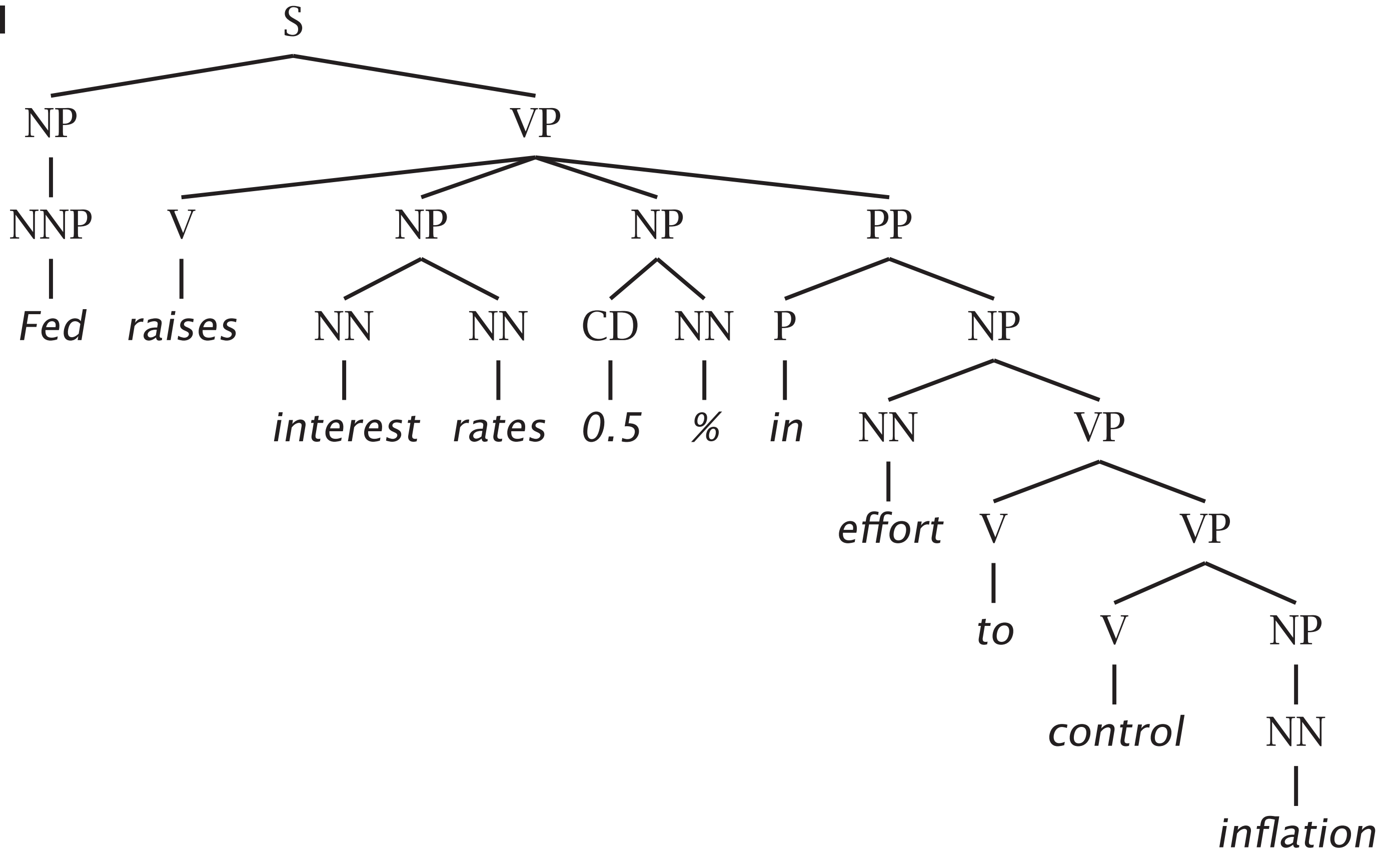
Part of Speech Tagging

FSNLP, chapters 9 and 10

Christopher Manning and

Hinrich Schütze

The problem of POS ambiguity

- Structures for: *Fed raises interest rates 0.5% in effort to control inflation* (*NYT* headline 17 May 2000)
- 

Part-of-speech ambiguities

		VB					
	VBZ	VBP	VBZ				
NNP	NNS	NN	NNS	CD	NN		
Fed	*raises*	*interest*	*rates*	*0.5*	*%*	*in*	*effort*
						to	*control*
							inflation

Part-of-speech examples

NN	noun	baby, toy
VB	verb	see, kiss
JJ	adjective	tall, grateful, alleged
RB	adverb	quickly, frankly, . . .
IN	preposition	in, on, near
DT	determiner	the, a, that
WP	wh-pronoun	who, what, which, . . .
CC	conjunction	and, or

POS ambiguity

- Words often have more than one POS: *back*
  - *The **back** door* = JJ
  - *On my **back*** = NN
  - *Win the voters **back*** = RB
  - *Promised to **back** the bill* = VB
- The POS tagging problem is to determine the POS tag for a particular instance of a word.

Why should we care?

- The first statistical NLP task
- Been done to death by different methods
- Easy to evaluate (how many tags are correct?)
- Canonical sequence (finite-state model) task
- Can be done well with methods that look at local context
- Though should ŞreallyŤ do it by parsing!
- Fast linear task of considerable value

The task of part of speech tagging

- A lightweight (usually linear time) processing task, which can usefully empower other applications:
  - Knowing how to pronounce a word: *récord* [noun] vs. *recórd* [verb]; *lead* as noun vs. verb
  - Matching small phrasal chunks or particular word class patterns for tasks such as information retrieval, information extraction or terminology acquisition (collocation extraction). Ee.g., just matching nouns, compound nouns, and adjective noun patterns:
    - {A|N}* N
  - POS information can be used to lemmatize a word correctly (i.e., to remove inflections):
    - *saw* [n] → *saw*; *saw* [v] → *see*

The task of part of speech tagging

- - Can differentiate word senses that involve part of speech differences
  - POS can be used as backoff in various class-based models, when too little information is known about a particular word
  - Can be a preprocessor for a parser (often better, but more expensive, to let the parser do the tagging as well)
  - Tagged text helps linguists find interesting syntactic constructions in texts (*ssh* used as a verb)

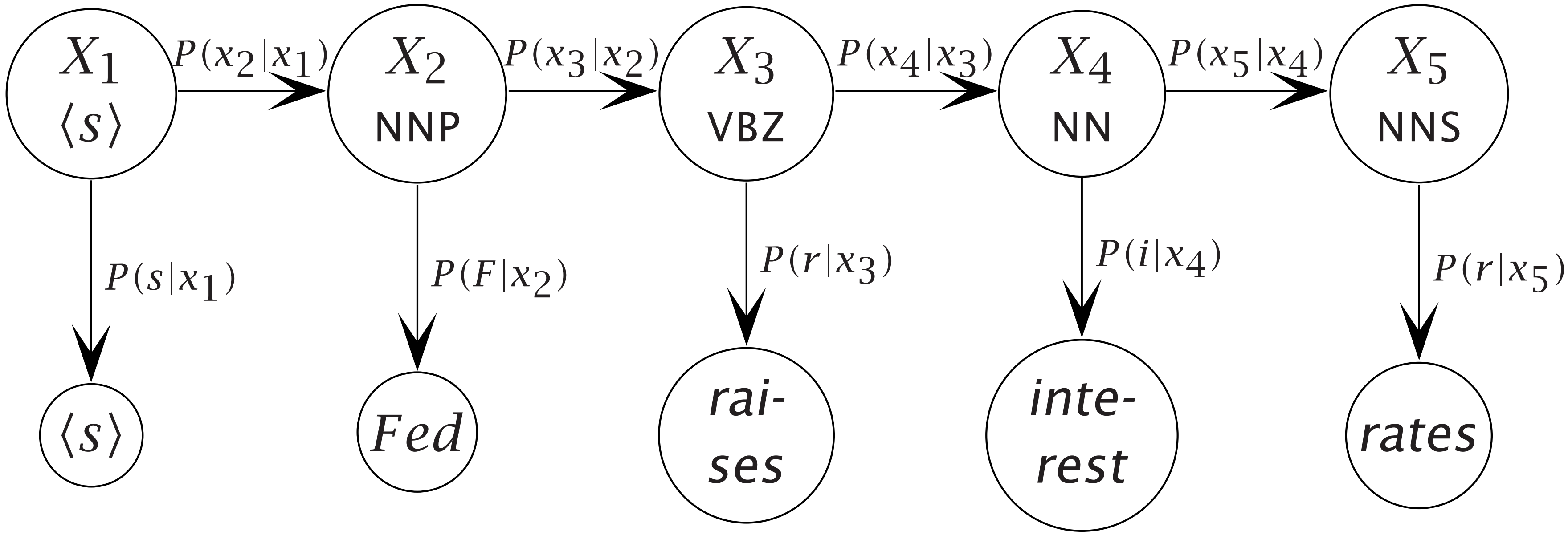
Part of speech tagging

Information sources:

- Sequence of words: syntagmatic information
  - Surprisingly weak information source
  - Many words have various parts of speech – cf. the example above
- Frequency of use of words
  - Surprisingly effective: gets 90+% performance by itself (for English)*
    - This acts as a baseline for performance

*Even up to 93.7%, based on the results of Toutanova et al. (2003).

Hidden Markov Models – POS example



- Top row is unobserved states, interpreted as POS tags
- Bottom row is observed output observations
- We normally do supervised training, and then (Bayesian network style) inference to decide POS tags