Computational Psycholinguistics

Lecture 1: Introduction

Matthew W. Crocker
crocker@coli.uni-sb.de

“To understand and model the processes that underlie the human capacity to understand language”

• How does the human language processor work?
• How is it realized in the brain?
• How can we model it computationally?
• Where does it come from?
• How does language interact with other cognitive systems and the environment?
What is it?

• Using computational techniques to better understand and model how people produce and comprehend language

• Competence: Principles that relate utterances to underlying meaning?

• Performance: How do people establish this relationship during on-line language processing?

• Computational psycholinguistics seeks cognitively plausible theories about both mental rules and representations, and about cognitive processes

• Computational psycholinguistics seeks to realize such theories as implemented computational models of human knowledge and behavior

Different from NLP?

• Early NLP (e.g. Winograd, 1983) clearly viewed itself as building models of human understanding

• Proposals were heavily informed by intuitions about how people understand, and linguistic theories about mental representations

• Modern NLP has shifted emphasis:

  • Application: do limited tasks accurately and robustly, often without real understanding (e.g. spam filters, IR, document clustering, summarization)

  • Deep NLU: Emphasis is on representations, coverage and efficiency. Little concern with cognitive plausibility
Areas of Psycholinguistics

• Speech perception and articulation
  • visual word recognition
• Lexical access and lexical choice
  • The mental lexicon
• Sentence processing:
  • syntactic, semantic, pragmatic
• Discourse and dialogue
  • Anaphora, priming, alignment
• Situated language processing:
  • interaction with task, context
  • the immediate environment
• Embodied language processing:
  • grounding language in action/perception systems of the brain
  • Language Acquisition and Development
  • Language Evolution

Models of Sentence Processing

• Language is complex & dynamic
  • multiple levels of representation & knowledge
  • each level has rich internal structure, unique constraints & representations
  • processing unfolds over time: both across levels, and in response to signal
  • levels interact in dynamically, and in complex ways
• We need computational models to understand ...
  • the dynamics & interactions of processing; the role of processing limitations
  • relate processing with empirical data; make predictions
So what ... 

- Speech streams include no discrete boundaries to indicate where one word ends and another begins.
- We understand non-fluent speech, fragments, interruptions and non-native speakers. Incomplete sentences are no problem for us.
- We deal with ambiguity all the time without breaking down. Computer parsers often maintain thousands of possible interpretations.
- We have a vocabulary of about 60,000 words. We access somewhere between 2-4 words/second (low error rates ~ 2/1000 words)
- We understand speech even faster than we can produce it. We are so fast, we can even finish each others sentences.
Human language processing

- People are highly **accurate** in understanding language
- People process language **rapidly**, in **real-time**
- People understand and produce language **incrementally**
- People even **anticipate** what's going to be said next
- People rapidly adjust to **context**, and are **robust**
- People achieve this despite **limitations** on processing resources
- People do make some interesting **errors**, and exhibit **breakdown** in certain situations ...

But things don’t alway go smoothly...

- Police police police police police.
  - *Internal affairs investigates the detectives that monitor other police.*
- The boat floated down the river sank.
  - *The boat that we floated down the river eventually sank.*
- The child put the candy on the table in his mouth.
  - *The child put the candy that was on the table in his mouth.*
- The editor authors the newspaper hired disliked resigned.
  - *The editor that reporters that the newspaper hired disliked resigned.*
- In New York, someone is hit by a car every 10 minutes..
  - *… and the poor guy is getting tired of it.*
Lexical access

• Visual & spoken word recognition
  • Central importance of lexical frequency
• Incremental & parallel access
  • words with similar onset & offset are activated (beetle vs beaker vs speaker)
• Multiple meanings
  • “Bug”: both insect & spy device senses are accessed initially
  • Rapid decay of non-preferred sense
• Key issue: Bottom-up versus Top-down “selection”

Sentence processing

• Sentence processing is the means by which the words of an utterance are combined to yield and interpretation
  • All people do it well
  • It is a difficult task: complexity and ambiguity
  • Unlike lexical access, it can’t simply be ‘retrieval’
• Compositional: interpretation must be constructed on-line, rapidly
  • Even for sentences with novel structures, or words used in novel positions
Context Free Grammars

• Context-free grammar rules:

\[
\begin{align*}
S & \rightarrow NP \ VP \\
PP & \rightarrow P \ NP \\
VP & \rightarrow V \ NP \\
VP & \rightarrow V \\
NP & \rightarrow NP \ PP \\
NP & \rightarrow Det \ N
\end{align*}
\]

\[
\begin{align*}
Det & \rightarrow \text{the} \\
Det & \rightarrow \text{every} \\
N & \rightarrow \text{man, woman} \\
N & \rightarrow \text{book} \\
P & \rightarrow \text{with} \\
V & \rightarrow \text{read, reads}
\end{align*}
\]

• Node admissibility criterion:

A tree is admitted by the grammar, if for each non-terminal node, N, with daughters Ds, there is a rule in the grammar of the form: \( N \rightarrow Ds \).

Simple example

![Diagram of a tree structure](attachment:tree.png)
Theories of Linguistic Knowledge

- Theories of Syntax
  - **Representations**: Trees, feature structures, dependencies
  - **Structure building**: PS-rules, transformations, unification, composition, tree substitution
  - **Constraints on representations**: Case marking, theta-Criterion, c-command, binding principles, head-foot principle

- Competence Hypothesis
  - The mechanisms of language comprehension directly utilize the rules and representations of the linguistic theory

The Competence Hypothesis

- Knowledge: **Competence hypothesis**
  - Need to recover the meaning of sentences/utterances
  - Assumptions about (levels of) representations
    - Linguistic theory is isomorphic to human linguistic knowledge
    - Comprehension and production share same knowledge
  - **Weak competence**: people recover representations that are isomorphic to those of linguistic theories
  - **Strong competence**: people directly use the grammatical knowledge & principles of linguistic theories
The Modularity Issue

• Is language distinct from other cognitive & perceptual processes?
  • e.g. vision, smell, reasoning ...

• Do distinct modules exist within the language processor?
  • e.g. word segmentation, lexical access, syntax ...

• What is a module anyway!?

Architectures and Mechanisms

• What does “distinct” mean:
  • Representational autonomy: e.g. phonological versus logical representations
    • Possibly interactive processes
  • Procedural autonomy: e.g. lexical access versus parsing
    • Possibly shared representations

• How is the language module organized/interact with other systems?
  • Does architecture affect possible mechanisms?

  • What is the interface and bandwidth between modules?
Strong competence & modularity

• Fodor’s proposals emphasis language as a module, distinct from other perceptual cognitive abilities

• Linguistic theories suggest that language itself may consist of sub-levels: phonology, morphology, syntax, semantics ...

• Each with different rules and representations

• Do these correspond to distinct processes?

• Are these processes modules?

A Modular Architecture
Support for Linguistic Modularity

- Modular lexical access versus syntax: Forster
  - all possible word meanings temporarily available
  - no immediate influence of syntactic context

- Modular syntax versus semantics: Frazier
  - initial attachment ambiguities resolved by purely structural preferences
  - no immediate effect of semantics or context

- Dissociation in language impairment at different levels
  - lexical, syntactic, semantic; production versus comprehension

Attachment Preferences

```
NP  VP
  S
  NP
  VP
  S

The reporter
said
Hillary
will run
AdvP
last night
```
Against linguistic modularity

- Empirical evidence from on-line methods
  - “immediate” influence of animacy, frequency, plausibility, context …
    - The woman sent the flowers was pleased
    - The patient sent the flowers was pleased

- Appropriate computational frameworks:
  - symbolic constraint-satisfaction systems
  - connectionist systems & competitive activation models

- Homogenous/Integrative Linguistic Theory: HPSG
  - multiple levels of representation within a unified formalism

Human Language Processing

- We understand language incrementally, word-by-word
  - How do people construct interpretations?

- We must resolve local and global ambiguity
  - How do people resolve lexical ambiguity?
  - How do people decide upon a particular interpretation?

- Decisions are sometimes wrong!
  - What information is used to identify we made a mistake?
  - How do we find an alternative interpretation?

- Answers can reveal important details about the underlying mechanisms
Roadmap

• Theories of sentence processing:
  • modularity, parsing strategies, information sources, reanalysis

• Symbolic parsing models:
  • incremental parsing, ambiguity resolution, memory load, probabilistic models

• Rational, probabilistic parsing models:
  • Symbolic parsers augmented with probabilities, derived from experience

• Information theoretic approaches:
  • Modeling communication as a bounded rational probabilistic problem

Tutorials

• We'll be using various software packages and programs to make some of the concepts more concrete:
  • Prolog implementations of incremental parsing algorithms
  • Prolog implement of incremental HMM POS tagging
  • TnT statistical POS tagger
  • Roark’s incremental statistical parser

• For Wednesday: Install SWI-Prolog on your laptops, and bring them … if you can’t, then partner with someone.
  • http://www.swi-prolog.org

SWI Prolog
Course details

- Weekly lectures (Mon 2-4pm) and tutorials (Wed 2-4pm)
  - Participation in, and completion of, all tutorials is required!
- Assessment: Final Exam (100%), Date: Mon, February 3, 2020
  - All tutorial assignments must be successfully completed to sit the exam
- Course materials (overheads and most readings) will be made available on the course homepage (linked from general course page)
- Contact: please e-mail first – crocker@coli.uni-sb.de