

Computational psycholinguistics

Roger Levy, Kinton Bicknell, and Nathaniel Smith

University of California, San Diego

1. Brief Description

Over the last two decades, computational linguistics has been revolutionized as a result of three closely related developments, two empirical and one theoretical: increases in computing power, the new availability of large linguistic datasets, and a paradigm shift toward the view that language processing by computers is best approached through the tools of statistical inference. During roughly the same time frame, there have been similar theoretical developments in cognitive psychology towards a view of major aspects of human cognition as instances of rational statistical inference. Developments in these two fields have set the stage for renewed interest in computational approaches to human language processing. In this tutorial, we briefly survey some of the key theoretical issues at the forefront of this interdisciplinary field today, and show how modeling techniques from NLP are being employed, extended, and coupled with experimental techniques from psycholinguistics to further our understanding of real-time human language use.

2. Tutorial structure

1. Introduction & summary of current state of key areas in psycholinguistics

(a) Key empirical findings involving ambiguity resolution, prediction, integration of diverse information sources, and speaker choice in real-time language comprehension and production (~10 minutes)

(b) Framing of empirical findings and concomitant theoretical issues in terms that can be cleanly related to leading NLP models and algorithms (~10 minutes)

2. Review of exact inference techniques for stochastic grammatical formalisms

(a) Weighted finite-state automata and context-free grammars (~10 minutes)

(b) Probabilistic Earley algorithm (~10 minutes)

(c) Weighted intersection of FSA and CFG (~10 minutes)

3. Modeling key results in ambiguity resolution and expectation-based facilitation

(a) Global disambiguation preferences (~5 minutes)

(b) Measuring online processing difficulty: intro to self-paced reading (~5 minutes)

(c) Expectation-based facilitation in unambiguous contexts (~5–10 minutes)

4. Coffee break

5. Online production: speaker choice

- (a) Zipf's second law (frequency & word length) and information-theoretic interpretations (~10 minutes)
- (b) Phonetic duration & reduction effects in online word production (~5 minutes)
- (c) Morphological- and lexical-level reduction phenomena: modeling and empirical investigation (~10 minutes)

6. Cognitive limitations and implications for modeling

- (a) Memory limitations & garden pathing: empirical results (~10 minutes)
- (b) Modeling approach I: incremental beam search & garden pathing (~5 minutes)
- (c) Modeling approach II: stochastic incremental search & "digging-in" effects (~10 minutes)

7. Additional theoretical challenges

- (a) Bounds on rationality in real-time language use? "Good-enough" comprehension effects and "local-coherence" effects (~10 minutes)
- (b) Possible avenues of attack: more refined models introducing input uncertainty (~10 minutes)
- (c) More sophisticated experimental tools: eye-tracking (~5 minutes)
- (d) New experimental findings on input uncertainty, "hallucinated" garden paths (~5 minutes)
- (e) Future directions (~5 minutes)

8. Summary and questions (~5–10 minutes)

3. Instructor

Roger Levy, rlevy@ling.ucsd.edu

My research focuses on theoretical and applied questions in the processing of natural language. Inherently, linguistic communication involves the resolution of uncertainty over a potentially unbounded set of possible signals and meanings. How can a fixed set of knowledge and resources be deployed to manage this uncertainty? To address these questions I use a combination of computational modelling and psycholinguistic experimentation. This work furthers our understanding of the cognitive underpinning of language processing, and helps us design models and algorithms that will allow machines to process human language.

Klinton Bicknell and Nathaniel Smith are PhD students at the University of California, San Diego.