

Graphene: Semantically-Linked Propositions in Open Information Extraction

COLING 2018

Matthias Cetto¹, Christina Niklaus¹, André Freitas²
and Siegfried Handschuh¹

¹Natural Language Processing and Semantic Computing
University of Passau

²School of Computer Science
University of Manchester

August 23, 2018

“The Treasury will announce details of the November refunding.”



⟨The Treasury; will announce; details of the November refunding⟩

Key features of Open IE (Stanovsky and Dagan, 2016):

1. Assertedness
2. Minimal Propositions
3. Completeness and Open Lexicon

“The Treasury will announce details of the November refunding.”



⟨The Treasury; will announce; details of the November refunding⟩

Key features of Open IE (Stanovsky and Dagan, 2016):

1. Assertedness
2. Minimal Propositions
3. Completeness and Open Lexicon

“The Treasury will announce details of the November refunding.”



⟨The Treasury; will announce; details of the November refunding⟩

Key features of Open IE (Stanovsky and Dagan, 2016):

1. **Assertedness**
2. **Minimal Propositions**
3. **Completeness and Open Lexicon**

“Although the Treasury will announce details of the November refunding on Monday, the funding will be delayed if Congress and President Bush fail to increase the Treasury’s borrowing capacity.”



“Although the Treasury will announce details of the November refunding on Monday, the funding will be delayed if Congress and President Bush fail to increase the Treasury’s borrowing capacity.”



► **incorrect extractions**

⟨Congress and President Bush; increase; the Treasury’s borrowing capacity⟩

incorrect extractions → assertedness?

“Although the Treasury will announce details of the November refunding on Monday, the funding will be delayed if Congress and President Bush fail to increase the Treasury’s borrowing capacity.”



- ▶ **incorrect extractions**
- ▶ **missing context**

⟨ The funding; will be delayed; ⟩
missing context → assertedness?

“Although the Treasury will announce details of the November refunding on Monday, the funding will be delayed if Congress and President Bush fail to increase the Treasury’s borrowing capacity.”



- ▶ **incorrect extractions**
- ▶ **missing context**
- ▶ **long arguments**

⟨The funding; will be delayed; if Congress and President Bush fail to increase the Treasury’s borrowing capacity Although the Treasury will announce details of the November refunding on Monday⟩

long arguments → minimality?

"Although the Treasury will announce details of the November refunding on Monday, the funding will be delayed if Congress and President Bush fail to increase the Treasury's borrowing capacity."



- ▶ **incorrect extractions**
- ▶ **missing context**
- ▶ **long arguments**
- ▶ **missing relations**

⟨Bush; is; president⟩

missing relations → completeness?

Lightweight semantic representation of propositions:



idea

- ▶ two-layered hierarchy in the form of *core relational tuples* and accompanying *contextual information*
- ▶ *semantically linked* via rhetorical relations

Lightweight semantic representation of propositions:



idea

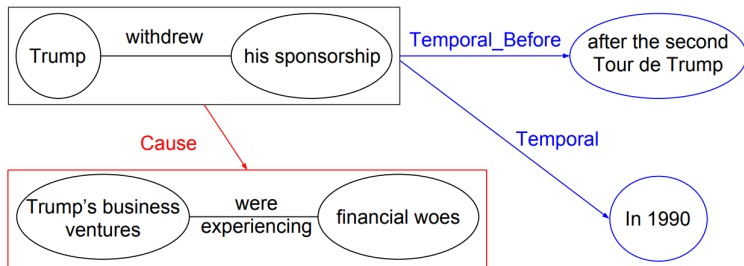
- ▶ two-layered hierarchy in the form of *core relational tuples* and accompanying *contextual information*
- ▶ *semantically linked* via rhetorical relations

Lightweight semantic representation of propositions:



- ▶ two-layered hierarchy in the form of *core relational tuples* and accompanying *contextual information*
- ▶ *semantically linked* via rhetorical relations

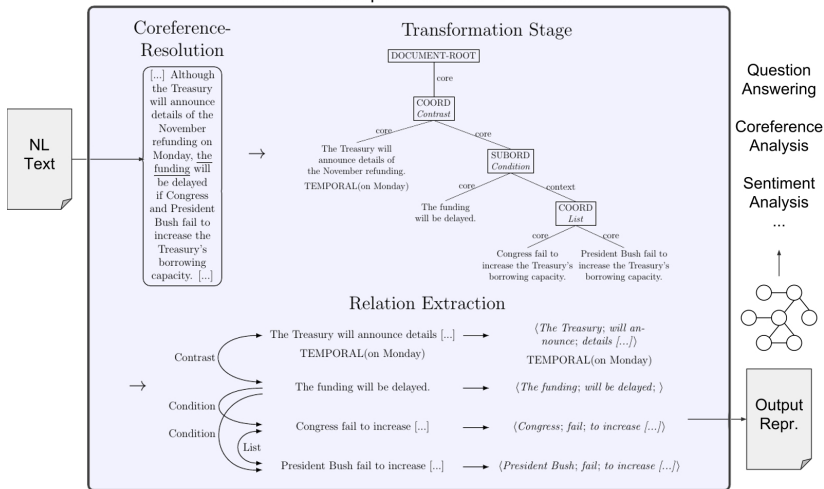
“Trump withdrew his sponsorship after the second Tour de Trump in 1990 because his business ventures were experiencing financial woes.”

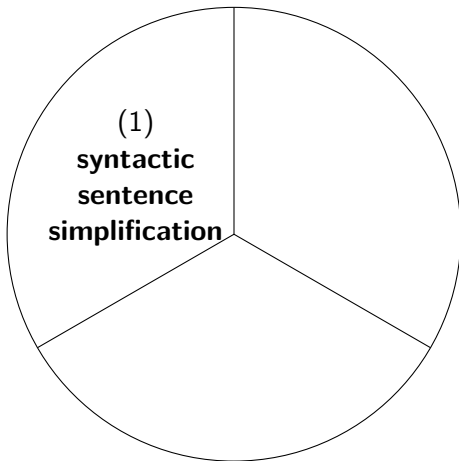


Framework for Open Information Extraction

Objective:
Generating semantically linked propositions

Framework for Open Information Extraction





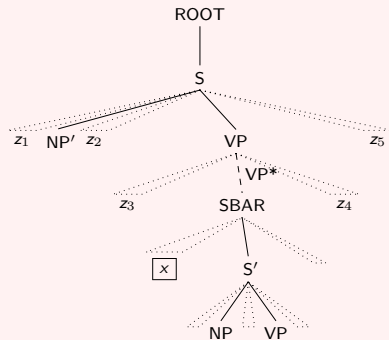
- ▶ **Recursive** transformation of syntactically complex sentences into simplified, compact structures using a **rule-based sentence splitting approach** for
 - ▶ clausal disembedding, and
 - ▶ phrasal disembedding
- ▶ *"Trump withdrew his sponsorship after the second Tour de Trump in 1990 because his business ventures were experiencing financial woes."*
 - ▶ CORE: *"T. withdrew his sponsorship."*
 - ▶ CONTEXT: *"His business ventures were experiencing financial woes."*
 - ▶ CONTEXT: *"This was after the second Tour de Trump."*
 - ▶ CONTEXT: *"This was in 1990."*

- ▶ **Recursive** transformation of syntactically complex sentences into simplified, compact structures using a **rule-based sentence splitting approach** for
 - ▶ clausal disembedding, and
 - ▶ phrasal disembedding
- ▶ “*Trump withdrew his sponsorship after the second Tour de Trump in 1990 because his business ventures were experiencing financial woes.*”
 - ▶ CORE: “*T. withdrew his sponsorship.*”
 - ▶ CONTEXT: “*His business ventures were experiencing financial woes.*”
 - ▶ CONTEXT: “*This was after the second Tour de Trump.*”
 - ▶ CONTEXT: “*This was in 1990.*”

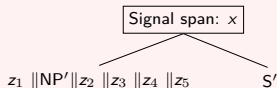
- ▶ **Recursive** transformation of syntactically complex sentences into simplified, compact structures using a **rule-based sentence splitting approach** for
 - ▶ clausal disembedding, and
 - ▶ phrasal disembedding
- ▶ “*Trump withdrew his sponsorship after the second Tour de Trump in 1990 because his business ventures were experiencing financial woes.*”
 - ▶ CORE: “*T. withdrew his sponsorship.*”
 - ▶ CONTEXT: “*His business ventures were experiencing financial woes.*”
 - ▶ CONTEXT: “*This was after the second Tour de Trump.*”
 - ▶ CONTEXT: “*This was in 1990.*”

Rule: Subordinated Clauses with Closing Subordinative Clauses

Phrasal Pattern:

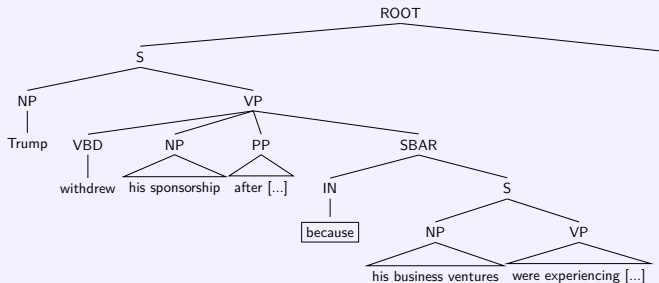


Extraction:



Rule: Subordinated Clauses with Closing Subordinative Clauses

Phrasal Pattern:



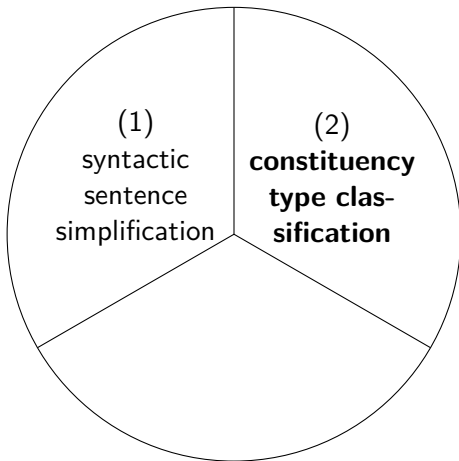
Rule: Subordinated Clauses with Closing Subordinative Clauses

Extraction:

"because"

Trump withdrew his
sponsorship after the second
Tour de Trump in 1990.

His business ventures
were experiencing
financial woes.



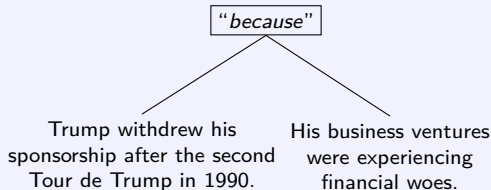
- ▶ Depicts the **contextual hierarchy** between the simplified sentences
- ▶ Adopts the concept of **nuclearity from RST** (Mann and Thompson, 1988):
 - ▶ coordinate sentences represent nucleus spans (CORE)
 - ▶ subordinate sentences represent satellite spans (CONTEXT)

- ▶ Depicts the **contextual hierarchy** between the simplified sentences
- ▶ Adopts the concept of **nuclearity from RST** (Mann and Thompson, 1988):
 - ▶ coordinate sentences represent nucleus spans (CORE)
 - ▶ subordinate sentences represent satellite spans (CONTEXT)

- ▶ Depicts the **contextual hierarchy** between the simplified sentences
- ▶ Adopts the concept of **nuclearity from RST** (Mann and Thompson, 1988):
 - ▶ coordinate sentences represent nucleus spans (CORE)
 - ▶ subordinate sentences represent satellite spans (CONTEXT)

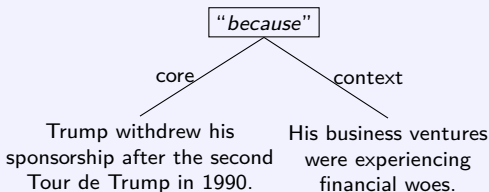
- ▶ Depicts the **contextual hierarchy** between the simplified sentences
- ▶ Adopts the concept of **nuclearity from RST** (Mann and Thompson, 1988):
 - ▶ coordinate sentences represent nucleus spans (CORE)
 - ▶ subordinate sentences represent satellite spans (CONTEXT)

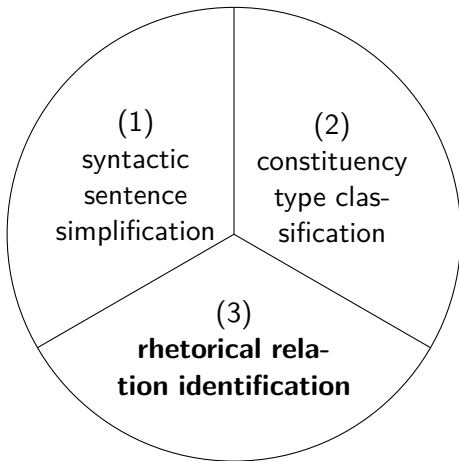
Subordinated Clauses with Closing Subordinative Clauses



- ▶ Depicts the **contextual hierarchy** between the simplified sentences
- ▶ Adopts the concept of **nuclearity from RST** (Mann and Thompson, 1988):
 - ▶ coordinate sentences represent nucleus spans (CORE)
 - ▶ subordinate sentences represent satellite spans (CONTEXT)

Subordinated Clauses with Closing Subordinative Clauses





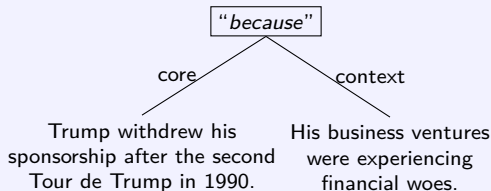
- ▶ Use of **syntactic and lexical features** to extract cue phrases from sentences
- ▶ **Mapping** of cue phrases to rhetorical relations based on the work of Taboada and Das (2013)

- ▶ Use of **syntactic and lexical features** to extract cue phrases from sentences
- ▶ **Mapping** of cue phrases to rhetorical relations based on the work of Taboada and Das (2013)

- ▶ Use of **syntactic and lexical features** to extract cue phrases from sentences
- ▶ **Mapping** of cue phrases to rhetorical relations based on the work of Taboada and Das (2013)

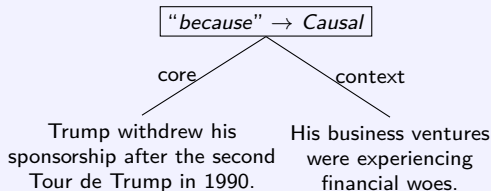
- ▶ Use of **syntactic and lexical features** to extract cue phrases from sentences
- ▶ **Mapping** of cue phrases to rhetorical relations based on the work of Taboada and Das (2013)

Subordinated Clauses with Closing Subordinative Clauses



- ▶ Use of **syntactic and lexical features** to extract cue phrases from sentences
- ▶ **Mapping** of cue phrases to rhetorical relations based on the work of Taboada and Das (2013)

Subordinated Clauses with Closing Subordinative Clauses



DOCUMENT-ROOT

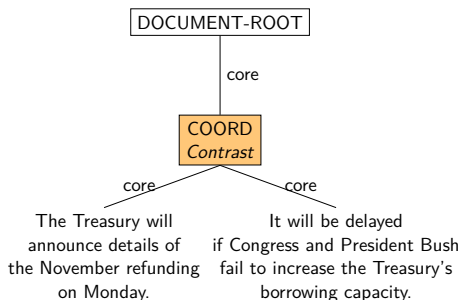
core

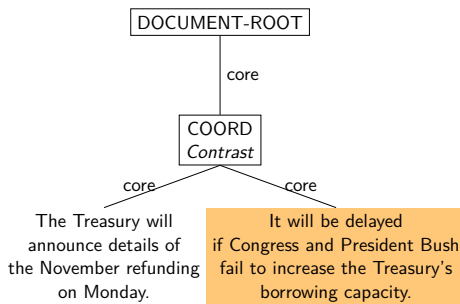
Although the Treasury will announce details of the November refunding on Monday,
the funding will be delayed if Congress and President Bush
fail to increase the Treasury's borrowing capacity.

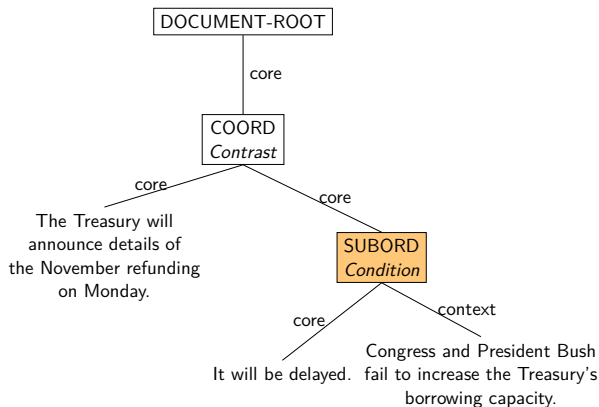
DOCUMENT-ROOT

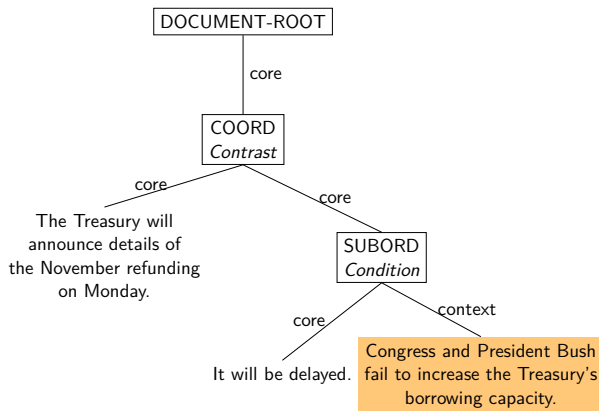
core

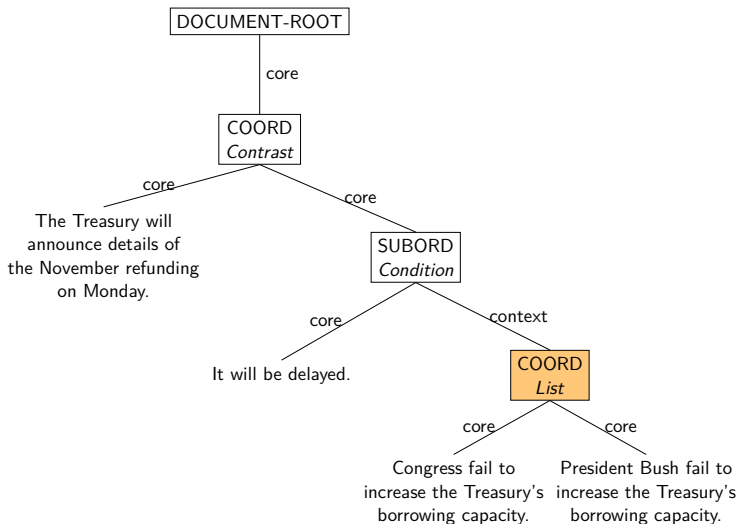
Although the Treasury will announce details of the November refunding on Monday, the funding will be delayed if Congress and President Bush fail to increase the Treasury's borrowing capacity.

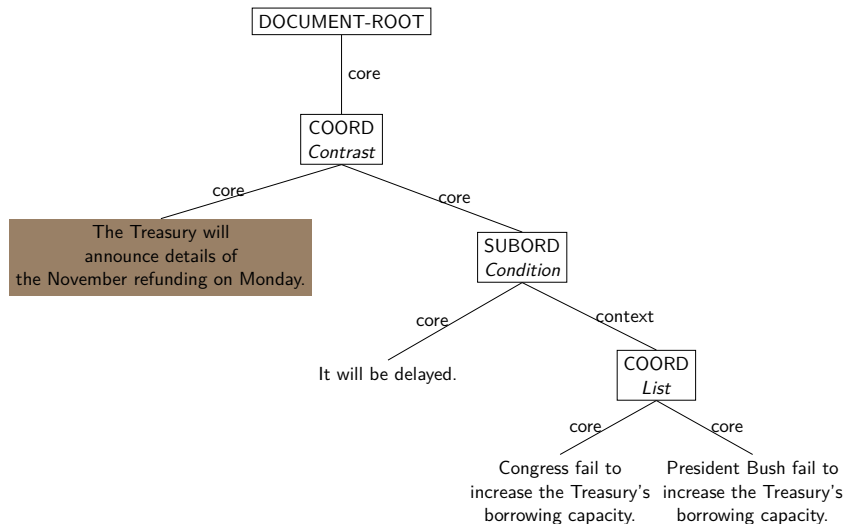


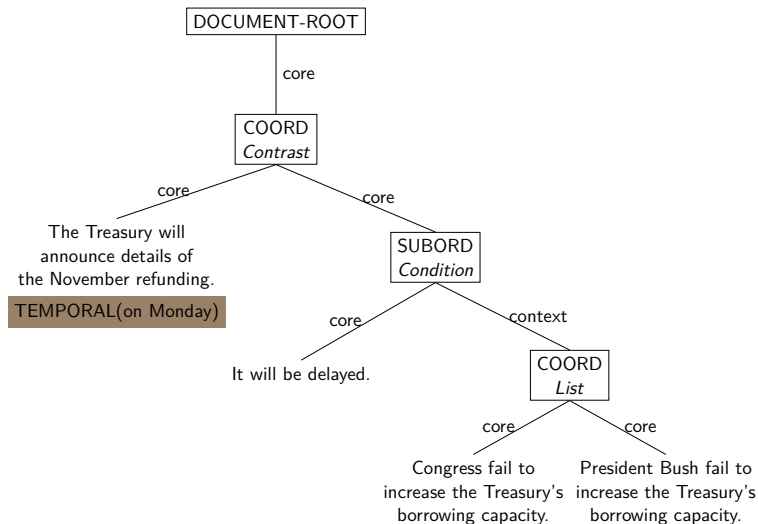




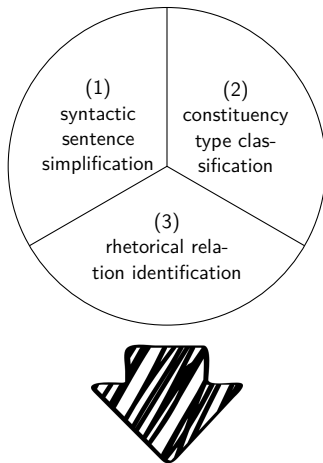






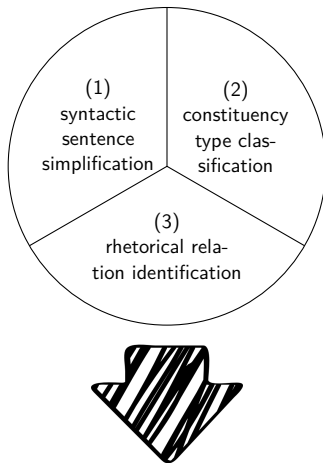


Step 1: Transformation Stage



Hierarchy of semantically linked sentences that present a
simple syntax in the form of a discourse tree

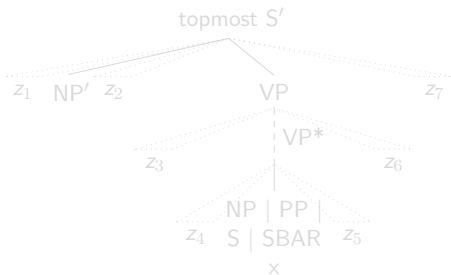
Step 1: Transformation Stage



Hierarchy of semantically linked sentences that present a simple syntax in the form of a discourse tree

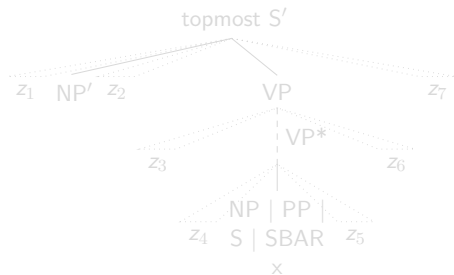
Step 2: Relation Extraction

- ▶ Convert each (simplified) sentence into a binary relational tuple: $\langle arg_1, rel, arg_2 \rangle$
- ▶ state-of-the-art Open IE systems can serve as potential implementations
- ▶ RE implementation for **Graphene**:



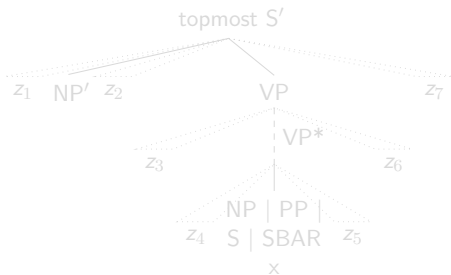
$rel \leftarrow z_3 \parallel z_4$
 $arg_1 \leftarrow NP'$
 $arg_2 \leftarrow x \parallel z_5$

- ▶ Convert each (simplified) sentence into a binary relational tuple: $\langle arg_1, rel, arg_2 \rangle$
- ▶ state-of-the art Open IE systems can serve as potential implementations
- ▶ RE implementation for **Graphene**:



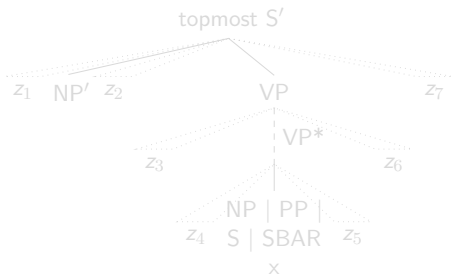
$rel \leftarrow z_3 \parallel z_4$
 $arg_1 \leftarrow NP'$
 $arg_2 \leftarrow x \parallel z_5$

- ▶ Convert each (simplified) sentence into a binary relational tuple: $\langle arg_1, rel, arg_2 \rangle$
- ▶ state-of-the art Open IE systems can serve as potential implementations
- ▶ RE implementation for **Graphene**:



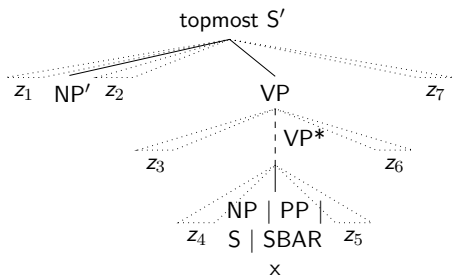
$rel \leftarrow z_3 \parallel z_4$
 $arg_1 \leftarrow NP'$
 $arg_2 \leftarrow x \parallel z_5$

- ▶ Convert each (simplified) sentence into a binary relational tuple: $\langle arg_1, rel, arg_2 \rangle$
- ▶ state-of-the art Open IE systems can serve as potential implementations
- ▶ RE implementation for **Graphene**:



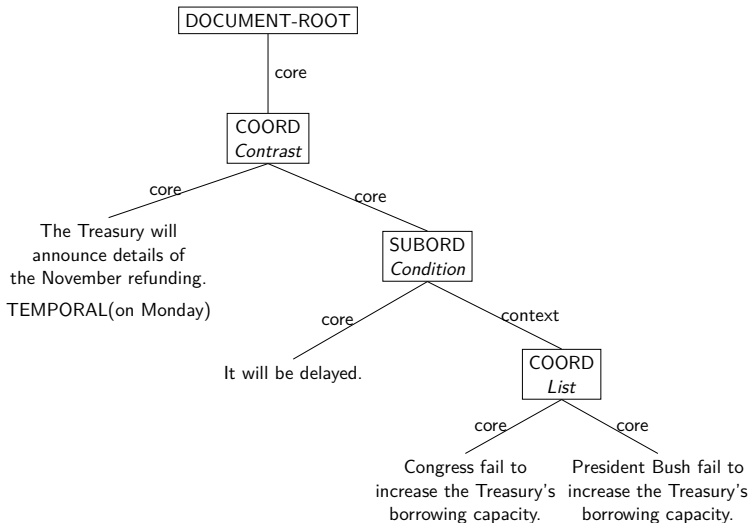
$rel \leftarrow z_3 \parallel z_4$
 $arg_1 \leftarrow NP'$
 $arg_2 \leftarrow x \parallel z_5$

- ▶ Convert each (simplified) sentence into a binary relational tuple: $\langle arg_1, rel, arg_2 \rangle$
- ▶ state-of-the art Open IE systems can serve as potential implementations
- ▶ RE implementation for **Graphene**:

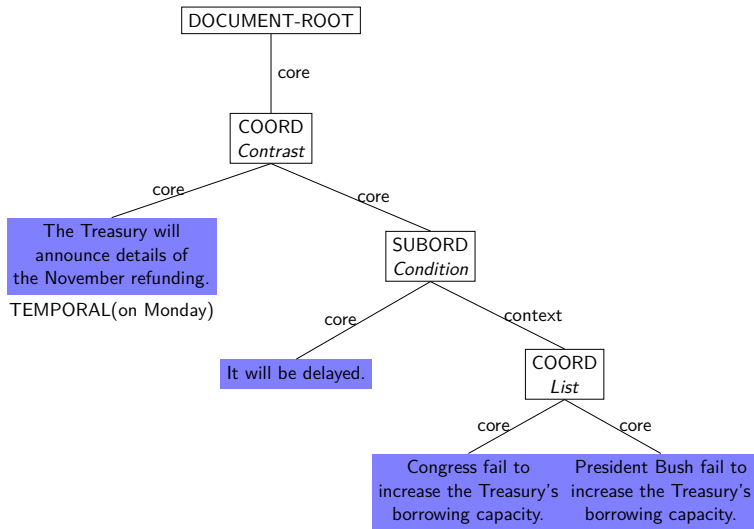


$rel \leftarrow z_3 \parallel z_4$
 $arg_1 \leftarrow NP'$
 $arg_2 \leftarrow x \parallel z_5$

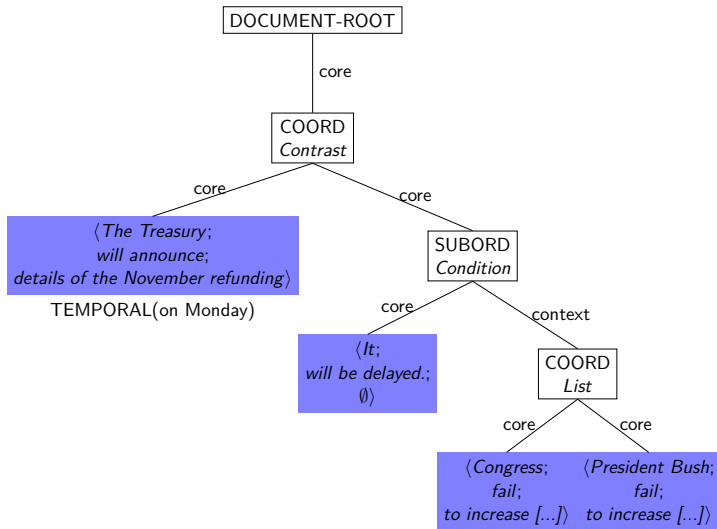
Step 2: Relation Extraction



Step 2: Relation Extraction



Step 2: Relation Extraction



"Although the Treasury will announce details of the November refunding on Monday, the funding will be delayed if Congress and President Bush fail to increase the Treasury's borrowing capacity."

#1 0 the Treasury will announce details of the November refunding
S:TEMPORAL on Monday
L:CONTRAST #2

#2 0 it will be delayed
L:CONTRAST #1
L:CONDITION #3
L:CONDITION #4

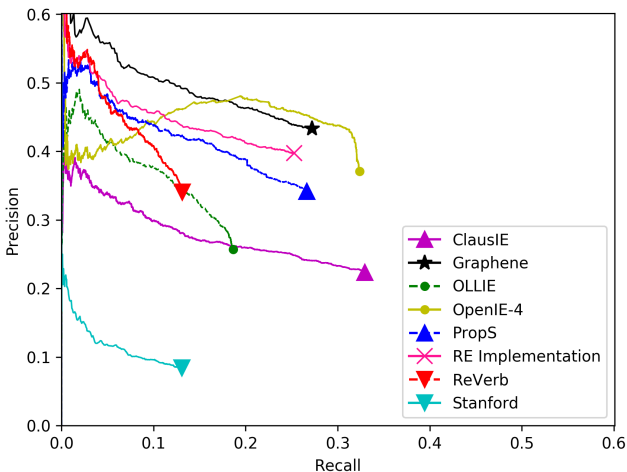
#3 1 Congress fail to increase the Treasury's borrowing capacity
L:LIST #4

#4 1 president Bush fail to increase the Treasury's borrowing [...]
L:LIST #3

Comparison of Graphene with state-of-the-art Open IE systems based on a large benchmark dataset for Open IE (Stanovsky and Dagan, 2016):

- ▶ OLLIE (Mausam et al., 2012)
- ▶ ClausIE (Del Corro and Gemulla, 2013)
- ▶ Stanford Open IE (Angeli et al., 2015)
- ▶ PropS (Stanovsky et al., 2016)
- ▶ OpenIE-4 (Mausam, 2016)
- ▶ ReVerb (Fader et al., 2011)

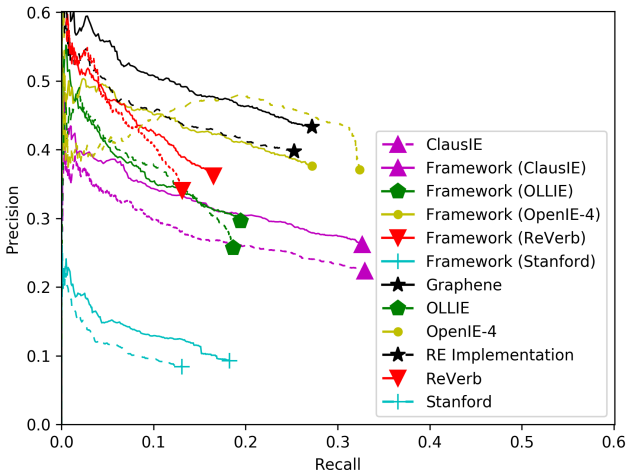
Performance of Graphene:



Performance of Graphene:

System	avg. P	R	AUC
OpenIE-4	0.446	0.324	0.145
Graphene	0.501	0.272	0.136
<i>RE implementation</i>	0.455	0.253	0.115
PropS	0.424	0.267	0.113
ClausIE	0.282	0.330	0.093
OLLIE	0.376	0.187	0.070
RE _{VERB}	0.466	0.131	0.061
Stanford Open IE	0.120	0.131	0.016

Improvements of state-of-the-art systems when operating as RE component of our framework:



- ▶ Graphene achieves best average precision (50.1%) with comparable recall (27.2%) to that of other high-precision Open IE systems
- ▶ lightweight semantic representation:
 - ▶ increases expressiveness of extracted propositions
 - ▶ avoids the problem of overspecified argument phrases
 - ▶ yields a more compact structure
- ▶ state-of-the-art Open IE systems benefit from clausal, phrasal and rhetorical disembedding as a preprocessing step with increased AUC score up to 63% improvement

- ▶ Graphene achieves best average precision (50.1%) with comparable recall (27.2%) to that of other high-precision Open IE systems
- ▶ lightweight semantic representation:
 - ▶ increases expressiveness of extracted propositions
 - ▶ avoids the problem of overspecified argument phrases
 - ▶ yields a more compact structure
- ▶ state-of-the-art Open IE systems benefit from clausal, phrasal and rhetorical disembedding as a preprocessing step with increased AUC score up to 63% improvement

- ▶ Graphene achieves best average precision (50.1%) with comparable recall (27.2%) to that of other high-precision Open IE systems
- ▶ lightweight semantic representation:
 - ▶ increases expressiveness of extracted propositions
 - ▶ avoids the problem of overspecified argument phrases
 - ▶ yields a more compact structure
- ▶ state-of-the-art Open IE systems benefit from clausal, phrasal and rhetorical disembedding as a preprocessing step with increased AUC score up to 63% improvement

- ▶ Graphene achieves best average precision (50.1%) with comparable recall (27.2%) to that of other high-precision Open IE systems
- ▶ lightweight semantic representation:
 - ▶ increases expressiveness of extracted propositions
 - ▶ avoids the problem of overspecified argument phrases
 - ▶ yields a more compact structure
- ▶ state-of-the-art Open IE systems benefit from clausal, phrasal and rhetorical disembedding as a preprocessing step with increased AUC score up to 63% improvement