

# NLP

# Parsing

*Introduction and recap*

# Parsing Programming Languages

```
#include <stdio.h>

int main()
{
    int n, reverse = 0;

    printf("Enter a number to reverse\n");
    scanf("%d", &n);

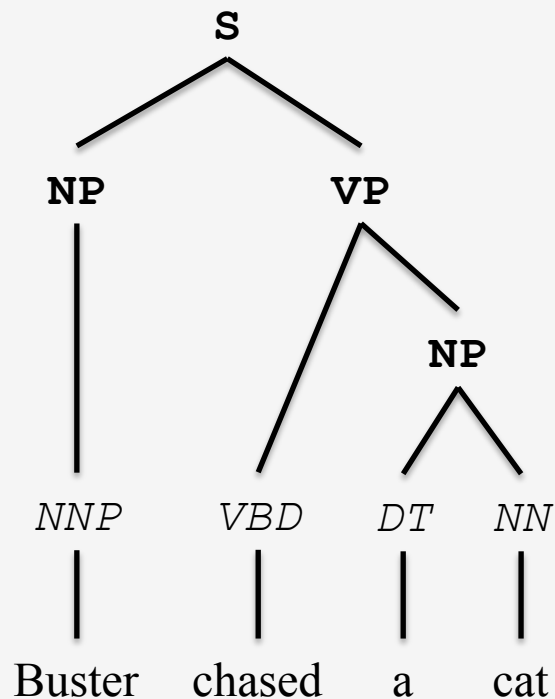
    while (n != 0)
    {
        reverse = reverse * 10;
        reverse = reverse + n%10;
        n = n/10;
    }
    printf("Reverse of entered number is = %d\n", reverse);

    return 0;
}
```

# Parsing Human Language

- **Coordination scope**: Small boys and girls are playing.
- **Prepositional phrase attachment**: I saw the man with the telescope.
- **Gaps**: Mary likes Physics but hates Chemistry.
- **Particles vs. prepositions**: She ran up a large bill.
- **Gerund vs. adjective**: Playing cards can be expensive.

# Phrase Structure



# Parsing

*Parsing noun sequences*

## Noun-noun Compounds

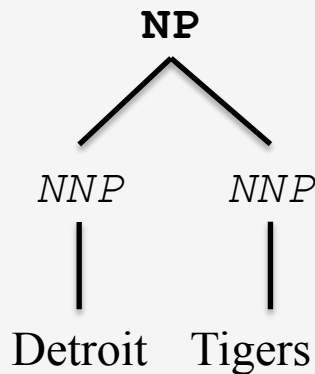
- Fish tank = tank that holds fish
- Fish net = net used to catch fish
- Fish soup = soup made with fish
- Fish oil = oil extracted from fish
- Fish sauce = sauce for fish dishes? sauce made of fish?

# Noun-noun Compounds

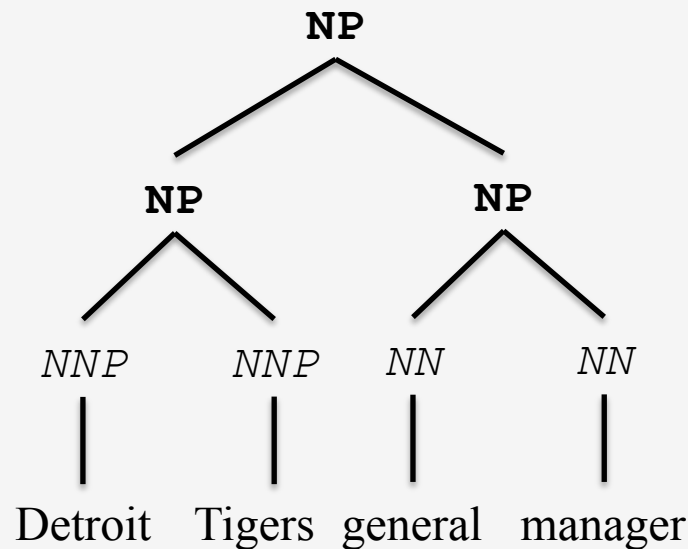
- Head of the compound
  - College junior – a kind of junior
  - Junior college – a kind of college
- Head first?
  - Attorney general
- Adjectives?
  - New Mexico, general manager
- More than two nouns?
  - luxury car dealership



# Noun Phrase Consisting Of Two Nouns



# Noun Phrase Consisting Of Four Nouns



# Representation Using Parentheses

- ((Salt Lake) City)
  - (Salt (Lake City))
- 
- Salt Lake City mayor?

# Solution

- (((Salt Lake) City) mayor)

# Representation Using Parentheses

- (((Salt Lake) City) mayor)
- ((Detroit Tigers) (general manager))
- Leland Stanford Junior University?

# Solution

- (((Leland Stanford) Junior) University)

# Combinatorics

- $n=2$   
(A B)
- $n=3$   
((A B) C)  
(A (B C))
- $n=4$   
((A B)(C D))

# Solution

- $n=4$

((A B)(C D))

(A (B (C D)))

(A ((B C) D))

((A (B C)) D)

((((A B) C) D))



# What About $n > 4$ ?

- $n=5$

$((A\ B)((C\ D)E))$

...

## Solution

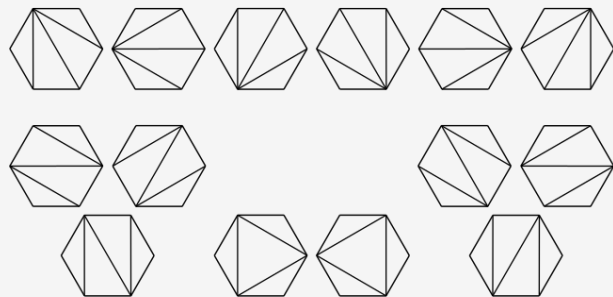
- The general solution is  $C(n)$ , a notation for the  $n^{\text{th}}$  Catalan number

$$C_n = \frac{1}{n+1} \binom{2n}{n}, \text{ for } n \geq 0$$

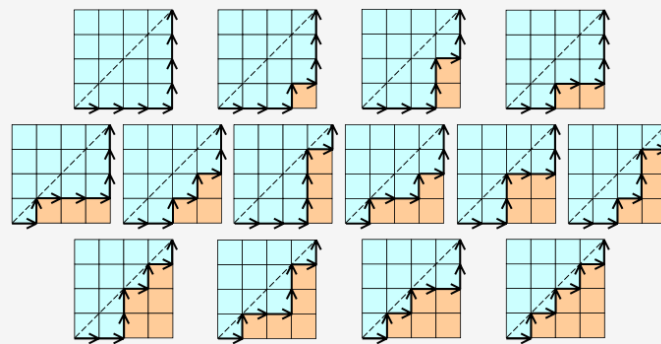
- 1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, 742900, ...
- Sequence A000108 in the On-Line Encyclopedia of Integer Sequences® (OEIS®)
- <https://oeis.org/>

# Other Uses Of Catalan Numbers

- the number of different ways a convex polygon with  $n + 2$  sides can be cut into triangles by connecting vertices with straight lines.
- the number of monotonic paths along the edges of a grid with  $n \times n$  square cells, which do not pass above the diagonal.



<http://en.wikipedia.org/wiki/File:Catalan-Hexagons-example.svg>



[http://en.wikipedia.org/wiki/File:Catalan\\_number\\_4x4\\_grid\\_example.svg](http://en.wikipedia.org/wiki/File:Catalan_number_4x4_grid_example.svg)

# NLP