Let’s do it “again”: A First Computational Approach to Detecting Adverbial Presupposition Triggers

ANDRE CIANFLONE*, YULAN FENG*, JAD KABBARA* & JACKIE CK CHEUNG

(* EQUAL CONTRIBUTION)
“Again”

Heard on the campaign trail:

Make the middle class mean something *again*, with rising incomes and broader horizons.

Make America great *again*.
What is presupposition?

• **Presuppositions**: assumptions shared by discourse participants in an utterance (Frege 1892, Strawson 1950, Stalnaker 1973, Stalnaker 1998).

• **Presupposition triggers**: expressions that indicate the presence of presuppositions.

• Example:

Oops! I did it *again* ← Trigger

• Presupposes Britney did it before
Linguistic Analysis

• Presuppositions are preconditions for statements to be true or false (Kaplan 1970; Strawson, 1950).

• Classes of construction that can trigger presupposition (Zare et al., 2012):
  – Definite descriptions (Kabbara et al., 2016), e.g.: “The queen of the United Kingdom”.
  – Stressed constituents (Krifka, 1998), e.g.: “Yes, Peter did eat pasta.”
  – Factive verbs, e.g.: “Michael regrets eating his mother’s cookies.”
  – Implicative verbs, e.g.: “She managed to make it to the airport on time.”
  – Relations between verbs (Tremper and Frank, 2013; Bos, 2003), e.g.: won >> played.
Motivation & Applications

• Interesting testbed for pragmatic reasoning: investigating presupposition triggers requires understanding preceding context.

• Presupposition triggers influencing political discourse:
  - The abundant use of presupposition triggers helps to better communicate political messages and consequently persuade the audience (Liang and Liu, 2016).

• To improve the readability and coherence in language generation applications (e.g., summarization, dialogue systems).
Adverbial Presupposition Triggers

• Adverbial presupposition triggers such as again, also, and still.

• Indicate the recurrence, continuation, or termination of an event in the discourse context, or the presence of a similar event.

• The most commonly occurring presupposition triggers (after existential triggers) (Khaleel, 2010).

• Little work has been done on these triggers in the computational literature from a statistical, corpus-driven perspective.
This Work

• **Computational approach** to detecting presupposition triggers.

• Create **new datasets** for the task of detecting adverbial presupposition triggers.

• **Control for potential confounding factors** such as class balance and syntactic governor of the triggering adverb.

• Present a new **weighted pooling attention mechanism** for the task.
Outline

Task Definition
Learning Model
Experiments & Results
Task

- Detect contexts in which adverbial presupposition triggers can be used.
- Requires detecting recurring or similar events in the discourse context.
- Five triggers of interest: *too, again, also, still, yet*.
- Frame the learning problem as a binary classification for predicting the presence of an adverbial presupposition (as opposed to the identity of the adverb).
Sample Configuration

- 3-tuple: label, list of tokens, list of POS tags.
- Back to our example:

  Make America great again.
Sample Configuration

• 3-tuple: label, list of tokens, list of POS tags.

• Back to our example:

Make America great again. ← Trigger
Sample Configuration

• 3-tuple: label, list of tokens, list of POS tags.

• Back to our example:

Make America great again. ← Trigger

Headword (aka governor of ”again”)
Sample Configuration

• 3-tuple: label, list of tokens, list of POS tags.

• Back to our example:

@@@ Make America great again.  

Trigger

Headword
(aka governor of "again")

• Special token: to identify the candidate context in the passage to the model.
Sample Configuration

• 3-tuple: label, list of tokens, list of POS tags.

• Back to our example: REMOVE ADVERBS

@@@ Make America great again.  Trigger

Headword
(aka governor of "again")
Sample Configuration

• 3-tuple: label, list of tokens, list of POS tags.
• Back to our example:

```
( 'again', Trigger
[‘@@@@@’, ‘Make’, ‘America’, ‘great’],
```
Positive vs Negative Samples

• Negative samples
  - Same governors as in the positive cases but without triggering presupposition.

• Example of positive sample:
  - Juan is coming to the event too.

• Example of negative sample:
  - Whitney is coming tomorrow.
Extracting Positive Samples

• Scan through all the documents to search for target adverbs.

• For each occurrence of a target adverb:
  - Store the location and the governor of the adverb.
  - Extract 50 unlemmatized tokens preceding the governor, together with the tokens right after it up to the end of the sentence (where the adverb is).
  - Remove adverb.
Extracting Negative Samples

• Extract sentences containing the same governors (as in the positive cases) but not any of the target adverbs.
  - Number of samples in the positive and negative classes roughly balanced.

• Negative samples are extracted/constructed in the same manner as the positive examples.
Position-Related Confounding Factors

We try to control position-related confounding factors by two randomization approaches:

1. Randomize the order of documents to be scanned.
2. Within each document, start scanning from a random location in the document.
Learning Model

- Presupposition involves reasoning over multiple spans of text.
- At a high level, our model extends a bidirectional LSTM model by:
  1. Computing correlations between the hidden states at each timestep.
  2. Applying an attention mechanism over these correlations.
- **No new parameters** compared to standard bidirectional LSTM.
Learning Model: Overview

- **Input:**
  - POS tag one-hot encoding
  - Word embedding lookup
  - The Old Granary... @@ included Bertrand Russell

- **biLSTM:**
  - Hidden states concatenated to form matrix H

- **Column-wise softmax**
  - $M = H^TH$

- **Row-wise softmax**

- **Column-wise average**

- **Output:**
  - Fully connected layer
  - Sum over time
  - Weighted H matrix
  - Attention vector from vector/matrix multiplication
Learning Model: Input

- **Embed** input.
- Optionally concatenate with **POS tags**.

**Embedding + POS**

- POS tag one-hot encoding
- Word embedding lookup
Learning Model: RNN

• **Bidirectional LSTM:**
  Matrix $H = [h_1||h_2|| ... ||h_T]$ concatenates all hidden states.

• E.g.:

We continue to feel that the stock market is the place to be for long-term appreciation.
Learning Model: Matching Matrix

• Pair-wise matching matrix $M$

\[ M = H^T H \]
Learning Model: Softmax

• **Column-wise** softmax: Learn how to aggregate.
Learning Model: Softmax

- **Column-wise** softmax: Learn how to aggregate.
- **Row-wise** softmax: Attention distribution over words.

```
Column-wise softmax

softmax

M = H^TH

biLSTM: hidden states concatenated to form matrix H

Input: encoding and embeddings concatenated

POS tag one-hot encoding

word embedding lookup

The Old Granary... @@@@ included Bertrand Russell
```
Learning Model: Attention Score

- The columns of $M^r$ are then averaged, forming vector $\beta$. 

![Diagram of Learning Model: Attention Score](image)

- biLSTM: hidden states concatenated to form matrix $H$
- Input: encoding and embeddings concatenated
- Column-wise softmax
- Row-wise softmax

\[ M = H^T H \]
Learning Model: Attention Score

• The columns of $M^r$ are then averaged, forming vector $\beta$.

• Final attention vector $\alpha$:

$$\alpha = M^c \beta$$

based on (Cui et al., 2017).
Learning Model: Attend

• **Attend:**
  
  \[ c = \sum_{t=1}^{T} \alpha_t h_t . \]

• A form of **self-attention** (Paulus 2017, Vaswani 2017).
Learning Model: Predict

- **Predict:**
  - Dense layer:
    \[ z = \sigma(W_z c + b_z). \]
  - Softmax:
    \[ y = s(W_0 z + b_0). \]
Datasets

New datasets extracted from:

• The **English Gigaword** corpus:
  - Individual sub-datasets (i.e., presence of each adverb vs. absence).
  - ALL (i.e., presence of one of the 5 adverbs vs. absence).

• The **Penn Tree Bank (PTB)** corpus:
  - ALL.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Training</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTB</td>
<td>5,175</td>
<td>482</td>
</tr>
<tr>
<td>Gigaword yet</td>
<td>63,843</td>
<td>15840</td>
</tr>
<tr>
<td>Gigaword too</td>
<td>85,745</td>
<td>21501</td>
</tr>
<tr>
<td>Gigaword again</td>
<td>85,944</td>
<td>21762</td>
</tr>
<tr>
<td>Gigaword still</td>
<td>194,661</td>
<td>48741</td>
</tr>
<tr>
<td>Gigaword also</td>
<td>537,626</td>
<td>132928</td>
</tr>
</tbody>
</table>
Results Overview

• Our model outperforms all other models in **10 out of 14 scenarios** (combinations of datasets and whether or not POS tags are used).

• **WP outperforms regular LSTM** without introducing additional parameters.

• For all models, we find that **including POS tags benefits** the detection of adverbial presupposition triggers in Gigaword and PTB datasets.
Results – WSJ

- **WP** best on WSJ.
- **RNNs** outperform baselines by large margin.

<table>
<thead>
<tr>
<th>Models</th>
<th>Variants</th>
<th>All adverbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFC</td>
<td>-</td>
<td>51.66</td>
</tr>
<tr>
<td>LogReg</td>
<td>+ POS</td>
<td>52.81</td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td>54.47</td>
</tr>
<tr>
<td>CNN</td>
<td>+ POS</td>
<td>58.84</td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td>62.16</td>
</tr>
<tr>
<td>LSTM</td>
<td>+ POS</td>
<td>74.23</td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td>73.18</td>
</tr>
<tr>
<td>WP</td>
<td>+ POS</td>
<td><strong>76.09</strong></td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td><strong>74.84</strong></td>
</tr>
</tbody>
</table>

**MFC**: Most Frequent Class

**LogReg**: Logistic Regression

**LSTM**: bidirectional LSTM

**CNN**: Convolutional Network based on (Kim 2014)
Results – Gigaword

• Baselines

<table>
<thead>
<tr>
<th>Models</th>
<th>Variants</th>
<th>All adverbs</th>
<th>Again</th>
<th>Still</th>
<th>Too</th>
<th>Yet</th>
<th>Also</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFC</td>
<td>-</td>
<td>50.24</td>
<td>50.25</td>
<td>50.29</td>
<td>65.06</td>
<td>50.19</td>
<td>50.32</td>
</tr>
<tr>
<td>LogReg</td>
<td>+ POS</td>
<td>53.65</td>
<td>59.49</td>
<td>56.36</td>
<td>69.77</td>
<td><strong>61.05</strong></td>
<td>52.00</td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td>52.86</td>
<td>58.60</td>
<td>55.29</td>
<td>67.60</td>
<td><strong>58.60</strong></td>
<td>56.07</td>
</tr>
<tr>
<td>CNN</td>
<td>+ POS</td>
<td>59.12</td>
<td>60.26</td>
<td>59.54</td>
<td>67.53</td>
<td>59.69</td>
<td>61.53</td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td>57.21</td>
<td>57.28</td>
<td>56.95</td>
<td>67.84</td>
<td>56.53</td>
<td>59.76</td>
</tr>
</tbody>
</table>
Results – Gigaword

- LSTM and LSTM with Attention (WP)

<table>
<thead>
<tr>
<th>Models</th>
<th>Variants</th>
<th>All adverbs</th>
<th>Again</th>
<th>Still</th>
<th>Too</th>
<th>Yet</th>
<th>Also</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFC</td>
<td>-</td>
<td>50.24</td>
<td>50.25</td>
<td>50.29</td>
<td>65.06</td>
<td>50.19</td>
<td>50.32</td>
</tr>
<tr>
<td>LogReg</td>
<td>+ POS</td>
<td>53.65</td>
<td>59.49</td>
<td>56.36</td>
<td>69.77</td>
<td><strong>61.05</strong></td>
<td>52.00</td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td>52.86</td>
<td>58.60</td>
<td>55.29</td>
<td>67.60</td>
<td><strong>58.60</strong></td>
<td>56.07</td>
</tr>
<tr>
<td>CNN</td>
<td>+ POS</td>
<td>59.12</td>
<td>60.26</td>
<td>59.54</td>
<td>67.53</td>
<td>59.69</td>
<td>61.53</td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td>57.21</td>
<td>57.28</td>
<td>56.95</td>
<td>67.84</td>
<td>56.53</td>
<td>59.76</td>
</tr>
<tr>
<td>LSTM</td>
<td>+ POS</td>
<td><strong>60.58</strong></td>
<td><strong>61.81</strong></td>
<td>60.72</td>
<td><strong>69.70</strong></td>
<td>59.13</td>
<td><strong>81.48</strong></td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td><strong>58.86</strong></td>
<td><strong>59.93</strong></td>
<td><strong>58.97</strong></td>
<td><strong>68.32</strong></td>
<td>55.71</td>
<td><strong>81.16</strong></td>
</tr>
<tr>
<td>WP</td>
<td>+ POS</td>
<td><strong>60.62</strong></td>
<td>61.59</td>
<td><strong>61.00</strong></td>
<td>69.38</td>
<td>57.68</td>
<td><strong>82.42</strong></td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td><strong>58.87</strong></td>
<td>58.49</td>
<td><strong>59.03</strong></td>
<td><strong>68.37</strong></td>
<td>56.68</td>
<td><strong>81.64</strong></td>
</tr>
</tbody>
</table>
Results – Gigaword

- **WP** outperforms in 10 out of 14 cases.
- Better performance with **POS**.

<table>
<thead>
<tr>
<th>Models</th>
<th>Variants</th>
<th>All adverbs</th>
<th>Again</th>
<th>Still</th>
<th>Too</th>
<th>Yet</th>
<th>Also</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFC</td>
<td>-</td>
<td>50.24</td>
<td>50.25</td>
<td>50.29</td>
<td>65.06</td>
<td>50.19</td>
<td>50.32</td>
</tr>
<tr>
<td>LogReg</td>
<td>+ POS</td>
<td>53.65</td>
<td>59.49</td>
<td>56.36</td>
<td>69.77</td>
<td><strong>61.05</strong></td>
<td>52.00</td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td>52.86</td>
<td>58.60</td>
<td>55.29</td>
<td>67.60</td>
<td><strong>58.60</strong></td>
<td>56.07</td>
</tr>
<tr>
<td>CNN</td>
<td>+ POS</td>
<td>59.12</td>
<td>60.26</td>
<td>59.54</td>
<td>67.53</td>
<td>59.69</td>
<td>61.53</td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td>57.21</td>
<td>57.28</td>
<td>56.95</td>
<td>67.84</td>
<td>56.53</td>
<td>59.76</td>
</tr>
<tr>
<td>LSTM</td>
<td>+ POS</td>
<td>60.58</td>
<td><strong>61.81</strong></td>
<td>60.72</td>
<td><strong>69.70</strong></td>
<td>59.13</td>
<td>81.48</td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td><strong>58.86</strong></td>
<td><strong>59.93</strong></td>
<td><strong>58.97</strong></td>
<td><strong>68.32</strong></td>
<td>55.71</td>
<td>81.16</td>
</tr>
<tr>
<td>WP</td>
<td>+ POS</td>
<td><strong>60.62</strong></td>
<td>61.59</td>
<td><strong>61.00</strong></td>
<td>69.38</td>
<td>57.68</td>
<td><strong>82.42</strong></td>
</tr>
<tr>
<td></td>
<td>- POS</td>
<td><strong>58.87</strong></td>
<td>58.49</td>
<td><strong>59.03</strong></td>
<td><strong>68.37</strong></td>
<td>56.68</td>
<td><strong>81.64</strong></td>
</tr>
</tbody>
</table>
Qualitative Analysis

• Positive sample:
  ... We continue to feel that the stock market is the place to be for long-term appreciation.

• Negative sample:
  ... Careers count most for the well-to-do. Many affluent people place personal success and money above family.
Conclusion

• New task, detection of adverbial presupposition triggers
• New datasets for the task.
• New attention model tailored for the task.
• Our model outperforms other strong baselines without additional parameters over the standard LSTM model.
Future Directions

• Incorporate such a system in an NLG pipeline (e.g., dialogue or summarization with text rewriting).
• Discourse analysis with presupposition (e.g., political speech).
• Investigate other types of presupposition.
Thank you to our co-authors:

Yulan Feng  Prof. Jackie CK Cheung

Thank you to our sponsors: