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:

Hillary Clinton

Donald Trump

`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, Stalnaker1998).
- **Presupposition triggers**: expressions that indicate the presence of presuppositions.
- Example:

`Oops! I did it` *`again`* ← Trigger

- Presupposes Britney did it before

Linguistic Analysis

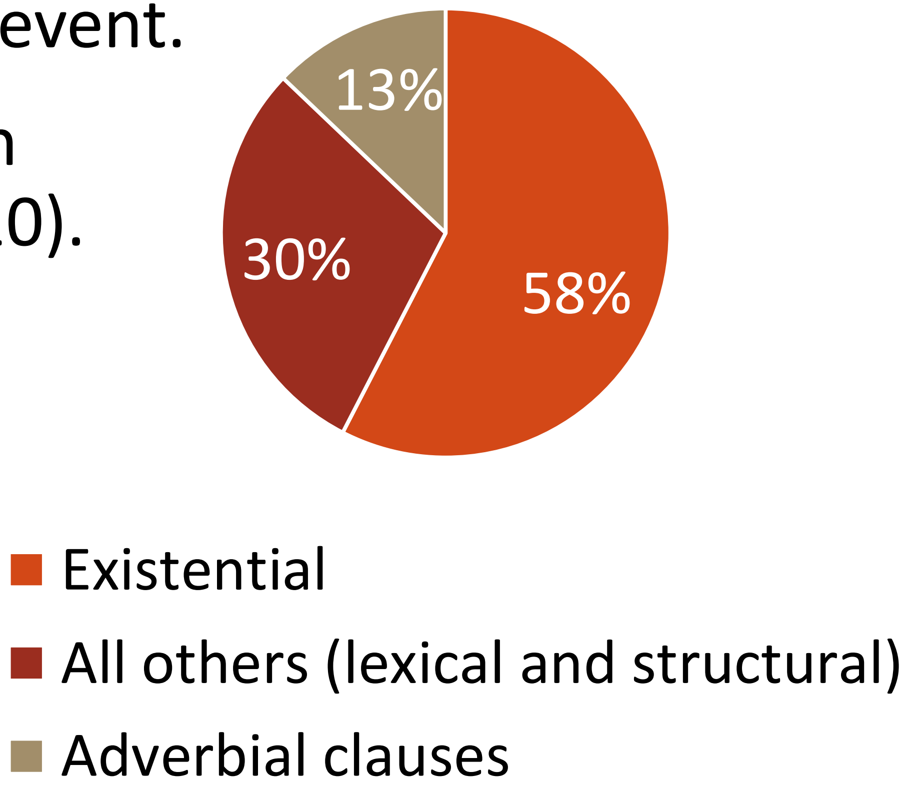
- Presuppositions are preconditions for statements to be true or false (Kaplan 1970; Strawson, 1950).
- Classes of construction that can trigger presupposition (Zare et al., 2012):
  - Definite descriptions (Kabbara et al., 2016), e.g.: “The queen of the United Kingdom”.
  - Stressed constituents (Krifka, 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.

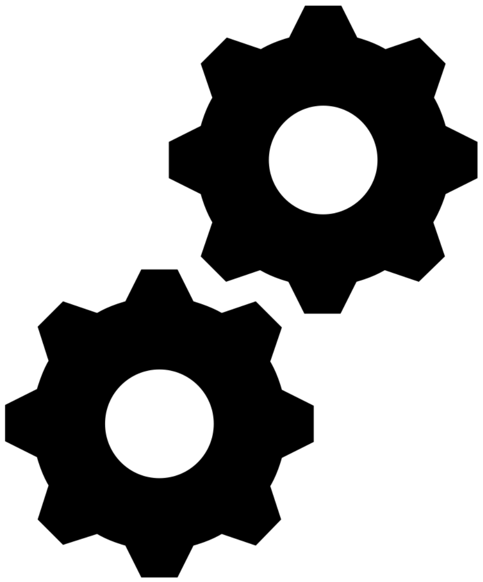


This Work

- **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'],  ← Tokens
  ['@@@@',  'VB',   'NNP',     'JJ'  ] ) ← POS tags
```

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

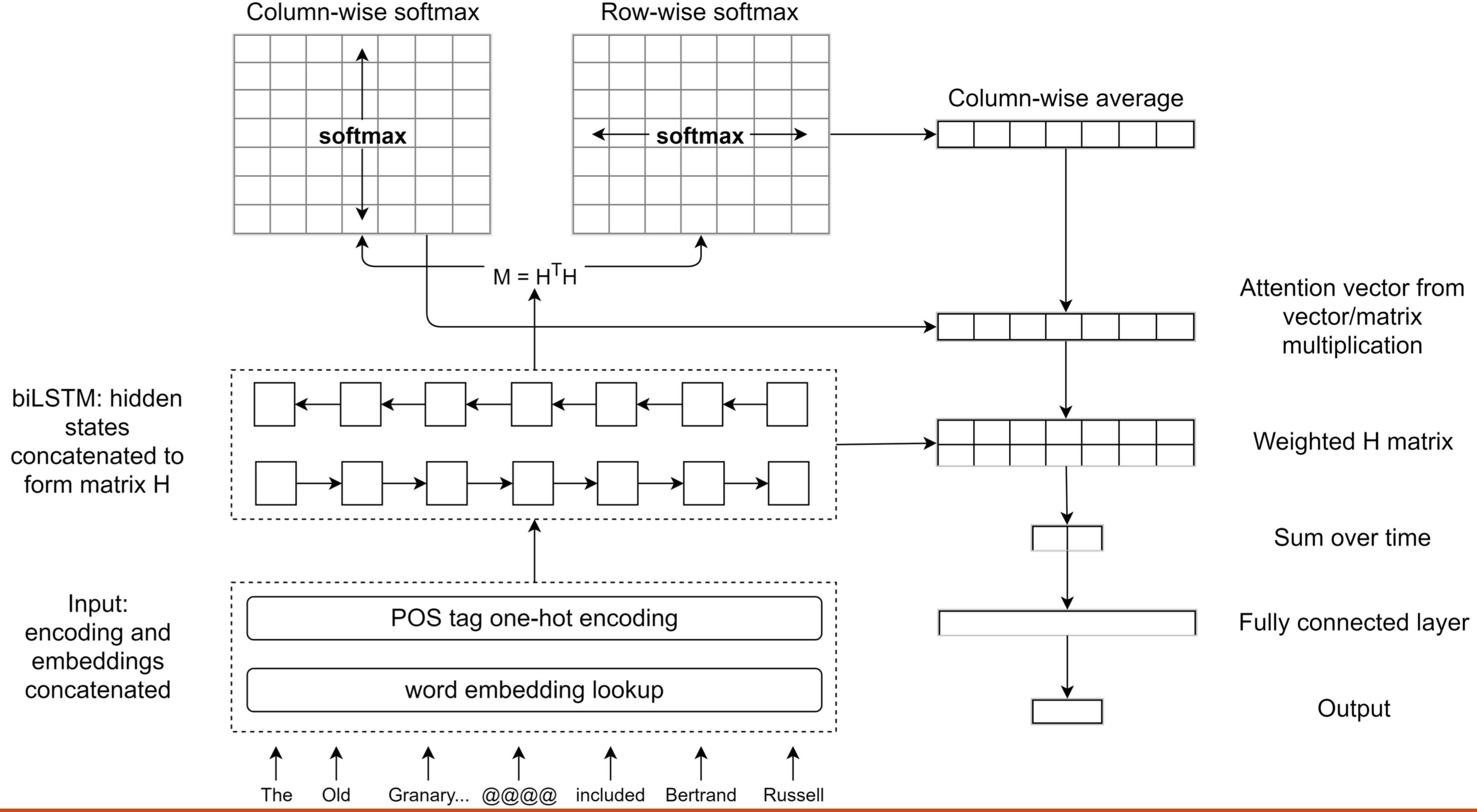
We try to control position-related confounding factors by two randomization approaches:

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.

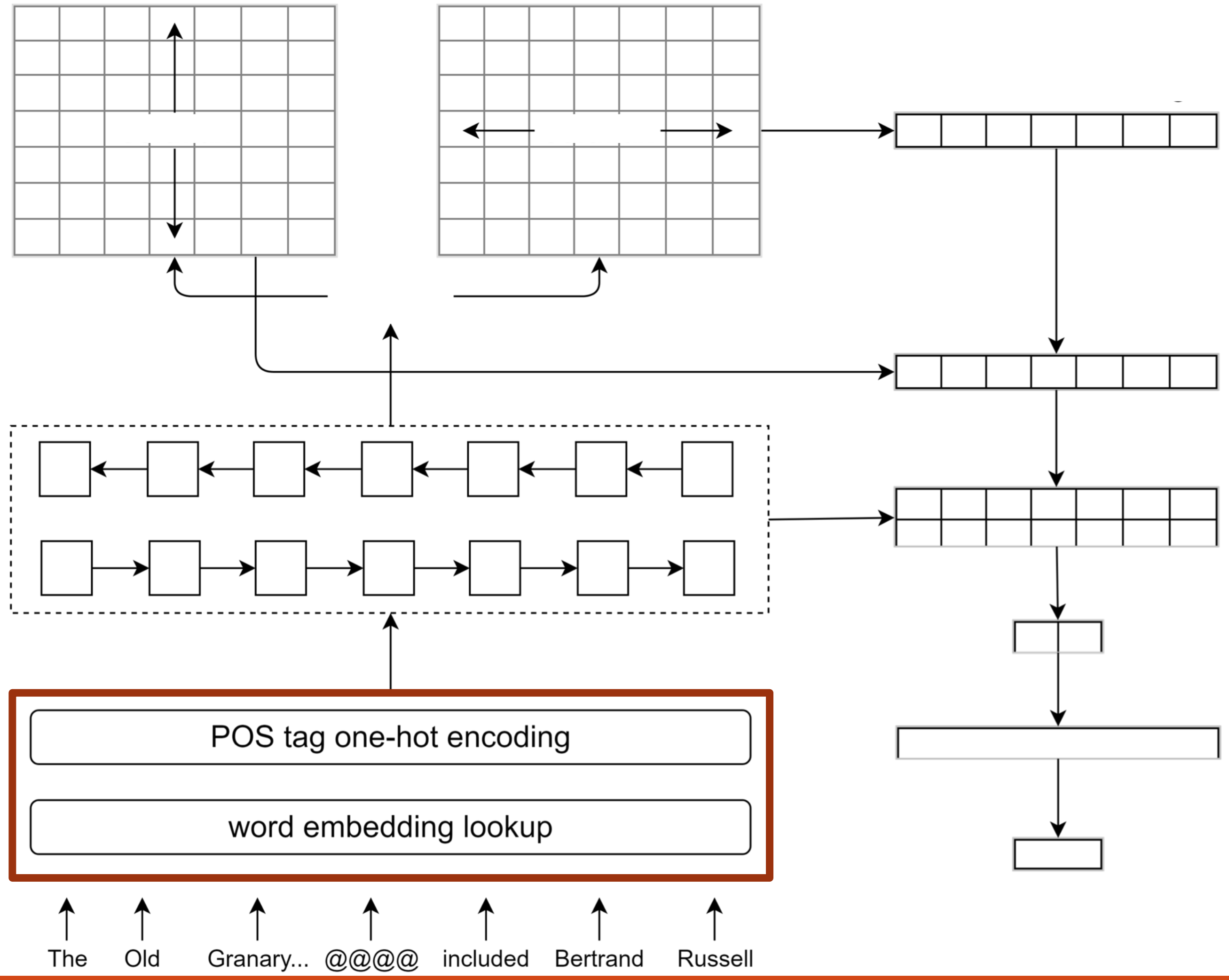
Learning Model: Overview



Learning Model: Input

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

Embedding + POS

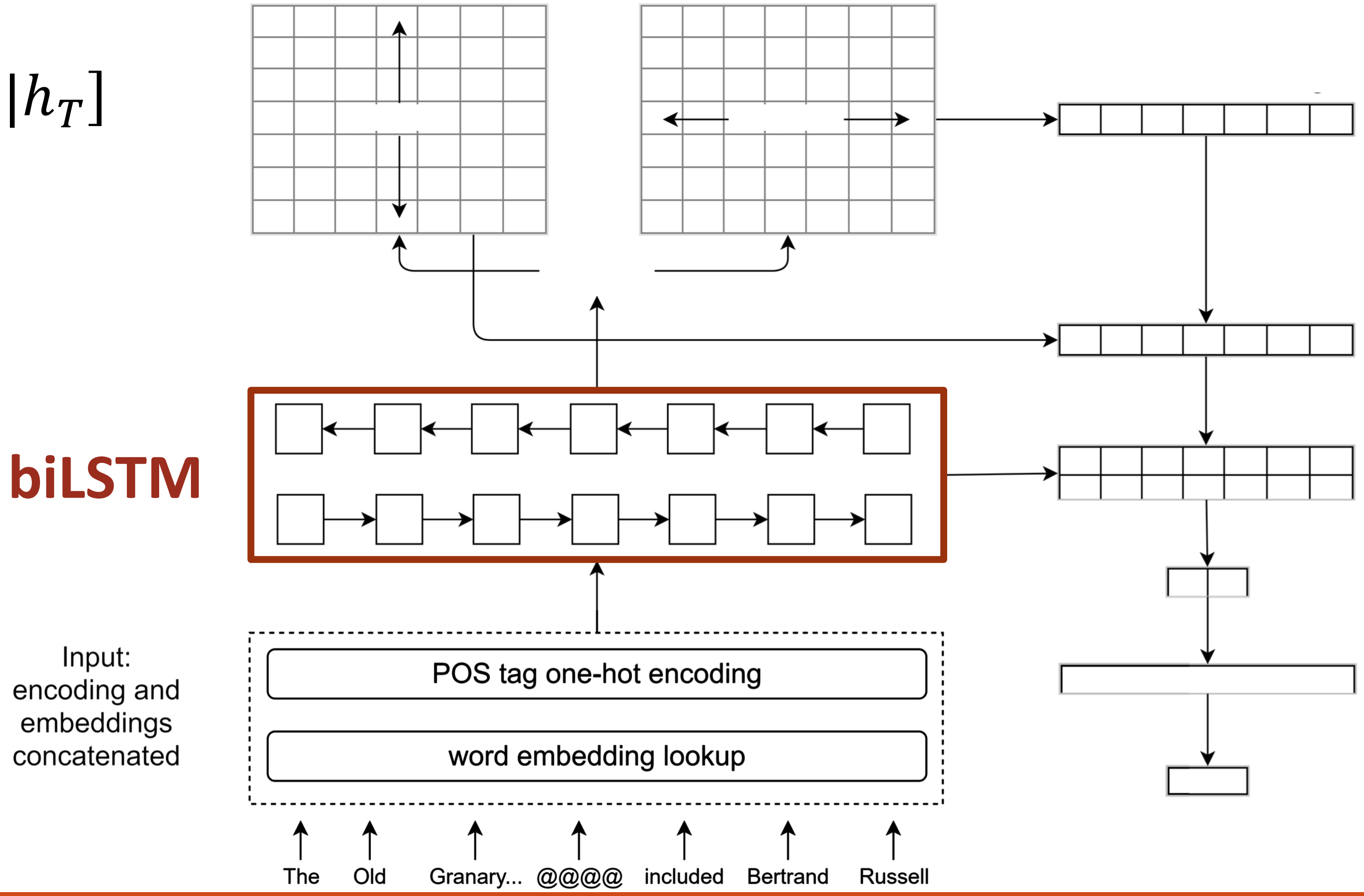


Learning Model: RNN

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

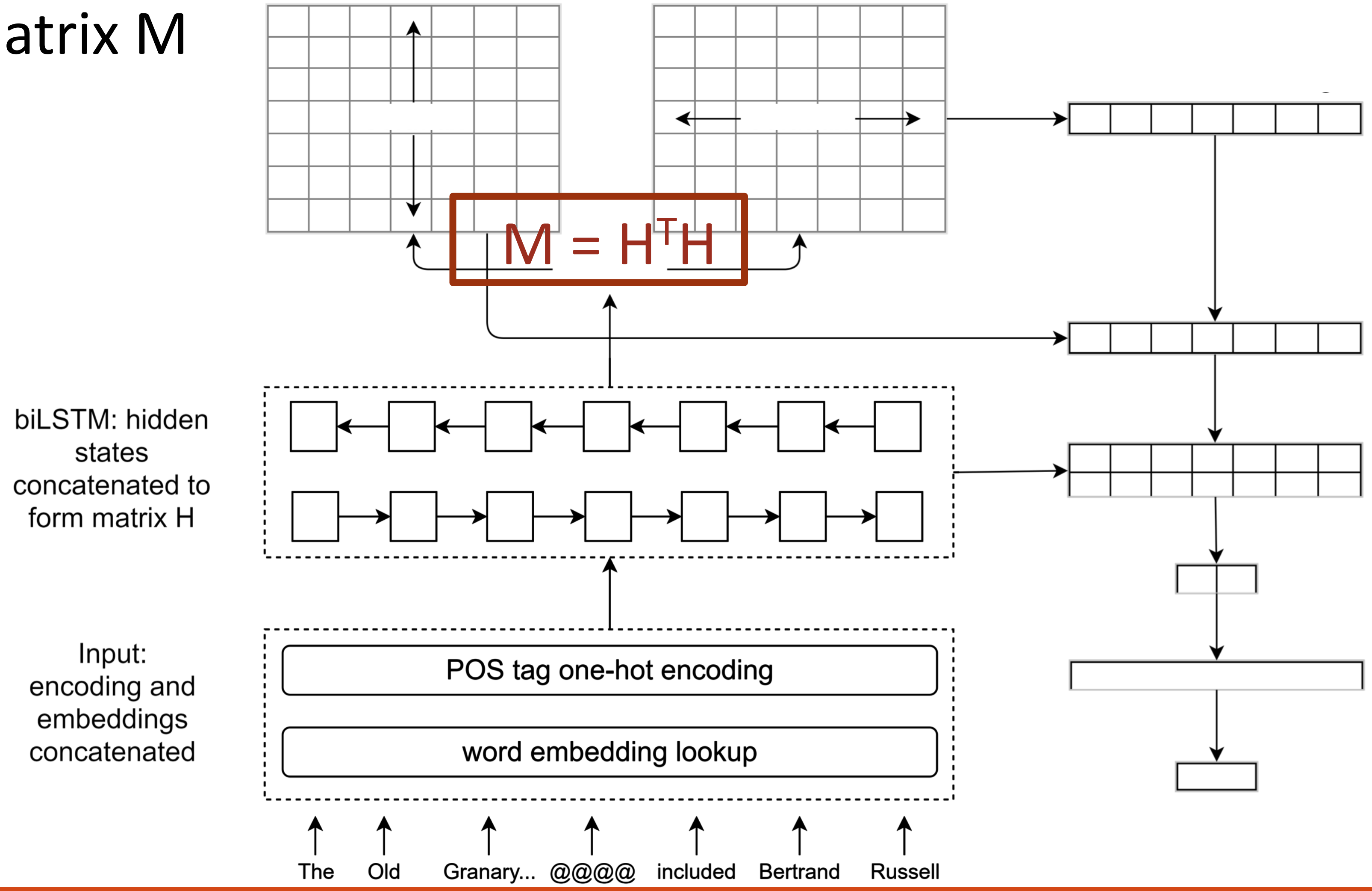
- E.g.:

We continue to feel that the stock market is the @@@@ place to be for long-term appreciation.



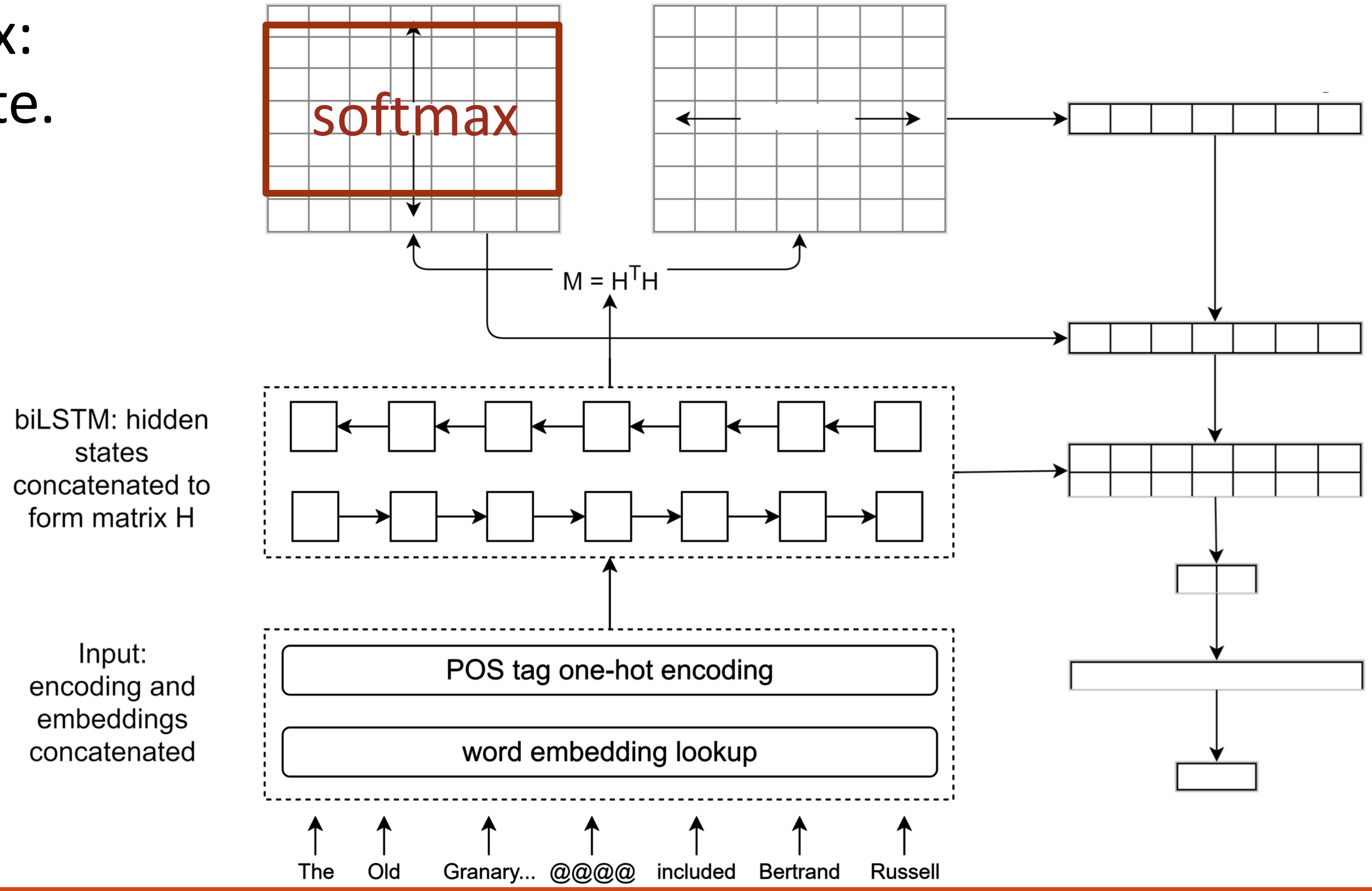
Learning Model: Matching Matrix

- Pair-wise matching matrix M



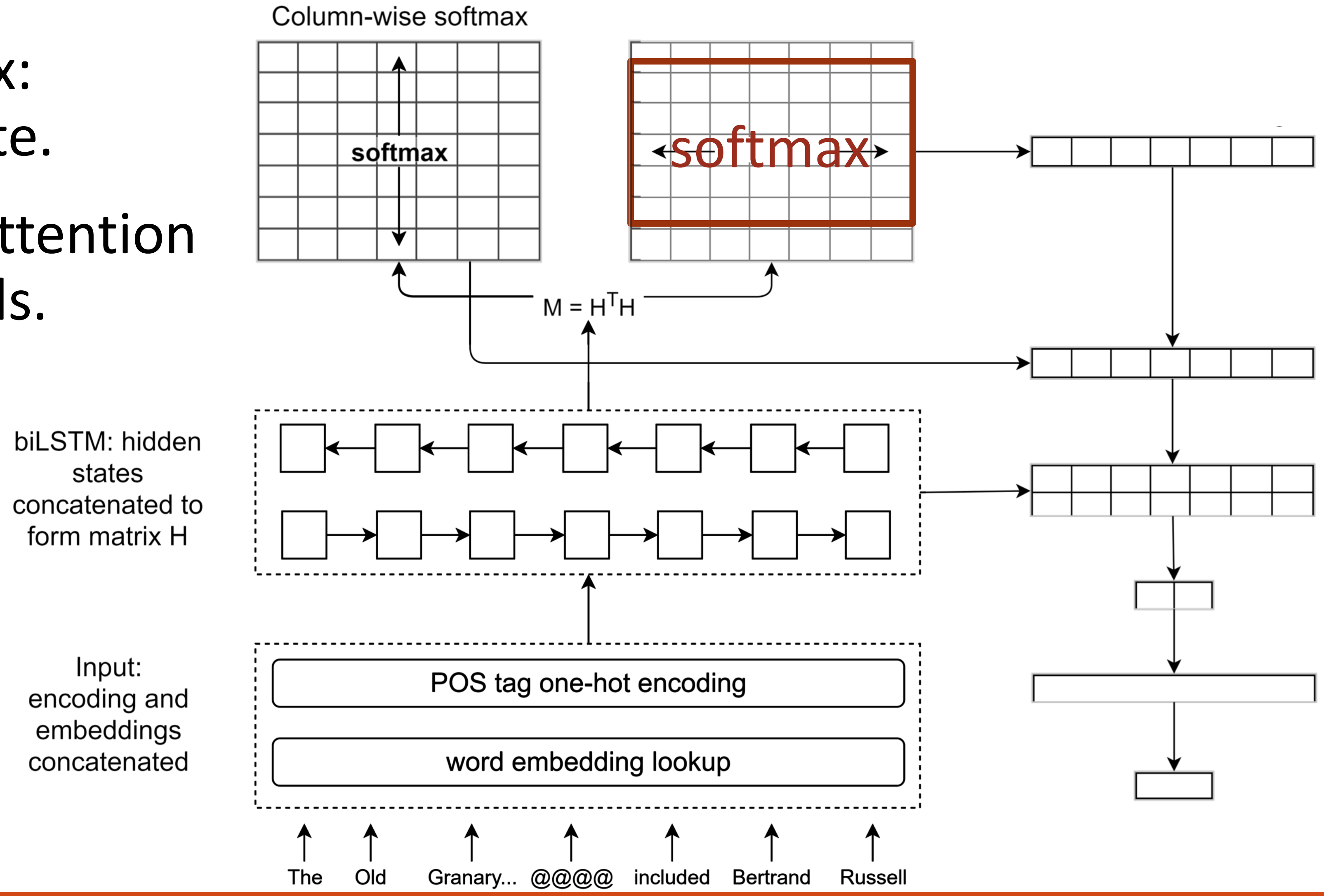
Learning Model: Softmax

- **Column-wise** softmax:
  Learn how to aggregate.



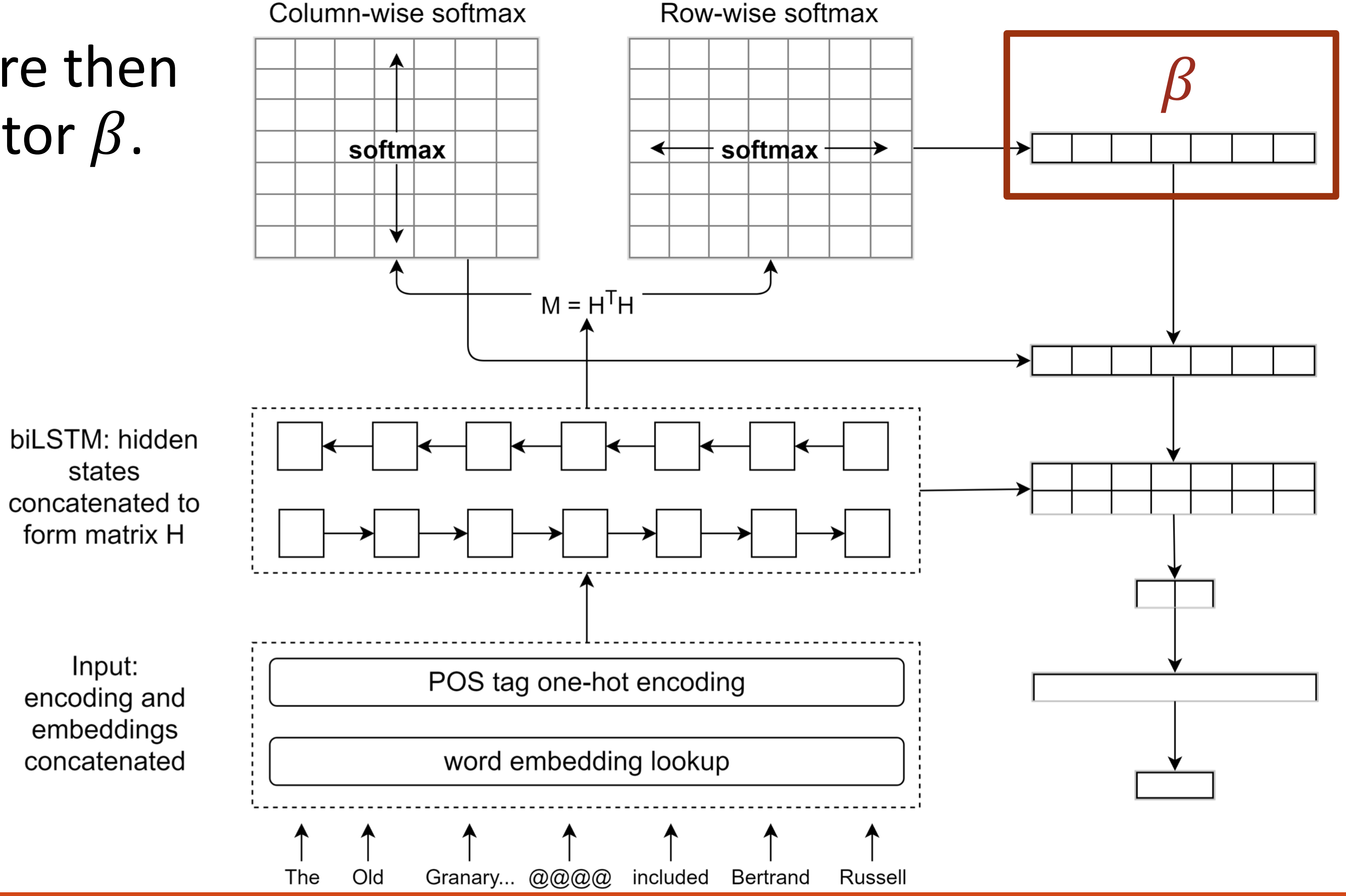
Learning Model: Softmax

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



Learning Model: Attention Score

- The columns of M^r are then **averaged**, forming vector β.

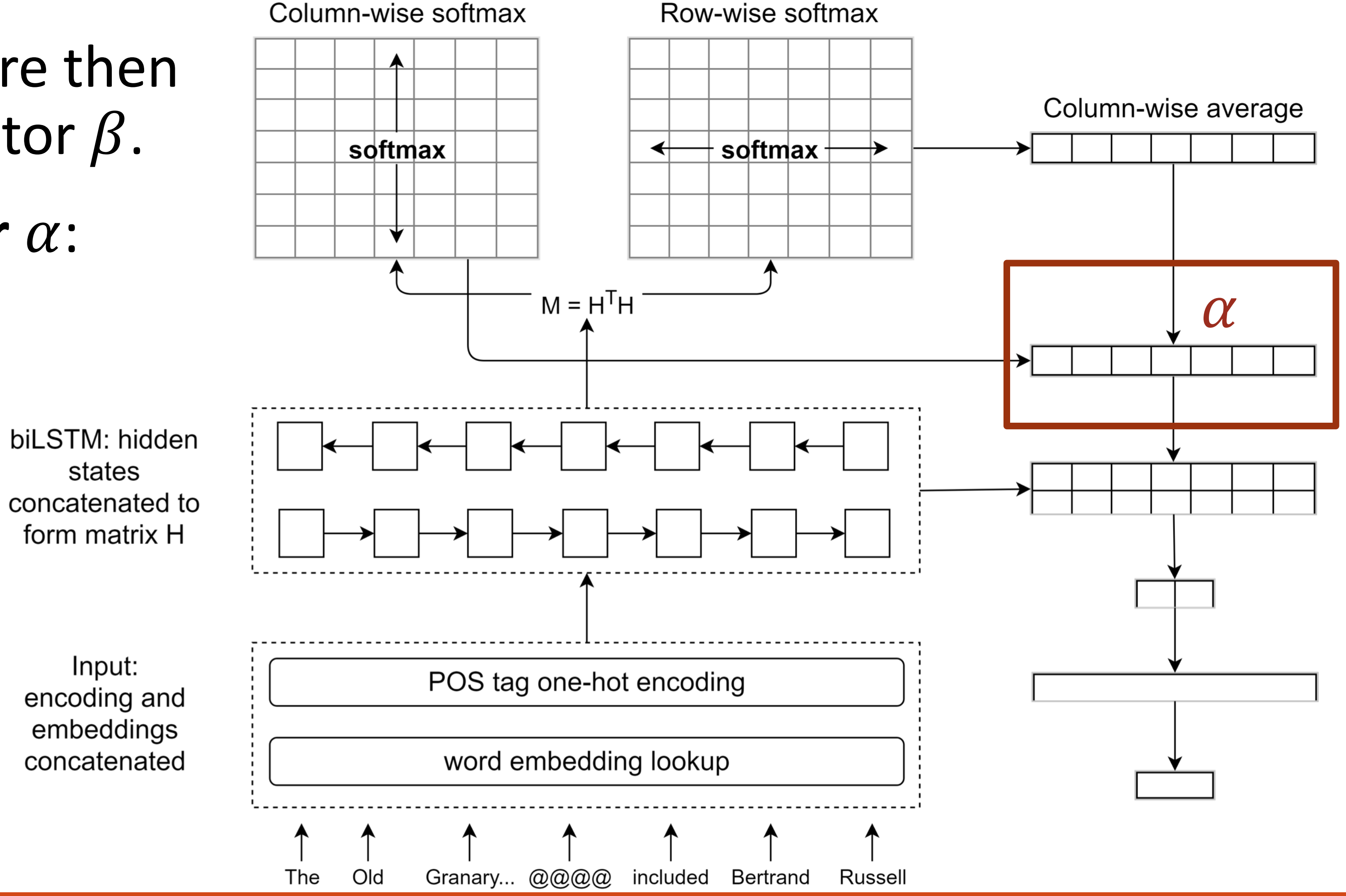


Learning Model: Attention Score

- The columns of M^r are then **averaged**, forming vector β.
- Final **attention vector** α:

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

based on (Cui et al., 2017).

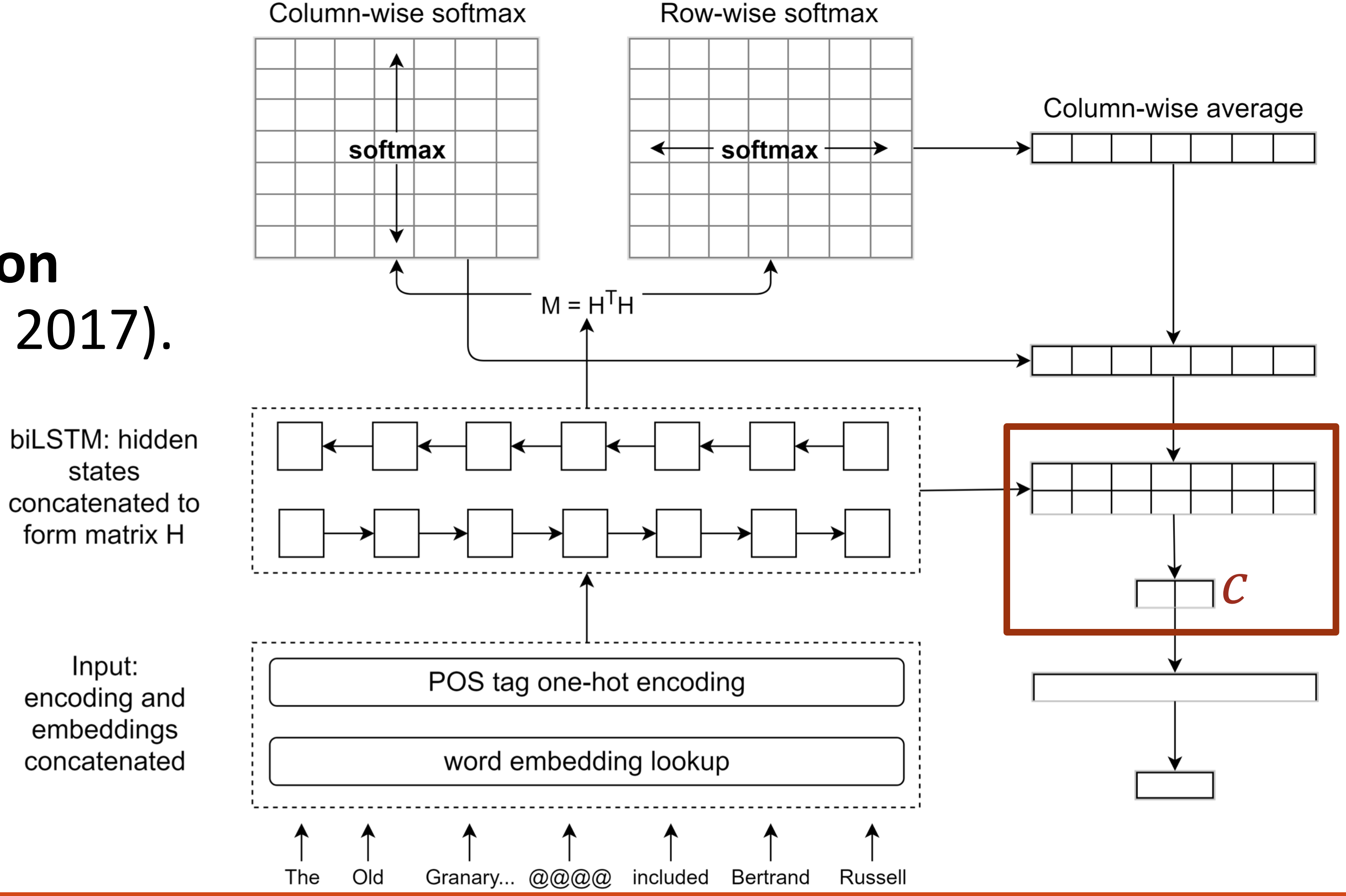


Learning Model: Attend

- **Attend**:

$$c = \sum_{t=1}^{T} \alpha_t h_t \,.$$

- A form of **self-attention** (Paulus 2017, Vaswani 2017).

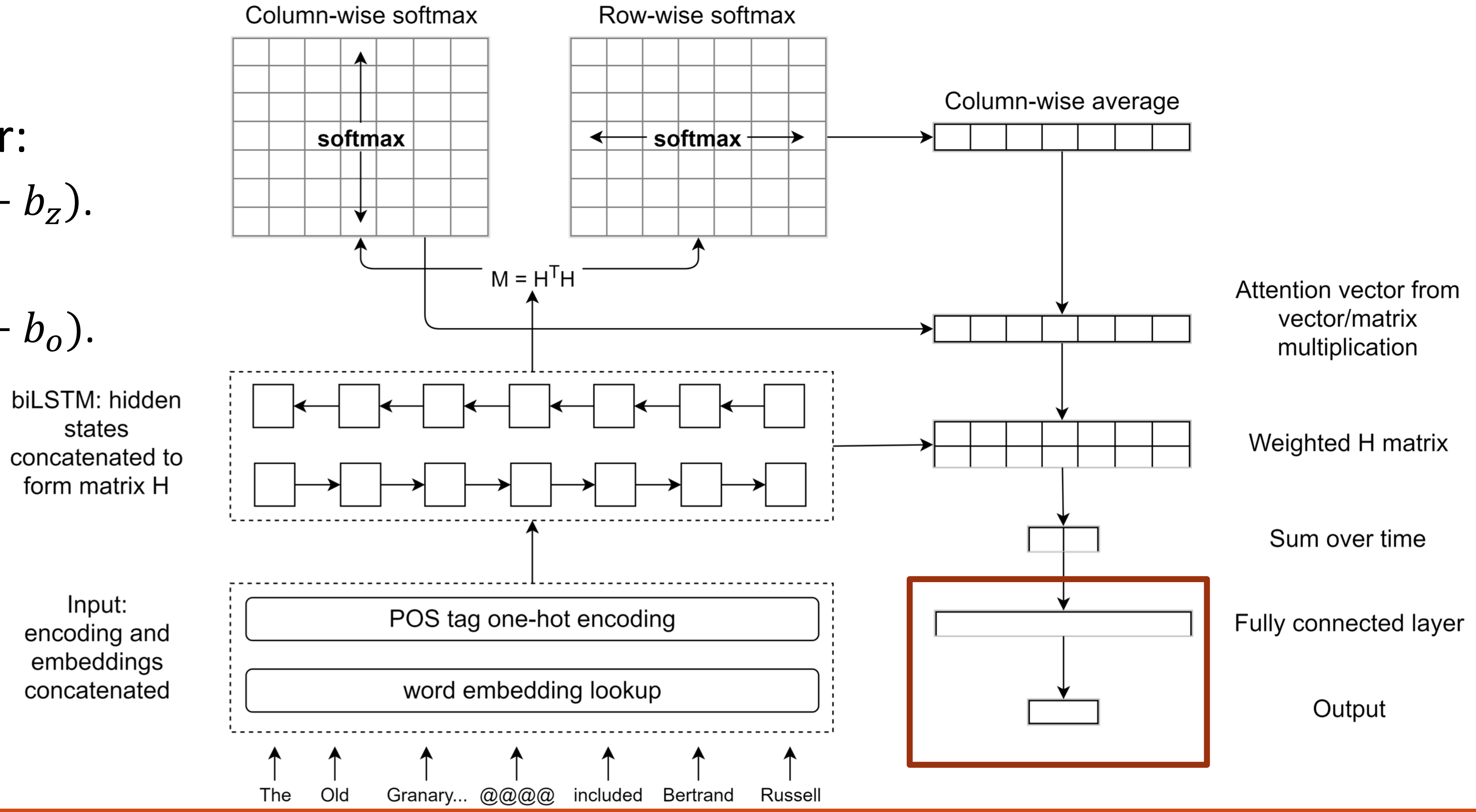


Learning Model: Predict

- **Predict**:
  - Dense layer:

    $z = \sigma(W_z c + b_z).$
  - Softmax:

    $y = s(W_o z + b_o).$



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.

Corpus	Training	Test
PTB	5,175	482
Gigaword yet	63,843	15840
Gigaword too	85,745	21501
Gigaword again	85,944	21762
Gigaword still	194,661	48741
Gigaword also	537,626	132928

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.

		WSJ - Accuracy
Models	Variants	All adverbs
MFC	-	51.66
LogReg	+ POS	52.81
	- POS	54.47
CNN	+ POS	58.84
	- POS	62.16
LSTM	+ POS	74.23
	- POS	73.18
WP	+ POS	**76.09**
	- POS	**74.84**

MFC: Most Frequent Class

LogReg: Logistic Regression

LSTM: bidirectional LSTM

CNN: Convolutional Network based on (Kim 2014)

Results – Gigaword

- Baselines

		Gigaword - Accuracy					
Models	Variants	All adverbs	Again	Still	Too	Yet	Also
MFC	-	50.24	50.25	50.29	65.06	50.19	50.32
LogReg	+ POS	53.65	59.49	56.36	69.77	**61.05**	52.00
	- POS	52.86	58.60	55.29	67.60	**58.60**	56.07
CNN	+ POS	59.12	60.26	59.54	67.53	59.69	61.53
	- POS	57.21	57.28	56.95	67.84	56.53	59.76

Results – Gigaword

- LSTM and LSTM with Attention (WP)

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

Results – Gigaword

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

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

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.

McGill
School of Computer Science

Thank you! ☺

Thank you to our co-authors:

Yulan Feng

Prof. Jackie CK Cheung

Thank you to our sponsors:

CRIM

NSERC
CRSNG

Fonds de recherche
sur la nature
et les technologies
Québec