State of the Art in Timeline Extraction

Steven Bethard

University of Alabama at Birmingham

i2b2 AUG NLP Workshop
9 Jul 2014

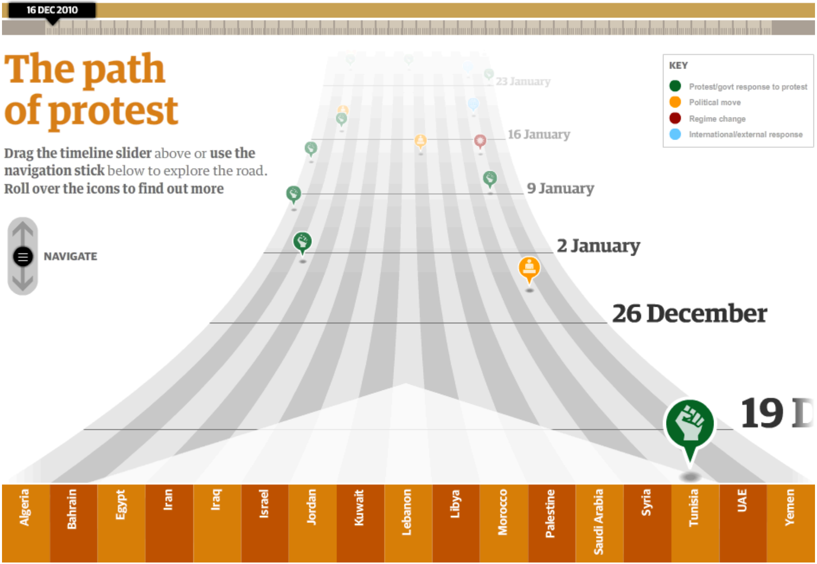
A Middle East Timeline

At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today…

Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire when police confiscated his fruits and vegetables for selling without a permit…

Mobs have since attacked public buildings and the local office of the party of President Zine El Abidine Ben Ali.

A Middle East Timeline



www.guardian.co.uk/world/interactive/2011/mar/22/middle-east-protest-interactive-timeline

A Clinical Record Timeline

A 32-year-old woman was admitted to the hospital because of left subcostal pain. Three months before admission an evaluation elsewhere included a CT scan of the abdomen. She had a history of eczema and of asthma. She had lost 18 kg in weight during the preceding 18 months. An abdominal examination revealed a soft systolic bruit...

- History: eczema, asthma
- Last 18 months: lost 18kg
- 3 months ago: CT scan
- Current: left subcostal pain
- Current: soft systolic bruit

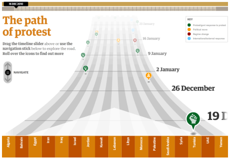
Outline

1. Introduction
2. Timeline extraction as supervised learning
   - Identifying events and times
   - Normalizing times
   - Linking events and times
3. Improving timelines with unannotated data
   - Temporal information via web queries
   - Latent structure from unsupervised models
4. Improving the model of temporal links
   - Links from linguistic constructions
   - Links from narrative containers
   - Links as dependency trees
   - Dense links through transitivity
5. Conclusions

A timeline model should predict like humans

At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today. Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire...

h (At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today. Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire...)



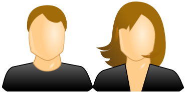
Models $h(x)$ are learned from annotations

At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today. Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire…

Manual Annotation

… on fire → rioting → … → died …

17 December → on fire; today → died



Feature Extraction

…	…
part-of-speech=verb	0
syntax-path=VP→S←NP	1
distance	7.9
…	…

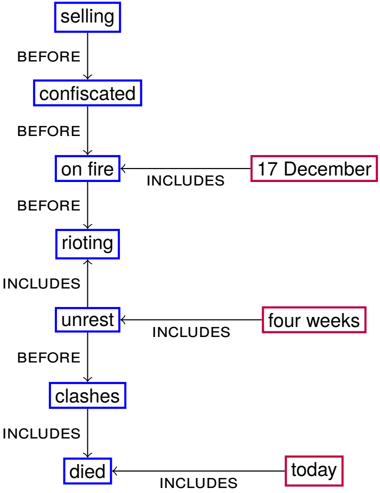
→ Model Trainer

↓

$h(x)$

Timelines are annotated as graphs

h (At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today. Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire when police confiscated his fruits and vegetables for selling without a permit. . .) =



ISO-TimeML (Pustejovsky, Lee, et al. 2010)

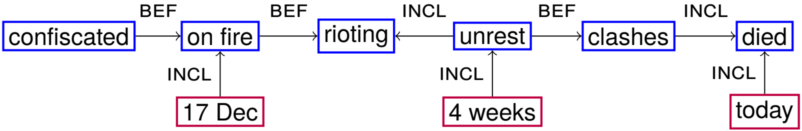
Timeline extraction as a classification pipeline

At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today. Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire when police confiscated his fruits...

$$h_{\text{time+event}} \downarrow$$

At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today. Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire when police confiscated his fruits...

$$h_{\text{link}} \downarrow$$



Events and times as word classification

x	$h_{\text{time+event}}(x)$
…	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
…	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* → LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^\top x}}$$
- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
...	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
...	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* $\rightarrow$ LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^{\top}x}}$$
- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
…	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
…	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g. *Dec* $\rightarrow$ LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:

$$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^\top x}}$$

- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
...	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
...	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* $\longrightarrow$ LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:

  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^{\top}x}}$$

- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
…	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
…	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* $\rightarrow$ LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^{\top}x}}$$
- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
...	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
...	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* $\longrightarrow$ LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:

$$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^\top x}}$$

- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
…	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
…	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* $\rightarrow$ LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^{\top}x}}$$
- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
…	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
…	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* $\longrightarrow$ LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:

$$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^\top x}}$$

- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
…	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
…	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g. *Dec* $\rightarrow$ LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:

$$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^\top x}}$$

- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
...	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
...	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* $\rightarrow$ LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:

$$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^{\top}x}}$$

- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
…	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
…	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* $\longrightarrow$ LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^\top x}}$$
- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
…	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
…	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* → LuLLLL
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:

$$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^\top x}}$$

- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
...	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
...	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* → LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:

$$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^\top x}}$$

- Support vector machines
- Conditional random fields

Events and times as word classification

x	$h_{\text{time+event}}(x)$
...	
after	O
four	B-TIME
weeks	I-TIME
of	O
unrest	B-EVENT
,	O
it	O
was	O
reported	B-EVENT
today	B-TIME
...	

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. VERB
- Character category patterns, e.g *Dec* → LuLlLl
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^\top x}}$$
- Support vector machines
- Conditional random fields

Normalizing times using rules

Normalizing time expressions:

- *December 5, 2007* ⇒ 2007-12-05
- *the day before yesterday* ⇒ 2014-07-07

Modeled with hand-constructed grammars (e.g. Bethard 2013):

Normalizing times using rules

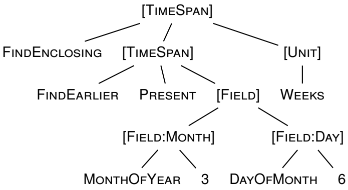
Normalizing time expressions:

- *December 5, 2007* ⇒ 2007-12-05
- *the day before yesterday* ⇒ 2014-07-07

Modeled with hand-constructed grammars (e.g. Bethard 2013):

the week of March 6

⇓



⇓

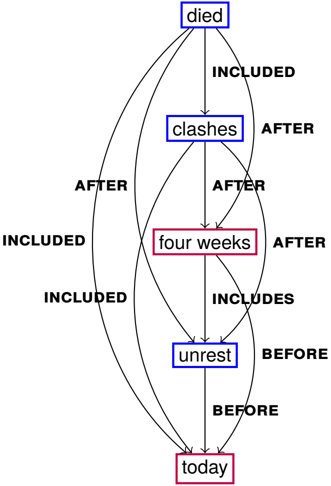
2014-W10 ⇐ *today*=2014-07-09

Linking events and times as pair classification

x	$h_{\text{link}}(x)$
(died, clashes)	INCLUDED
(died, four weeks)	AFTER
(died, unrest)	AFTER
(died, today)	INCLUDED
(clashes, four weeks)	AFTER
(clashes, unrest)	AFTER
(clashes, today)	INCLUDED
(four weeks, unrest)	INCLUDES
(four weeks, today)	BEFORE
(unrest, today)	BEFORE

Linking events and times as pair classification

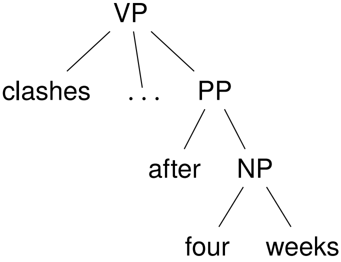
x	$h_{\text{link}}(x)$
(died, clashes)	INCLUDED
(died, four weeks)	AFTER
(died, unrest)	AFTER
(died, today)	INCLUDED
(clashes, four weeks)	AFTER
(clashes, unrest)	AFTER
(clashes, today)	INCLUDED
(four weeks, unrest)	INCLUDES
(four weeks, today)	BEFORE
(unrest, today)	BEFORE



Linking events and times as pair classification

Features $x = (f_1, f_2, \ldots, f_m)$

- Event$_1$ part-of-speech
- Event$_2$ part-of-speech
- Bag of words between
- …
- Path in syntactic tree



Learning h_{link}:

- Support vector machine
- Logistic regression
- …

How well does it work?

Shared tasks: TempEval 2007, TempEval 2010, TempEval 2013

- Common annotated training set: events, times, links
- Common unannotated test set for system predictions

Evaluation metrics:

Precision	$\frac{\text{\# correct predictions}}{\text{\# predictions}}$
Recall	$\frac{\text{\# correct predictions}}{\text{\# actual}}$
F_1	$\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$

How well does it work?

TempEval 2007:

- Link classification: 55-80% accuracy

TempEval 2010:

- Event identification: 81% precision, 86% recall
- Time identification: 90% precision, 82% recall
- Time normalization: 85% accuracy
- Link classification: 55-81% accuracy

TempEval 2013:

- Event identification: 81% precision, 81% recall
- Time identification: 86% precision, 80% recall
- Time normalization: 86% accuracy
- Link identification: 37% precision, 35% recall

Some remaining challenges

Current features aren't predictive enough:

- *He* $\boxed{\textit{waited}}$ *there and* $\boxed{\textit{looked}}$ *around.* (INCLUDES)
- *He* $\boxed{\textit{arrived}}$ *there and* $\boxed{\textit{looked}}$ *around.* (BEFORE)

TempEval data is small and sparse:

- no examples of *autumn*
- only 1 example of *winter*

TempEval links are incomplete:

- E.g., *Farkas was* $\boxed{\textit{ordered}}$ *home and* $\boxed{\textit{retired}}$.

Some remaining challenges

Current features aren't predictive enough:

- *He* $\boxed{\textit{waited}}$ *there and* $\boxed{\textit{looked}}$ *around.* (INCLUDES)
- *He* $\boxed{\textit{arrived}}$ *there and* $\boxed{\textit{looked}}$ *around.* (BEFORE)

TempEval data is small and sparse:

- no examples of *autumn*
- only 1 example of *winter*

TempEval links are incomplete:

- E.g., *Farkas was* $\boxed{\textit{ordered}}$ *home and* $\boxed{\textit{retired}}$.

Some remaining challenges

Current features aren't predictive enough:

- *He $\boxed{\text{waited}}$ there and $\boxed{\text{looked}}$ around.* (INCLUDES)
- *He $\boxed{\text{arrived}}$ there and $\boxed{\text{looked}}$ around.* (BEFORE)

TempEval data is small and sparse:

- no examples of *autumn*
- only 1 example of *winter*

TempEval links are incomplete:

- E.g., *Farkas was $\boxed{\text{ordered}}$ home and $\boxed{\text{retired}}$.*

Developing features that capture semantics

(Gusev, Chambers, et al. 2011)

Intuition: Duration information should help

- *He* $\boxed{\textit{waited}}$ *there and* $\boxed{\textit{looked}}$ *around.* (INCLUDES)
- *He* $\boxed{\textit{arrived}}$ *there and* $\boxed{\textit{looked}}$ *around.* (BEFORE)

Approach: ask the web how long *waiting* takes

Web patterns as classifier > training on 1700 annotations

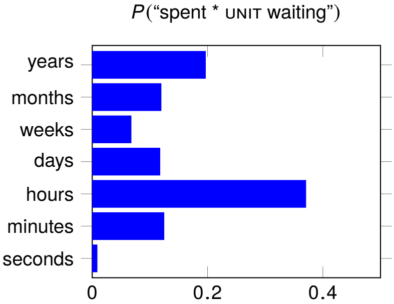
Developing features that capture semantics

(Gusev, Chambers, et al. 2011)

Intuition: Duration information should help

- *He* $\boxed{\textit{waited}}$ *there and* $\boxed{\textit{looked}}$ *around.* (INCLUDES)
- *He* $\boxed{\textit{arrived}}$ *there and* $\boxed{\textit{looked}}$ *around.* (BEFORE)

Approach: ask the web how long *waiting* takes



Web patterns as classifier > training on 1700 annotations

Developing features that capture semantics

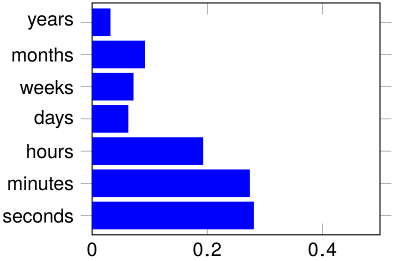
(Gusev, Chambers, et al. 2011)

Intuition: Duration information should help

- *He* $\boxed{\textit{waited}}$ *there and* $\boxed{\textit{looked}}$ *around.* (INCLUDES)
- *He* $\boxed{\textit{arrived}}$ *there and* $\boxed{\textit{looked}}$ *around.* (BEFORE)

Approach: ask the web how long *waiting* takes

P("spent * UNIT waiting"|"spent * UNIT")



looking

Web patterns as classifier > training on 1700 annotations

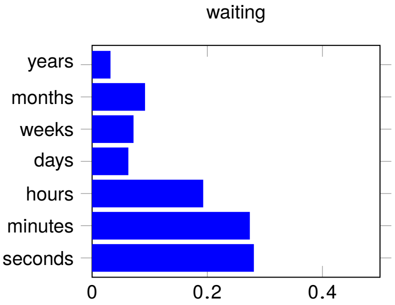
Developing features that capture semantics

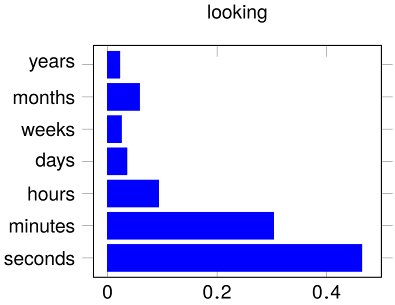
(Gusev, Chambers, et al. 2011)

Intuition: Duration information should help

- *He* $\boxed{\textit{waited}}$ *there and* $\boxed{\textit{looked}}$ *around.* (INCLUDES)
- *He* $\boxed{\textit{arrived}}$ *there and* $\boxed{\textit{looked}}$ *around.* (BEFORE)

Approach: ask the web how long *waiting* takes





Web patterns as classifier > training on 1700 annotations

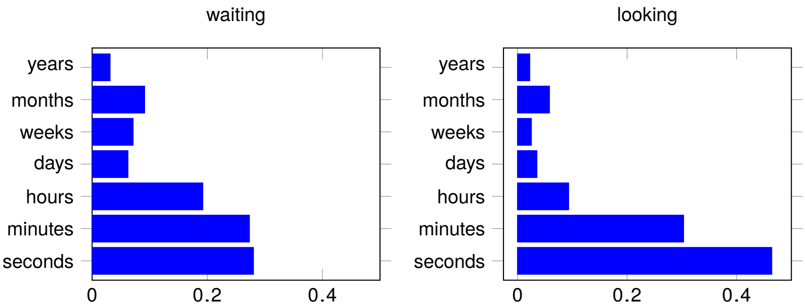
Developing features that capture semantics

(Gusev, Chambers, et al. 2011)

Intuition: Duration information should help

- *He* $\boxed{\textit{waited}}$ *there and* $\boxed{\textit{looked}}$ *around.* (INCLUDES)
- *He* $\boxed{\textit{arrived}}$ *there and* $\boxed{\textit{looked}}$ *around.* (BEFORE)

Approach: ask the web how long *waiting* takes



Web patterns as classifier > training on 1700 annotations

Managing sparsity with unsupervised models

(Kolomiyets, Bethard, et al. 2011)

TempEval data is small and sparse:

- no examples of *autumn*
- only 1 example of *winter*

Latent Words Language Model (Deschacht and Moens 2009)

Generate training examples

- TempEval train, TempEval test: 0.820 $\rightarrow$ 0.861 F_1
- TempEval train, Reuters test: 0.773 $\rightarrow$ 0.826 F_1
- Better than synonym expansion with WordNet

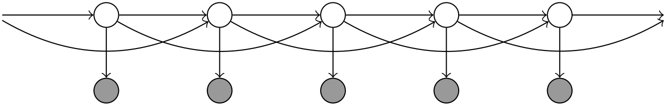
Managing sparsity with unsupervised models

(Kolomiyets, Bethard, et al. 2011)

TempEval data is small and sparse:

- no examples of *autumn*
- only 1 example of *winter*

Latent Words Language Model (Deschacht and Moens 2009)



Generate training examples

- TempEval train, TempEval test: 0.820 → 0.861 F_1
- TempEval train, Reuters test: 0.773 → 0.826 F_1
- Better than synonym expansion with WordNet

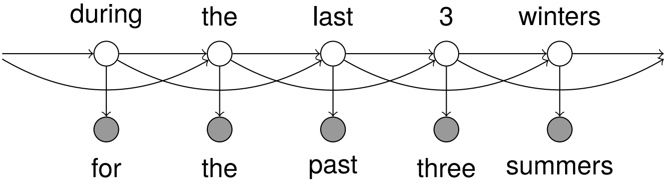
Managing sparsity with unsupervised models

(Kolomiyets, Bethard, et al. 2011)

TempEval data is small and sparse:

- no examples of *autumn*
- only 1 example of *winter*

Latent Words Language Model (Deschacht and Moens 2009)



Generate training examples

- TempEval train, TempEval test: 0.820 → 0.861 F_1
- TempEval train, Reuters test: 0.773 → 0.826 F_1
- Better than synonym expansion with WordNet

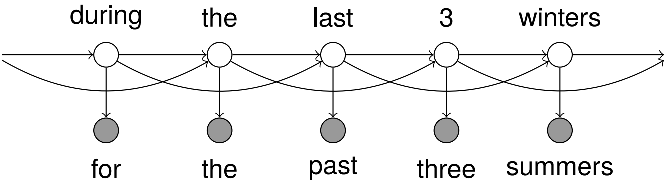
Managing sparsity with unsupervised models

(Kolomiyets, Bethard, et al. 2011)

TempEval data is small and sparse:

- no examples of *autumn*
- only 1 example of *winter*

Latent Words Language Model (Deschacht and Moens 2009)



Generate training examples

- TempEval train, TempEval test: 0.820 → 0.861 F_1
- TempEval train, Reuters test: 0.773 → 0.826 F_1
- Better than synonym expansion with WordNet

TempEval relation annotations are unintuitive

Arbitrary links in TempEval 2007:

Turning its back on 210 years of loyalty to the British royal family, a constitutional convention voted overwhelmingly Friday to make Australia a republic under its own president.

210 years $\xrightarrow{\textit{BEFORE}}$ *make*

Missing links in TempEval 2010

The World Court Friday rejected U.S. and British objections to a Libyan World Court case that has blocked the trial of two Libyans suspected of blowing up a Pan Am jumbo jet over Scotland in 1988.

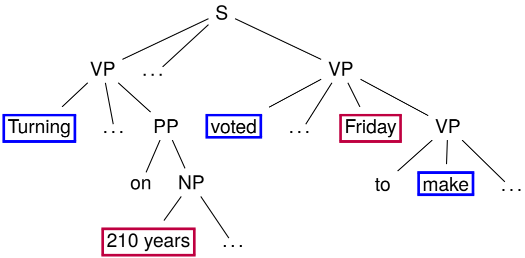
rejected $\xrightarrow{\textit{AFTER}}$ *objections*   *suspected* $\xrightarrow{\textit{AFTER}}$ *blowing*

More linguistically plausible links

(Bethard, Martin, et al. 2007)

More natural links using syntax:

Turning its back on 210 years of loyalty to the British royal family, a constitutional convention voted overwhelmingly Friday to make Australia a republic under its own president.

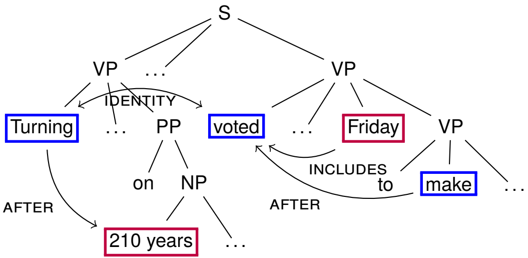


More linguistically plausible links

(Bethard, Martin, et al. 2007)

More natural links using syntax:

Turning its back on 210 years of loyalty to the British royal family, a constitutional convention voted overwhelmingly Friday to make Australia a republic under its own president.

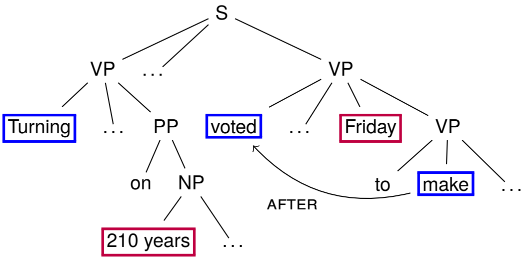


More linguistically plausible links

(Bethard, Martin, et al. 2007)

More natural links using syntax:

Turning its back on 210 years of loyalty to the British royal family, a constitutional convention voted overwhelmingly Friday to make Australia a republic under its own president.



Case study on 132 document corpus:

- Easier for annotators (90% agreement)
- Easier for models (89% accuracy)

Narrative containers

(Pustejovsky and Stubbs 2011)

President Obama paid tribute Sunday to 29 workers killed in an explosion at a West Virginia coal mine earlier this month, saying they died "in pursuit of the American dream." The blast at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years.

earlier this month
killed explosion died blast

Sunday
paid saying

Case study re-annotating the TimeBank:

- Easier for annotators (Kappa 0.74)

Case study with clinical narratives:

- Tractable for models (Miller, Bethard, et al. 2013)

Narrative containers

(Pustejovsky and Stubbs 2011)

President Obama paid tribute Sunday to 29 workers killed in an explosion at a West Virginia coal mine earlier this month, saying they died "in pursuit of the American dream." The blast at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years.

earlier this month
killed explosion died blast

Sunday
paid saying

Case study re-annotating the TimeBank:

- Easier for annotators (Kappa 0.74)

Case study with clinical narratives:

- Tractable for models (Miller, Bethard, et al. 2013)

Narrative containers

(Pustejovsky and Stubbs 2011)

President Obama paid tribute Sunday to 29 workers killed in an explosion at a West Virginia coal mine earlier this month, saying they died "in pursuit of the American dream." The blast at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years.

earlier this month			
killed	explosion	died	blast

Sunday	
paid	saying

Case study re-annotating the TimeBank:

- Easier for annotators (Kappa 0.74)

Case study with clinical narratives:

- Tractable for models (Miller, Bethard, et al. 2013)

Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were $\boxed{\textit{on}}$ *the road together, when a bear suddenly* $\boxed{\textit{appeared}}$ *on the scene. Before he* $\boxed{\textit{observed}}$ *them, one* $\boxed{\textit{made}}$ *for a tree at the side of the road, and* $\boxed{\textit{climbed}}$ *up into the branches and* $\boxed{\textit{hid}}$*. The other ...* $\boxed{\textit{threw}}$ *himself on the ground and* $\boxed{\textit{pretended}}$ *to be dead.*

Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other . . . threw himself on the ground and pretended to be dead.

on

Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were $\boxed{\text{on}}$ the road together, when a bear suddenly $\boxed{\text{appeared}}$ on the scene. Before he $\boxed{\text{observed}}$ them, one $\boxed{\text{made}}$ for a tree at the side of the road, and $\boxed{\text{climbed}}$ up into the branches and $\boxed{\text{hid}}$. The other … $\boxed{\text{threw}}$ himself on the ground and $\boxed{\text{pretended}}$ to be dead.

$\boxed{\text{on}}$ $\boxed{\text{appeared}}$

Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other … threw himself on the ground and pretended to be dead.

INCLUDES

on appeared

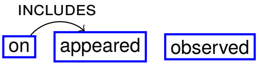
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other … threw himself on the ground and pretended to be dead.



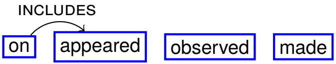
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other … threw himself on the ground and pretended to be dead.



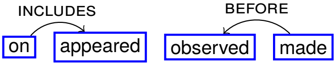
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other … threw himself on the ground and pretended to be dead.



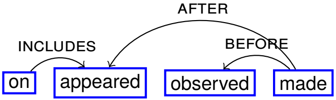
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other . . . threw himself on the ground and pretended to be dead.



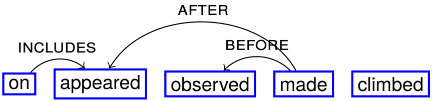
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were *on* *the road together, when a bear suddenly* *appeared* *on the scene. Before he* *observed* *them, one* *made* *for a tree at the side of the road, and* *climbed* *up into the branches and* *hid*. *The other . . .* *threw* *himself on the ground and* *pretended* *to be dead.*



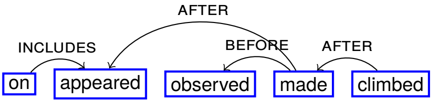
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were $\boxed{\textit{on}}$ *the road together, when a bear suddenly* $\boxed{\textit{appeared}}$ *on the scene. Before he* $\boxed{\textit{observed}}$ *them, one* $\boxed{\textit{made}}$ *for a tree at the side of the road, and* $\boxed{\textit{climbed}}$ *up into the branches and* $\boxed{\textit{hid}}$*. The other . . .* $\boxed{\textit{threw}}$ *himself on the ground and* $\boxed{\textit{pretended}}$ *to be dead.*



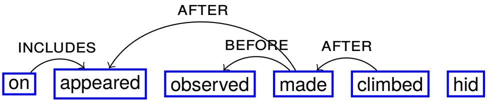
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other . . . threw himself on the ground and pretended to be dead.



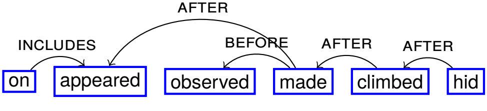
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other . . . threw himself on the ground and pretended to be dead.



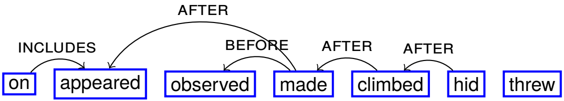
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were [on] *the road together, when a bear suddenly* [appeared] *on the scene. Before he* [observed] *them, one* [made] *for a tree at the side of the road, and* [climbed] *up into the branches and* [hid]*. The other . . .* [threw] *himself on the ground and* [pretended] *to be dead.*



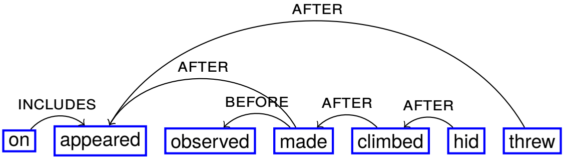
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other . . . threw himself on the ground and pretended to be dead.



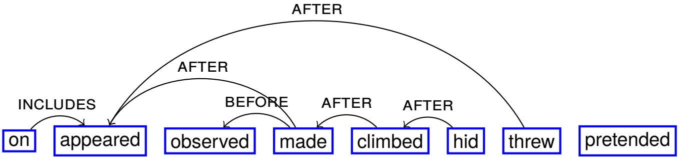
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other . . . threw himself on the ground and pretended to be dead.



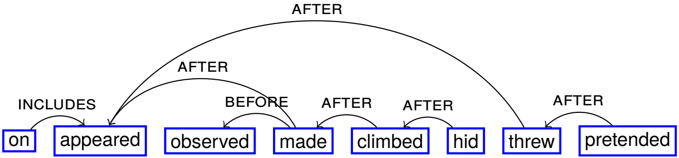
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

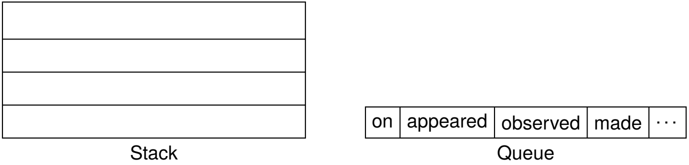
(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other … threw himself on the ground and pretended to be dead.



Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

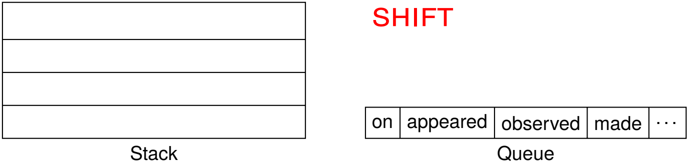


Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

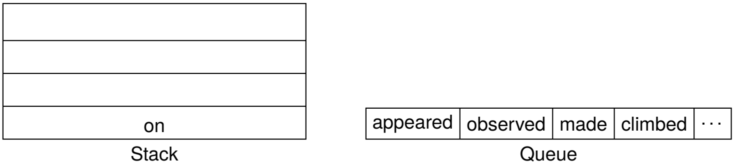


Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

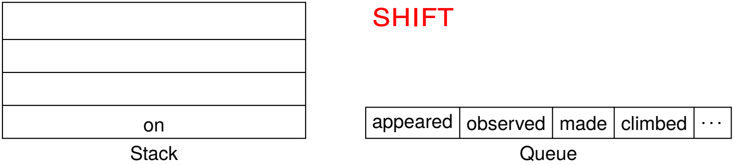


Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

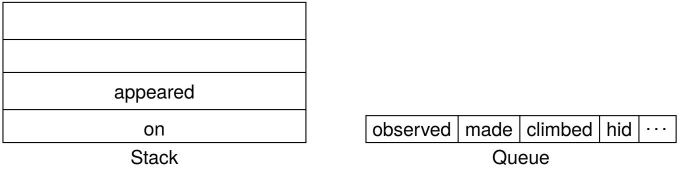


Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)



Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Stack
appeared
on

Stack

REDUCE$_{\text{INCLUDES}}$

observed	made	climbed	hid	···

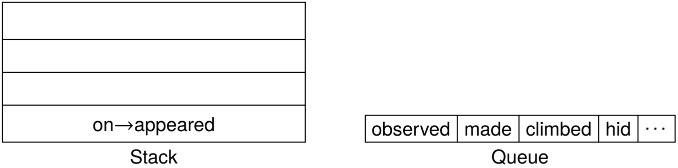
Queue

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)



Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Stack
on→appeared

SHIFT

observed	made	climbed	hid	⋯

Queue

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

observed
on→appeared

Stack

made	climbed	hid	threw	· · ·

Queue

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

SHIFT

observed
on→appeared

Stack

made	climbed	hid	threw	⋯

Queue

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

made
observed
on→appeared

Stack

climbed	hid	threw	pretended	···

Queue

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

made
observed
on→appeared

Stack

REDUCE$_{\text{BEFORE}}$

climbed	hid	threw	pretended	···

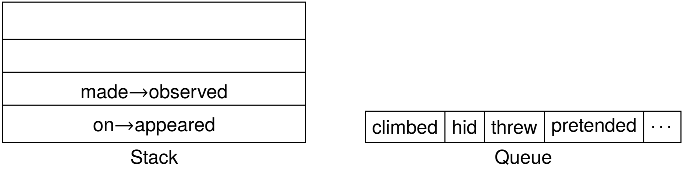
Queue

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

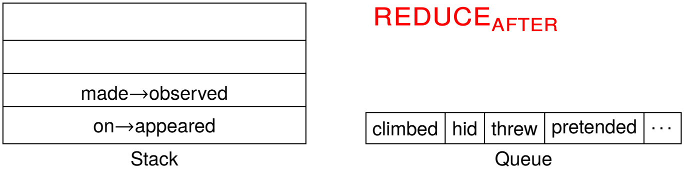


Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

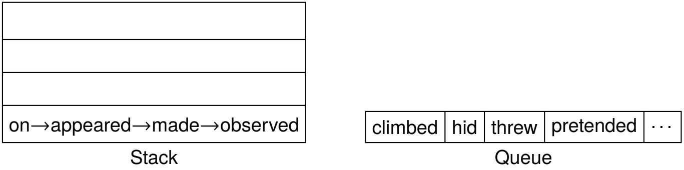


Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

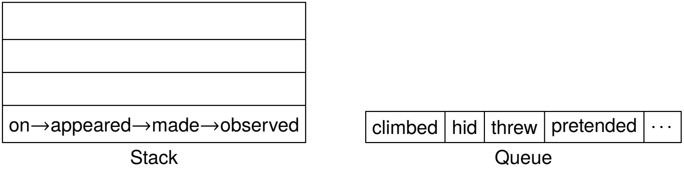


Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)



Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Dense links through transitivity

(Chambers, Cassidy, et al. 2014; Cassidy, McDowell, et al. 2014)

There were four or five people inside and they just started $\boxed{\textit{firing}}$*. Ms. Sanders was* $\boxed{\textit{hit}}$ *several times and was* $\boxed{\textit{pronounced}}$ $\boxed{\textit{dead}}$ *at the scene.*

Dense annotation:

Dense links through transitivity

(Chambers, Cassidy, et al. 2014; Cassidy, McDowell, et al. 2014)

There were four or five people inside and they just started $\boxed{\textit{firing}}$. Ms. Sanders was $\boxed{\textit{hit}}$ several times and was $\boxed{\textit{pronounced}}$ $\boxed{\textit{dead}}$ at the scene.

Dense annotation:

$$\boxed{\text{firing}} \xrightarrow{\text{INCLUDES}} \boxed{\text{hit}}$$

Dense links through transitivity

(Chambers, Cassidy, et al. 2014; Cassidy, McDowell, et al. 2014)

There were four or five people inside and they just started firing. Ms. Sanders was hit several times and was pronounced dead at the scene.

Dense annotation:

firing $\xrightarrow{\text{INCLUDES}}$ hit

firing $\xrightarrow{\text{BEFORE}}$ pronounced

Dense links through transitivity

(Chambers, Cassidy, et al. 2014; Cassidy, McDowell, et al. 2014)

There were four or five people inside and they just started firing. Ms. Sanders was hit several times and was pronounced dead at the scene.

Dense annotation:

firing $\xrightarrow{\text{INCLUDES}}$ hit

firing $\xrightarrow{\text{BEFORE}}$ pronounced

hit $\xrightarrow{\text{BEFORE}}$ pronounced

Dense links through transitivity

(Chambers, Cassidy, et al. 2014; Cassidy, McDowell, et al. 2014)

There were four or five people inside and they just started firing. Ms. Sanders was hit several times and was pronounced dead at the scene.

Dense annotation:

firing $\xrightarrow{\text{INCLUDES}}$ hit

firing $\xrightarrow{\text{BEFORE}}$ pronounced

hit $\xrightarrow{\text{BEFORE}}$ pronounced

firing $\xrightarrow{\text{BEFORE}}$ dead

Dense links through transitivity

(Chambers, Cassidy, et al. 2014; Cassidy, McDowell, et al. 2014)

There were four or five people inside and they just started $\boxed{\textit{firing}}$*. Ms. Sanders was* $\boxed{\textit{hit}}$ *several times and was* $\boxed{\textit{pronounced}}$ $\boxed{\textit{dead}}$ *at the scene.*

Dense annotation:

$\boxed{\text{firing}} \xrightarrow{\text{INCLUDES}} \boxed{\text{hit}}$

$\boxed{\text{firing}} \xrightarrow{\text{BEFORE}} \boxed{\text{pronounced}}$

$\boxed{\text{hit}} \xrightarrow{\text{BEFORE}} \boxed{\text{pronounced}}$

$\boxed{\text{firing}} \xrightarrow{\text{BEFORE}} \boxed{\text{dead}}$

$\boxed{\text{hit}} \xrightarrow{\text{BEFORE}} \boxed{\text{dead}}$

Dense links through transitivity

(Chambers, Cassidy, et al. 2014; Cassidy, McDowell, et al. 2014)

There were four or five people inside and they just started $\boxed{\textit{firing}}$. Ms. Sanders was $\boxed{\textit{hit}}$ several times and was $\boxed{\textit{pronounced}}$ $\boxed{\textit{dead}}$ at the scene.

Dense annotation:

$\boxed{\text{firing}} \xrightarrow{\text{INCLUDES}} \boxed{\text{hit}}$

$\boxed{\text{firing}} \xrightarrow{\text{BEFORE}} \boxed{\text{pronounced}}$

$\boxed{\text{hit}} \xrightarrow{\text{BEFORE}} \boxed{\text{pronounced}}$

$\boxed{\text{firing}} \xrightarrow{\text{BEFORE}} \boxed{\text{dead}}$

$\boxed{\text{hit}} \xrightarrow{\text{BEFORE}} \boxed{\text{dead}}$

$\boxed{\text{pronounced}} \xrightarrow{\text{AFTER}} \boxed{\text{dead}}$

Dense links through transitivity

(Chambers, Cassidy, et al. 2014; Cassidy, McDowell, et al. 2014)

There were four or five people inside and they just started firing. Ms. Sanders was hit several times and was pronounced dead at the scene.

Dense annotation:

firing $\xrightarrow{\text{INCLUDES}}$ hit

firing $\xrightarrow{\text{BEFORE}}$ pronounced

hit $\xrightarrow{\text{BEFORE}}$ pronounced

firing $\xrightarrow{\text{BEFORE}}$ dead

hit $\xrightarrow{\text{BEFORE}}$ dead

pronounced $\xrightarrow{\text{AFTER}}$ dead

Cascade model:

- Rule-based & learned classifiers
- Classifiers sorted by precision
- Transitivity after each classifier
- Earlier classifiers constrain later

On dense link annotations, the cascade model outperforms the top system from TempEval 2013

Conclusions

Summary:

- Timelines represented as event-time graphs
- Events and times accurately identified with supervised classifiers
  - Times normalized using rules
  - Web patterns, unsupervised models for semantics
- Linking events and times is challenging
  - Syntax, narrative containers suggest better links
  - Dependency parsing and using transitivity improve coverage

Open issues:

- Define “event” in a domain-independent way
- Normalize event-relative times, e.g. “3 weeks postop”
- Achieve high agreement + coverage in relation annotation

Thanks!

University of Colorado:

- Jim Martin
- Martha Palmer
- Tammy Sumner
- Sara Klingenstein
- Will Corvey
- Will Styler

Stanford University:

- Dan Jurafsky
- Andrey Gusev

KU Leuven, Belgium:

- Marie-Francine Moens
- Oleksandr Kolomiyets

United States Naval Academy:

- Nate Chambers

Brandeis University:

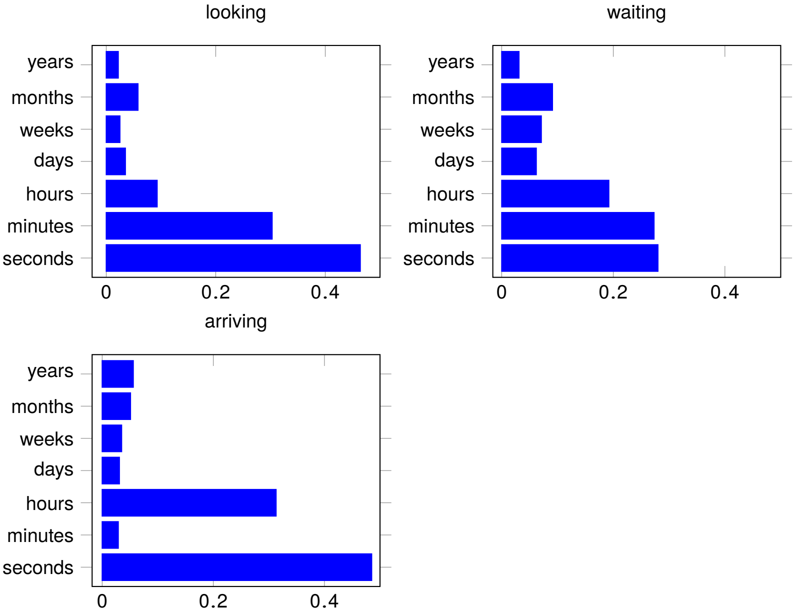
- James Pustejovsky
- Marc Verhagen

Children's Hospital Boston:

- Guergana Savova
- Tim Miller
- Dima Dligach
- Chen Lin
- Sameer Pradhan

And to i2b2 AUG NLP Workshop for inviting me here!

Event durations: looking, waiting, arriving



References I

Cassidy, Taylor, Bill McDowell, Nathanael Chambers, and Steven Bethard (2014). "An Annotation Framework for Dense Event Ordering". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. (Acceptance rate 25%). Baltimore, Maryland: Association for Computational Linguistics, pp. 501–506.

Chambers, Nathanael, Taylor Cassidy, Bill McDowell, and Steven Bethard (2014). "Dense Event Ordering with a Multi-Pass Architecture". In: *Transactions of the Association for Computational Linguistics* 2.

Bethard, Steven (2013). "A Synchronous Context Free Grammar for Time Normalization". In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. (Acceptance rate 24%). Seattle, Washington, USA: Association for Computational Linguistics, pp. 821–826.

Miller, Timothy, Steven Bethard, Dmitriy Dligach, Sameer Pradhan, Chen Lin, and Guergana Savova (2013). "Discovering Temporal Narrative Containers in Clinical Text". In: *Proceedings of the 2013 Workshop on Biomedical Natural Language Processing*. Sofia, Bulgaria: Association for Computational Linguistics, pp. 18–26.

Kolomiyets, Oleksandr, Steven Bethard, and Marie-Francine Moens (2012). "Extracting Narrative Timelines as Temporal Dependency Structures". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. (Acceptance rate 19%). Jeju Island, Korea: Association for Computational Linguistics, pp. 88–97.

Gusev, Andrey, Nathanael Chambers, Divye Raj Khilnani, Pranav Khaitan, Steven Bethard, and Dan Jurafsky (2011). "Using Query Patterns to Learn the Duration of Events". In: *International Conference on Computational Semantics*. (Acceptance rate 42%), pp. 145–154.

References II

Kolomiyets, Oleksandr, Steven Bethard, and Marie-Francine Moens (2011). "Model-Portability Experiments for Textual Temporal Analysis". In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. (Acceptance rate 26%), pp. 271–276.

Pustejovsky, James and Amber Stubbs (2011). "Increasing Informativeness in Temporal Annotation". In: *Proceedings of the 5th Linguistic Annotation Workshop*. Portland, Oregon, USA: Association for Computational Linguistics, pp. 152–160.

Pustejovsky, James, Kiyong Lee, Harry Bunt, and Laurent Romary (2010). "ISO-TimeML: An International Standard for Semantic Annotation". In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Ed. by Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias. Valletta, Malta: European Language Resources Association (ELRA). ISBN: 2-9517408-6-7.

Deschacht, Koen and Marie-Francine Moens (2009). "The Latent Words Language Model". In: *18th Annual Belgian-Dutch Conference on Machine Learning (Benelearn)*.

Bethard, Steven, James H. Martin, and Sara Klingenstein (2007). "Finding Temporal Structure in Text: Machine Learning of Syntactic Temporal Relations". In: *International Journal of Semantic Computing (IJSC)* 1.4, pp. 441–458.

Brewer, William F. and Edward H. Lichtenstein (1982). "Stories are to entertain: A structural-affect theory of stories". In: *Journal of Pragmatics* 6.5-6, pp. 473–486.

Johnson-Laird, P.N. (1980). "Mental Models in Cognitive Science". In: *Cognitive Science* 4.1, pp. 71–115.