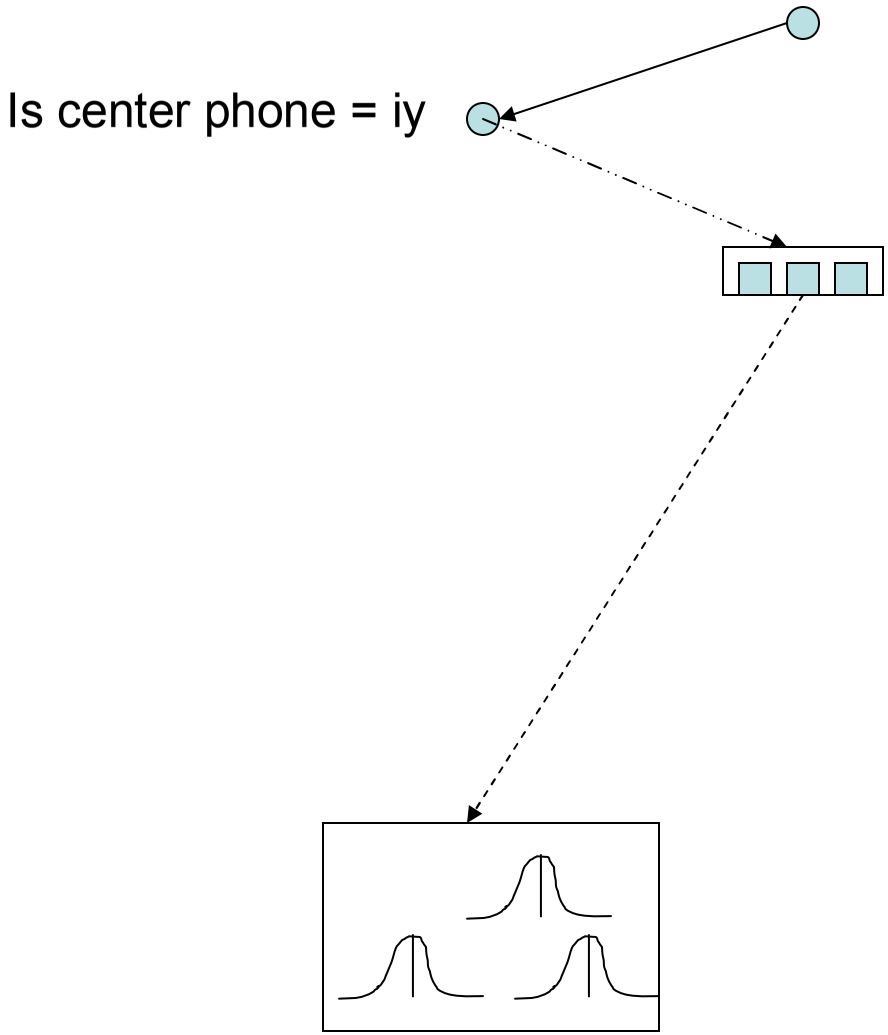
Phonetic decision tree in Janus

Based on the paper
from Michael Finke and Ivica Rogina
“Wide Context Acoustic Modeling in Read vs. Spontaneous Speech”

Train a context dependent (CD) system with Janus

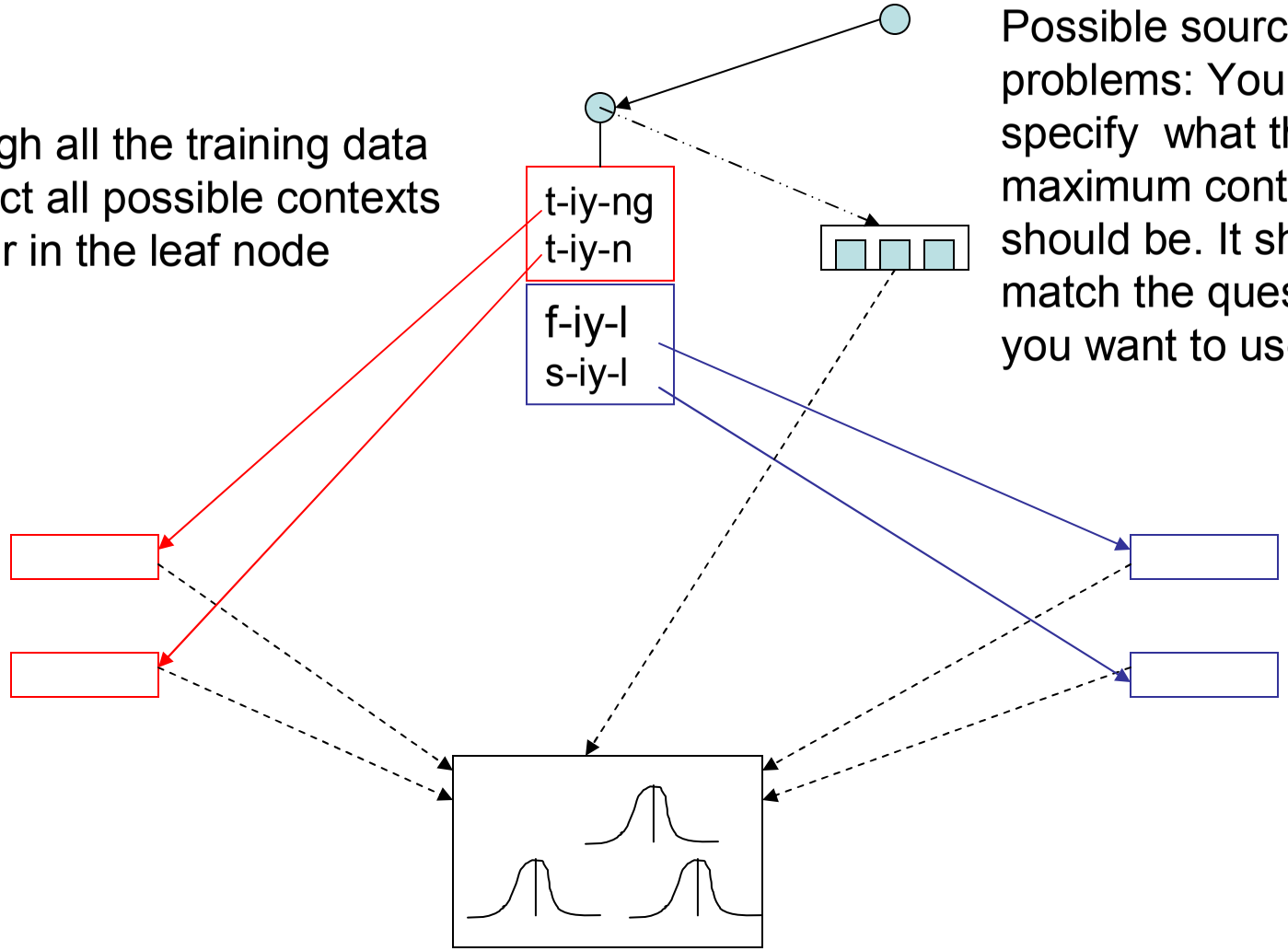
- Create context independent (CI) decision tree (simple script)
- Train CI system
- Add polyphone tree to CI - decision tree
- Train the distribution of polyphone tree
- Cluster polyphones → CD – decision tree
- Do training with the CD - decision tree

Trained CI system



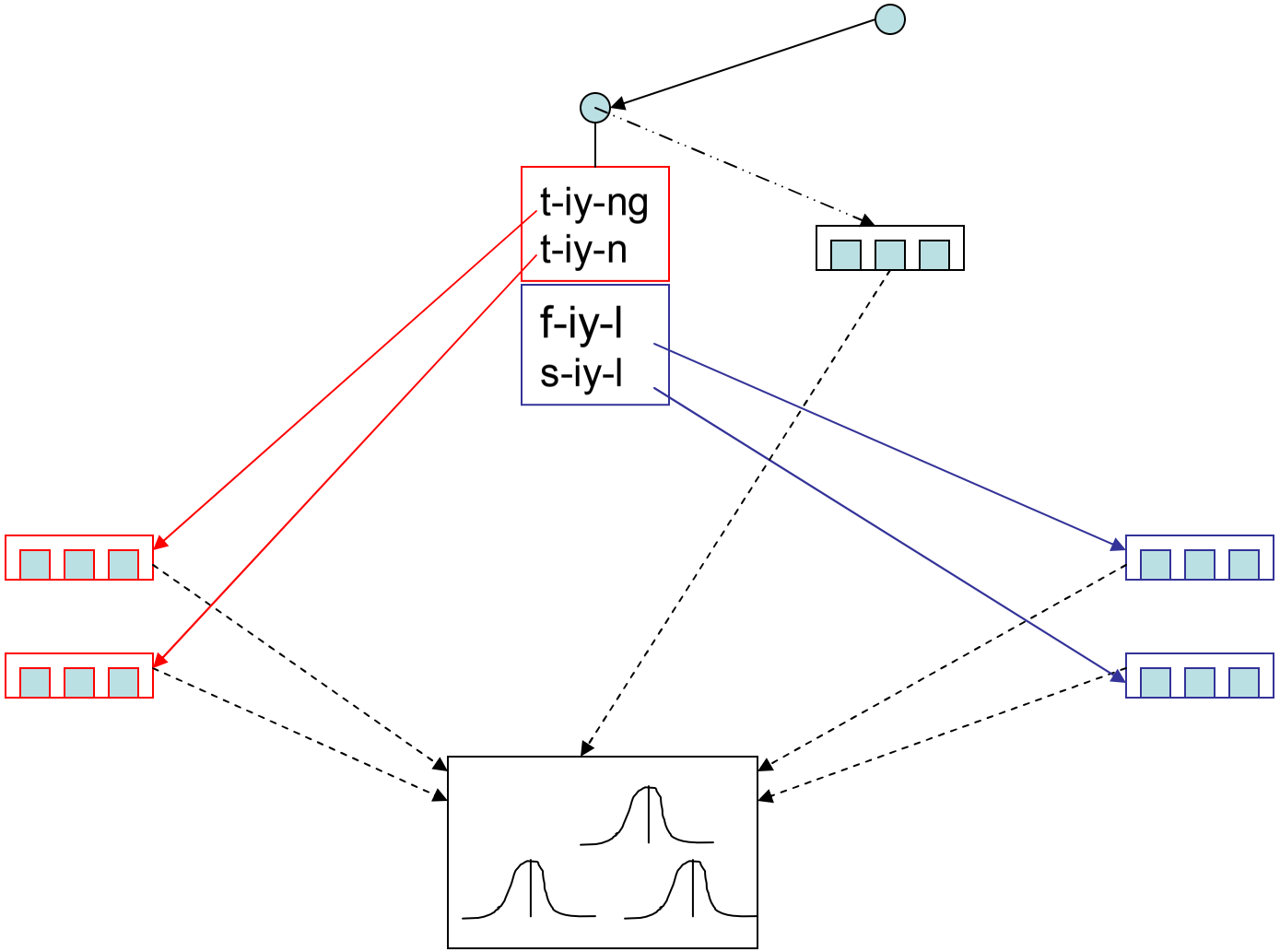
Add the polyphone tree (PTree)

Go through all the training data and collect all possible contexts that occur in the leaf node



Possible source of problems: You have to specify what the maximum context size should be. It should match the questions you want to use.

Train the distribution of PTree



Cluster criterion

Counts of component *i* in model *m*

$$p_i^l = \frac{1}{\gamma^l} \sum_{m \in L} \gamma_m \alpha_{mi}, \quad \gamma^l = \sum_{m \in L} \gamma_m$$

$$p_i^r = \frac{1}{\gamma^r} \sum_{m \in R} \gamma_m \alpha_{mi}, \quad \gamma^r = \sum_{m \in R} \gamma_m$$

Counts of model *m*

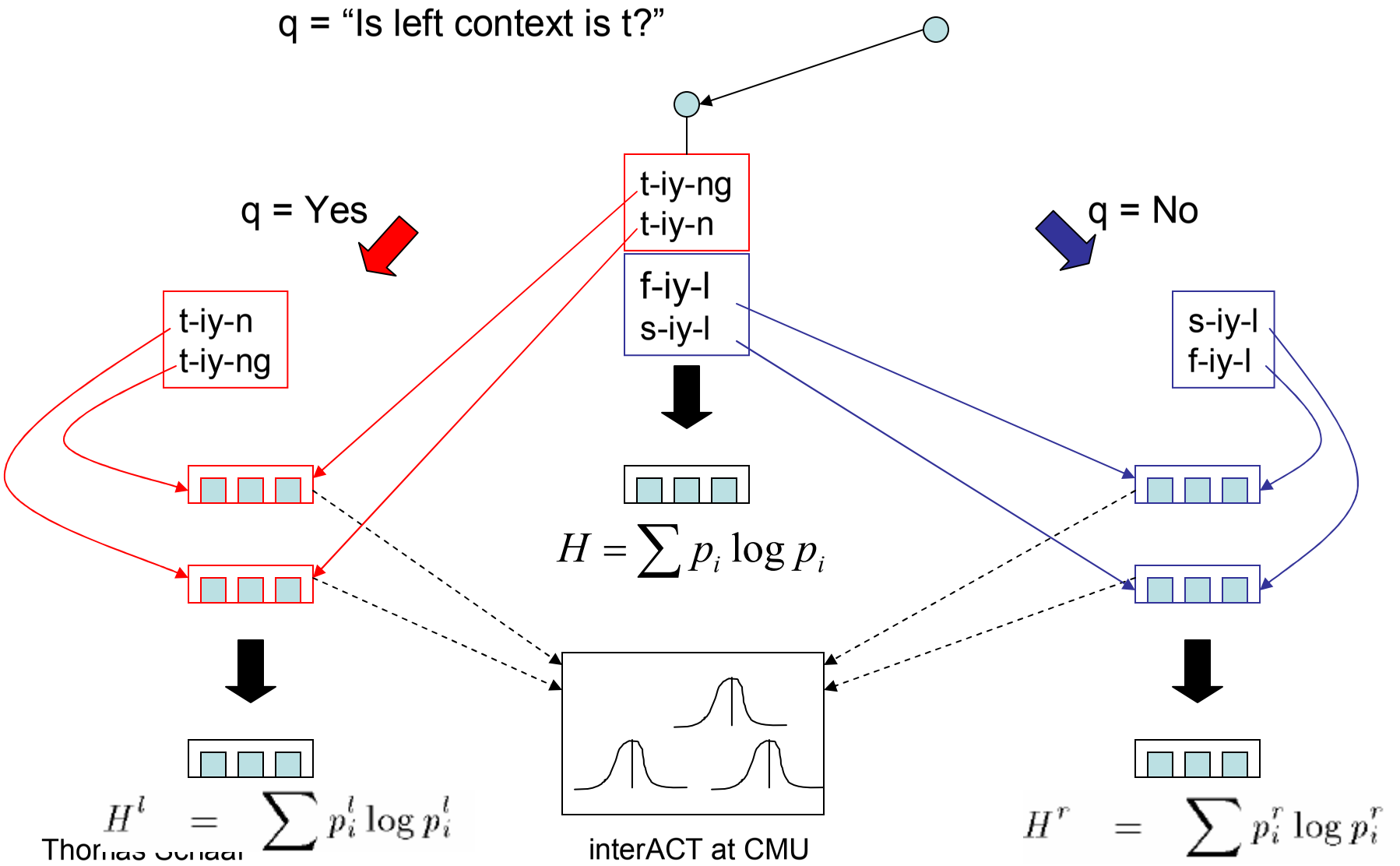
$$D(q) = \gamma^l H^l + \gamma^r H^r - \gamma H$$

$$-H^l = \sum p_i^l \log p_i^l$$

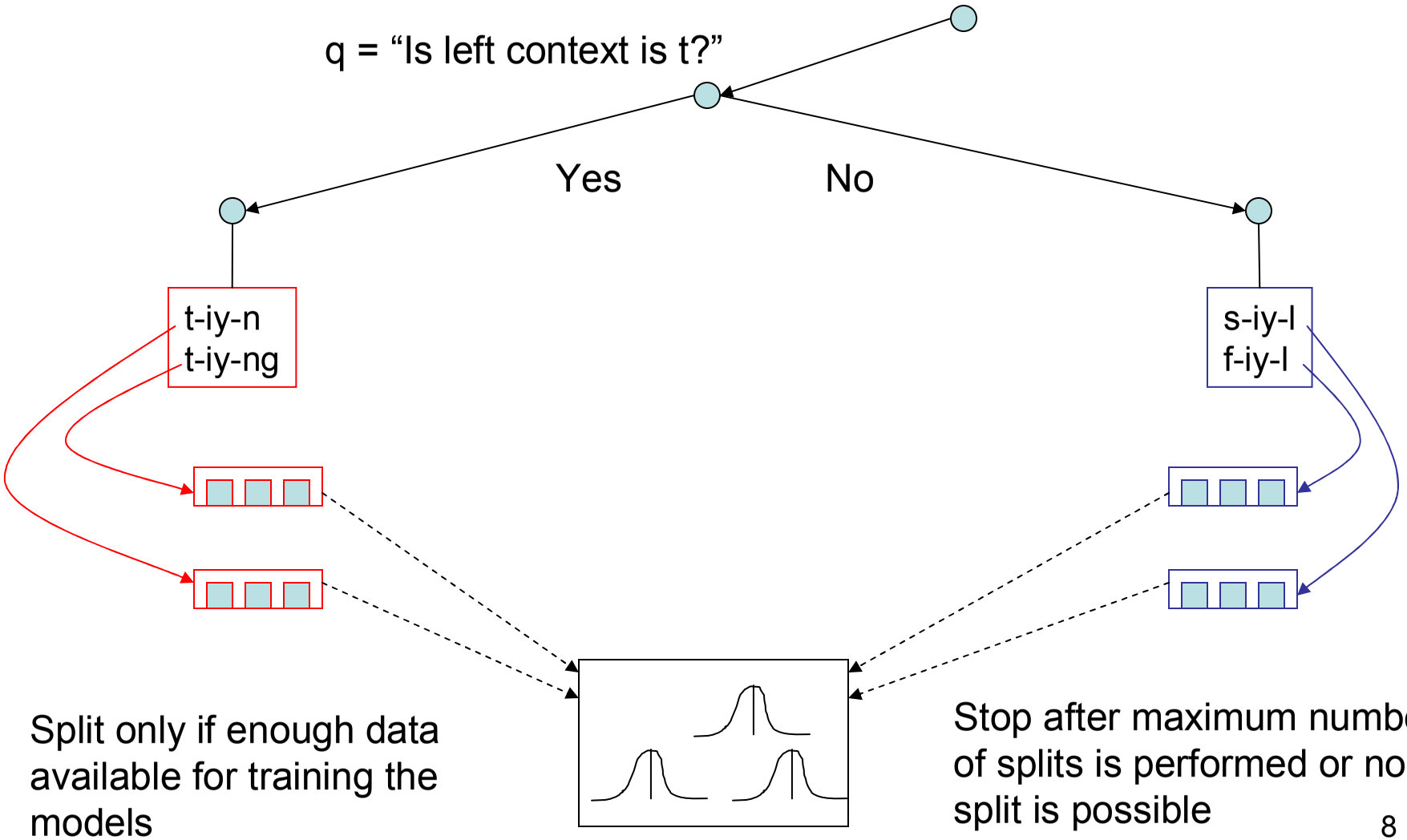
$$-H^r = \sum p_i^r \log p_i^r$$

Question splits distribution into *L* and *R*

Example: Question q splits context



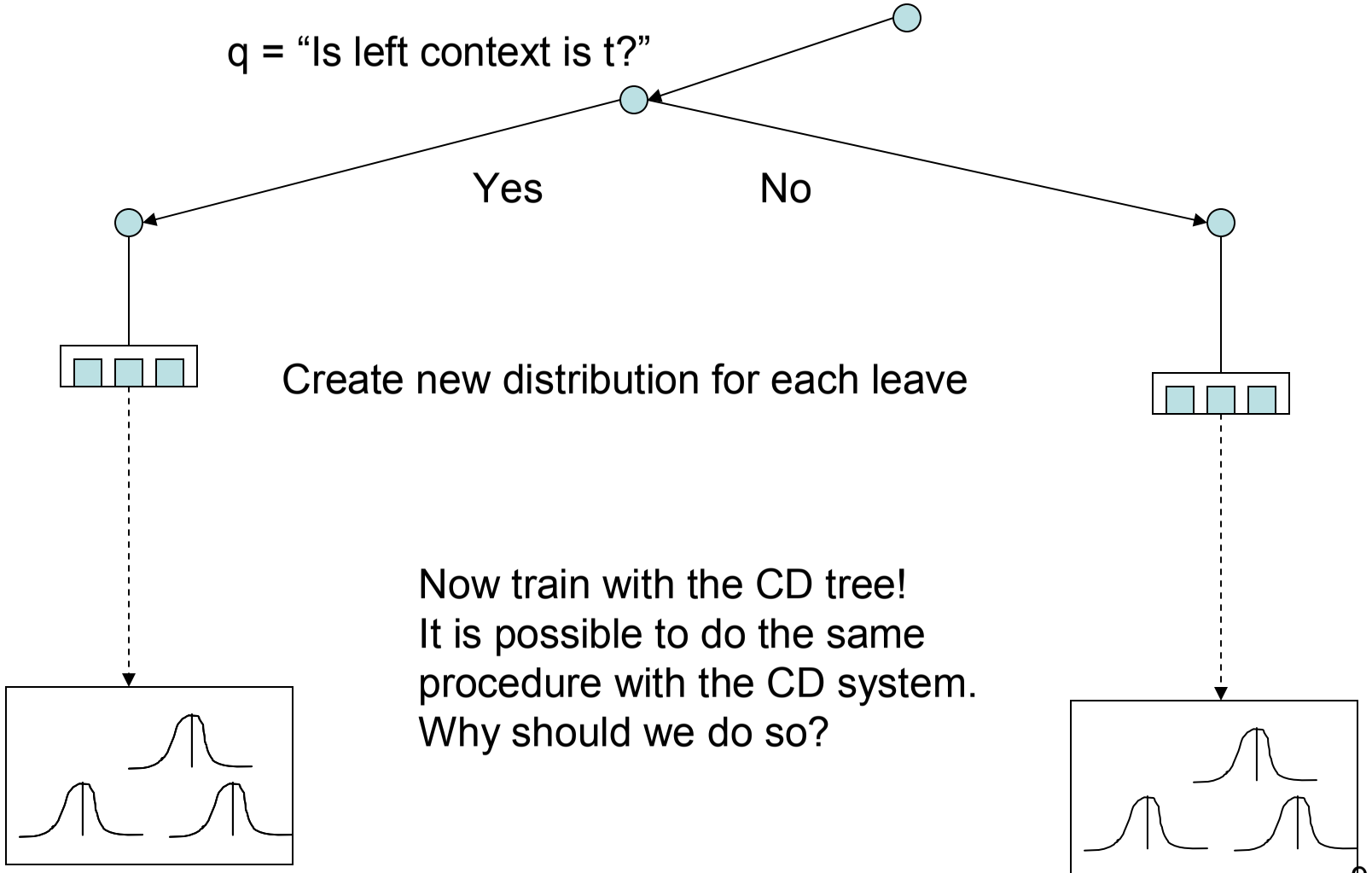
Example: Question q splits context



Split only if enough data available for training the models

Stop after maximum number of splits is performed or no split is possible

Remove PTree and split codebooks



Now train with the CD tree!
It is possible to do the same procedure with the CD system.
Why should we do so?

Results over context width

Task	Context ±1	Context ±2	Context ±3
WSJ	20.9% WE	20.2% WE	19.9% WE
SWB	46.0% WE	43.6% WE	

Main differences between HTK and Janus

- Distance:
  - HTK: Measures improvement in log-likelihood. Uses a single Gaussian to do so.
  - Janus: Maximum Entropy gain on the mixture weight distribution.
- Parameter reduction:
  - HTK: agglomerative clustering after spliting
  - Janus: do clustering in two steps

Some questions not discussed

- What is a useful context width?
- How many Gaussians should I use in the CI system?
- Can I do merge an split training for the CI system?
- What questions should I ask?

Two Stage Clustering