Modeling ... on-line sentence comprehension

Matthew W. Crocker
Computational Linguistics
Saarland University
Modeling ... on-line sentence comprehension

Matthew W. Crocker
Computational Linguistics
Saarland University
Area of Research
Area of Research

- The processes that underlie the human capacity to understand language
Area of Research

- The processes that underlie the human capacity to understand language
- How does the human language processor work?
  - Architectures, mechanisms
Area of Research

- The processes that underlie the human capacity to understand language
- How does the human language processor work?
  - Architectures, mechanisms
- How can we model it computationally?
  - Understanding: *Competence*
  - Behaviour: *Performance*
Area of Research

- The processes that underlie the human capacity to understand language
- How does the human language processor work?
  - Architectures, mechanisms
- How can we model it computationally?
  - Understanding: *Competence*
  - Behaviour: *Performance*
- Interaction of language with other cognitive systems and the environment
Human Sentence Comprehension
Human Sentence Comprehension

A Wide Coverage Model of Semantic Plausibility
Human Sentence Comprehension

A Wide Coverage Model of Semantic Plausibility

Visually Situated Comprehension & Attention
Human Sentence Comprehension

A Wide Coverage Model of Semantic Plausibility

Visually Situated Comprehension & Attention

Outline
Traditional Approaches
Traditional Approaches

- Experimental research:
  - Reading: self-paced and eye-tracking paradigms
  - Measure: reading times = processing complexity
Traditional Approaches

- Experimental research:
  - Reading: self-paced and eye-tracking paradigms
  - Measure: reading times = processing complexity

- Psycholinguistic theories:
  - Emphasis on linguistic processing (lexical, syntactic)
  - Theories strive to explain processing complexity
Traditional Approaches

- **Experimental research:**
  - Reading: self-paced and eye-tracking paradigms
  - Measure: reading times = processing complexity

- **Psycholinguistic theories:**
  - Emphasis on linguistic processing (lexical, syntactic)
  - Theories strive to explain processing complexity

- Emphasis on the weaknesses of human comprehension
Traditional Approaches

- **Experimental research:**
  - Reading: self-paced and eye-tracking paradigms
  - Measure: reading times = processing complexity

- **Psycholinguistic theories:**
  - Emphasis on linguistic processing (lexical, syntactic)
  - Theories strive to explain processing complexity
  - Emphasis on the weaknesses of human comprehension
  - Failure to situate the human language processor
Human Language Processing
Performance Paradox: “How is it people understand language so accurately and effortlessly given it’s complexity and ambiguity?”

- We understand language incrementally, word-by-word, in real-time.
- Sometimes we even anticipate what’s coming next.
- We rapidly resolve ambiguity and revised misinterpretations.
Performance Paradox: “How is it people understand language so accurately and effortlessly given it’s complexity and ambiguity?”

- We understand language incrementally, word-by-word, in real-time.
- Sometimes we even anticipate what’s coming next.
- We rapidly resolve ambiguity and revised misinterpretations.

(Partial) Solution:

- We optimize based on our long-term linguistic experience and knowledge.
- We adaptively exploit our immediate (non-)linguistic context.
Sentence Processing Models
Some challenges
Some challenges

Coverage

- Account for *garden path* and *garden variety language*
Some challenges

- Coverage
  - Account for garden path and garden variety language

- Semantics
  - Plausibility, selectional restrictions, priming, associates, inferences
Some challenges

- Coverage
  - Account for \textit{garden path} and \textit{garden variety language}

- Semantics
  - Plausibility, selectional restrictions, priming, associates, inferences

- Linguistic and non-linguistic context
  - Exploiting common ground, visual environment, etc.
Some challenges

- Coverage
  - Account for *garden path* and *garden variety language*

- Semantics
  - Plausibility, selectional restrictions, priming, associates, inferences
  - Linguistic and non-linguistic context
    - Exploiting common ground, visual environment, etc.

- Linking hypotheses
  - Quantitative predictions of multiple measures
Probabilistic sentence processing
Probabilistic sentence processing

Empirical: lots of evidence for the importance of *frequency* in comprehension

- lexical, subcategorization, structural preferences
Probabilistic sentence processing

- Empirical: lots of evidence for the importance of frequency in comprehension
  - lexical, subcategorization, structural preferences
- Provides a framework for expressing preferences & biases
  - (including those where we don’t have a clue as to their proper treatment)
Probabilistic sentence processing

- Empirical: lots of evidence for the importance of **frequency** in comprehension
  - lexical, subcategorization, structural preferences
- Provides a framework for expressing **preferences & biases**
  - (including those where we don’t have a clue as to their proper treatment)
- Good techniques for **training**, possible to extend **coverage**
Probabilistic sentence processing

- Empirical: lots of evidence for the important of **frequency** in comprehension
  - lexical, subcategorization, structural preferences
- Provides a framework for expressing **preferences & biases**
  - (including those where we don’t have a clue as to their proper treatment)
- Good techniques for **training**, possible to extend **coverage**
- Unified explanation of **ambiguity resolution**
Probabilistic sentence processing

- Empirical: lots of evidence for the important of **frequency** in comprehension
  - lexical, subcategorization, structural preferences
- Provides a framework for expressing **preferences & biases**
  - (including those where we don’t have a clue as to their proper treatment)
- Good techniques for **training**, possible to extend **coverage**
- Unified explanation of **ambiguity resolution**
- Interesting linking hypotheses
  - beam search, **rank/re-rank**, **surprisal** and entropy reduction
Goal: Optimize accurate incremental interpretation
**Goal:** Optimize accurate incremental interpretation

**Function:** Adopt the most likely interpretation:

$$\arg\max_{s_i} P(s_i) \text{ for all } s_i \in S$$
**Goal:** Optimize accurate incremental interpretation

**Realization:**
Likelihood $\approx$ Experience $\approx$ Corpora

**Function:** Adopt the most likely interpretation:
$$\arg\max_{s_i} P(s_i) \text{ for all } s_i \in S$$
**Goal:** Optimize accurate incremental interpretation

**Realization:**
Likelihood $\approx$ Experience $\approx$ Corpora

**Function:** Adopt the most likely interpretation:
$$\arg\max_i P(s_i) \text{ for all } s_i \in S$$
**Goal:** Optimize accurate incremental interpretation

**Realization:**
Likelihood $\approx$ Experience $\approx$ Corpora

**Function:** Adopt the most likely interpretation:
$$\arg\max_i P(s_i) \text{ for all } s_i \in S$$

**Lexical Category Disambiguation**

Optimally assign syntactic parts-of-speech to each word in the sentence:
$$P(t_0, \ldots, t_n, w_0, \ldots, w_n) \approx \prod_{i=1}^{n} P(w_i \mid t_i) P(t_i \mid t_{i-1})$$
- Trained on Suzanne (120K) & BNC (10M)
- High accuracy (> 90%)
- Explained a broad range of data

Corley & Crocker, 2000
Crocker & Corley, 2002
Goal: Optimize accurate incremental interpretation

Realization:
Likelihood ≈ Experience ≈ Corpora

Function: Adopt the most likely interpretation:
\[
\arg\max_i P(s_i) \text{ for all } s_i \in S
\]

Lexical Category Disambiguation

Optimally assign syntactic parts-of-speech to each word in the sentence:
\[
P(t_0, \ldots, t_n, w_0, \ldots w_n) \approx \prod_{i=1}^{n} P(w_i | t_i) P(t_i | t_{i-1})
\]
- Trained on Suzanne (120K) & BNC (10M)
- High accuracy (> 90%)
- Explained a broad range of data

Wide-coverage probabilistic parsing

Incrementally assign the most likely syntactic structure as each word \(i\) is encountered:
\[
\hat{t}_{1..i} = \arg\max_j P(s_{1..i, j}) \text{ for all } s_{1..i, j} \in TREES
\]
- Approximated using PCFG & ICMM
- Trained on Penn TreeBank & Negra corpora
- Wide coverage & models parse preferences
Reduced Relative Clause
Reduced Relative Clause

“The man raced to the station was innocent”
Reduced Relative Clause

"The man raced to the station was innocent"
Reduced Relative Clause

“The man raced to the station was innocent”
Reduced Relative Clause

“The man raced to the station was innocent”
Reduced Relative Clause

“The man raced to the station was innocent”
Reduced Relative Clause
Reduced Relative Clause

“The man held at the station was innocent”
Reduced Relative Clause

“The man held at the station was innocent”
Reduced Relative Clause

“The man held at the station was innocent”
Reduced Relative Clause

“The man held at the station was innocent”
Reduced Relative Clause

“The man held at the station was innocent”
Reduced Relative Clause

“The man held at the station was innocent”
Human Sentence Comprehension

A Wide Coverage Model of Semantic Plausibility

Visually Situated Comprehension & Attention

Outline
Outline

Human Sentence Comprehension

A Wide Coverage Model of Semantic Plausibility

Visually Situated Comprehension & Attention
Using Plausibility

The doctor cured...
Using Plausibility

The doctor cured ...

The doctor cured by the ...
Using Plausibility

The doctor cured the patient.

The doctor cured the patient by the...
Using Plausibility

The doctor cured the patient.

The patient cured the doctor.
Using Plausibility

The doctor cured the ... patient.

Arg: patient
GF: Subject
Sense: heal
Role: Agent
Modeling Plausibility
Modeling Plausibility

- Goal 1: a graded model of verb-argument plausibility
  - automatically trained from a semantically annotated corpus
Modeling Plausibility

- **Goal 1**: a graded model of verb-argument plausibility
  - automatically trained from a semantically annotated corpus

- **Evaluation of the semantic model**:
  - model human plausibility judgments and selectional preferences
Modeling Plausibility

Goal 1: a graded model of verb-argument plausibility
- automatically trained from a semantically annotated corpus

Evaluation of the semantic model:
- model human plausibility judgments and selectional preferences

Goal 2: Integration with a parsing account:
- use semantic model during parsing; predict reading complexity
Modeling Plausibility

- Goal 1: a graded model of verb-argument plausibility
  - automatically trained from a semantically annotated corpus

- Evaluation of the semantic model:
  - model human plausibility judgments and selectional preferences

- Goal 2: Integration with a parsing account:
  - use semantic model during parsing; predict reading complexity

- Evaluation of the integrated parsing model
  - correlate with RT data from several 8 studies (4 phenomena)
The Semantic Model
The Semantic Model

Let plausibility be the likelihood of a given verb, argument, role:

\[
Plaus = P(r,a,v,c,gf) = P(v) \cdot P(c \mid v) \cdot P(gf \mid v,c) \cdot P(r \mid v,c,gf) \cdot P(a \mid v,c,gf,r)
\]

role, argument head, verb, verb sense, grammatical function
Let plausibility be the likelihood of a given verb, argument, role:

\[
Plaus = P(r,a,v,c,\text{gf}) = P(v) \cdot P(c|v) \cdot P(\text{gf}|v,c) \cdot P(r|v,c,\text{gf}) \cdot P(a|v,c,\text{gf},r)
\]

role, argument head, verb, verb sense, grammatical function

Smoothing:

- Good-Turing smoothing for 1st 4 terms
- Class-based smoothing for the final term
- generalize from word tokens to word classes
Training + Smoothing
Training + Smoothing

<table>
<thead>
<tr>
<th>Training Method</th>
<th>PropBank/FrameNet</th>
<th>Sentence Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PropBank</td>
<td>cure.1</td>
<td>[The doctor $Arg0$] cured [the patient $Arg1$]</td>
</tr>
<tr>
<td>FrameNet</td>
<td>Healing</td>
<td>[The doctor $Healer$] cured [the patient $Patient$]</td>
</tr>
</tbody>
</table>
Training + Smoothing

<table>
<thead>
<tr>
<th>PropBank</th>
<th>cure.1</th>
<th>[The doctor Arg0] cured [the patient Arg1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>FrameNet</td>
<td>Healing</td>
<td>[The doctor Healer] cured [the patient Patient]</td>
</tr>
</tbody>
</table>

- Semantically annotated corpora are too sparse
  - unseen <verb,role,arg> triples for $P(a|v,c,gf,r)$

Training + Smoothing

<table>
<thead>
<tr>
<th>PropBank</th>
<th>cure.1</th>
<th>[The doctor Arg0] cured [the patient Arg1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>FrameNet</td>
<td>Healing</td>
<td>[The doctor Healer] cured [the patient Patient]</td>
</tr>
</tbody>
</table>

- Semantically annotated corpora are too sparse
  - unseen <verb,role,arg> triples for \( P(a|v,c,gf,r) \)
- Class based smoothing:
  - Nouns: WordNet's lowest level (synset) ontology
  - Verbs: ID/IB soft clustering algorithm
  - FrameNet: 57K propositions, 2K verbs

Semantically annotated corpora are too sparse

unseen \langle \text{verb}, \text{role}, \text{arg} \rangle \text{ triples for } P(a|v,c,\text{gf},r)

Class based smoothing:

Nouns: WordNet's lowest level (synset) ontology

Verbs: ID/IB soft clustering algorithm

FrameNet: 57K propositions, 2K verbs

PropBank | cure.1 | [The doctor Arg0] cured [the patient Arg1]
FrameNet  | Healing | [The doctor Healer] cured [the patient Patient]

Verb Classes
Verb Classes

Aim: Induce classes of semantically similar verbs

Verb Classes

- Aim: Induce classes of semantically similar verbs
- Corpora: PropBank & FrameNet
Verb Classes

- Aim: Induce classes of semantically similar verbs
- Corpora: PropBank & FrameNet
- Features: for each verb-argument occurrence
  - arg head, role label, verb sense, syntactic path, path-role

Verb Classes

- **Aim**: Induce classes of semantically similar verbs
- **Corpora**: PropBank & FrameNet
- **Features**: for each verb-argument occurrence
  - arg head, role label, verb sense, syntactic path, path-role
- **Soft-clustering**: Information Bottleneck / Information Distortion
  - soft-clustering, probabilistic class membership

Verb Classes

- Aim: Induce classes of semantically similar verbs
- Corpora: PropBank & FrameNet
- Features: for each verb-argument occurrence
  - arg head, role label, verb sense, syntactic path, path-role
- Soft-clustering: Information Bottleneck / Information Distortion
  - soft-clustering, probabilistic class membership
- FN: 57K propositions, 2K verbs
  - 13 clusters, similar size, clear semantic similarity

Verb Classes

- **Aim:** Induce classes of semantically similar verbs
- **Corpora:** PropBank & FrameNet
- **Features:** for each verb-argument occurrence
  - arg head, role label, verb sense, syntactic path, path-role
- **Soft-clustering:** Information Bottleneck
  - soft-clustering, probabilistic class membership
- **FN:** 57K propositions, 2K verbs
  - 13 clusters, similar size, clear semantic similarity

<table>
<thead>
<tr>
<th>Move</th>
<th>Communicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>cycle</td>
<td>tell</td>
</tr>
<tr>
<td>follow</td>
<td>advise</td>
</tr>
<tr>
<td>travel</td>
<td>confide</td>
</tr>
<tr>
<td>lead</td>
<td>inform</td>
</tr>
<tr>
<td>chase</td>
<td>urge</td>
</tr>
<tr>
<td>accompany</td>
<td>confess</td>
</tr>
<tr>
<td>escort</td>
<td>write</td>
</tr>
<tr>
<td>commute</td>
<td>address</td>
</tr>
<tr>
<td>trail</td>
<td>ask</td>
</tr>
<tr>
<td>usher</td>
<td>request</td>
</tr>
</tbody>
</table>

Predicting Human Judgments

- *McRae et al, 1998: 100 data points
good & bad fillers for agent & patient
  - No seen triples

- *Own study: 18 verbs from PB & FN
  - 6 exp-thm, 6 ag-rec, 6 ag-pat
  - 3 frequent fillers for each role
  - 25% seen triples
Predicting Human Judgments

Human judgments

- McRae et al, 1998: 100 data points
good & bad fillers for agent & patient
  - No seen triples

- Own study: 18 verbs from PB & FN
  - 6 exp-thm, 6 ag-rec, 6 ag-pat
  - 3 frequent fillers for each role
  - 25% seen triples

<table>
<thead>
<tr>
<th>Verb</th>
<th>Argument</th>
<th>Role</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>cure</td>
<td>doctor</td>
<td>healer</td>
<td>6.3</td>
</tr>
<tr>
<td>cure</td>
<td>doctor</td>
<td>patient</td>
<td>3.8</td>
</tr>
<tr>
<td>cure</td>
<td>doctor</td>
<td>healer</td>
<td>1.4</td>
</tr>
<tr>
<td>cure</td>
<td>doctor</td>
<td>patient</td>
<td>6.1</td>
</tr>
</tbody>
</table>

(McRae et al, JML, 1998)
Predicting Human Judgments

Human judgments

- McRae et al, 1998: 100 data points
  - good & bad fillers for agent & patient
  - No seen triples

- Own study: 18 verbs from PB & FN
  - 6 exp-thm, 6 ag-rec, 6 ag-pat
  - 3 frequent fillers for each role
  - 25% seen triples

<table>
<thead>
<tr>
<th>Verb</th>
<th>Argument</th>
<th>Role</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>cure</td>
<td>doctor</td>
<td>healer</td>
<td>6.3</td>
</tr>
<tr>
<td>cure</td>
<td>doctor</td>
<td>patient</td>
<td>3.8</td>
</tr>
<tr>
<td>cure</td>
<td>doctor</td>
<td>healer</td>
<td>1.4</td>
</tr>
<tr>
<td>cure</td>
<td>doctor</td>
<td>patient</td>
<td>6.1</td>
</tr>
</tbody>
</table>

(McRae et al, JML, 1998)

Predicting Judgments:

<table>
<thead>
<tr>
<th>Judgements</th>
<th>N</th>
<th>Coverage</th>
<th>Correlation r</th>
</tr>
</thead>
<tbody>
<tr>
<td>McRae</td>
<td>64</td>
<td>87.5%</td>
<td>0.415*</td>
</tr>
<tr>
<td>Pado</td>
<td>414</td>
<td>98.5%</td>
<td>0.522***</td>
</tr>
</tbody>
</table>

Predicting Human Judgments

- **Human judgments**
  - McRae et al, 1998: 100 data points
    - good & bad fillers for agent & patient
  - No seen triples
  - Own study: 18 verbs from PB & FN
    - 6 exp-thm, 6 ag-rec, 6 ag-pat
    - 3 frequent fillers for each role
    - 25% seen triples

<table>
<thead>
<tr>
<th>Verb</th>
<th>Argument</th>
<th>Role</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>cure</td>
<td>cure</td>
<td>doctor</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>healer</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>patient</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.1</td>
</tr>
</tbody>
</table>

- Compared with role-labelers (Moschitti) on labeling task:
  - Semantic model performs better due to semantic smoothing, and less reliance on syntactic features

- Compared to selectional preference models:

<table>
<thead>
<tr>
<th>Train</th>
<th>Test</th>
<th>Model</th>
<th>Coverage</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FN</td>
<td>McRae</td>
<td>Plaus. Model</td>
<td>88%</td>
<td>0.415, **</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Resnik</td>
<td>81%</td>
<td>0.025, ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clark&amp;Weir</td>
<td>81%</td>
<td>-0.038, ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Li&amp;Abe</td>
<td>88%</td>
<td>-0.056, ns</td>
</tr>
<tr>
<td>Padó</td>
<td></td>
<td>Plaus. Model</td>
<td>97%</td>
<td>0.515, ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Resnik</td>
<td>93%</td>
<td>0.031, ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clark&amp;Weir</td>
<td>93%</td>
<td>0.165, **</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Li&amp;Abe</td>
<td>97%</td>
<td>0.112, *</td>
</tr>
</tbody>
</table>

ns: not significant, *: p<0.05, **: p<0.01.

- Correlation $r$
  - 0.415*
  - 0.522***

Predicting Human Judgments

Human judgments

- *McRae et al, 1998*: 100 data points
good & bad fillers for agent & patient
  - No seen triples

- *Own study*: 18 verbs from PB & FN
  - 6 exp-thm, 6 ag-rec, 6 ag-pat
  - 3 frequent fillers for each role
  - 25% seen triples

Predicting Judgments:

<table>
<thead>
<tr>
<th>Judgements</th>
<th>N</th>
<th>Coverage</th>
<th>Correlation r</th>
</tr>
</thead>
<tbody>
<tr>
<td>McRae</td>
<td>64</td>
<td>87.5%</td>
<td>0.415*</td>
</tr>
<tr>
<td>Pado</td>
<td>414</td>
<td>98.5%</td>
<td>0.522***</td>
</tr>
</tbody>
</table>


SynSem Integration Model

Semantic Model

Semantic Ranking:
1.
2.
...

Syntactic Model

Syntactic Ranking:
1.
2.
...

Global Ranking:
1.
2.
...

Difficulty Prediction

SynSem Integration Model

- **Syntactic Model:**
  - Incremental probabilistic parser [Roark 2001]
  - Head-lexicalised

SynSem Integration Model

- **Syntactic Model:**
  - Incremental probabilistic parser [Roark 2001]
  - Head-lexicalised

- **Semantic Model**
SynSem Integration Model

- **Syntactic Model:**
  - Incremental probabilistic parser [Roark 2001]
  - Head-lexicalised

- **Semantic Model**

- **Parameters:**
  - Interpolation factor to determine global ranking
  - Selection of cost functions: Conflict and Revision

“The critic wrote the painting had been ...”
“The critic wrote the painting had been ...”
“The critic wrote the painting had been ...”

**Conflict:** cost of competing constraints in ambiguous region

\[ \text{Cost}_{\text{conflict}} = \text{abs} ( \text{rank}_{\text{syn}}(gp) - \text{rank}_{\text{sem}}(gp) ) \]
Conflict: cost of competing constraints in ambiguous region

\[ \text{Cost}_{\text{conflict}} = \text{abs}( \text{rank}_{\text{syn}}(gp) - \text{rank}_{\text{sem}}(gp) ) \]

Revision: cost of disambiguation

\[ \text{Cost}_{\text{revision}} = 1 \text{ iff semantic change non-monotonically, and} \]
\[ \text{the revised semantics has lower probability}, \]
\[ = 0 \text{ otherwise} \]
Evaluation NP/S
Good Object: The critic \textit{wrote/argued} the \textbf{book} had been ...
Evaluation NP/S

- Good Object: The critic *wrote/argued* the **book** had been ...

- Bad Object: The critic *wrote/argued* the **painting** had been ...

Evaluation NP/S

**Good Object:** The critic *wrote/argued* the **book** had been ... 

**Bad Object:** The critic *wrote/argued* the **painting** had been ... 

Evaluation NP/S

- Good Object: The critic wrote/argued the book had been ...
- Bad Object: The critic wrote/argued the painting had been ...

Garnsey et al., 1997.
Evaluation NP/S

- **Good Object**: The critic *wrote/argued* the *book* had been ...
- **Bad Object**: The critic *wrote/argued* the *painting* had been ...

Evaluation: Combined
Evaluation: Combined

- Pooled data from two NP/S studies
Evaluation: Combined

- Pooled data from two NP/S studies
- Pooled data from eight studies
- Two studies each for NP/S, NP/0, MC/RR, PP Attachment

Evaluation: Combined

- Pooled data from two NP/S studies
- Pooled data from eight studies
  - Two studies each for NP/S, NP/0, MC/RR, PP Attachment

<table>
<thead>
<tr>
<th>Data</th>
<th>N</th>
<th>Pearson’s r</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP/S</td>
<td>12</td>
<td>$r = 0.688, p&lt;0.05$</td>
</tr>
<tr>
<td>8 Studies</td>
<td>36</td>
<td>$r = 0.700, p&lt;0.001$</td>
</tr>
</tbody>
</table>

Interim Summary
Interim Summary

- **Probabilistic models of plausibility** provide a good account of human judgment findings
  - distinguishing both good and bad role fillers
Interim Summary

- **Probabilistic models of plausibility** provide a good account of human judgment findings
  - distinguishing both good and bad role fillers
- Wide coverage and good performance in general:
  - rational and experience-based
  - lexical category assignment, syntactic analysis and disambiguation
  - thematic role assignment and plausibility
Interim Summary

- **Probabilistic models of plausibility** provide a good account of human judgment findings
  - distinguishing both good and bad role fillers
- Wide coverage and good performance in general:
  - rational and experience-based
  - lexical category assignment, syntactic analysis and disambiguation
  - **thematic role assignment and plausibility**
- Good account of garden variety and garden path phenomena!
Outline

Human Sentence Comprehension

A Wide Coverage Model of Semantic Plausibility

Visually Situated Comprehension & Attention
Outline

Human Sentence Comprehension

A Wide Coverage Model of Semantic Plausibility

Visually Situated Comprehension & Attention
Situated Language Processing
Situated Language Processing

Psycholinguistic experiments reveal the close temporal interaction of language understanding and visual attention mechanisms
Psycholinguistic experiments reveal the close temporal interaction of language understanding and visual attention mechanisms:

- Utterance-mediated eye movements in related scenes:
  - Referential expressions trigger looks to scene entities
  - Compositional interpretation triggers anticipatory looks to role fillers
    - Reflects use of syntactic and semantic constraints
  - Intonation can influence both of the above, during understanding
  - Influence of scene information on spoken comprehension
German 101
German 101

- Subject Verb Object (SVO) and Object Verb Subject (OVS)
German 101

- **Subject Verb Object (SVO) and Object Verb Subject (OVS)**

- Case-marking reflects grammatical function:
  - “Der Hase” (rabbit): Nominative/Subject
  - “Den Hasen” (rabbit): Accusative/Object
  - “Die Prinzessin” (princess): Nominative or Accusative
German 101

- **Subject Verb Object (SVO)** and **Object Verb Subject (OVS)**

- Case-marking reflects grammatical function:
  - “*Der Hase*” (rabbit): **Nominative/Subject**
  - “*Den Hasen*” (rabbit): **Accusative/Object**
  - “*Die Prinzessin*” (princess): **Nominative** or **Accusative**

- Preferred SVO word order, OVS is marked:
  - **SVO**: “*Die Prinzessin sah den Hasen*” **easy**
  - **OVS**: “*Die Prinzessin sah der Hase*” **difficult**
Anticipation in Visual Worlds
Anticipation in Visual Worlds
Anticipation in Visual Worlds

Anticipation in Visual Worlds

**SVO:** Der Hase frisst gleich den Kohl
“The rabbit eats soon the cabbage”

**OVS:** Den Hasen frisst gleich der Fuchs
“The rabbit is eaten soon by the fox”

Anticipation in Visual Worlds

SVO: Der Hase frisst gleich den Kohl
"The rabbit eats soon the cabbage"

OVS: Den Hasen frisst gleich der Fuchs
"The rabbit is eaten soon by the fox"
Anticipation in Visual Worlds

**SVO:** Der Hase frisst gleich den Kohl
"The rabbit eats soon the cabbage"

**OVS:** Den Hasen frisst gleich der Fuchs
"The rabbit is eaten soon by the fox"
Anticipation in Visual Worlds

**SVO**: Der Hase frisst gleich den Kohl
“The rabbit eats soon the cabbage”

**OVS**: Den Hasen frisst gleich der Fuchs
“The rabbit is eaten soon by the fox”
**Anticipation in Visual Worlds**

**SVO:** Der Hase frisst gleich *den Kohl*

“The rabbit eats soon the cabbage”

**OVS:** Den Hasen frisst gleich *der Fuchs*

“The rabbit is eaten soon by the fox”

---

**Experiment 1**

**SO-condition**

Normalized Cumulative Gaze Probability

- der Hase
- frisst gleich
- NP2
Anticipation in Visual Worlds

**SVO:** Der Hase frisst gleich **den Kohl**
"The rabbit eats soon the cabbage"

**OVS:** Den Hasen frisst gleich **der Fuchs**
"The rabbit is eaten soon by the fox"
Anticipation in Visual Worlds

SVO: Der Hase frisst gleich **den Kohl**
“The rabbit eats soon *the cabbage*”

OVS: Den Hasen frisst gleich **der Fuchs**
“The rabbit is eaten soon by the fox”

Scene influence on interpretation

**SVO:** Die Prinzessin wäscht gleich den Pirat
“The princess washes soon the pirate”

**OVS:** Die Prinzessin malt gleich der Fechter
“The princess is painted possibly by the fencer”

Scene influence on interpretation

**SVO:** Die Prinzessin wäscht gleich den Pirat
"The princess washes soon the pirate"

**OVS:** Die Prinzessin malt gleich der Fechter
"The princess is painted possibly by the fencer"
Scene influence on interpretation

**SVO:** *Die Prinzessin wäscht gleich den Pirat*
“The princess washes soon the pirate”

**OVS:** *Die Prinzessin malt gleich der Fechter*
“The princess is painted possibly by the fencer”
**Scene influence on interpretation**

**SVO:** *Die Prinzessin wäscht gleich den Pirat*

“The princess washes soon the pirate”

**OVS:** *Die Prinzessin malt gleich der Fechter*

“The princess is painted possibly by the fencer”

Scene influence on interpretation

**SVO:** Die Prinzessin wäscht gleich den Pirat
“The princess washes soon the pirate”

**OVS:** Die Prinzessin malt gleich der Fechter
“The princess is painted possibly by the fencer”

Knowledge versus Scene

- What is the relative priority of information
  - Stored world knowledge lead to anticipation of typical Agent or Theme
  - Scene events with lead to anticipation of depicted Agent or Patient
    Knoeferle, Crocker, Pickering & Scheepers (2005)

- What happens when Knowledge and Scene cause different predictions about likely role fillers?
Knowledge versus Scene

- What is the relative priority of information
  - Stored world knowledge lead to anticipation of typical Agent or Theme
  - Scene events with lead to anticipation of depicted Agent or Patient
    Knoeferle, Crocker, Pickering & Scheepers (2005)

- What happens when Knowledge and Scene cause different predictions about likely role fillers?
Scene vs Knowledge

When scene or knowledge resolve Agent

**Depict:** Den Piloten verköstigt gleich der Detektiv
“The pilot is served soon by the detective”

**Typical:** Den Piloten verzaubert gleich der Zauberer
“The pilot is jinxed soon by the magician”

Scene vs Knowledge

When scene or knowledge resolve Agent

**Depict:** *Den Piloten verköstigt gleich der Detektiv*

“The pilot is **served** soon by the detective”

**Typical:** *Den Piloten verzaubert gleich der Zauberer*

“The pilot is **jinxed** soon by the magician”

Scene vs Knowledge

When scene or knowledge resolve Agent

**Depict:** Den Piloten verköstigt gleich der Detektiv
“The pilot is served soon by the detective”

**Typical:** Den Piloten verzaubert gleich der Zauberer
“The pilot is jinxed soon by the magician”
Scene vs Knowledge

When scene or knowledge resolve Agent

Depict: Den Piloten verköstigt gleich der Detektiv
“The pilot is served soon by the detective”

Typical: Den Piloten verzaubert gleich der Zauberer
“The pilot is jinxed soon by the magician”
Scene vs Knowledge

When scene or knowledge resolve Agent

**Depict:** *Den Piloten verköstigt gleich der Detektiv*
“The pilot is served soon by the detective”

**Typical:** *Den Piloten verzaubert gleich der Zauberer*
“The pilot is jinxed soon by the magician”

Scene vs Knowledge

When scene **and** knowledge resolve Agent

**Depict:** Den Piloten *bespitzelt* gleich *der Zauberer*
“*The pilot is served* soon by the detective”

**Typical:** Den Piloten *bespitzelt* gleich *der Detektiv*
“*The pilot is jinxed* soon by the magician”

Scene vs Knowledge

When scene and knowledge resolve Agent

**Depict:** *Den Piloten bespitzelt gleich der Zauberer*
“The pilot is served soon by the detective”

**Typical:** *Den Piloten bespitzelt gleich der Detektiv*
“The pilot is jinxed soon by the magician”

Scene vs Knowledge

When scene and knowledge resolve Agent

**Depict:** *Den Piloten bespitzelt gleich der Zauberer*

“The pilot is served soon by the detective”

**Typical:** *Den Piloten bespitzelt gleich der Detektiv*

“The pilot is jinxed soon by the magician”

Scene vs Knowledge

When scene and knowledge resolve Agent

**Depict:** *Den Piloten bespitzelt gleich der Zauberer*

“The pilot is served soon by the detective”

**Typical:** *Den Piloten bespitzelt gleich der Detektiv*

“The pilot is jinxed soon by the magician”

Scene vs Knowledge

When scene **and** knowledge resolve Agent

**Depict:** *Den Piloten bespitzelt gleich der Zauberer*
“The pilot is **served** soon by the detective”

**Typical:** *Den Piloten bespitzelt gleich der Detektiv*
“The pilot is **jinxed** soon by the magician”

Summary
Exp 1-2: Rapid use of linguistic and scene information guide interpretation, and direct visual attention
Exp 1-2: Rapid use of linguistic and scene information guide interpretation, and direct visual attention

Exp 3: Anticipation of role-fillers relies preferentially upon depicted events over stored knowledge
Summary

- Exp 1-2: Rapid use of linguistic and scene information guide interpretation, and direct visual attention

- Exp 3: Anticipation of role-fillers relies preferentially upon depicted events over stored knowledge

- The “coordinated interplay” of language, scene and knowledge
  - Utterance ➔ Attention ➔ Scene ➔ Interpretation
  - Priority for scene over knowledge: “believe your eyes”
  - Possibly originates from bootstrapping during acquisition
Coordinated Interplay Account
Coordinated Interplay Account
Coordinated Interplay Account

Utterance ➔ Attention

Knoeferle & Crocker, *J of Memory and Language*, in press.
Coordinated Interplay Account

Utterance

Attention

Visual Scene

Knoeferle & Crocker, *J of Memory and Language*, in press.
Coordinated Interplay Account

Incremental Interpretation:
- Prosodic, lexical, syntactic, semantic constraints

Forward Inferencing:
- knowledge-driven expectations
- scene-driven by events and affordances

Knoeferle & Crocker, *J of Memory and Language*, in press.
Coordinated Interplay Account

Incremental Interpretation:
- Prosodic, lexical, syntactic, semantic constraints

Forward Inferencing:
- knowledge-driven expectations
- scene-driven by events and affordances

Referential:
- rapid (180ms) looks to depicted noun and verb referents
- parallel lexical access

Anticipatory:
- looks to object likely to be mentioned based on scene or knowledge

Knoeferle & Crocker, *J of Memory and Language*, in press.
Coordinated Interplay Account

**Utterance**

**Incremental Interpretation:**
- Prosodic, lexical, syntactic, semantic constraints

**Forward Inferencing:**
- knowledge-driven expectations
- scene-driven by events and affordances

**Attention**

**Referential:**
- rapid (180ms) looks to depicted noun and verb referents
- parallel lexical access

**Anticipatory:**
- looks to object likely to be mentioned based on scene or knowledge

**Visual Scene**

**Referential Context:**
- objects in the scene

**Scene events:**
- depicted actions and participants

**Affordances:**
- events and relationships supported by the scene
Coordinated Interplay Account

Incremental Interpretation:
- Prosodic, lexical, syntactic, semantic constraints

Forward Inferencing:
- knowledge-driven expectations
- scene-driven by events and affordances

Utterance

Visual Scene

Referential Context:
- objects in the scene

Scene events:
- depicted actions and participants

Affordances:
- events and relationships supported by the scene

Attention

Referential:
- rapid (180ms) looks to depicted noun and verb referents
- parallel lexical access

Anticipatory:
- looks to object likely to be mentioned based on scene or knowledge

Coordinated Interplay of Comprehension and Attention

Knoeferle & Crocker, *J of Memory and Language*, in press.
Coordinated Interplay Account

Incremental Interpretation:
- Prosodic, lexical, syntactic, semantic constraints

Forward Inferencing:
- knowledge-driven expectations
- scene-driven by events and affordances

Priority of scene events over stored knowledge

Utterance

Visual Scene

Referential Context:
- objects in the scene

Scene events:
- depicted actions and participants

Affordances:
- events and relationships supported by the scene

Referential:
- rapid (180ms) looks to depicted noun and verb referents
- parallel lexical access

Anticipatory:
- looks to object likely to be mentioned based on scene or knowledge

Attention

Coordinated Interplay of Comprehension and Attention

Knoeferle & Crocker, *J of Memory and Language*, in press.
Adaptive Models
Adaptive Models

Requirements

- Seemless integration of diverse information sources & modalities
- Experience-based: learn selectional/stereotypical role fillers
- Rational: optimal use of immediate information resources & experience
Adaptive Models

Requirements

- Seemless integration of diverse information sources & modalities
- Experience-based: learn selectional/stereotypical role fillers
- Rational: optimal use of immediate information resources & experience

Probabilistic Models:

- Bayesian Networks: Narayanan & Jurafsky (2006)
Adaptive Models

Requirements

- Seemless integration of diverse information sources & modalities
- Experience-based: learn selectional/stereotypical role fillers
- Rational: optimal use of immediate information resources & experience

Probabilistic Models:

- Bayesian Networks: Narayanan & Jurafsky (2006)

Connectionist Models: Simple Recurrent Networks (Elman, 1990)
Simple Recurrent Networks

Feed-forward connectionist architecture

Words are input incrementally

Context layer is a copy of the hidden layer at \( t-1 \)

Supervised learning using back-propagation of error

Output can be anything from next word prediction to full MRS representations
CIAnet

Model the use of attention and priority of the scene:

Gating vector implements attention: learned automatically during training

Multiplied element-wise over event constituents

Binds together events constituents

Implicit inhibition of one event over other

CIAnet

Model the use of attention and priority of the scene:

Gating vector implements attention: learned automatically during training

Multiplied element-wise over event constituents

Binds together events constituents

Implicit inhibition of one event over other

CIAnet

Model the use of attention and priority of the scene:

Gating vector implements attention: learned automatically during training

Multiplied element-wise over event constituents

Binds together events constituents

Implicit inhibition of one event over other

CIAnet

Model the use of attention and priority of the scene:

Goals is to model:
- experience
- immediate scene
- sentence alone
- priority of the scene
- anticipatory behaviour

CIAnet

Model the use of attention and priority of the scene:

Gating vector implements attention: learned automatically during training

Multiplied element-wise over event constituents

Binds together events constituents

Implicit inhibition of one event over other
CIANet Architecture

CIANet Architecture

CIANet Architecture

CIANet Architecture

CIANet Architecture

Scene

Language

input layer

gate(t−1)

context layer

hidden layer

Attention

just−now

SRN trained with BPTT
- Enhanced with encoding of scene
- Agent/action/patient weights shared
- Gating vector implements attention explicitly
- Developed automatically during training
- Multiplied element-wise over event constituents

CIANet Architecture

Predicting Human Behavior
Predicting Human Behavior

- Training data: common lexicon and grammar covering both experiments
  - OVS & SVO sentences generated from experimental materials (26K sentences)
  - Trained on final interpretation
- Conflicting conditions held out

Predicting Human Behavior

- Training data: common lexicon and grammar covering both experiments
  - OVS & SVO sentences generated from experimental materials (26K sentences)
  - Trained on final interpretation
- **Conflicting conditions held out**
- Scene provided as context 50% of time (14K distinct scene events)
  - Unbiased approximation to language experience
  - Adaptive operation with & without scene

Average activation of the attention vector

High variance on first NP: units in the attentional vector are similar

Verb immediate shifts attention to the depicted event, if present

Only the stereotypical, no-conflict condition is the stereotypical event attended
Average activation of the attention vector

High variance on first NP: units in the attentional vector are similar

Verb immediate shifts attention to the depicted event, if present

Only the stereotypical, no-conflict condition is the stereotypical event attended
Normalized Euclidean of the output Subject to depicted Agent

Linguistic expectations dominate at verb: no attention shift yet

Scene dominates at adverb: Attention shifts to scene overriding language

Final interpretation overrides anticipation
Normalized Euclidean of the output Subject to depicted Agent

Linguistic expectations dominate at verb: no attention shift yet

Scene dominates at adverb: Attention shifts to scene overriding language

Final interpretation overrides anticipation
Conclusions
Conclusions

- Coordinated interplay:
  - incremental utterance driven attention to scene objects & events
  - rapid use of event information: reflected in attention & ERP measures
  - priority of scene
Conclusions

- Coordinated interplay:
  - incremental utterance driven attention to scene objects & events
  - rapid use of event information: reflected in attention & ERP measures
  - priority of scene

- Adaptive computational model: CIANet
  - learns linguistic constraints, selectional restrictions & use of scene
  - attention mechanism improves performance and models behaviour
  - predicts scene priority is an emergent behaviour
Directions
Directions

- Models of language processing and language acquisition need to “situated”
  - Beyond lexical and syntactic processes
Directions

- Models of language processing and language acquisition need to “situated”
  - Beyond lexical and syntactic processes

Challenges

- Better models of learning: unsupervised or cognitively plausible supervision
- Grounding in the environment
  - acquisition & enriched comprehension
- Realtime integration of multi-modal information
  - speech, scene, attention, gesture, joint attention ...