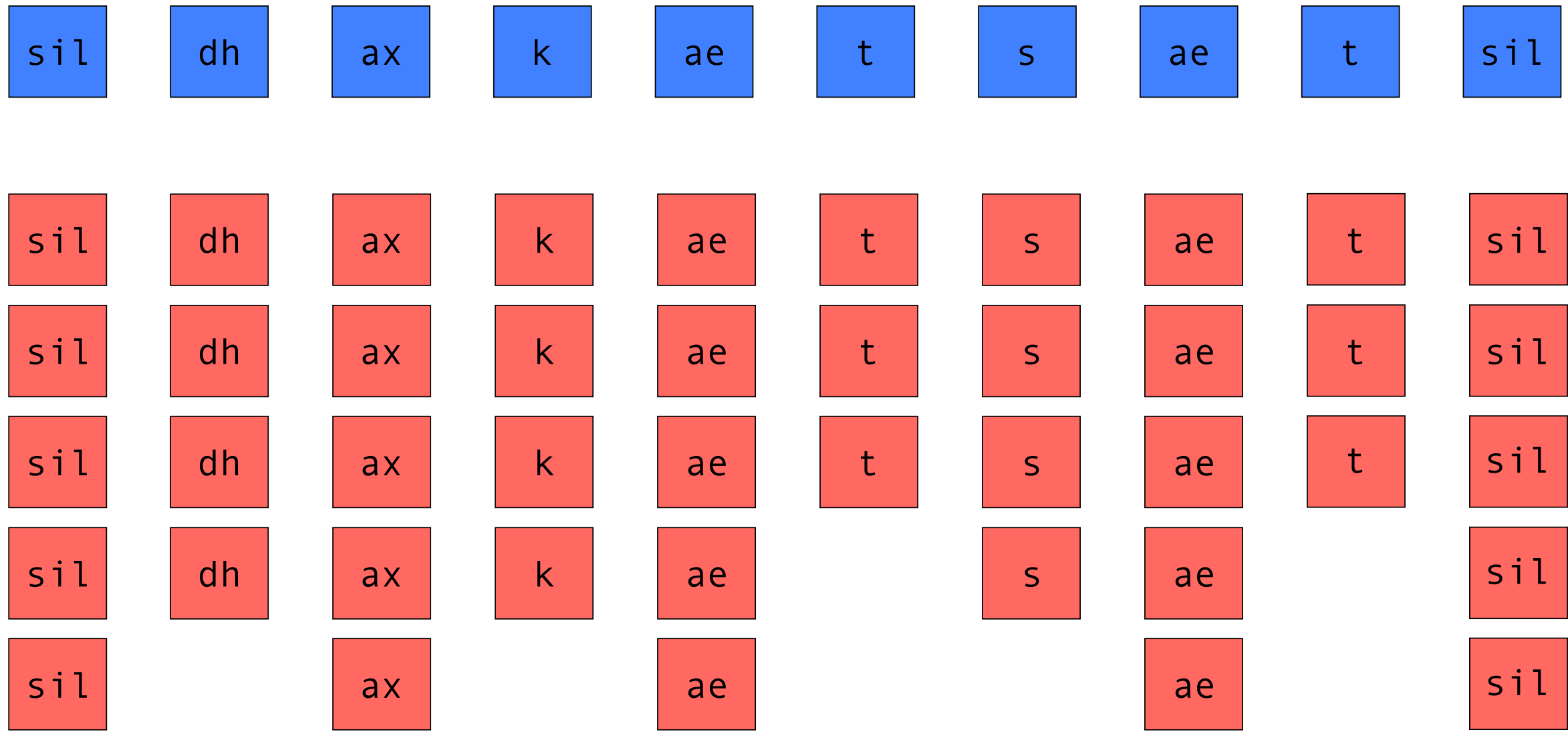
Can you make the search *faster*?

- What *dimensions* of the problem does computation time depend on?
- exactly **how** does compute time depend on them (linear? worse than linear?)
- Can you **reduce** those dimensions?
- What will the effect be on **quality**?



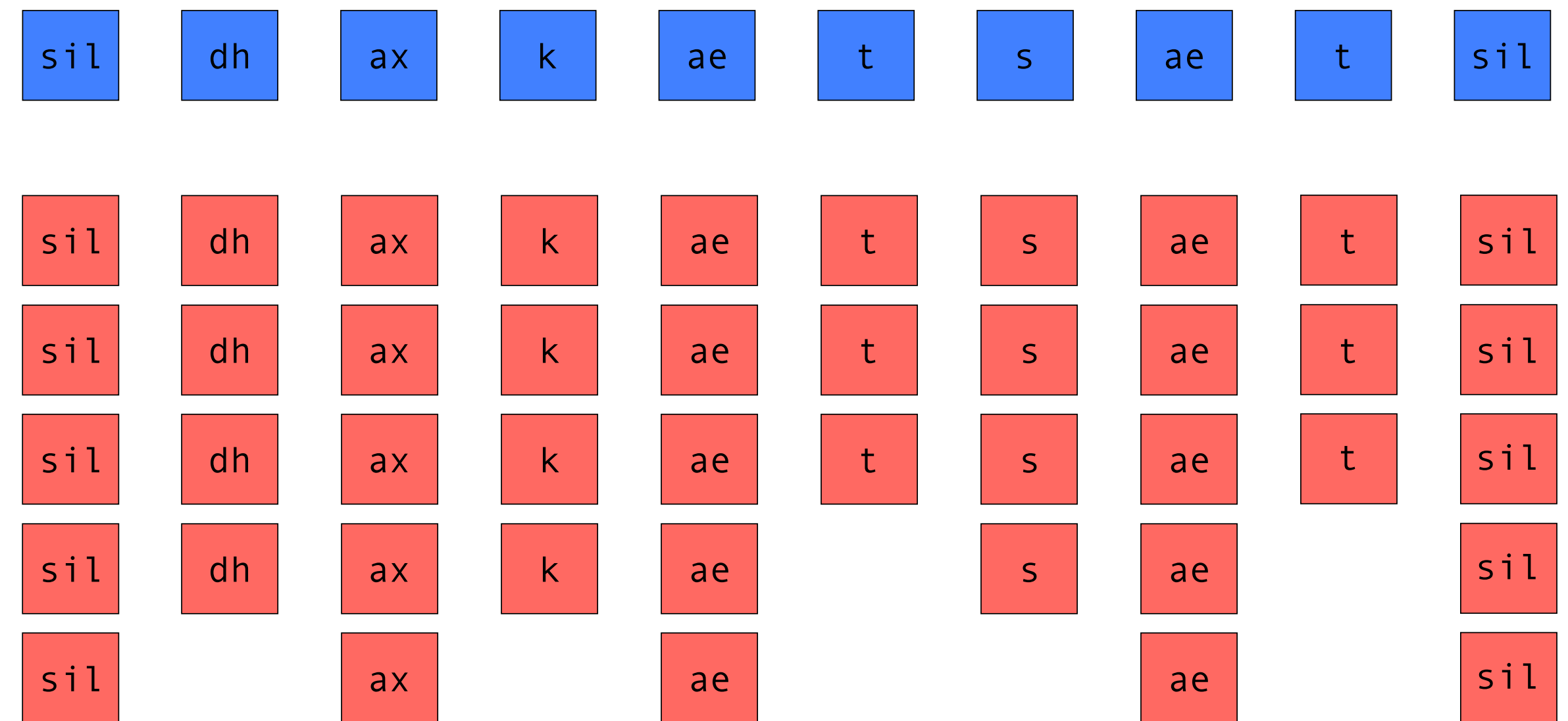
Can you make the search *faster*?

- Can we **cache** the target and/or join costs?
- i.e., save them, then re-use later for a different sentence, *without* recomputing



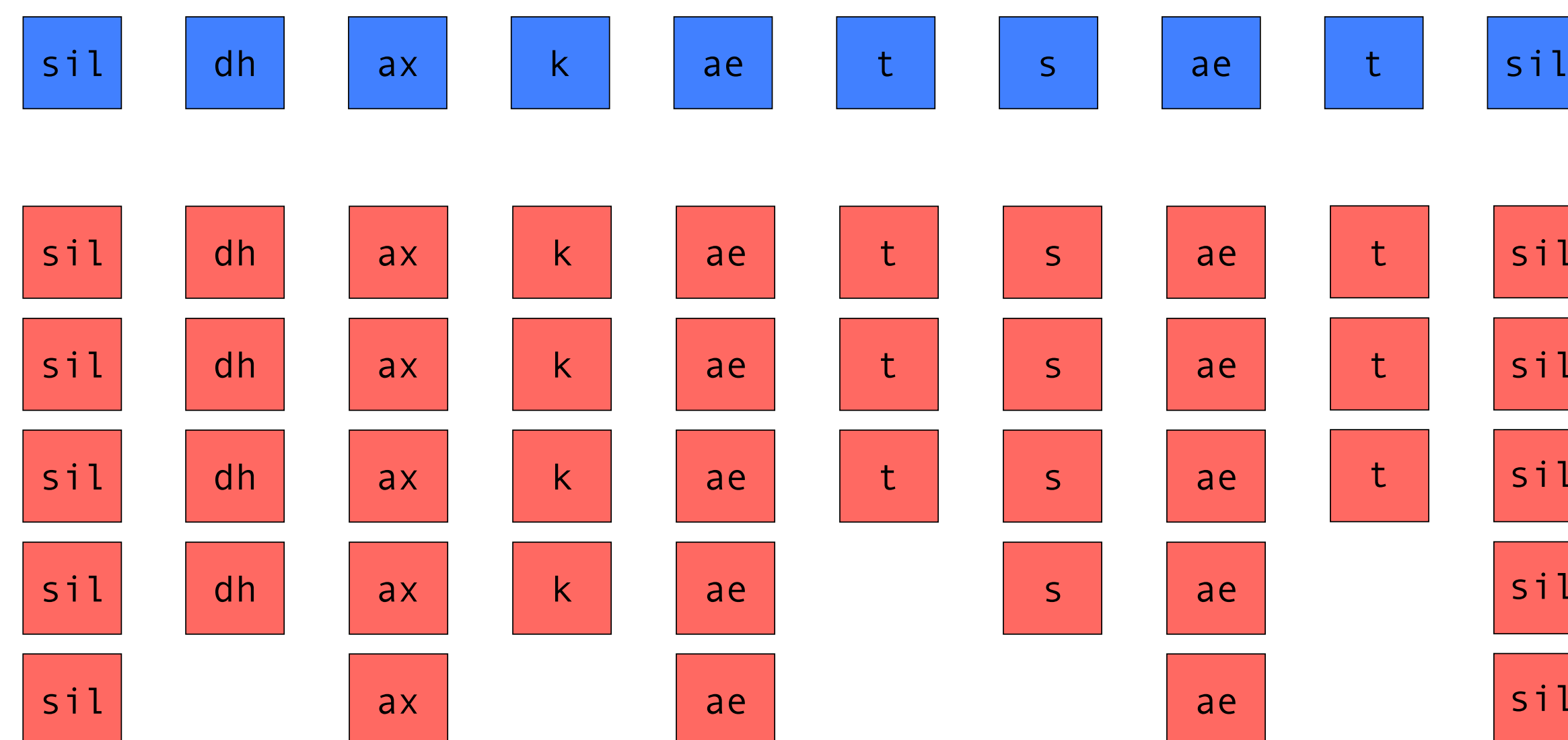
Target cost

- What would happen if the target cost function considered the linguistic specifications of **neighbouring target units**?
- What would happen if the target cost function considered the linguistic specifications of **neighbouring candidate units**?



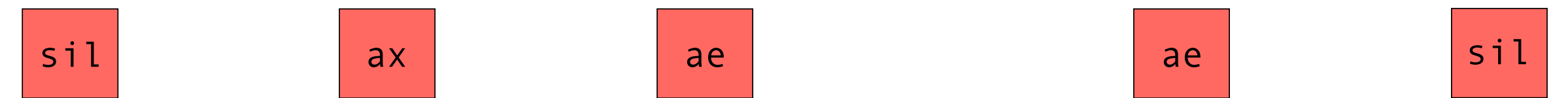
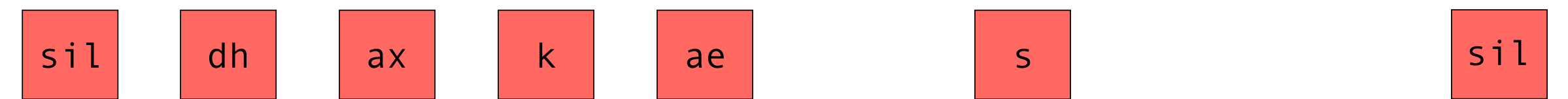
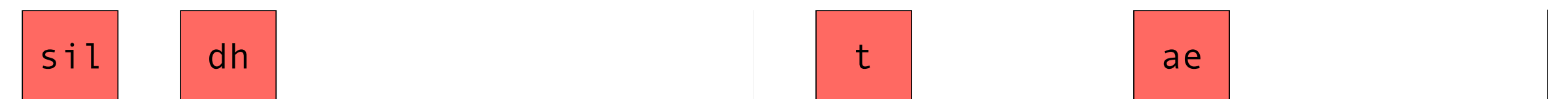
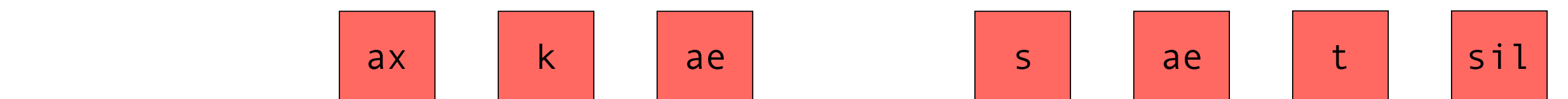
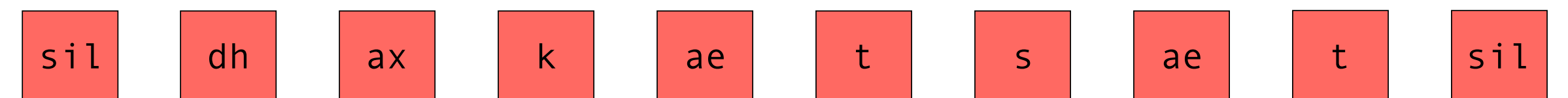
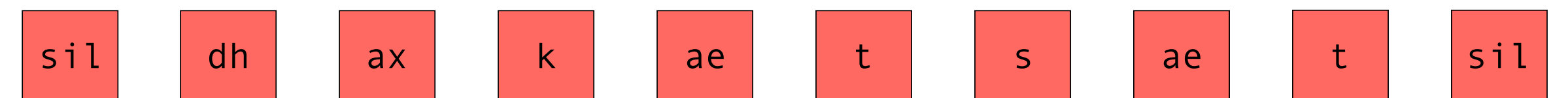
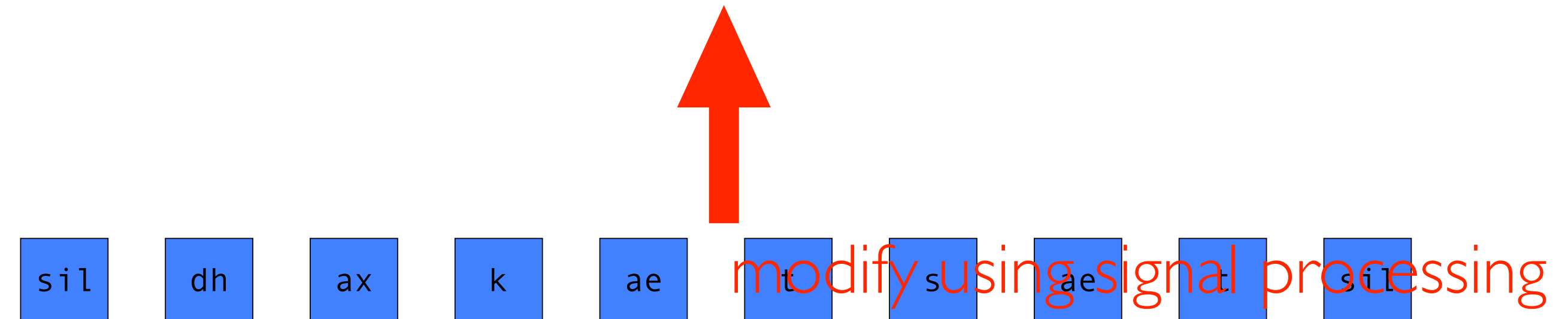
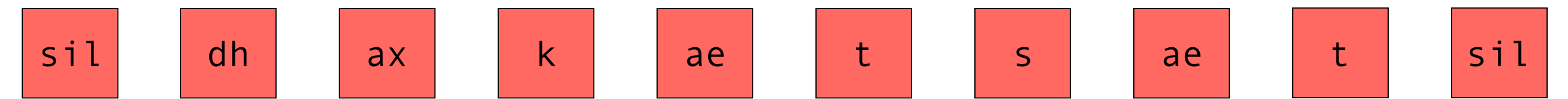
Join cost

- What would happen if the join cost function considered the acoustic properties of **more than two adjacent candidate units**?



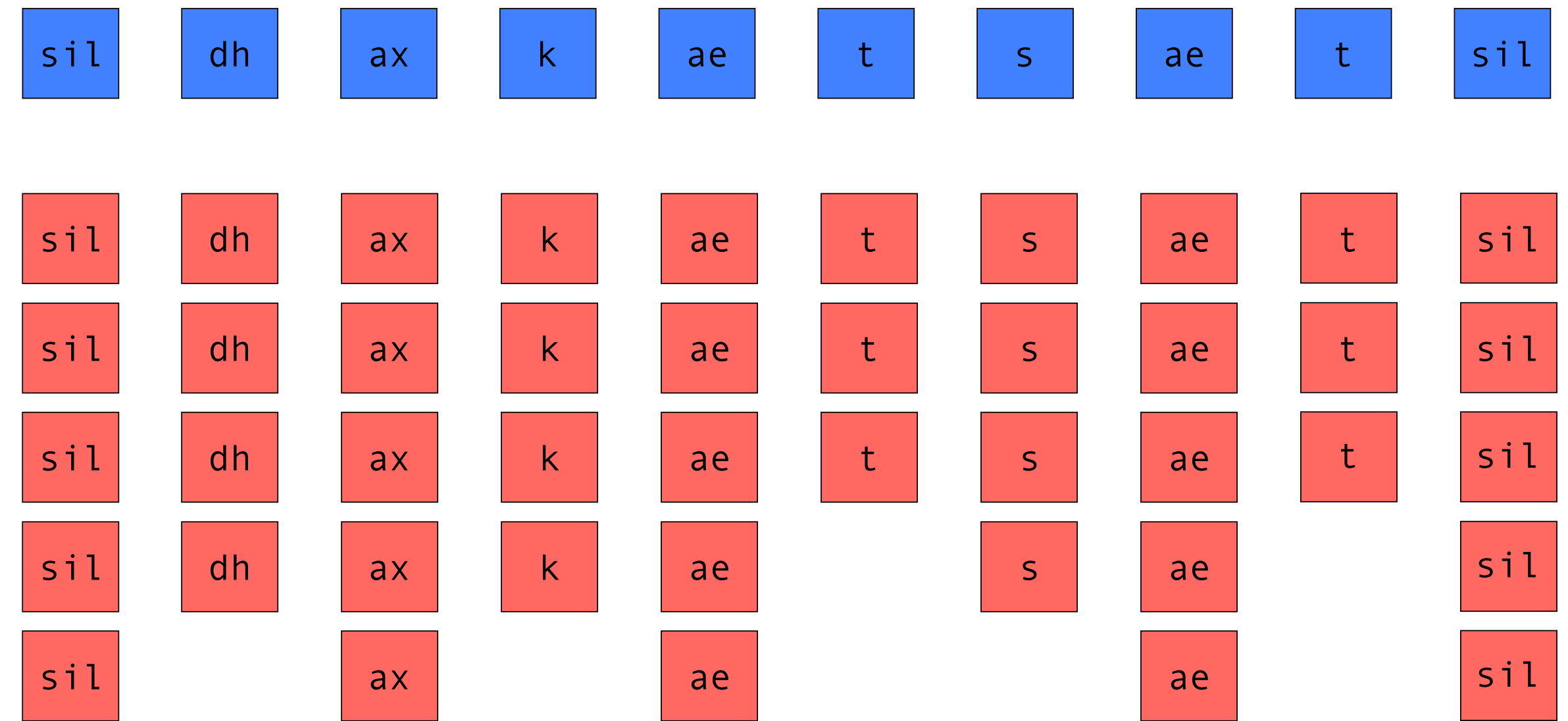
Signal processing

- We can **modify** the selected sequence of candidates after they have been selected
 - join smoothing
 - prosody modification
-
- What effect, if any, might this have on the selection procedure?
 - join cost
 - target cost
 - search



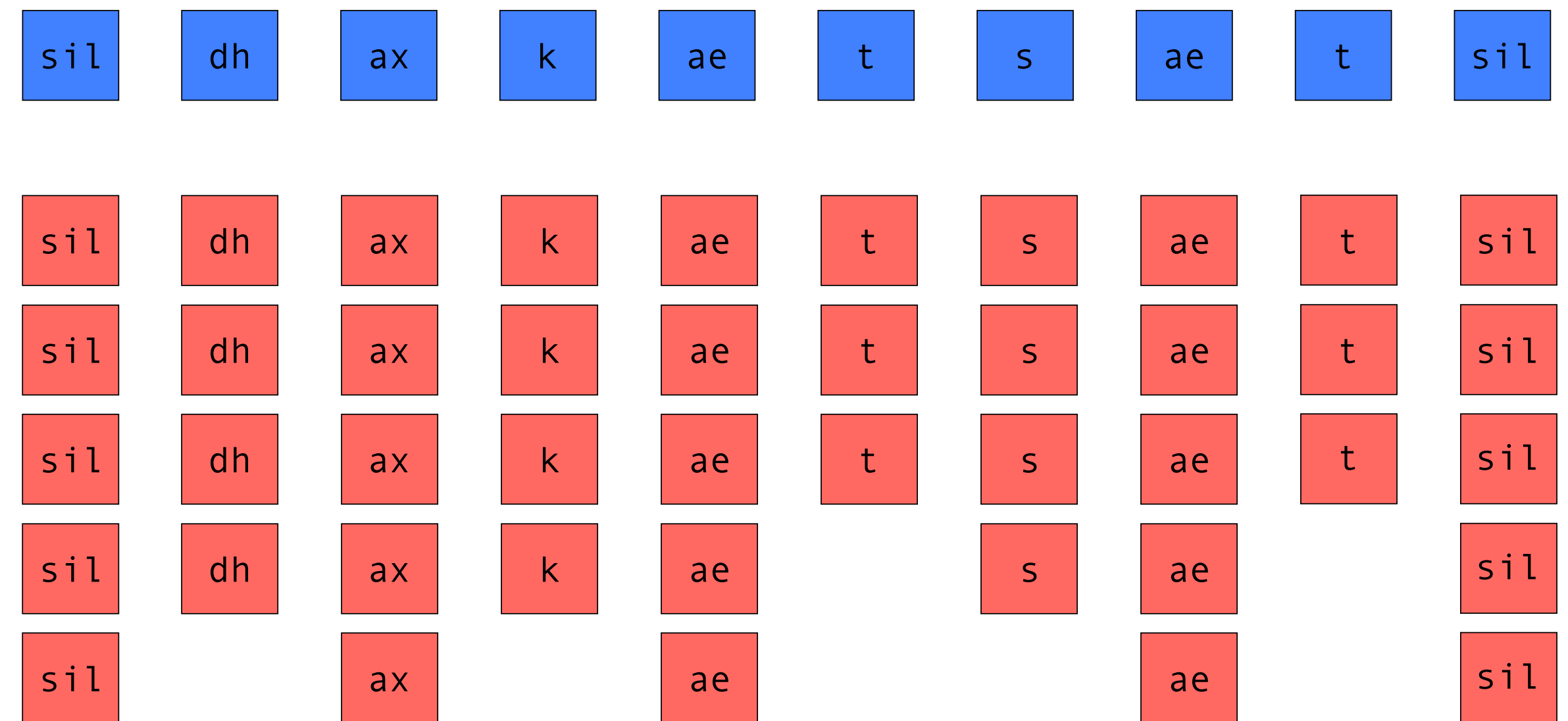
“Looking forward and backwards”

- The search proceeds left-to-right
- Does this mean that the cost functions only “look” in one direction?

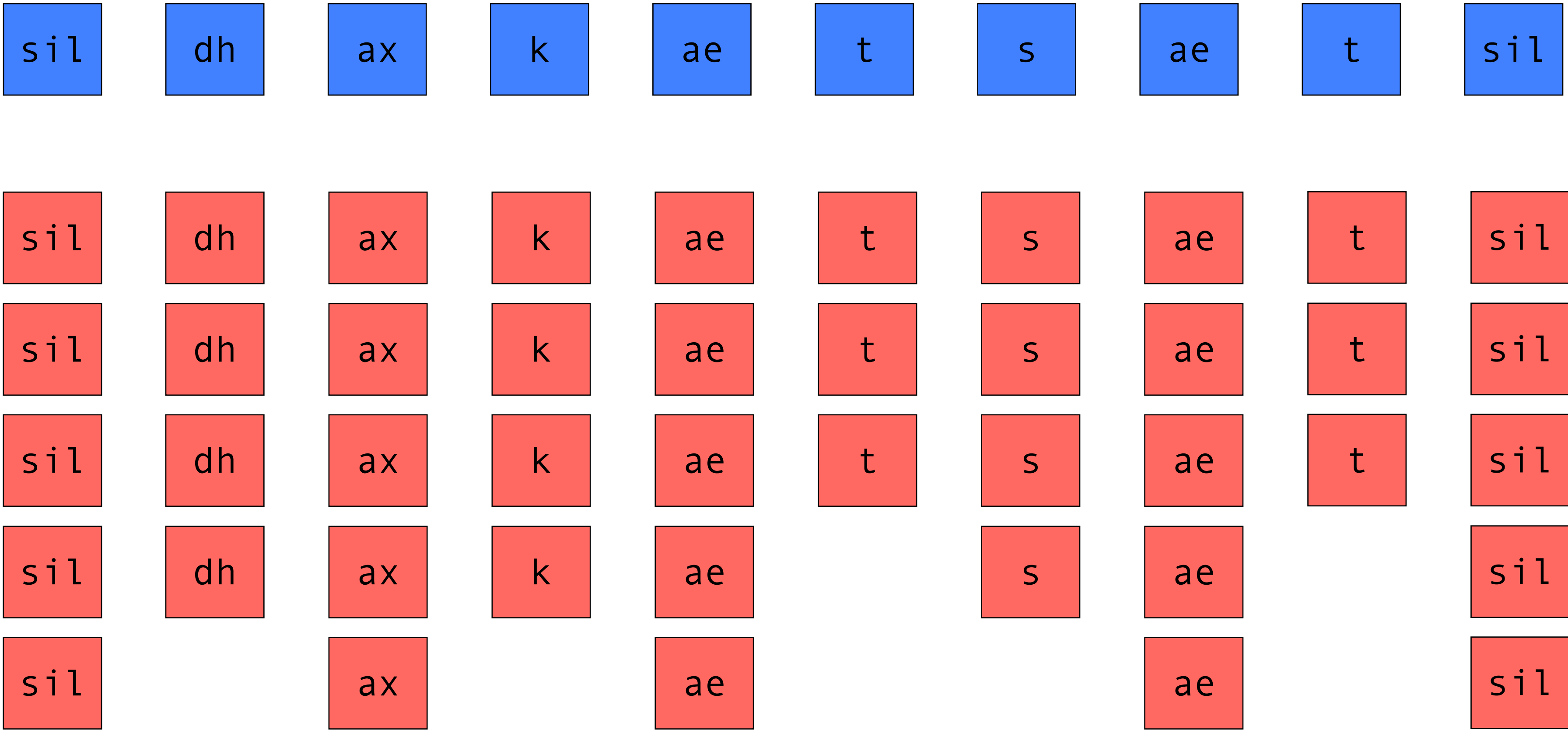


Relative importance: target vs. join

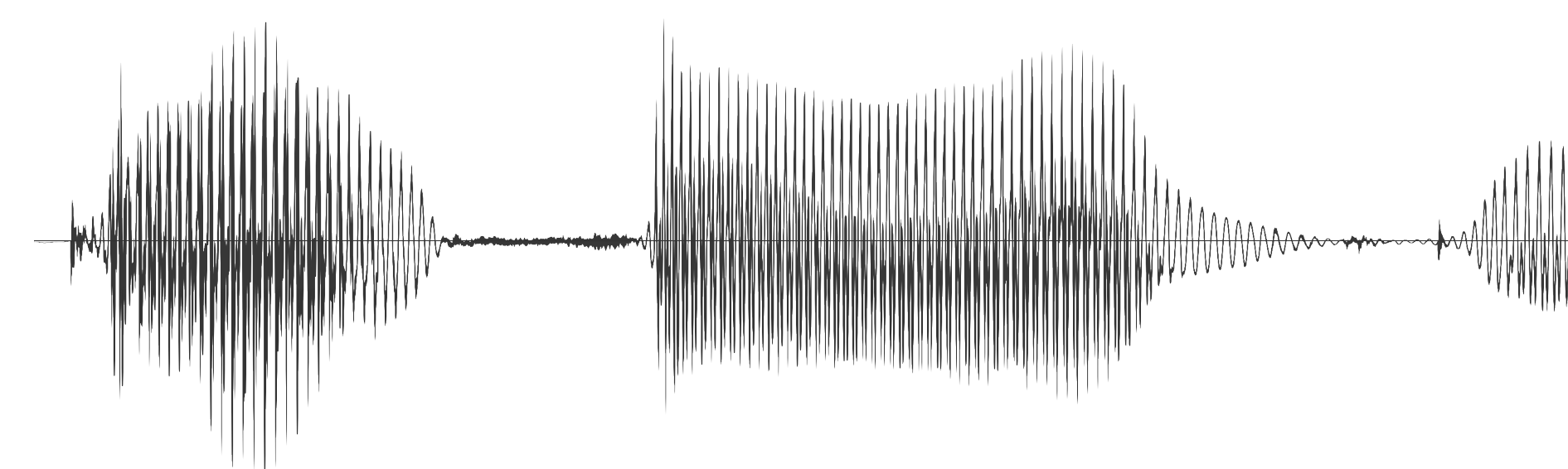
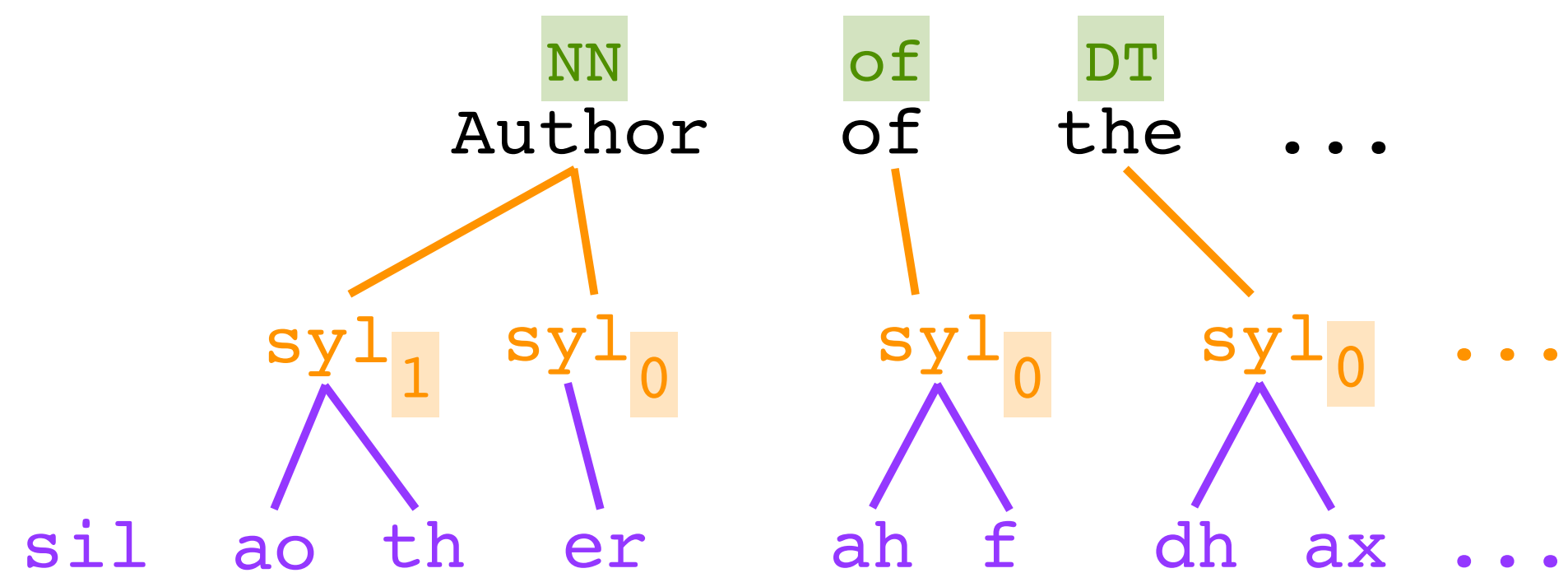
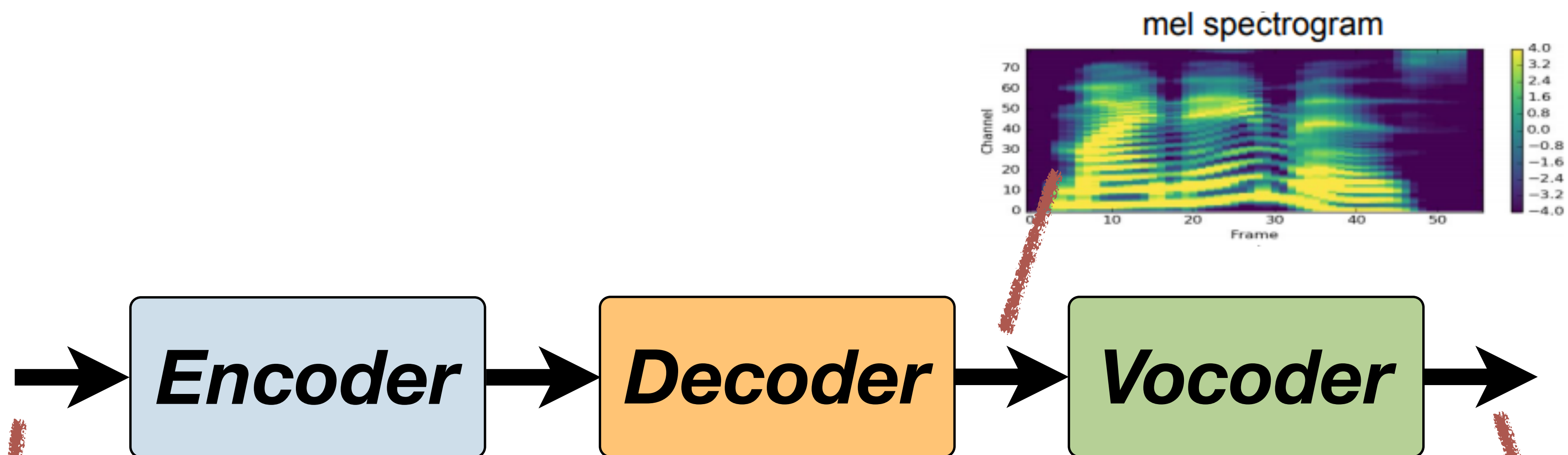
- The overall cost is a sum of target and join costs
- We can weight one relative to the other
- What would be the effects of
 - higher weighting on the **target** cost
 - higher weighting on the **join** cost



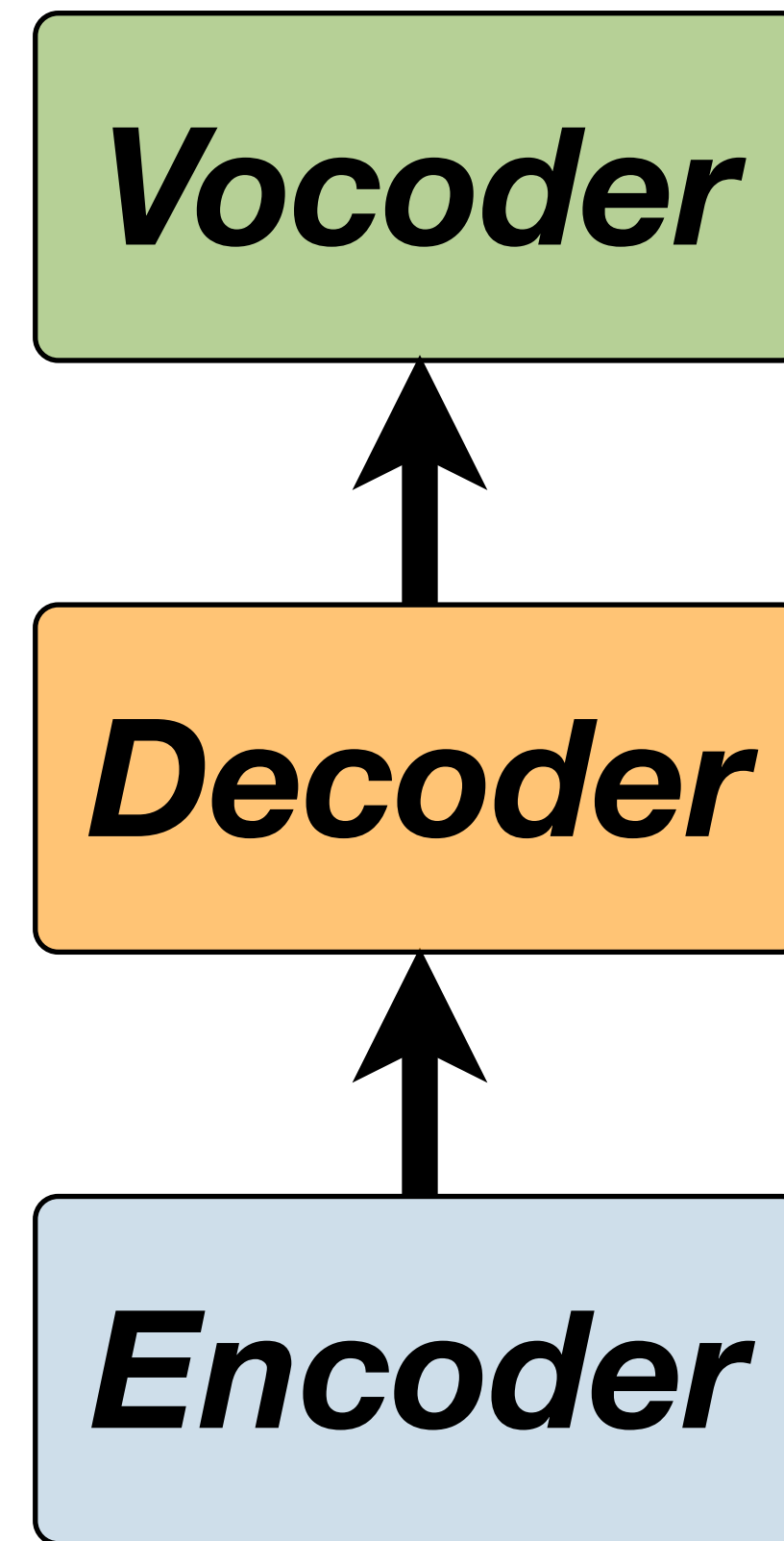
How does this diagram compare to HMM-based ASR ?



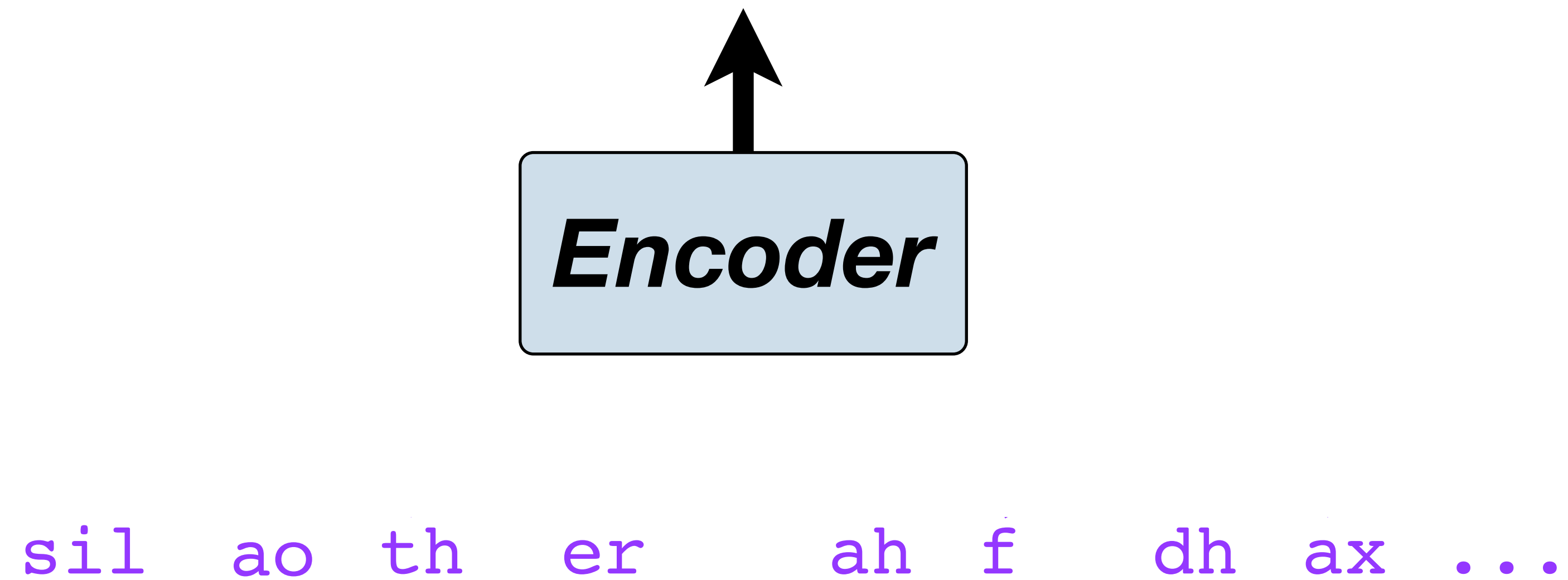
From unit selection to a sequence-to-sequence model



From unit selection to a sequence-to-sequence model



From unit selection to a sequence-to-sequence model



From unit selection to a sequence-to-sequence model

left context: ao th
right context: <w> ah
position in phrase: initial
syllable stress: unstressed

sil ao th er ah f dh ax ...