

# Statistical Natural Language Parsing

# An exponential number of attachments



# Attachment ambiguities

- A key parsing decision is how we ‘attach’ various constituents
  - PPs, adverbial or participial phrases, infinitives, coordinations, etc.

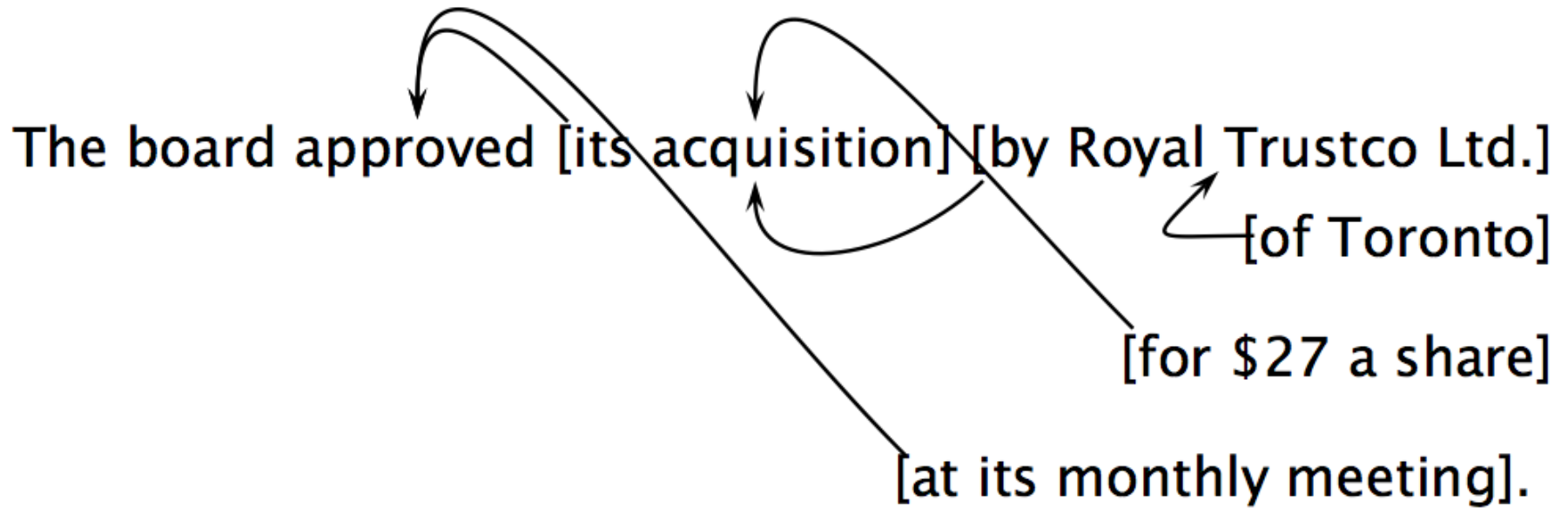
The board approved [its acquisition] [by Royal Trustco Ltd.]  
[of Toronto]  
[for \$27 a share]  
[at its monthly meeting].

- Catalan numbers:  $C_n = (2n)! / [(n+1)!n!]$
- An exponentially growing series, which arises in many tree-like contexts:
  - E.g., the number of possible triangulations of a polygon with  $n+2$  sides
    - Turns up in triangulation of probabilistic graphical models....



## Attachment ambiguities

- A key parsing decision is how we ‘attach’ various constituents
  - PPs, adverbial or participial phrases, infinitives, coordinations, etc.



- Catalan numbers:  $C_n = (2n)! / [(n+1)!n!]$
- An exponentially growing series, which arises in many tree-like contexts:
  - E.g., the number of possible triangulations of a polygon with  $n+2$  sides
    - Turns up in triangulation of probabilistic graphical models....



## Quiz Question!

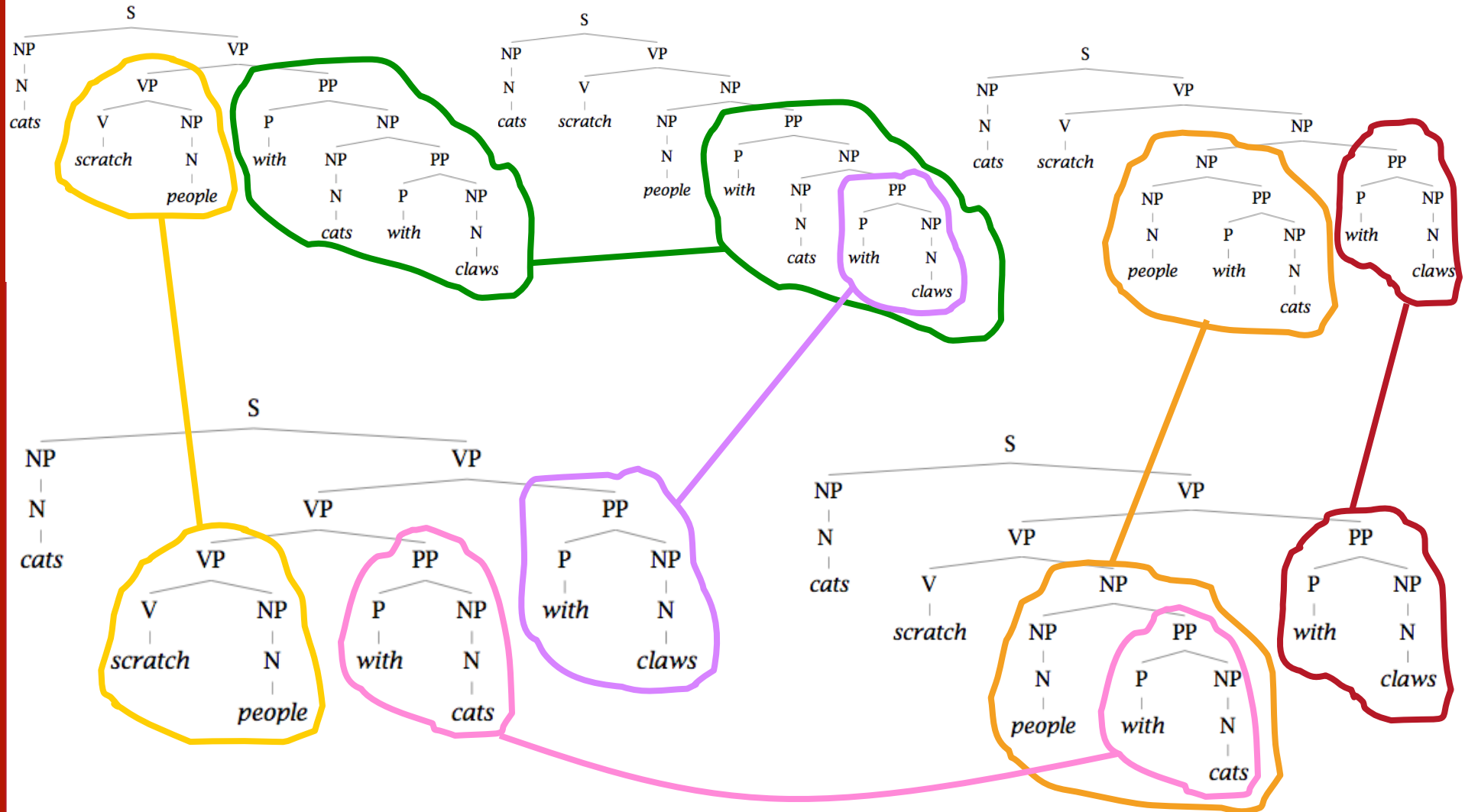
- How many distinct parses does the following sentence have due to PP attachment ambiguities?
  - A PP can attach to any preceding V or N within the verb phrase, subject only to the parse still being a tree.
  - (This is equivalent to there being no crossing dependencies, where if  $d_2$  is a dependent of  $d_1$  and  $d_3$  is a dependent of  $d_2$ , then the line  $d_2-d_3$  begins at  $d_2$  under the line from  $d_1$  to  $d_2$ .)

John wrote the book with a pen in the room.



# Two problems to solve:

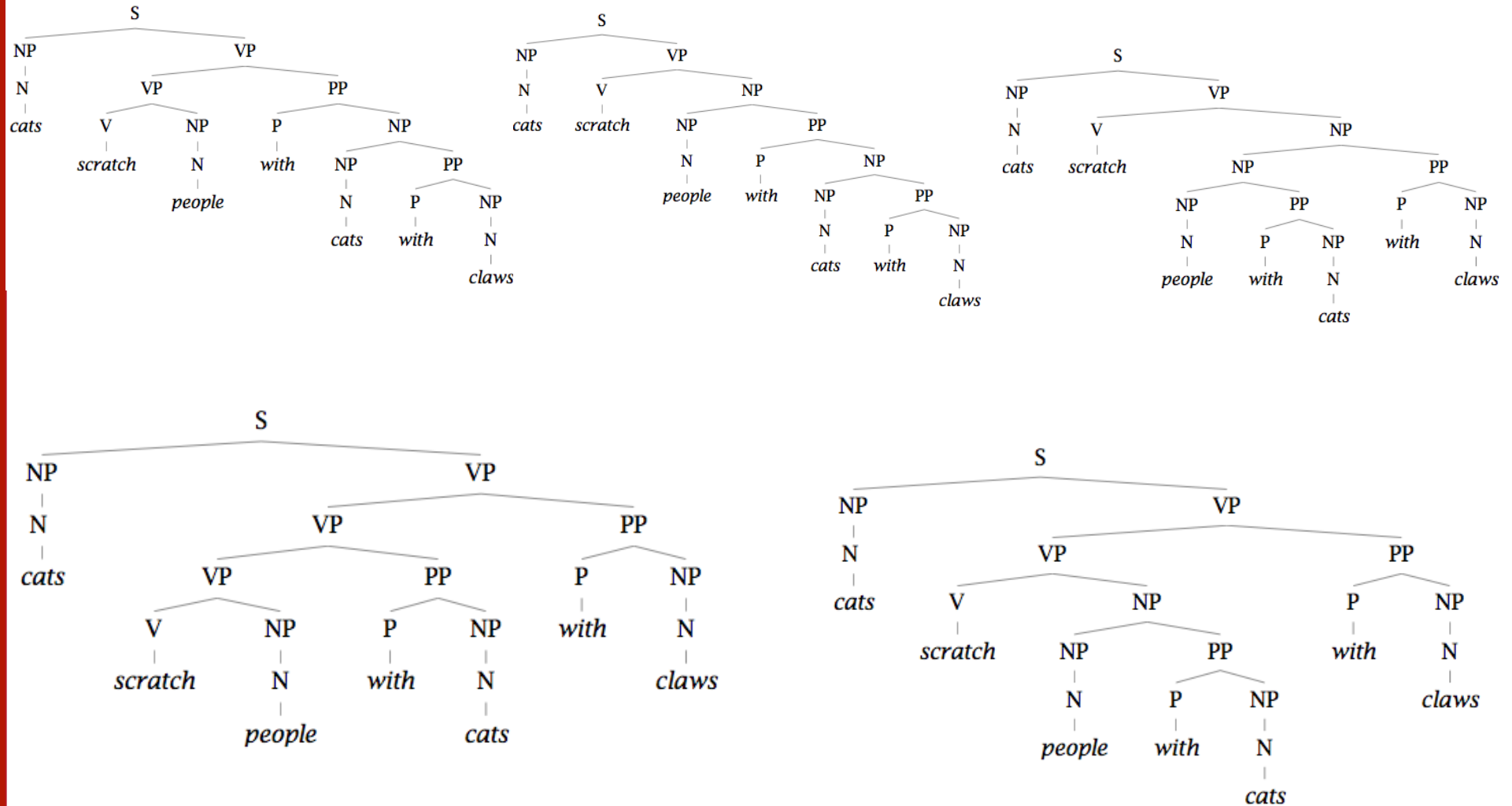
## 1. Repeated work...





# Two problems to solve:

## 1. Repeated work...





## Two problems to solve:

### 2. Choosing the correct parse

- How do we work out the correct attachment:
  - She saw the man with a telescope
- Is the problem 'AI complete'? Yes, but ...
- Words are good predictors of attachment
  - Even absent full understanding
  - Moscow sent more than 100,000 soldiers into Afghanistan ...
  - Sydney Water breached an agreement with NSW Health ...
- Our statistical parsers will try to exploit such statistics.

# Statistical Natural Language Parsing

# An exponential number of attachments