A Distributed, Large Scale Connectionist Model of the Interaction of Lexical and Semantic Constraints

Michael W. Harm†, Robert Thornton‡ and Maryellen C. MacDonald‡

†CNBC, Carnegie Mellon University
‡Neuroscience Program, University of Southern California
Constraint Based Theories of Sentence Processing

- Focus on the integration of multiple soft constraints:
  - Syntactic / Structural
  - Pragmatic
  - Lexical frequency biases
Problem: How are Constraints Weighed?

- Claim is that multiple constraints matter, but data suggest that some constraints are weighed more heavily than others.

- “Grain” problem: infinite number of possible statistics could be calculated.

- Challenge is to specify computational mechanism underlying constraint integration.
Need for Computational Models

- Specify precise weighing of constraints according to:
  - Informativeness of cue
  - Prior world knowledge about relationships of objects and events, etc.
  - Current state of the system: what is and isn’t known from previous words
• Prior computational models (e.g., Christiansen & Chater, 1999; Elman, 1990; Tabor et al., 1997) examined interactions of lexical and contingent frequency constraints, but did not include semantics.

• Semantics crucial to grain problem and constraint based theories in general
Goals of Current Work

- Recast grain problem
  - Not “Which grain level does the system compute statistics over,“
  - But rather: “What factors affect the relative influence of simple and contingent statistical regularities?”

- Develop a computational model with distributed semantic representations to investigate complex constraint interactions
Semantic Specificity and Statistical Regularity

- Distributional regularities over multiple levels of semantic specificity:
  - Structural: regularities over coarse-grained semantics (i.e., object, action) correspond to syntactic categories
  - Pragmatic: regularities over mid-level semantics yield pragmatic constraints. For example, animate objects participate in mental events, etc.
  - Lexical: statistics calculated over fine-grained semantics yield word-specific constraints
Specificity and Regularity (cont.)

- **Action Entity Determiner**
- **Mental Physical Life Form**
- **Object**
- **Mental Physical Construct Violent**
- **Definite Indefinite**

Bias Against

Bias For

"Structural"

"Pragmatic"

"Lexical"
Example: Syntactic Category Ambiguities

- Good domain to look for local statistical influences
- Rich interactions of lexical and pragmatic biases
  - Corporation fires (us/are)
  - Warehouse fires (us/are)
- Words have lexical frequency biases for head noun versus modifying noun status
- Pragmatics: warehouses burn, corporations fire people
- MacDonald (1993) manipulated frequency bias of first word (bias for modifier or head noun), frequency of co-occurrence of words, and semantic plausibility
- Found effects of all factors. Evidence for online use of multiple sources of information in disambiguating
Simulating Syntactic Category Ambiguities

- Distributed Semantic Representations
  - WordNet online semantic database (Miller, 1990)
  - 8,207 hierarchically structured features, including:
    → High-level semantic (entity, modifier, state)
    → Mid-level category (living thing, tool, plant, human)
    → Lexical, item specific
    → Morphological (plural, past tense)

- Distributed word form representations
  - Localist syllables
  - Morphological affixes
  - 8,210 word forms, 3,421 features
Task: Compute Semantics of Current Word
Training Set

- Large training set: 20,000 Word triplets from tagged WSJ and Brown corpora
- Tokens fitting NOUN-NOUN-XXX or NOUN-VERB-XXX template
- Broad range of ambiguous items and phrases:
  - Modifying noun vs. head noun biases ranged from heavily biased to equibiased
    (e.g., *Tax* is very heavily biased to be a modifier. *Tariff* is weak. Semantics are the same.)
  - Semantic biases ranged from heavy to weak
    (e.g., *<mental actions>* require an *<living>* noun. *<physical actions>* can be performed by *<living>* or *<object>* nouns.)
Training Regime

- Word forms presented one at a time
- Target is semantics for current word
- Interpretation nodes: target is N/N or N/V over whole duration of triplet
- Activity builds up gradually over time using continuous time backprop: 42 samples, 10 units of whole time, one word every 14 samples. Under time pressure
- Trained for 400,000 presentations of randomly selected triplets
Scoring the Model

- Measured accuracy of semantic features after each word, and error on interpretation nodes at end of triplet
- Mean total semantic error: 0.66 per word
- 99.99% of features correct
- Mean error on interpretation nodes: 0.03
Behavior of the Model

- Structural constraints: *man dog bites* not coherently assimilated - interpretation nodes unstable. Novel (though perhaps implausible) grammatical phrases (e.g., *truth bites man*) generally are stable.

- Pragmatics: *dog bites man* recognized more quickly than *truth bites man*.

- Lexical: *tax cuts are* recognized more quickly than *tariff cuts are*. 
Replication of MacDonald (1993)

- 2x2 design: selected 24 items, crossed bias of first word (to modifying noun vs. head noun) with combinatorial semantic constraint

- Identified combinatorial semantic constraint by computing conditional entropy of semantic features: degree to which features of both words constrain features of second word (e.g., abstraction -> communication, as in injury claims (N/N bias) versus company claims (N/V bias))
• Dependent variable: integrated difference in activity of interpretation features for N/N and N/V over time.

• Results: Reliable effect of lexical frequency \((p < 0.003)\) and plausibility \((p < 0.007)\)
Conclusions

- Modeling pragmatic effects in empirical studies requires rich semantics

- World knowledge approximated with rich semantics and large training set of real-world utterances

- Potential to look at much more complex phenomena involving deeper world knowledge, e.g., conceptual combination, more complex ambiguities

- Framework allows examination of effects of multiple cues computed at multiple levels. Recasts “grain problem.”
References


