Multi-Sentence Argument Linking

Seth Ebner*, Patrick Xia*, Ryan Culkin, Kyle Rawlins, Benjamin Van Durme
Motivation
More than 50 protesters were gunned down in the Maidan, the center of the popular uprising.

Traditional SRL:
in-sentence

<table>
<thead>
<tr>
<th>Firearm-Attack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attacker:  ∅</td>
</tr>
<tr>
<td>Target:  More than 50 protesters</td>
</tr>
<tr>
<td>Instrument:  ∅</td>
</tr>
<tr>
<td>Place:  Maidan</td>
</tr>
</tbody>
</table>
More than 50 protesters were gunned down in the Maidan, the center of the popular uprising. Until now the identity of the killers has been a mystery.
Cross-Sentence Links are Common
Cross-Sentence Links are Common

Beyond NomBank (Gerber and Chai, 2012)

71% relative increase in role coverage on NomBank
Cross-Sentence Links are Common

Beyond NomBank (Gerber and Chai, 2012)

71% relative increase in role coverage on NomBank

DARPA AIDA Phase 1

38% of events have an argument outside the same sentence as the trigger
Cross-Sentence Links are Common

Beyond NomBank (Gerber and Chai, 2012)
71% relative increase in role coverage on NomBank

DARPA AIDA Phase 1
38% of events have an argument outside the same sentence as the trigger

90% of arguments can be found within 2 sentences of the trigger
Motivation

Data: Roles Across Multiple Sentences

Model

Results
Roles Across Multiple Sentences
Roles Across Multiple Sentences

Based on DARPA AIDA event ontology
Roles Across Multiple Sentences

Based on DARPA AIDA event ontology

9K+ examples from 4K news articles
Roles Across Multiple Sentences

Based on DARPA AIDA event ontology

9K+ examples from 4K news articles

Annotators mark spans in 5-sentence context window
Roles Across Multiple Sentences

Based on DARPA AIDA event ontology

9K+ examples from 4K news articles

Annotators mark spans in 5-sentence context window

Protocol and analysis in the paper!
When Russian aircraft bombed a remote garrison in southeastern Syria last month, alarm bells sounded at the Pentagon and the Ministry of Defense in London.

The Russians weren’t bombing a run-of-the-mill rebel outpost, according to U.S. officials.
Broad + Diverse Coverage of Ontology
Broad + Diverse Coverage of Ontology
Broad + Diverse Coverage of Ontology

![Graphs showing frequency and CDF of event types with different methods: RAMS train, AIDA-1, BNB.](image)
Motivation

Data: Roles Across Multiple Sentences

Model

Results
Approach
Approach

Roles are evoked by event triggers, forming *implicit arguments* (implicit discourse referents)
Approach

Roles are evoked by event triggers, forming *implicit arguments* (implicit discourse referents)

the **place** of a *particular* attack event
Approach

Roles are evoked by event triggers, forming *implicit arguments* (implicit discourse referents)

the place of a *particular* attack event

Implicit arguments linked to explicit mentions in text
Model
Model

1) Learn span representations for each trigger span and candidate argument span
Model

1) Learn span representations for each trigger span and candidate argument span

2) For each trigger, prune to top-\(k\) candidate arguments
Model

1) Learn span representations for each trigger span and candidate argument span

2) For each trigger, prune to top-\(k\) candidate arguments

3) Score reps of implicit arguments with reps of explicit arguments using a decomposable scoring function
Enumerate candidate argument spans up to a fixed width
Create argument representations for every span

Span representation:
- BiLSTM hidden states at start and end of span (contextualized and type-level embeddings)
- Soft “head word” vector from attention mask over type-level embeddings
- Learned representation of span width (Lee et al., 2017; i.a.)
Prune to find good spans

Score each representation, and keep the top $\lambda_A n$

$n$: document length
$\lambda_A$: hyperparameter
Create event trigger representations for every span

Span representation:
- BiLSTM hidden states at start and end of span (contextualized and type-level embeddings)
- Soft “head word” vector from attention mask over type-level embeddings
- Learned representation of span width (Lee et al., 2017; i.a.)
Score each representation, and keep the top $\lambda_E n$.

$n$: document length

$\lambda_E$: hyperparameter

Prune to find good spans
Further prune to find good arguments for a particular event
Further prune to find good arguments for a particular event
Trigger representation

Argument representation

Role embedding

Implicit argument representation

\[ l(a, \tilde{a}_e, r) = s_{E,R}(e, r) + s_{A,R}(a, r) + s_l(a, \tilde{a}_e, r) + s_c(e, a), \quad a \neq \epsilon \]
Motivation

Data: Roles Across Multiple Sentences

Model

Results
RAMS Results

F1

- Most common
- Fixed trigger
- Context as trigger
RAMS Results vs. Distance
RAMS Results vs. Distance

Performance does not degrade over distance

![Graph showing F1 scores over different distances.](image)
Koum grew up in the Ukraine under Soviet rule before immigrating to the US as a teenager…
ST. LOUIS (KTVI) – The St. Circuit Attorney’s Office charged a 50-year-old Friday for allegedly gunning another man down in a north city neighborhood back in December. According to Schron Jackson, a spokeswoman for the St. Louis Metropolitan Police Department, the shooting happened at 12:10 p.m. on December 2, 2015 in the 3000 block of N. 20th Street in the St. Louis Place neighborhood. The victim, identified as 32-year-old Leonard Arnold, was found lying unconscious outside. Arnold had been shot multiple times and was pronounced dead at the scene, Jackson said. The suspect, Maurice Alexander, was charged with first-degree murder, first-degree assault, and two counts of armed criminal action. …

Victim Name: Leonard Arnold
Shooter Name: Maurice Alexander
Location: 3000 block of N. 20th Street
Weapon: ∅
# Shots: shot multiple times
Summary
Summary

Model finds argument spans anywhere in a document that fill roles for events
Summary

Model finds argument spans anywhere in a document that fill roles for events

RAMS dataset has broader coverage than prior datasets
Summary

Model finds argument spans anywhere in a document that fill roles for events

RAMS dataset has broader coverage than prior datasets

Model works well on other tasks too (more results in the paper!)
Thanks!

Data, code, and models

nlp.jhu.edu/rams
Additional Slides

Note: these slides may be either outdated or even incorrect (based on much earlier versions of the model or buggy graphing code). Use at your own risk.
Motivation

{ Communicator, Recipient, Place }  { Attacker, Target, Instrument, Place }
Roles Across Multiple Sentences (RAMS)

Gloss

the banks lent $380 million to Trump for the benefit of Trump at someplace

Context

Meanwhile, another Trump disaster was brewing.

Eastern Air Lines, which had been struggling, put its northeastern air shuttle up for sale.

Trump persuaded the banks to lend him $380 million to purchase the route, and in June 1989 the Trump Shuttle began flying.

Trump introduced the airline with his usual style — by insulting the competition.

At an elegant event at Logan Airport in Boston, Trump took the stage and suggested that the other airline with a northeastern shuttle, Pan Am, few unsafe planes.

Event type: Transaction,TransferMoney,BorrowLend

Event definition: The borrowing or lending of money

Arguments

<table>
<thead>
<tr>
<th>Role</th>
<th>Argument</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giver</td>
<td>someone or something</td>
<td>the banks</td>
</tr>
<tr>
<td>Recipient</td>
<td>someone or something</td>
<td>Trump</td>
</tr>
<tr>
<td>Beneficiary</td>
<td>someone</td>
<td>Trump</td>
</tr>
<tr>
<td>Money</td>
<td>some money</td>
<td>$380 million</td>
</tr>
<tr>
<td>Place</td>
<td>someplace</td>
<td></td>
</tr>
</tbody>
</table>
RAMS Results

- Our model
- + type-constrained decoding

Metrics:
- P
- R
- F1
Error Analysis
## Ablations

<table>
<thead>
<tr>
<th>Model</th>
<th>Greedy</th>
<th>TCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our model</td>
<td>69.9</td>
<td>75.1</td>
</tr>
<tr>
<td>- distance score</td>
<td>69.0</td>
<td>74.3</td>
</tr>
<tr>
<td>- ( s_l(a, \tilde{a}_e, r) )</td>
<td>54.9</td>
<td>58.4</td>
</tr>
<tr>
<td>- ( s_{A,R}(a, r) )</td>
<td>68.6</td>
<td>73.8</td>
</tr>
<tr>
<td>+ ( s_{E,R}(e, r) )</td>
<td>69.5</td>
<td>74.4</td>
</tr>
<tr>
<td>+ ( s_c(e, a) )</td>
<td>65.9</td>
<td>70.6</td>
</tr>
<tr>
<td>w/ argmax decoding</td>
<td>69.9</td>
<td>75.1</td>
</tr>
<tr>
<td>BERT 6–9</td>
<td>69.6</td>
<td><strong>75.3</strong></td>
</tr>
<tr>
<td>ELMo</td>
<td>68.5</td>
<td>75.2</td>
</tr>
</tbody>
</table>

Table 3: \( F_1 \) on RAMS dev data when link score components are separately included/excluded (Equation 2) or other contextualized encoders are used in the best performing model. TCD = type-constrained decoding.
Ablations

\[ l(a, \tilde{a}_{e,r}) = s_{E,R}(e,r) + s_{A,R}(a,r) + s_{l}(a, \tilde{a}_{e,r}) + s_{c}(e,a), \quad a \neq \epsilon \]
RAMS Results vs. Distance

<table>
<thead>
<tr>
<th>Dist.</th>
<th># Gold</th>
<th># Predict</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>79</td>
<td>69</td>
<td>81.2</td>
<td>70.9</td>
<td><strong>75.7</strong></td>
</tr>
<tr>
<td>-1</td>
<td>164</td>
<td>151</td>
<td>76.8</td>
<td>70.7</td>
<td><strong>73.7</strong></td>
</tr>
<tr>
<td>0</td>
<td>1,811</td>
<td>1,688</td>
<td>77.7</td>
<td>72.4</td>
<td><strong>75.0</strong></td>
</tr>
<tr>
<td>1</td>
<td>87</td>
<td>83</td>
<td>78.3</td>
<td>74.7</td>
<td><strong>76.5</strong></td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>39</td>
<td>87.2</td>
<td>72.3</td>
<td><strong>79.1</strong></td>
</tr>
<tr>
<td>Total</td>
<td>2,189</td>
<td>2,030</td>
<td>78.0</td>
<td>72.3</td>
<td><strong>75.1</strong></td>
</tr>
</tbody>
</table>

Table 4: Performance breakdown by distance (number of sentences) between argument and event trigger for our model using TCD over the development data.

Performance doesn’t degrade over distance
RAMS works as pre-training

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Dev. F₁</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>No pre-training</td>
<td>25.0</td>
<td>36.6</td>
<td>12.9</td>
<td>19.1</td>
</tr>
<tr>
<td>No pre-training&lt;sup&gt;T&lt;/sup&gt;</td>
<td>27.1</td>
<td>53.5</td>
<td>11.0</td>
<td>18.2</td>
</tr>
<tr>
<td>RAMS pre-training</td>
<td>34.1</td>
<td>43.9</td>
<td>16.9</td>
<td>24.4</td>
</tr>
<tr>
<td>RAMS pre-training&lt;sup&gt;T&lt;/sup&gt;</td>
<td>34.2</td>
<td>62.5</td>
<td>15.4</td>
<td>24.8</td>
</tr>
</tbody>
</table>

Table 8: P(reision), R(ecall), and F₁ on AIDA-1 En-glish development and test data. T designates the use of ontology-aware type-constrained decoding.
Figure 5: Distances between triggers and arguments in RAMS and proportion of arguments at that distance (counts are shown above each bar). Negative distances indicate that the argument occurs before the trigger.
<table>
<thead>
<tr>
<th>Threshold</th>
<th>Conjunctive</th>
<th>Disjunctive</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>55.3</td>
<td>78.0</td>
<td>59.8</td>
<td>73.5</td>
</tr>
<tr>
<td>1</td>
<td>69.9</td>
<td>80.3</td>
<td>74.9</td>
<td>75.3</td>
</tr>
<tr>
<td>2</td>
<td>73.9</td>
<td>82.0</td>
<td>78.2</td>
<td>77.8</td>
</tr>
<tr>
<td>3</td>
<td>76.4</td>
<td>83.6</td>
<td>80.9</td>
<td>79.1</td>
</tr>
<tr>
<td>4</td>
<td>78.8</td>
<td>84.3</td>
<td>82.7</td>
<td>80.4</td>
</tr>
</tbody>
</table>

Table 6: Pairwise span boundary inter-annotator agreement statistics for various span difference thresholds.
Figure 8: Number of event types for which a given percentage of roles are filled in RAMS train set.
## RAMS Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Dev. $F_1$</th>
<th>$P$</th>
<th>$R$</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our model</td>
<td>69.9</td>
<td>62.8</td>
<td>74.9</td>
<td>68.3</td>
</tr>
<tr>
<td>Our model$^{TCD}$</td>
<td>75.1</td>
<td>78.1</td>
<td>69.2</td>
<td>73.3</td>
</tr>
<tr>
<td>Most common</td>
<td>17.3</td>
<td>15.7</td>
<td>15.7</td>
<td>15.7</td>
</tr>
<tr>
<td>Fixed trigger$^{TCD}$</td>
<td>60.2</td>
<td>83.7</td>
<td>41.9</td>
<td>55.8</td>
</tr>
<tr>
<td>Context as trigger$^{TCD}$</td>
<td>62.1</td>
<td>80.5</td>
<td>45.8</td>
<td>58.4</td>
</tr>
</tbody>
</table>

Table 2: $P$(recision), $R$(ecall), and $F_1$ on RAMS development and test data. $TCD$ designates the use of ontology-aware type-constrained decoding.
<table>
<thead>
<tr>
<th>Model</th>
<th>Dev. $F_1$</th>
<th>$P$</th>
<th>$R$</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our model</td>
<td>69.9</td>
<td>62.8</td>
<td>74.9</td>
<td>68.3</td>
</tr>
<tr>
<td>Our model$^{TCD}$</td>
<td>75.1</td>
<td>78.1</td>
<td>69.2</td>
<td>73.3</td>
</tr>
<tr>
<td>Most common</td>
<td>17.3</td>
<td>15.7</td>
<td>15.7</td>
<td>15.7</td>
</tr>
<tr>
<td>Fixed trigger$^{TCD}$</td>
<td>60.2</td>
<td>83.7</td>
<td>41.9</td>
<td>55.8</td>
</tr>
<tr>
<td>Context as trigger$^{TCD}$</td>
<td>62.1</td>
<td>80.5</td>
<td>45.8</td>
<td>58.4</td>
</tr>
</tbody>
</table>

Table 2: $P$recision, $R$ecall, and $F_1$ on RAMS development and test data. $TCD$ designates the use of ontology-aware type-constrained decoding.
# RAMS Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Dev. F$_1$</th>
<th>P</th>
<th>R</th>
<th>F$_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our model</td>
<td>69.9</td>
<td>62.8</td>
<td>74.9</td>
<td>68.3</td>
</tr>
<tr>
<td>Our model$^{\text{TCD}}$</td>
<td>75.1</td>
<td>78.1</td>
<td>69.2</td>
<td>73.3</td>
</tr>
<tr>
<td>Most common</td>
<td>17.3</td>
<td>15.7</td>
<td>15.7</td>
<td>15.7</td>
</tr>
<tr>
<td>Fixed trigger$^{\text{TCD}}$</td>
<td>60.2</td>
<td>83.7</td>
<td>41.9</td>
<td>55.8</td>
</tr>
<tr>
<td>Context as trigger$^{\text{TCD}}$</td>
<td>62.1</td>
<td>80.5</td>
<td>45.8</td>
<td>58.4</td>
</tr>
</tbody>
</table>

Table 2: P(reiction), R(ecall), and F$_1$ on RAMS development and test data. TCD designates the use of ontology-aware type-constrained decoding.
Roles Across Multiple Sentences (RAMS)

**Context**
Meanwhile, another Trump disaster was brewing.
Eastern Air Lines, which had been struggling, put its northeastern air shuttle up for sale.
Trump persuaded the banks to lend him $380 million to purchase the route, and in June 1989 the Trump Shuttle began flying.

At an elegant event at Logan Airport in Boston, Trump took the stage and suggested that the other airline with a northeastern shuttle, Pan Am, flew unsafe planes.

**Event type:** Transaction, TransferMoney, BorrowLend

**Event definition:** The borrowing or lending of money

**Arguments**
- **Giver:** someone or something, the banks
- **Recipient:** someone or something, Trump
- **Beneficiary:** someone, Trump
- **Money:** some money, $380 million
- **Place:** someplace
Roles Across Multiple Sentences (RAMS)

Gloss
the banks lent $380 million to Trump for the benefit of Trump at someplace

Context
Meanwhile, another Trump disaster was brewing.
Eastern Air Lines, which had been struggling, put its northeastern air shuttle up for sale.

Trump persuaded the banks to lend him $380 million to purchase the route, and in June 1989 the Trump Shuttle began flying.

Trump introduced the airline with his usual style — by insulting the competition.

At an elegant event at Logan Airport in Boston, Trump took the stage and suggested that the other airline with a northeastern shuttle, Pan Am, few unsafe planes.

Event type: Transaction,TransferMoney,BorrowLend
Event definition: The borrowing or lending of money

Arguments

<table>
<thead>
<tr>
<th>Role</th>
<th>Argument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giver</td>
<td>someone or something the banks</td>
</tr>
<tr>
<td>Recipient</td>
<td>someone or something Trump</td>
</tr>
<tr>
<td>Beneficiary</td>
<td>someone Trump</td>
</tr>
<tr>
<td>Money</td>
<td>some money $380 million</td>
</tr>
<tr>
<td>Place</td>
<td>someplace</td>
</tr>
</tbody>
</table>

- Giver: the banks
- Recipient: Trump
- Beneficiary: Trump
- Money: $380 million
- Place: someplace
Roles Across Multiple Sentences (RAMS)

Gloss
the banks lent $300 million to Trump for the benefit of Trump at someplace

Context
Meanwhile, another Trump disaster was brewing.
Eastern Air Lines, which had been struggling, put its northeastern air shuttle up for sale.
Trump persuaded the banks to lend him $300 million to purchase the route, and in June 1989 the Trump Shuttle began flying.
Trump introduced the airline with his usual style—by insulting the competition.
At an elegant event at Logan Airport in Boston, Trump took the stage and suggested that the other airline with a northeastern shuttle, Pan Am, had unsafe planes.

Event type: Transaction, TransferMoney, BorrowLend

Event definition: The borrowing or lending of money

Arguments

<table>
<thead>
<tr>
<th>Role</th>
<th>Argument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giver</td>
<td>the banks</td>
</tr>
<tr>
<td>Recipient</td>
<td>Trump</td>
</tr>
<tr>
<td>Beneficiary</td>
<td>Trump</td>
</tr>
<tr>
<td>Money</td>
<td>$300 million</td>
</tr>
<tr>
<td>Place</td>
<td>someplace</td>
</tr>
</tbody>
</table>
Roles Across Multiple Sentences (RAMS)

**Gloss**

the banks lent $380 million to Trump for the benefit of Trump at someplace

**Context**

Meanwhile, another Trump disaster was brewing.

Eastern Air Lines, which had been struggling, put its northeastern air shuttle up for sale.

Trump persuaded the banks to lend him $380 million to purchase the route, and in June 1989 the Trump Shuttle began flying.

Trump introduced the airline with his usual style — by insulting the competition.

At an elegant event at Logan Airport in Boston, Trump took the stage and suggested that the other airline with a northeastern shuttle, Pan Am, flew unsafe planes.

**Event type:** Transaction.TransferMoney.BorrowLend

**Event definition:** The borrowing or lending of money

**Arguments**

<table>
<thead>
<tr>
<th>Role</th>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giver</td>
<td>someone or something</td>
<td>the banks</td>
</tr>
<tr>
<td>Recipient</td>
<td>someone or something</td>
<td>Trump</td>
</tr>
<tr>
<td>Beneficiary</td>
<td>someone</td>
<td>Trump</td>
</tr>
<tr>
<td>Money</td>
<td>some money</td>
<td>$380 million</td>
</tr>
<tr>
<td>Place</td>
<td>someplace</td>
<td></td>
</tr>
</tbody>
</table>

 □ Argument not present
Roles Across Multiple Sentences (RAMS)

Gloss
the banks lent $380 million to Trump for the benefit of Trump at someplace

Context
Meanwhile, another Trump disaster was brewing.
Eastern Air Lines, which had been struggling, put its northeastern air shuttle up for sale.
Trump persuaded the banks to lend him $380 million to purchase the route, and in June 1989 the Trump Shuttle began flying.
Trump introduced the airline with his usual style — by insulting the competition.
At an elegant event at Logan Airport in Boston, Trump took the stage and suggested that the other airline with a northeastern shuttle, Pan Am, few unsafe planes.

Event type: Transaction,TransferMoney,BorrowLend
Event definition: The borrowing or lending of money

Arguments
Giver
someone or something
the banks
Recipient
someone or something
Trump
Beneficiary
someone
Trump
Money
some money
$380 million
Place
someplace
Roles Across Multiple Sentences (RAMS)

**Context**

Meanwhile, another Trump disaster was brewing.

Eastern Air Lines, which had been struggling, put its northeastern air shuttle up for sale.

Trump persuaded the banks to lend him $380 million to purchase the route, and in June 1989 the Trump Shuttle began flying.

Trump introduced the airline with his usual style — by insulting the competition.

At an elegant event at Logan Airport in Boston, Trump took the stage and suggested that the other airline with a northeastern shuttle, Pan Am, flew unsafe planes.

**Event type:** Transaction, TransferMoney, BorrowLend

**Event definition:** The borrowing or lending of money

**Arguments**

<table>
<thead>
<tr>
<th>Role</th>
<th>Argument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giver</td>
<td>someone or something</td>
</tr>
<tr>
<td>Recipient</td>
<td>the banks</td>
</tr>
<tr>
<td>Beneficiary</td>
<td>Trump</td>
</tr>
<tr>
<td>Money</td>
<td>some money</td>
</tr>
<tr>
<td>Place</td>
<td>someplace</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Role</th>
<th>Argument</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$380 million</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Role</th>
<th>Argument</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Argument not present</td>
</tr>
</tbody>
</table>