

# Ling 5801: Lecture Notes 17

## Semantic Parsing

CFG parsers and Markov chains are generally weak predictors. Let's combine and improve them!

### Contents

17.1 Problem: parse trees depend on meaning . . . . .	1
17.2 Expensive Solution: word vectors . . . . .	1
17.3 Cheaper Solution: semantic class vectors . . . . .	3
17.4 Better Solution: dimensionality reduction . . . . .	5

### 17.1 Problem: parse trees depend on meaning

Consider the prepositional phrases **with ...** in the following sentences:

- **She ate** [<sub>N</sub> [<sub>N</sub> **pasta**] **with pesto**]. – attaches at noun phrase
- **She** [<sub>V-aN</sub> [<sub>V-aN</sub> **ate pasta**] **with a fork**]. – attaches at verb phrase

Preferred attachment seems to depend on the difference between words **pesto** and **a fork**.

However, [<sub>N</sub> [<sub>N</sub> **pesto**] **and** [<sub>N</sub> **a fork**]] coordinate just fine, so they are not different categories.

### 17.2 Expensive Solution: word vectors

We want to model preferred attachment, so we will use ‘prod\_pair’ and ‘max\_argmax’ operators, but marginalize over words.

Use a linear algebraic marginal (over words) inside a Viterbi parse (over trees and categories)!

Assume  $a = \langle p_a, \tau_a, \mathbf{v}_a \rangle \in \mathbb{R} \times \text{Tree} \times \mathbb{R}^D$  and  $b = \langle p_b, \tau_b, \mathbf{v}_b \rangle \in \mathbb{R} \times \text{Tree} \times \mathbb{R}^D$ :

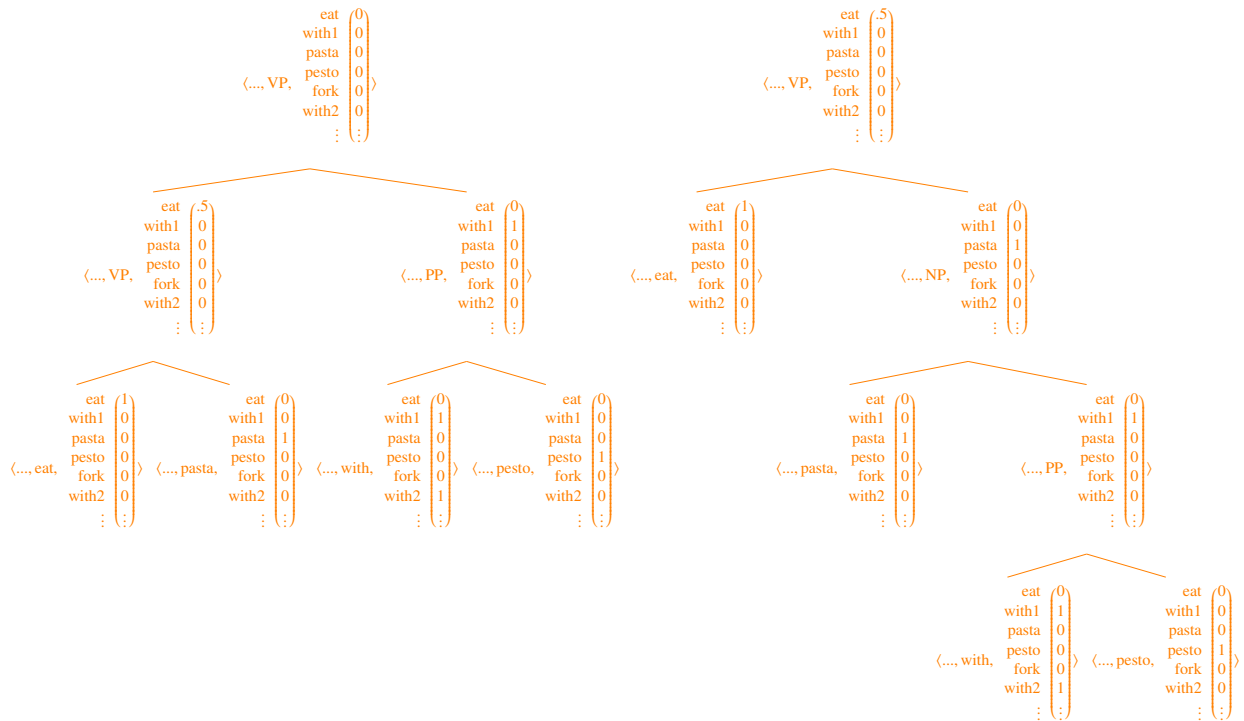
$$\text{prod\_pair}(a, b) = \langle p_a \cdot p_b, \langle \tau_a, \tau_b \rangle, \mathbf{G}(\mathbf{v}_a \otimes \mathbf{v}_b) \rangle$$

$$\text{max\_argmax}(a, b) = \begin{cases} a & \text{if } p_a \mathbf{v}_a^\top \mathbf{1} \geq p_b \mathbf{v}_b^\top \mathbf{1} \\ b & \text{if } p_a \mathbf{v}_a^\top \mathbf{1} < p_b \mathbf{v}_b^\top \mathbf{1} \end{cases}$$

Use matrix **G** to combine vectors of word senses:

$$\mathbf{G} = \begin{matrix} & \dots & \text{eat pasta} & \dots & \text{eat with2} & \dots & \text{with1 pesto} & \dots & \text{pasta with1} & \dots & \text{with2 fork} & \dots \\ \text{eat} & \dots & .5 & \dots & .5 & \dots & 0 & \dots & 0 & \dots & 0 & \dots \\ \text{with1} & \dots & 0 & \dots & 0 & \dots & 1 & \dots & 0 & \dots & 0 & \dots \\ \text{pasta} & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 1 & \dots & 0 & \dots \\ \text{pesto} & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots \\ \text{fork} & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots \\ \text{with2} & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 1 & \dots \\ \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots & \dots \end{matrix}$$

Preferred and dispreferred trees for *... ate pasta with pesto*:



Preferred and dispreferred trees for *... ate pasta with a fork*:





## 17.4 Better Solution: dimensionality reduction

Shrink  $\mathbf{G}$  matrix by capturing variation in fewer dimensions...