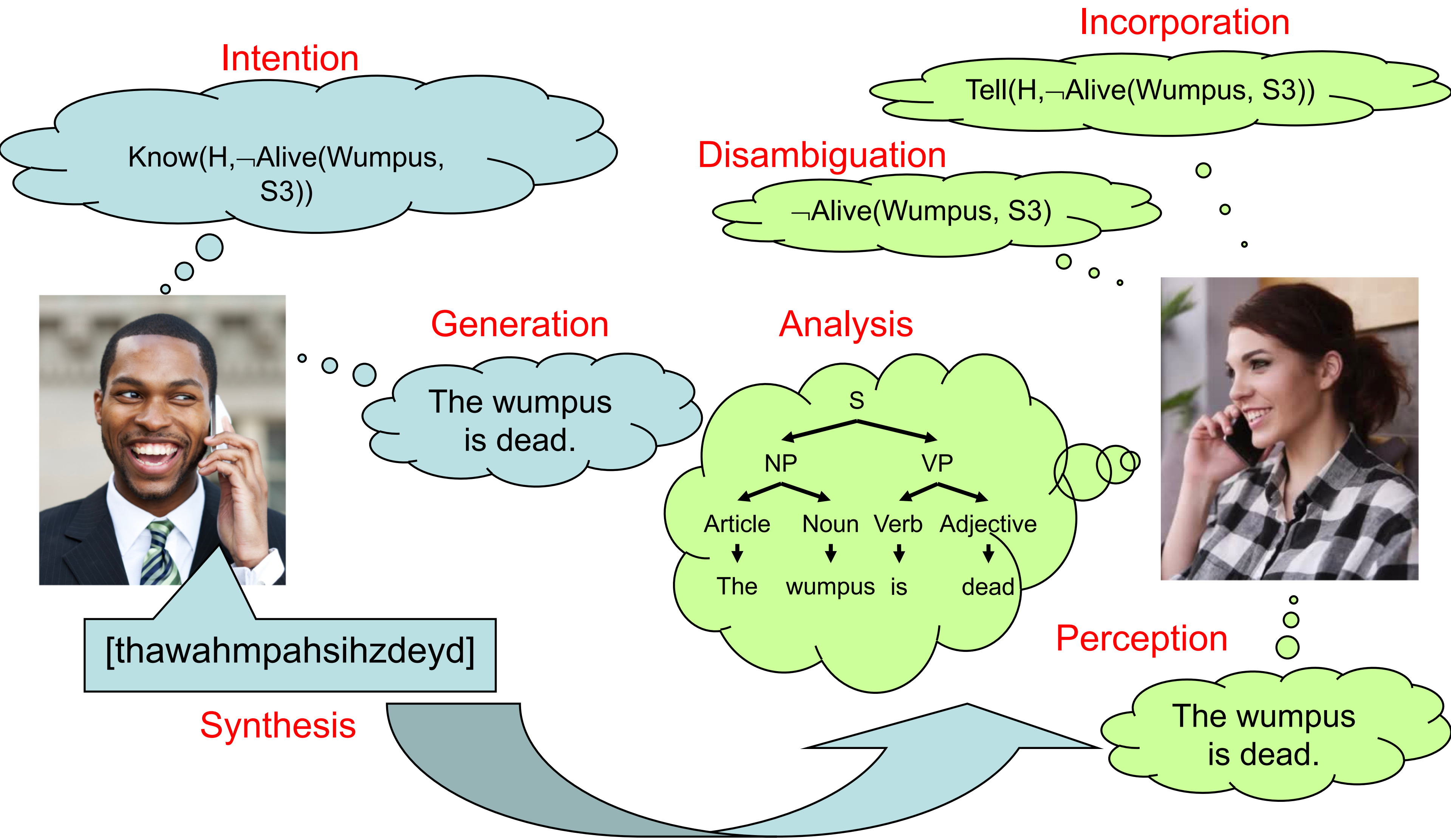
Communication II

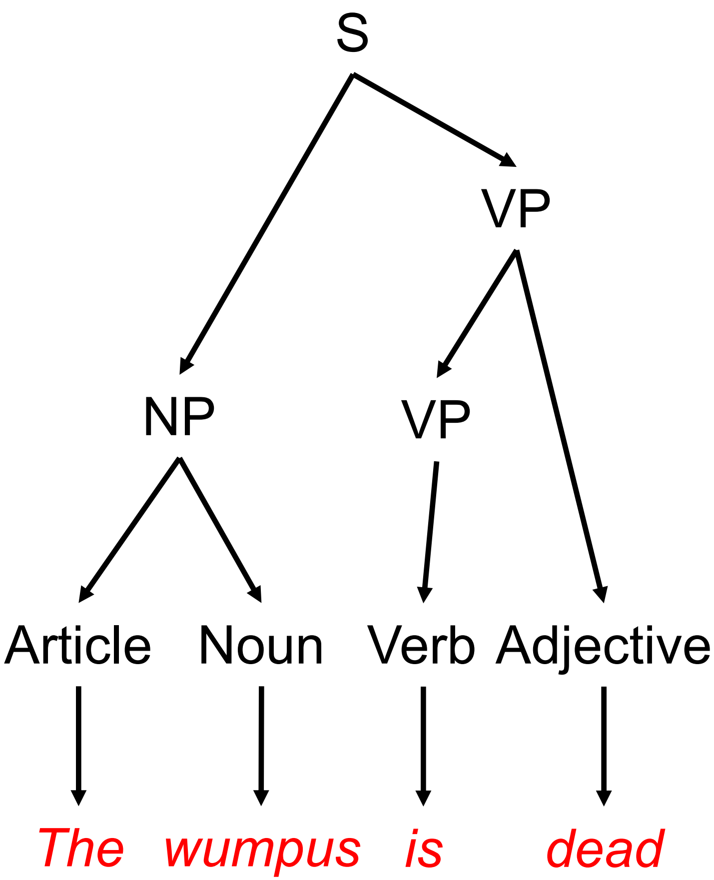
CPSC 473 – Artificial Intelligence

Brian Scassellati

Component Steps of Communication

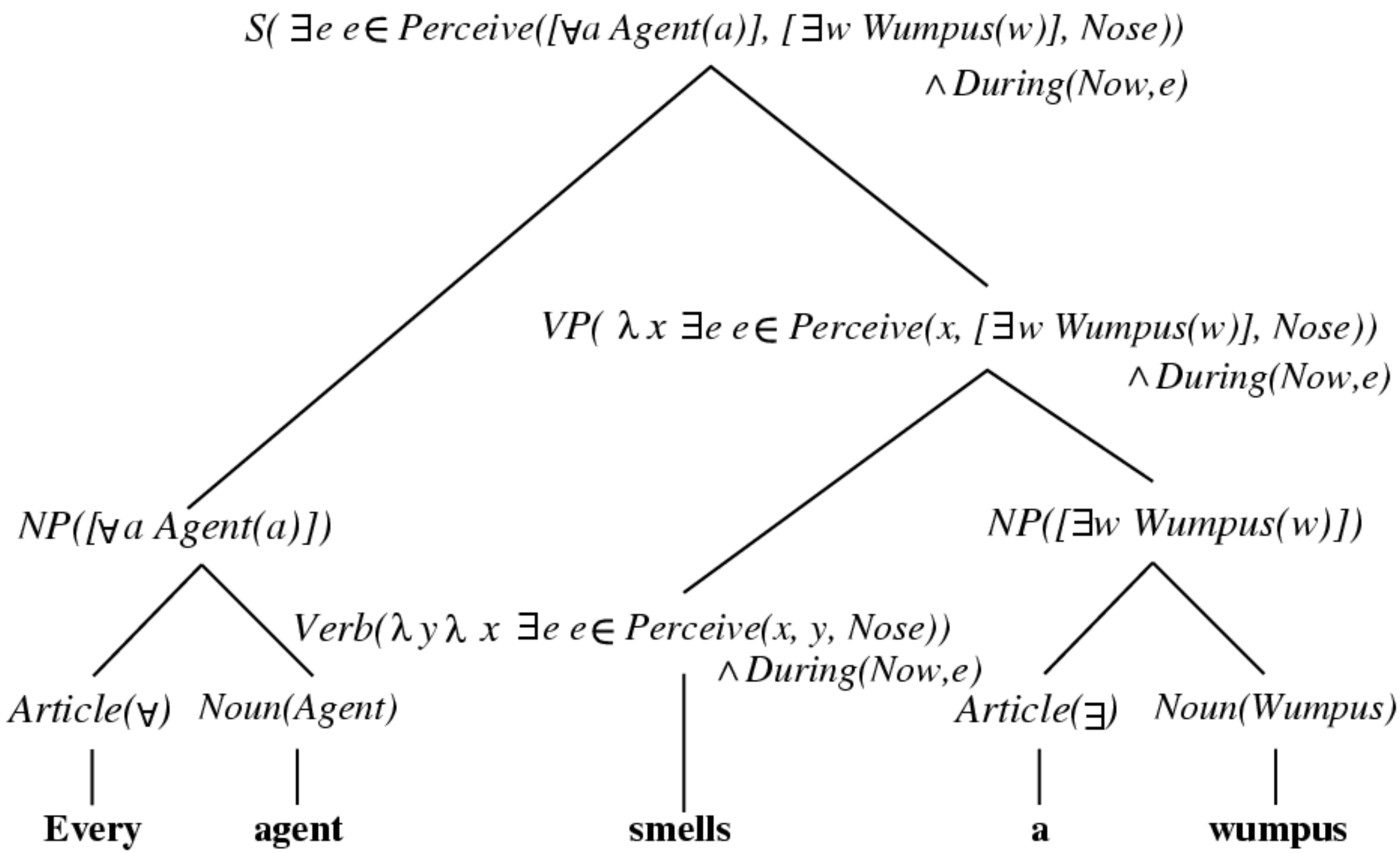


Bottom-Up Parsing Example



Forest	**Rule being applied**
The wumpus is dead	Article → *the*
Article *wumpus is dead*	Noun → *wumpus*
Article Noun *is dead*	NP → Article Noun
NP *is dead*	Verb → *is*
NP Verb *dead*	Adjective → *dead*
NP Verb Adjective	VP → Verb
NP VP Adjective	VP → VP Adjective
NP VP	S → NP VP
S	

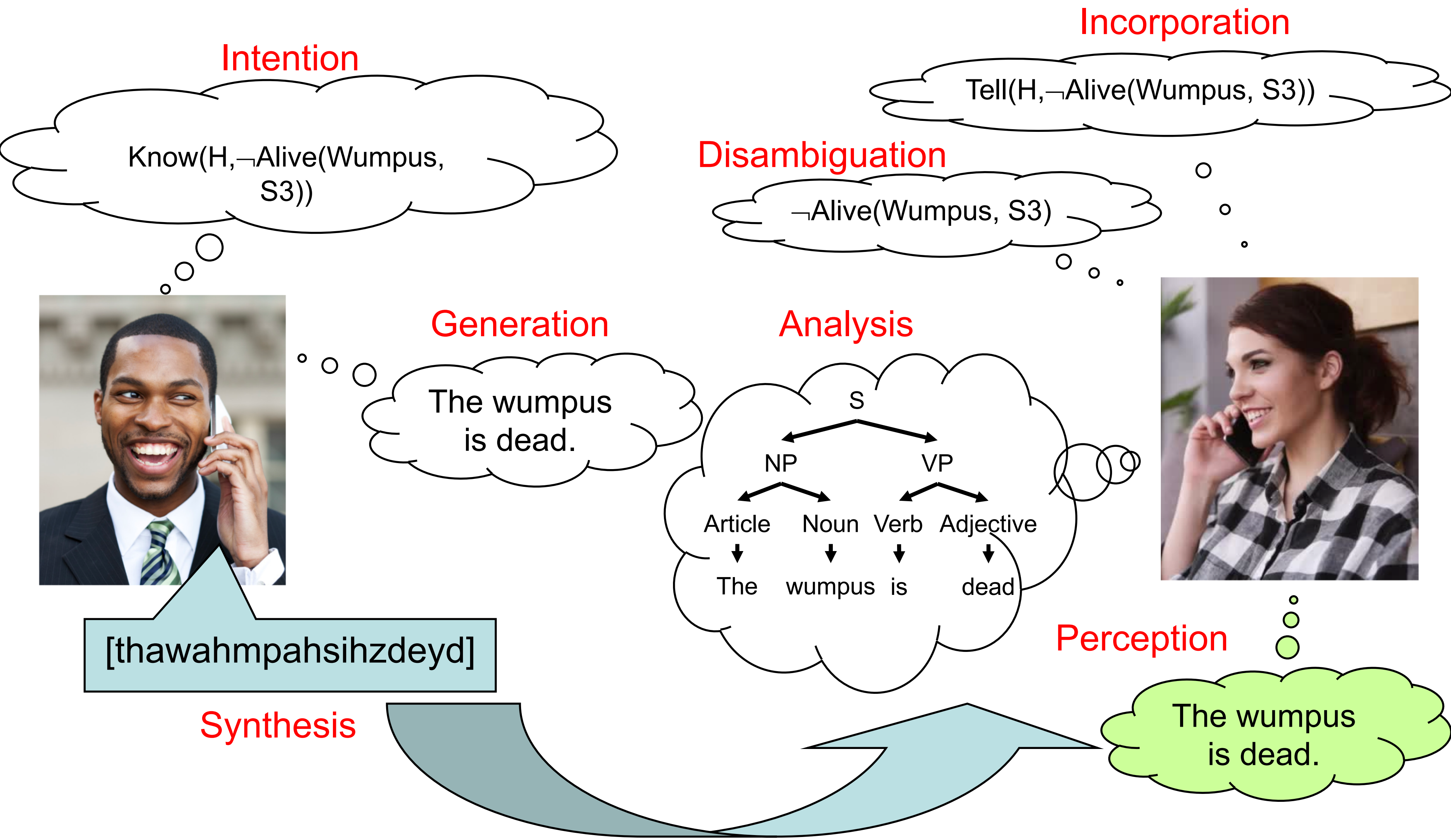
Parsing with Syntax and Semantics



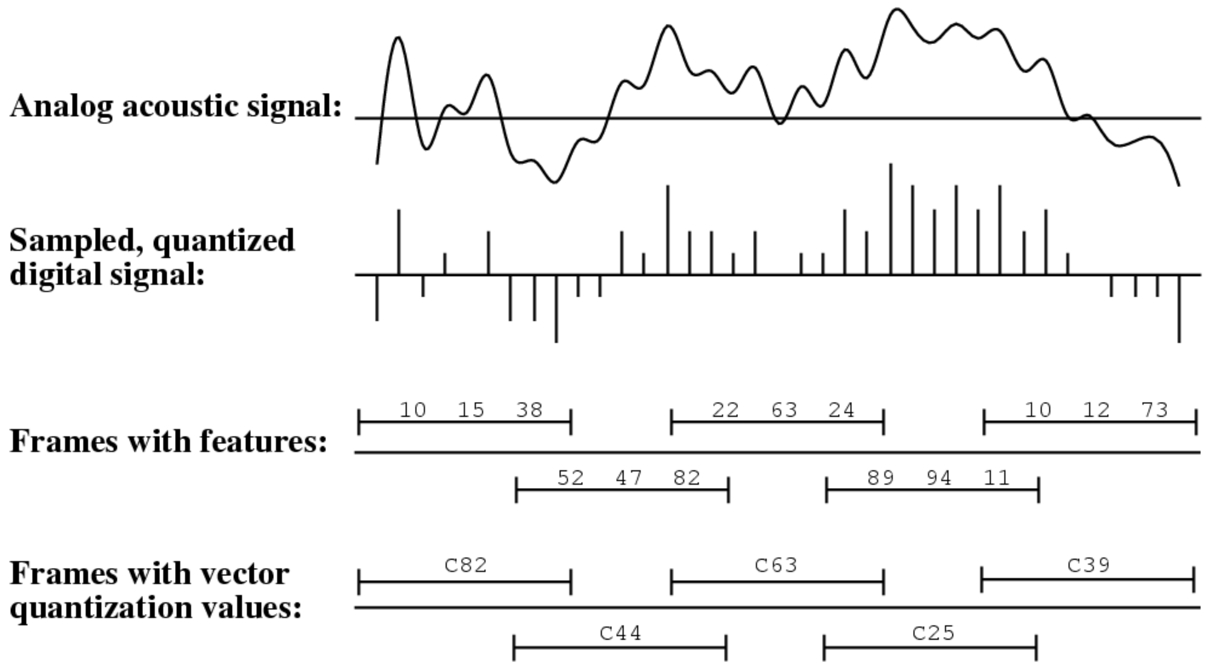
Ambiguity

- Ambiguous newspaper headlines
  - Squad helps dog bite victim.
  - Red-hot star to wed astronomer.
  - Helicopter powered by human flies.
  - American pushes bottle up Germans.
- Many places that ambiguity can arise
  - Lexical ambiguity (*star* has more than one meaning)
  - Syntactic ambiguity (is *dog* an adjective or a noun)
  - Semantic ambiguity (A *coast road* can either lead to the coast or run along the coast)
  - Pragmatic ambiguity (*I'll meet you next Friday…* is *Friday* two days or nine days away?)

Component Steps of Communication



From Analog Audio to Digital Features



- Analog signal is too noisy, contains too much data, and is not in a representation that is easy to manipulate
- Digitize and reduce the dimensionality using quantization

The Speech Recognition Problem

- Recover the words that produce a given acoustic signal
- Given a signal, identify the sequence of words that maximizes P(words | signal)

$$P(words \mid signal) = \frac{P(words)P(signal \mid words)}{P(signal)}$$

  - P(words) is the language model
  - P(signal | words) is the acoustic model
  - P(signal) is a normalizing constant

The Language Model: P(words)

- How to get the probability of a sequence of words?

$$
\begin{aligned}
P(w_1 \ldots w_n) &= P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_1 w_2)\ldots P(w_n \mid w_1 \ldots w_{n-1}) \\
&= \prod_{i=1}^{n} P(w_i \mid w_1 \ldots w_{i-1})
\end{aligned}
$$

- But this gets really, really complicated

  P(the rat ate cheese) = P(the) * P(rat|the) * P(ate|the rat) * P(cheese|the rat ate)

- Approximate with a bigram model that depends only on pairs of words

$$
\begin{aligned}
P(w_1 \ldots w_n) &= P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_2)\ldots P(w_n \mid w_{n-1}) \\
&= \prod_{i=1}^{n} P(w_i \mid w_{i-1})
\end{aligned}
$$

- Easier to compute values

  P(the rat ate cheese) = P(the) * P(rat|the) * P(ate|rat) * P(cheese|ate)

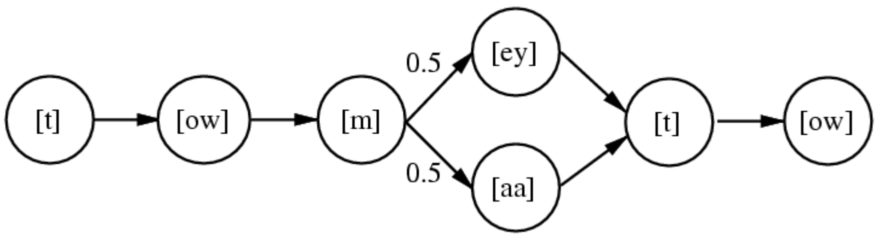
Building a bigram model

Word	Unigram count	Previous words									
		OF	IN	IS	ON	TO	FROM	THAT	WITH	LINE	VISION
THE	367	179	143	44	44	65	35	30	17	0	0
ON	69	0	0	1	0	0	0	0	0	0	0
OF	281	0	0	2	0	1	0	3	0	4	0
TO	212	0	0	19	0	0	0	0	0	0	1
IS	175	0	0	0	0	0	0	13	0	1	3
A	153	36	36	33	23	21	14	3	15	0	0
THAT	124	0	3	18	0	1	0	0	0	0	0
WE	105	0	0	0	1	0	0	12	0	0	0
LINE	17	1	0	0	0	1	0	0	0	0	0
VISION	13	3	0	0	1	0	1	0	0	0	0

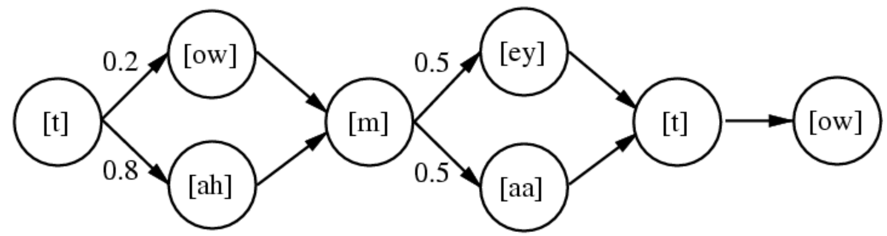
- Construct a bigram table just by counting word frequencies
  - This table is taken from chapter 24 in the text
- Can also use higher-order models (trigram, etc.)
  - Distinguish “ate a banana” from “ate a bandana”

The Acoustic Model: P(signal | words)

Word model with dialect variation:



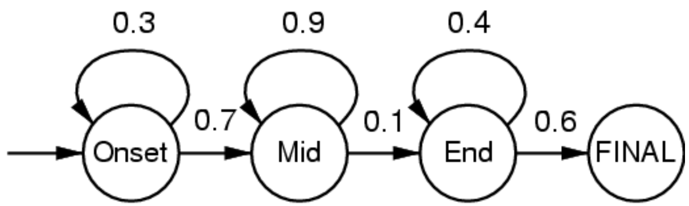
Word model with coarticulation and dialect variations:



- Markov models for generating a word from phones
  - States give a unique output symbol
    - Total output is a sequence of output symbols (or state names)
  - Links have a probability associated with them
    - Unlabelled links have a probability of 1
    - Markov property: history does not matter
  - Probability of a pronunciation is the product of the probabilities along the paths

The Acoustic Model: P(signal | words) Hidden Markov Models

Phone HMM for [m]:



Output probabilities for the phone HMM:

Onset:	Mid:	End:
C1: 0.5	C3: 0.2	**C1: 0.1**
C2: 0.2	C4: 0.7	C6: 0.5
C3: 0.3	C5: 0.1	C7: 0.4

- Output of a state is determined by a probability distribution
- Multiple states can share the same output symbols
- True state of the system is “hidden” from the user

- Computes P(signal | phone)
- P([C1,C4,C6] | [m]) = prob. of going from O→M→E by the output probs.
  (0.7 x 0.1 x 0.6) x (0.5 x 0.7 x 0.5) = 0.0075
- P([C1,C3,C4,C6] | [m]) = P(O→O→M→E) + P(O→M→M→E) =
  (0.3 x 0.7 x 0.1 x 0.6) x (0.5 x 0.3 x 0.7 x 0.5) +
  (0.7 x 0.9 x 0.1 x 0.6) x (0.5 x 0.2 x 0.7 x 0.5)
  = 0.0006615 + 0.001323 = 0.0019845

The Acoustic Model: P(signal | words) Putting it all together

- Language bigram model gives us

  $P(word_i \mid word_{i-1})$ -or- P(word | words)

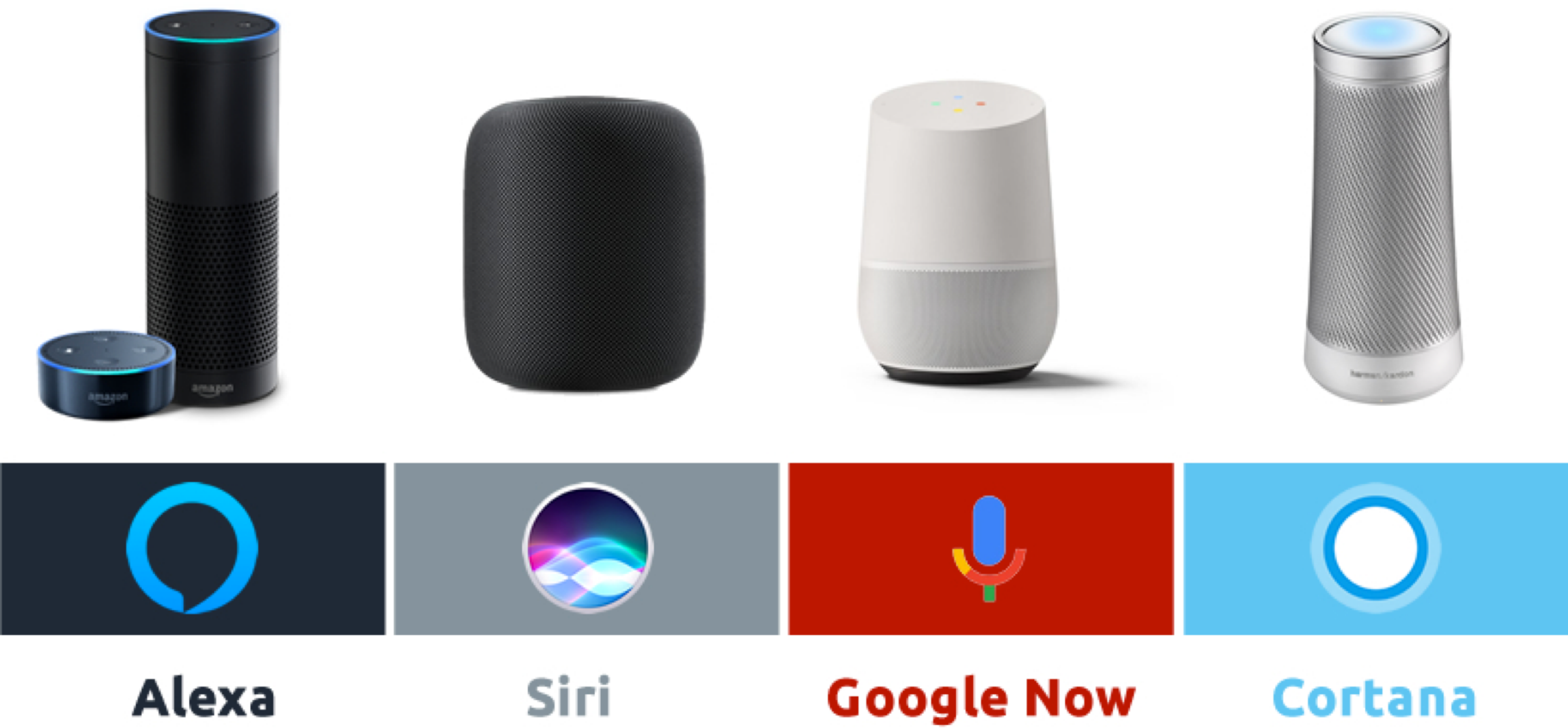
  – Like an HMM in which each word is a state and each bigram probability is a transition between states
- Word pronunciation HMM gives us

  P(phones | word)
- Phone HMM gives us

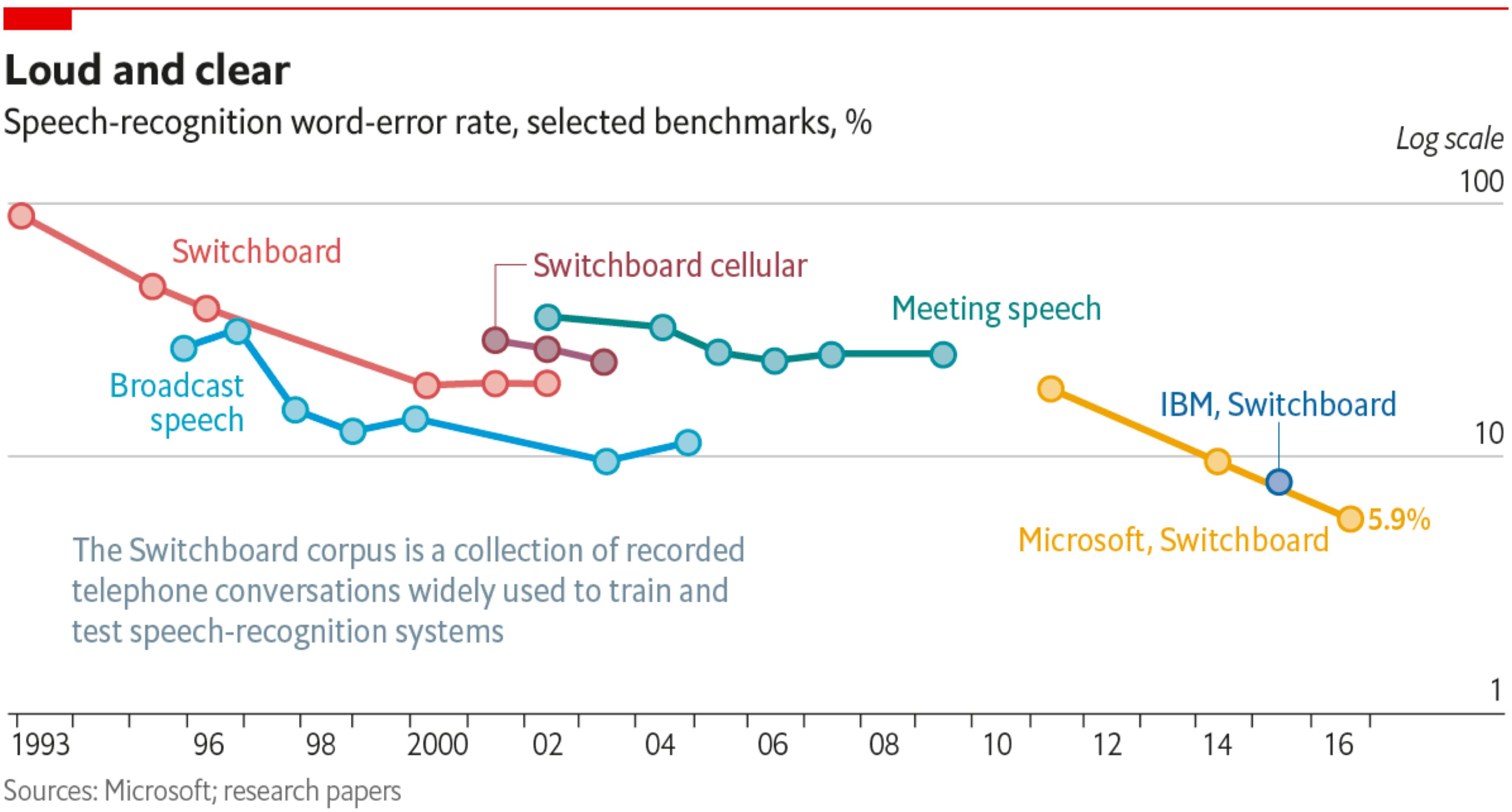
  P(signal | phones)
- Put them all together into one big HMM

  P(signal | words) = P(signal | phones) * P(phones | word) * P(word | words)

Modern Voice Assistants



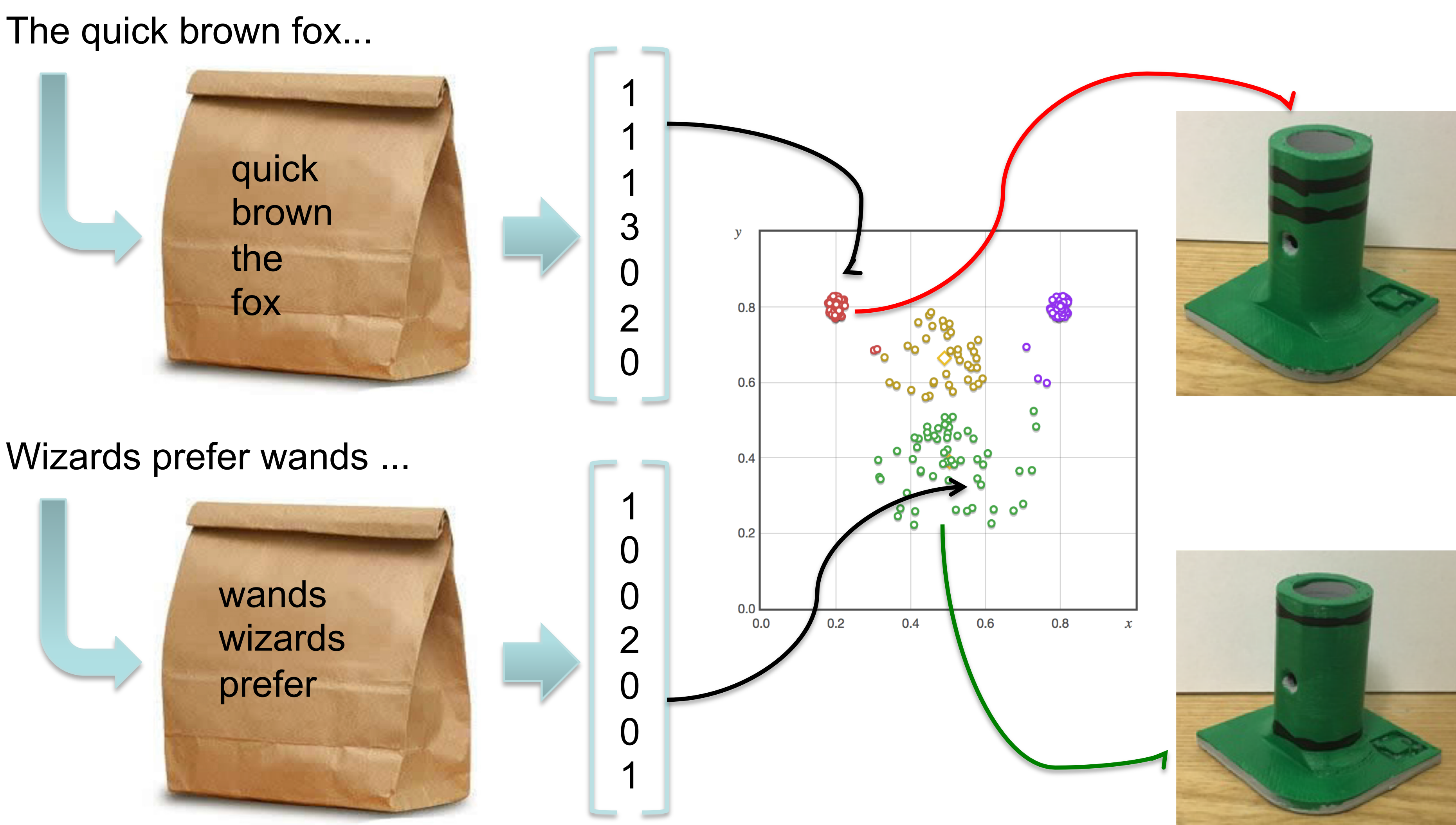
Word Recognition Error Rates



Source: https://www.economist.com/technology-quarterly/2017-05-01/language

Context-Free Approaches: Bag-of-Words Models

Context-Free Approaches: Bag-of-Words Models



Feedback is Critical

- Unsupervised learning: no indication is given whether an output was correct or incorrect
- Supervised learning: when an error occurs, agent receives the correct output
- Reinforcement learning: when an error occurs, agent receives an evaluation of its output, but is not told the correct output

Goals of Unsupervised Learning

- To find useful representations of the data, for example:
  - finding clusters, e.g. ***k*-means**, ART
  - dimensionality reduction, e.g. PCA, Hebbian learning, multidimensional scaling (MDS)
  - building topographic maps, e.g. elastic networks, Kohonen maps
  - finding the hidden causes or sources of the data
  - modeling the data density

Practical Uses of Unsupervised Learning

- Data compression
- Outlier detection
- Classification
- Make other learning tasks easier
- Model human learning and perception

Overview: *K*-Means

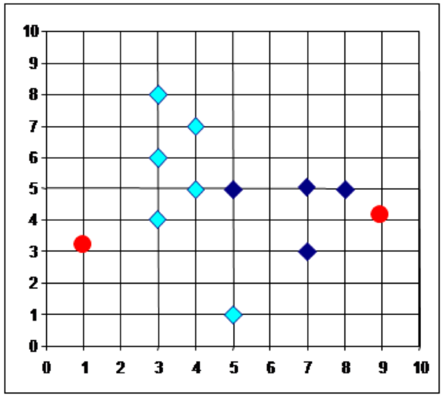
- Clustering is the process of partitioning a group of data points into a small number of clusters.
- In general, we have n data points $\boldsymbol{x}_i$, i=1…n to partition into k clusters.
- *K*-means aims to find the positions u_i, i=1…k that minimize the distance from the data points to the cluster, where c_i is the set of points belonging to to cluster i

$$\arg\min_{c} \sum_{i=1}^{k} \sum_{\mathrm{x} \in c_i} d(\mathrm{x}, u_i) = \arg\min_{c} \sum_{i=1}^{k} \sum_{\mathrm{x} \in c_i} \left\| \mathrm{x} - u_i \right\|_2^2$$

d= *s*quared Euclidian distance

- This is NP hard; *K*-means hopes to find global minimum

K-means example



K=2

Arbitrarily choose K object as initial cluster center

K-Means Algorithm (Lloyd's)

1. Initialize the center of the clusters

$$u_i = \text{ some value}, \; i = 1,\ldots,k$$

Since the algorithm stops in a local minimum, the initial position of the clusters is very important!

1. Attribute the closest cluster to each data point

$$\mathrm{c}_i = \left\{ j : d\left(\mathrm{x}_j, u_i\right) \le d\left(\mathrm{x}_j, u_l\right),\; l \ne i,\; j = 1,\ldots,n \right\}$$

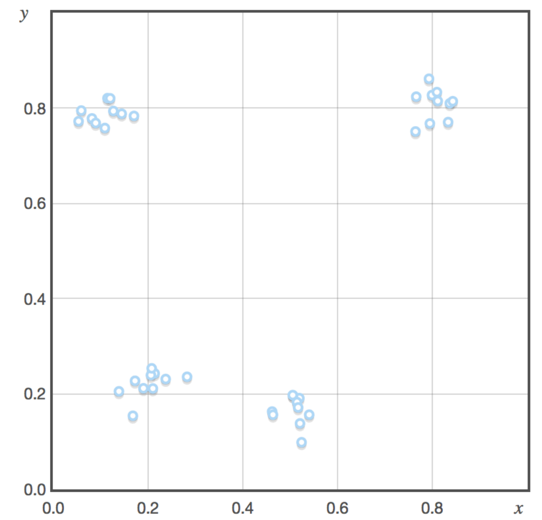
1. Set the position of each cluster to the mean of all data points belonging to that cluster

$$u_i = \frac{1}{\left|c_i\right|} \sum_{j \in c_i} \mathrm{x}_j,\; \forall i \qquad \text{where } \left|c\right| = \#\text{ elements in } c$$

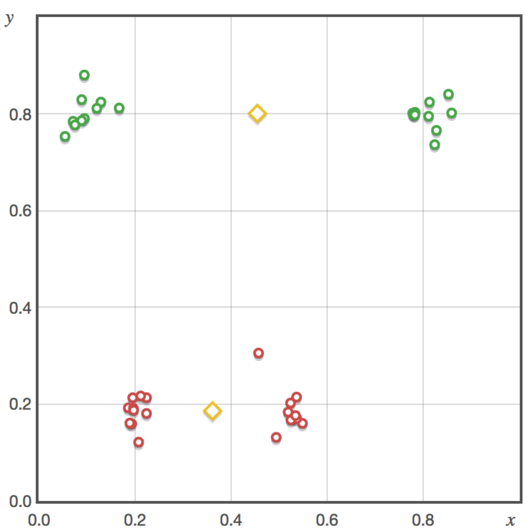
1. Repeat steps 2-3 until convergence.

K-Means: Example #1

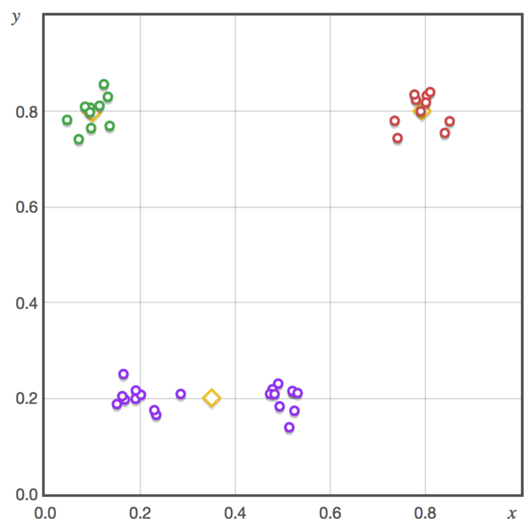
- Arbitrarily choose *K* objects as the initial cluster centers
- Repeat until no change:
  - Assign data points to closer cluster
  - Calculate center of each cluster



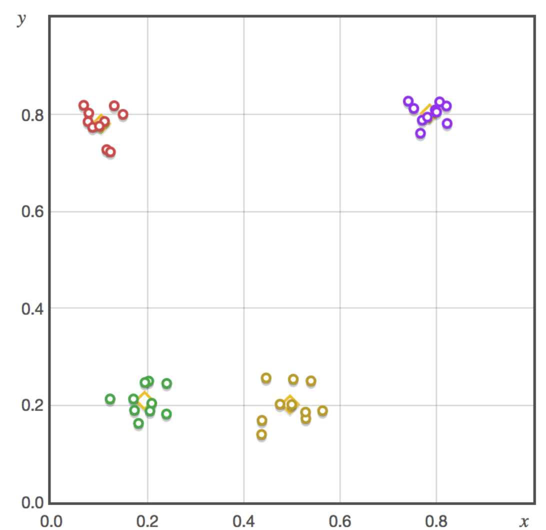
K=2



K=3

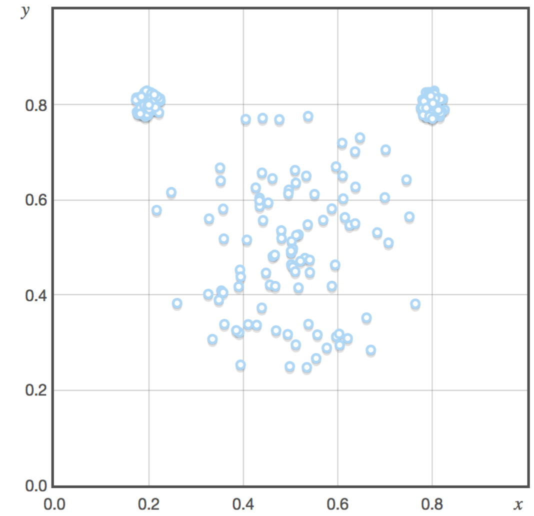


K=4

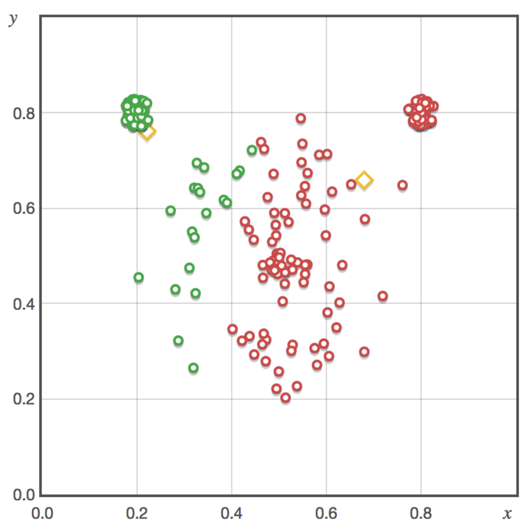


K-Means: Example #2

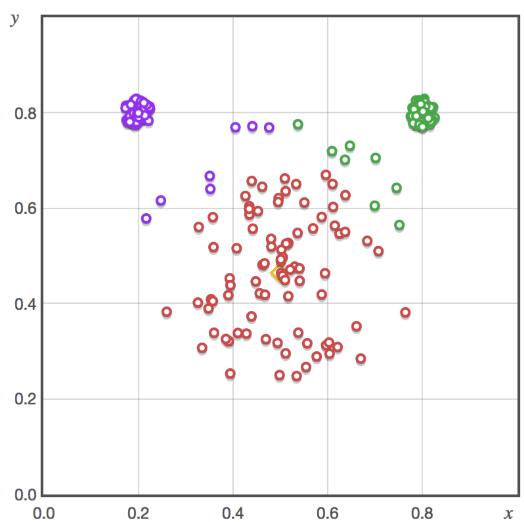
- Arbitrarily choose *K* objects as the initial cluster centers
- Repeat until no change:
  - Assign data points to closer cluster
  - Calculate center of each cluster



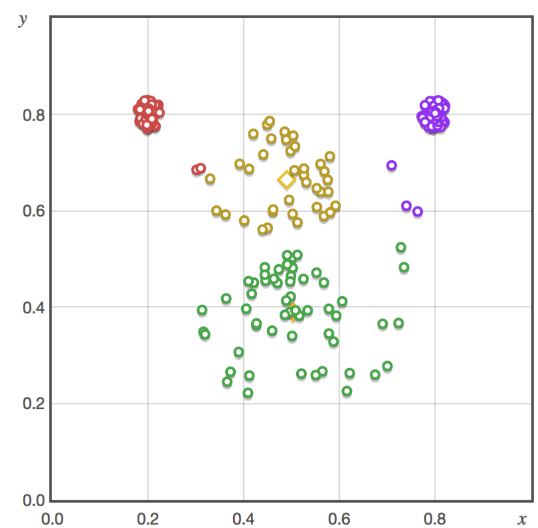
K=2



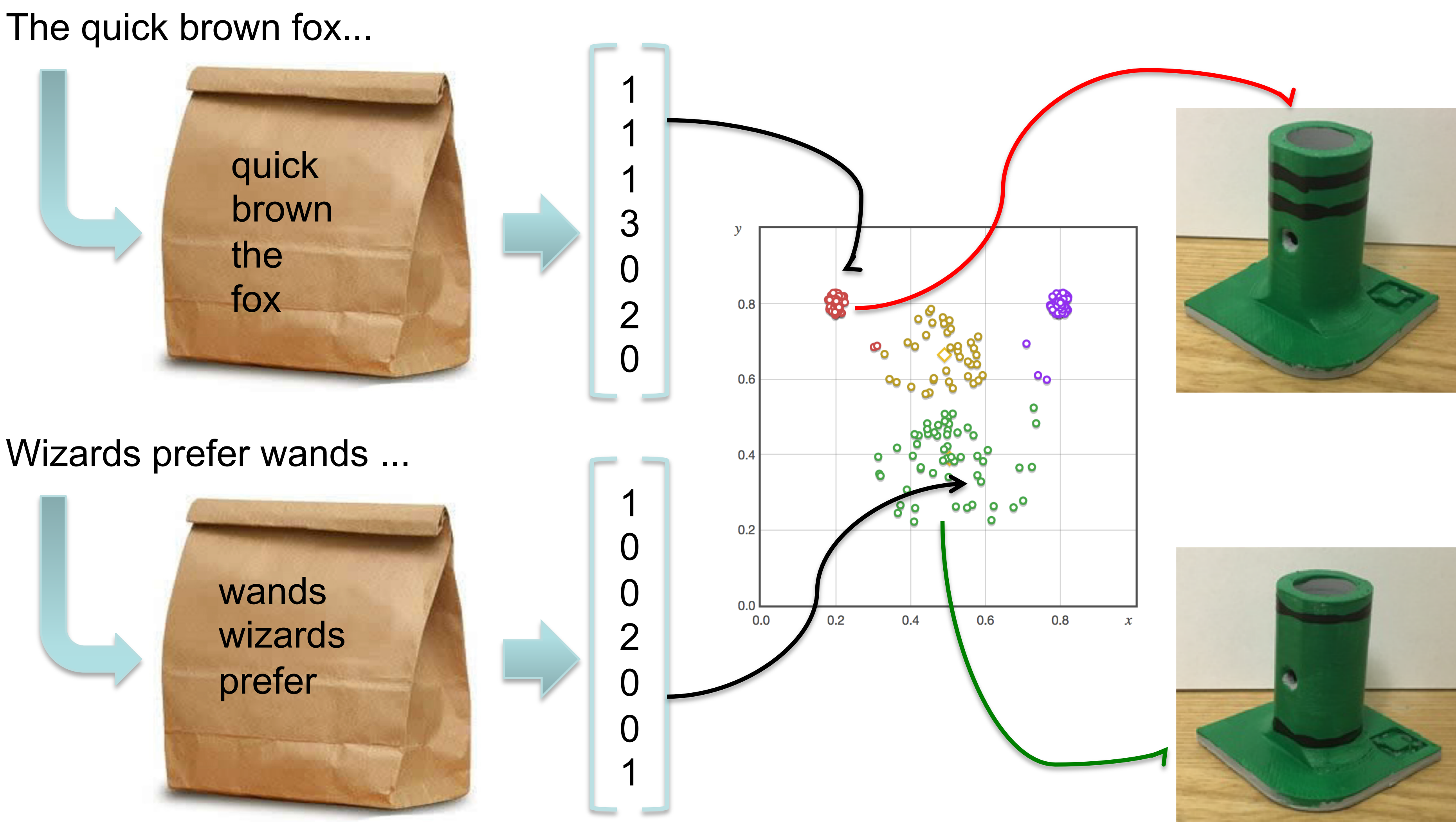
K=3



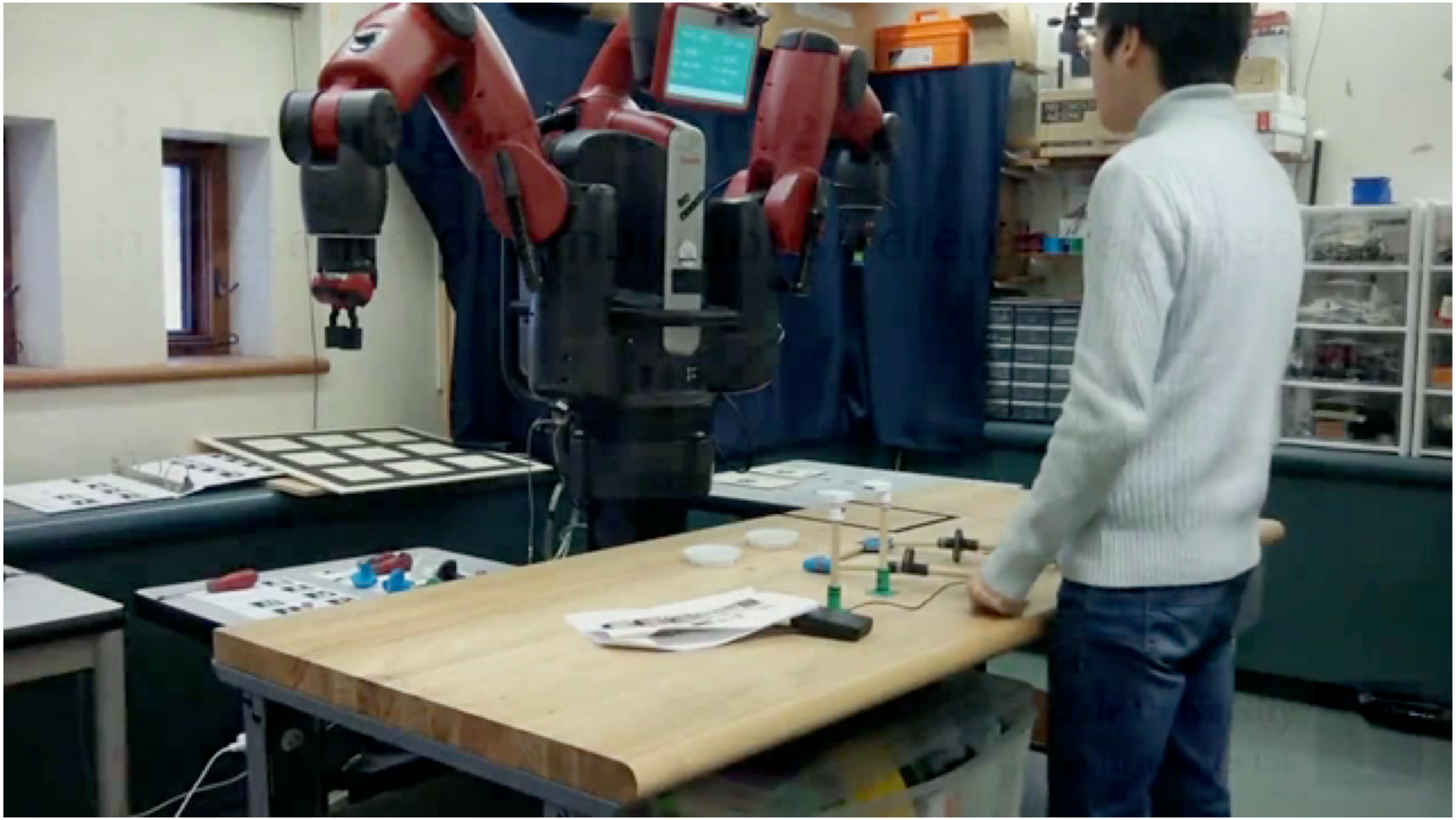
K=4



Context-Free Approaches: Bag-of-Words Models



Bag-of-Words for Object Selection



(Brawer, Widder, Roncone, Mangin & Scassellati, ICRA 2018)

Administrivia

- Next week
  - Perception (mostly vision)