

# **Time Expression Recognition Using a Constituent-based Tagging Scheme**

Xiaoshi Zhong and Erik Cambria

Nanyang Technological University  
{xszhong, cambria}@ntu.edu.sg

# Outline

- Time expression analysis
  - Datasets: TimeBank, Gigaword, WikiWars, Tweets
  - Findings: loose structure, differentiable
- Time expression recognition
  - TOMN: a constituent-based tagging scheme
  - Baselines: HeidelTime, SUTime, SynTime, ClearTK, UWTime
  - Datasets: TE-3, WikiWars, Tweets

# Time Expression - Examples

Today

Friday

September

Last week

2 years ago

September 2006

2006 September

January 30, 1998

1 September 2006

the third quarter of 1984

# Time Expression - Constituents

Today

Friday

September

Last week

2 years ago

September 2006

2006 September

January 30, 1998

1 September 2006

the third quarter of 1984

- Time token
  - Explicitly express time information
- Modifier
  - Modify time tokens
- Numeral
  - Numbers and ordinals (except year)

# Time Expression - Constituents

Today

Friday

September

Last week

2 years ago

September 2006

2006 September

January 30, 1998

1 September 2006

the third quarter of 1984

- Time token
  - Explicitly express time information
- Modifier
  - Modify time tokens
- Numeral
  - Numbers and ordinals (except year)

# Time Expression - Constituents

Today

Friday

September

Last week

2 years ago

September 2006

2006 September

January 30, 1998

1 September 2006

the third quarter of 1984

- Time token
  - Explicitly express time information
- Modifier
  - Modify time tokens
- Numeral
  - Numbers and ordinals (except year)

# Time Expression - Constituents

Today

Friday

September

Last week

2 years ago

September 2006

2006 September

January 30, 1998

1 September 2006

the third quarter of 1984

- Time token
  - Explicitly express time information
- Modifier
  - Modify time tokens
- Numeral
  - Numbers and ordinals (except year)

# Time Expression Analysis

- Datasets
  - TimeBank
  - Gigaword
  - WikiWars
  - Tweets
- Findings
  - Loose structure
  - Differentiable



# Time Expression Analysis - Datasets

- Datasets
  - TimeBank: a benchmark dataset used in TempEval evaluations
  - Gigaword: a large dataset with automatically generated labels
  - WikiWars: a war domain dataset collected from Wikipedia
  - Tweets: a tweet dataset collected from Twitter
- Dataset statistics

<b>Dataset</b>	<b>#Docs</b>	<b>#Words</b>	<b>#Timex</b>
TimeBank	183	61,418	1,243
Gigaword	2,452	666,309	12,739
WikiWars	22	119,468	2,671
Tweets	942	18,199	1,127

The datasets differ in size, source, domain, and text type, but their time expressions demonstrate similar characteristics

# Time Expression Analysis - Finding 1

- **Loose structure:** time expressions are formed by loose structure

# Time Expression Analysis - Finding 1

- **Loose structure:** time expressions are formed by loose structure
  - Loose collocation  
September  
September 2006  
1 September 2006
  - Exchangeable order  
September 2006  
2006 September

# Time Expression Analysis - Finding 1

- **Loose structure:** time expressions are formed by loose structure

- Loose collocation

September/B

September/B 2006/I

1/B September/I 2006/I

- Exchangeable order

September/B 2006/I

2006/B September/I

In perspective of position within time expressions, under BIO scheme, 'September' may appear as (i) Beginning or (ii) Inside word of time expressions

# Time Expression Analysis - Finding 1

- **Loose structure:** time expressions are formed by loose structure

- Loose collocation

September/U

September/B 2006/L

1/B September/I 2006/L

- Exchangeable order

September/B 2006/L

2006/B September/L

Under BILOU scheme, ‘September’ may appear as (1) Unit-word time expressions, (2) **B**eginning, (3) **I**nside, or (4) **L**ast word of time expressions

# Time Expression Analysis - Finding 1

- **Loose structure:** time expressions are formed by loose structure

Percentage of distinct time tokens and modifiers that appear in *different positions* within time expressions

Dataset	BIO Scheme		BILOU Scheme	
	Time Token	Modifier	Time Token	Modifier
TimeBank	58.18	33.33	63.64	33.33
Gigaword	61.29	45.83	77.05	46.00
WikiWars	53.57	26.19	61.40	29.55
Tweets	67.21	27.59	72.58	27.59

# Time Expression Analysis - Finding 2

- **Differentiable:** time tokens can differentiate time expressions from common text

# Time Expression Analysis - Finding 2

- **Differentiable:** time tokens can differentiate time expressions from common text

Percentage of time expression's constituents that appear in time expressions ( $P_{timex}$ ) and in common text ( $P_{text}$ )

Dataset	Time Token		Modifier		Numeral	
	$P_{timex}$	$P_{text}$	$P_{timex}$	$P_{text}$	$P_{timex}$	$P_{text}$
TimeBank	94.61	0.34	47.39	22.56	22.61	3.16
Gigaword	96.44	0.65	28.05	22.82	20.24	2.03
WikiWars	91.81	0.14	31.64	26.14	38.01	9.82
Tweets	96.01	0.50	21.38	13.03	18.81	0.128

$$P_{timex}(T) = \frac{\#timex \text{ that contain } T}{\#total \text{ timex}}$$

$$P_{text}(T) = \frac{\#tokens \text{ that are } T}{\#total \text{ tokens}}$$



# Fundamental Problem - Inconsistent Tag Assignment

- Position-based tagging scheme
  - BIO scheme: **B**eginning or **I**nside word of time expressions, **O**utside time expressions
  - BILOU scheme: **U**nit-word time expressions, **B**eginning, **I**nside, **L**ast word of multi-word time expressions, **O**utside time expressions
- Inconsistent tag assignment
  - During training, a word is assigned with different tags simply because the word appears in different positions within labeled chunks

# Inconsistent Tag Assignment

- Position-based tagging scheme
  - BILOU scheme: **B**eginning, **I**nside, **L**ast word of multi-word time expressions, **O**utside time expressions

# Inconsistent Tag Assignment

- Position-based tagging scheme
  - BILOU scheme: **B**eginning, **I**nside, **L**ast word of multi-word time expressions, **O**utside time expressions

1) September/U                      2) September/B 2006/L  
3) 2006/B September/L          4) 1/B September/I 2006/L

# Inconsistent Tag Assignment

- Position-based tagging scheme
  - BILOU scheme: **B**eginning, **I**nside, **L**ast word of multi-word time expressions, **O**utside time expressions

1) September/U                      2) September/B 2006/L  
3) 2006/B September/L          4) 1/B September/I 2006/L

- 1) (... , w=September, ..., U)
- 2) (... , w=September, ..., B)
- 3) (... , w=September, ..., L)
- 4) (... , w=September, ..., I )

# Inconsistent Tag Assignment

- Position-based tagging scheme
  - BILOU scheme: **U**nit-word time expressions, **B**eginning, **I**nside, **L**ast word of multi-word time expressions, **O**utside time expressions

1) September/U                      2) September/B 2006/L  
3) 2006/B September/L            4) 1/B September/I 2006/L

- 1) (... , w=September, ..., U)
- 2) (... , w=September, ..., B)
- 3) (... , w=September, ..., L)
- 4) (... , w=September, ..., I)

Inconsistent tag assignment reduces the predictive power of ‘September’

This contradicts finding 2 that time tokens can differentiate time expressions from common text



# Review

- Two findings
    - Finding 1: time expressions are formed by loose structure
    - Finding 2: time tokens can differentiate time expressions from common text
  - Finding 1 leads BIOES scheme to inconsistent tag assignment
    - Reduce the predictive power of time tokens
- 1) September/U                      2) September/B 2006/L  
3) 2006/B September/L          4) 1/B September/I 2006/L
- Under BIOES scheme, Finding 1 contradicts Finding 2

# Overcome Inconsistent Tag Assignment

- Constituent-based tagging scheme
  - TOMN scheme: **T**ime token, **M**odifier, **N**umeral, **O**utside time expressions
  - TOMN scheme assigns a word with a tag according to its constituent role

# Overcome Inconsistent Tag Assignment

- Constituent-based tagging scheme
  - TOMN scheme: **T**ime token, **M**odifier, **N**umeral, **O**utside time expressions
  - TOMN scheme assigns a word with a tag according to its constituent role

1) September/T

2) September/T 2006/T

3) 2006/T September/T

4) 1/N September/T 2006/T



# Overcome Inconsistent Tag Assignment

- Constituent-based tagging scheme
  - TOMN scheme: **T**ime token, **M**odifier, **N**umeral, **O**utside time expressions
  - TOMN scheme assigns a word with a tag according to its constituent role

1) September/T                      2) September/T 2006/T  
3) 2006/T September/T          4) 1/N September/T 2006/T

- 1) (... , w=September, ..., T)  
2) (... , w=September, ..., T)  
3) (... , w=September, ..., T)  
4) (... , w=September, ..., T)

# Overcome Inconsistent Tag Assignment

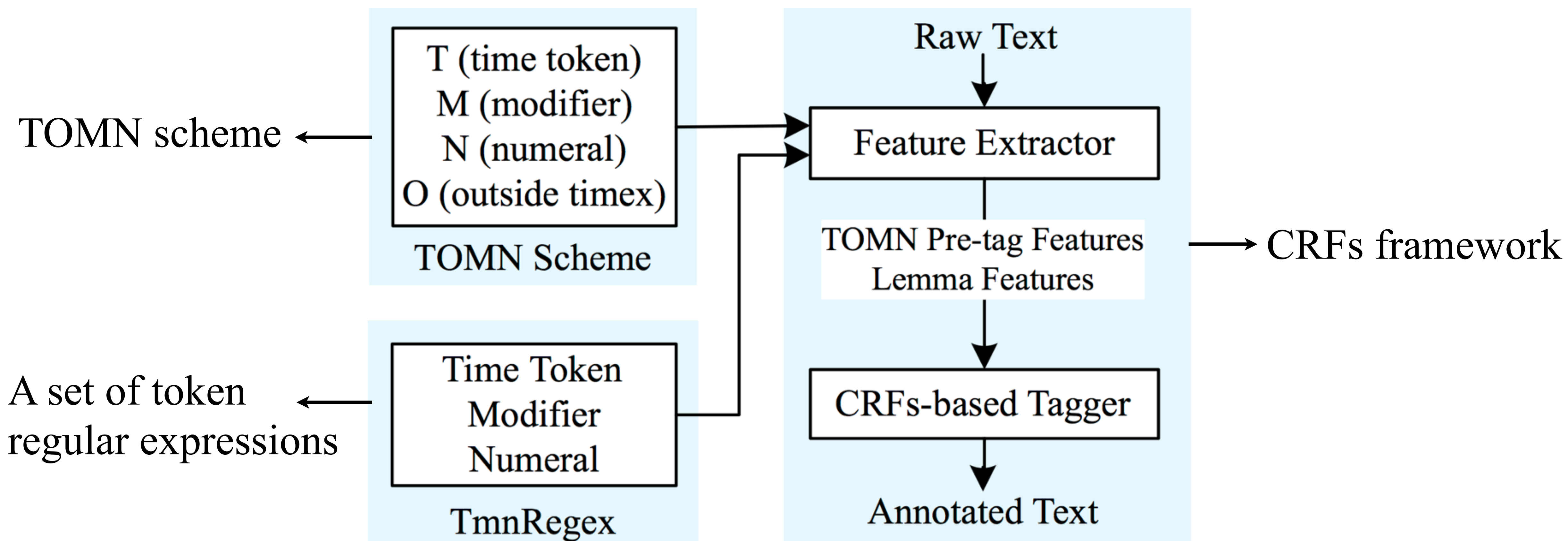
- Constituent-based tagging scheme
  - TOMN scheme: **T**ime token, **M**odifier, **N**umeral, **O**utside time expressions
  - TOMN scheme assigns a word with a tag according to its constituent role

1) September/T	2) September/T 2006/T
3) 2006/T September/T	4) 1/N September/T 2006/T

- 1) (... , w=September, ... , T)
- 2) (... , w=September, ... , T)
- 3) (... , w=September, ... , T)
- 4) (... , w=September, ... , T)

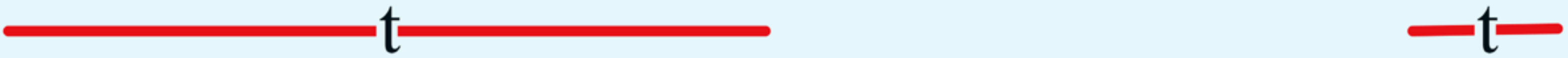
Consistent tag assignment protects  
'September's predictive power

# Time Expression Recognition - TOMN



# Time Expression Recognition - Examples

- Non-O words that appear together form a time expression

  
On/o September/T 1/N ,/M 1939/T ,/O ... state/o in/o 1939/T ./O

  
... in/o a/M few/M days/T and/M weeks/T respectively/o ./O

# Time Expression Recognition - Experiments

- Our method: TOMN
- Baselines
  - HeidelTime: rule-based
  - SUTime: rule-based
  - SynTime: type-based
  - ClearTK: learning-based
  - UWTime: learning-based
- Datasets
  - TE-3, WikiWars, Tweets



Performance of TOMN and baselines. **Best results** are in boldface and second best are underlined.

Dataset	Method	Strict Match			Relexed Match		
		<i>Pr.</i>	<i>Re.</i>	<i>F1</i>	<i>Pr.</i>	<i>Re.</i>	<i>F1</i>
TE-3	HeidelTime (Strotgen et al., 2013)	83.85	78.99	81.34	93.08	87.68	90.30
	SUTime (Chang and Manning, 2013)	78.72	80.43	79.57	89.36	91.30	90.32
	SynTime (Zhong et al., 2017)	<u>91.43</u>	<b>92.75</b>	<b>92.09</b>	94.29	<b>95.65</b>	<b>94.96</b>
	ClearTK (Bethard, 2013)	85.90	79.70	82.70	93.75	86.96	90.23
	UWTime (Lee et al., 2014)	86.10	80.40	83.10	<u>94.60</u>	88.40	91.40
	TOMN	<b>92.59</b>	90.58	91.58	<b>95.56</b>	93.48	94.51
WikiWars	HeidelTime (Strotgen et al., 2013)	<b>88.20</b>	78.50	<u>83.10</u>	95.80	85.40	90.30
	SUTime	78.61	76.69	76.64	95.74	89.57	92.55
	SynTime (Zhong et al., 2017)	80.00	80.22	80.11	92.16	<b>92.41</b>	92.29
	ClearTK	87.69	<u>80.28</u>	<b>83.82</b>	<u>96.80</u>	90.54	<u>93.54</u>
	UWTime (Lee et al., 2014)	<u>87.70</u>	78.80	83.00	<b>97.60</b>	87.60	92.30
	TOMN	84.57	<b>80.48</b>	82.47	96.23	<u>92.35</u>	<b>94.25</b>
Tweets	HeidelTime	<b>91.67</b>	74.26	82.05	<u>96.88</u>	78.48	86.71
	SUTime	77.69	79.32	78.50	88.84	90.72	89.77
	SynTime (Zhong et al., 2017)	89.52	<u>94.07</u>	<u>91.74</u>	93.55	<b>98.31</b>	<b>95.87</b>
	ClearTK	86.83	75.11	80.54	96.59	83.54	89.59
	UWTime	88.36	70.76	78.59	<b>97.88</b>	78.39	87.06
	TOMN	<u>90.69</u>	<b>94.51</b>	<b>92.56</b>	93.52	97.47	95.45

Performance of TOMN and baselines. **Best results** are in boldface and second best are underlined.

Dataset	Method	Strict Match			Relexed Match		
		<i>Pr.</i>	<i>Re.</i>	<i>F1</i>	<i>Pr.</i>	<i>Re.</i>	<i>F1</i>
TE-3	HeidelTime (Strotgen et al., 2013)	83.85	78.99	81.34	93.08	87.68	90.30
	SUTime (Chang and Manning, 2013)	78.72	80.43	79.57	89.36	91.30	90.32
	SynTime (Zhong et al., 2017)	<u>91.43</u>	<b>92.75</b>	<b>92.09</b>	94.29	<b>95.65</b>	<b>94.96</b>
	ClearTK (Bethard, 2013)	85.90	79.70	82.70	93.75	86.96	90.23
	UWTime (Lee et al., 2014)	86.10	80.40	83.10	<u>94.60</u>	88.40	91.40
	TOMN	<b>92.59</b>	90.58	91.58	<b>95.56</b>	93.48	94.51
WikiWars	HeidelTime (Strotgen et al., 2013)	<b>88.20</b>	78.50	<u>83.10</u>	95.80	85.40	90.30
	SUTime	78.61	76.69	76.64	95.74	89.57	92.55
	SynTime (Zhong et al., 2017)	80.00	80.22	80.11	92.16	<b>92.41</b>	92.29
	ClearTK	87.69	<u>80.28</u>	<b>83.82</b>	<u>96.80</u>	90.54	<u>93.54</u>
	UWTime (Lee et al., 2014)	<u>87.70</u>	78.80	83.00	<b>97.60</b>	87.60	92.30
	TOMN	84.57	<b>80.48</b>	82.47	96.23	<u>92.35</u>	<b>94.25</b>
Tweets	HeidelTime	<b>91.67</b>	74.26	82.05	<u>96.88</u>	78.48	86.71
	SUTime	77.69	79.32	78.50	88.84	90.72	89.77
	SynTime (Zhong et al., 2017)	89.52	<u>94.07</u>	<u>91.74</u>	93.55	<b>98.31</b>	<b>95.87</b>
	ClearTK	86.83	75.11	80.54	96.59	83.54	89.59
	UWTime	88.36	70.76	78.59	<b>97.88</b>	78.39	87.06
	TOMN	<u>90.69</u>	<b>94.51</b>	<b>92.56</b>	93.52	97.47	95.45

# Time Expression Recognition - Experiments

- Cross-dataset experiments
  - Train on one dataset's training set
  - Test on other datasets' test sets
- Learning-based baselines
  - ClearTK
  - UWTime
- Datasets
  - TE-3
  - WikiWars
  - Tweets



# Cross-dataset performance on **test set of TE-3**

(Color background indicates single-dataset results)

Training Set	Method	Strict Match			Relexed Match		
		<i>Pr.</i>	<i>Re.</i>	<i>F1</i>	<i>Pr.</i>	<i>Re.</i>	<i>F1</i>
TE-3	ClearTK	85.90	79.70	82.70	93.75	86.96	90.23
	UWTime	86.10	80.40	83.10	94.60	88.40	91.40
	TOMN	<b>92.59</b>	<b>90.58</b>	<b>91.58</b>	<b>95.56</b>	<b>93.48</b>	<b>94.51</b>
WikiWars	ClearTK	65.67	63.77	64.71	87.31	84.78	86.03
	UWTime	76.92	72.46	74.63	88.46	83.33	85.82
	TOMN	<b>84.06</b>	<b>84.06</b>	<b>84.06</b>	<b>93.48</b>	<b><u>93.48</u></b>	<b>93.48</b>
Tweets	ClearTK	72.59	71.01	71.79	<b>93.33</b>	91.30	92.31
	UWTime	80.00	72.46	76.05	92.80	84.06	88.21
	TOMN	<b>85.42</b>	<b>89.13</b>	<b>87.23</b>	91.67	<b>95.65</b>	<b>93.62</b>

# Cross-dataset performance on **test set of WikiWars**

Training Set	Method	Strict Match			Relexed Match		
		<i>Pr.</i>	<i>Re.</i>	<i>F1</i>	<i>Pr.</i>	<i>Re.</i>	<i>F1</i>
TE-3	ClearTK	74.38	60.76	66.89	<b>97.54</b>	79.68	87.71
	UWTime	<b>87.01</b>	<b>79.34</b>	<b>83.00</b>	96.07	87.60	91.64
	TOMN	82.18	75.65	79.07	96.26	<b>87.93</b>	<b>91.90</b>
WikiWars	ClearTK	87.69	80.28	<b>83.82</b>	96.80	90.54	93.54
	UWTime	<b>87.70</b>	78.80	83.00	<b>97.60</b>	87.60	92.30
	TOMN	84.57	<b>80.48</b>	82.47	96.23	<b>92.35</b>	<b>94.25</b>
Tweets	ClearTK	57.75	54.73	56.20	91.93	87.12	<b>89.46</b>
	UWTime	<b>80.28</b>	62.81	<b>70.48</b>	<b>94.37</b>	73.83	82.84
	TOMN	60.29	<b>66.00</b>	63.02	84.74	<b>92.76</b>	88.57

# Cross-dataset performance on **test set of Tweets**

Training Set	Method	Strict Match			Relexed Match		
		<i>Pr.</i>	<i>Re.</i>	<i>F1</i>	<i>Pr.</i>	<i>Re.</i>	<i>F1</i>
TE-3	ClearTK	81.16	47.26	59.73	97.10	56.54	71.47
	UWTime	89.66	65.82	75.91	94.83	69.62	80.29
	TOMN	<b>92.92</b>	<b>88.61</b>	<b>90.71</b>	<b>96.90</b>	<b>92.41</b>	<b>94.60</b>
WikiWars	ClearTK	72.48	45.57	55.96	95.30	59.92	73.58
	UWTime	<b>87.43</b>	61.60	72.28	<b>95.81</b>	67.61	79.21
	TOMN	85.00	<b>86.08</b>	<b>85.53</b>	93.75	<u><b>94.94</b></u>	<b>94.34</b>
Tweets	ClearTK	86.83	75.11	80.54	96.59	83.54	89.59
	UWTime	88.36	70.76	78.59	<b>97.88</b>	78.39	87.06
	TOMN	<b>90.69</b>	<b>94.51</b>	<b>92.56</b>	93.52	<b>97.47</b>	<b>95.45</b>

# Time Expression Recognition - Efficiency

- TOMN is more efficient

Runtime of going through a whole process (unit: seconds)

Method	TE-3	WikiWars	Tweets
ClearTK	152	223	86
UWTime	864	1,050	160
TOMN	36	48	42

# Summary

- Have two findings
  - Loose structure
  - Differentiable
- Reveal a fundamental problem in position-based tagging scheme
  - Inconsistent tag assignment
- Define a constituent-based tagging scheme
  - Good results
  - Less time