

Joint Learning for Event Coreference Resolution

Jing Lu and Vincent Ng

Human Language Technology Research Institute

University of Texas at Dallas



The world has lost a great man, former Prime Minister Julia Gillard said.



Event Coreference Subtasks

- **Trigger Detection:** Determine trigger word and subtype

Life.Die

Nelson Mandela has **died** at age 95.

The world has **lost** a great man, **former** Prime Minister Julia Gillard **said**.

Life.Die

Personnel.End-Position

Contact.Broadcast



Event Coreference Subtasks

- **Anaphoricity Detection:** Determine if an event mention is anaphoric

Not Anaphoric

Nelson Mandela has **died** at age 95.

The world has **lost** a great man, **former** Prime Minister Julia Gillard **said**.

Anaphoric

Not Anaphoric

Not Anaphoric



Event Coreference Subtasks

- **Coreference:** Determine which event mentions refer to the same real-world event

Nelson Mandela has **died** at age 95.

The world has **lost** a great man, **former** Prime Minister Julia Gillard **said**.



Trigger

Anaphoricity

Coref

These three event subtasks are **inter-dependent**



Trigger-Coreference Dependencies

Nelson Mandela has **died** aged 95 surrounded

Trigger

Anaphoricity

Coref

The world has **lost** a great man...

Life.Die Personnel.End-Position Life.Injury NULL



Trigger-Coreference Dependencies

Life.Die

Nelson Mandela has **died** aged 95 surrounded

Trigger

Anaphoricity

Coref

The world has **lost** a great man...

Life.Die Personnel.End-Position Life.Injury NULL



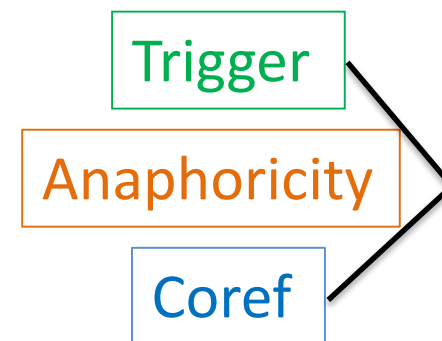
Trigger-Coreference Dependencies

Life.Die
Nelson Mandela has **died** aged 95 surrounded



The world has **lost** a great man...

Life.Die Personnel.End-Position ~~Life.Injury~~ ~~NULL~~



Coreferent event mentions must have same subtype



Trigger-Anaphoricity Dependencies

Trigger

Anaphoricity

Coref

The world has **lost** a great man...

Life.Die Personnel.End-Position Life.Injury NULL



Trigger-Anaphoricity Dependencies

Trigger

Anaphoricity

Coref

The world has **lost** a great man...

Life.Die Personnel.End-Position Life.Injury NULL

Anaphoric



Trigger-Anaphoricity Dependencies

Trigger

Anaphoricity

Coref

The world has **lost** a great man...

Life.Die Personnel.End-Position Life.Injury ~~NULL~~

Anaphoric

Anaphoric event mentions cannot have subtype NULL



Coref-Anaphoricity Dependencies

Trigger

Anaphoricity

Coref

The world has **lost** a great man, **former...**

Non-anaphoric



Coref-Anaphoricity Dependencies

Trigger

Anaphoricity

Coref

The world has **lost** a great man, former...

Non-anaphoric

Non-anaphoric event mentions cannot have antecedents



Top Event Coreference Resolvers in KBP'16

- Largely capture inter-dependencies with a pipeline architecture



- Weakness: error propagation

Lu and Ng (2016), Liu et al. (2016), Nguyen et al. (2016)



Goal

- Improve event coreference performance by **jointly learning three subtasks**: trigger detection, anaphoricity detection, and coreference
 - Motivated by Durrett and Klein's (2014) model and its success in jointly modeling the key tasks in the entity stack
 - Event anaphoricity detection introduced as an auxiliary task to mediate the interactions between trigger detection and coreference



Plan for the Talk

- Models
 - **Independent model** (unary, per-task factors)

Trigger

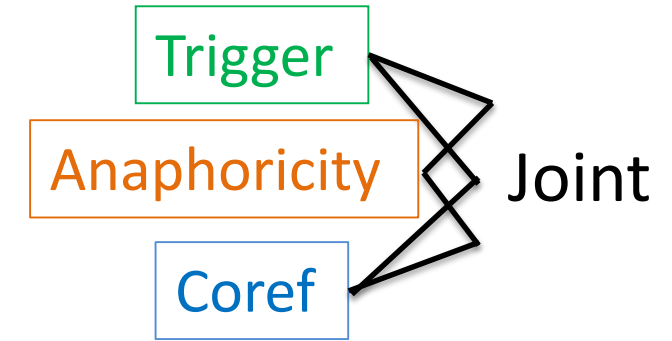
Anaphoricity

Coref



Plan for the Talk

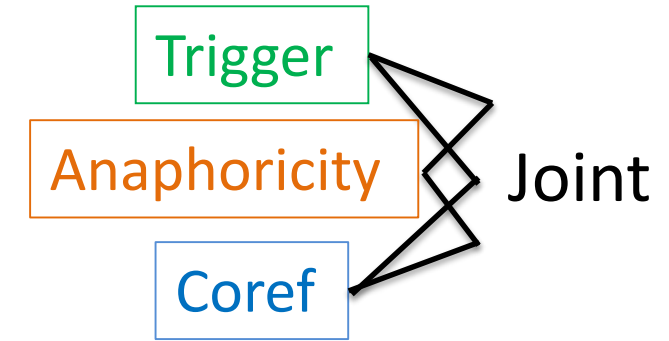
- Models
 - **Independent model** (unary, per-task factors)
 - **Joint model** (higher-order, cross-task factors)





Plan for the Talk

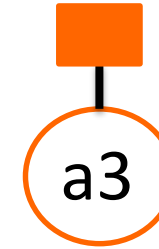
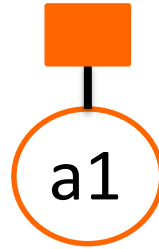
- Models
 - **Independent model** (unary, per-task factors)
 - **Joint model** (higher-order, cross-task factors)
- Evaluation



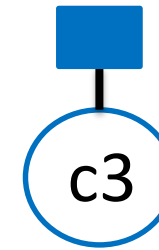
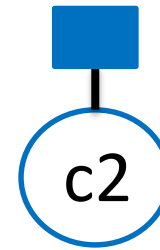
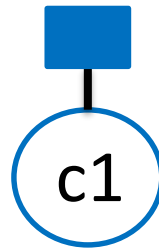


Model

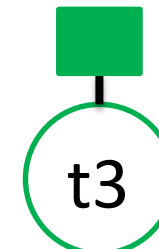
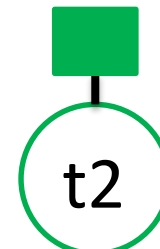
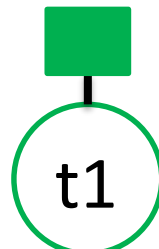
Anaphoricity



Coref



Trigger

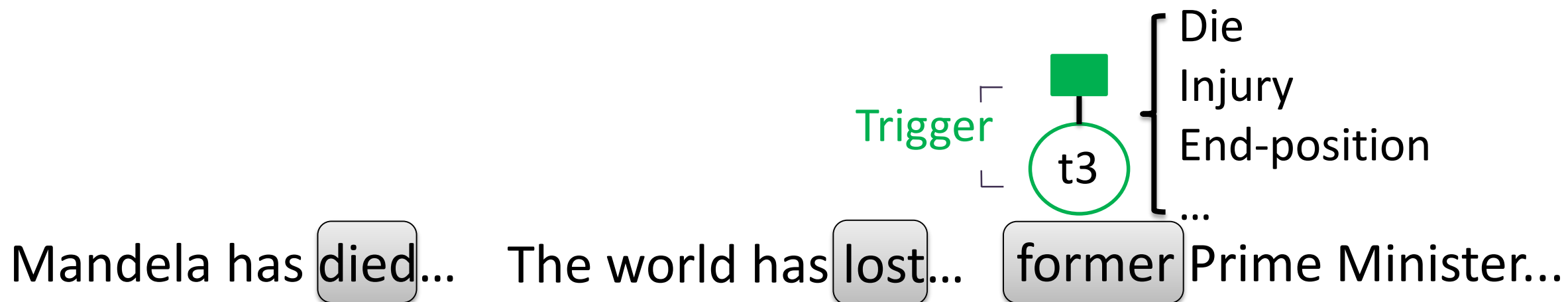


Mandela has **died**...The world has **lost**...

former Prime...

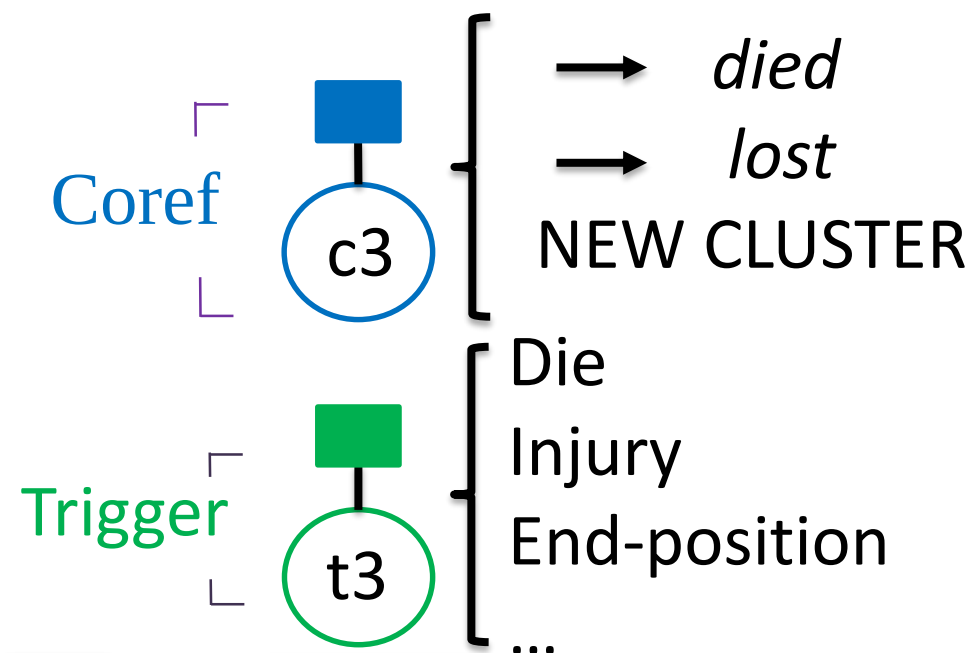


Model





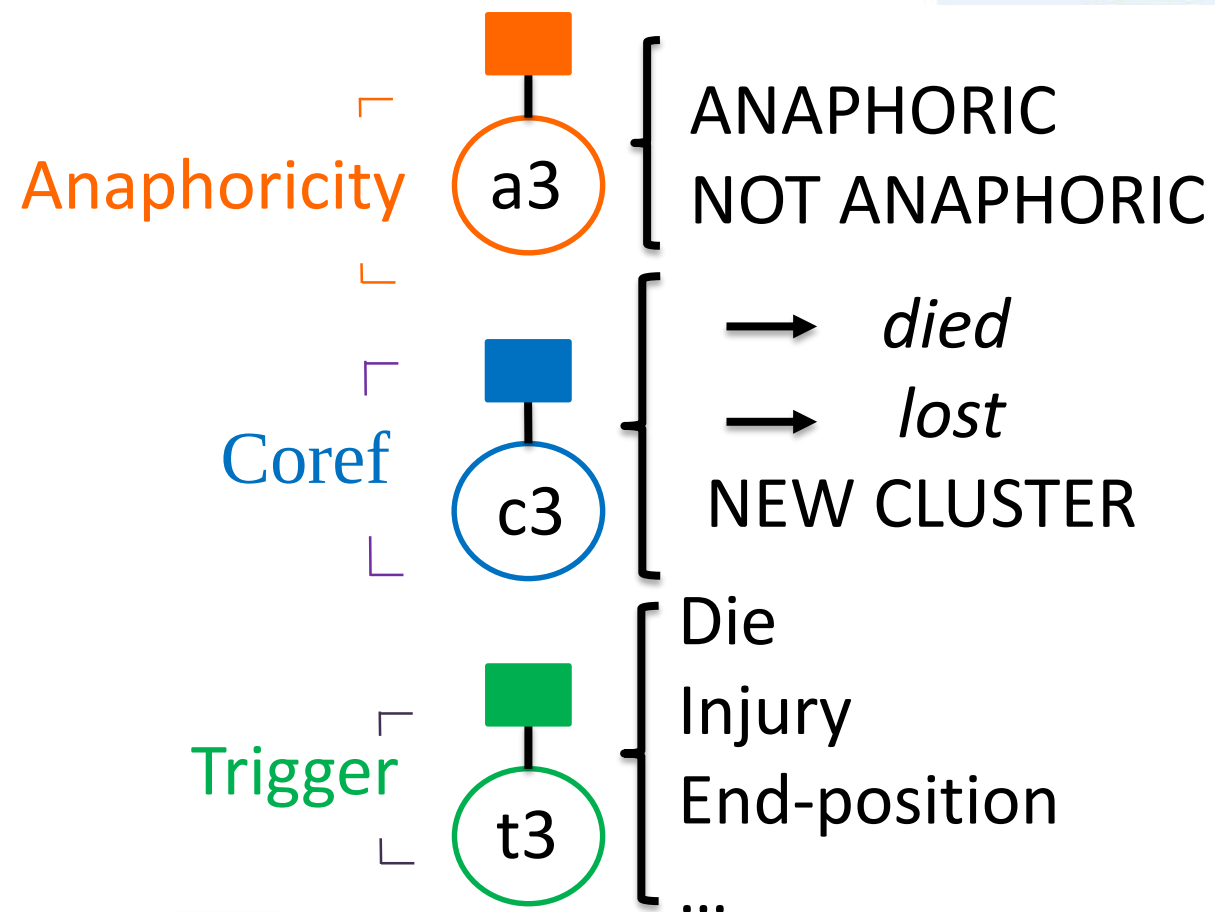
Model



Mandela has died... The world has lost... former Prime Minister...



Model



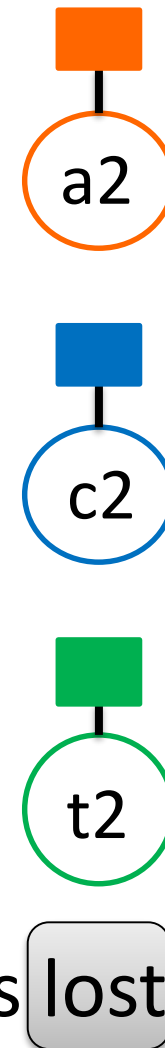
Mandela has died... The world has lost... former Prime Minister...



Model

- Log-linear probability distribution over the variables

$$p(t, a, c | x; \Theta) \propto \exp\left(\sum_i \theta_i f_i(t, a, c, x)\right)$$



The world has lost...

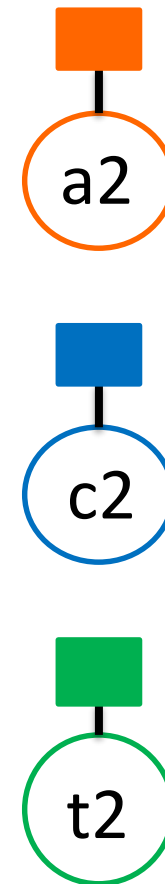


Model

- Log-linear probability distribution over the variables

$$p(t, a, c|x; \Theta) \propto \exp\left(\sum_i \theta_i f_i(t, a, c, x)\right)$$

Input document
context



The world has lost...



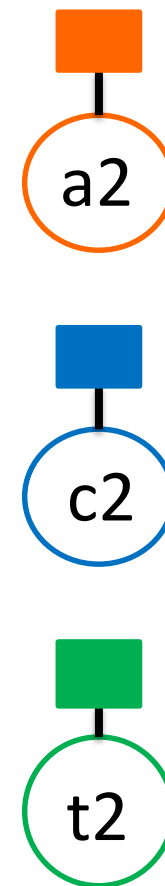
Model

- Log-linear probability distribution over the variables

$$p(t, a, c | x; \Theta) \propto \exp\left(\sum_i \theta_i f_i(t, a, c, x)\right)$$

Feature
function

Input document
context



The world has lost...



Model

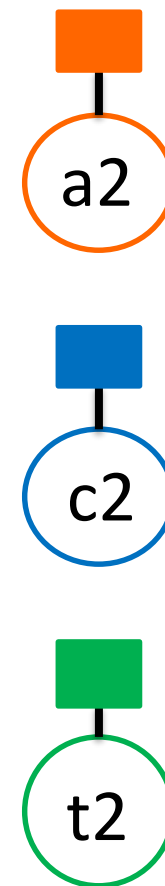
- Log-linear probability distribution over the variables

$$p(t, a, c | x; \Theta) \propto \exp\left(\sum_i \theta_i f_i(t, a, c, x)\right)$$

Weight vector

Feature function

Input document context



The world has lost...

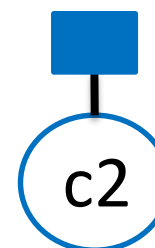


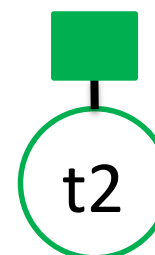
Independent Model

$$p(t, a, c | x; \Theta) \propto \exp(\Theta^T F(t, a, c, x))$$

$$F_{Indep} = \sum_{i=1}^n f(a_i, x) + f(c_i, x) + f(t_i, x)$$

$$f(a_i, x)$$

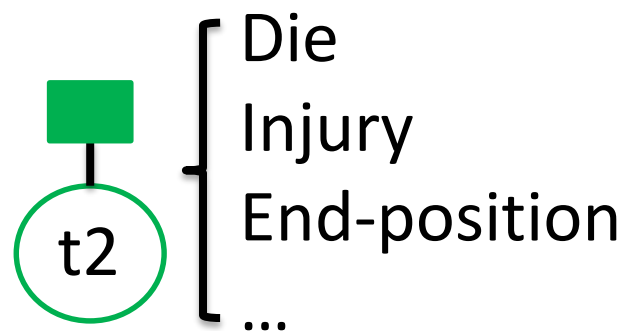

$$f(c_i, x)$$


$$f(t_i, x)$$


The world has lost...



Trigger Detection Features



The world has **lost** a great man, former Prime Minister...



Trigger Detection Features



The world has **lost** a great man, former Prime Minister...

Trigger word

lost



Trigger Detection Features



The world has **lost** a great man, former Prime Minister...

Trigger word	lost
Trigger Lemma	lose



Trigger Detection Features



The world has lost a great man, former Prime Minister...

Trigger word	lost
Trigger Lemma	lose
Bigram	has_lost, lost_a



Trigger Detection Features



The world has **lost** a great man, former Prime Minister...

Trigger word	lost
Trigger Lemma	lose
Bigram	has_lost, lost_a
Nearest Entity (Left)	world
Nearest Entity (Right)	man

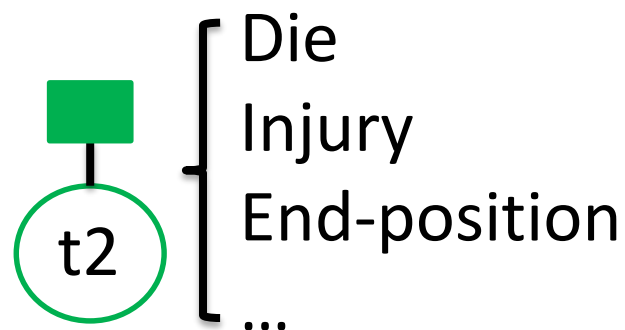


Trigger Detection Features



The world has **lost** a great man, former Prime Minister...

Trigger word	lost
Trigger Lemma	lose
Bigram	has_lost, lost_a
Nearest Entity (Left)	world
Nearest Entity (Right)	man
Dependencies	World, man



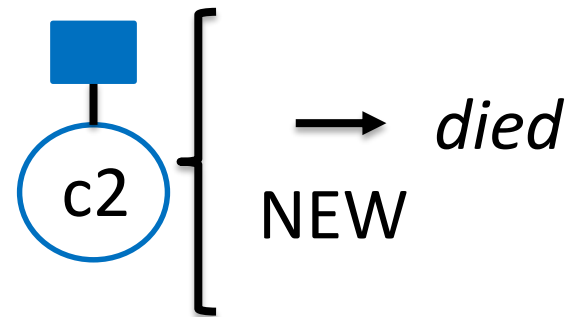
The world has **lost** a great man, former Prime Minister...

Trigger word	lost
Trigger Lemma	lose
Bigram	has_lost, lost_a
Nearest Entity (Left)	world
Nearest Entity (Right)	man
Dependencies	World, man

7/31/2017



Event Coreference

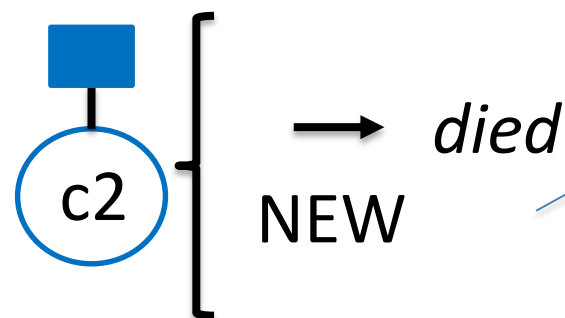


The world has lost...



Event Coreference Features

Features representing the current mention



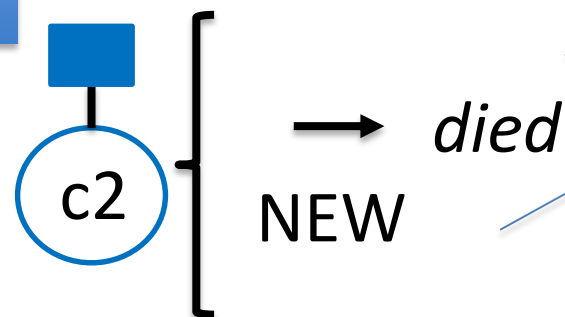
The world has **lost...**



Event Coreference Features

Features representing the current mention

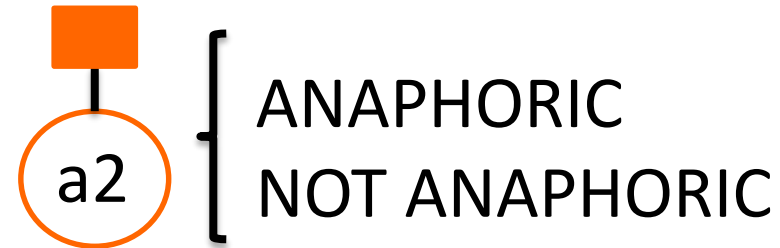
Features representing a preceding mention and its relationship with the current mention



The world has **lost...**



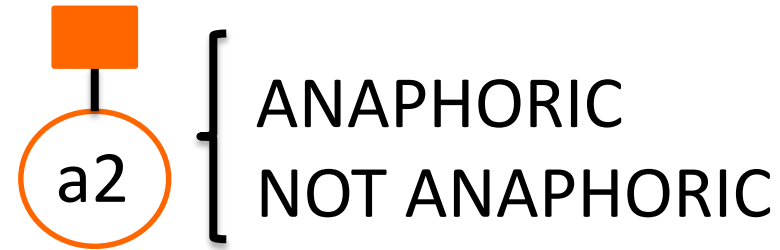
Anaphoricity Detection Features



The world has lost a great man, former Prime Minister...



Anaphoricity Detection Features

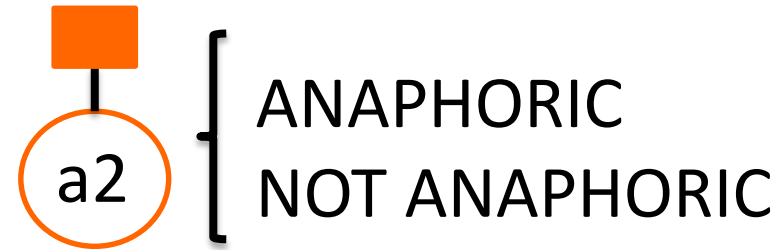


The world has **lost** a great man, former Prime Minister...

Trigger words of preceding mentions died



Anaphoricity Detection Features

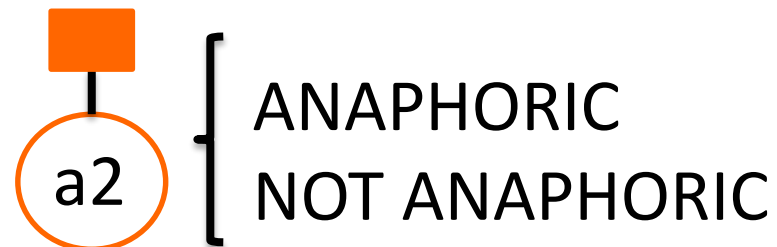


The world has lost a great man, former Prime Minister...

Trigger words of preceding mentions	died
Same lemma	No



Anaphoricity Detection Features



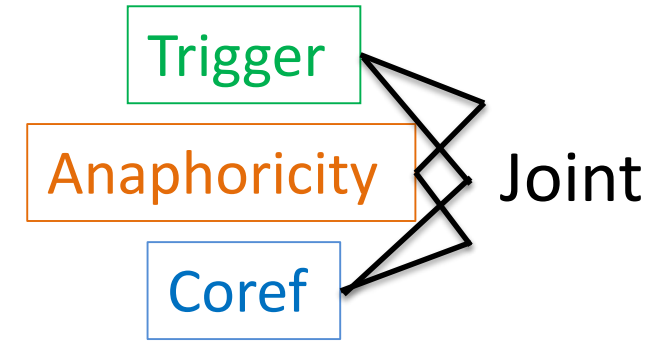
The world has **lost** a great man, former Prime Minister...

Trigger words of preceding mentions	died
Same lemma	No
Anaphoricity probability	0.75



Plan for the Talk

- Models
 - Independent model (unary, per-task factors)
 - **Joint model** (higher-order, cross-task factors)
- Evaluation

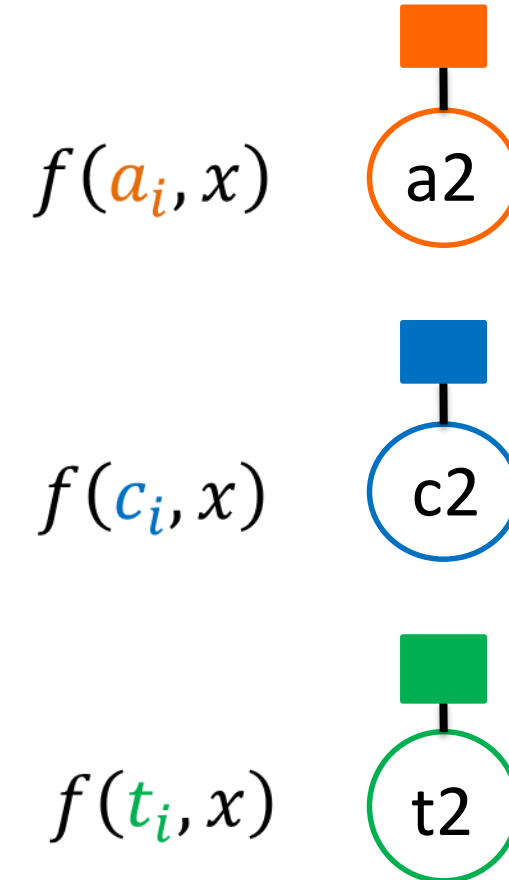




Joint Model

$$p(t, a, c | x; \Theta) \propto \exp(\Theta^T F(t, a, c, x))$$

$$F_{Indep} = \sum_{i=1}^n f(a_i, x) + f(c_i, x) + f(t_i, x)$$



The world has lost...

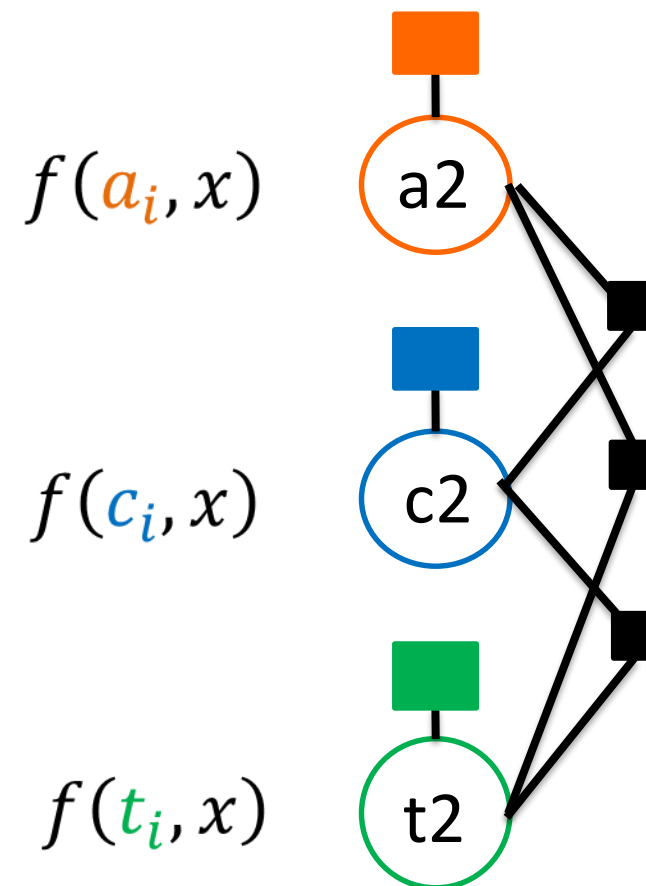


Joint Model

$$p(t, a, c | x; \Theta) \propto \exp(\Theta^T F(t, a, c, x))$$

$$F_{Indep} = \sum_{i=1}^n f(a_i, x) + f(c_i, x) + f(t_i, x)$$

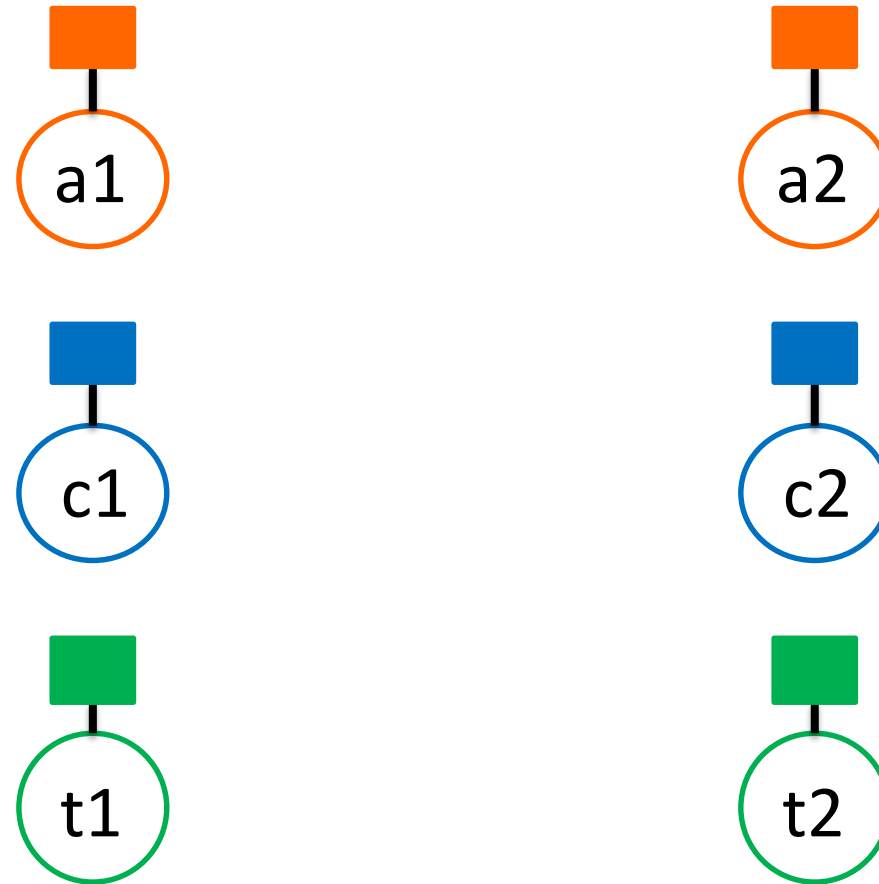
$$F_{Joint} = F_{Indep} + \sum \text{joint factors}$$



The world has lost...



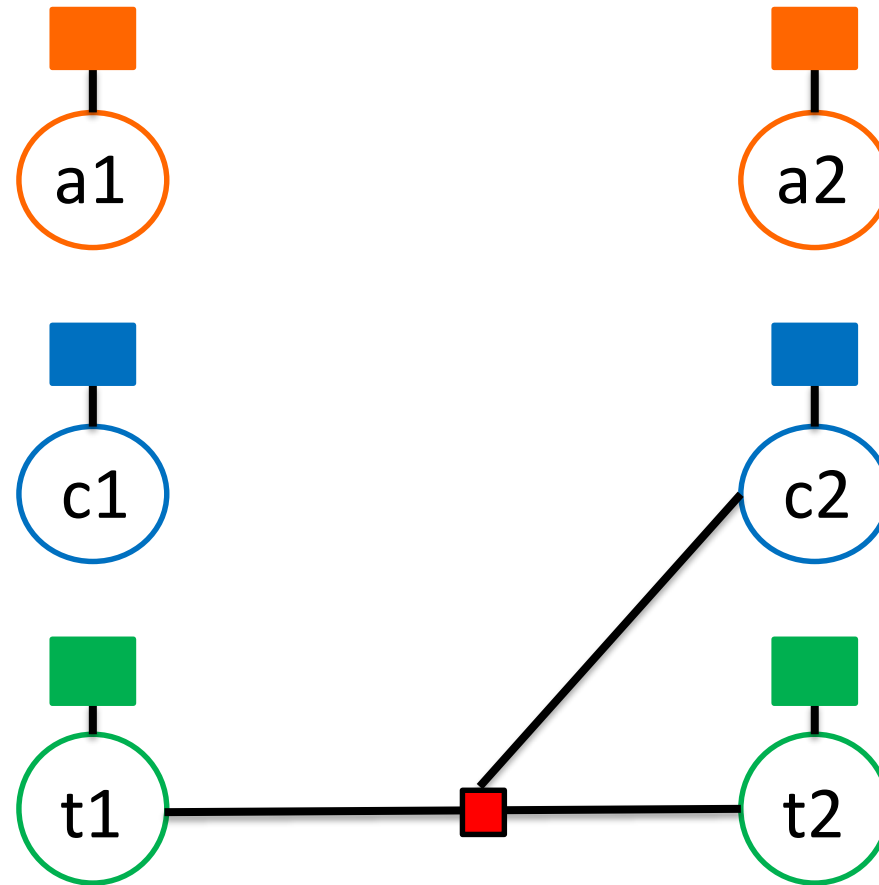
Trigger – Coref Factors



Mandela has died... The world has lost...



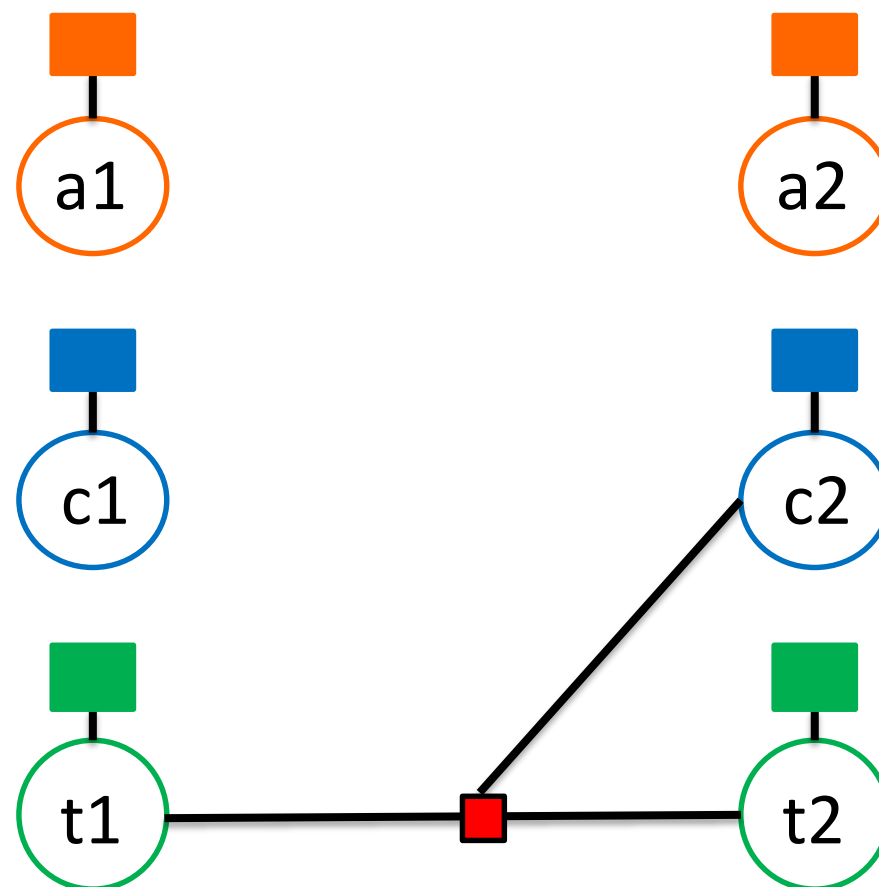
Trigger – Coref Factors



Mandela has died... The world has lost...

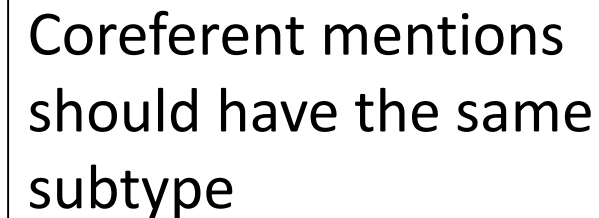


Trigger – Coref Factors



Coreferent mentions should have the same subtype

Mandela has died... The world has lost...

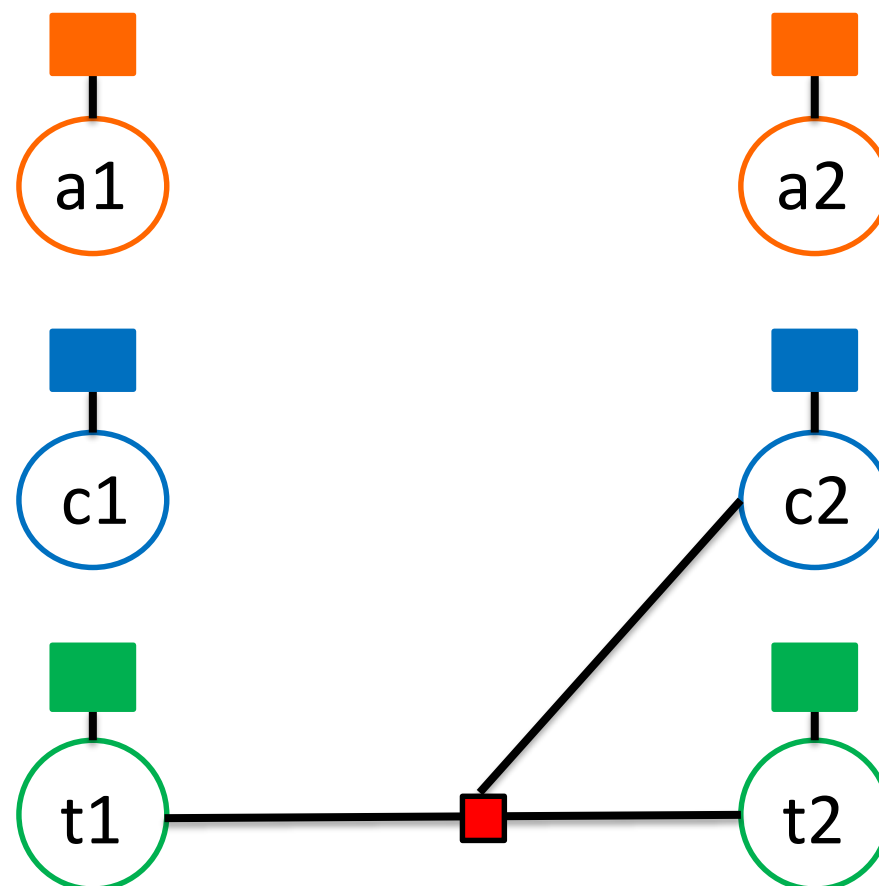


Subtypes=Die-Die

7/31/2017



Trigger – Coref Factors



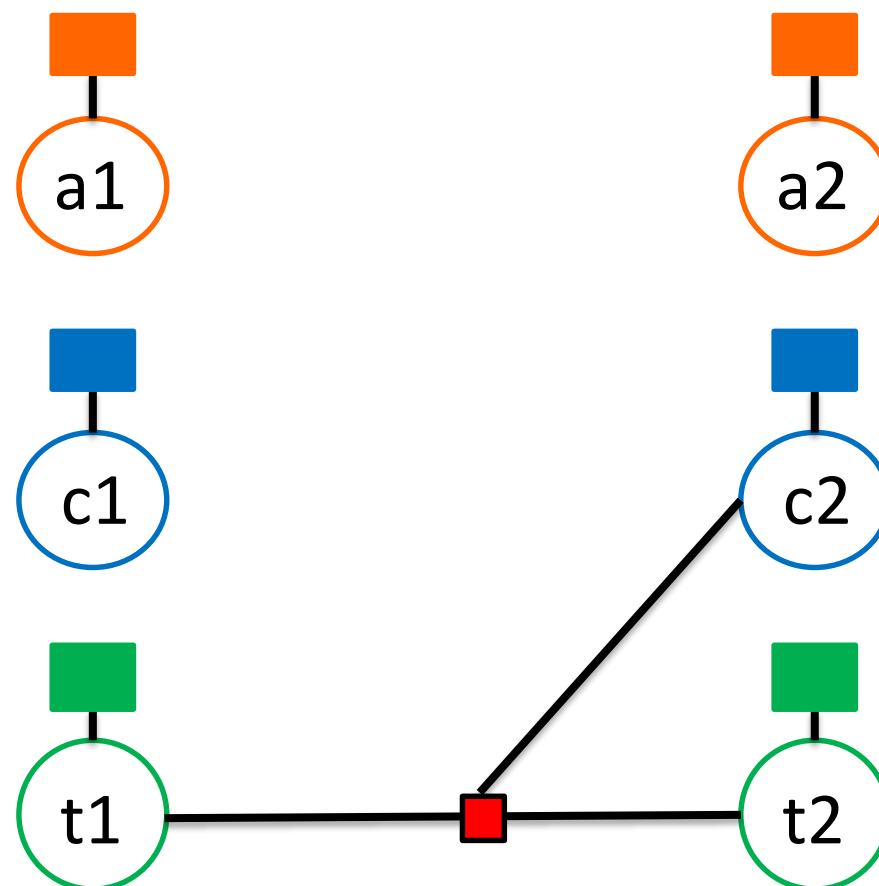
Coreferent mentions
should have the same
subtype

Subtypes=Die-Die
CurrType=Die, Prev = died

Mandela has died... The world has lost...



Trigger – Coref Factors



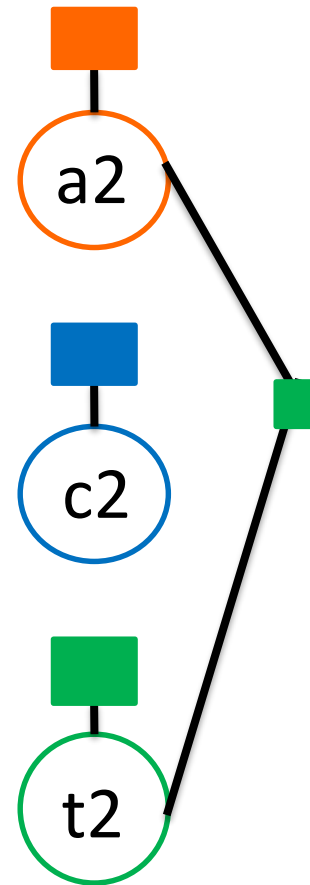
Coreferent mentions
should have the same
subtype

Subtypes=Die-Die
CurrType=Die, Prev = died
PrevType=Die, Curr = lost

Mandela has died... The world has lost...



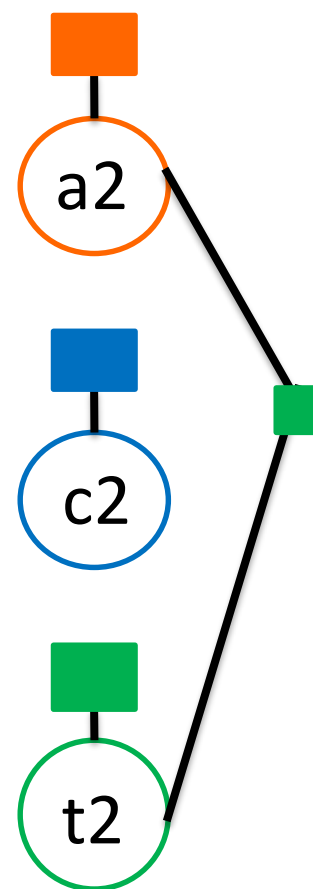
Trigger - Anaphoricity Factors



The world has lost...



Trigger - Anaphoricity Factors

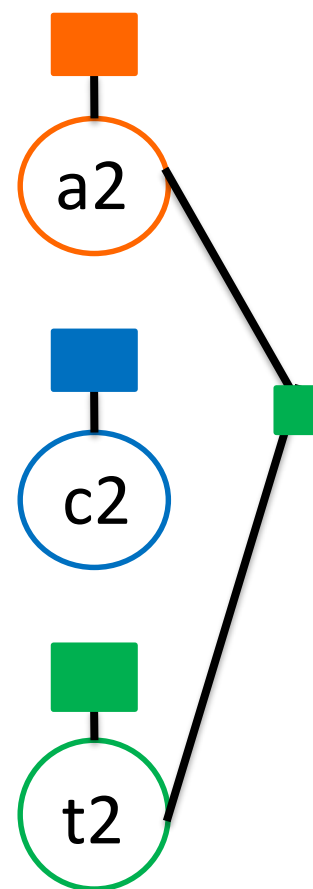


Anaphoric event
mention cannot have
subtype NULL

The world has lost...



Trigger - Anaphoricity Factors



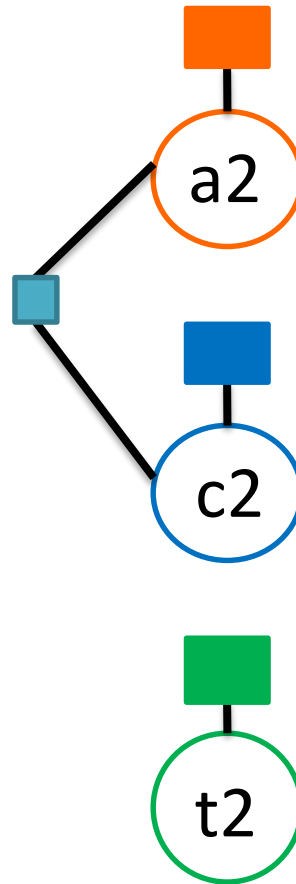
Anaphoric event
mention cannot have
subtype NULL

Subtypes=Die, Lemma=lose,
Anaphoricity = Yes

The world has lost...



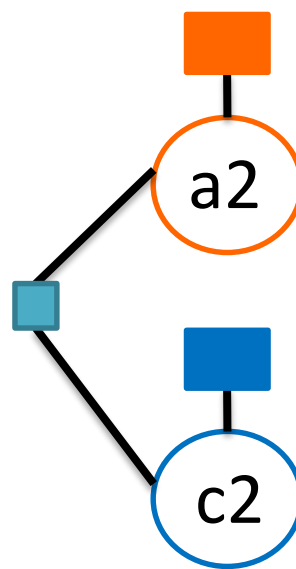
Coref - Anaphoricity Factors



The world has lost...



Coref - Anaphoricity Factors

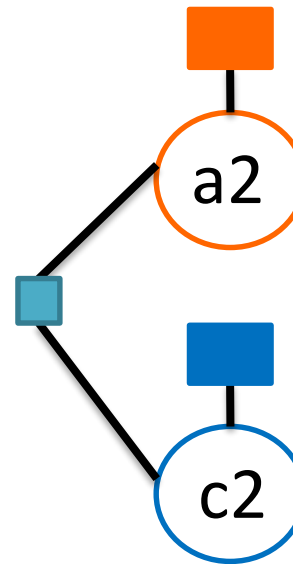


Anaphoric event mentions
should have antecedents

The world has 

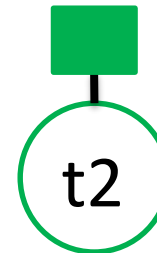


Coref - Anaphoricity Factors



Anaphoric event mentions
should have antecedents

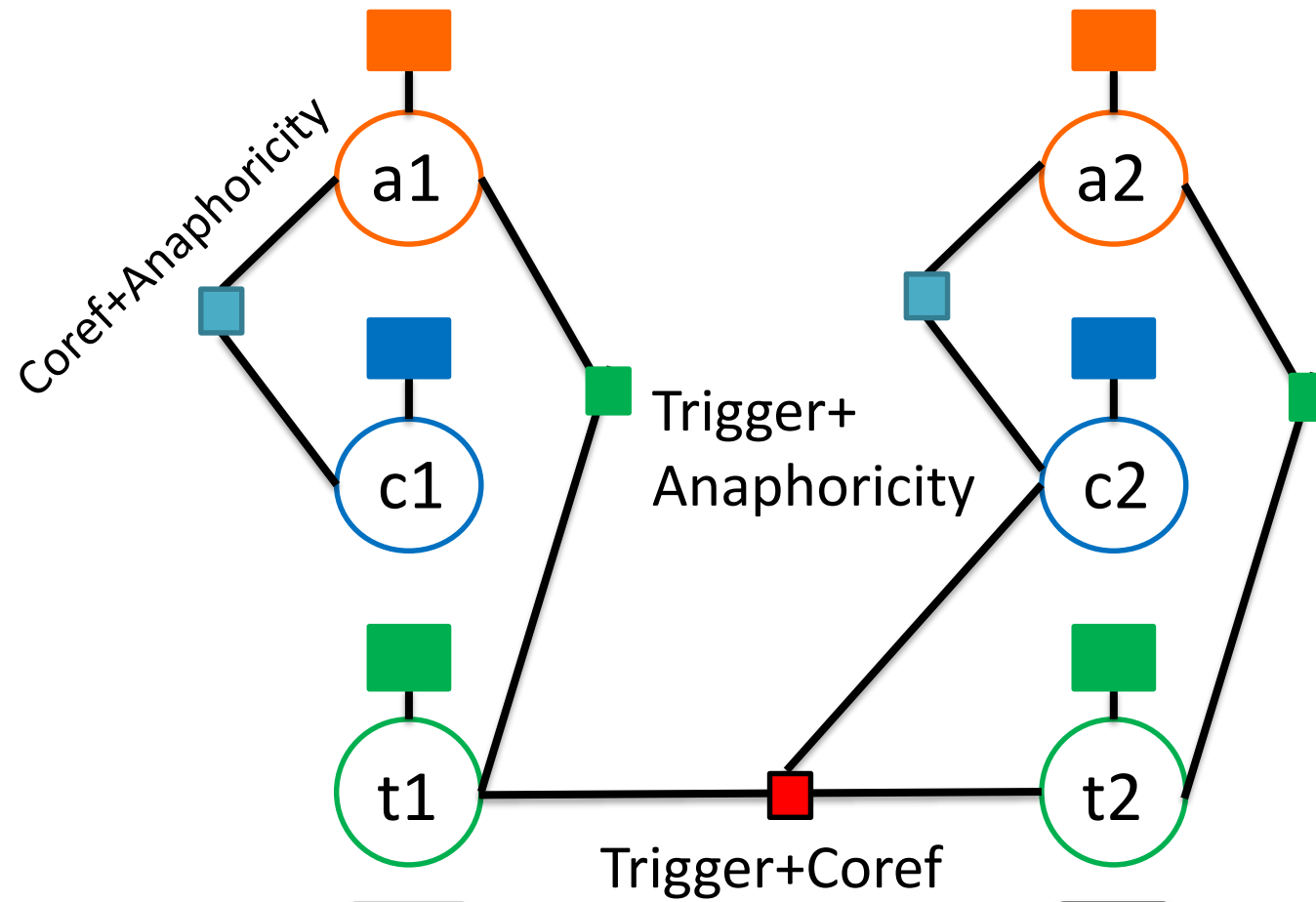
Anaphoricity = Yes, antecedent = Yes



The world has lost...



Joint Model



Mandela has died... The world has lost...



Training

- Log-linear probability distribution over the variables

$$p(t, a, c|x; \Theta) \propto \exp\left(\sum_i \theta_i f_i(t, a, c, x)\right)$$

- Conditional log-likelihood of the training data with L1 regularization

$$L(\Theta) = \sum_{i=1}^d \log \sum_{\text{latent stuff}} p'(t_i^*, a_i^*, c_i^* | x_i; \Theta) + \lambda \|\Theta\|_1$$



Training

- Log-linear probability distribution over the variables

$$p(t, a, c | x; \Theta) \propto \exp\left(\sum_i \theta_i f_i(t, a, c, x)\right)$$

- Conditional log-likelihood of the training data with L1 regularization

$$L(\Theta) = \sum_{i=1}^d \log \sum_{\text{latent stuff}} p'(t_i^*, a_i^*, c_i^* | x_i; \Theta) + \lambda \|\Theta\|_1$$

$$p'(t, a, c | x_i; \theta) \propto p(t, a, c | x_i; \theta) \exp[\alpha_t l_t(t, t^*) + \alpha_a l_a(a, a^*) + \alpha_c l_c(c, C^*)]$$



- $$p(t, a, c|x; \Theta) \propto \exp(\sum_i \theta_i f_i(t, a, c, x))$$

- $$L(\Theta) = \sum_{i=1}^d \log \sum_{\text{latent stuff}} p'(t_i^*, a_i^*, c_i^* | x_i; \Theta) + \lambda \|\Theta\|_1$$

7/31/2017



Inference: Belief Propagation

- computes the marginals for a variable or a set of variables to which a factor connects
- encourages compatibility among component models
- To improve scalability
 - Restrict the domains of the coreference variables
 - Restrict the domains of the subtype variables

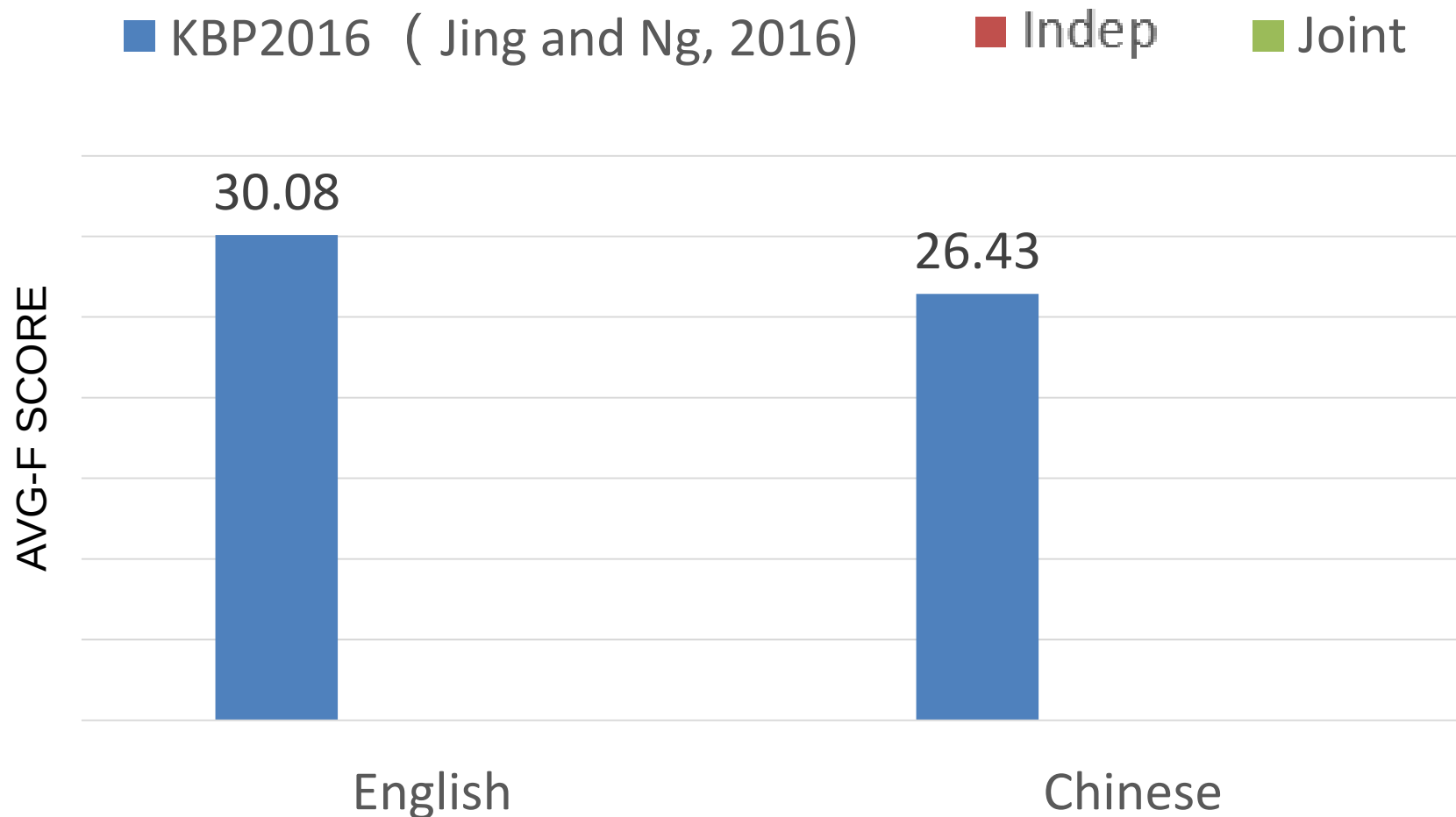


Evaluation

- Dataset:
 - KBP 2016 **English** and **Chinese** corpora
- Evaluation metrics:
 - **Event Coreference**: MUC, B₃, CEAF_e, BLANC , and average of their F-scores
 - **Trigger detection**: F-score
 - **Anaphoricity detection**: F-score

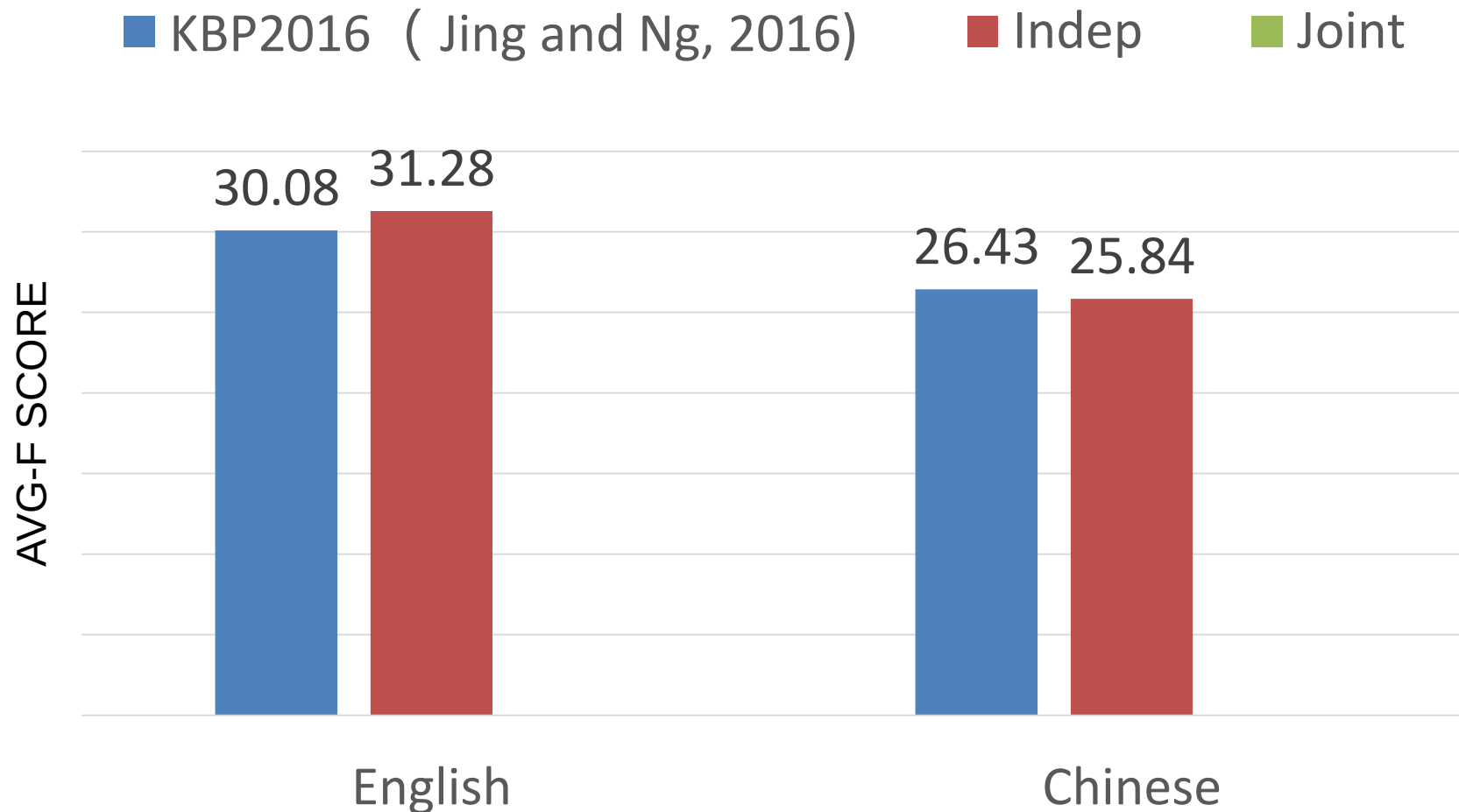


Event Coreference Results



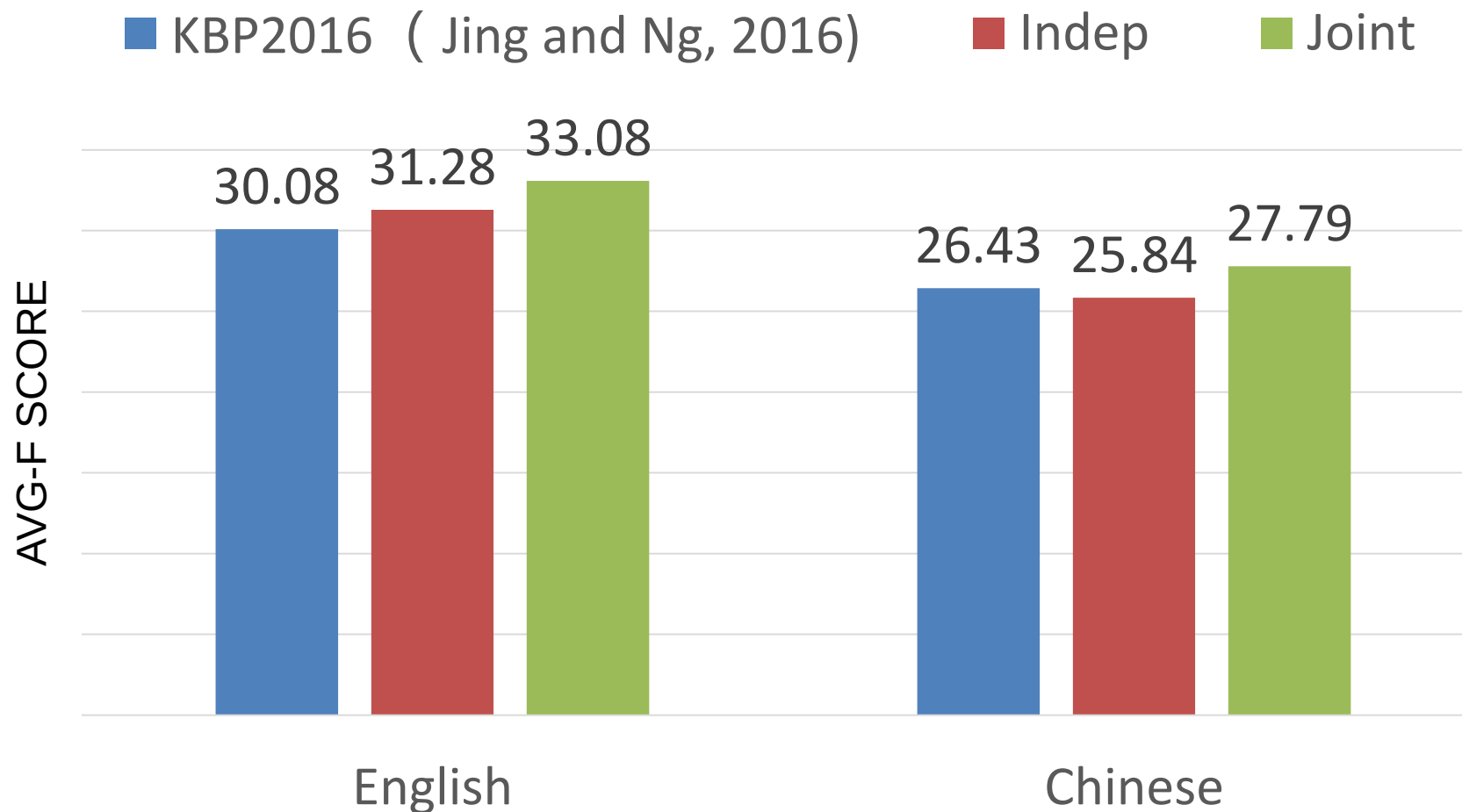


Event Coreference Results



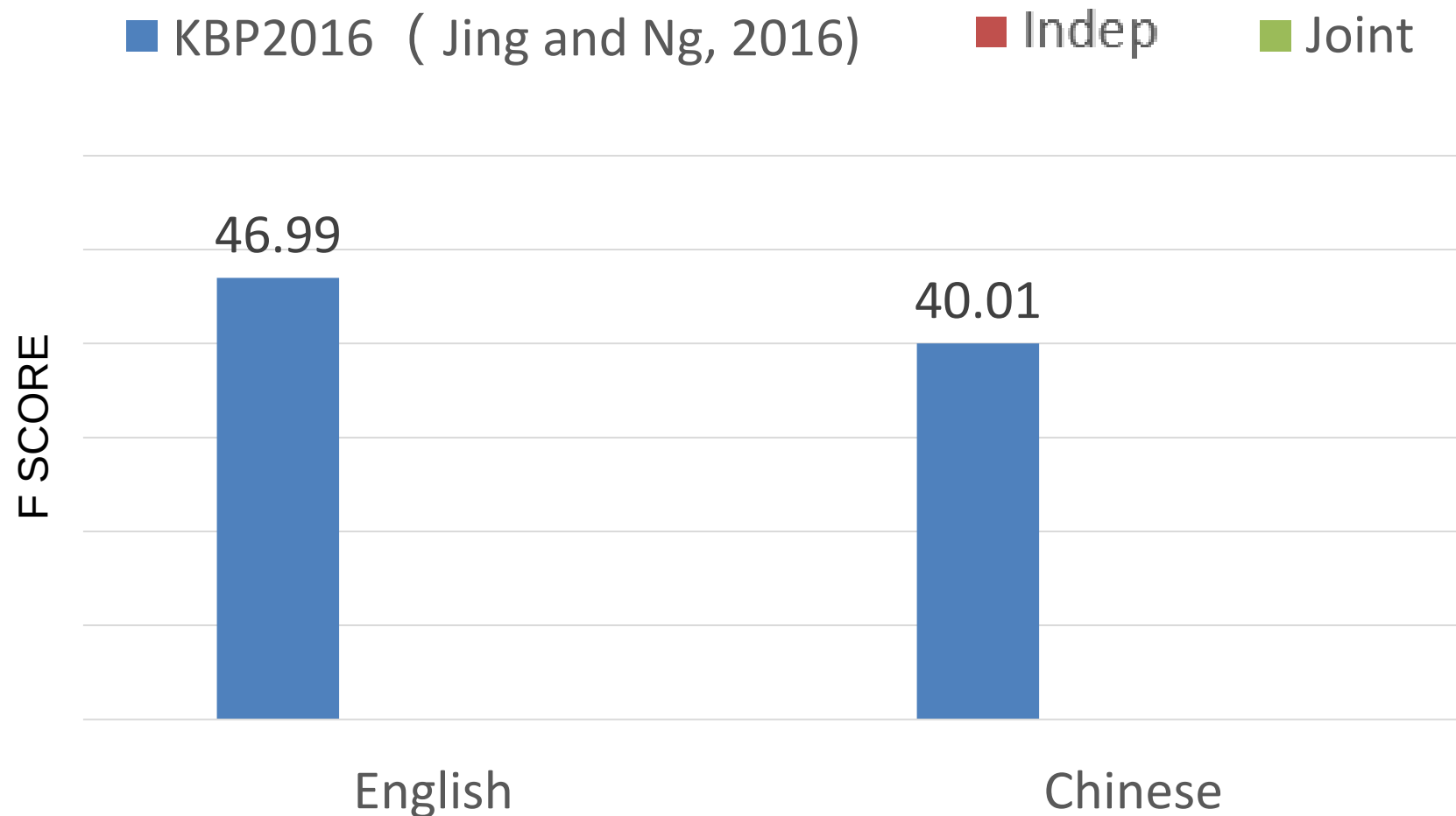


Event Coreference Results



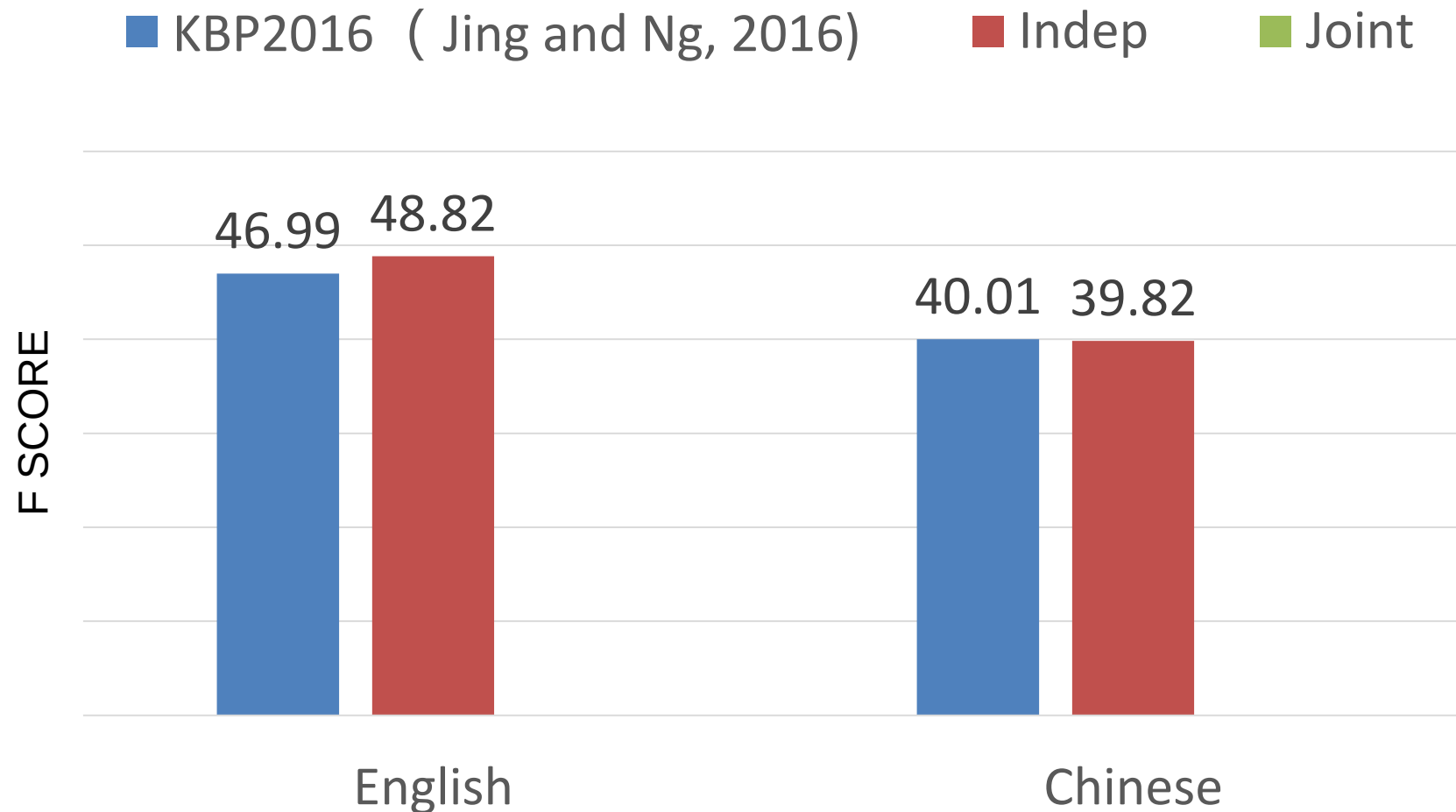


Trigger Detection Results



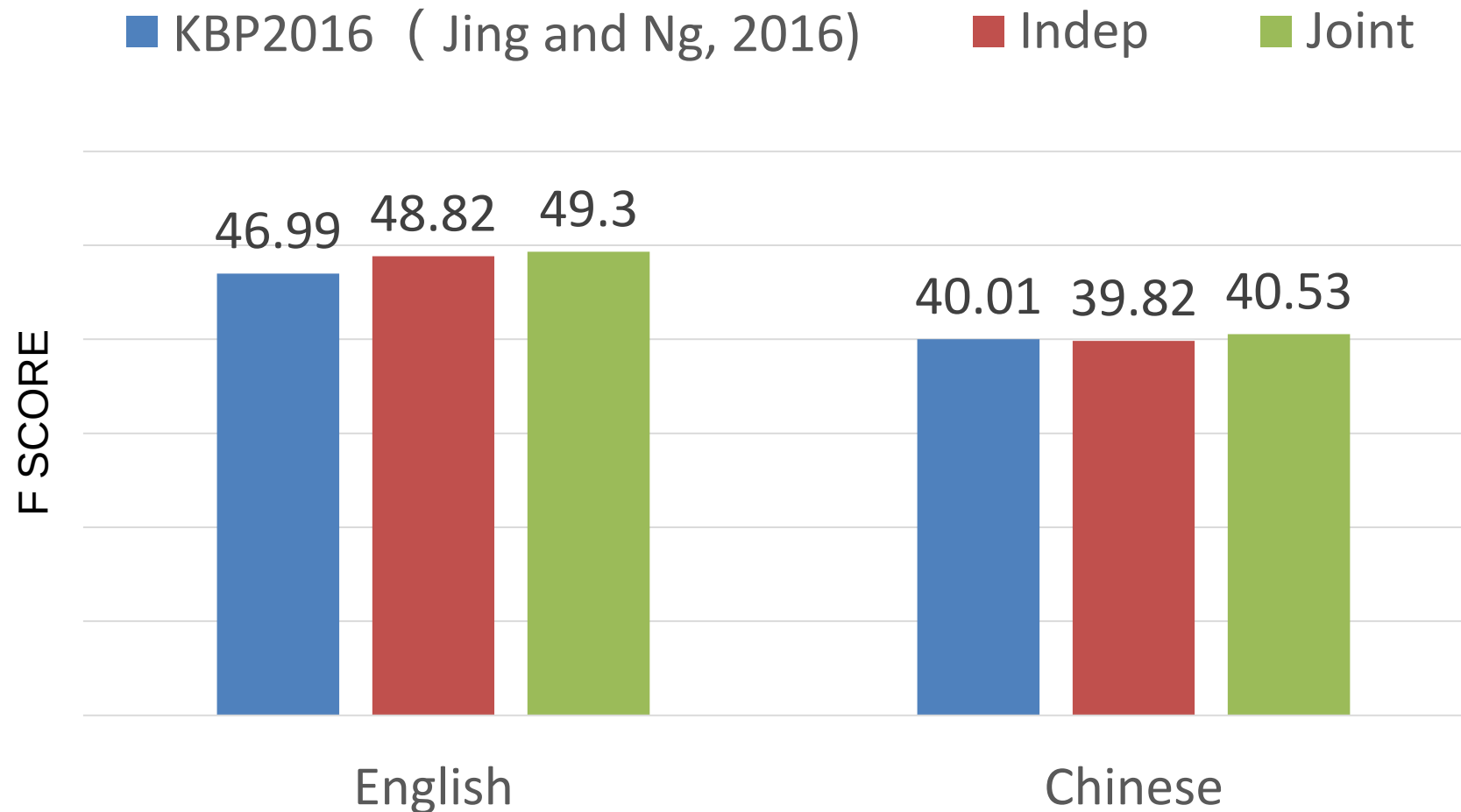


Trigger Detection Results



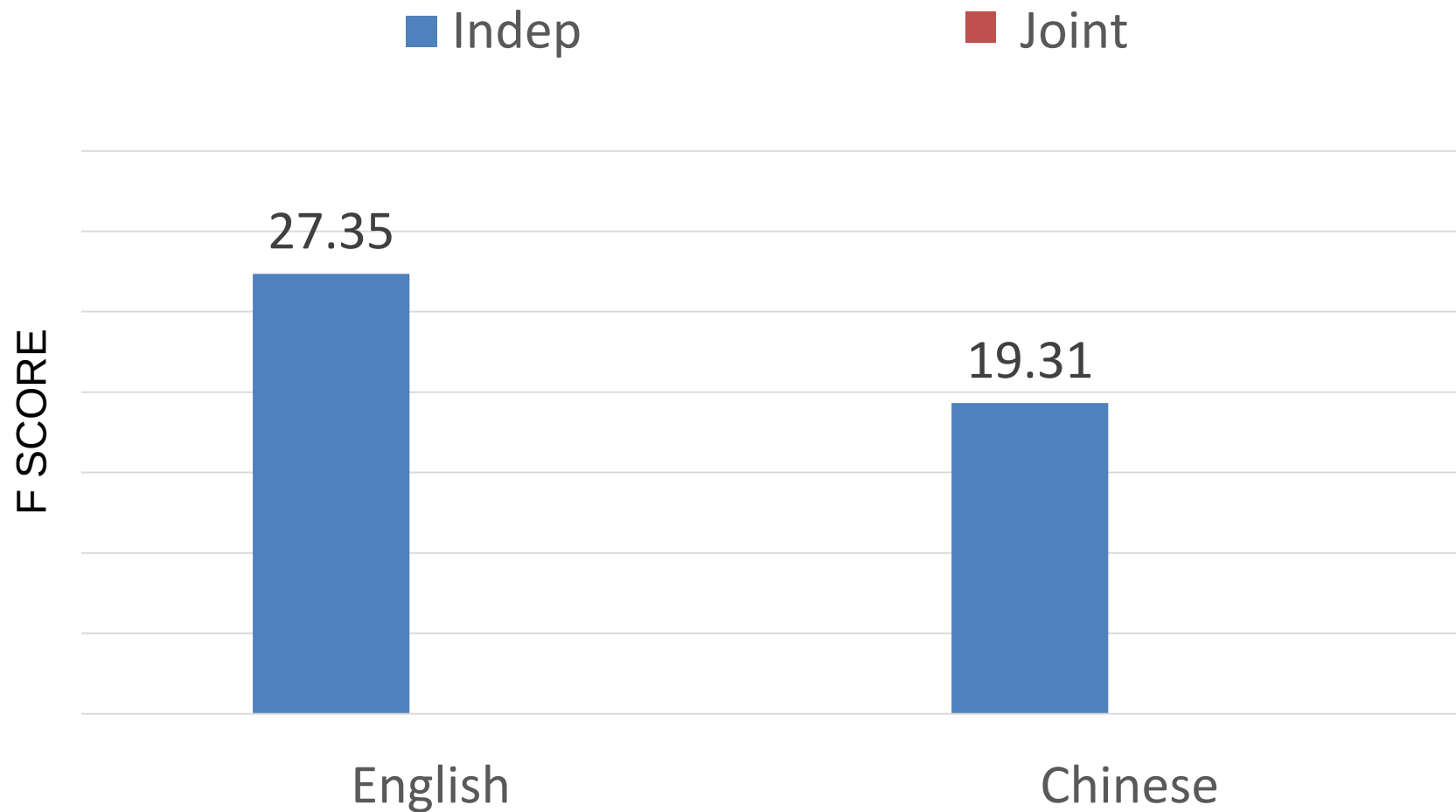


Trigger Detection Results



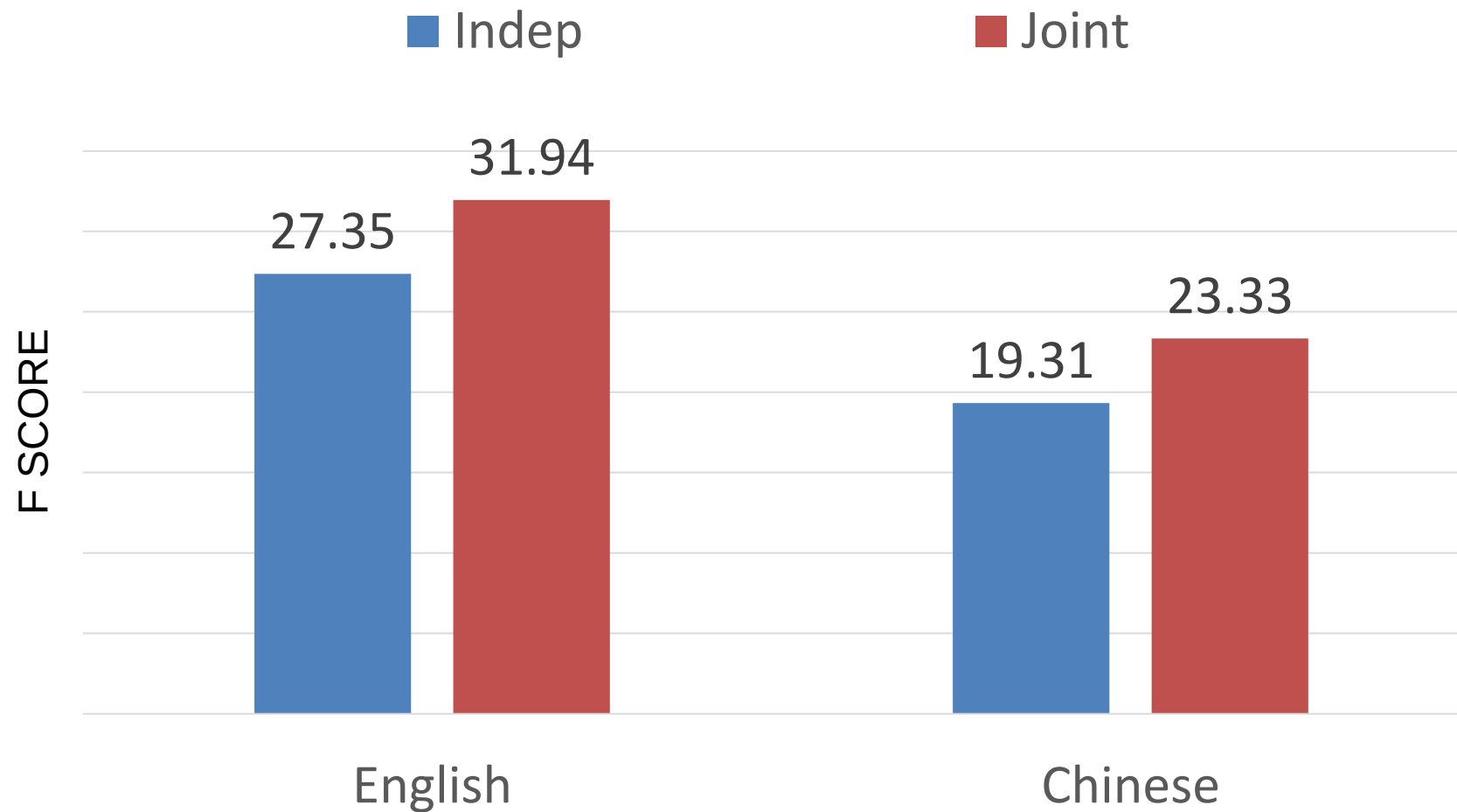


Anaphoricity Detection Results





Anaphoricity Detection Results





Conclusion

- Joint predictions on trigger detection, event coreference resolution, and event anaphoricity
- Joint models effectively improve the performance of event coreference resolution
 - achieves the best results to date on KBP 2016 English and Chinese corpora



Model Ablation

- evaluate the importance of each of the three types of joint factors
 - add each type of joint factors to the independent model
 - remove each type of joint factors from the full joint model



Model Ablation: Coref-Trigger interactions

	English			Chinese		
	Coref	Trigger	Anaph	Coref	Trigger	Anaph
INDEP.	31.28	48.82	27.35	25.84	39.82	19.31
INDEP.+CorefTrigger	+0.39	+0.42	−0.05	+0.95	+0.56	−0.37
INDEP.+CorefAnaph	+0.40	−0.08	+3.45	+0.37	+0.04	−0.11
INDEP.+TriggerAnaph	+0.11	+0.38	+1.35	+0.14	+0.52	+0.02
JOINT−CorefTrigger	+0.56	−0.06	+4.41	+0.19	+0.35	+3.34
JOINT−CorefAnaph	+0.63	+0.66	+1.46	+1.50	+0.88	+0.28
JOINT−TriggerAnaph	+1.89	+0.50	+4.01	+1.65	+0.50	+1.79
JOINT	+1.80	+0.48	+4.59	+1.95	+0.71	+4.02

Contributes the most to coreference performance, and trigger detection



Model Ablation: Coref-Anaphoricity interactions

	English			Chinese		
	Coref	Trigger	Anaph	Coref	Trigger	Anaph
INDEP.	31.28	48.82	27.35	25.84	39.82	19.31
INDEP.+CorefTrigger	+0.39	+0.42	-0.05	+0.95	+0.56	-0.37
INDEP.+CorefAnaph	+0.40	-0.08	+3.45	+0.37	+0.04	-0.11
INDEP.+TriggerAnaph	+0.11	+0.38	+1.35	+0.14	+0.52	+0.02
JOINT-CorefTrigger	+0.56	-0.06	+4.41	+0.19	+0.35	+3.34
JOINT-CorefAnaph	+0.63	+0.66	+1.46	+1.50	+0.88	+0.28
JOINT-TriggerAnaph	+1.89	+0.50	+4.01	+1.65	+0.50	+1.79
JOINT	+1.80	+0.48	+4.59	+1.95	+0.71	+4.02

In isolation: improve coreference performance

In combination: improves both tasks, particularly anaphoricity determination



Model Ablation: Trigger-Anaphoricity interactions

	English			Chinese		
	Coref	Trigger	Anaph	Coref	Trigger	Anaph
INDEP.	31.28	48.82	27.35	25.84	39.82	19.31
INDEP.+CorefTrigger	+0.39	+0.42	-0.05	+0.95	+0.56	-0.37
INDEP.+CorefAnaph	+0.40	-0.08	+3.45	+0.37	+0.04	-0.11
INDEP.+TriggerAnaph	+0.11	+0.38	+1.35	+0.14	+0.52	+0.02
JOINT-CorefTrigger	+0.56	-0.06	+4.41	+0.19	+0.35	+3.34
JOINT-CorefAnaph	+0.63	+0.66	+1.46	+1.50	+0.88	+0.28
JOINT-TriggerAnaph	+1.89	+0.50	+4.01	+1.65	+0.50	+1.79
JOINT	+1.80	+0.48	+4.59	+1.95	+0.71	+4.02

In isolation: improve both tasks

In combination: improves anaphoricity determination



Error Analysis

- Precision Error
 - Erroneous and mistyped triggers
 - assign the same subtype to event mentions triggered by the same word.
 - assigns the wrong subtype to triggers that possess different subtypes in different contexts
 - » Example: trigger "leave" sometimes has the subtype Movement.Transport-Person and sometimes has the subtype Personnel.End-Position.
 - words that are only sometimes used as triggers are often wrongly posited as triggers when they are not
 - Failure to extract arguments
 - we only exploit dependency relations for argument extraction
 - Sometimes arguments that do not appear in the same sentence as their trigger.



Error Analysis

- Recall Error
 - Missing triggers
 - fails to identify trigger words that are unseen or rarely occurring in the training data.
 - Lack of entity coreference information.
 - the corresponding arguments of two coreferent event mentions are typically coreferent
 - This issue is particularly serious in discussion forum documents,
 - pronouns often serve as subjects and objects of event mentions
 - More serious in Chinses documents due to the zero pronouns
 - Lack of contextual understanding
 - Additional contextual information present in neighboring sentences is not used
 - when describing the fact that Tim Cook will attend Apple's Istanbul store opening
 - "Cook is expected to **return** to Turkey for the store opening"
 - "Tim **travels** abroad YET AGAIN to be feted by the not-so-high-and-mighty".