A Cost Metric for Pronoun Resolution: Uncertainty Increases Processing Cost

Olga Seminck & Pascal Amsili

(1) Université catholique de Louvain, Institute of Neuroscience, Media innovation and intelligibility Lab, Centre de traitement automatique du langage

(2) Université Sorbonne Nouvelle - Paris 3, Laboratoire Lattice, CNRS

olga.seminck@uclouvain.be

Saarbrücken, 25 October 2019
Introduction
Pronoun resolution is a form of anaphora resolution.

NP $\alpha_1$ takes NP $\alpha_2$ as its anaphoric antecedent if $\alpha_1$ depends on $\alpha_2$ for its interpretation. (Van Deemter and Kibble 2000)

A secret’s worth depends on the people from whom it must be kept.

*The Shadow of the Wind*, Carlos Ruiz Zafón

Pronoun resolution is the process of finding the antecedent of an anaphoric pronoun.
What leads to cognitive cost in pronoun resolution?

- **Multiple Factors**
  - Grammatical role of the antecedent (Crawley et al. 1990)
  - Parallel grammatical roles (Smyth 1994)
  - Frequency of the antecedent (Shillcock 1982)
  - Distance between the pronoun and the antecedent (Clark and Sengul 1979)
  - ...  

For modelisation of cognitive cost on corpus, difficult to take all factors into account

According to some theories, these factors reflect more abstract but broader processes
Cost Metric
Cost Metrics

Cost Metric: formula that predicts processing cost
- Translates hypothesis into prediction

Example: surprisal
- Hypothesis: unexpected events are harder to process
- Cost metric: Difficulty(event) = − log(P(event))
A cost metric to predict the difficulty of pronouns

Prediction for pronouns resolution:
More uncertainty about the antecedent $\rightarrow$ more processing cost

Entropy: measure of uncertainty

$$H(X) = - \sum_{j \in X} p(X = j) \cdot \log_2(p(X = j))$$

Entropy

Applies to a random variable: antecedent of a pronoun
Relative Entropy

Entropy increases with the number of antecedent candidates.

- Keep scores comparable through the text
- ‘Normalise’ the entropy

Normalisation: relative entropy
‘Distance’ between actual probability distribution & flat distribution

\[ H_{relative}(P||Q) = \sum_{i \in P \land i \in Q} P(i) \log \frac{P(i)}{Q(i)} \]  

Larger distance \( \Rightarrow \) less uncertainty \( \Rightarrow \) less processing cost
Experiment
Test of the Entropy Cost Metric on Corpus

Does the cost metric make relevant predictions about reading behaviour on corpus?

- Dundee Corpus annotated for anaphoric pronouns
- Automatic coreference resolution system
  - Estimate relative entropy of pronouns
Eye-movements of 10 native English speakers
Reading 65 texts
From the Independent (newspaper)
Total: 50 000 tokens
Annotated with part of speech (Frank 2010) and dependency relations (Barrett et al. 2015)

Annotation of the antecedent of all 1 109 anaphorical pronouns.
A data-set to study pronoun resolution in natural data.
Measuring reading time for pronouns: a hard problem

Pronouns are fixated only 20 - 30% of the time
(Ehrlich and Rayner 1983)

Pronouns are very short

Spill-over effects

Other experiments:
(Seminck and Amsili 2018; Von der Malsburg 2018)
Take a window of words around the pronoun.

► ... at a time \textbf{[when they are at greatest risk]}, and then ...
► ... on it; \textbf{[but it would seriously degrade the]} quality ...

Problems:
► Need multiple models
► Few data-points per pronoun
Solution: binomial metric

A simpler reading metric was more suited: Is the pronoun fixated or not?

Binomial outcome: yes/no answer.

Advantages:

▶ More data points
▶ There is only one point to measure

“a word is skipped because it has been identified on the previous fixation” (Brysbaert and Vitu 1998)

Hypothesis: a fixated pronoun indicates more processing difficulty.
An NLP-system gives the probability distribution

The Red House tells the story of a mysterious, tormented individual who breaks into toy shops and museums to steal dolls and puppets. Once they are in his power...

1. Probability distribution from parameters of resolution system

2. Calculate relative entropy over this probability distribution

<table>
<thead>
<tr>
<th>Antecedent of they</th>
<th>Probability</th>
<th>Relative entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Red House</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>a mysterious, tormented individual</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>toy shops and museums</td>
<td>0.31</td>
<td>0.83</td>
</tr>
<tr>
<td>dolls and puppets</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>
The resolution system

Obtain probabilities from a state of the art NLP-system (Lee et al. 2017)

Lee et al.’s system:

- End to end: without pre-processing
- Neural-network architecture
- Ranking system
Result
A statistical model predicted whether the pronoun is fixated or not.

Is the relative entropy of importance to this prediction?

Mixed effects model:
\[
\text{fixated} \sim \text{length} + \text{frequency} + \text{comma} + \text{punctuation} + \text{rel}_\text{ent} \\
+ (1 + \text{rel}_\text{ent} | \text{participant}) + (1 | \text{item})
\]
The entropy cost metric predicts reading behaviour

The relative entropy was a predictor in reading behaviour.

A lower distance between the entropy and the maximal entropy
⇒ more participants fixating the pronoun

Estimate: -0.07 (95% Credible interval = [-0.01, -0.13])

Conclusion:
More uncertainty about the antecedent of the pronoun leads to
more people fixating it.
Conclusion
Conclusion

Uncertainty about the antecedent leads to more cognitive cost in pronoun resolution

Notions from Information Theory are also relevant for pronoun resolution

NLP-systems can be used to estimate probabilities that are relevant to human language processing


