FEATURE SPACES

Prof. Alexander Huth

10/12/2017

REMINDER

* Homework 1 due Tuesday! (10/17)

* Office hours Monday 9:30-11am in NHB 3.134

SYSTEM IDENTIFICATION

$$Y = f(X)$$

* What kind of a function is f?

SYSTEM IDENTIFICATION

* Linear model

$$Y = X\beta$$

* Linearized model

$$Y = \mathbb{L}(X)\beta$$

* Nonlinear model

$$Y = \Theta(X)$$

SYSTEM IDENTIFICATION

* **Linear model**
  * cheap, pointless
* **Linearized model**
  * sweet spot, but requires **hypothesis**!
* **Nonlinear model**
  * wildly expensive, difficult

LINEARIZING TRANSFORMATION

=

FEATURE SPACE

=

HYPOTHESIS

LINEARIZED MODELS

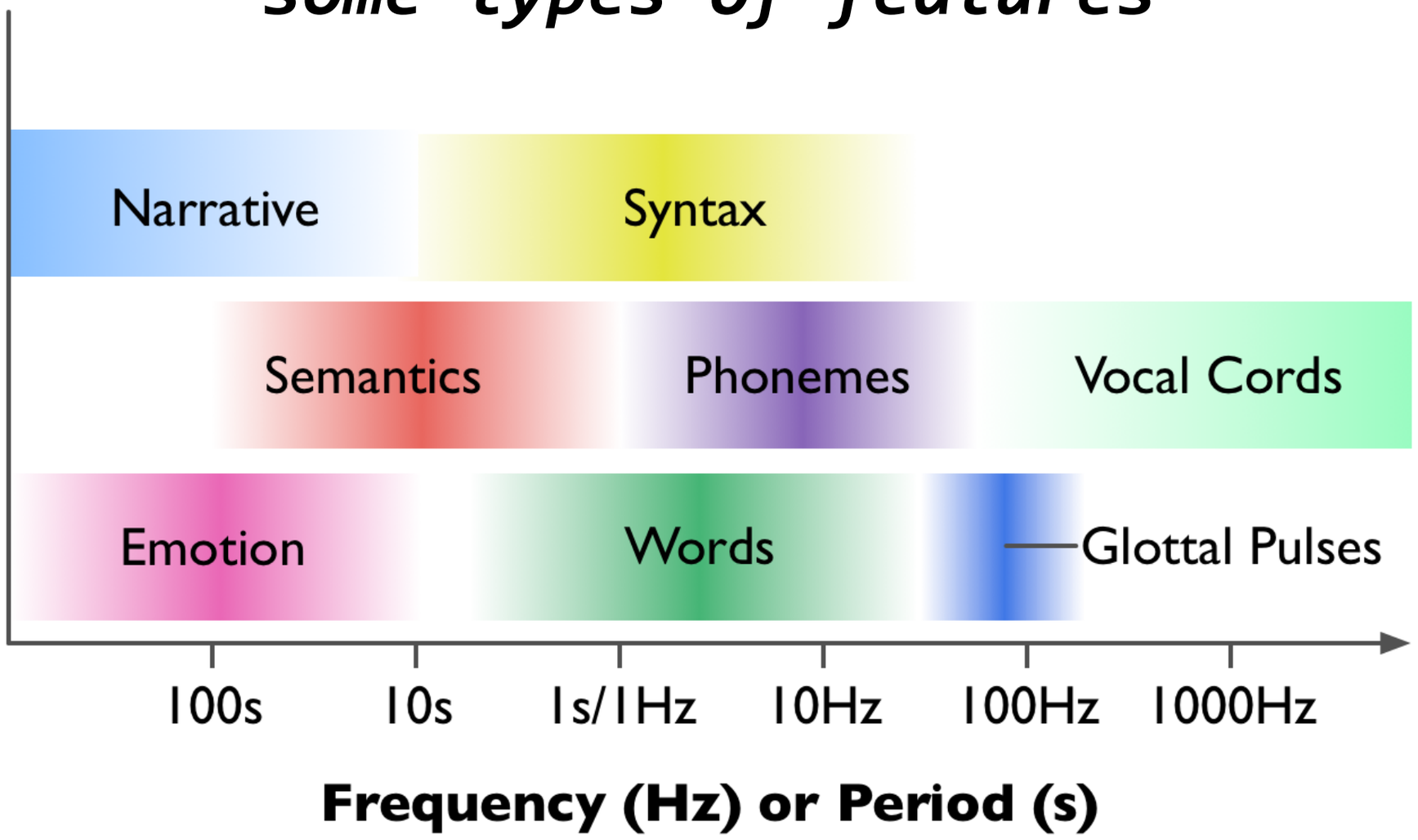
LANGUAGE

“Now this is a story
all about how my
life got flipped-
turned upside down…”

what kinds of features?

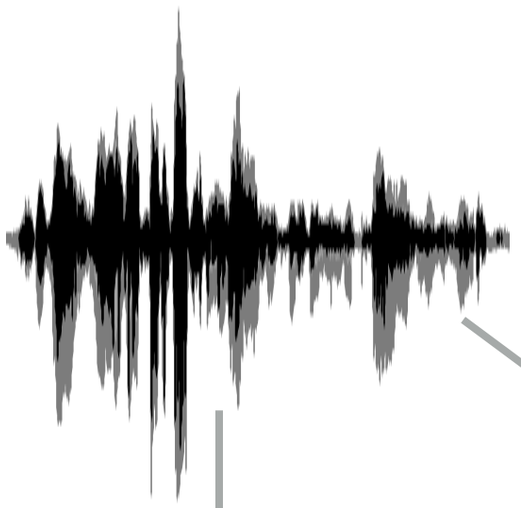
LANGUAGE

some types of features



LANGUAGE

"Now this is a story all about how my life got flipped-turned upside down…"



words

syntax

semantics

narrative

emotion

waveform

intonation

spectrum

phonemes

articulation

HOW DO WE GET FEATURES?

* **Linguistics**
  * Parts of speech, thematic roles, etc.
* **NLP / machine learning**
  * Models of syntax, semantics, etc.

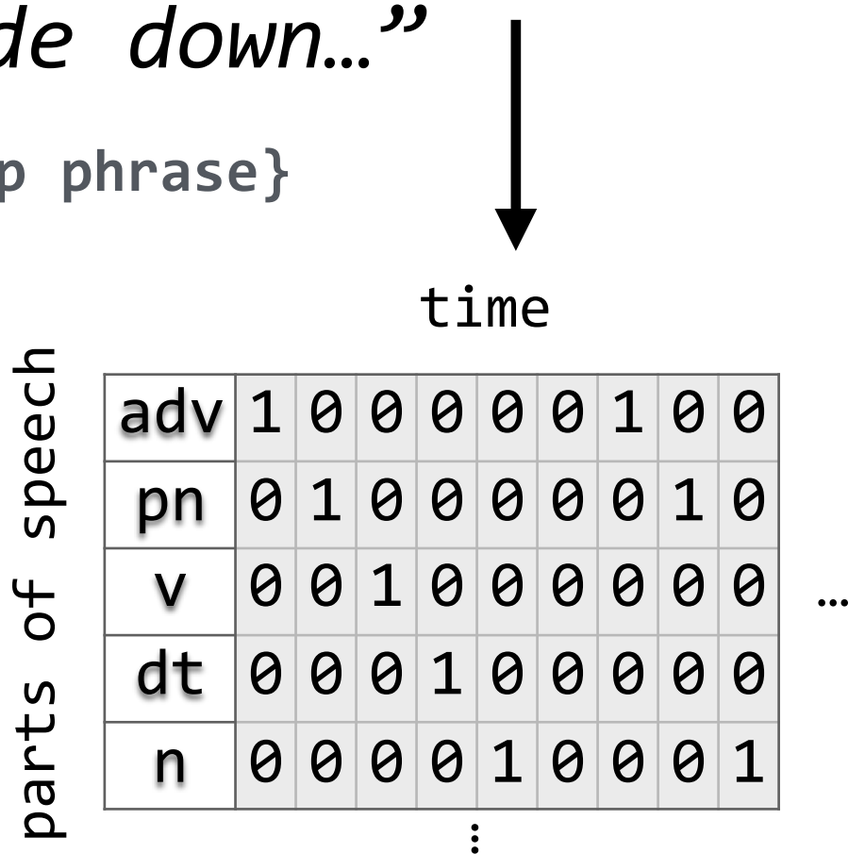
SYNTAX - PART OF SPEECH

“Now this is a story all about how my life

{adv} {pn} {v} {dt} {n} {adj} {prep} {adv} {pn} {n}

got flipped-turned upside down…”

{v-p} {v-p} {v-p} {prep phrase}

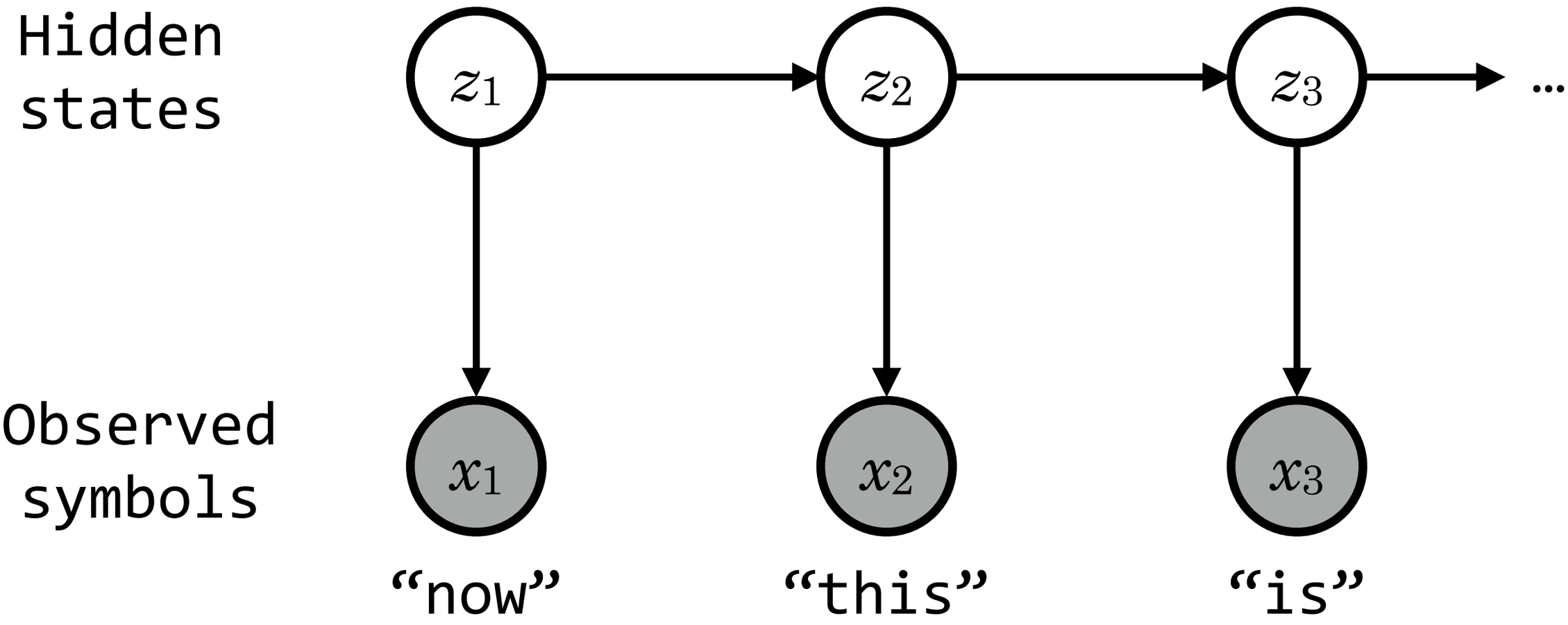


SYNTAX - HMM

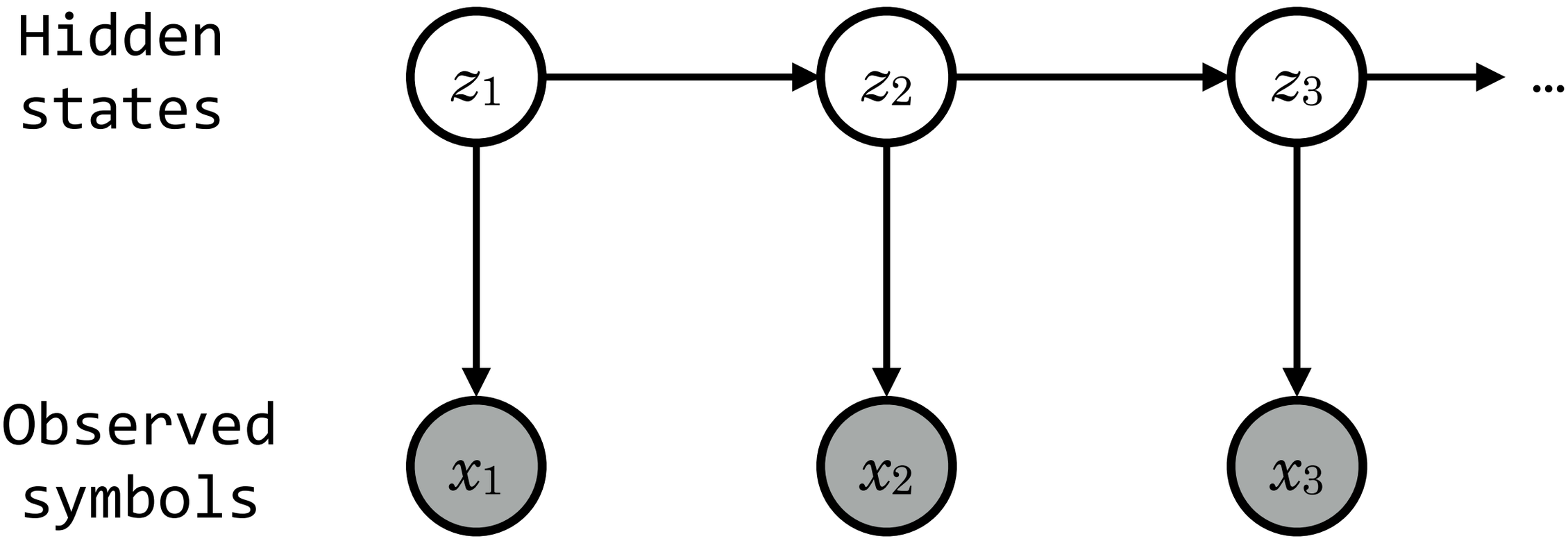
* *We get:* a sequence of observed symbols
  [“now”, “this”, “is”, “a”, “story”, …]

* *We think:* there are hidden, underlying states
  [a, b, c, d, e, …]

SYNTAX - HMM

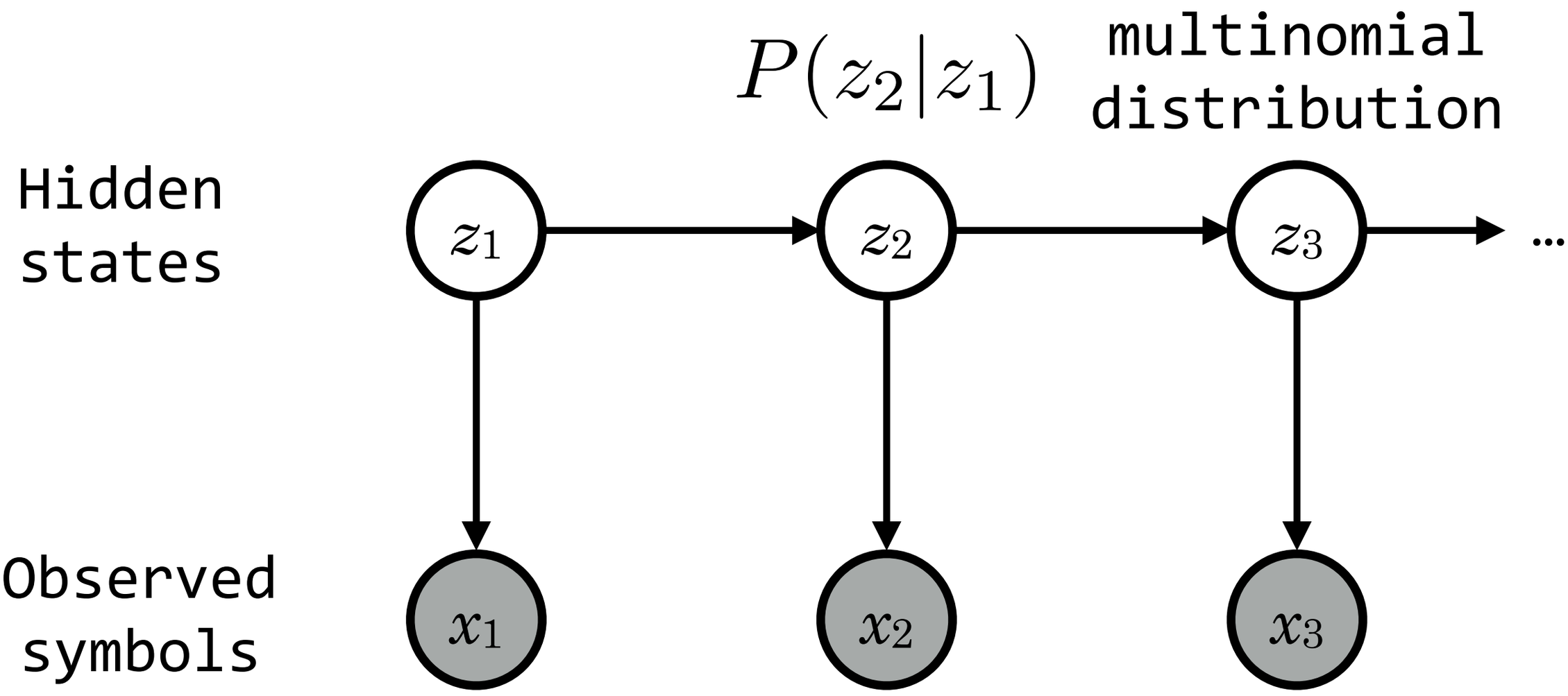


SYNTAX - HMM



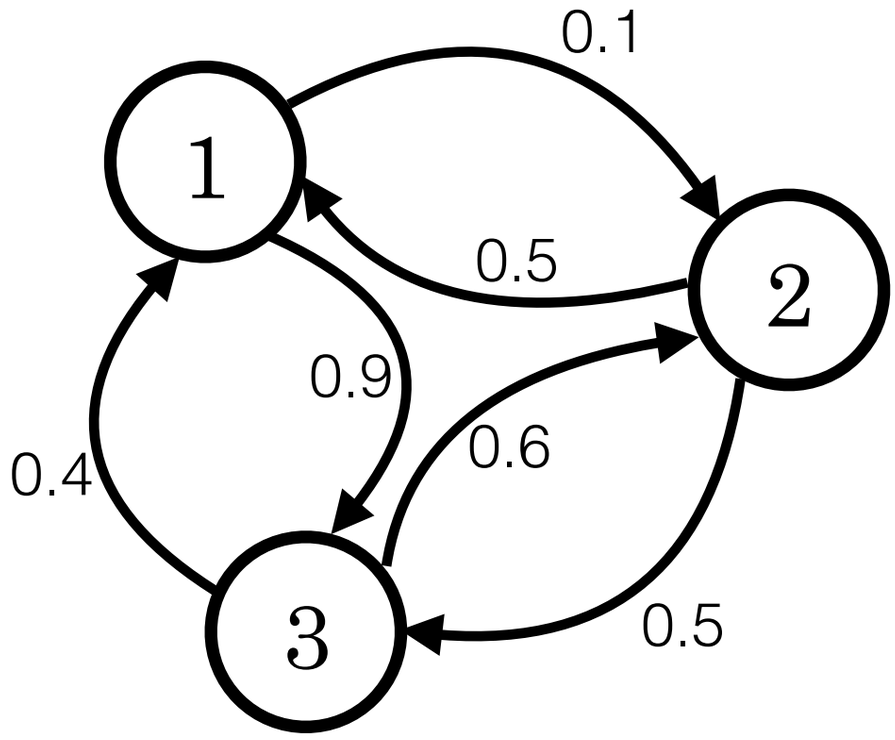
$P(x_1|z_1)$ multinomial distribution

SYNTAX - HMM


$P(z_2|z_1)$
multinomial distribution
Hidden states
z_1
z_2
z_3
…
Observed symbols
x_1
x_2
x_3
$P(x_1|z_1)$
multinomial distribution

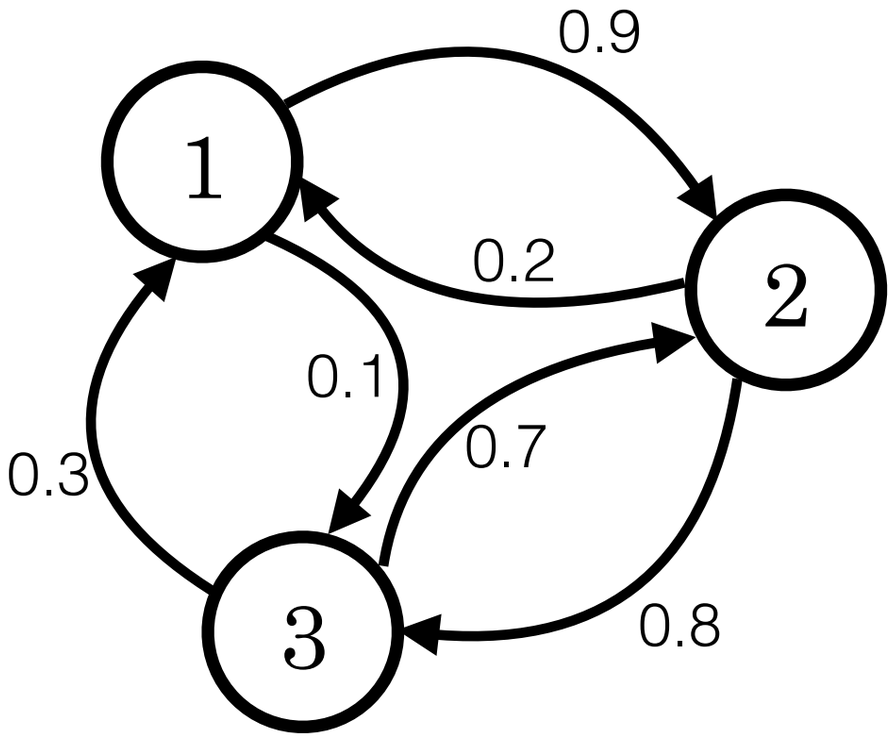
SYNTAX - HMM

Hidden state transitions



SYNTAX - HMM

Hidden state transitions θ



Emission ϕ

	z=1	z=2	z=3
x=now	0.7	0.0	0.1
x=this	0.0	0.6	0.0
x=is	0.3	0.0	0.9
x=a	0.0	0.4	0.0

SYNTAX - HMM

* **Learning time:** we know x, what are theta and phi?

* Put differently: find theta and phi that maximize probability of observed x

* The easy way: Markov chain Monte Carlo (MCMC)

  * (Different kind of Markov chain)

SYNTAX - HMM

* **Inference time:** we know x, we know theta, we know phi; what is $P(z|x)$?

* Finally, use inferred state probabilities as features in a linearized model!

NEXT TIME

* Semantics!