Computational Cognitive Science

Lecture 13: Sentence Processing

Chris Lucas
(Slides adapted from Frank Keller's)

School of Informatics
University of Edinburgh
clucas2@inf.ed.ac.uk

7 November, 2017

1. Background
   - Grammars and Processing
   - Ambiguity and Garden Paths
2. Modeling Human Parsing
   - Probabilistic Parallel Parser
   - Probabilistic Context-free Grammars
   - Results

Reading: Jurafsky (1996) (you can skip Section 3).

Introduction

In a previous lecture dealt with *lexical processing,* which turns a sound wave into a sequence of words.

⇒

look at the
yellow dog

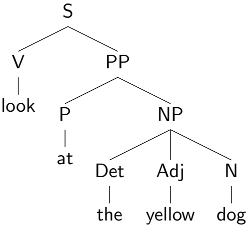
Introduction

In a previous lecture dealt with *lexical processing*, which turns a sound wave into a sequence of words.

⇒

look at the yellow dog

⇒



In this lecture, we will look at *sentence processing* (parsing), which turns a sequence of words into a syntactic representation.

Syntactic representations make explicit how the words in a sentence relate to each other.

Context-free Grammar

In order to build syntactic representations, we need a grammar. A simple type of grammar is a *context-free grammar:*

Phrasal categories:
S: sentence, NP: noun phrase, VP: verb phrase

Lexical categories (parts of speech):
Det: determiner, N: noun, V: verb

Context-free rules:

S	$\rightarrow$	NP VP	Det	$\rightarrow$	the
NP	$\rightarrow$	Det N	N	$\rightarrow$	kitten
VP	$\rightarrow$	V NP	N	$\rightarrow$	dog
			V	$\rightarrow$	bit

Derivations and Syntax Trees

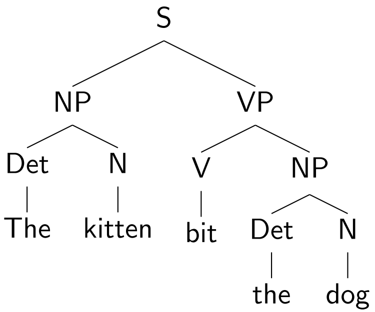
A *derivation* is the sequence of strings that results from applying a sequence of grammar rules, starting from a start symbol, here S:

Derivation

S $\Rightarrow$ NP VP $\Rightarrow$ NP V NP $\Rightarrow$ NP V Det N $\Rightarrow$ NP bit Det N $\Rightarrow$ NP bit Det dog $\Rightarrow$ NP bit the dog $\Rightarrow$ Det N bit the dog $\Rightarrow$ the N bit the dog $\Rightarrow$ the kitten bit the dog

Derivations and Syntax Trees

Derivations are represented as *syntax trees:*



A sentence can have multiple syntax trees, which typically correspond to different interpretations: *syntactic ambiguity.*

Global Ambiguity

Examples of ambiguous sentences:

(1)
a. She sat on the chair covered in dust.
b. I put the book on the table in the kitchen.
c. Kids make nutritious snacks.
d. Milk drinkers are turning to powder.
e. Old school pillars are replaced by alumni.

These are cases of *global ambiguity:* the sentence has multiple syntax trees, independently of how the trees are computed.

Examples from http://www.fun-with-words.com/ambiguous_garden_path.html

Human Sentence Processing

A *parser* is an algorithm that computes the syntax trees of a sentence, given a grammar.

The human parser is *incremental:* it builds trees word by word as the input arrives.

Incrementality can cause an additional form of ambiguity. Example:

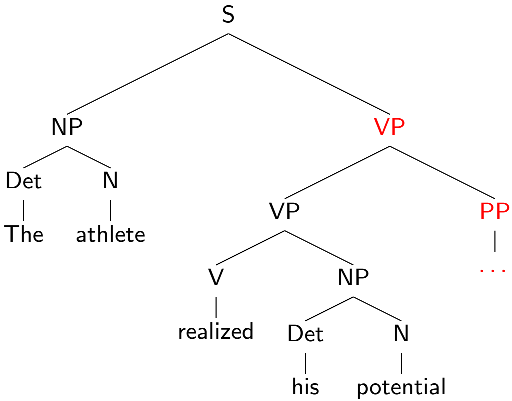
(2) The athlete realized his potential ...

a. ... at the competition.

b. ... could make him a world-class sprinter.

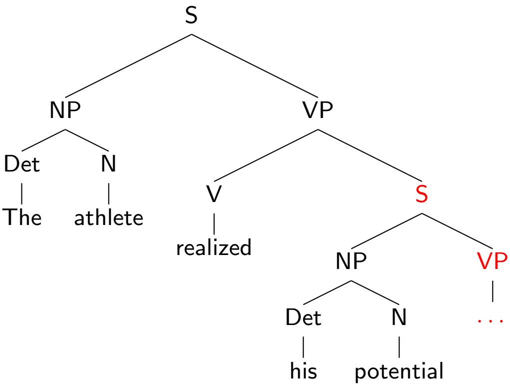
This is *local ambiguity:* the input up to the current word is compatible with more than one syntax tree; the ambiguity is resolved once the rest of the input is processed.

Local Ambiguity: Tree 1 (VP $\rightarrow$ V NP)



(3) The athlete realized his potential *at the competition.*

Local Ambiguity: Tree 2 (VP $\rightarrow$ V S)



(4) The athlete realized his potential *could make him a world-class sprinter.*

Garden Paths

This is an example of a *garden path:*

- both trees are compatible with the input up until *potential*; only the next word disambiguates;
- however, the processor commits to a single (wrong) tree early on, and trips up when later input is inconsistent with that tree;
- presumably, the processor now has to compute a new tree that is consistent with the input;
- garden path sentences result in longer reading times, reverse eye-movements, lower comprehension accuracies, etc.

Garden Paths

Examples of garden paths triggered by local ambiguity:

(5) a. The complex houses married and single students.
    b. The horse raced past the barn fell.

But: the following examples are not garden paths, even though they have the same syntactic structure:

(6) a. The warehouse fires a dozen employees each year.
    b. The bird found in the room died.

Our model needs to be able to explain this.

Probabilistic Parallel Parser

Jurafsky (1996) proposes an *incremental, parallel parsing model:*

- all (partial) parses compatible with the current input are constructed in parallel;
- each parse is assigned a probability;
- parses are pruned from the search space if their probability is a factor of α below the most probable parse: *beam search;*
- garden paths occur when the tree that is ultimately correct has been pruned.

How are parse probabilities determined?

Computing Parse Probabilities

Jurafsky, 1996 focuses on two sources of information:

- *construction probability:* the probability of a syntactic tree;
- *valence probability:* the probability of particular syntactic categories as arguments for a specific verbs.

Assume that construction probability and valence probability are independent:

$$P(tree) = P(construction) \cdot P(valence)$$

We can use a *probabilistic context-free grammar* to compute $P(construction)$.

Probabilistic Context-free Grammars

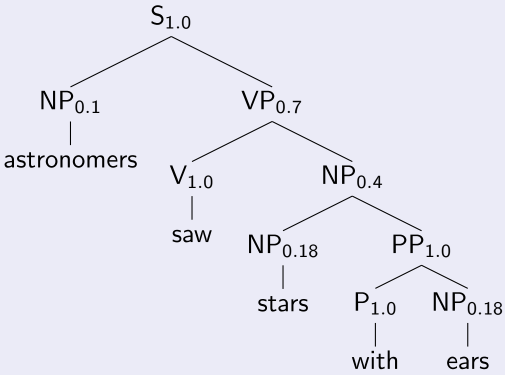
Example (Manning & Schütze, 1999)

S $\rightarrow$ NP VP	1.0	NP $\rightarrow$ NP PP	0.4
PP $\rightarrow$ P NP	1.0	NP $\rightarrow$ astronomers	0.1
VP $\rightarrow$ V NP	0.7	NP $\rightarrow$ ears	0.18
VP $\rightarrow$ VP PP	0.3	NP $\rightarrow$ saw	0.04
P $\rightarrow$ with	1.0	NP $\rightarrow$ stars	0.18
V $\rightarrow$ saw	1.0	NP $\rightarrow$ telescopes	0.1

- $P(A \rightarrow B\ C)$ is defined as $P(B\ C|A)$, the probability of deriving the right-hand side $B\ C$ given the left-hand side is A;
- the probabilities of all rules with the same LHS sum to one;
- $P(construction) = \prod_i P(rule_i)$ for all rules applied in tree.

Probabilistic Context-free Grammar

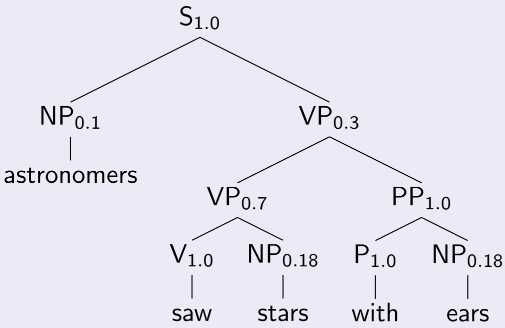
Example (Manning & Schütze, 1999)



$$P(t_1) = 1.0 \cdot 0.1 \cdot 0.7 \cdot 1.0 \cdot 0.4 \cdot 0.18 \cdot 1.0 \cdot 1.0 \cdot 0.18 = 0.0009072$$

Probabilistic Context-free Grammar

Example (Manning & Schütze, 1999)



$P(t_2) = 1.0 \cdot 0.1 \cdot 0.3 \cdot 0.7 \cdot 1.0 \cdot 0.18 \cdot 1.0 \cdot 1.0 \cdot 0.18 = 0.0006804$

Hence t_1 *more probable than* t_2 according to our PCFG.

Estimating PCFG Probabilities

The probabilities of a PCFG can be estimated from a *treebank,* a corpus annotated with syntactic trees.

We can use the *relative frequency* with which a rule $A \rightarrow B\ C$ occurs in a corpus to estimate its probability:

$$P(A \rightarrow B\ C) = \frac{c(A \rightarrow B\ C)}{\sum_{\xi} c(A \rightarrow \xi)}$$

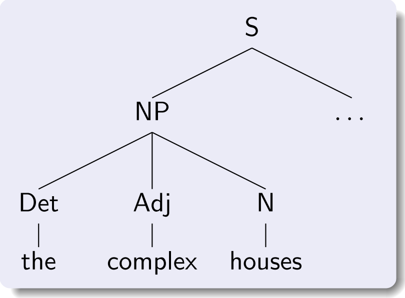
Where $c(A \rightarrow \xi)$ counts how often $A \rightarrow \xi$ occurs in the corpus.

Now we can use this approach to explain the examples of *local ambiguity* that we saw earlier.

Modeling Garden Path Effects

Garden path caused by construction probabilities:

S → NP ...	0.92	N → houses	0.00055
NP → Det Adj N	0.28	Adj → complex	0.00086
Det → the	0.71		

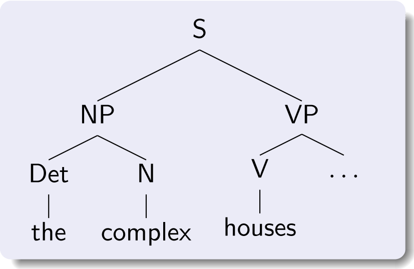


$P(t_1) = 8.5 \cdot 10^{-8}$ (preferred)

Modeling Garden Path Effects

Garden path caused by construction probabilities:

NP $\rightarrow$ Det N	0.63	V $\rightarrow$ houses	0.000052
S $\rightarrow$ [NP $_{\text{VP}}$[V ...	0.48	Det $\rightarrow$ the	0.71
N $\rightarrow$ complex	0.000029		

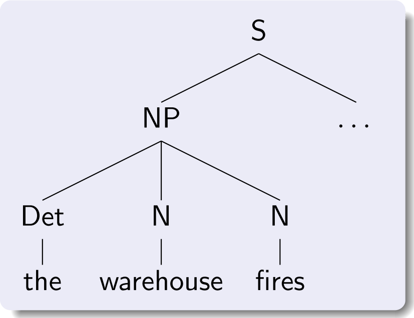


$P(t_2) = 3.2 \cdot 10^{-10}$ (grossly dispreferred)

Modeling Ambiguity

Ambiguous construction, no garden path:

S $\rightarrow$ NP ...	0.92	N $\rightarrow$ fires	0.00017
NP $\rightarrow$ Det N N	0.28		

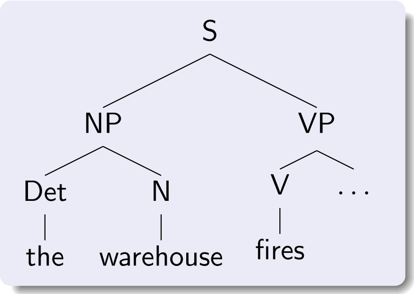


$P(t_1) = 4.2 \cdot 10^{-5}$ (preferred)

Modeling Ambiguity

Ambiguous construction, no garden path:

NP $\rightarrow$ Det N	0.63	V $\rightarrow$ fires	0.000036
S $\rightarrow$ [NP $_{VP}$[V ...	0.48		

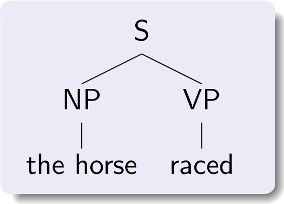


$P(t_2) = 1.1 \cdot 10^{-5}$ (mildly dispreferred)

Combining Valence and Construction Probabilities

Garden path caused by construction probabilities and valence probabilities:

$P(\text{race}, \langle\text{agent}\rangle) = 0.92$



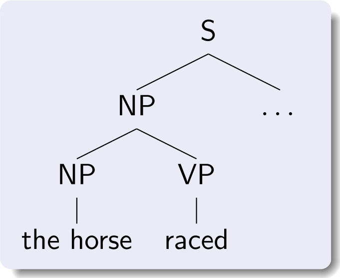
$P(t_1) = 0.92$ (preferred)

Combining Valence and Construction Probabilities

Garden path caused by construction probabilities and valence probabilities:

$P(\text{race}, \langle \text{agent}, \ \text{theme} \rangle) = 0.08$

NP $\rightarrow$ NP XP 0.14

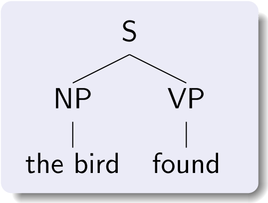


$P(t_2) = 0.0112$ (grossly dispreferred)

Combining Valence and Construction Probabilities

Disambiguation using construction probabilities and valence probabilities, no garden path:

$P(\text{find}, \langle\text{agent}\rangle) = 0.38$



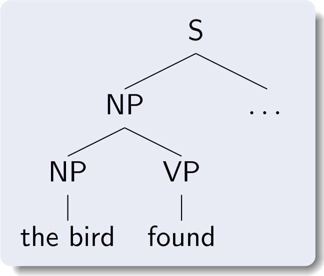
$P(t_1) = 0.38$ (preferred)

Combining Valence and Construction Probabilities

Disambiguation using construction probabilities and valence probabilities, no garden path:

$P(\text{find}, \langle \text{agent}, \ \text{theme} \rangle) = 0.62$

NP $\rightarrow$ NP XP   0.14



$P(t_2) = 0.0868$ (mildly dispreferred)

Setting the Beam Width

Crucial assumption: if the relative probability of a tree falls below a certain value, then it will be pruned.

sentence	probability ratio
the complex houses ...	267:1
the horse raced ...	82:1
the warehouse fires ...	3.8:1
the bird found ...	3.7:1

Assumption: a garden path occurs if the probability ratio is greater than threshold α (here: $\alpha \approx 5$).

Summary

- Sentences processing (parsing) is the task of assigning a syntax tree to a string of words;
- human sentence processing is incremental (word by word); this can lead to local ambiguity;
- garden paths derive from local ambiguities that are hard to resolve; they lead to longer processing times;
- Jurafsky, 1996 proposes an incremental, probabilistic parser as a model of human sentence processing;
- key modeling assumptions:
  - combination of PCFG probabilities and valence probabilities;
  - pruning of improbable trees (beam search);
  - garden paths happen when correct tree has been pruned.

References

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*(2), 137–194.

Manning, C. D. & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.