Linguists for Deep Learning; or: How I Learned to Stop Worrying and Love Neural Networks

Christopher Potts

Stanford Linguistics

*Sem 2018, June 5–6, New Orleans
Linguists for Deep Learning; or: How I Learned to Stop Worrying and Love Neural Networks

Christopher Potts

Stanford Linguistics

*Sem 2018, June 5–6, New Orleans
Linguists for Deep Learning; or: How I Learned to Stop Worrying and Love Neural Networks

Christopher Potts

Stanford Linguistics

*Sem 2018, June 5–6, New Orleans
Signs of the apocalypse?


*NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened.*
Signs of the apocalypse?


*NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened.*

Yann LeCun in 2015 [link]

*The next frontier for Deep Learning is natural language understanding.*
Signs of the apocalypse?

*NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened.*

Yann LeCun in 2015 [link]
The next frontier for Deep Learning is natural language understanding.

Yann LeCun at Stanford in 2018 [link]
*I would say language is number 300 in the list of 500 problems that we need to face.*
Signs of the apocalypse?


*NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened.*

Yann LeCun in 2015 [link]

*The next frontier for Deep Learning is natural language understanding.*

Yann LeCun at Stanford in 2018 [link]

*I would say language is number 300 in the list of 500 problems that we need to face.*

**Did deep learning swerve off the road instead?**
Signs of the apocalypse?


*NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened.*

But what does this mean?
Signs of the apocalypse?


*NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened.*

But what does this mean?

- If deep learning brings useful tools, ideas, and insights to another field, has it thereby damaged that field? I’d say the opposite!
Signs of the apocalypse?

NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened.

But what does this mean?

- If deep learning brings useful tools, ideas, and insights to another field, has it thereby damaged that field? I’d say the opposite!
- So what potential does deep learning have to improve the science of language?
Signs of the apocalypse?


*NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened.*

But what does this mean?

- If deep learning brings useful tools, ideas, and insights to another field, has it thereby damaged that field? I’d say the opposite!
- So what potential does deep learning have to improve **the science of language**?

My argument today

Deep learning has much to offer the study of **linguistic meaning and communication**.
<table>
<thead>
<tr>
<th>Signs of the apocalypse?</th>
<th><strong>Lexical semantics</strong></th>
<th>Compositional semantics</th>
<th>Pragmatics</th>
<th>The next frontier</th>
</tr>
</thead>
</table>

**Lexical semantics**
# Dimensions of lexical meaning

<table>
<thead>
<tr>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$w_4$</th>
<th>$w_5$</th>
<th>$w_6$</th>
<th>$\vdots$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>$c_2$</td>
<td>$c_3$</td>
<td>$c_4$</td>
<td>$c_5$</td>
<td>$\cdots$</td>
</tr>
</tbody>
</table>
### Dimensions of lexical meaning

<table>
<thead>
<tr>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
<th>C₄</th>
<th>C₅</th>
<th>⋮</th>
</tr>
</thead>
<tbody>
<tr>
<td>W₁</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W₄</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W₅</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W₆</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>⋮</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **The stock deteriorated.**

---

**Diagram**: The structure for the English pronoun *she* is shown in (2):
Dimensions of lexical meaning

<table>
<thead>
<tr>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
<th>C₄</th>
<th>C₅</th>
<th>⋯</th>
</tr>
</thead>
<tbody>
<tr>
<td>W₁</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W₄</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W₅</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W₆</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The stock deteriorated.
The neglect of lexical meaning in semantics

Foreword

In the spring of 1976, Terry Parsons and Barbara Partee taught a course on Montague grammar, which I attended. On the second to the final day of class, Terry went around the room asking the students if there were any questions at all that remained unanswered, and promised to answer them on the last day of class. I asked if he really meant ANY question at all, which he emphatically said that he meant. As I had encountered a few questions in my lifetime that remained at least partially unresolved, I decided to ask one of them. What is life? What is the meaning of life? After all, Barbara and Terry had promised to provide answers to any question at all.

On the final day of class Barbara wore her Montague grammar T-shirt, and she and Terry busied themselves answering our questions. At long last, they came to my question. I anticipated a protracted and involved answer, but their reply was crisp and succinct. First Barbara, chalk in hand, showed me the meaning of life.

\[ \text{"life"} \]

Terry then stepped up and showed me what life really is.

\[ \text{"^life"} \]

As we were asked to show on a homework assignment earlier in the year, this is equivalent to: \text{life}'.

Leaving me astounded that I had been living in such darkness for all these years, the class then turned to the much stickier problem of pronouns.
The neglect of lexical meaning in semantics

Thomason (1974)

The problems of a semantic theory should be distinguished from those of lexicography [...]

The neglect of lexical meaning in semantics

Thomason (1974)

The problems of a semantic theory should be distinguished from those of lexicography [...] A central goal of (semantics) is to explain how different kinds of meanings attach to different syntactic categories; another is to explain how the meanings of phrases depend on those of their components. [...]
The neglect of lexical meaning in semantics

Thomason (1974)

The problems of a semantic theory should be distinguished from those of lexicography [...] A central goal of (semantics) is to explain how different kinds of meanings attach to different syntactic categories; another is to explain how the meanings of phrases depend on those of their components. [...] But we should not expect a semantic theory to furnish an account of how any two expressions belonging to the same syntactic category differ in meaning.
The neglect of lexical meaning in semantics

Thomason (1974)
The problems of a semantic theory should be distinguished from those of lexicography [...] A central goal of (semantics) is to explain how different kinds of meanings attach to different syntactic categories; another is to explain how the meanings of phrases depend on those of their components. [...] But we should not expect a semantic theory to furnish an account of how any two expressions belonging to the same syntactic category differ in meaning. “Walk” and “run,” for instance, and “unicorn” and “zebra” certainly do differ in meaning, and we require a dictionary of English to tell us how. But the making of a dictionary demands considerable knowledge of the world.
Jerrold Katz (1972) on meaning

The arbitrariness of the distinction between form and matter reveals itself [...]

The question “What is meaning?” broken down:

• What is synonymy?
• What is antonymy?
• What is superordination?
• What is semantic ambiguity?
• What is semantic truth?
• What is a possible answer to a question?
Jerrold Katz (1972) on meaning

The arbitrariness of the distinction between form and matter reveals itself [...]

The question “What is meaning?” broken down:

- What is synonymy?
- What is antonymy?
- What is superordination?
- What is semantic ambiguity?
- What is semantic truth?
- What is a possible answer to a question?
- ...
Children are situated word learners

Children learn word meanings

1. with incredible speed
2. despite relatively few inputs
3. by using cues from
   - contrast inherent in the forms they hear
   - social cues
   - assumptions about the speaker’s goals
   - regularities in the physical environment.

(Frank et al. 2012; Frank & Goodman 2014)
Purely distributional meaning
Purely distributional meaning

- High-dimensional

😊
Purely distributional meaning

- High-dimensional
- Meaning from dense linguistic inter-relationships 😊
Purely distributional meaning

- High-dimensional 😊
- Meaning from dense linguistic inter-relationships 😊
- Meaning *solely* from (*n*-th-order) co-occurrence 😞
Purely distributional meaning

- High-dimensional
- Meaning from dense linguistic inter-relationships
- Meaning *solely* from \((n\text{-th order})\) co-occurrence
- No grounding in physical or social contexts
Purely distributional meaning

- High-dimensional 😊
- Meaning from dense linguistic inter-relationships 😊
- Meaning *solely* from \((n\text{-th order})\) co-occurrence 😞
- No grounding in physical or social contexts 😞
- Not symbolic 😞
Faruqui et al. (2015): Retrofitting to graphs

\[ \sum_{i \in V} \alpha_i \| q_i - \hat{q}_i \|^2 + \sum_{(i,j,r) \in E} \beta_{ij} \| q_i - q_j \|^2 \]

Balances fidelity to the original vector \( \hat{q}_i \) against looking more like one’s graph neighbors.

Forces are balanced with \( \alpha = 1 \) and \( \beta = \frac{1}{\text{Degree}(i)} \)
Faruqui et al. (2015): Retrofitting to graphs

\[ \sum_{i \in V} \alpha_i \| q_i - \hat{q}_i \|^2 + \sum_{(i,j,r) \in E} \beta_{ij} \| q_i - q_j \|^2 \]

Balances **fidelity to the original vector** \( \hat{q}_i \) **against looking more like one’s graph neighbors.**

Forces are balanced with \( \alpha = 1 \) and \( \beta = \frac{1}{\text{Degree}(i)} \)
What retrofitting to WordNet might do
What retrofitting to WordNet might do

- Cluster *mammal* with *dog* and *puppy* even though *mammal* has a different, unusual distribution.
What retrofitting to WordNet might do

- Cluster *mammal* with *dog* and *puppy* even though *mammal* has a different, unusual distribution.

- Avoid polarity mistakes like modeling *superb* and *awful* as similar (though beware antonym edges!).
What retrofitting to WordNet might do

- Cluster *mammal* with *dog* and *puppy* even though *mammal* has a different, unusual distribution.

- Avoid polarity mistakes like modeling *superb* and *awful* as similar (though beware antonym edges!).

- Holistic consistency:

![Figure 3: Two-dimensional PCA projections of 100-dimensional SG vector pairs holding the “adjective to adverb” relation, before (left) and after (right) retrofitting.](image-url)
Concerns about identity retrofitting

- No attention to edge semantics; edges mean ‘similar to’.
- Presupposes a uniform initial embedding space.
- No modeling of missing edges.
Hand-build functions from Mrkšić et al. (2016)

- **AntonymRepel:**
  \[ \sum_{(i,j) \in A} \text{ReLU}(1.0 - d(q_i, q_j)) \]

- **SynonymAttract:**
  \[ \sum_{(i,j) \in S} \text{ReLU}(d(q_i, q_j) - 0) \]

- **VectorSpacePreservation:**
  \[ \sum \sum \text{ReLU}(d(q_i, q_j) - d(\hat{q}_i, \hat{q}_j)) \]
Functional relations (Lengerich et al. 2018)

Framework

\[ \sum_{i \in \mathcal{V}} \alpha_i \| q_i - \hat{q}_i \|^2 + \sum_{(i,j,r) \in \mathcal{E}} \beta_{ijr} f_r(q_i, q_j) - \sum_{(i,j,r) \in \mathcal{E}^-} \beta_{ijr} f_r(q_i, q_j) + \lambda \sum_r \rho(f_r) \]
Functional relations (Lengerich et al. 2018)

Framework

\[
\sum_{i \in V} \alpha_i \| q_i - \hat{q}_i \|^2 + \sum_{(i,j,r) \in E} \beta_{ijr} f_r(q_i, q_j) - \sum_{(i,j,r) \in E^-} \beta_{ijr} f_r(q_i, q_j) + \lambda \sum_r \rho(f_r)
\]

Faruqui et al.

\[
f_r(q_i, q_j) = \| q_i - q_j \|^2
\]

with \( \beta_{ijr} = 0 \)
Functional relations (Lengerich et al. 2018)

Framework

\[ \sum_{i \in \mathcal{V}} \alpha_i \|q_i - \hat{q}_i\|^2 + \sum_{(i,j,r) \in \mathcal{E}} \beta_{ijr} f_r(q_i, q_j) - \sum_{(i,j,r) \in \mathcal{E}^-} \beta_{ijr} f_r(q_i, q_j) + \lambda \sum_r \rho(f_r) \]

Linear

\[ f_r(q_i, q_j) = \|A_r q_j + b_r - q_i\|^2 \]

- \( \rho(f_r) = \|A_r\|^2 \)
- We initialize \( A_r = 1 \) and \( b_r = 0 \)
- Initialization can be different for different relations, e.g., \( A_{\text{antonym}} = -1 \)
Functional relations (Lengerich et al. 2018)

Framework

$$\sum_{i \in \mathcal{V}} \alpha_i \| q_i - \hat{q}_i \|^2 + \sum_{(i,j,r) \in \mathcal{E}} \beta_{ijr} f_r(q_i, q_j) - \sum_{(i,j,r) \in \mathcal{E}^-} \beta_{ijr} f_r(q_i, q_j) + \lambda \sum_r \rho(f_r)$$

Simplest neural (akin to Sutskever et al. 2009)

$$f_r(q_i, q_j) = \tanh(q_i^T A_r q_j)$$
Functional relations (Lengerich et al. 2018)

Framework

\[
\sum_{i \in \mathcal{V}} \alpha_i \|q_i - \hat{q}_i\|^2 + \sum_{(i,j,r) \in \mathcal{E}} \beta_{ijr} f_r(q_i, q_j) - \sum_{(i,j,r) \in \mathcal{E}^-} \beta_{ijr} f_r(q_i, q_j) + \lambda \sum_r \rho(f_r)
\]

Neural Tensor Network (akin to Socher et al. 2013)

\[
f_r(q_i, q_j) = u_r^\top \tanh(q_i^\top A_r q_j)
\]

where \(A_r \in \mathbb{R}^{d \times d \times k}\) and \(\rho(f_r) = \|A_r\|^2 + \|u_r\|^2\)
Functional relations (Lengerich et al. 2018)

Framework

$$\sum_{i \in \mathcal{V}} \alpha_i \|q_i - \hat{q}_i\|^2 + \sum_{(i,j,r) \in \mathcal{E}} \beta_{ijr} f_r(q_i, q_j) - \sum_{(i,j,r) \in \mathcal{E}^-} \beta_{ijr} f_r(q_i, q_j) + \lambda \sum_r \rho(f_r)$$

Your favorite graph embedding method
FrameNet evaluation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>87.58</td>
<td>88.59</td>
<td>85.60</td>
<td>91.24</td>
<td>89.59</td>
</tr>
<tr>
<td>Faruqui et al.</td>
<td>90.79</td>
<td>87.87</td>
<td>87.02</td>
<td>94.50</td>
<td>94.24</td>
</tr>
<tr>
<td>FR-Linear</td>
<td><strong>92.92</strong></td>
<td>92.04</td>
<td><strong>89.37</strong></td>
<td>94.65</td>
<td><strong>94.73</strong></td>
</tr>
<tr>
<td>FR-Neural</td>
<td>92.46</td>
<td><strong>92.54</strong></td>
<td><strong>89.57</strong></td>
<td><strong>95.65</strong></td>
<td>94.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>‘Precedes’ (220/136)</th>
<th>‘See Also’ (268/76)</th>
<th>‘Causative Of’ (204/36)</th>
<th>‘Inchoative Of’ (60/16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>87.30</td>
<td>85.11</td>
<td>86.11</td>
<td>82.50</td>
</tr>
<tr>
<td>Faruqui et al.</td>
<td>85.26</td>
<td>83.81</td>
<td>84.49</td>
<td>78.33</td>
</tr>
<tr>
<td>FR-Linear</td>
<td>87.00</td>
<td><strong>91.93</strong></td>
<td><strong>92.09</strong></td>
<td><strong>82.50</strong></td>
</tr>
<tr>
<td>FR-Neural</td>
<td><strong>89.16</strong></td>
<td><strong>93.25</strong></td>
<td><strong>94.33</strong></td>
<td><strong>85.00</strong></td>
</tr>
</tbody>
</table>
A drug–disease knowledge graph

Faruqui et al.

FR-Linear

<table>
<thead>
<tr>
<th>Model</th>
<th>‘Treats’ (9152/2490)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>72.02 ± 0.50</td>
</tr>
<tr>
<td>Faruqui et al.</td>
<td>72.93 ± 0.82</td>
</tr>
<tr>
<td>FR-Linear</td>
<td><strong>84.22 ± 0.82</strong></td>
</tr>
<tr>
<td>FR-Neural</td>
<td>73.52 ± 0.89</td>
</tr>
</tbody>
</table>
## Functional complexity in the lexicon

<table>
<thead>
<tr>
<th>Category</th>
<th>Semantic type</th>
</tr>
</thead>
<tbody>
<tr>
<td>noun</td>
<td>properties</td>
</tr>
<tr>
<td>intransitive verbs</td>
<td>properties</td>
</tr>
<tr>
<td>transitive verbs</td>
<td>entities to properties</td>
</tr>
<tr>
<td>adjectives</td>
<td>properties to properties</td>
</tr>
<tr>
<td>prepositions</td>
<td>entities to (properties to properties)</td>
</tr>
<tr>
<td>determiner</td>
<td>properties to sets of properties</td>
</tr>
</tbody>
</table>

Vector space models tend to be monotyped, but see Clark et al. 2011.
Some lexical generalizations

1. Some transitive verbs entail the existence of their direct object (see) and some do not (seek).

2. Across languages, verbs lexicalize manner or result, but not both (Rappaport Hovav & Levin 2010):
   - **Manner**: nibble, scribble, sweep, flutter
   - **Result**: clean, cover, empty, fill

3. Some adjectives predicate distributively across their arguments, others do not (Glass 2018):
   - **Box A and Box B are new.** (entails both are new)
   - **Box A and Box B are heavy.** (does not entail both are heavy)

Can we develop deep learning systems that derive such generalizations? No training against them; that's just restating them!
Some lexical generalizations

1. Some transitive verbs entail the existence of their direct object (see) and some do not (seek).
Some lexical generalizations

1. Some transitive verbs entail the existence of their direct object (see) and some do not (seek).

2. Across languages, verbs lexicalize manner or result, but not both (Rappaport Hovav & Levin 2010):
   - **Manner:** nibble, scribble, sweep, flutter
   - **Result:** clean, cover, empty, fill
Some lexical generalizations

1. Some transitive verbs entail the existence of their direct object (*see*) and some do not (*seek*).

2. Across languages, verbs lexicalize manner or result, but not both (Rappaport Hovav & Levin 2010):
   - **Manner**: nibble, scribble, sweep, flutter
   - **Result**: clean, cover, empty, fill

3. Some adjectives predicate distributively across their arguments, others do not (Glass 2018):
   - *Box A and Box B are new.*
     (entails both are new)
   - *Box A and Box B are heavy.*
     (does not entail both are heavy)
Some lexical generalizations

1. Some transitive verbs entail the existence of their direct object (see) and some do not (seek).

2. Across languages, verbs lexicalize manner or result, but not both (Rappaport Hovav & Levin 2010):
   - **Manner**: nibble, scribble, sweep, flutter
   - **Result**: clean, cover, empty, fill

3. Some adjectives predicate distributively across their arguments, others do not (Glass 2018):
   - *Box A and Box B are new.*
     (entails both are new)
   - *Box A and Box B are heavy.*
     (does not entail both are heavy)

Can we develop deep learning systems that derive such generalizations? No training against them; that’s just restating them!
Compositional semantics
A semanticist’s ideal

Every student attended a lecture

\[ \forall z \, ((\text{student} \, z) \rightarrow (\exists x \, (\text{lecture} \, x) \land (\text{attended} \, x \, z))) \]

\[ \lambda g \, \forall z \, ((\text{student} \, z) \rightarrow (g \, z)) \land (\exists x \, (\text{lecture} \, x) \land (\text{attended} \, x \, y)) \]

\[ \lambda f \, \lambda g \, \forall z \, ((f \, z) \rightarrow (g \, z)) \quad \text{student} \]

\[ \lambda f \, \lambda g \, (\exists x \, (\text{lecture} \, x) \land (g \, x)) \]

\[ \lambda f \, \lambda g \, (\exists x \, (f \, x) \land (g \, x)) \quad \text{lecture} \]
A semanticist’s ideal

Every student attended a lecture

\[ \forall z \ ((\text{student } z) \rightarrow (\exists x \ (\text{lecture } x) \land (\text{attended } x \ z))) \]

\[ \lambda g \ \forall z \ ((\text{student } z) \rightarrow (g z)) \ \lambda y \ (\exists x \ (\text{lecture } x) \land (\text{attended } x \ y)) \]

\[ \lambda f \ \lambda g \ \forall z \ ((f z) \rightarrow (g z)) \ \text{student} \]

\[ \text{attended} \]

\[ \lambda f \ \lambda g \ (\exists x \ (f x) \land (g x)) \ \text{lecture} \]

But is this really so ideal?
Complete semantic representations?

MacCartney & Manning (2009)

*The difficulty is plain: truly natural language is fiendishly complex. [...] Consider for a moment the difficulty of fully and accurately translating*

1. **Every firm polled saw costs grow more than expected, even after adjusting for inflation.**

to a formal meaning representation.
Sparse, fragmented feature representations

The NYT reported the deal fell through

| the  | 2     | source_NYT | T |
| NYT  | 1     | embedded_implicit_neg | T |
| report | 1     | deal_neg | 1 |
| length | 7     | vocab | 6 |
| …     | …     | … | … |
The answer from deep learning

$$S_1 = f(NP_1, VP_1)$$

$$NP_1 = f(The, NYT)$$

$$VP_1 = f(report, S_2)$$

$$S_2 = f(NP_2, VP_2)$$

$$NP_2 = f(the, deal)$$

$$VP_2 = f(fell, through)$$
The answer from deep learning

\[ h_1 = f(h_0, \text{The}) \]

\[ h_2 = f(h_1, \text{NYT}) \]

\[ h_3 = f(h_2, \text{reported}) \]

\[ h_4 = f(h_3, \text{the}) \]

\[ h_5 = f(h_4, \text{deal}) \]

\[ h_6 = f(h_5, \text{fell}) \]

\[ h_7 = f(h_6, \text{through}) \]
The answer from deep learning

\[ h_0 \rightarrow h_1 \rightarrow h_2 \rightarrow h_3 \rightarrow h_4 \rightarrow h_5 \rightarrow h_6 \rightarrow h_7 \]

The NYT reported the deal fell through
The answer from deep learning

All our parses are wrong, but perhaps we can discover the right one(s).
A new perspective on compositionality
A new perspective on compositionality

Compositionality

The meaning of a complex phrase is a function of the meaning of its constituent phrases.
A new perspective on compositionality

Compositionality
The meaning of a complex phrase is a function of the meaning of its constituent phrases.

Partee (1984):
Context-dependence, Ambiguity, and Challenges to Local, Deterministic Compositionality
A new perspective on compositionality

Compositionality
The meaning of a complex phrase is a function of the meaning of its constituent phrases.

Partee (1984):
Context-dependence, Ambiguity, and Challenges to Local, Deterministic Compositionality

\[ S_1 = \tanh(\text{NP}_1; \text{VP}_1 W + b) \]

\[ \text{NP}_1 \quad \text{VP}_1 = \tanh(\text{reported}; S_2 W + b) \]

\[ \text{reported} \quad S_2 \]
Compositional generalizations: monotonicity

(Bowman 2017)
Compositional generalizations: monotonicity

Kim smoked.
↑
Kim smoked cigars.

Kim didn’t smoke.
↓
Kim didn’t smoke cigars.

(Bowman 2017)
Compositional generalizations: monotonicity

Kim smoked.  Kim didn’t smoke.

Kim smoked cigars.  Kim didn’t smoke cigars.

A student smoked.  A student smoked cigars.

A Swedish student smoked.
Compositional generalizations: monotonicity

Kim smoked.  
↑
Kim smoked cigars.  
Kim didn’t smoke.  
↓
Kim didn’t smoke cigars.

A student smoked.  

A Swedish student smoked.  

A student smoked cigars.  

No student smoked.  

No Swedish student smoked.  

No student smoked cigars.

(Bowman 2017)
Compositional generalizations: monotonicity

Kim smoked.  Kim didn’t smoke.
↑  ↓
Kim smoked cigars.  Kim didn’t smoke cigars.

A student smoked.  A student smoked cigars.

A Swedish student smoked.  A student smoked cigars.

No student smoked.  No student smoked cigars.

No Swedish student smoked.  No student smoked cigars.

Every student smoked.  Every student smoked cigars.

Every Swedish student smoked.  Every student smoked cigars.

(Bowman 2017)
Compositional generalizations: monotonicity

Kim smoked.
↑
Kim smoked cigars.

Kim didn’t smoke.
↓
Kim didn’t smoke cigars.

A student smoked.

A Swedish student smoked.

A student smoked cigars.

No student smoked.

No Swedish student smoked.

No student smoked cigars.

Every student smoked.

Every Swedish student smoked.

Every student smoked cigars.

Most students smoked.

Most Swedish students smoked.

Most students smoked cigars.

(Bowman 2017)
Pragmatics
Natural language is situated and social
Natural language is situated and social

1. I am speaking.
Natural language is situated and social

1. I am speaking.

2. We won. [A team I’m on; a team I support; . . . ]
Natural language is situated and social

1. I am speaking.
2. We won. [A team I’m on; a team I support; . . . ]
3. I am here. [NAACL; New Orleans; planet earth; . . . ]
Natural language is situated and social

1. I am speaking.
2. We won. [A team I’m on; a team I support; . . . ]
3. I am here. [NAACL; New Orleans; planet earth; . . . ]
4. We are here. [pointing at a map]
Natural language is situated and social

1. I am speaking.
2. We won. [A team I’m on; a team I support; . . . ]
3. I am here. [NAACL; New Orleans; planet earth; . . . ]
4. We are here. [pointing at a map]
5. I’m not here now. [answering machine]
## Natural language is situated and social

1. I am speaking.
2. We won.  
   [A team I’m on; a team I support; ...]
3. I am here.  
   [NAACL; New Orleans; planet earth; ...]
4. We are here.  
   [pointing at a map]
5. I’m not here now.  
   [answering machine]
6. We went to a local bar after the workshop.
Natural language is situated and social

1. I am speaking.
2. We won. [A team I’m on; a team I support; . . . ]
3. I am here. [NAACL; New Orleans; planet earth; . . . ]
4. We are here. [pointing at a map]
5. I’m not here now. [answering machine]
6. We went to a local bar after the workshop.
7. three days ago, tomorrow, now, . . .
Natural language is situated and social

8. Where are you from?
Natural language is situated and social

8. Where are you from?
Natural language is situated and social

8. Where are you from?
   b. Stanford.
Natural language is situated and social

8. Where are you from?
   b. Stanford.
   c. The U.S.
Natural language is situated and social

8. Where are you from?
   b. Stanford.
   c. The U.S.
   d. Planet earth.
Natural language is situated and social

8. Where are you from?
   b. Stanford.
   c. The U.S.
   d. Planet earth.

9. If Kangaroos had no tails, they would fall over.
Natural language is situated and social

8. Where are you from?
   b. Stanford.
   c. The U.S.
   d. Planet earth.

9. If Kangaroos had no tails, they would fall over.
   a. True,
Natural language is situated and social

8. Where are you from?
   b. Stanford.
   c. The U.S.
   d. Planet earth.

9. If Kangaroos had no tails, they would fall over.
   a. True,
   b. as long as we don’t slip in the premise that they have jet packs.
Natural language is situated and social

8. Where are you from?
   b. Stanford.
   c. The U.S.
   d. Planet earth.

9. If Kangaroos had no tails, they would fall over.
   a. True,
   b. as long as we don’t slip in the premise that they have jet packs.

10. I didn’t see any.
Natural language is situated and social

8. Where are you from?
   b. Stanford.
   c. The U.S.
   d. Planet earth.

9. If Kangaroos had no tails, they would fall over.
   a. True,
   b. as long as we don’t slip in the premise that they have jet packs.

10. I didn’t see any.
    a. Are there typos in my slides?
Natural language is situated and social

8. Where are you from?
   b. Stanford.
   c. The U.S.
   d. Planet earth.

9. If Kangaroos had no tails, they would fall over.
   a. True,
   b. as long as we don’t slip in the premise that they have jet packs.

10. I didn’t see any.
    a. Are there typos in my slides?
    b. Are the cookies in cupboard.
Natural language is situated and social

8. Where are you from?
   b. Stanford.
   c. The U.S.
   d. Planet earth.

9. If Kangaroos had no tails, they would fall over.
   a. True,
   b. as long as we don’t slip in the premise that they have jet packs.

10. I didn’t see any.
    a. Are there typos in my slides?
    b. Are the cookies in cupboard.
    c. Are there bookstores downtown?
Natural language is situated and social

11. I think this is the way to the library.  [politeness]
Natural language is situated and social

11. I think this is the way to the library. [politeness]

12. Would you mind if I stole your pen for a second? [politeness]
Natural language is situated and social

11. I think this is the way to the library. [politeness]
12. Would you mind if I stole your pen for a second? [politeness]
13. He’s not exactly a genius/idiot. [irony]
Natural language is situated and social

11. I think this is the way to the library.  [politeness]
12. Would you mind if I stole your pen for a second?  [politeness]
13. He’s not exactly a genius/idiot.  [irony]
14. Great idea!  [sarcasm(?)]
Natural language is situated and social

11. I think this is the way to the library. [politeness]
12. Would you mind if I stole your pen for a second? [politeness]
13. He’s not exactly a genius/idiot. [irony]
14. Great idea! [sarcasm(?)]
15. Any chance we can sort this out here, officer? [bribery(?)]
Natural language is situated and social

11. I think this is the way to the library. [politeness]
12. Would you mind if I stole your pen for a second? [politeness]
13. He’s not exactly a genius/idiot. [irony]
14. Great idea! [sarcasm(?)]
15. Any chance we can sort this out here, officer? [bribery(?)]
16. It’d be a shame if something happened to your dog. [threat(?)]
**SHRDLU (Winograd 1972)**

Fully grounded system that parses the user’s input, maps it to logical form, interprets that logical form in its world, tries to take appropriate action, and generates responses.

One project did succeed. Terry Winograd’s program SHRDLU could use English intelligently, but there was a catch: the only subject you could discuss was a micro-world of simulated blocks.
SHRDLU (Winograd 1972)

1. Person: WHAT DOES THE BOX CONTAIN?
2. Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
3. Person: WHAT IS THE PYRAMID SUPPORTED BY?
4. Computer: THE BOX.
5. Person: HOW MANY BLOCKS ARE NOT IN THE BOX?
6. Computer: FOUR OF THEM.
7. Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?
9. Person: IS IT SUPPORTED?
10. Computer: YES, BY THE TABLE.
Example: sarcasm

![Diagram showing the process of determining if a statement is sarcastic.](attachment:sarcasm_diagram.png)
Example: sarcasm

SARC (Khodak et al. 2017); Hazarika et al. 2018
# Example: Colors in context

<table>
<thead>
<tr>
<th>Context</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Blue Color" /></td>
<td>blue</td>
</tr>
<tr>
<td><img src="image" alt="Purple Color" /></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Green Color" /></td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Example from the Colors in Context corpus from the Stanford Computation & Cognition Lab
Example: Colors in context

<table>
<thead>
<tr>
<th>Context</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Blue" /> <img src="image2" alt="Purple" /> <img src="image3" alt="Green" /></td>
<td>blue</td>
</tr>
<tr>
<td><img src="image1" alt="Blue" /> <img src="image2" alt="Pink" /> <img src="image3" alt="Blue" /></td>
<td>The darker blue one</td>
</tr>
</tbody>
</table>
Example: Colors in context

<table>
<thead>
<tr>
<th>Context</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Blue Square]</td>
<td>![Blue Square]</td>
</tr>
<tr>
<td>![Blue Square]</td>
<td>![Purple Square]</td>
</tr>
<tr>
<td>![Purple Square]</td>
<td>![Pink Square]</td>
</tr>
</tbody>
</table>

Table: Example from the Colors in Context corpus from the Stanford Computation & Cognition Lab
## Example: Colors in context

<table>
<thead>
<tr>
<th>Context</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Blue" /></td>
<td>blue</td>
</tr>
<tr>
<td><img src="image2" alt="Blue" /></td>
<td>The darker blue one</td>
</tr>
<tr>
<td><img src="image3" alt="Pink" /></td>
<td>dull pink not the super bright one</td>
</tr>
<tr>
<td><img src="image4" alt="Purple" /></td>
<td>Purple</td>
</tr>
</tbody>
</table>

**Table:** Example from the Colors in Context corpus from the Stanford Computation & Cognition Lab
## Example: Colors in context

<table>
<thead>
<tr>
<th>Context</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Blue" /></td>
<td>blue</td>
</tr>
<tr>
<td><img src="image2" alt="Darker Blue" /></td>
<td>The darker blue one</td>
</tr>
<tr>
<td><img src="image3" alt="Dull Pink" /></td>
<td>dull pink not the super bright one</td>
</tr>
<tr>
<td><img src="image4" alt="Purple" /></td>
<td>Purple</td>
</tr>
<tr>
<td><img src="image5" alt="Blue" /></td>
<td>blue</td>
</tr>
</tbody>
</table>

**Table:** Example from the Colors in Context corpus from the Stanford Computation & Cognition Lab
Literal neural speaker $S_0$
Neural literal listener $\mathcal{L}_0$
Neural pragmatic agents
Neural pragmatic agents

Neural pragmatic speaker (Andreas & Klein 2016)

\[ S_1(msg \mid c, C; \theta) = \frac{\mathcal{L}_0(c \mid msg, C; \theta)}{\sum_{msg' \in X} \mathcal{L}_0(c \mid msg', C; \theta)} \]
Neural pragmatic agents

Neural pragmatic speaker (Andreas & Klein 2016)

\[
S_1(\text{msg} \mid c, C; \theta) = \frac{\mathcal{L}_0(c \mid \text{msg}, C; \theta)}{\sum_{\text{msg}' \in X} \mathcal{L}_0(c \mid \text{msg}', C; \theta)}
\]

where $X$ is a sample from $S_0(\text{msg} \mid c, C; \theta)$ such that $\text{msg}^* \in X$. 
Neural pragmatic agents

Neural pragmatic speaker (Andreas & Klein 2016)

\[ S_1(msg | c, C; \theta) = \frac{\mathcal{L}_0(c | msg, C; \theta)}{\sum_{msg' \in X} \mathcal{L}_0(c | msg', C; \theta)} \]

where \( X \) is a sample from \( S_0(msg | c, C; \theta) \) such that \( msg^* \in X \).

Neural pragmatic listener

\[ \mathcal{L}_1(c | msg, C; \theta) \propto S_1(msg | c, C; \theta) \]
Example: Pragmatic image captioning

Mao et al. (2016); Vedantam et al. (2017): Captions that are true and distinguish their images from related ones.

S₀ caption: the dog is brown
S₁ caption: the head of a dog

Reasoning about all possible utterances/captions?

(Cohn-Gordon et al. 2018)
Example: Pragmatic image captioning

Mao et al. (2016); Vedantam et al. (2017): Captions that are true and distinguish their images from related ones.

S₀ caption: the dog is brown
S₁ caption: the head of a dog

Reasoning about all possible utterances/captions?
⇒ Sample from $S_0$

(Cohn-Gordon et al. 2018)
Example: Pragmatic image captioning

Mao et al. (2016); Vedantam et al. (2017): Captions that are true and distinguish their images from related ones.

S₀ caption: the dog is brown
S₁ caption: the head of a dog

Reasoning about all possible utterances/captions?

⇒ Full pragmatic reasoning about characters

(Cohn-Gordon et al. 2018)
Some pragmatic generalizations

1. **Scalar implicature**: general terms tend to signal that their more specific alternatives are pragmatically marked.

2. **I-implicature**: if a general term has prototypical instantiations in context, then it might be refined to pick out just those prototypes.

3. **Manner implicature**: unusual events are described with unusual language; normal events with normal language.

4. **Metaphor**: metaphorical language is pervasive and enables the speaker to highlight specific dimensions of meaning efficiently.

5. **Contextual refinement**: word and phrase meanings are flexible and respond to the social context.
The next frontier
The next frontier

The Human Speechome Project (Roy et al. 2006)
The next frontier

The Human Speechome Project (Roy et al. 2006)

And, not to play into stereotypes of linguists, but some symbolic reasoning would be useful!
The next frontier

The Human Speechome Project (Roy et al. 2006)

And, not to play into stereotypes of linguists, but some symbolic reasoning would be useful!

Thanks!
References


References


