Hidden Markov Models

Natalie Parde
UIC CS 421

What are Hidden Markov Models (HMMs)?

Probabilistic generative models for **sequences**

Make predictions based on an underlying set of **hidden states**

How does sequence labeling differ from other types of classification?

- Machine learning often addresses the problem of classifying text into discrete, predefined groups

Spam

Not Spam

Dear Esteemed Professor Dr. ***Natalie Parde***,
I am interested in applying to ***University of Illinois – Chicago*** for a **Ph.D.** in **Computer Science** in the area of ***Artificial Intelligence*** and ***Natural Language Processing***. I read your recent paper ***"Exploring MMSE Score Prediction Using Verbal and Non-Verbal Cues***" and see that you are interested in **Score Prediction** and **Verbal and Non-Verbal Cues**….

Standard Classification Assumption: Individual cases are disconnected and independent.

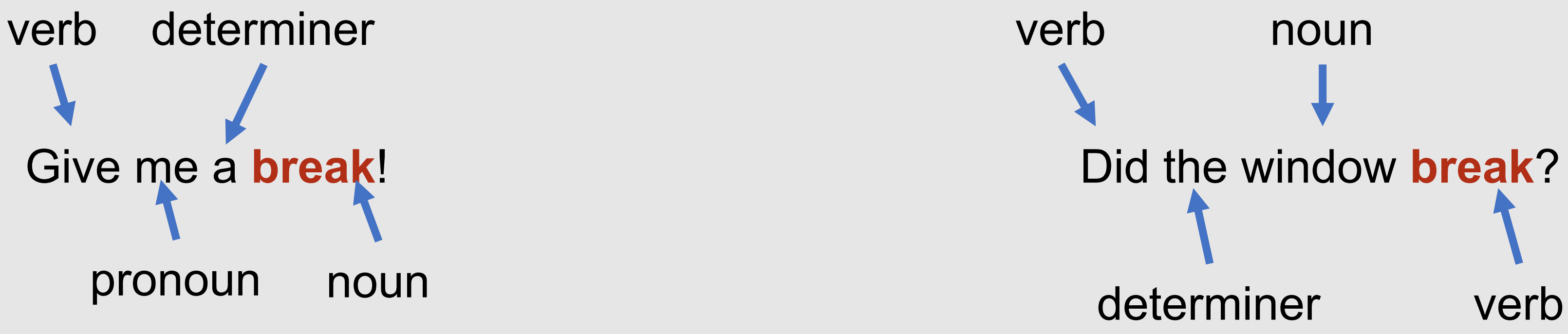
However, many NLP problems do not satisfy this assumption.

Instead, they involve many interconnected decisions, each of which resolve different ambiguities despite being mutually dependent.

For these problems, different learning and inference techniques are needed!

Sequence Labeling

- Many NLP problems can be viewed as **sequence labeling** tasks.
- Objective: Find the label for the next item, based on the labels of other items in the sequence.



Applications that can benefit from sequence labeling?

- Named entity recognition
- Semantic role labeling

person organization

Natalie Parde works at the **University of Illinois at Chicago** and lives in **Chicago, Illinois**.

location

agent source destination

Natalie drove for 15 hours from **Dallas** to **Chicago** in her hail-damaged **Honda Accord**.

instrument

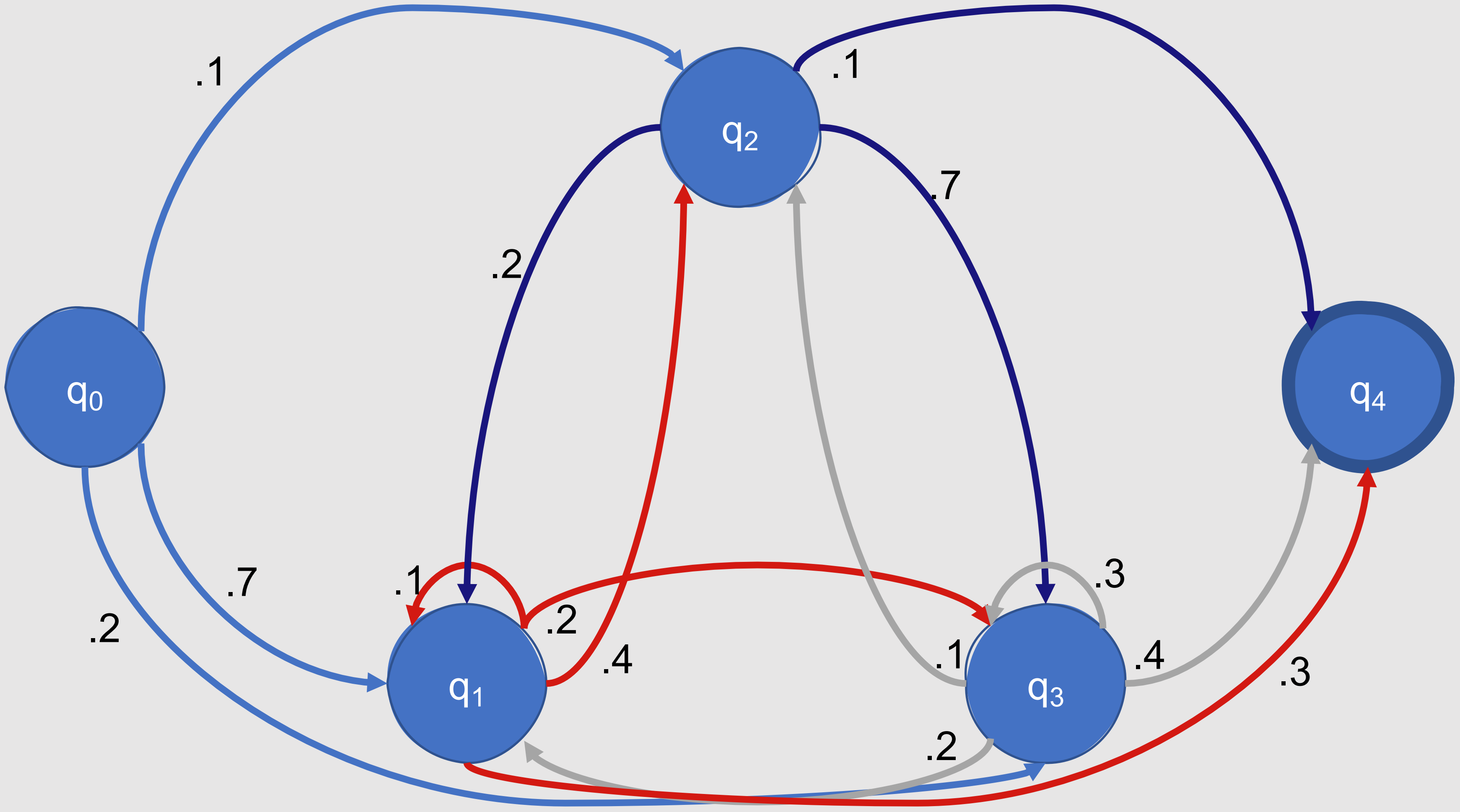
Probabilistic Sequence Models

- Allow uncertainties to be integrated over multiple, interdependent classifications
- These classifications collectively determine the most likely global assignment
- Two standard models:
  - Hidden Markov Models
  - Conditional Random Fields

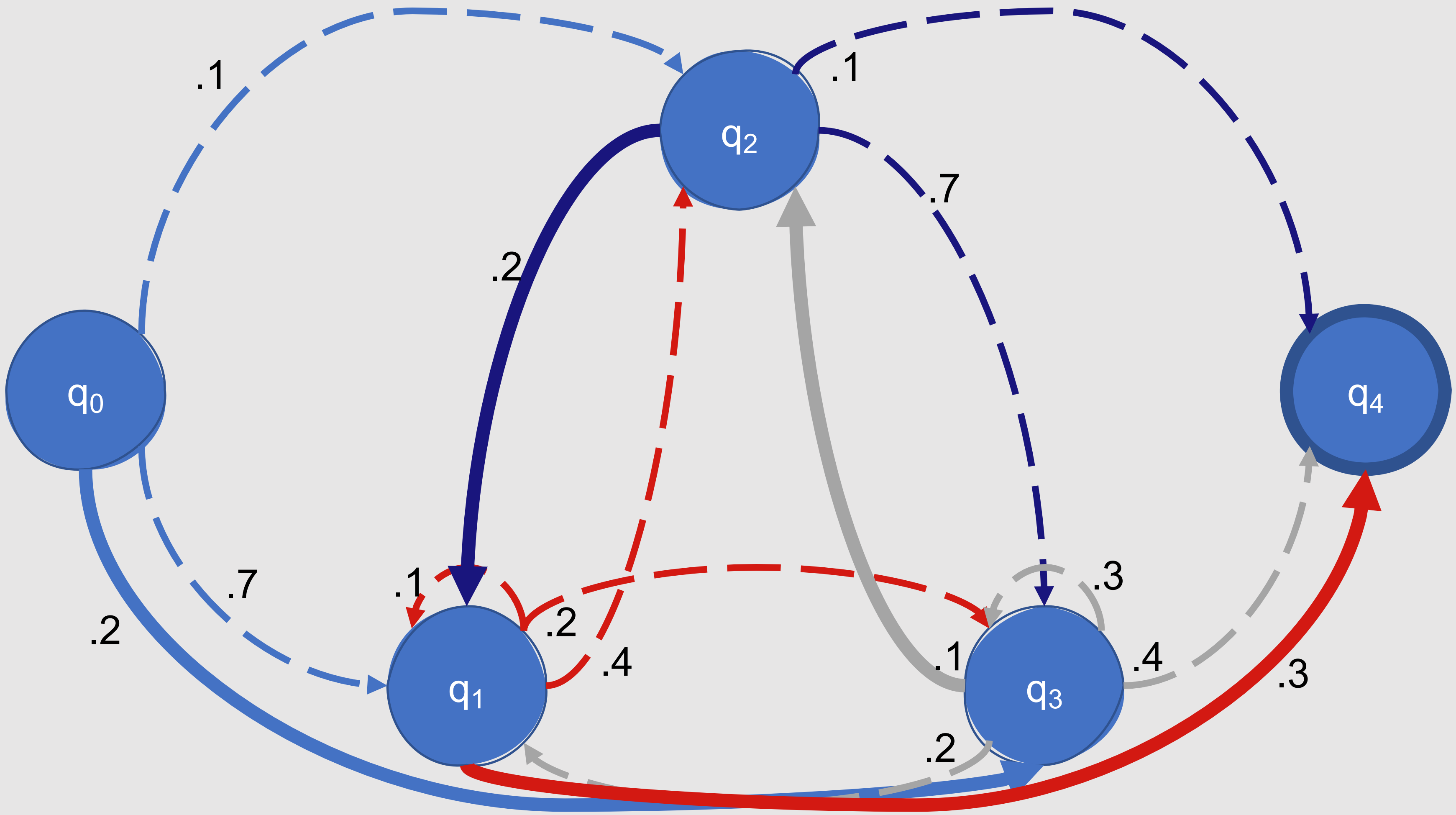
What are Markov Models?

- **Finite state automata with probabilistic state transitions**
- Markov Property: The future is independent of the past, given the present.
  - In other words, the next state only depends on the current state …it is independent of previous history.
- Also referred to as **Markov Chains**

Sample Markov Model



Sample Markov Model

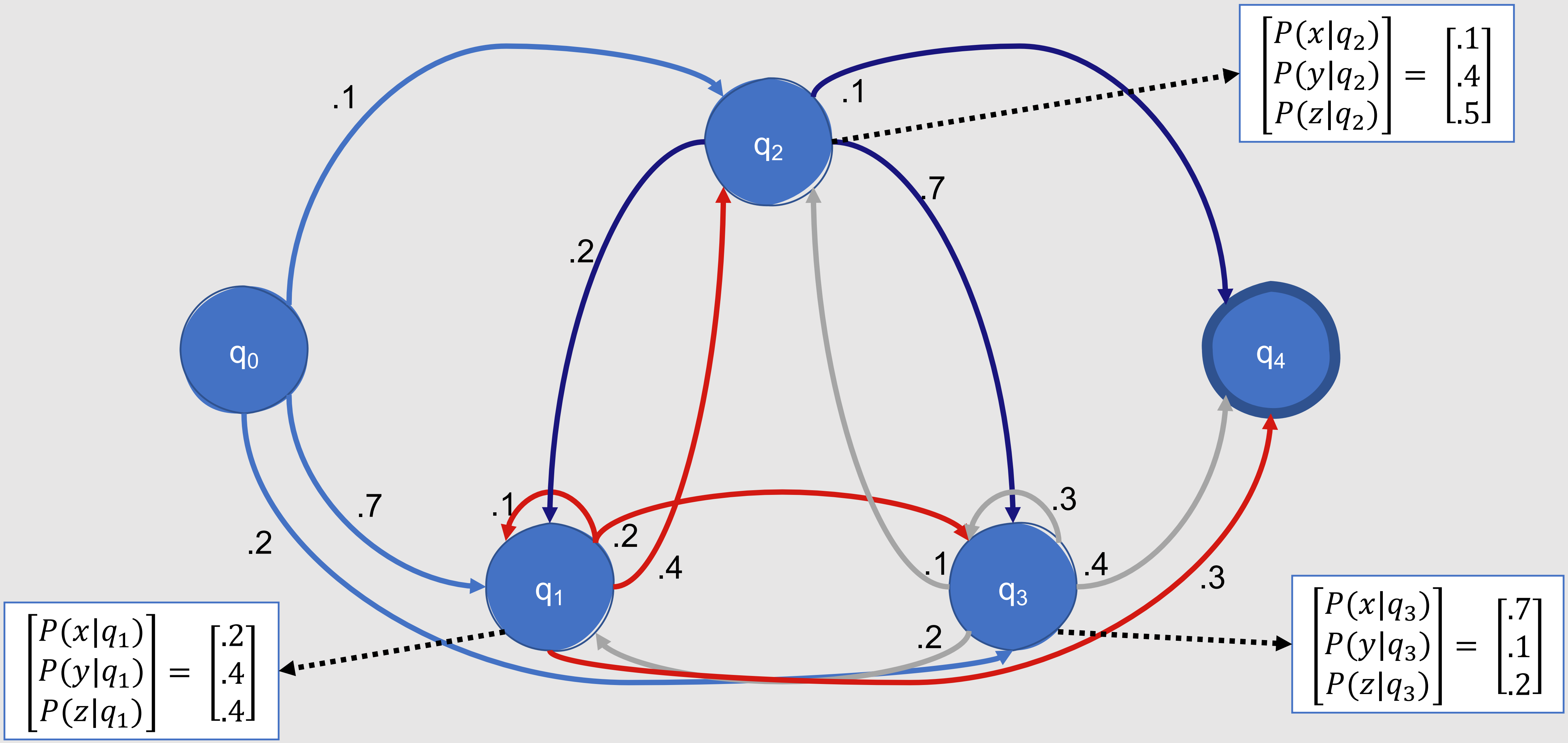


P(q_3 q_2 q_1 q_4)
= .2 * .1 * .2 * .3
= .0012

Hidden Markov Models

- Probabilistic generative models for sequences
- Assume an underlying set of hidden (unobserved) states in which the model can be
- Assume probabilistic transitions between states over time
- Assume probabilistic generation of items (e.g., tokens) from states

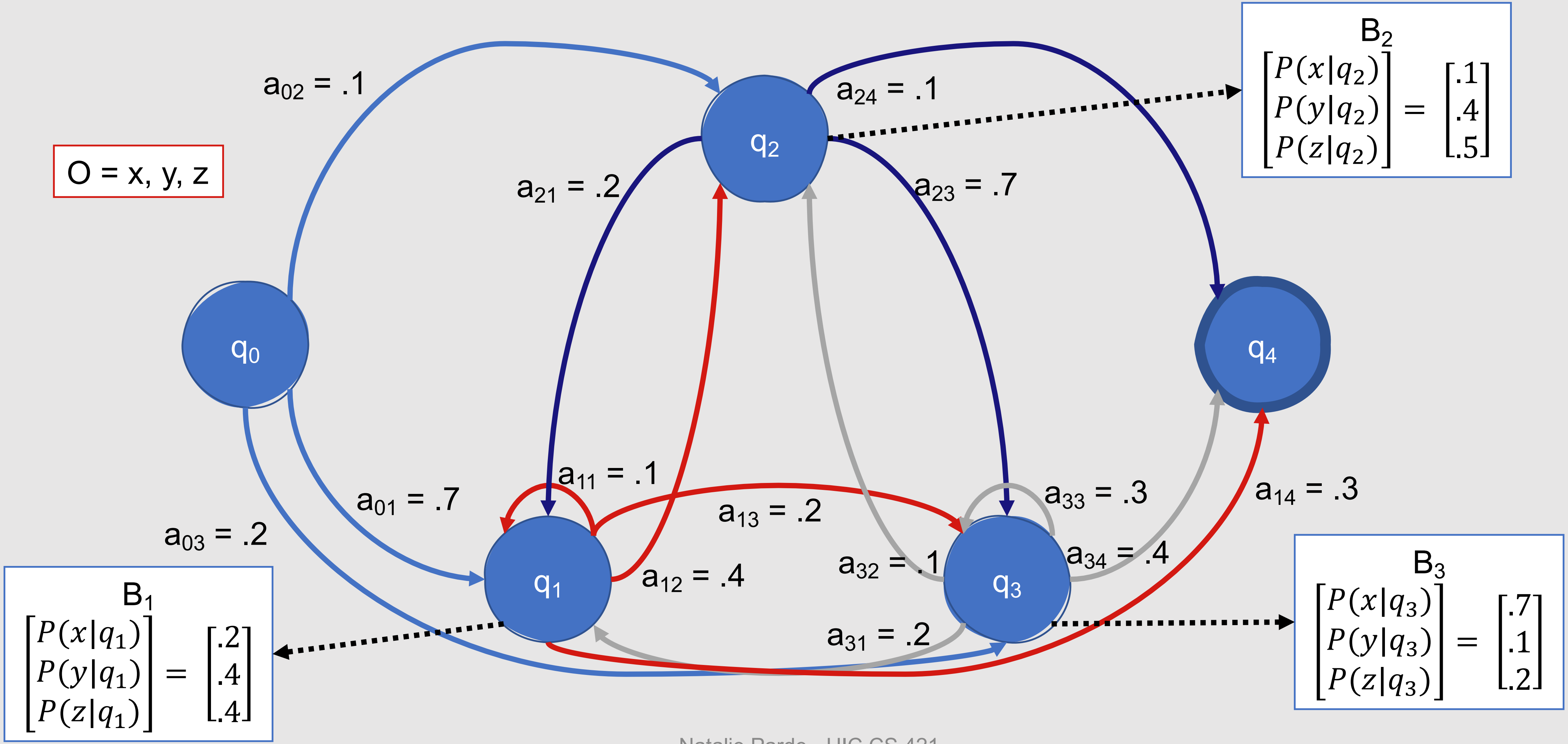
Sample Hidden Markov Model



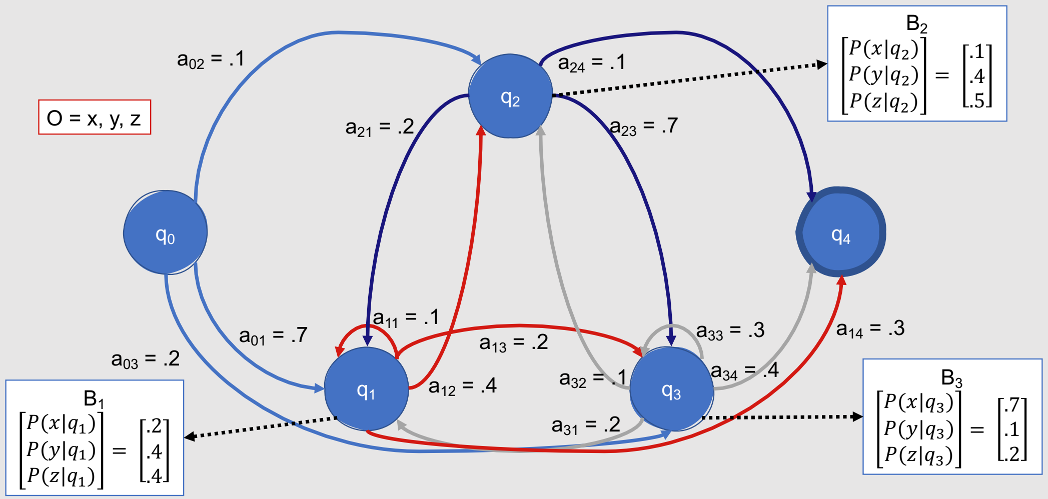
Formal Definition

- A Hidden Markov Model can be specified by enumerating the following properties:
  - The set of states, ***Q***
  - A transition probability matrix, ***A***, where each a_{ij} represents the probability of moving from state *i* to state *j*, such that $\sum_{j=1}^{n} a_{ij} = 1 \; \forall i$
  - A sequence of T observations, ***O***, each drawn from a vocabulary $V = v_1, v_2, \ldots, v_V$
  - A sequence of observation likelihoods, ***B***, also called emission probabilities, each expressing the probability of an observation o_t being generated from a state *i*
  - A start state, q_0, and final state, q_F, that are not associated with observations, together with transition probabilities out of q_0 and into q_F

Sample Hidden Markov Model



Corresponding Transition Matrix



	q0	q1	q2	q3	q4
q0	N/A	.7	.1	.2	N/A
q1	N/A	.1	.4	.2	.3
q2	N/A	.2	N/A	.7	.1
q3	N/A	.2	.1	.3	.4
q4	N/A	N/A	N/A	N/A	N/A

Practical Applications of HMMs

- One intuitive application: text generation
- More generally, you can generate a sequence of T observations: $O = o_1, o_2, \ldots, o_T$

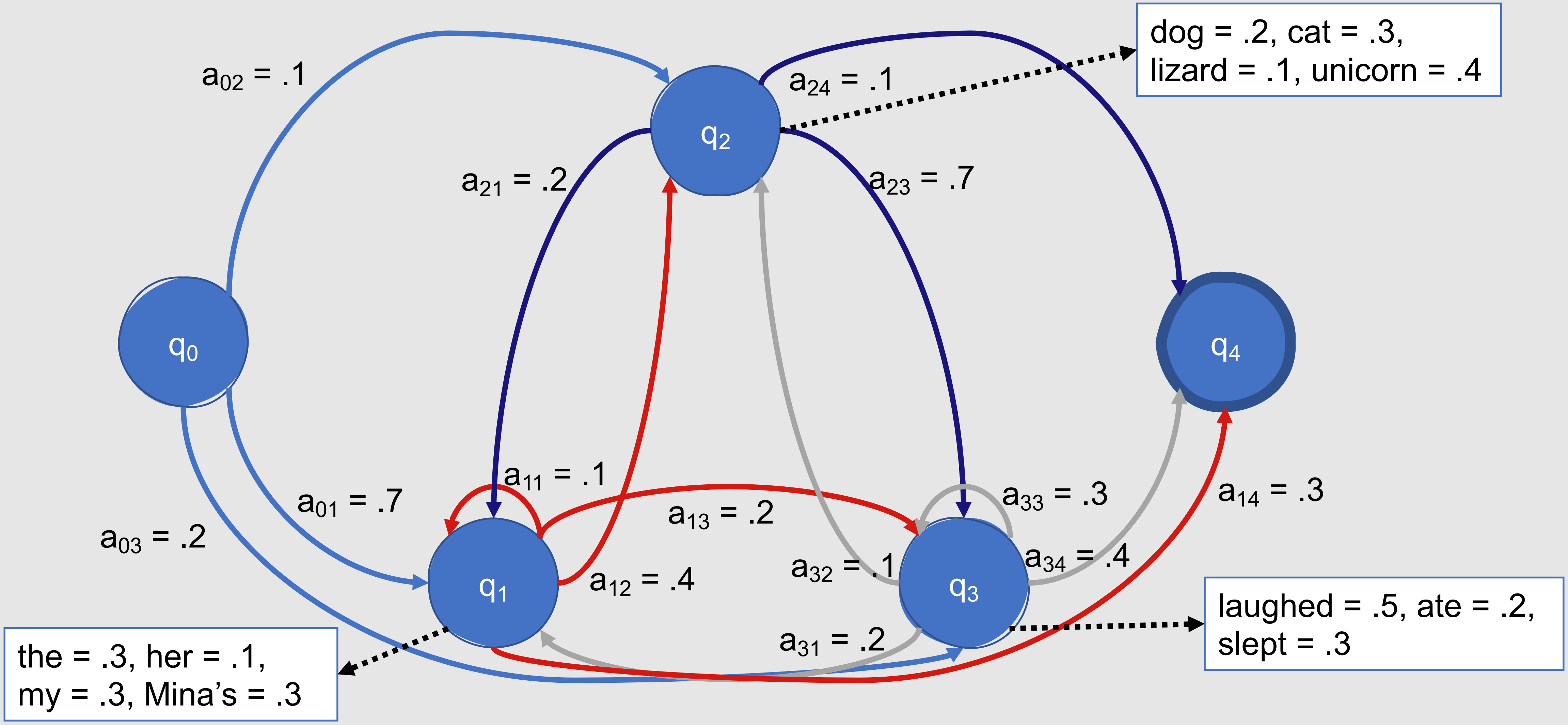
Begin in the start state

For t in [0, …, T]:

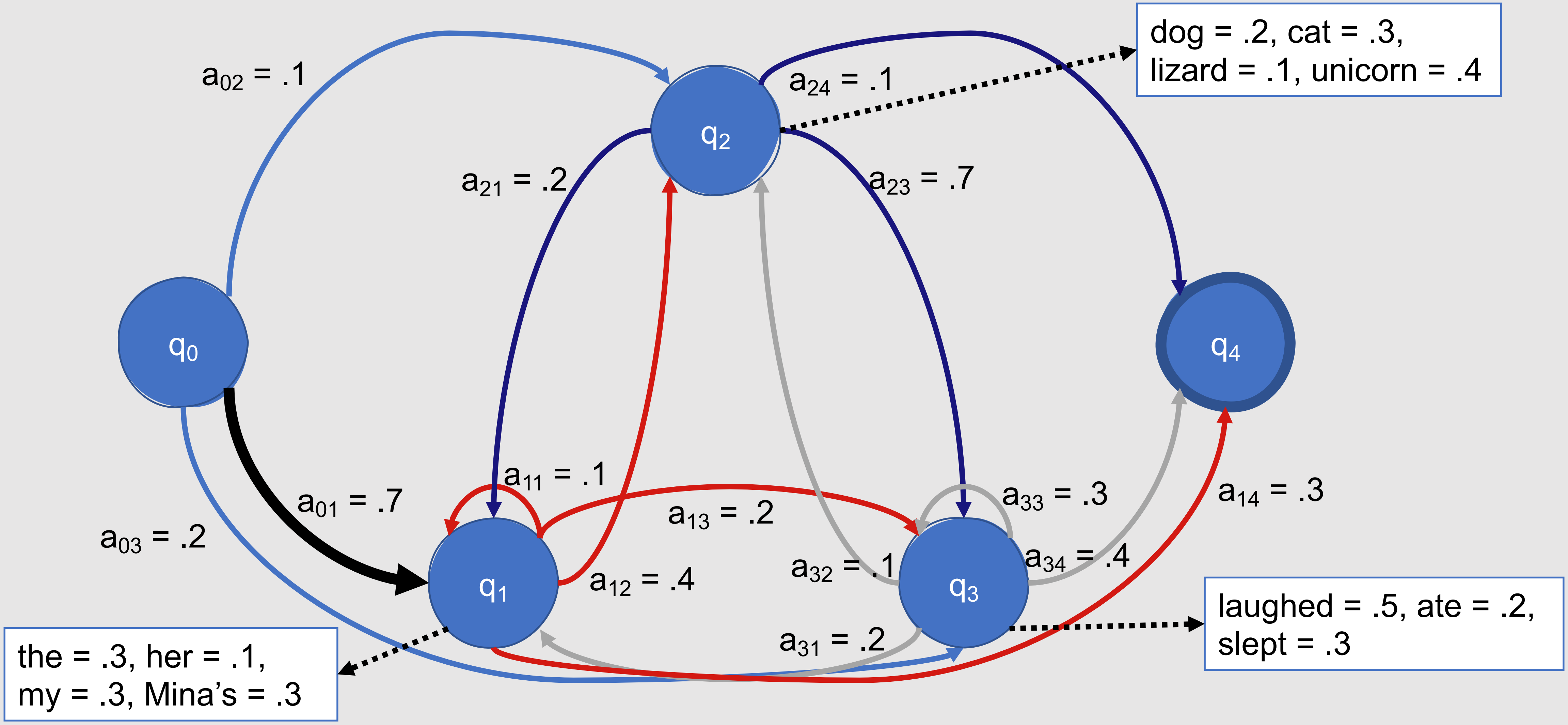
Randomly select a new state based on the transition distribution for the current state

Randomly select an observation from the new state based on the observation distribution for that state

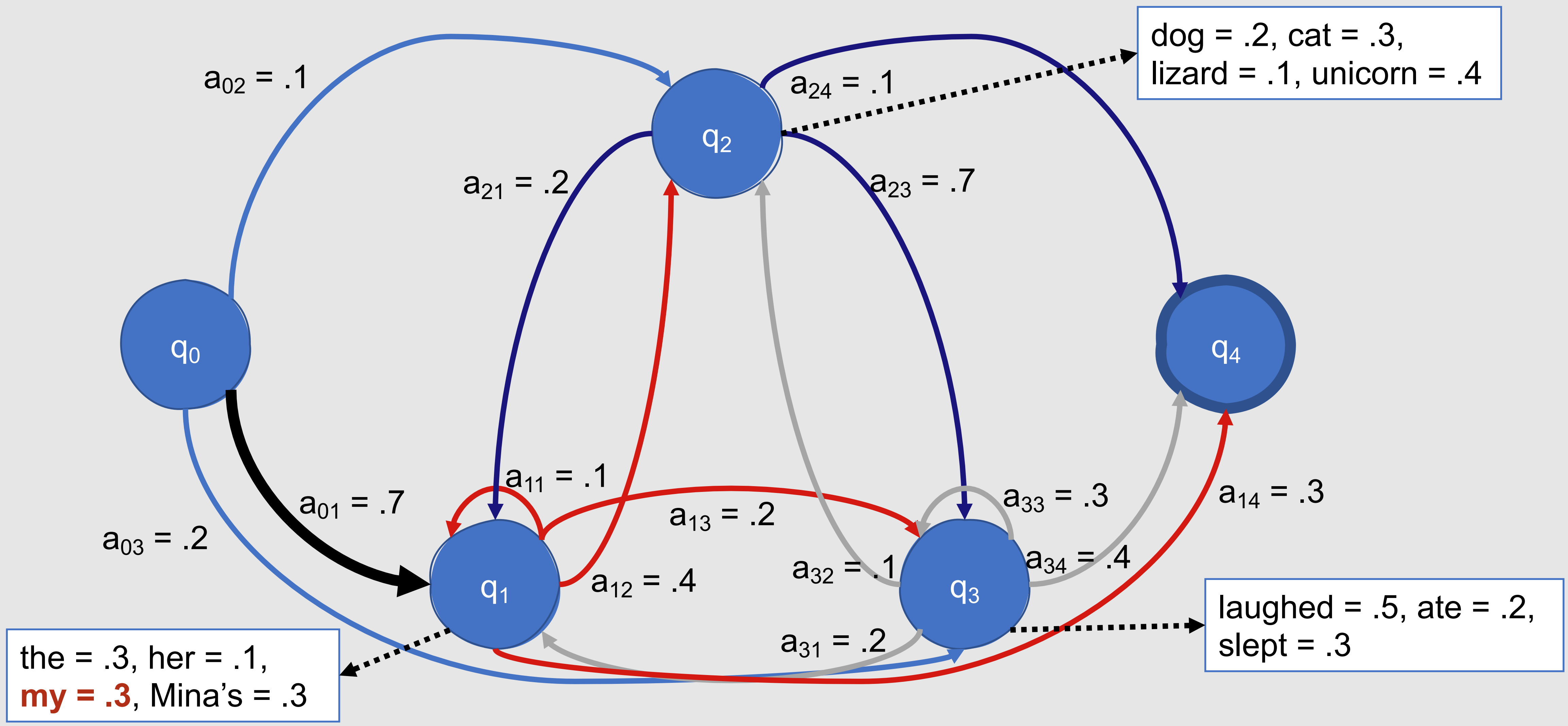
Sample Text Generation



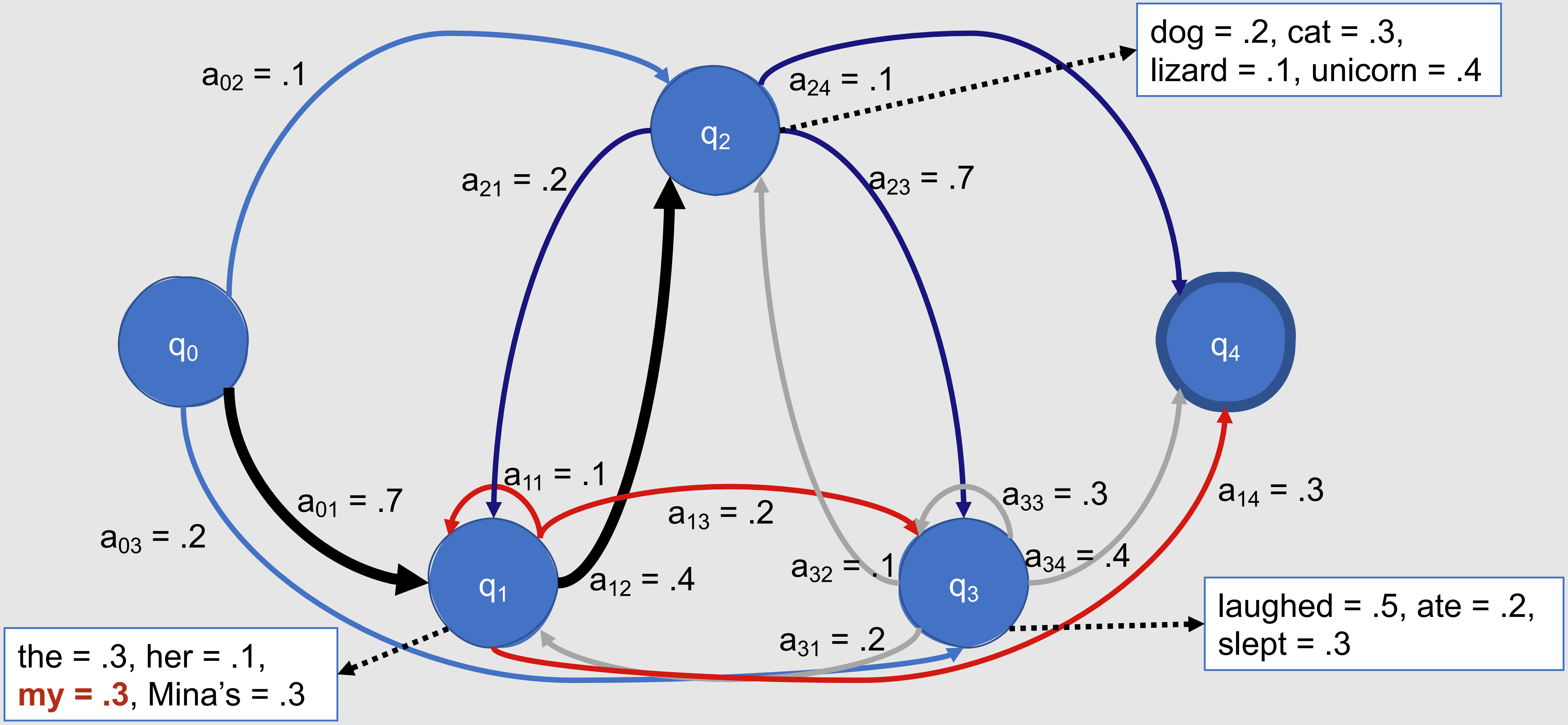
Sample Text Generation



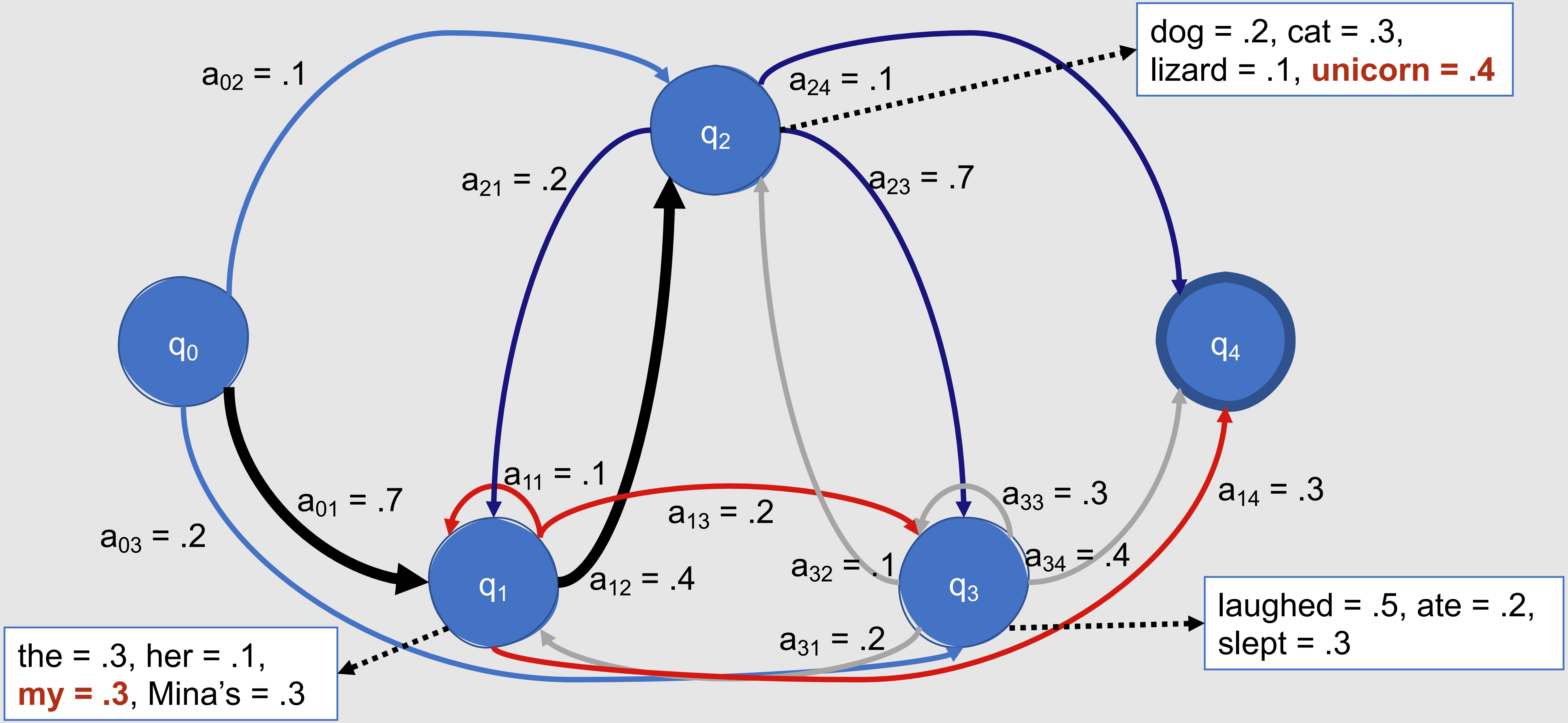
Sample Text Generation



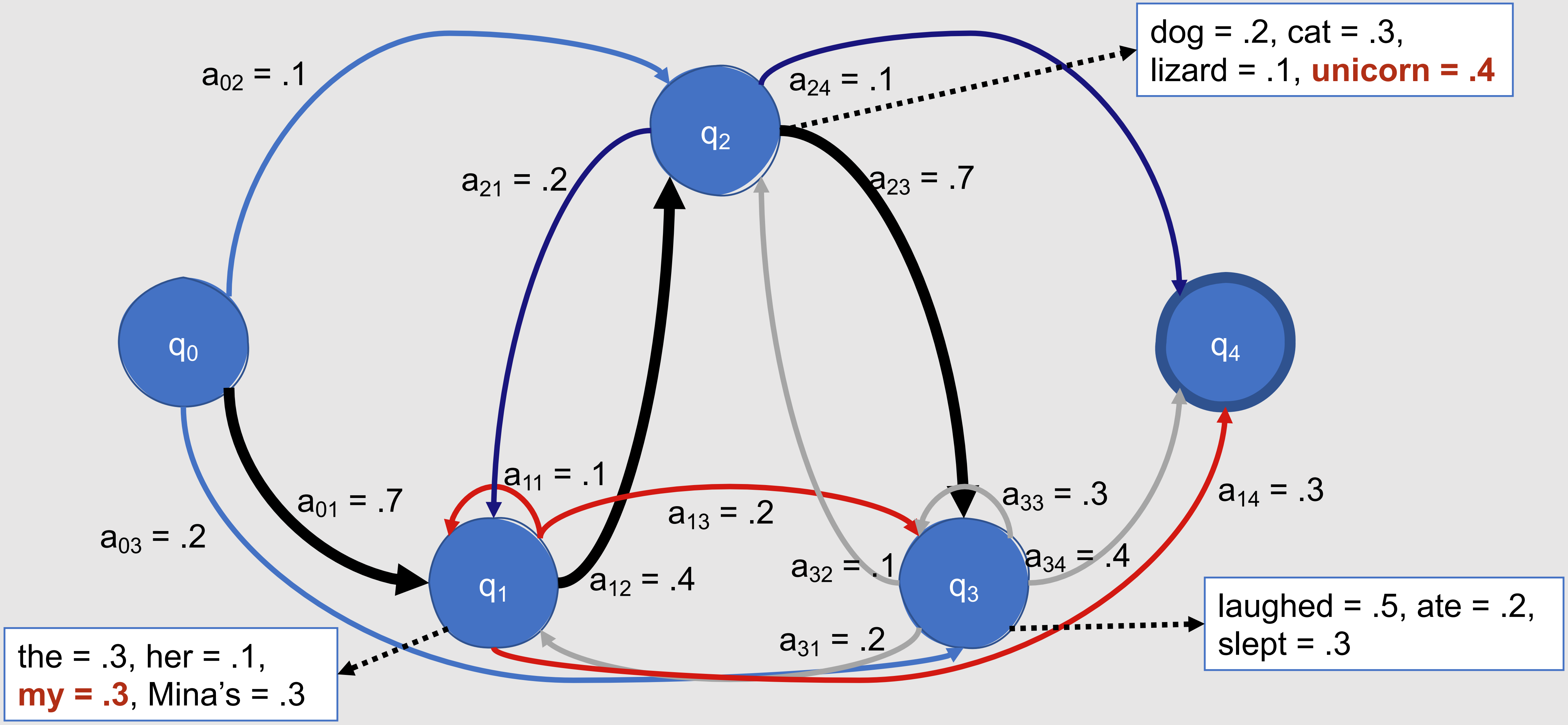
Sample Text Generation



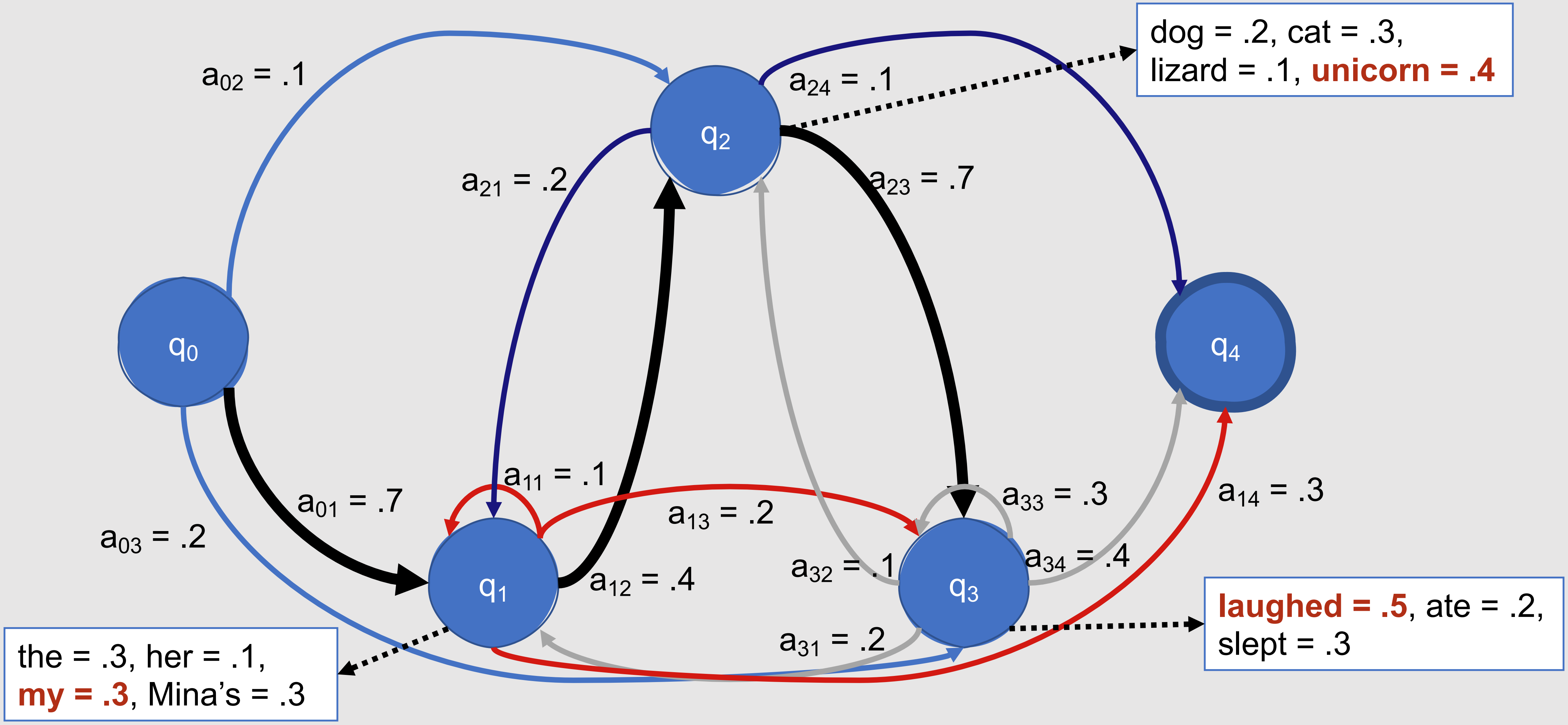
Sample Text Generation



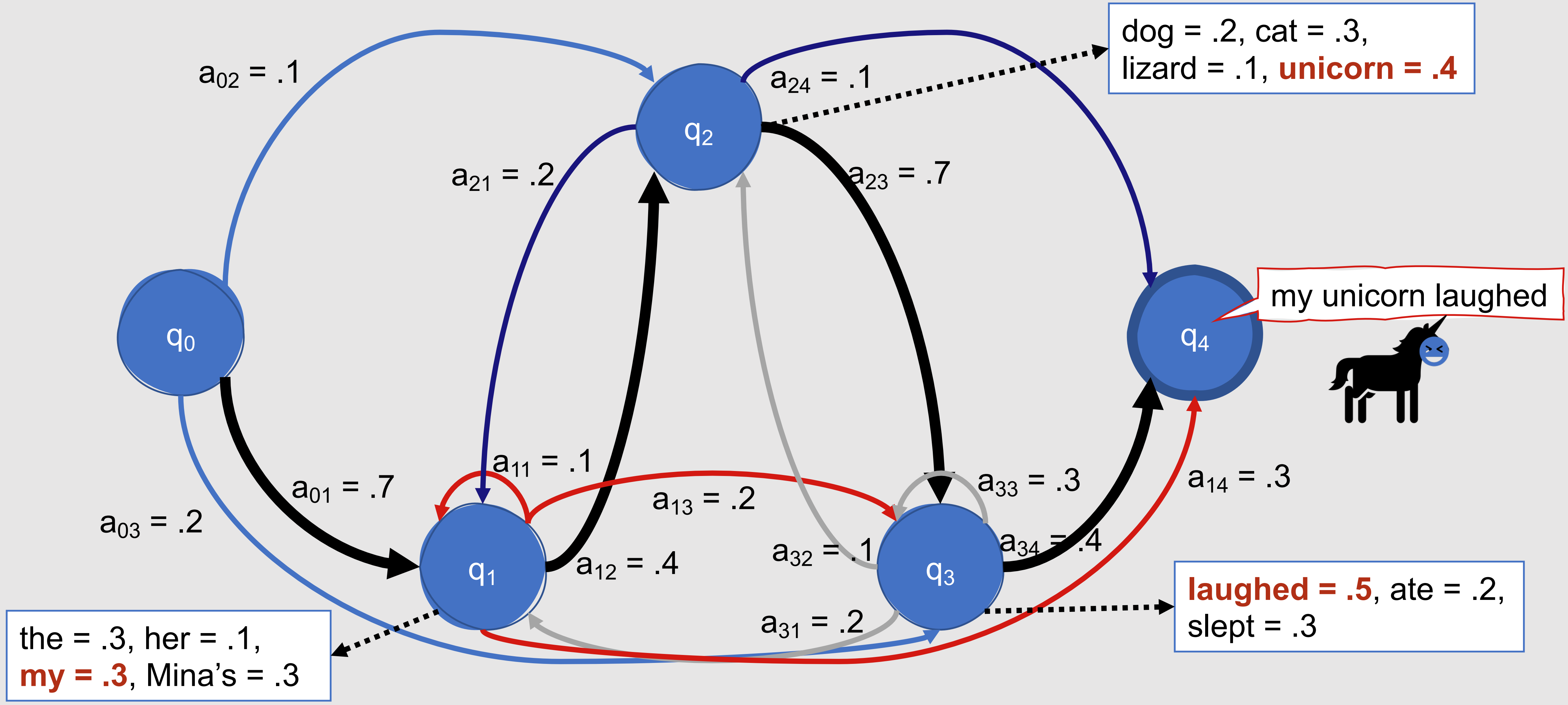
Sample Text Generation



Sample Text Generation



Sample Text Generation



Three Fundamental HMM Problems

- Observation Likelihood: How likely is a particular observation sequence to occur?
- Decoding: What is the best sequence of hidden states for an observed sequence?
  - What is the best sequence of labels for our test data?
- Learning: What are the transition probabilities and observation likelihoods that best fit the observation sequence and HMM states?
  - How do we empirically fit our training data?