Complexity (continued); Models of human parsing

Informatics 2A: Lecture 26

Adam Lopez

18 November 2016

1. Complexity (continued)
2. Models of human parsing
   - Cognitive Constraints
   - Garden Paths
   - A cognitive model of human parsing

Reading: J&M, ch. 9 (pp. 350–352), ch. 12 (pp. 467–473), ch. 13 (pp. 491–496).

The complexity of natural language

In the last lecture, we saw that English is non-regular and Swiss-German is non-context-free. Each proof required **specific assumptions**. Among other things, each proof assumed that the construction under investigation (centre-embedding and cross-serial dependencies, respectively) could be iterated unboundedly.

In the last lecture, we saw that English is non-regular and Swiss-German is non-context-free. Each proof required **specific assumptions**. Among other things, each proof assumed that the construction under investigation (centre-embedding and cross-serial dependencies, respectively) could be iterated unboundedly.

Question 1. What can we conclude if this assumption isn't true?

The complexity of natural language

In the last lecture, we saw that English is non-regular and Swiss-German is non-context-free. Each proof required **specific assumptions**. Among other things, each proof assumed that the construction under investigation (centre-embedding and cross-serial dependencies, respectively) could be iterated unboundedly.

Question 1. What can we conclude if this assumption isn't true?

Question 2. How can we test assumptions empirically?

Strong versus Weak Adequacy

Questions about the formal complexity of language are about the expressive power of syntax, as represented by a grammar that's adequate for it. But there is more than one way to define adequacy!

Strong versus Weak Adequacy

Questions about the formal complexity of language are about the expressive power of syntax, as represented by a grammar that's adequate for it. But there is more than one way to define adequacy!

A weakly adequate grammar

- generates all and only the strings of a language. This is the approach we took in lecture 25. If our assumptions don't hold, then the proof falls apart, and we might not be able to show even non-regularity.
- Counter: a regular account of the strings in a language doesn't necessarily give a correct (insightful) account of its syntactic structures.

Strong versus Weak Adequacy

Questions about the formal complexity of language are about the expressive power of syntax, as represented by a grammar that's adequate for it. But there is more than one way to define adequacy!

A weakly adequate grammar

- generates all and only the strings of a language. This is the approach we took in lecture 25. If our assumptions don't hold, then the proof falls apart, and we might not be able to show even non-regularity.
- Counter: a regular account of the strings in a language doesn't necessarily give a correct (insightful) account of its syntactic structures.

A strongly adequate grammar

- generates all and only the strings of the language;
- assigns them the "right" structures — ones that support a correct representation of meaning. (See previous lecture.)

These 'crossing dependencies' are non-context-free in a very strong sense: no CFG is even weakly adequate for modelling them.

Other phenomena can *in theory* be modelled using CFGs, though it seems unnatural to do so. E.g. a versus an in English.

a banana	an apple
a large apple	an exceptionally large banana

Over-simplifying a bit: a before consonants, an before vowels.

In theory, we could use a context-free grammar:

$$\begin{array}{rcl} NP & \rightarrow & \mathbf{a}\ NP1^c \\ NP1^c & \rightarrow & N^c \mid AP^c\ NP1 \\ AP^c & \rightarrow & A^c \mid Adv^c\ AP \end{array} \qquad \begin{array}{rcl} NP & \rightarrow & \mathbf{an}\ NP1^v \\ NP1^v & \rightarrow & N^v \mid AP^v\ NP1 \\ AP^v & \rightarrow & A^v \mid Adv^v\ AP \end{array}$$

But more natural to use context-sensitive rules, e.g.

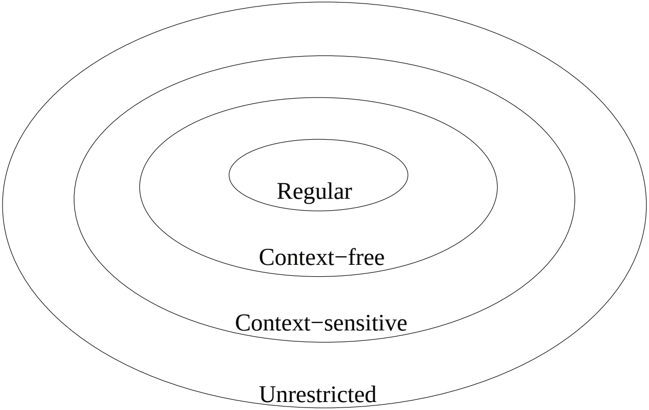
$$\begin{array}{rcl} \text{DET [c-word]} & \rightarrow & \textbf{a}\ \text{[c-word]} \\ \text{DET [v-word]} & \rightarrow & \textbf{an}\ \text{[v-word]} \end{array}$$

Review

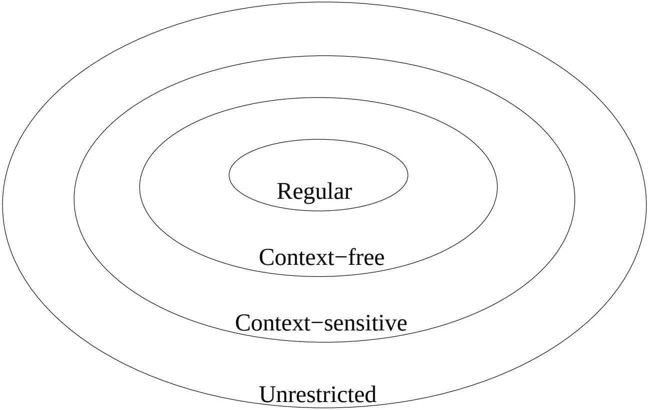
Chomsky Hierarchy: classifies languages on scale of complexity:

- Regular languages: those whose phrases can be 'recognized' by a finite state machine.
- Context-free languages: the set of languages accepted by pushdown automata. Many aspects of PLs and NLs can be described at this level;
- Context-sensitive languages: equivalent with a linear bounded nondeterministic Turing machine, also called a linear bounded automaton. Need this to capture e.g. *typing rules* in PLs.
- Unrestricted languages: *all* languages that can in principle be defined via mechanical rules.

Review



Review



Where do human languages fit within this complexity hierarchy?

Mild context sensitivity

A set $\mathcal{L}$ of languages is mildly context-sensitive if:

- $\mathcal{L}$ contains all context-free languages.
- $\mathcal{L}$ can describe cross-serial dependencies. There is an $n \geq 2$ such that $\{w^k | w \in T^*\} \in \mathcal{L}$ for all $k \leq n$.
- The languages in $\mathcal{L}$ are polynomially parsable.
- The languages in $\mathcal{L}$ have the constant growth property.

Let X be an alphabet and $L \subseteq X^*$. L has constant growth property iff there is a constant $c_0 > 0$ and a finite set of constants $C \subset \mathbb{N} \setminus \{0\}$ such that for all $w \in L$ with $|w| > c_0$, there is a $w' \in L$ with $|w| = |w'| + c$ for some $c \in C$

Example: the language $\{a^{2^n} | n \in \mathbb{N}\}$ does not have the constant growth property.

Summary: natural language complexity

- The ‘narrow’ language faculty involves a computational system that generates syntactic representations that can be mapped onto meanings.
- This raises the question of the complexity of this system (its position in the Chomsky hierarchy).
- A weakly adequate grammar generates the correct strings, while a strongly adequate one also generates the correct structures.
- NLs appear to surpass the power of context-free languages, but only just.
- The mild form of context-sensitivity seems weakly adequate for NL structures.

Human Parsing

So far, we looked at parsing from an engineering perspective. However, humans also do parsing to understand language.

The mathematical and algorithmic tools in this course can be used to analyze human parsing (human sentence processing). This is the domain of psycholinguistics.

Human Parsing

So far, we looked at parsing from an engineering perspective. However, humans also do parsing to understand language.

The mathematical and algorithmic tools in this course can be used to analyze human parsing (human sentence processing). This is the domain of psycholinguistics.

To study human parsing, we need:

- experimental data that tell us how humans parse;

Human Parsing

So far, we looked at parsing from an engineering perspective. However, humans also do parsing to understand language.

The mathematical and algorithmic tools in this course can be used to analyze human parsing (human sentence processing). This is the domain of psycholinguistics.

To study human parsing, we need:

- experimental data that tell us how humans parse;
- cognitive constraints derived from these data (e.g., incrementality, garden paths, memory limitations);

Human Parsing

So far, we looked at parsing from an engineering perspective. However, humans also do parsing to understand language.

The mathematical and algorithmic tools in this course can be used to analyze human parsing (human sentence processing). This is the domain of psycholinguistics.

To study human parsing, we need:

- experimental data that tell us how humans parse;
- cognitive constraints derived from these data (e.g., incrementality, garden paths, memory limitations);
- parsing models (and algorithms that implement them) that respect these constraints;

Human Parsing

So far, we looked at parsing from an engineering perspective. However, humans also do parsing to understand language.

The mathematical and algorithmic tools in this course can be used to analyze human parsing (human sentence processing). This is the domain of psycholinguistics.

To study human parsing, we need:

- experimental data that tell us how humans parse;
- cognitive constraints derived from these data (e.g., incrementality, garden paths, memory limitations);
- parsing models (and algorithms that implement them) that respect these constraints;
- an evaluation of the models against the data.

Incrementality

Parsing: extracting syntactic structure from a string; prerequisite for assigning a meaning to the string.

The human parser builds structures incrementally (word by word) as the input comes in.

This can lead to local ambiguity.

Example:

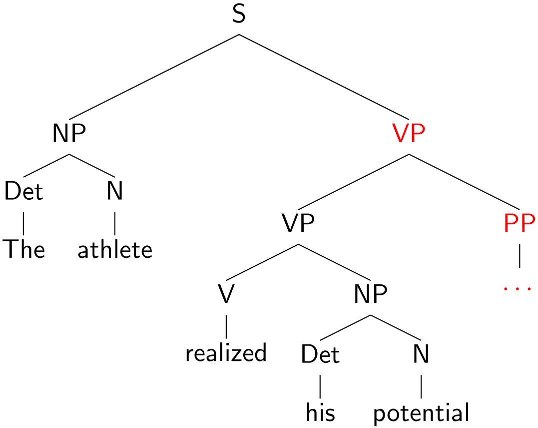
(1) The athlete realized his potential ...

a. ... at the competition.

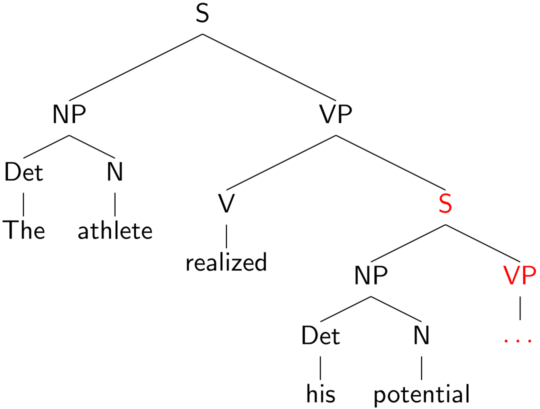
b. ... would make him a world-class sprinter.

Incrementality

Structure 1 (NP reading):



Structure 2 (S reading):



Garden Paths

- Early commitment: when it reaches *potential*, the processor has to decide which structure to build.
- If the parser makes the wrong choice (e.g., NP reading for sentence (1-b)) it needs to backtrack and revise the structure.
- A garden path occurs, which typically results in longer reading times (and reverse eye-movements).
- Some garden paths are so strong that the parser fails to recover from them.

More examples of garden paths:

(2)
a. The horse raced past the barn fell.
b. I convinced her children are noisy.
c. Until the police arrest the drug dealers control the street.
d. The old man the boat.
e. We painted the wall with cracks.
f. The cotton clothing is usually made of grows in Mississippi.
g. The prime number few.

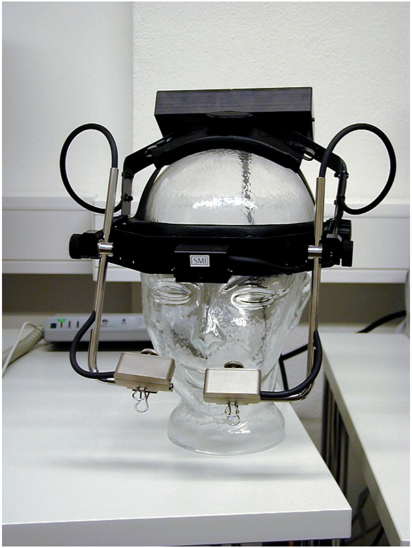
Eye-tracking

An eye-tracker makes it possible to record the eye-movements of subjects while they are performing a cognitive task:

- looking at a scene;
- driving a vehicle;
- using a computer;
- reading a text.

Mind's Eye Hypothesis: where subjects are looking indicates what they are processing. How long they are looking at it indicates how much processing effort is needed.

Eye-tracking



A head-mounted, video-based eye-tracker.

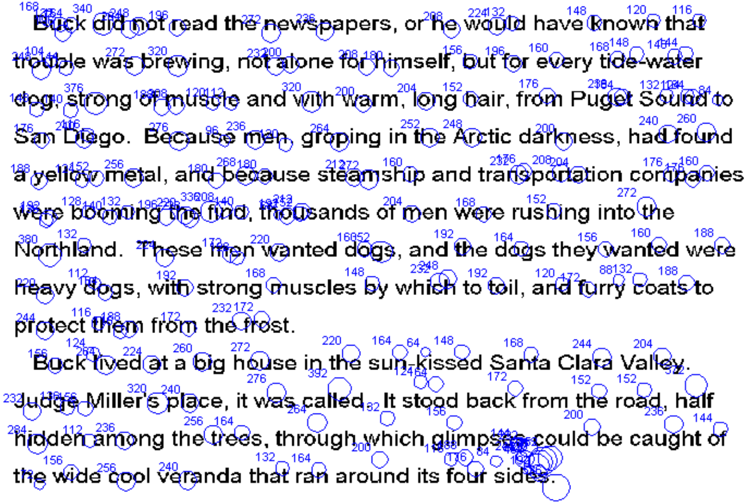
Eye-movements and Reading

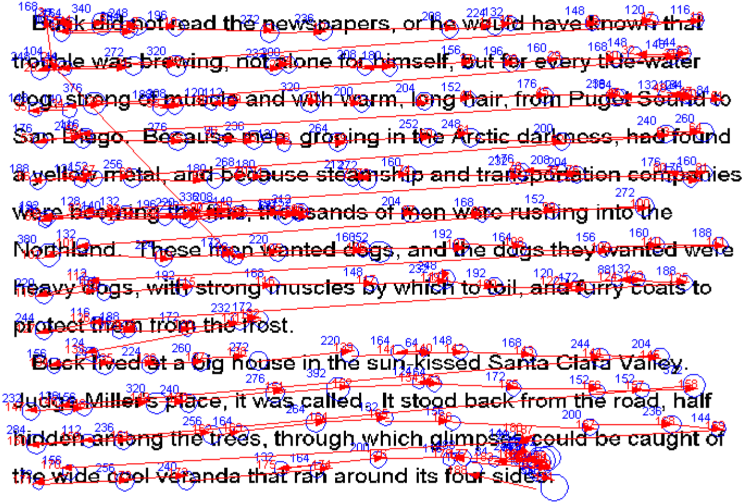
Let's look at eye-tracking data for reading in detail:

- eye-movements are recorded while subjects read texts;
- very high spatial (0.15° visual angle) and temporal (1 ms) accuracy;
- eye movements in reading are saccadic: a series of relatively stationary periods (fixations) between very fast movements (saccades);
- average fixation time is about 250 ms; can be longer or shorter, depending on ease or difficulty of processing;
- typically test a number of subjects, with a number of test sentences, and statistical analysis done on results.

Buck did not read the newspapers, or he would have known that trouble was brewing, not alone for himself, but for every tide-water dog, strong of muscle and with warm, long hair, from Puget Sound to San Diego. Because men, groping in the Arctic darkness, had found a yellow metal, and because steamship and transportation companies were booming the find, thousands of men were rushing into the Northland. These men wanted dogs, and the dogs they wanted were heavy dogs, with strong muscles by which to toil, and furry coats to protect them from the frost.

Buck lived at a big house in the sun-kissed Santa Clara Valley. Judge Miller's place, it was called. It stood back from the road, half hidden among the trees, through which glimpses could be caught of the wide cool veranda that ran around its four sides.


Buck did not read the newspapers, or he would have known that
trouble was brewing, not alone for himself, but for every tide-water
dog, strong of muscle and with warm, long hair, from Puget Sound to
San Diego. Because men, groping in the Arctic darkness, had found
a yellow metal, and because steamship and transportation companies
were booming the find, thousands of men were rushing into the
Northland. These men wanted dogs, and the dogs they wanted were
heavy dogs, with strong muscles by which to toil, and furry coats to
protect them from the frost.
Buck lived at a big house in the sun-kissed Santa Clara Valley.
Judge Miller's place, it was called. It stood back from the road, half
hidden among the trees, through which glimpses could be caught of
the wide cool veranda that ran around its four sides.


Buck did not read the newspapers, or he would have known that
trouble was brewing, not alone for himself, but for every tide-water
dog, strong of muscle and with warm, long hair, from Puget Sound to
San Diego. Because men, groping in the Arctic darkness, had found
a yellow metal, and because steamship and transportation companies
were booming the find, thousands of men were rushing into the
Northland. These men wanted dogs, and the dogs they wanted were
heavy dogs, with strong muscles by which to toil, and furry coats to
protect them from the frost.
Buck lived at a big house in the sun-kissed Santa Clara Valley.
Judge Miller's place, it was called. It stood back from the road, half
hidden among the trees, through which glimpses could be caught of
the wide cool veranda that ran around its four sides.

Eye-movements and Reading

We can use the data generated by eye-tracking experiments to investigate how the human parser works. For example:

- evidence for garden paths comes from increased reading times, and more reverse saccades, when reading certain words;
- evidence for incrementality comes from studies where participants view visual scenes while listening to sentences;
- evidence for interactivity comes from the fact that semantic properties of words influence reading times in the same way as syntactic ones.

We will sketch a model of these properties by building a parser that mimics human parsing behavior.

Clicker Question

Which of the following sentences in not a garden path?

1. The man returned to his house was happy.
2. The complex houses married and single soldiers and their families.
3. The tomcat that curled up on the cushion seemed friendly.
4. The sour drink from the ocean.

A cognitive model of human parsing

We've already seen an incremental parsing model: Earley!

Earley has Cognitively plausible incrementality: each word is integrated into the structure as it appears (no unconnected words).

Question. How might we adapt Earley to account for surprisal and garden paths?

A cognitive model of human parsing

We've already seen an incremental parsing model: Earley!

Earley has Cognitively plausible incrementality: each word is integrated into the structure as it appears (no unconnected words).

Question. How might we adapt Earley to account for surprisal and garden paths?

We've also seen the answer to this: probability!

Combine Earley with probabilities to simulate human parsing performance.

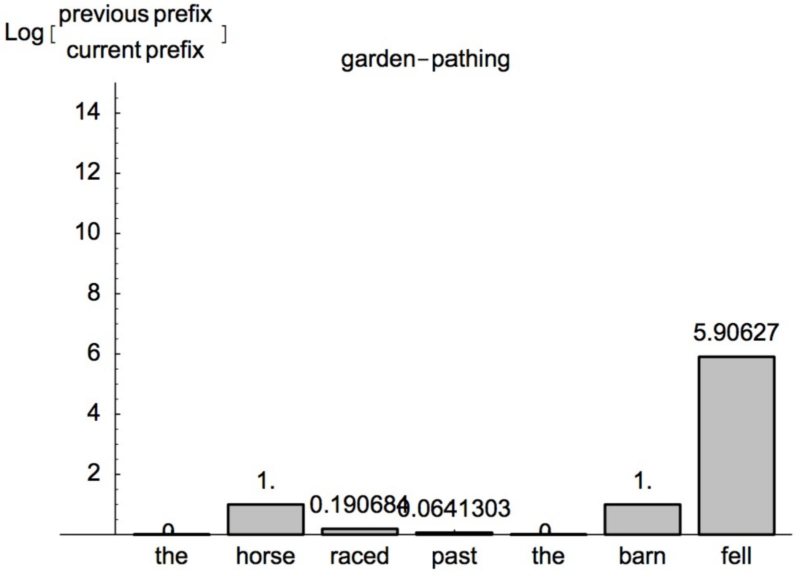
Probabilistic Earley parsing

Essential idea: use a probabilistic grammar (estimated in the usual way) to assign prefix probabilities: the probability of each prefix of the sentence.

Then measure the surprisal of each new word as a ratio of the prefix probability of a word and the next.

Places of high surprisal indicate a garden path.

Probabilistic Earley parsing simulates garden-pathing



Summary

- The human parser builds syntactic structure in response to strings of words;
- Parsing models have to capture the incrementality of human parsing and account for ambiguity resolution (garden paths);
- Known parsing algorithms (CYK, Earley) can be used, but...
- a simple bottom-up parser assumes limited incrementality, full parallelism: not cognitively plausible;
- Earley parsing models achieves a higher degree of incrementality;
- Probabilities model surprisal.