Computational Psycholinguistics Lecture 5: Probabilistic Accounts

Matthew W. Crocker

Marty Mayberry

Experience & Multiple Constraints

- The previous accounts adopt purely syntactic mechanisms for disambiguation
  - Initial parsing decisions are guided by syntax & subcategorization alone
  - Assume a modular parser & the "primacy" of syntax
- Does our prior **experience** with language, determine our preferences for interpreting the sentences we hear?
  - Tuning hypothesis: disambiguate structure based on how it has been most frequently disambiguated in the past.
- To what extent do non-syntactic **constraints** such as semantics, intonation, and context influence our resolution of ambiguity?

Against linguistic modularity

- Empirical evidence from on-line methods
  - evidence for "immediate" (very early) interaction effects of animacy, lexical frequency, plausibility, discourse context …
    - *The woman/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

Multiple constraints

"The doctor told the woman that ...

story

diet was unhealthy

he was in love with her husband

he was in love with to leave

story was about to leave

Prosody: intonation can assist disambiguation

Lexical preference: *that* = {Comp, Det, RelPro}

Subcat: *told* = { [_ NP NP] [_ NP S] [_ NP S'] [_ NP Inf] }

Semantics: Referential context, plausibility

- Reference may determine "argument attach" over "modifier attach"
- Plausibility of *story* versus *diet* as indirect object

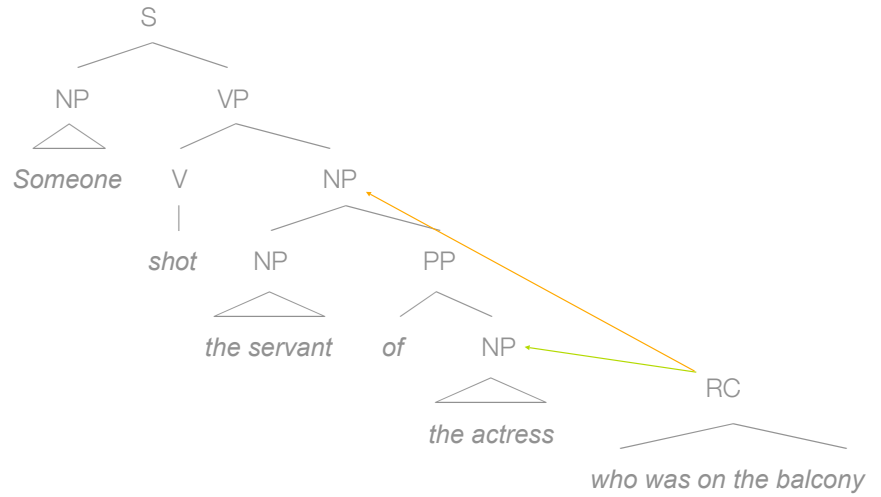
The Role of Experience

- Probabilistic models of sentence processing
  - Interactive, constraint-based accounts (connectionist)
  - Symbolic parsing models (statistical)
- Probabilistic Models: Breadth and Depth
  - SLCM: Maximal likelihood for category disambiguation
  - Statistical models of human parsing (Jurafsky)
  - Wide coverage probabilistic sentence processing (Crocker & Brants)
  - Criticisms of likelihood, and possible alternative: Informativity

Experienced-based Models

- Resolve ambiguities according to linguistic experience:
  - Lexical Guidance Hypothesis: (Ford et al)
    - Resolve subcategorisation ambiguities in favour of the most likely frame for the ambiguous verb
  - Linguistic Tuning Hypothesis: (Mitchell et al)
    - Resolve structural ambiguities according to the structure which has previously prevailed
- Relative clause attachment
  - Someone shot the servant of the actress who was on the balcony

Relative Clause Attachment



Cross-linguistic RC Preferences

Language	Off-line	On-line
Spanish	high	low
French	high	low
Italian	high	low
Dutch	high	
German	high	low(early), high(late)
English	low	low
Arabic	low	
Norwegian	low	
Swedish	low	
Romanian	low	

- Immediate low attachment, possibly revised quickly (even on-line) … seems the best account

Statistical Models of Language

- Statistics in linguistics [Abney, 1996]
  - Acquisition, change, and variation
  - Ambiguity and graded acceptability
  - Brings 'performance' back into linguistics
- Statistics in computational linguistics
  - Effective: accurate and robust
  - Eschews 'AI' problem
  - Trainable & efficient

Statistical Mechanisms

- Statistical information in the lexicon: frequencies or 'activations'
- Statistics in grammar and processing:
  - Association of grammatical knowledge with probabilistic weights
    - Could be used to model graded acceptability and/or disambiguation
  - Statistical processing mechanisms:
    - Sequences of parsing operations are probabilistic
  - Are complex structures associated with probabilities?
- Are statistics used "strategically" by the HSPM, or simply a product of the underlying architecture?

Motivating the Probabilistic HSPM

- Empirical: Evidence for the use of frequencies
  - Sense disambiguation [DM&R]
  - Category disambiguation [Corley&Crocker]
  - Subcategorization frame selection [TT&K, Garnsey]
  - Structural preferences [Mitchell et al]
- Rational: Near optimal heuristic behaviour
  - Select the “most likely” analysis
  - Ideal for modular architectures, where full knowledge isn’t available

Methodological

- Transparently combine symbolic and stochastic mechanisms
  - Associate probabilities with rules and representation
- Scaleable, predictive models
- Blurring the boundary between rational and empirical
  - Combines existing theories with mechanisms that learn from experience
  - Do probabilities encode “hidden” knowledge/representations?

Garden Path *vs.* Garden Variety

- Human Language Processing: **Garden Paths**
  - ✗ Incremental disambiguation process can fail
  - ✗ Memory limitations lead to breakdown
  - ✗ Garden paths lead to misinterpretations, complexity or breakdown
- Human Language Processing: **Garden Variety**
  - ✓ Accurate: typically recover the correct interpretation
  - ✓ Robust: are able to interpret ungrammatical & noisy input
  - ✓ Fast: people process utterances in real-time, incrementally

Is Understanding Rational?

- Hypothesis: In general people seem well-adapted for language.
  - Goal: Our models must account for, and explain:
    - Processing difficulty in specific circumstances
    - Effective performance in general
  - Method: Apply Rational Analysis
- Use probabilistic frameworks to reason about rational choice
  - Initial hypothesis: The optimal function is one which maximizes the likelihood of obtaining the correct interpretation of an utterance

Rational Analysis

- Hypothesis: People approach optimal language comprehension

  Rational Analysis: when a cognitive system is optimally adapted

  - Goals: Obtain the most likely interpretation
  - Environment: Input is incremental and ambiguous
  - Computational: Finiteness, 'foregrounded' interpretation

  Constructing a Rational Analysis:

  1. Derive the Optimal Function
  2. Test against the empirical data
  3. Revise the Optimal Function

Probabilistic Language Processing

- Task of comprehension: recover the correct interpretation
- Goal: Determine the most likely analysis for a given input:

$$\arg\max_i P(s_i) \text{ for all } s_i \in S$$

- *P* hides a multitude of sins:
  - P corresponds to the degree of belief in a particular interpretation
  - Influenced by recent utterances, experience, non-linguistic context
- P is usually determined by frequencies in corpora or completions
- To compare probabilities (of the S_i), we assume parallelism. How much?

Maximal Likelihood Models

- Language Technology: Broad coverage, high-accuracy parsing
  - Parse with the highest probability is usually correct
  - Also: speech recognition, POS tagging, semantic clustering, word sense
- Psycholinguistic evidence for the use of frequencies
  - Category disambiguation, word sense, subcategorization frame selection, structural preferences
- Psycholinguistic Models:
  - Constraint-based and connectionist (Tanenhaus, Macdonald, ...)
  - Jurafsky: probabilistic lexical access and disambiguation

Implementation

- Interpretation of probabilities
  - Likelihood of structure occurring, P can be determined by frequencies in corpora or human completions
- Estimation of probabilities
  - Infinite structural possibilities = sparse data
  - Associate probabilities with grammar (finite): e.g. PCFGs
- What mechanisms are required:
  - Incremental structure building and estimation of probabilities
  - Comparison of probabilities entails parallelism

The Grain Problem

- Experience-based models rely on frequency of prior linguistic exposure to determine preferences. What kinds of things do we count?
  - Actual sentence/structure occurences? Data too sparse?
  - Lexical: Verb subcategorization frequencies. Do we distinguish tenses? Senses?
    - Word level: specific word forms or lemmas? Part-of-speech, how detailed?
  - Tuning is structural: NP P NP RC (High) *vs* NP P NP RC (Low)
- Does all experience have equal weight (old vs. new)?
- Are more frequent “words” or “strings” (idioms) dealt with using finer grain statistics than rarer expressions?