A Joint Model for Entity Analysis: Coreference, Typing, and Linking

Greg Durrett and Dan Klein

UC Berkeley

The Entity Stack

Dell reported earnings at a press conference in Austin.

The Entity Stack

ORG EVENT LOC

Dell reported earnings at a press conference in Austin.

The Entity Stack

ORG EVENT LOC

Dell reported earnings at a press conference in Austin.

NER

The Entity Stack

ORG EVENT LOC

Dell reported earnings at a press conference in Austin.

The company had stronger than expected growth.

NER

The Entity Stack

ORG EVENT LOC

Dell reported earnings at a press conference in Austin.

The company had stronger than expected growth.

NER

The Entity Stack

ORG EVENT LOC

Dell reported earnings at a press conference in Austin.

The company had stronger than expected growth.

Coref

NER

The Entity Stack

ORG EVENT LOC

Dell reported earnings at a press conference in Austin.

The company had stronger than expected growth.

Coref

NER

The Entity Stack

ORG EVENT LOC

Dell reported earnings at a press conference in Austin.

The company had stronger than expected growth.

Link

Coref

NER

The Entity Stack

Dell reported earnings …

Link

Coref

NER

The Entity Stack

Dell reported earnings …

Link

Coref

NER

The Entity Stack

ORG

Dell reported earnings …

Link

Coref

NER

The Entity Stack

PER ORG

Dell reported earnings …

Link

Coref

NER

The Entity Stack

PER ORG

Dell reported earnings …

The company had …

Link

Coref

NER

The Entity Stack

PER ORG

Dell reported earnings …

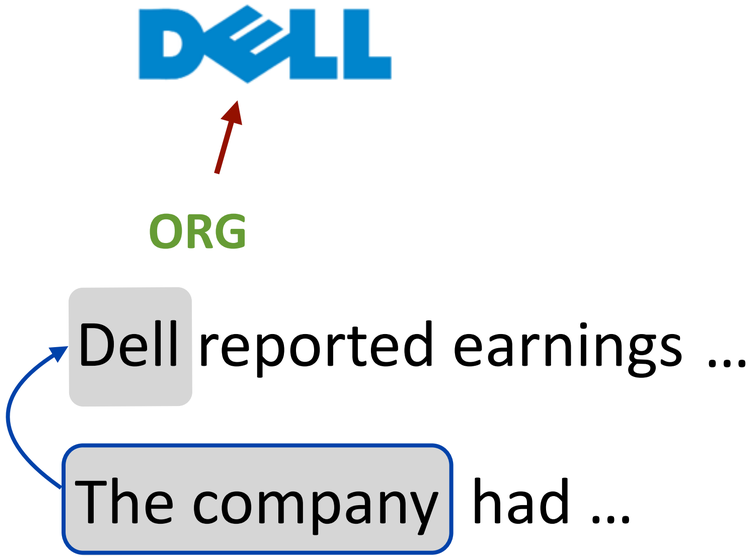
The company had …

Link

Coref

NER

The Entity Stack

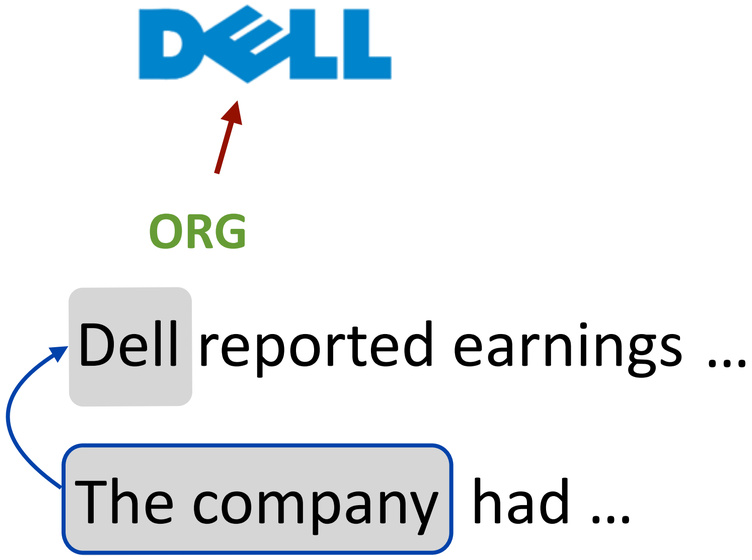


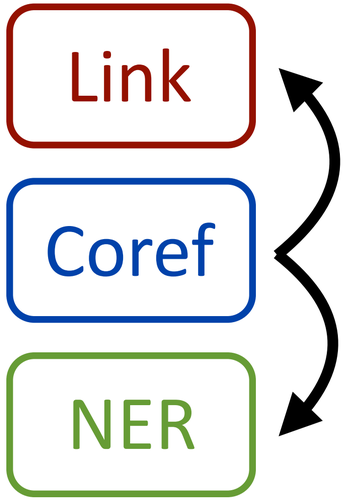
Link

Coref

NER

The Entity Stack

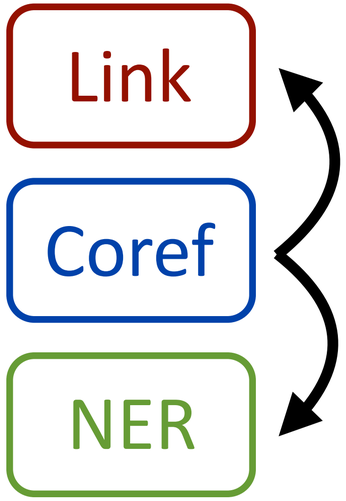




The Entity Stack

Dell reported earnings …

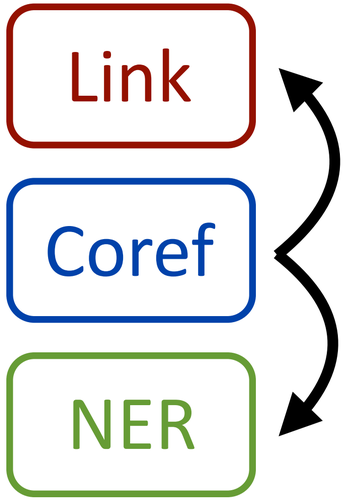
The company had …



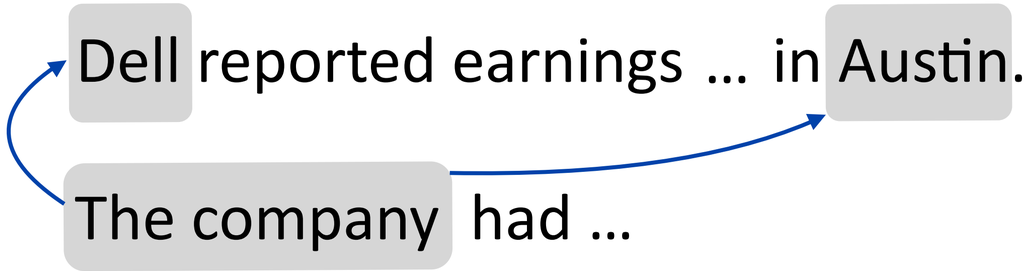
The Entity Stack

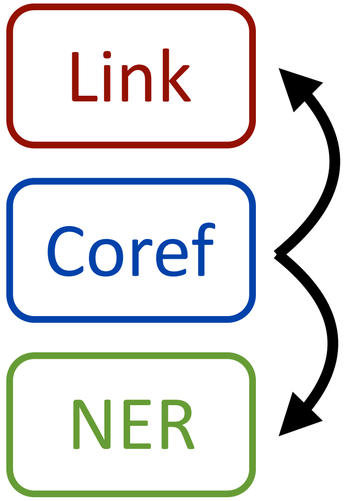
Dell reported earnings … in Austin.

The company had …

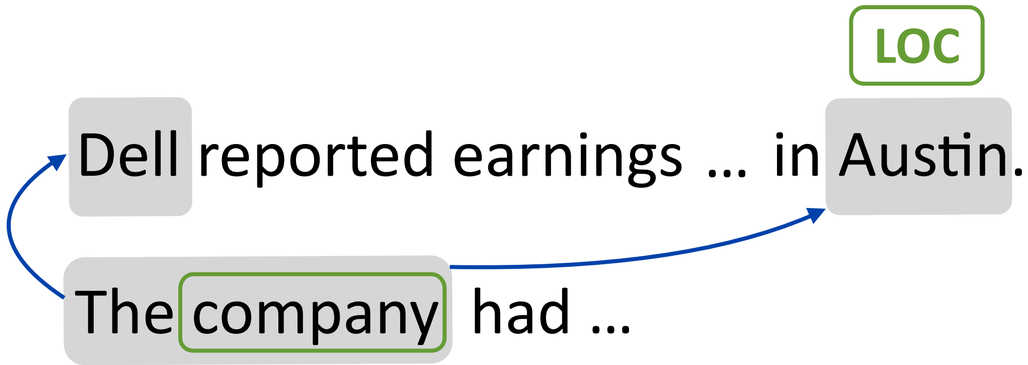


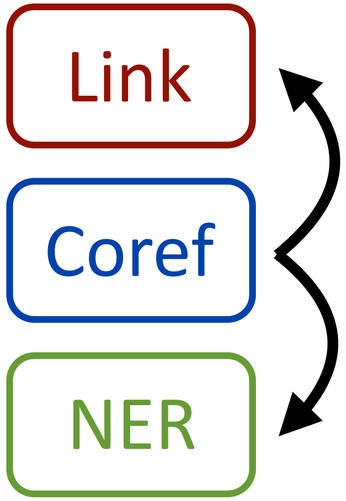
The Entity Stack



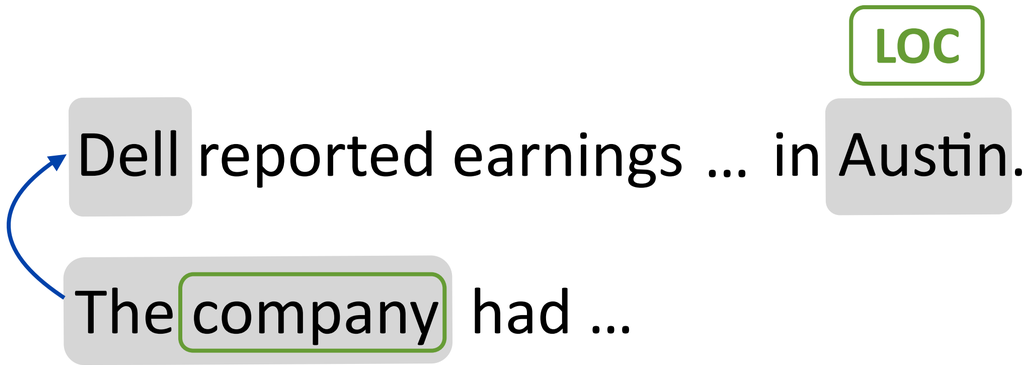


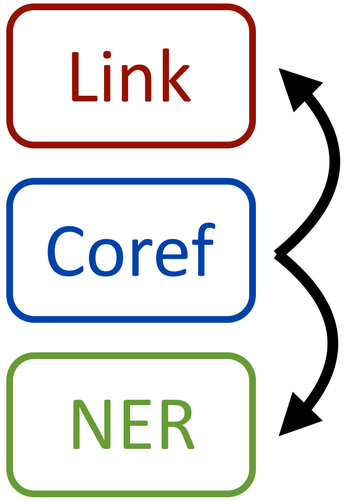
The Entity Stack



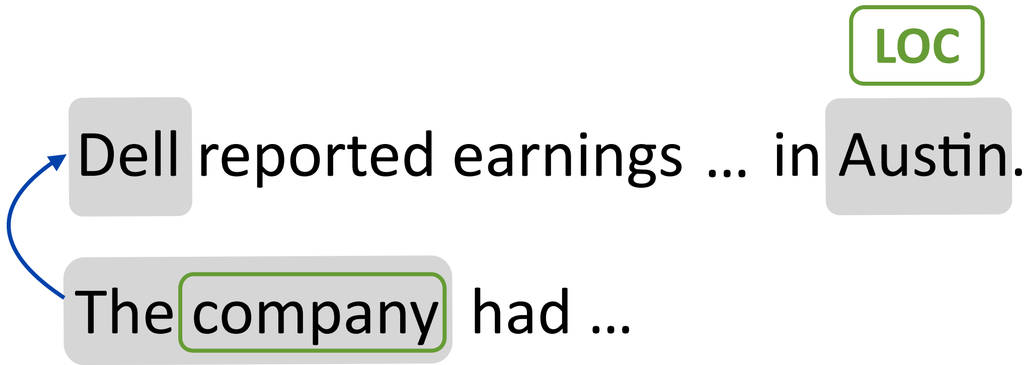


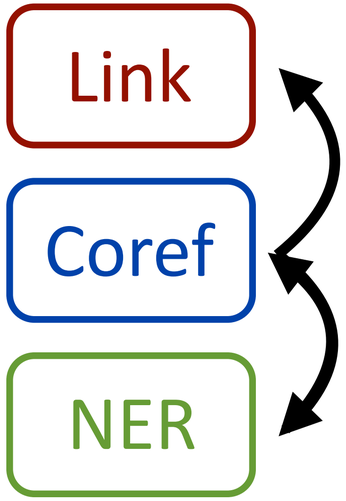
The Entity Stack



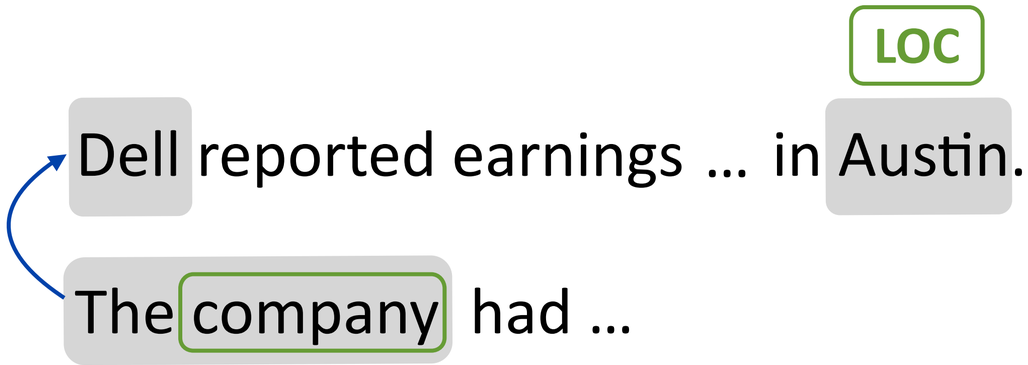


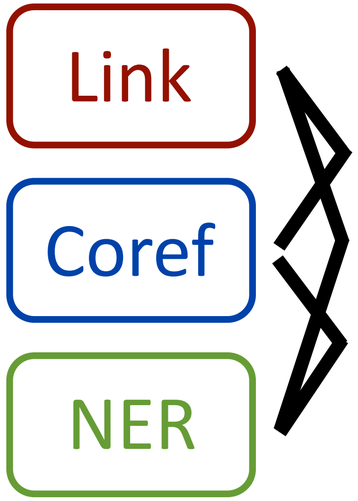
The Entity Stack





The Entity Stack



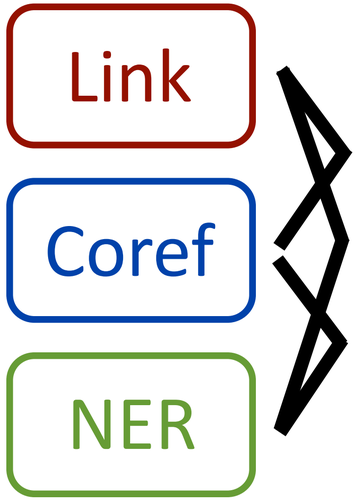


The Entity Stack

LOC

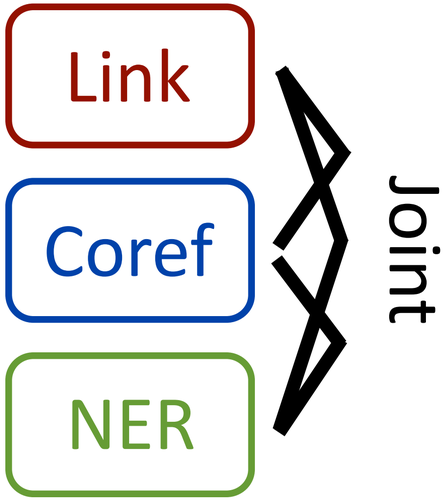
Dell reported earnings … in Austin.

The company had …



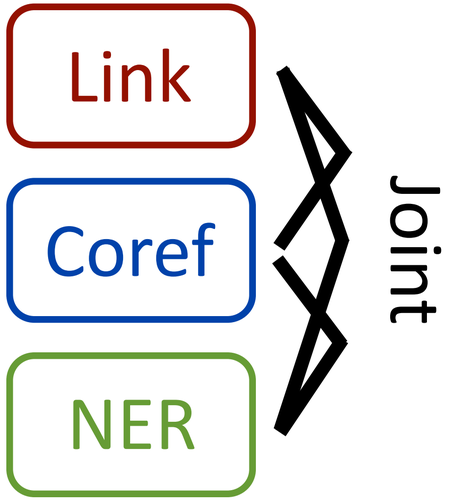
Model and predict the three tasks jointly

Outline



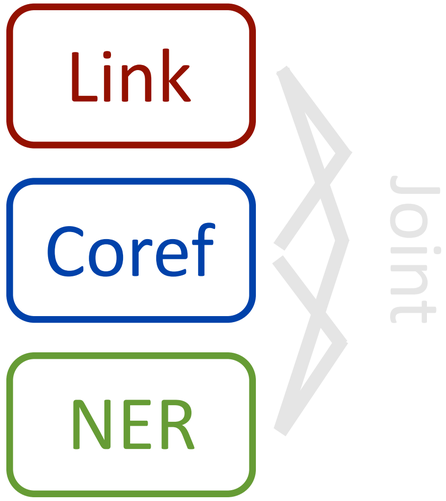
Outline

- Model



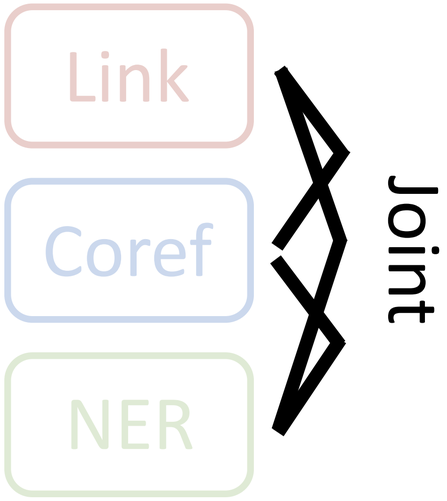
Outline

- Model
  - Independent models (unary, per-task factors)



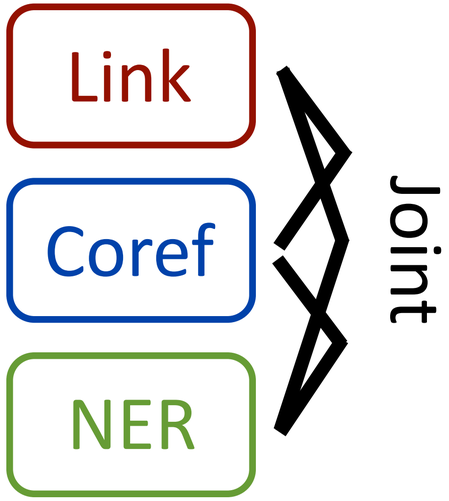
Outline

- Model
  - Independent models (unary, per-task factors)
  - Joint coupling (higher-arity, cross-task factors)



Outline

- Model
  - Independent models (unary, per-task factors)
  - Joint coupling (higher-arity, cross-task factors)
- Experiments



Model

Dell

Austin

The company

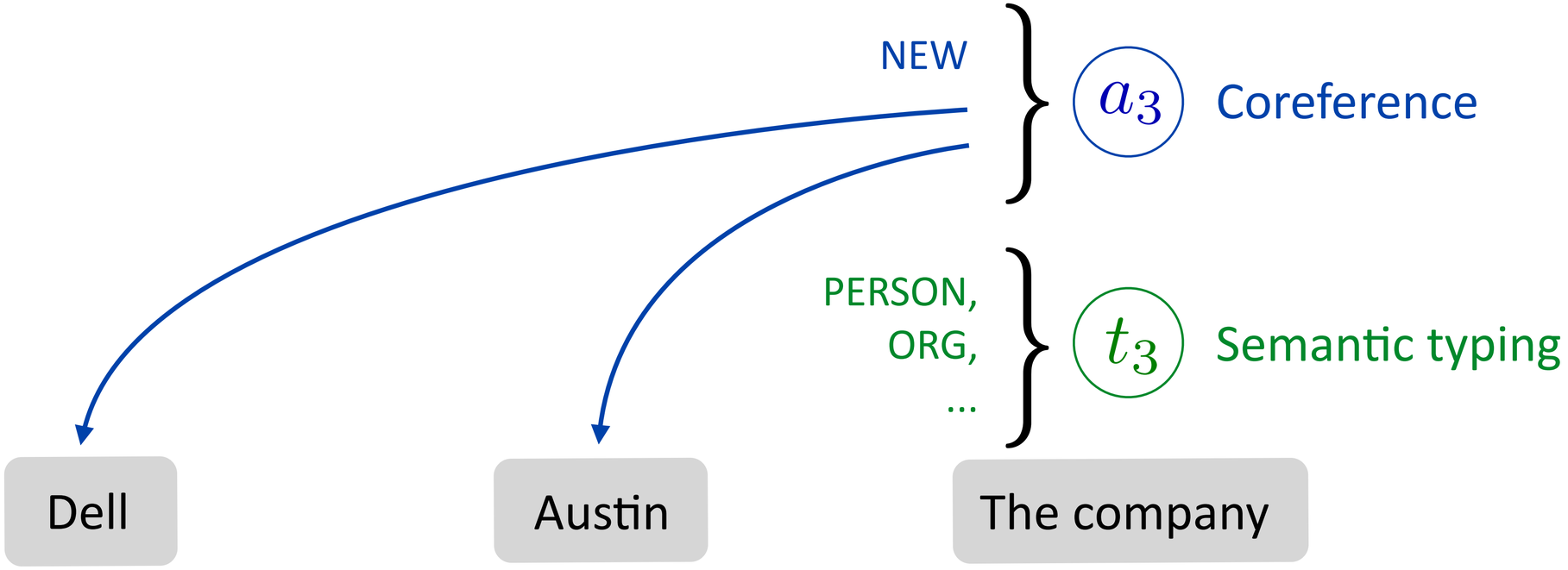
Model

Dell

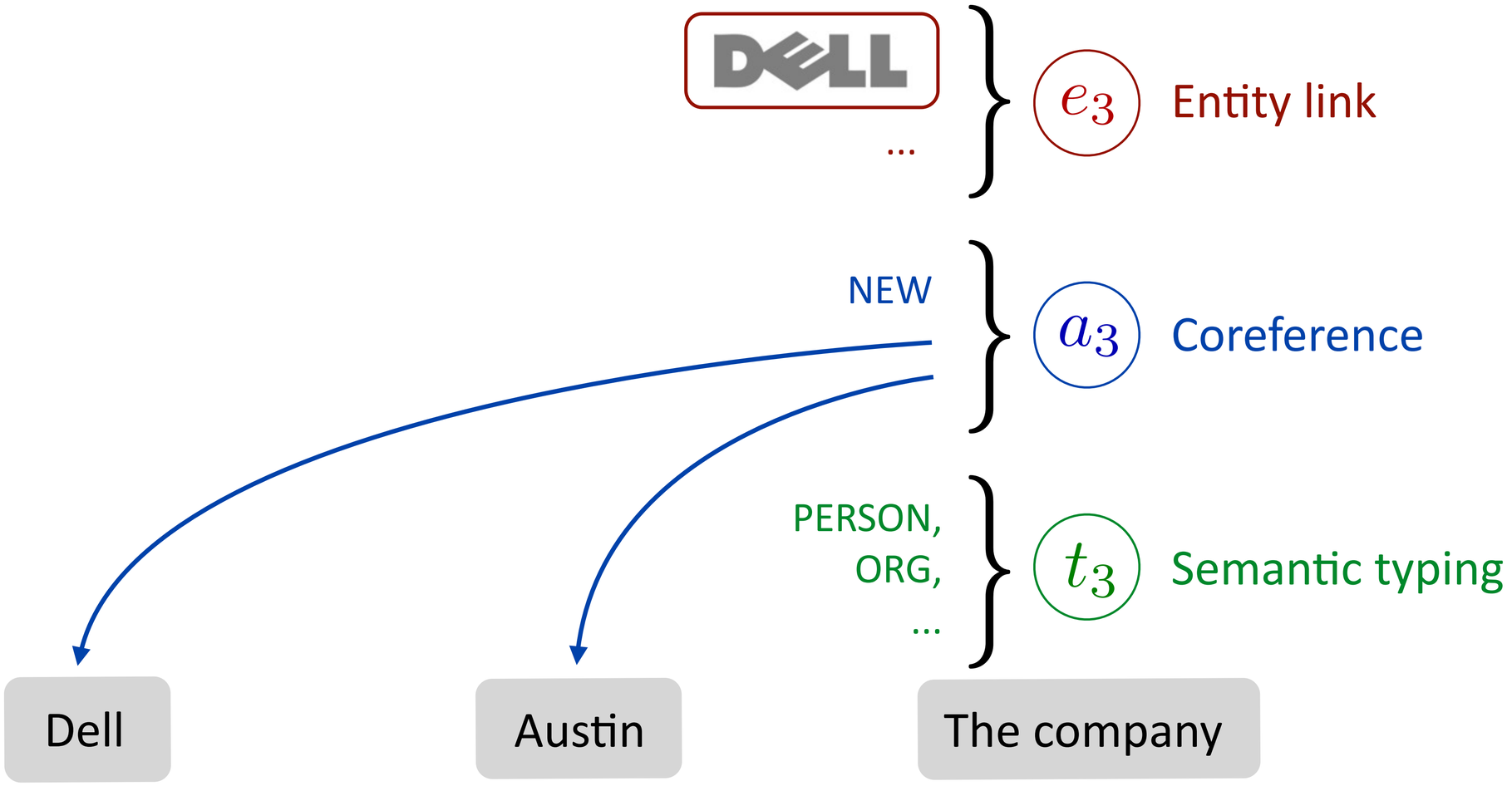
Austin

The company

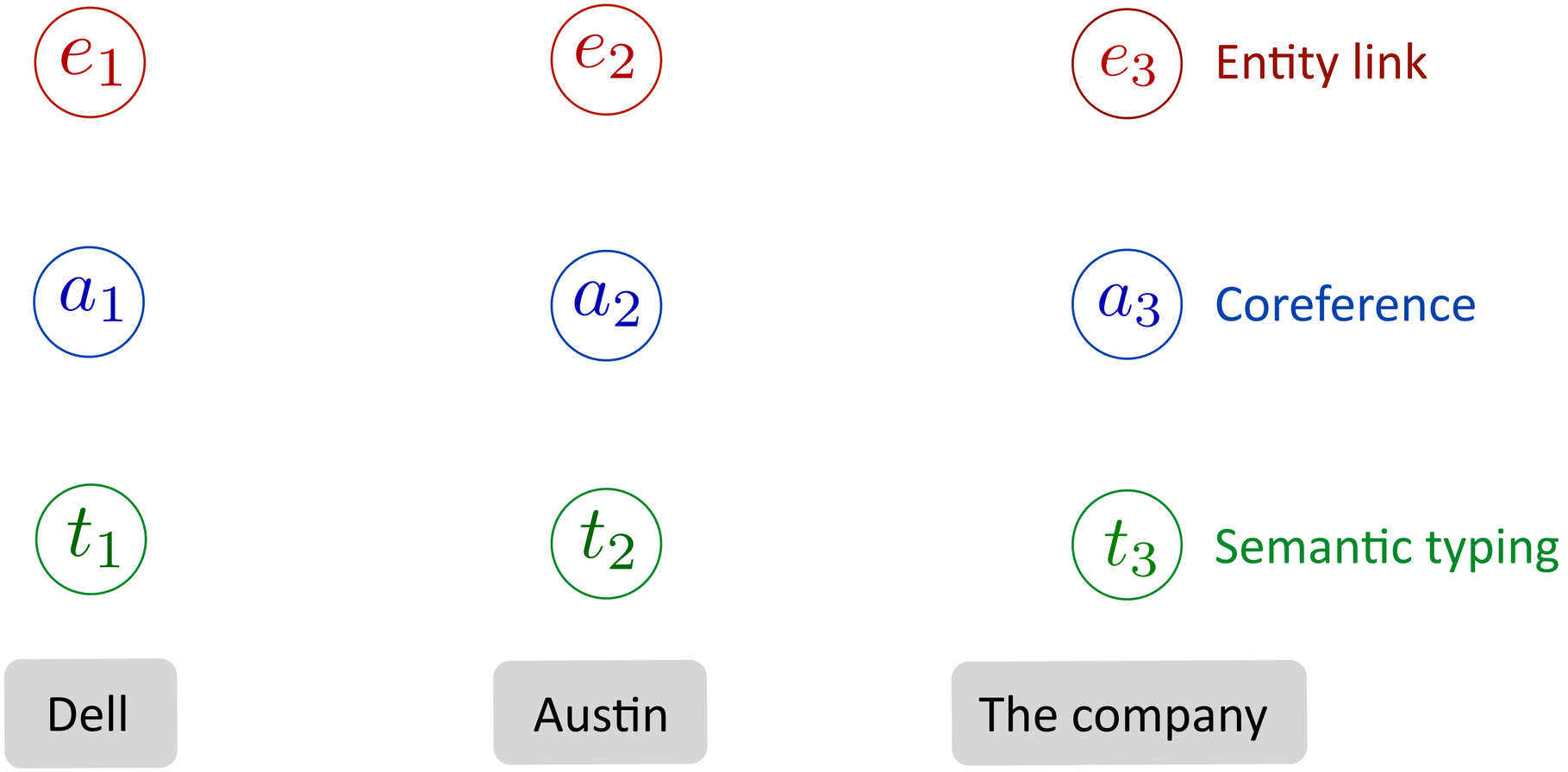
Model


NEW
a_3
Coreference
PERSON,
ORG,
...
t_3
Semantic typing
Dell
Austin
The company

Model


...
e_3
Entity link
NEW
a_3
Coreference
PERSON,
ORG,
...
t_3
Semantic typing
Dell
Austin
The company

Model


e_1
e_2
e_3
Entity link
a_1
a_2
a_3
Coreference
t_1
t_2
t_3
Semantic typing
Dell
Austin
The company

Model

e_1

a_1

t_1

Dell

Model

$$\mathbf{p}(\mathbf{t}, \mathbf{a}, \mathbf{e} | \mathbf{x}) \propto \exp\left(\mathbf{w}^\top \mathbf{f}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e})\right)$$

e_1

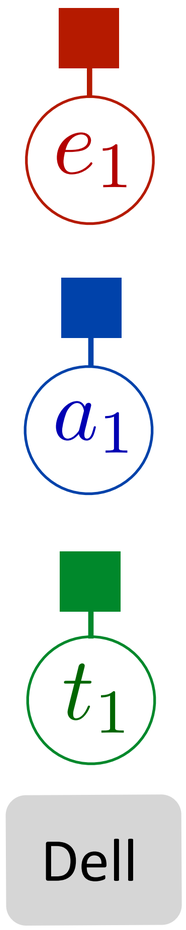
a_1

t_1

Dell

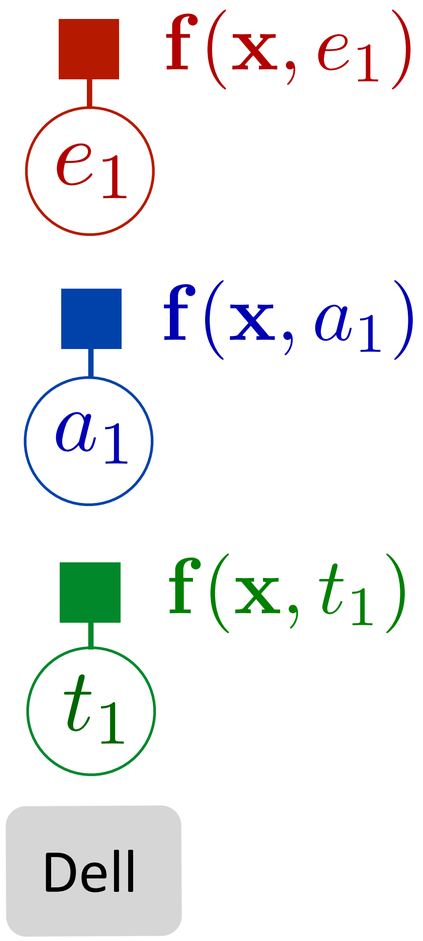
Model

$$\mathbf{p}(\mathbf{t}, \mathbf{a}, \mathbf{e} | \mathbf{x}) \propto \exp\left(\mathbf{w}^\top \mathbf{f}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e})\right)$$



Model

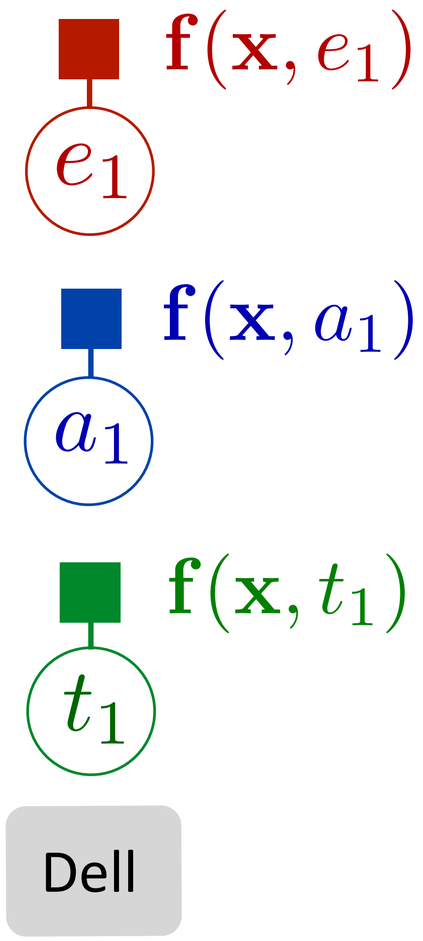
$$\mathbf{p}(\mathbf{t}, \mathbf{a}, \mathbf{e} | \mathbf{x}) \propto \exp\left(\mathbf{w}^\top \mathbf{f}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e})\right)$$



Model

$$\mathbf{p}(\mathbf{t}, \mathbf{a}, \mathbf{e} | \mathbf{x}) \propto \exp\left(\mathbf{w}^\top \mathbf{f}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e})\right)$$

$$\mathbf{f}_{\text{indep}}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e}) = \sum_{i=1}^{n} \left[\mathbf{f}(\mathbf{x}, t_i) + \mathbf{f}(\mathbf{x}, a_i) + \mathbf{f}(\mathbf{x}, e_i)\right]$$



Semantic Typing (NER)

t_1 { **PERSON**
ORG
…

<s> Dell reported earnings at a press conference …

Semantic Typing (NER)

t_1 { **PERSON**
ORG
…

<s> Dell reported earnings at a press conference …

Semantic Typing (NER)

t_1 { **PERSON**
ORG
…

POS tags

<s>	Dell	reported	earnings	at a press conference …
	NNP	VBZ	NNP	

Semantic Typing (NER)

t_1 { **PERSON**, **ORG**, **…** }

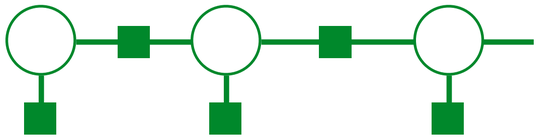
	<s> Dell	reported	earnings	at a press conference …
POS tags	NNP	VBZ	NNP	
Word shapes	X_	_ed	_s	

Semantic Typing (NER)

t_1 { **PERSON**, **ORG**, **…** }

	<s> Dell	reported	earnings	at a press conference …
POS tags	NNP	VBZ	NNP	
Word shapes	X_	_ed	_s	
Brown clusters	0001	1010	0011	

Semantic Typing (NER)



	<s>	Dell	reported	earnings
POS tags		NNP	VBZ	NNP
Word shapes		X_	_ed	_s
Brown clusters		0001	1010	0011

at a press conference …

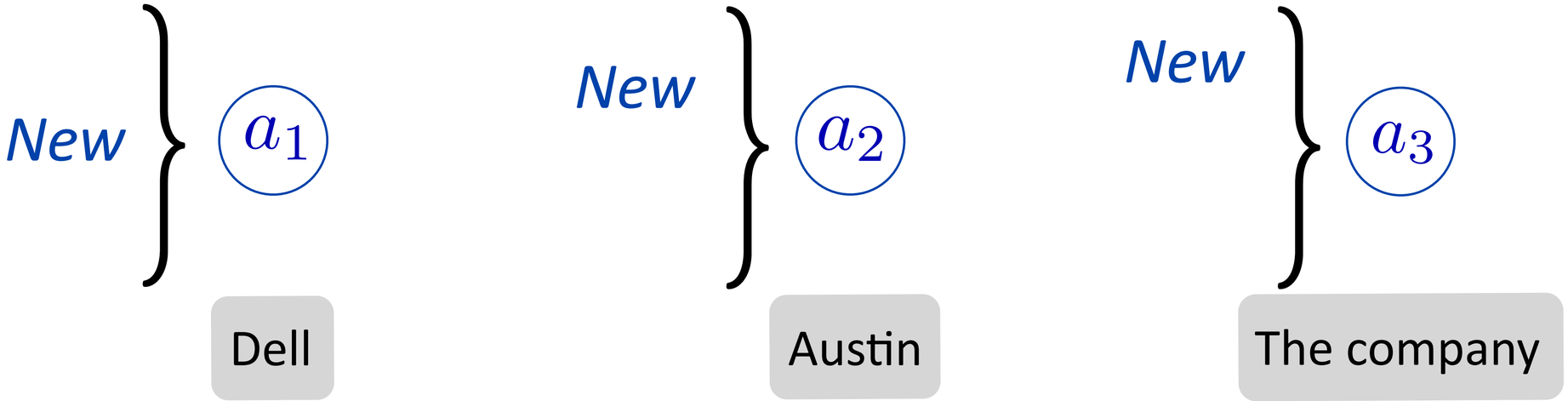
Coreference Resolution

a_1 a_2 a_3

Dell Austin The company

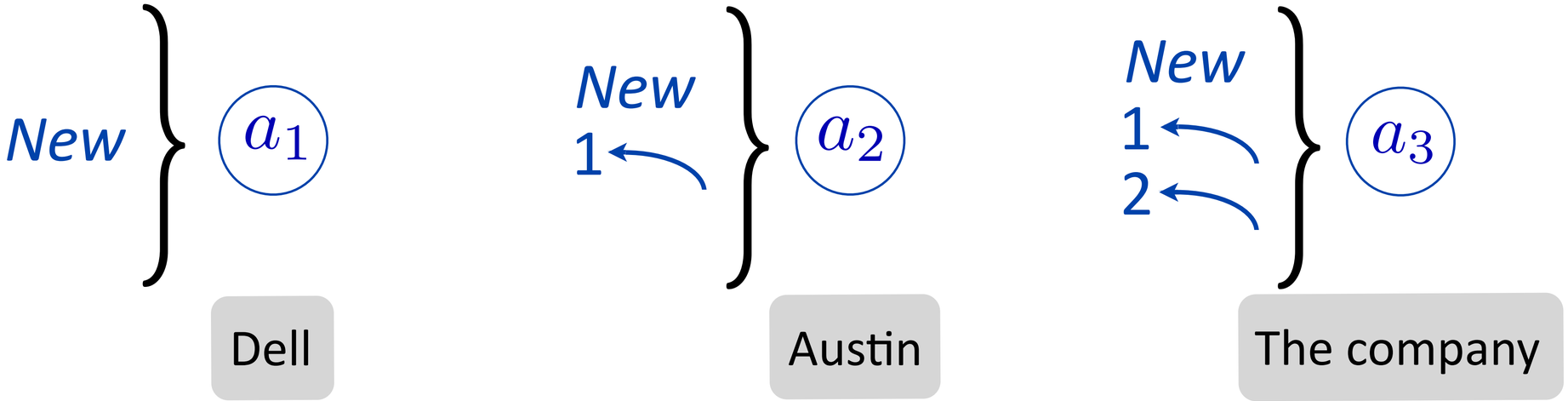
[Durrett and Klein (2013)]

Coreference Resolution



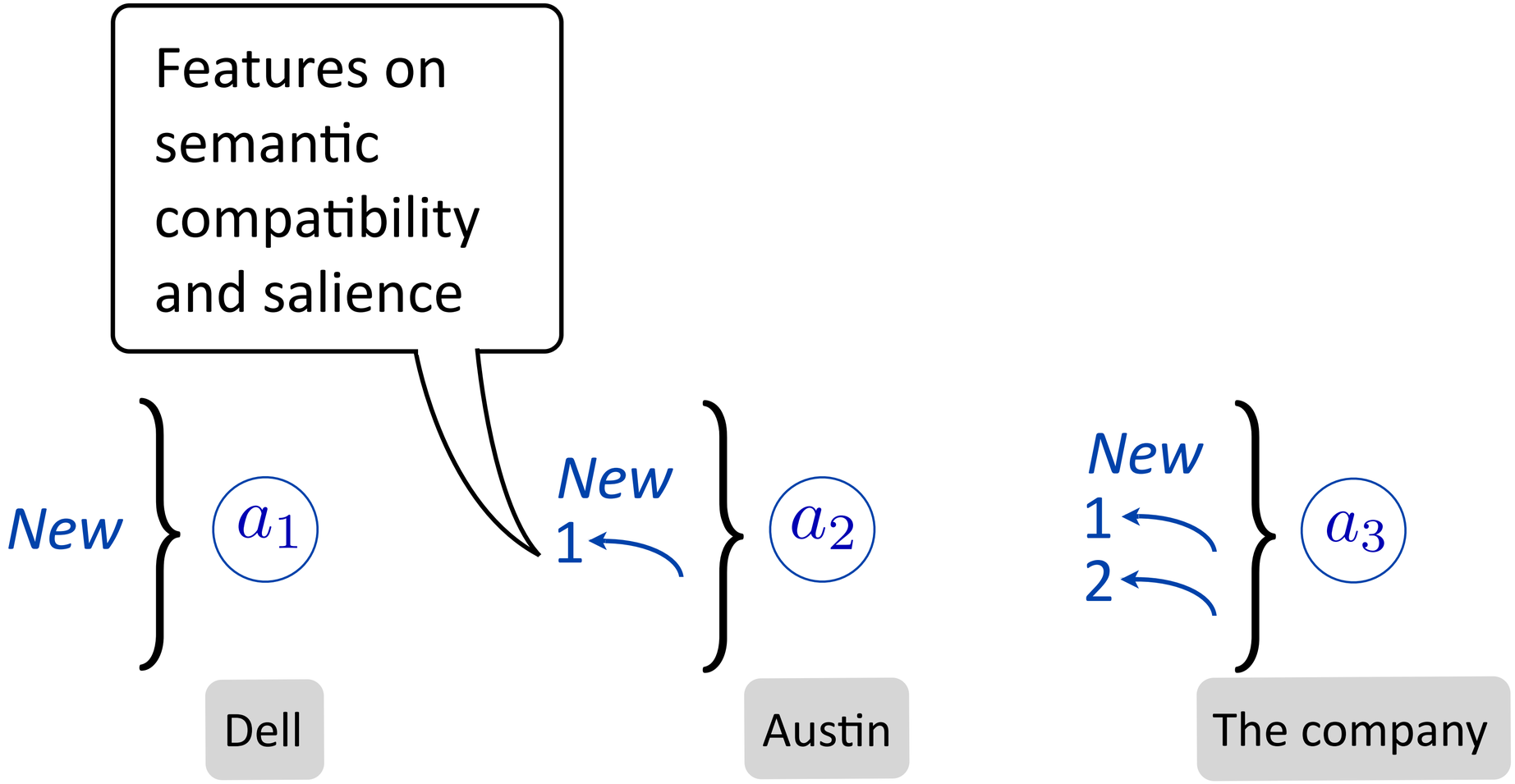
[Durrett and Klein (2013)]

Coreference Resolution



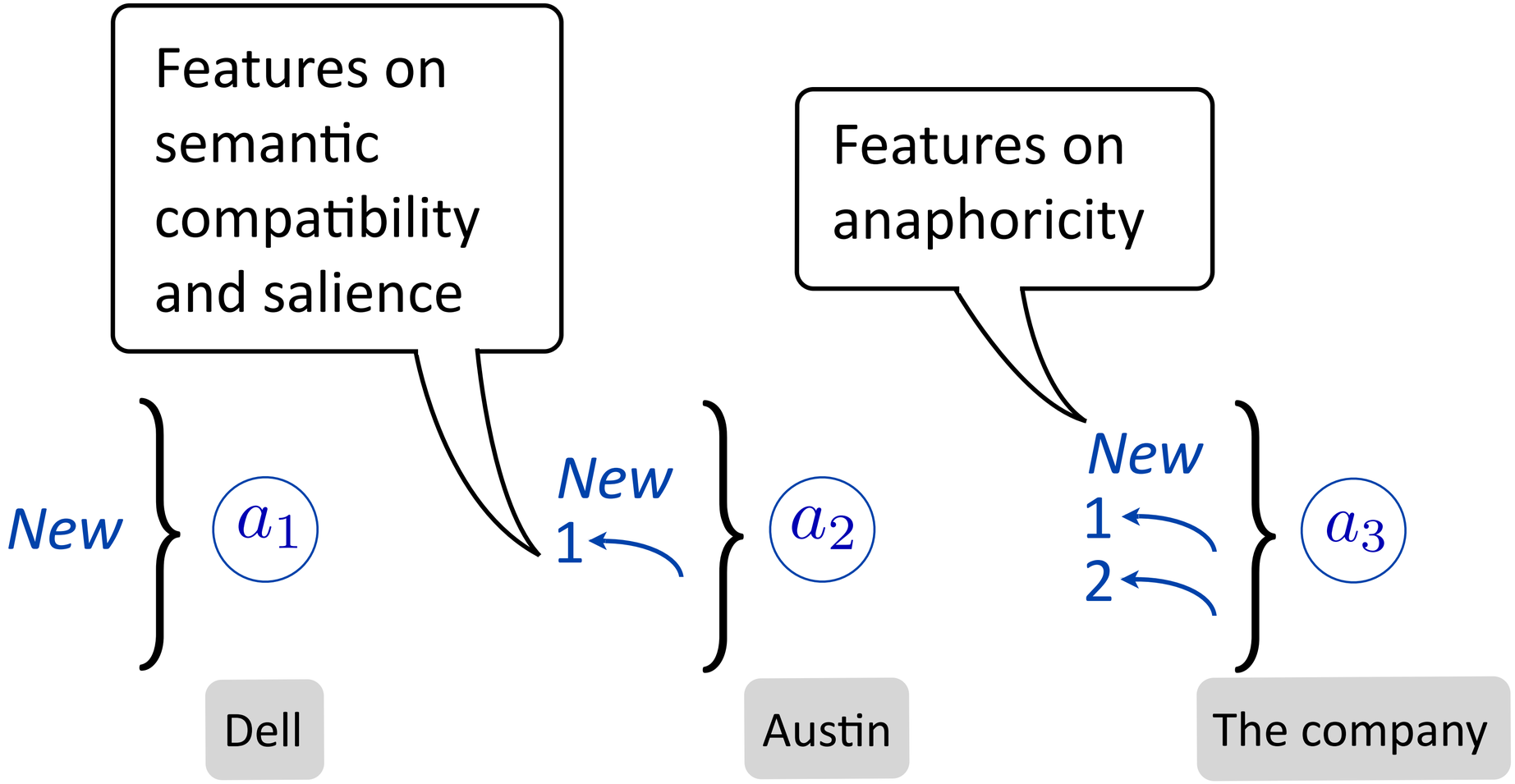
[Durrett and Klein (2013)]

Coreference Resolution



[Durrett and Klein (2013)]

Coreference Resolution



[Durrett and Klein (2013)]

Entity Linking

e_1

Dell

Entity Linking

e_1

Michael Dell

Dell, Inc.

Dell Comics

...

Dell

Entity Linking

WIKIPEDIA
The Free Encyclopedia

>4 million articles

e_1

Michael Dell
Dell, Inc.
Dell Comics
...

Dell

Aardvark
Aaron Burr
...
Zebra
Zimbabwe

Hyperlink Anchor Text

[Cucerzan 2007; Milne and Witten, 2008; Ratinov et al., 2011]

Hyperlink Anchor Text

Dell

[Cucerzan 2007; Milne and Witten, 2008; Ratinov et al., 2011]

Hyperlink Anchor Text

Dell

Round Rock, Texas

Round Rock is perhaps best known as the international headquarters of Dell, which employs ...

`Dell, Inc.`

[Cucerzan 2007; Milne and Witten, 2008; Ratinov et al., 2011]

Hyperlink Anchor Text

Dell

Round Rock, Texas

Round Rock is perhaps best known as the international headquarters of Dell, which employs ...

↓

`Dell, Inc.`

Western Publishing

With partners Dell and Simon & Schuster, the company ...

↓

`Dell Comics`

[Cucerzan 2007; Milne and Witten, 2008; Ratinov et al., 2011]

Hyperlink Anchor Text

Dell

Round Rock, Texas

Round Rock is perhaps best known as the international headquarters of Dell, which employs ...

↓

`Dell, Inc.`

Western Publishing

With partners Dell and Simon & Schuster, the company ...

↓

`Dell Comics`

...

[Cucerzan 2007; Milne and Witten, 2008; Ratinov et al., 2011]

Hyperlink Anchor Text

Dell

Round Rock, Texas

Round Rock is perhaps best known as the international headquarters of Dell, which employs ...

Dell, Inc.

Western Publishing

With partners Dell and Simon & Schuster, the company ...

Dell Comics

...

[Cucerzan 2007; Milne and Witten, 2008; Ratinov et al., 2011]

Query Selection

Query Selection

CEO Michael Dell

Query Selection

CEO Michael Dell

CEO Michael Dell [none]

Query Selection

CEO Michael Dell

CEO Michael Dell [none]

Dell

Query Selection

CEO Michael Dell

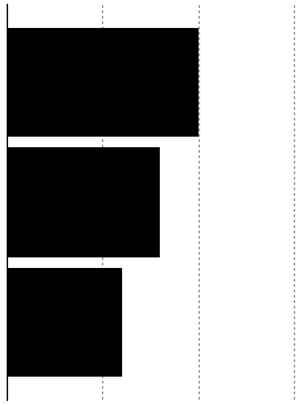
CEO Michael Dell [none]

Michael Dell Michael Dell

Dell Dell, Inc.

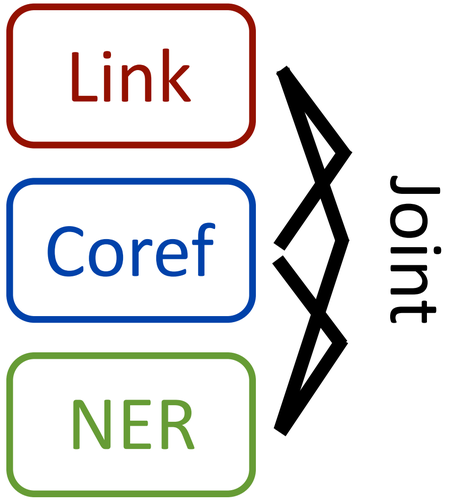
Michael Dell

Dell Comics



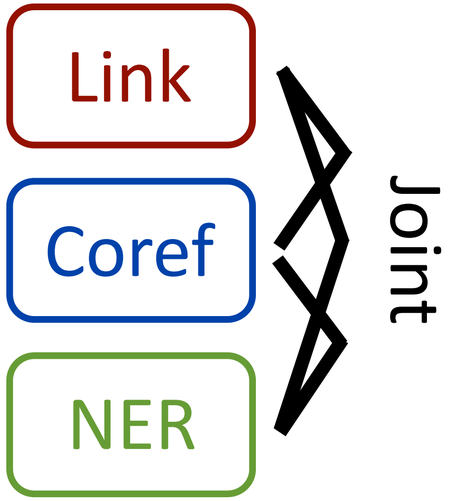
Outline

- Model
  - Independent models (unary, per-task factors)
  - Joint coupling (higher-arity, cross-task factors)
- Experiments



Outline

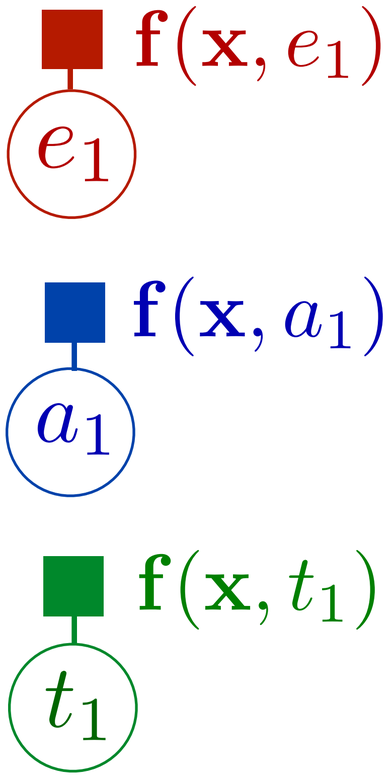
- Model
  - Independent models (unary, per-task factors)
  - Joint coupling (higher-arity, cross-task factors)
- Experiments



Model

$$\mathbf{p}(\mathbf{t}, \mathbf{a}, \mathbf{e} | \mathbf{x}) \propto \exp\left(\mathbf{w}^\top \mathbf{f}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e})\right)$$

$$\mathbf{f}_{\text{indep}}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e}) = \sum_{i=1}^{n} \left[\mathbf{f}(\mathbf{x}, t_i) + \mathbf{f}(\mathbf{x}, a_i) + \mathbf{f}(\mathbf{x}, e_i)\right]$$

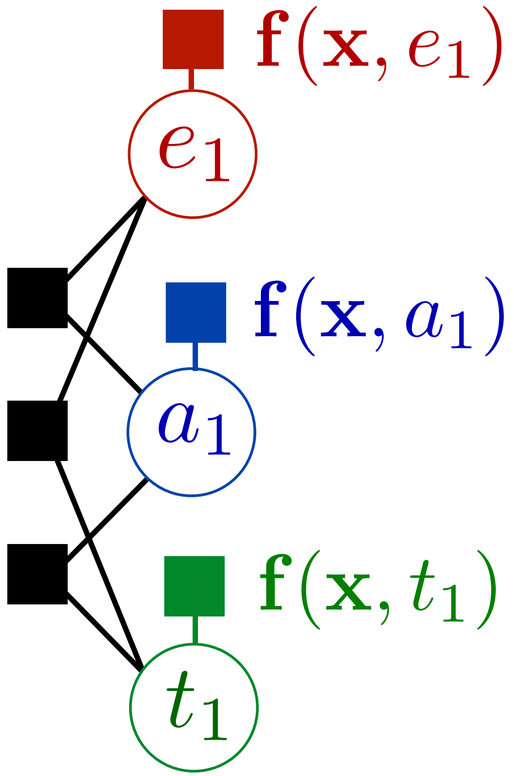


Dell

Model

$$\mathbf{p}(\mathbf{t}, \mathbf{a}, \mathbf{e} | \mathbf{x}) \propto \exp\left(\mathbf{w}^\top \mathbf{f}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e})\right)$$

$$\mathbf{f}_{\text{indep}}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e}) = \sum_{i=1}^{n} \left[\mathbf{f}(\mathbf{x}, t_i) + \mathbf{f}(\mathbf{x}, a_i) + \mathbf{f}(\mathbf{x}, e_i)\right]$$



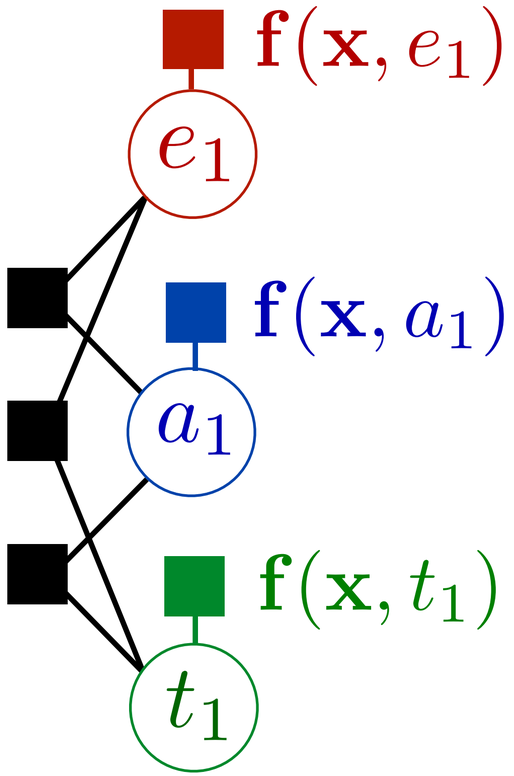
Dell

Model

$$\mathbf{p}(\mathbf{t}, \mathbf{a}, \mathbf{e} | \mathbf{x}) \propto \exp\left(\mathbf{w}^\top \mathbf{f}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e})\right)$$

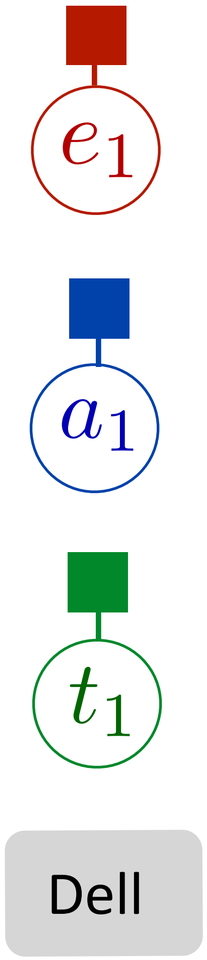
$$\mathbf{f}_{\text{indep}}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e}) = \sum_{i=1}^{n} \left[\mathbf{f}(\mathbf{x}, t_i) + \mathbf{f}(\mathbf{x}, a_i) + \mathbf{f}(\mathbf{x}, e_i)\right]$$

$$\mathbf{f}_{\text{joint}}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e}) = \mathbf{f}_{\text{indep}}(\mathbf{x}, \mathbf{t}, \mathbf{a}, \mathbf{e}) + \sum \text{joint factors}$$

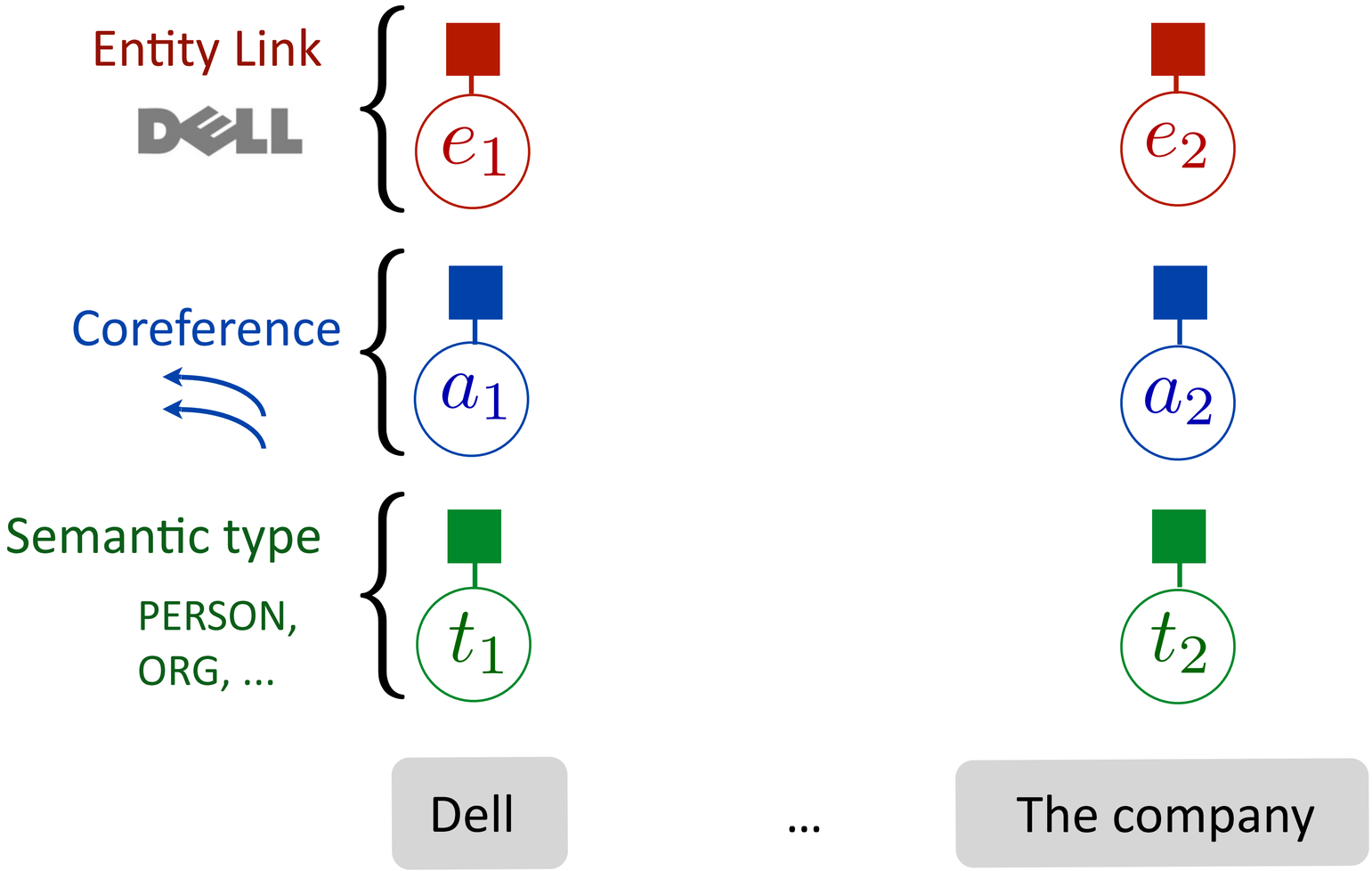


Dell

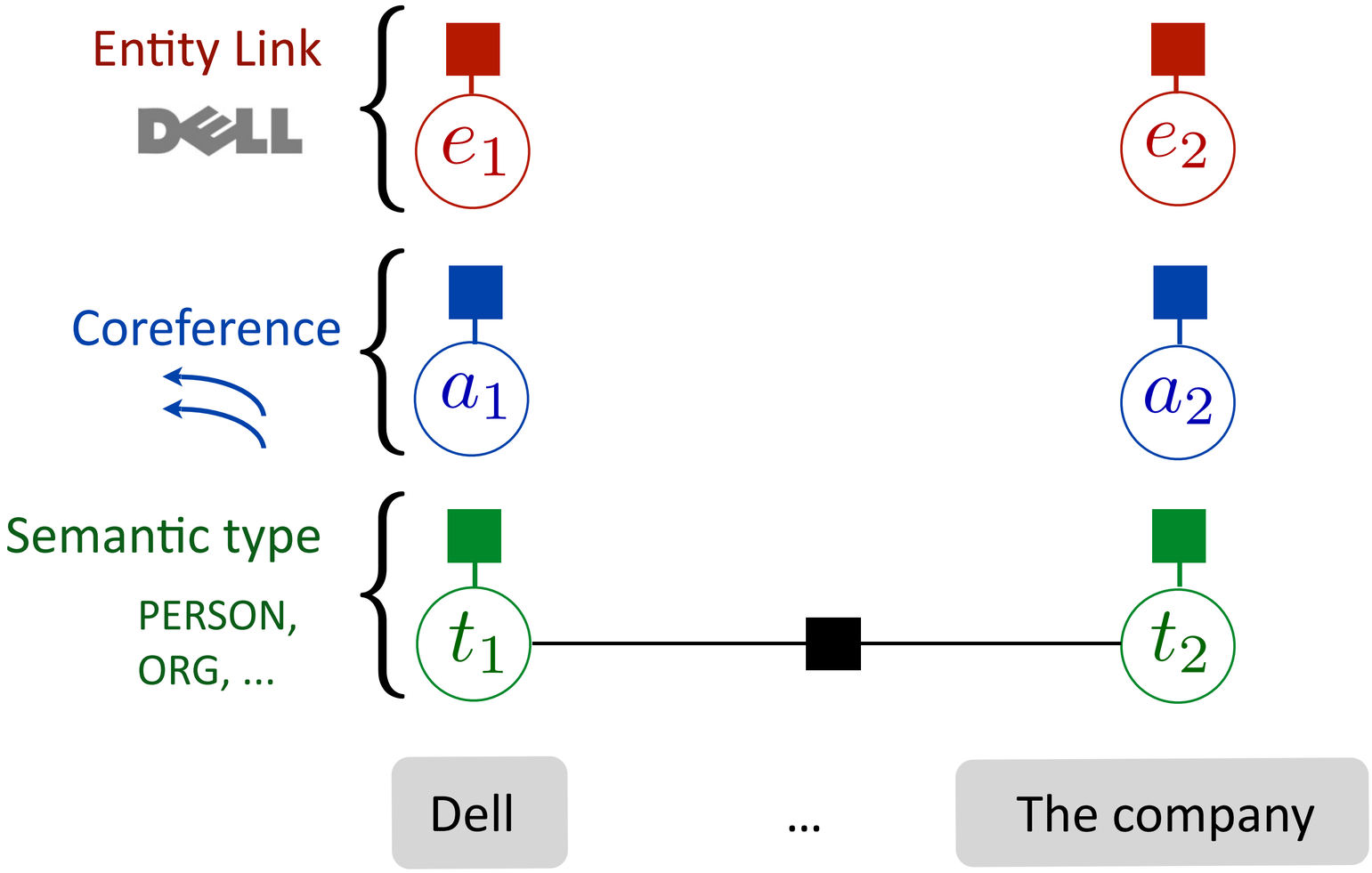
Model



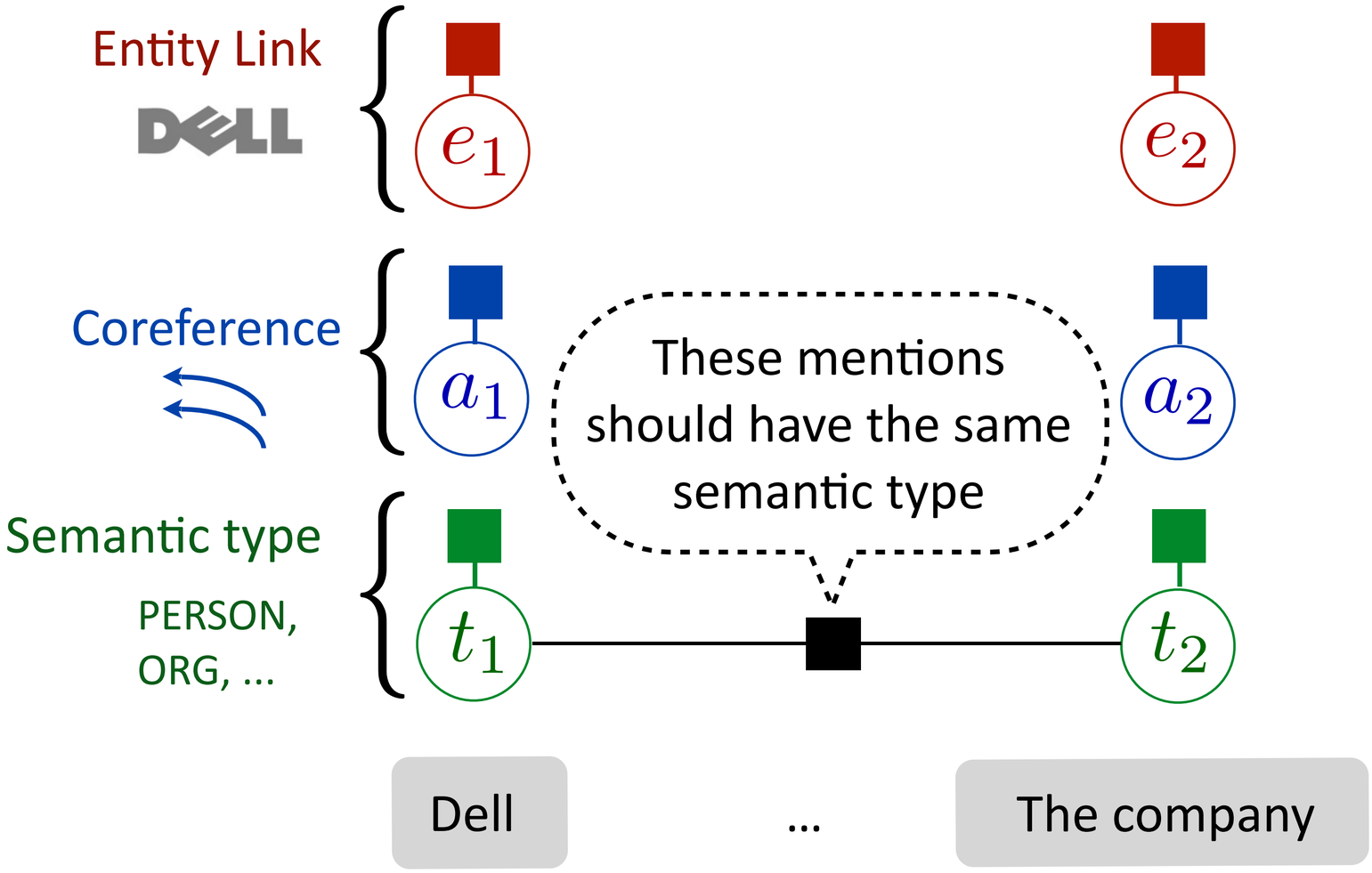
Coreference — Typing



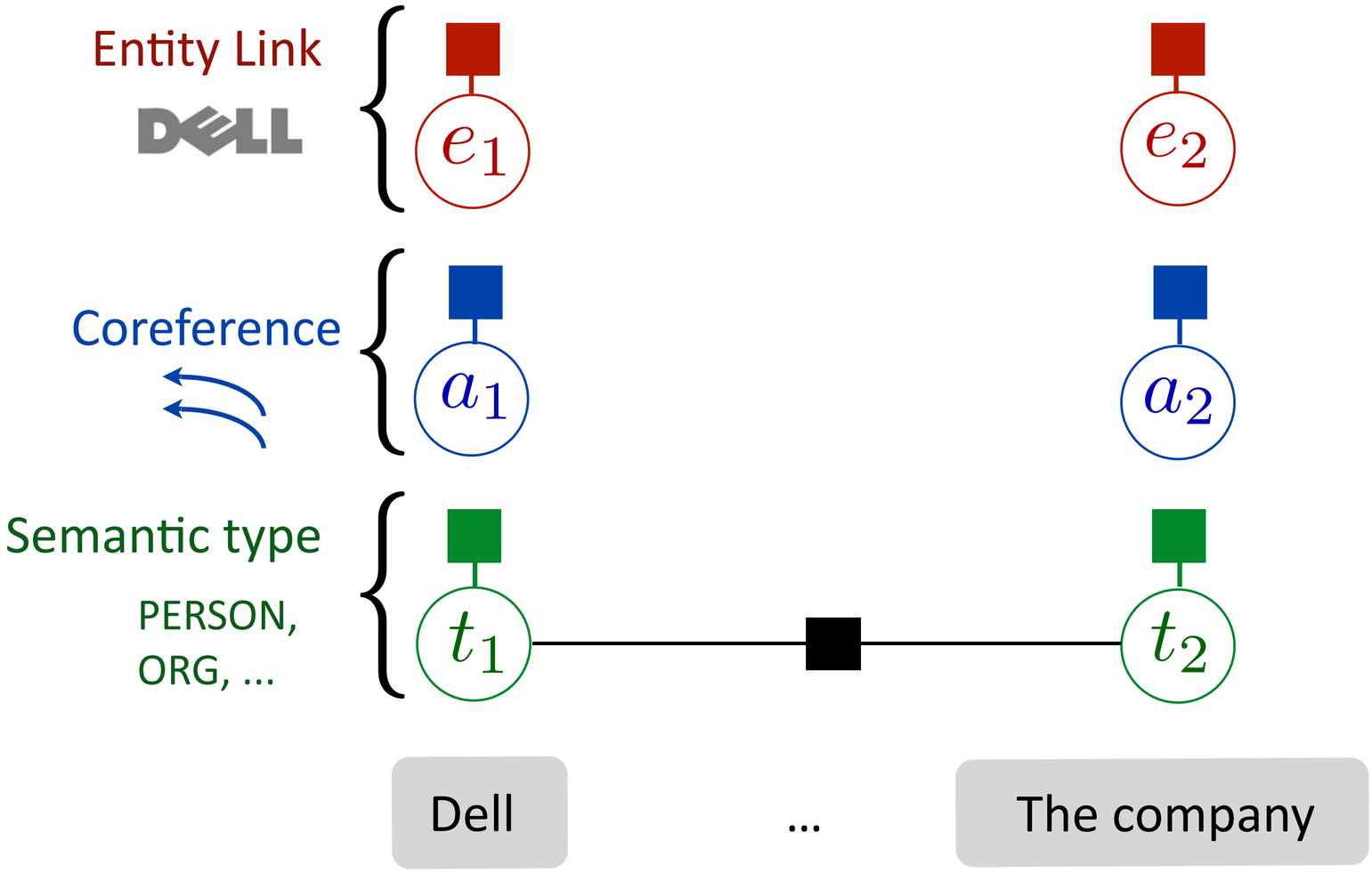
Coreference — Typing



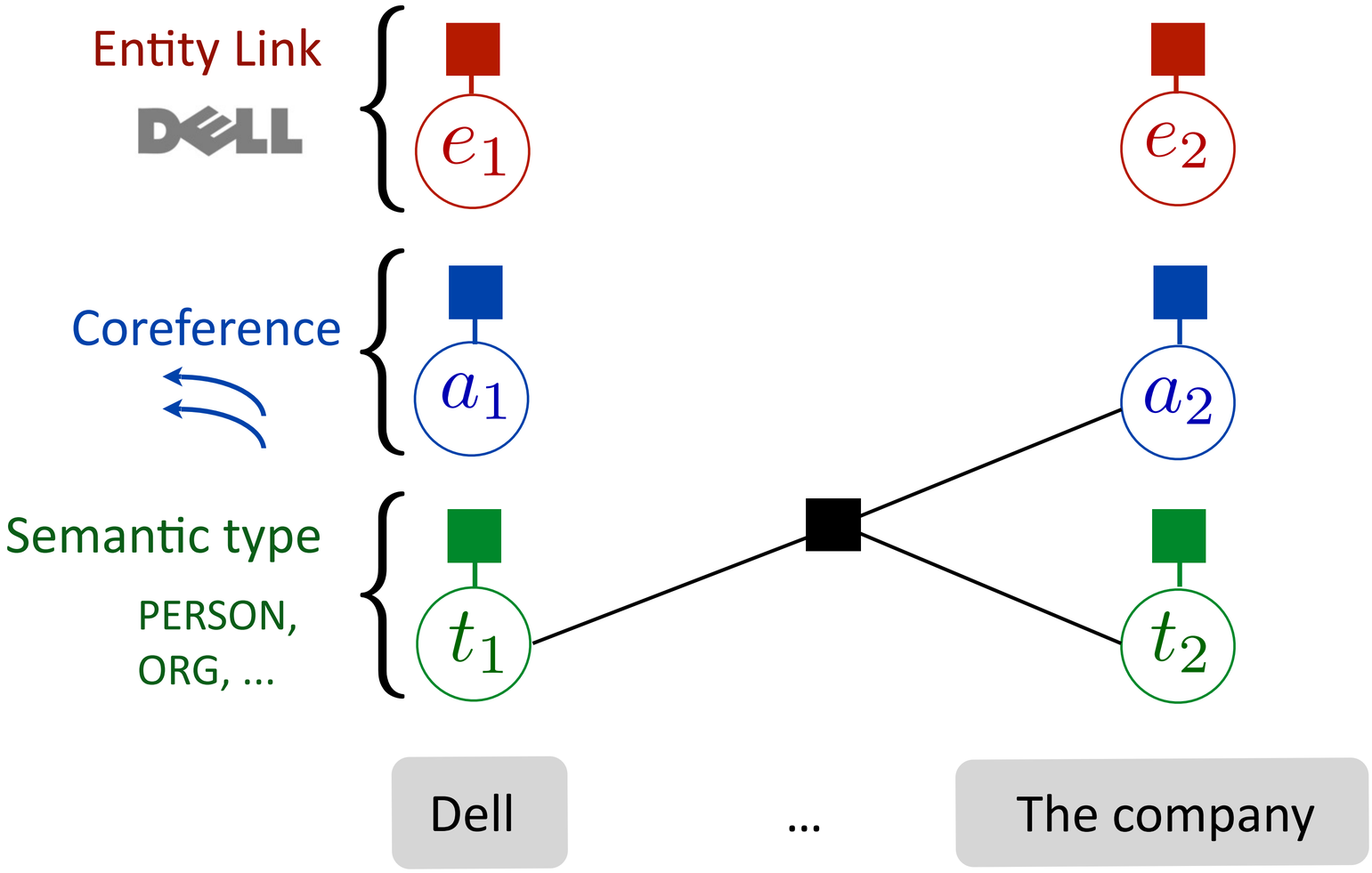
Coreference — Typing



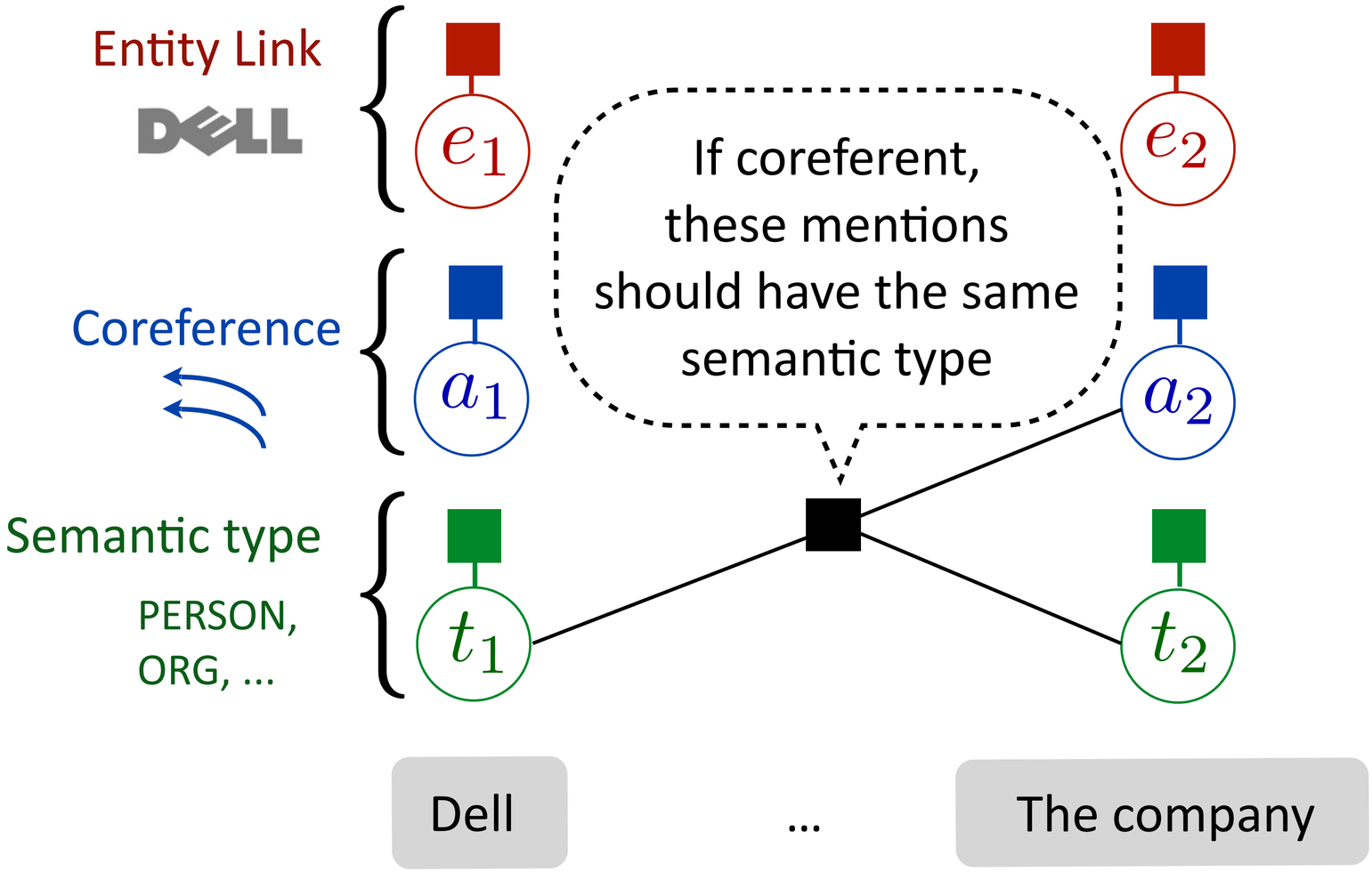
Coreference — Typing



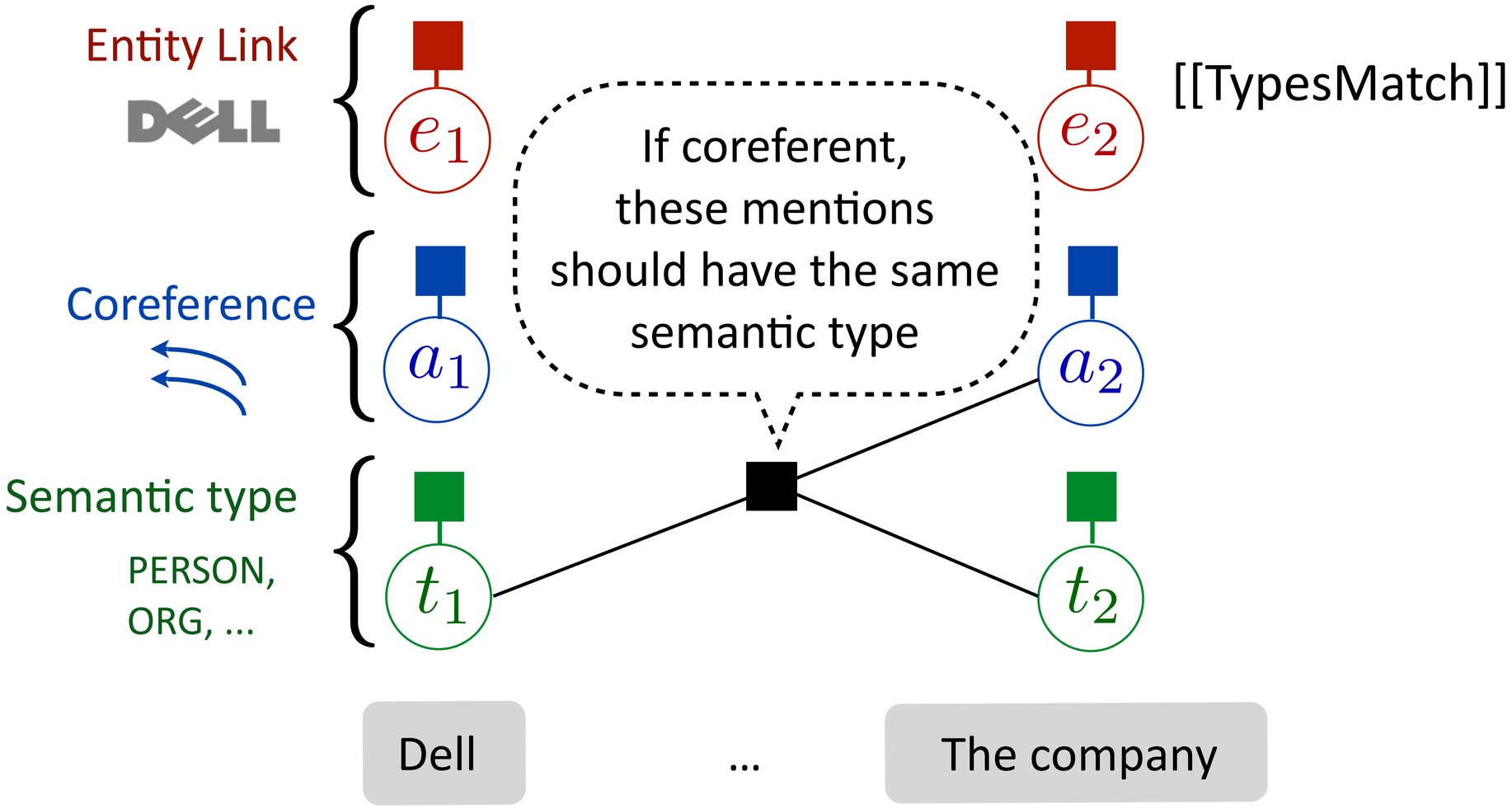
Coreference — Typing



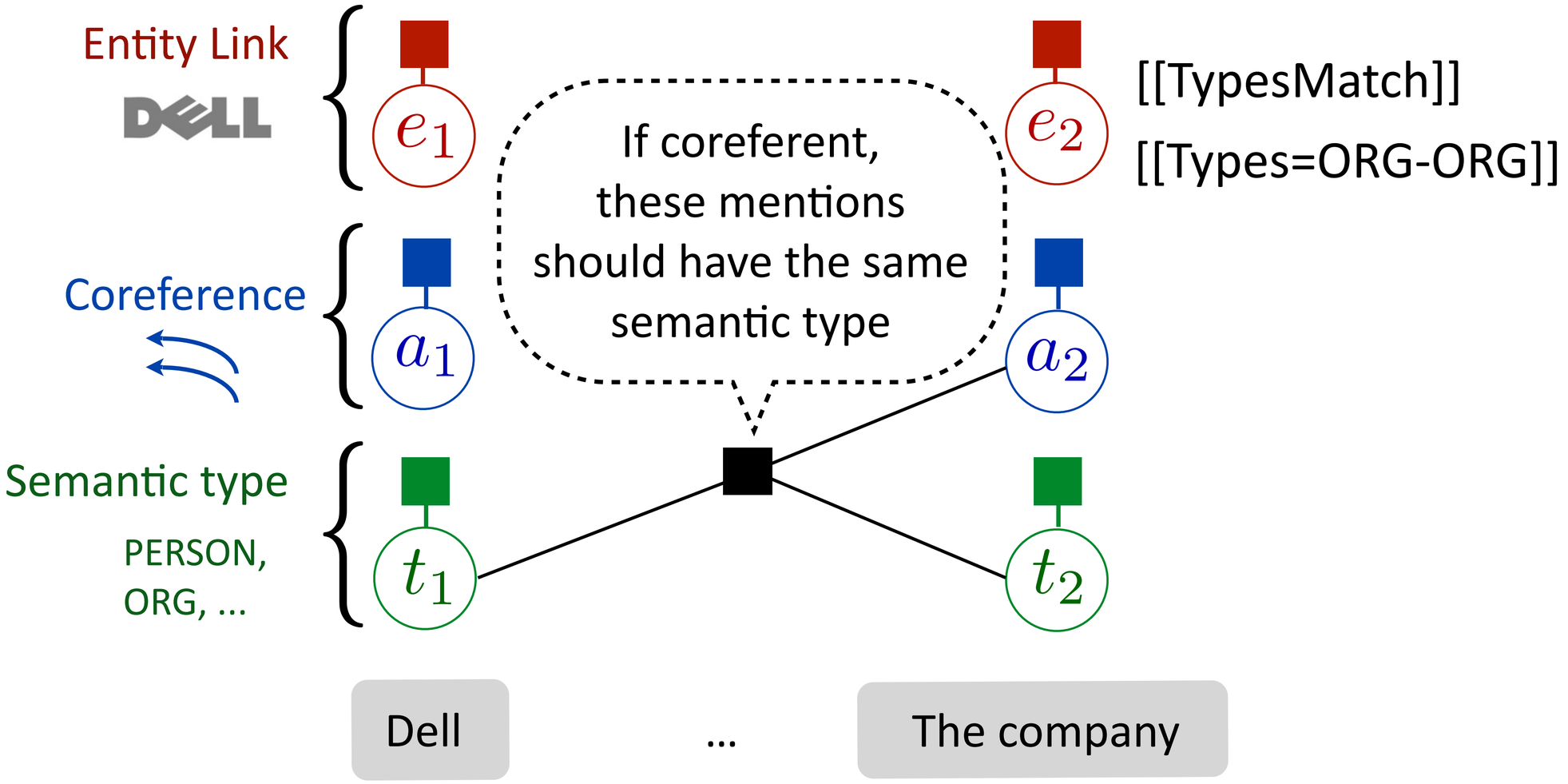
Coreference — Typing



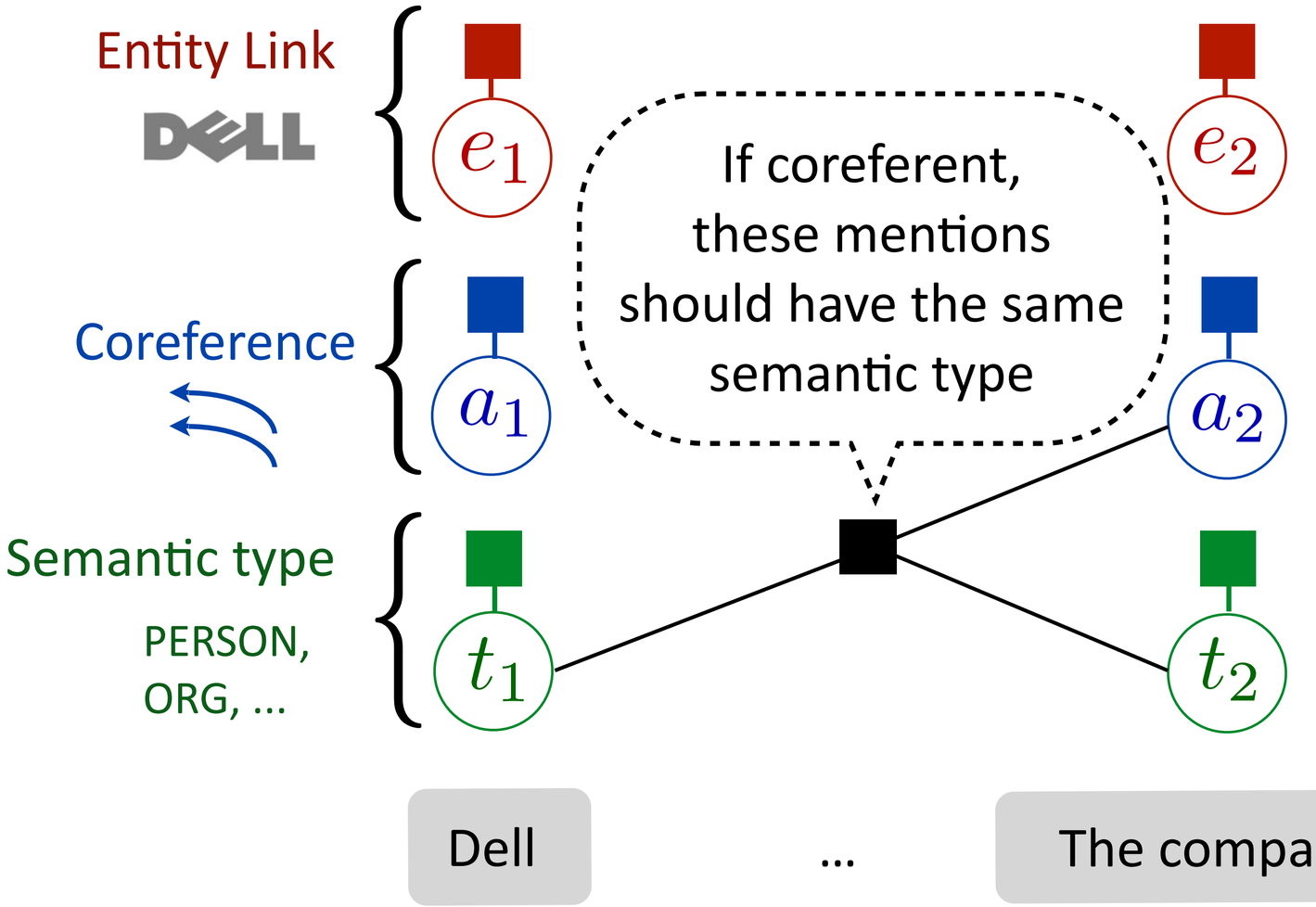
Coreference — Typing



Coreference — Typing



Coreference — Typing



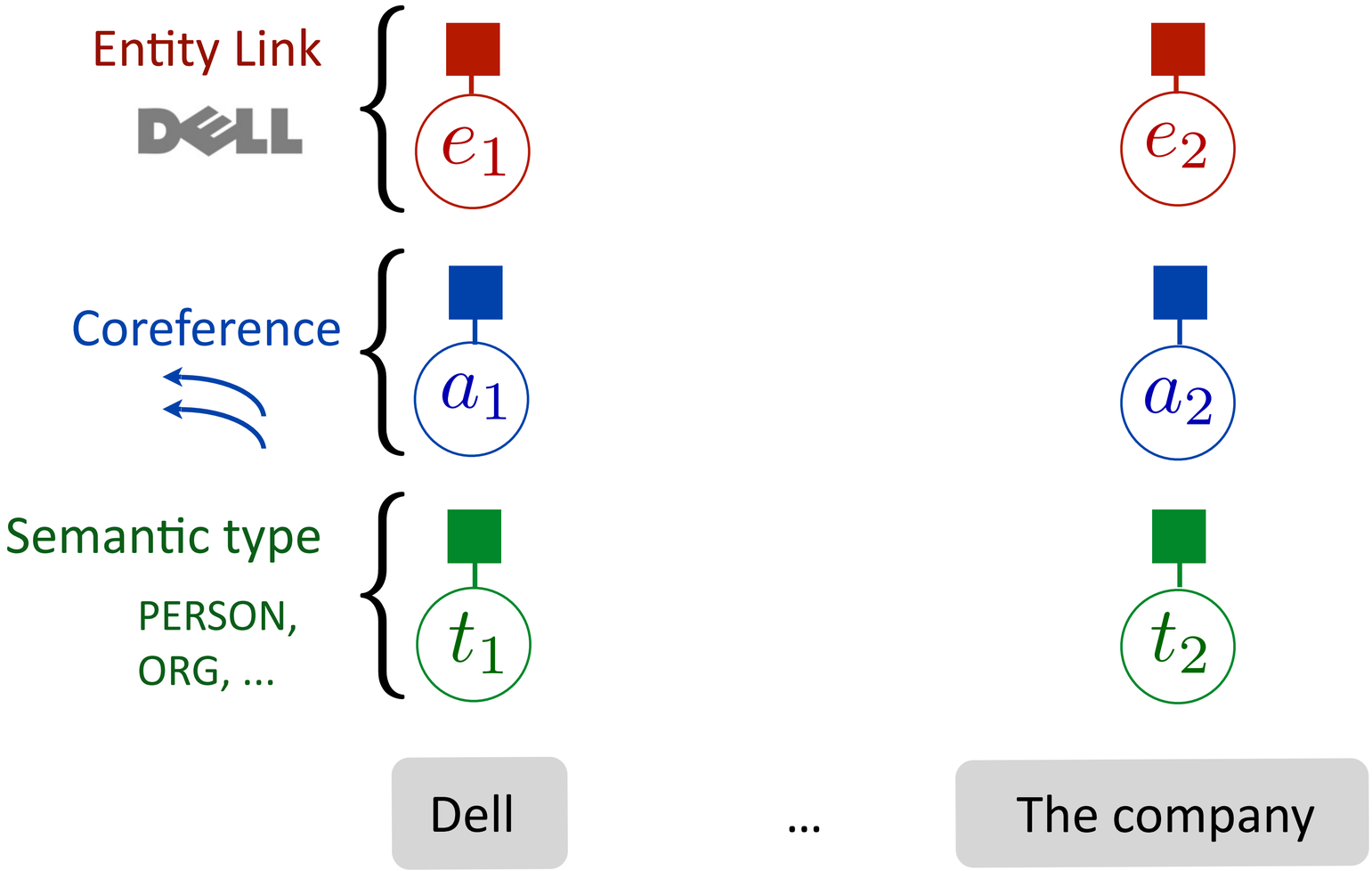
[[TypesMatch]]

[[Types=ORG-ORG]]

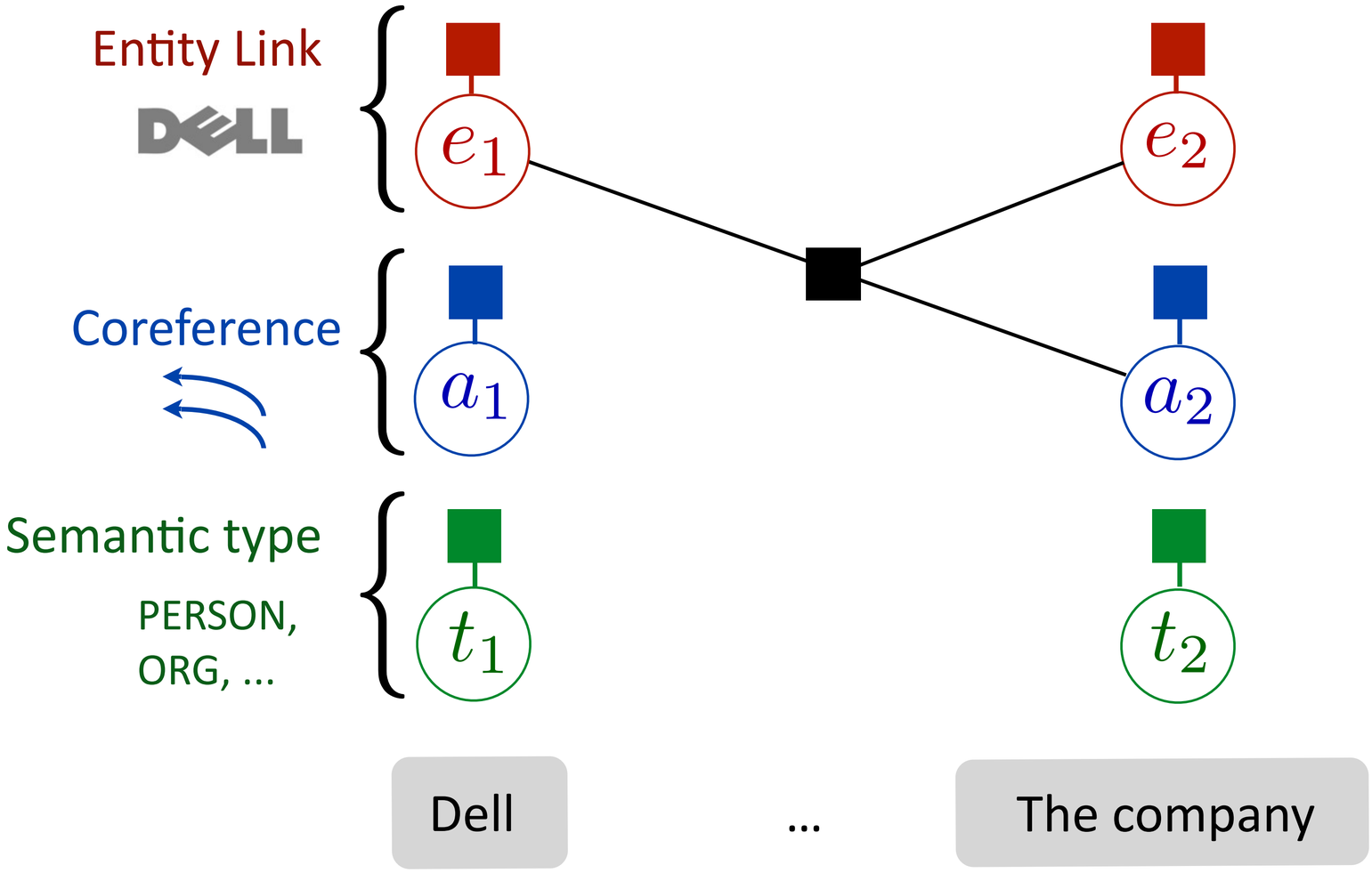
[[Curr=*company* PrevType=*ORG*]]

...

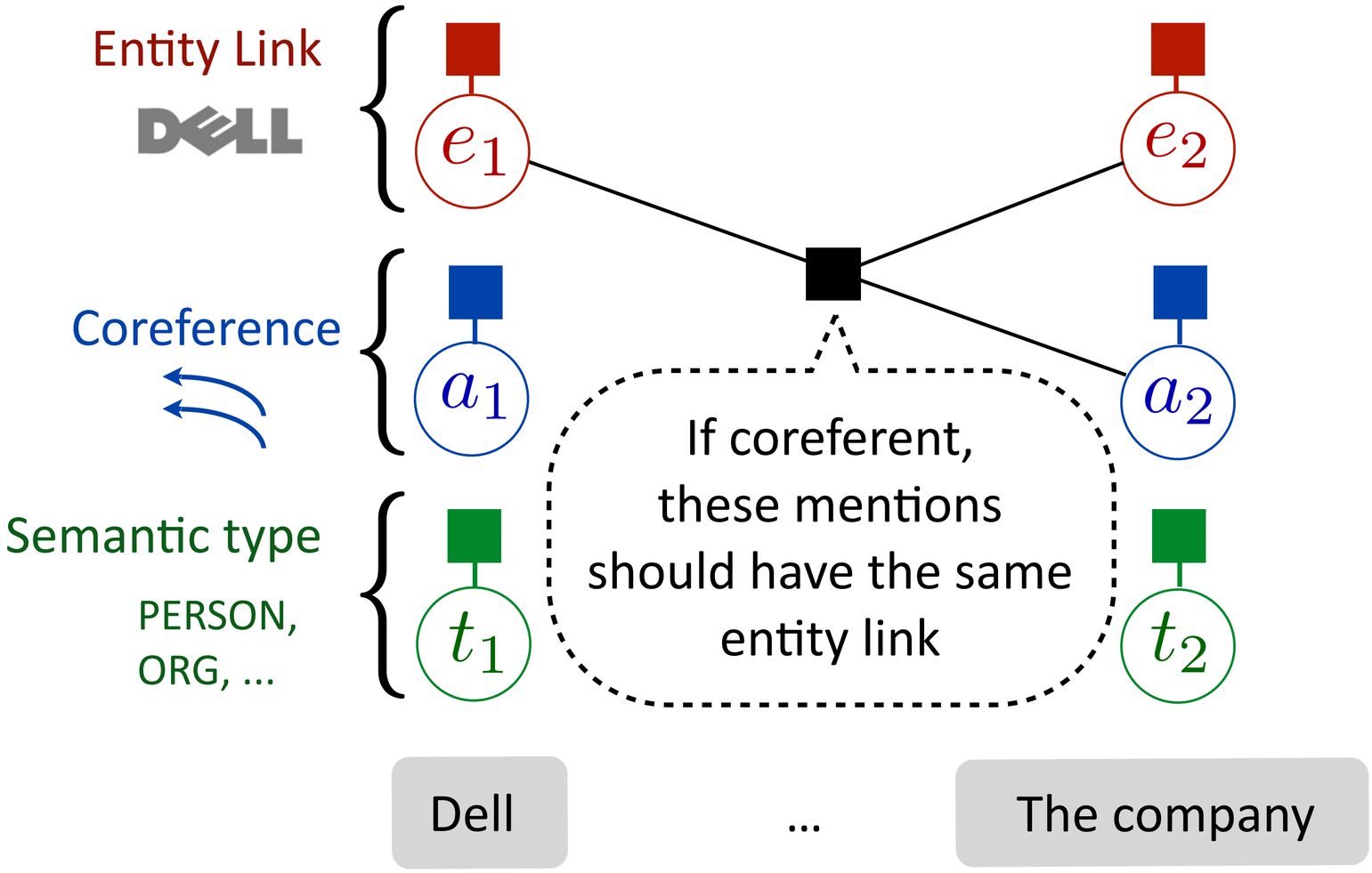
Coreference — Linking



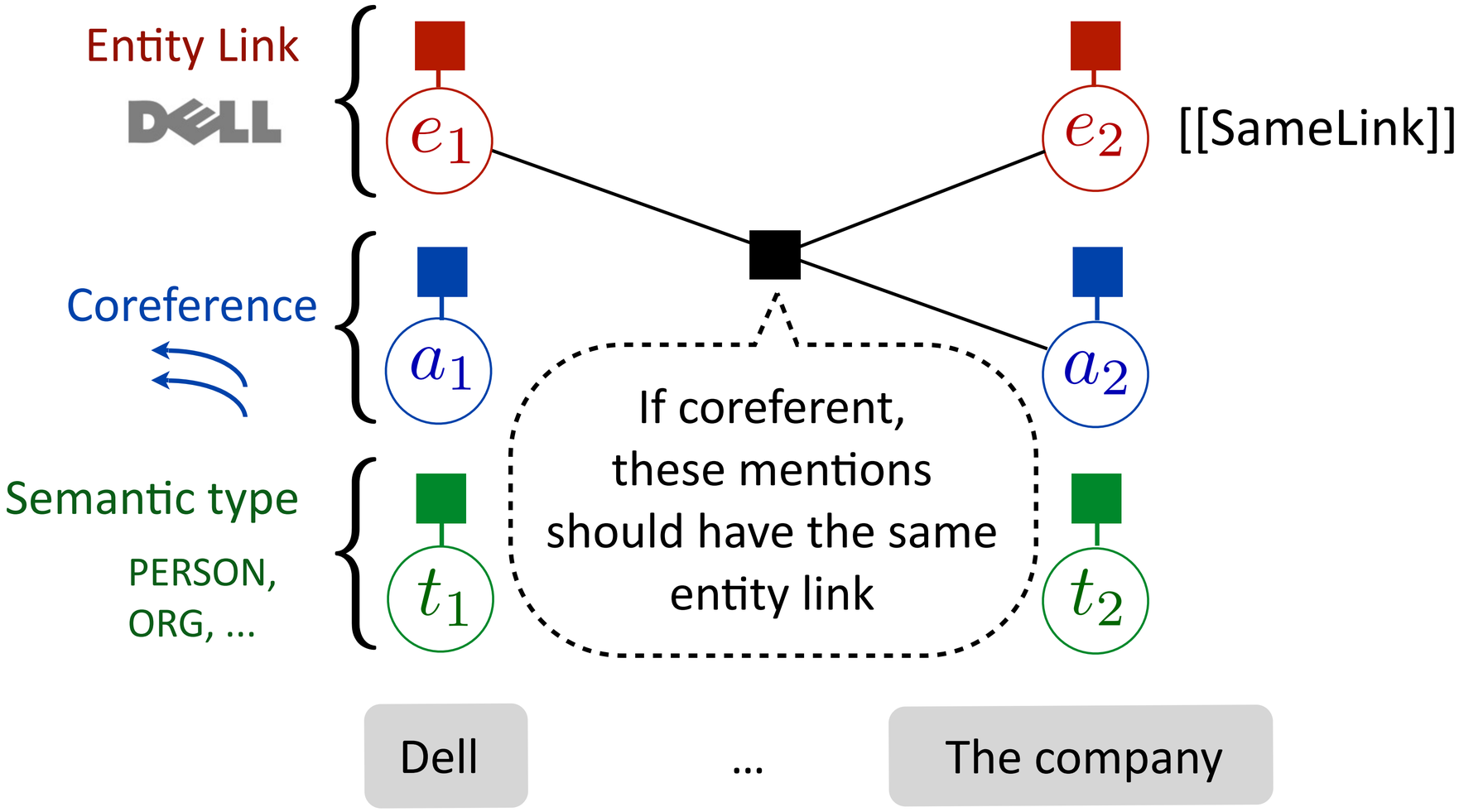
Coreference — Linking



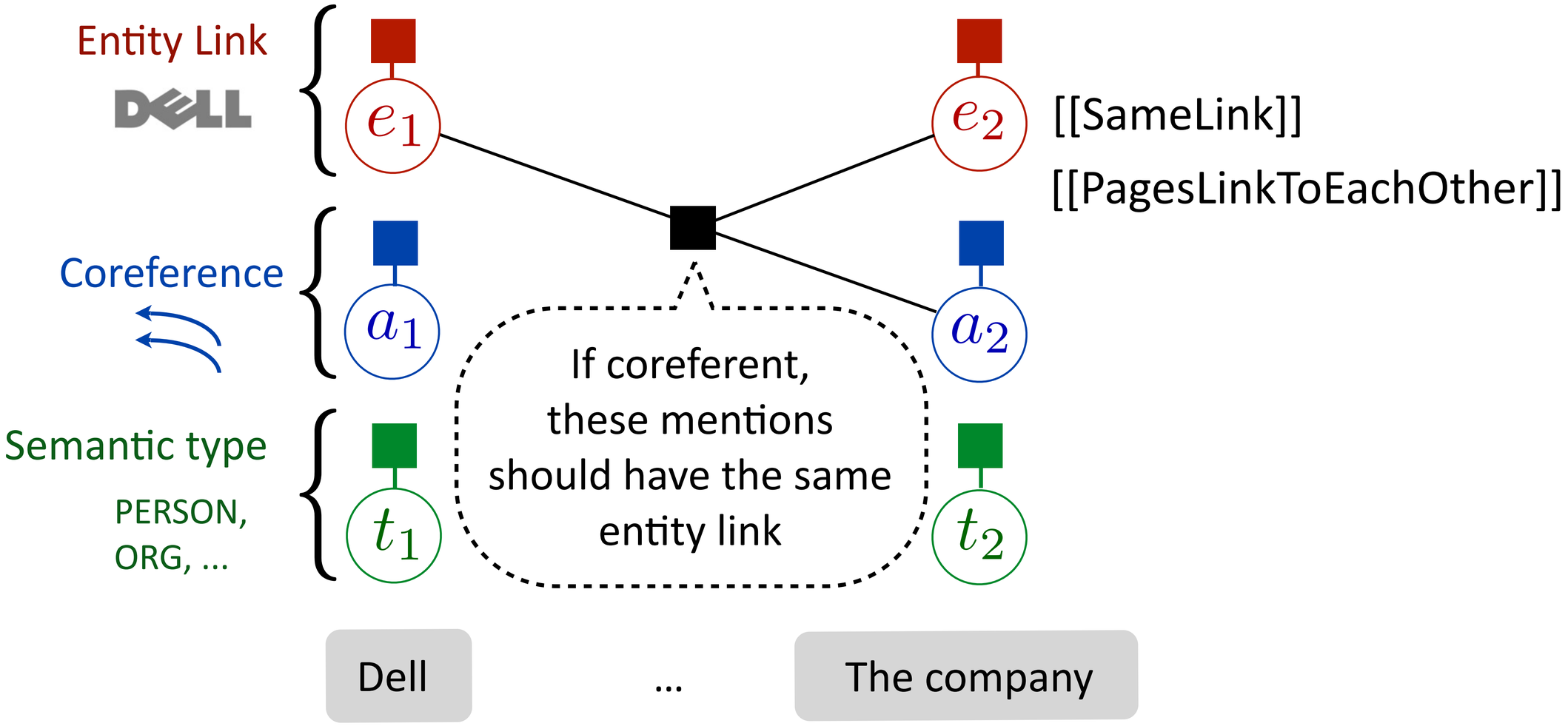
Coreference — Linking



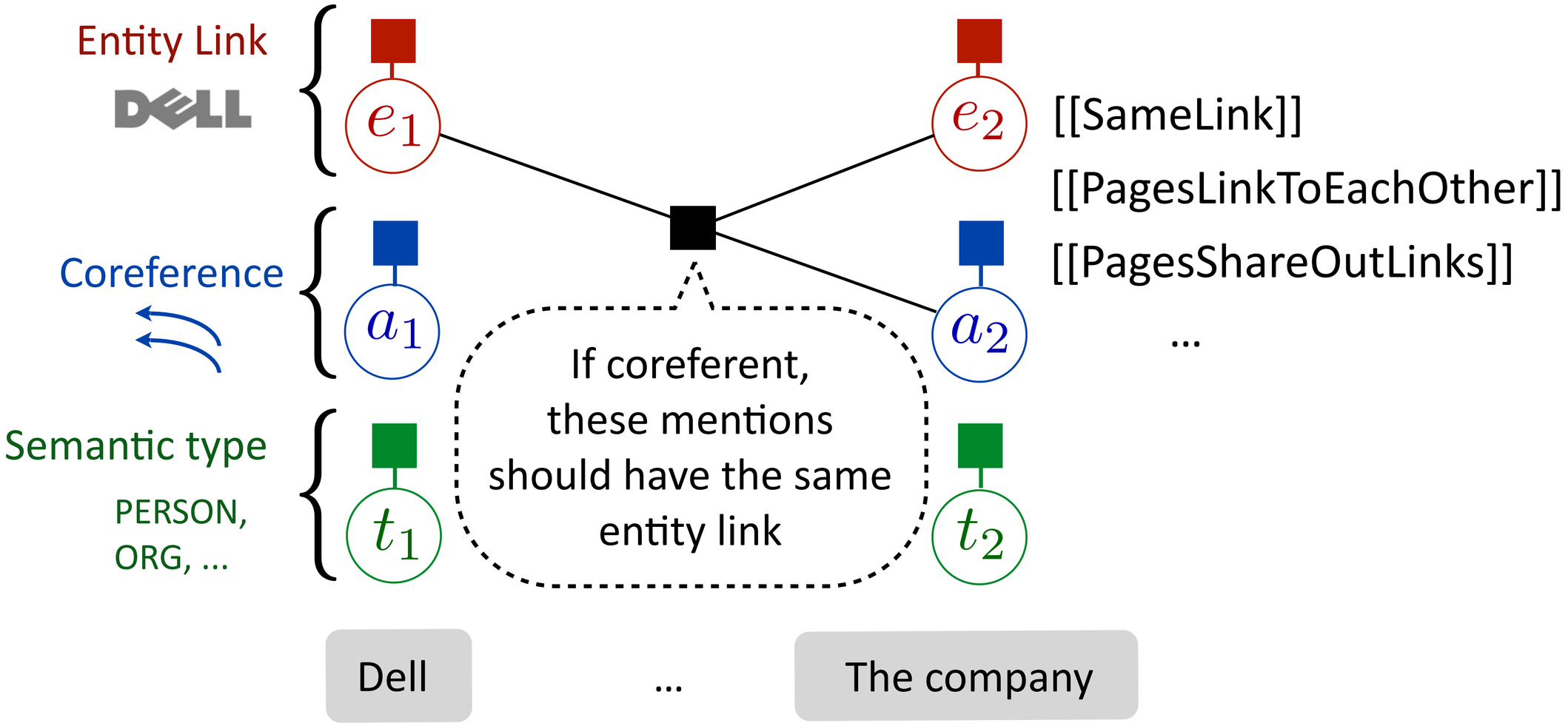
Coreference — Linking



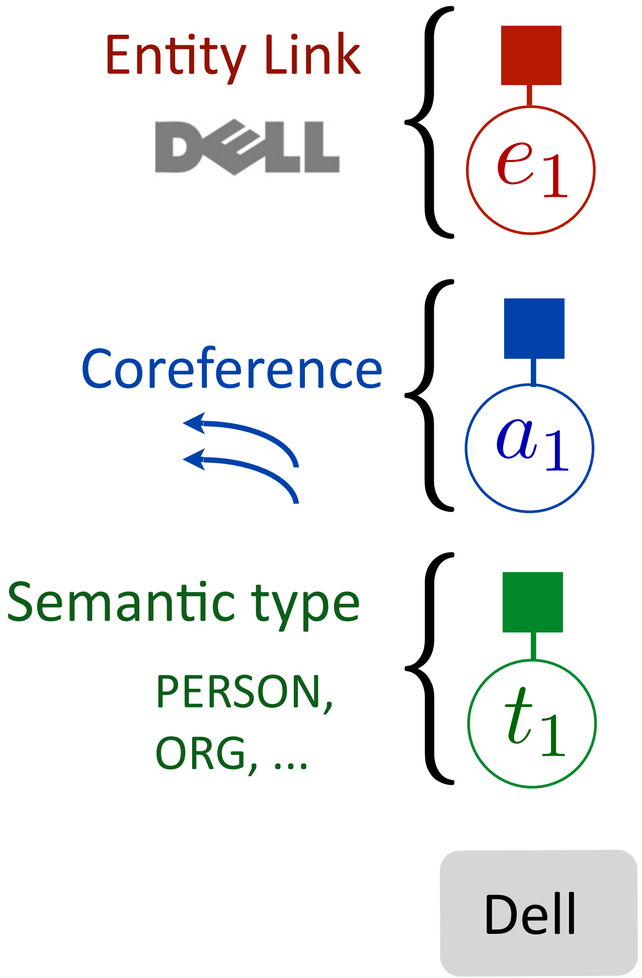
Coreference — Linking



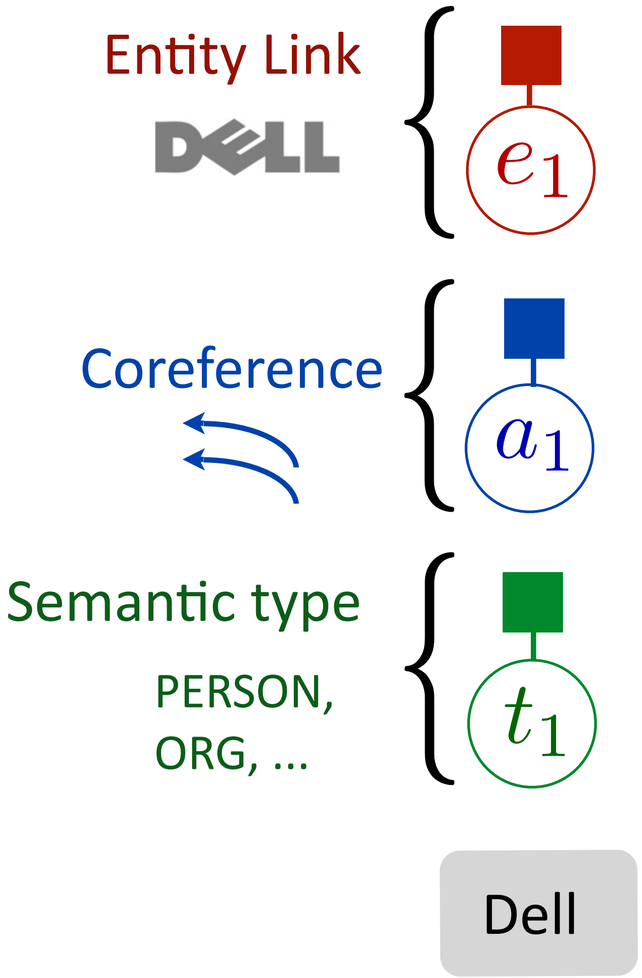
Coreference — Linking



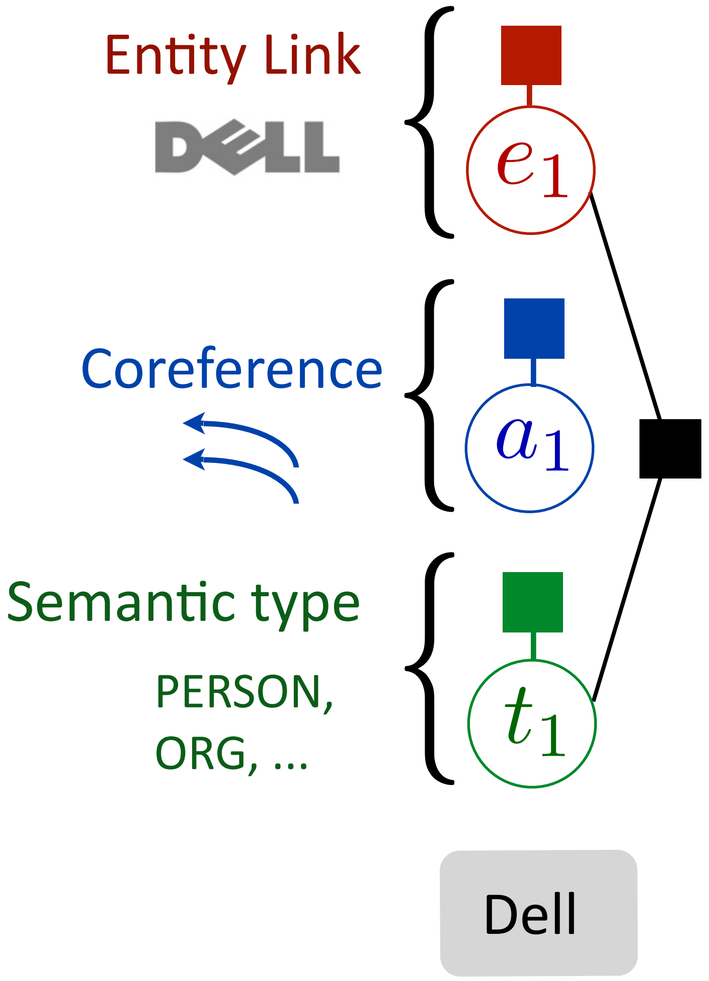
Typing — Linking



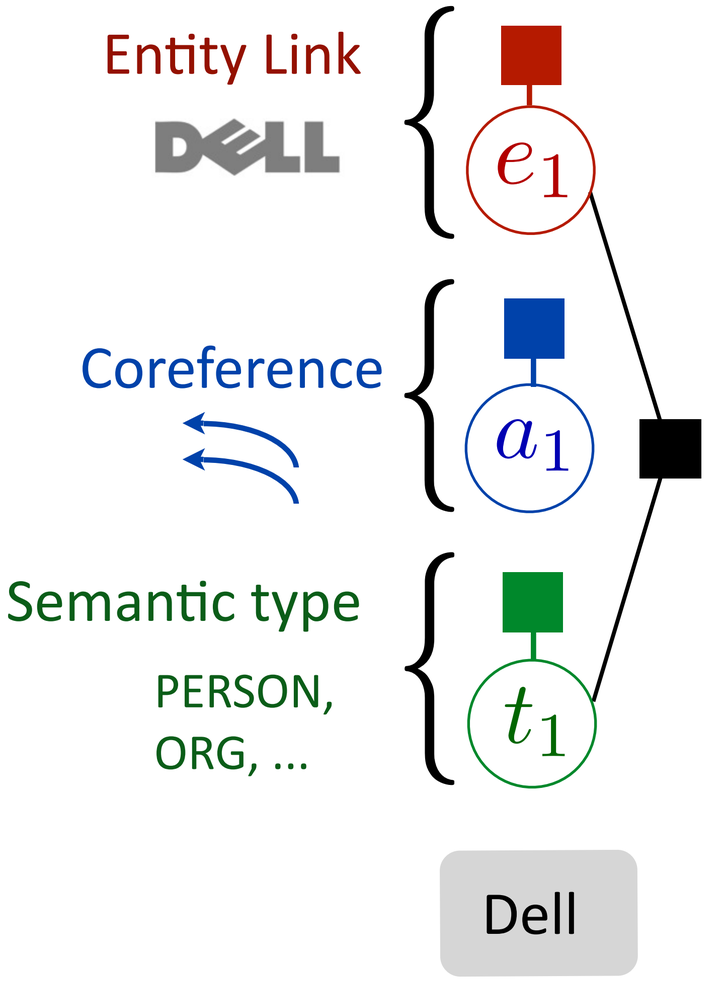
Typing — Linking



Typing — Linking



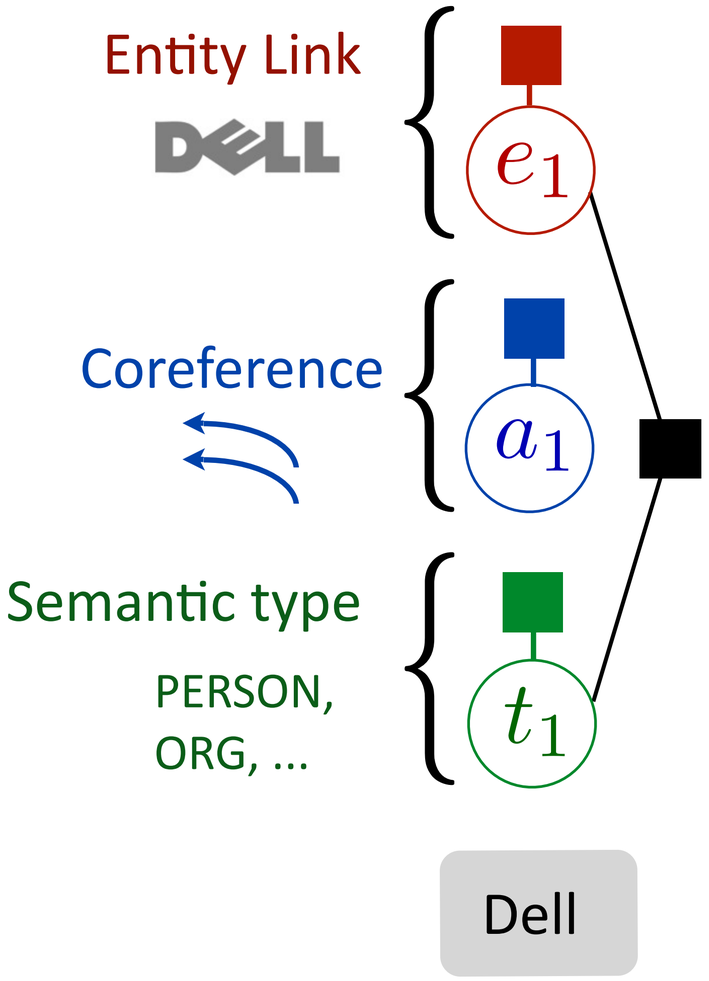
Typing — Linking



Dell

Dell Inc. is an American privately owned multinational computer technology company

Typing — Linking

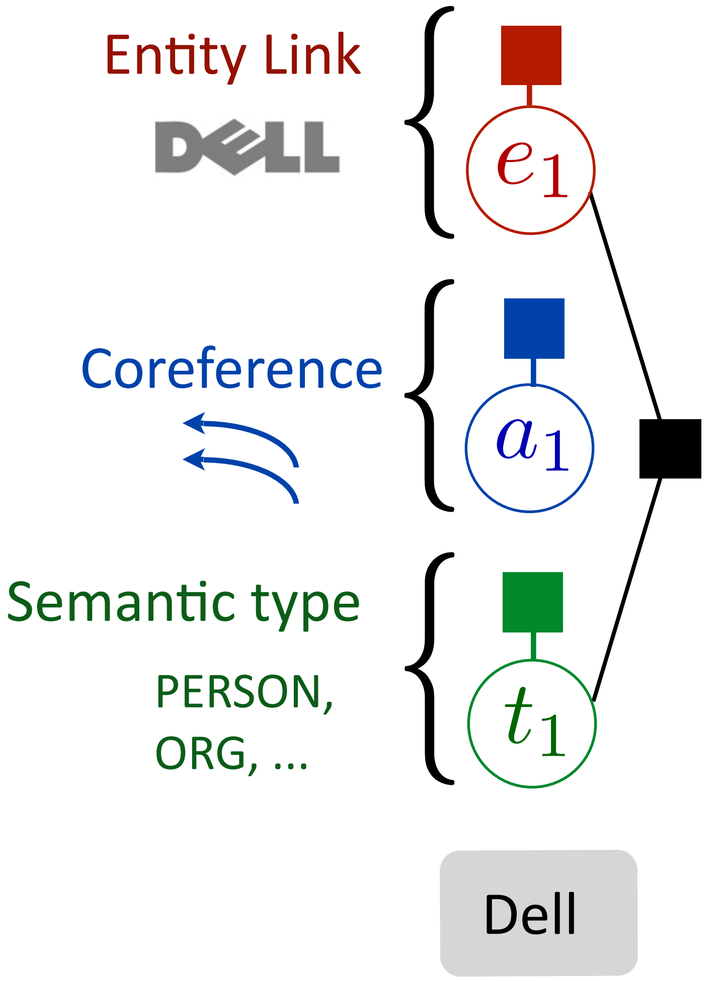


Dell

Dell Inc. is an American privately owned multinational computer technology company

[[ORG $\wedge$ PredNom=*company*]]

Typing — Linking



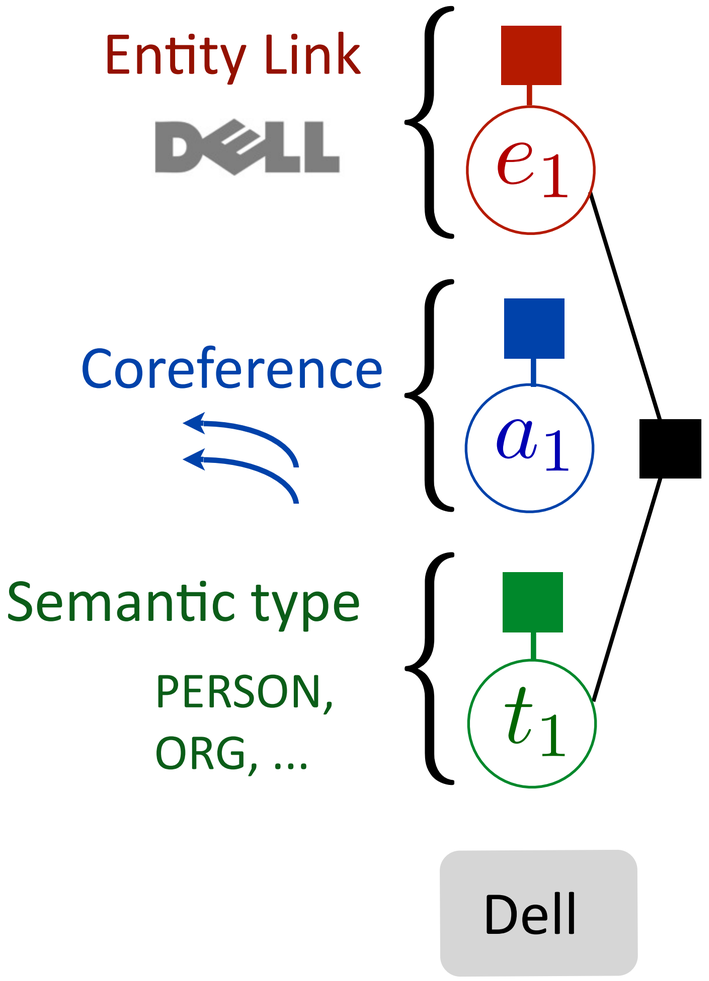
Dell

Dell Inc. is an American privately owned multinational computer technology company

Categories: Companies based in Austin, Texas

[[ORG $\wedge$ PredNom=*company*]]

Typing — Linking



Dell

Dell Inc. is an American privately owned multinational computer technology company

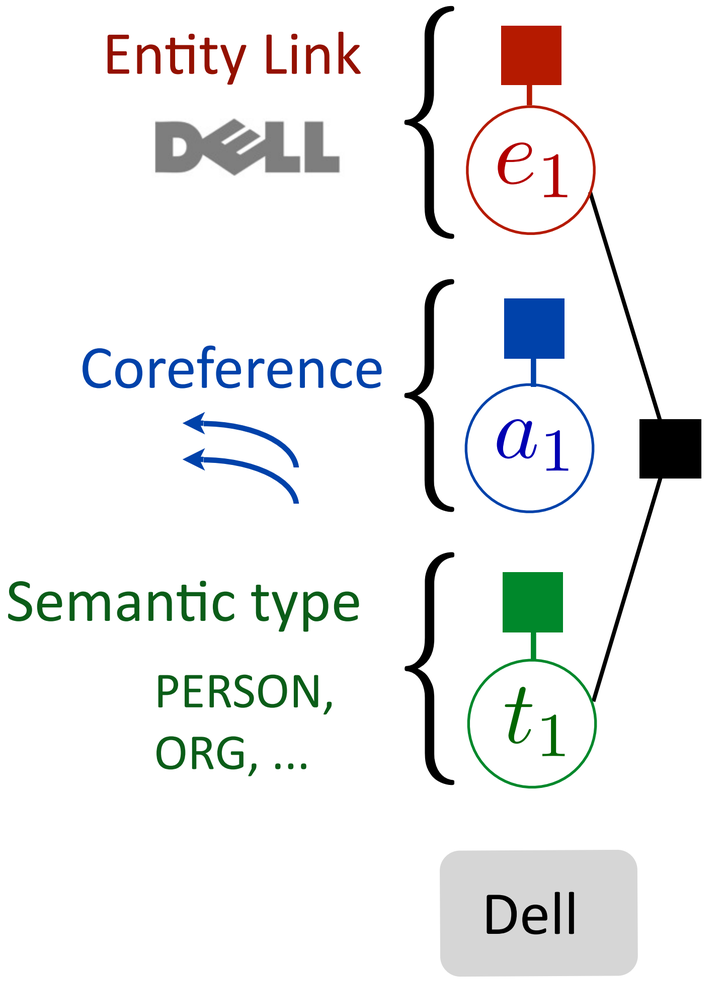
Categories: Companies based in Austin, Texas

[[ORG ∧ PredNom=*company*]]

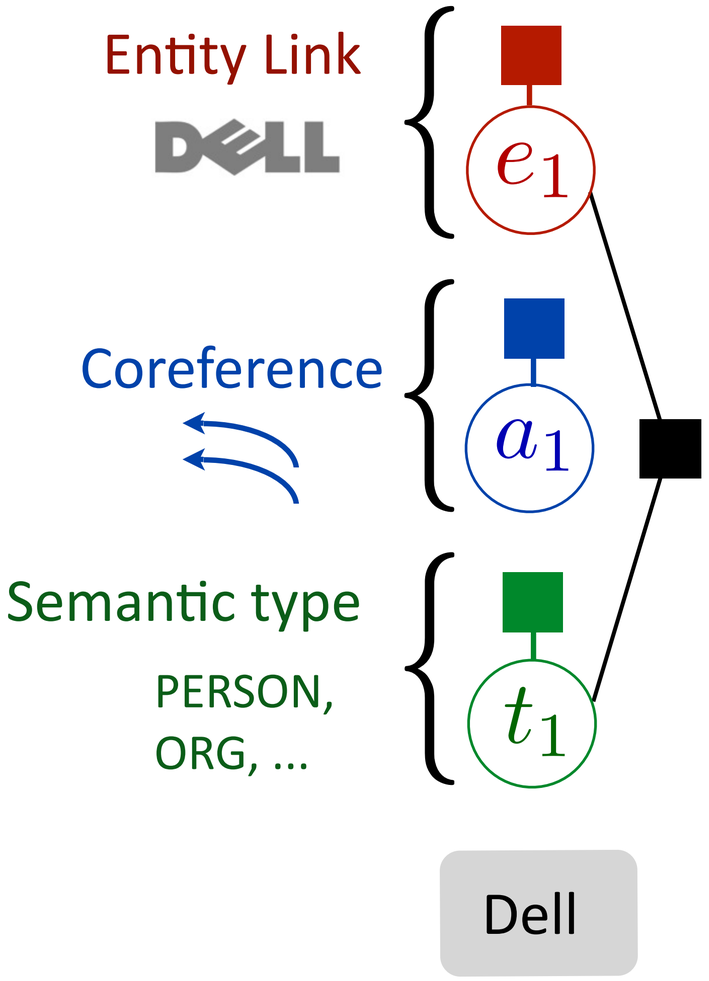
[[ORG ∧ Category=*Companies*]]

...

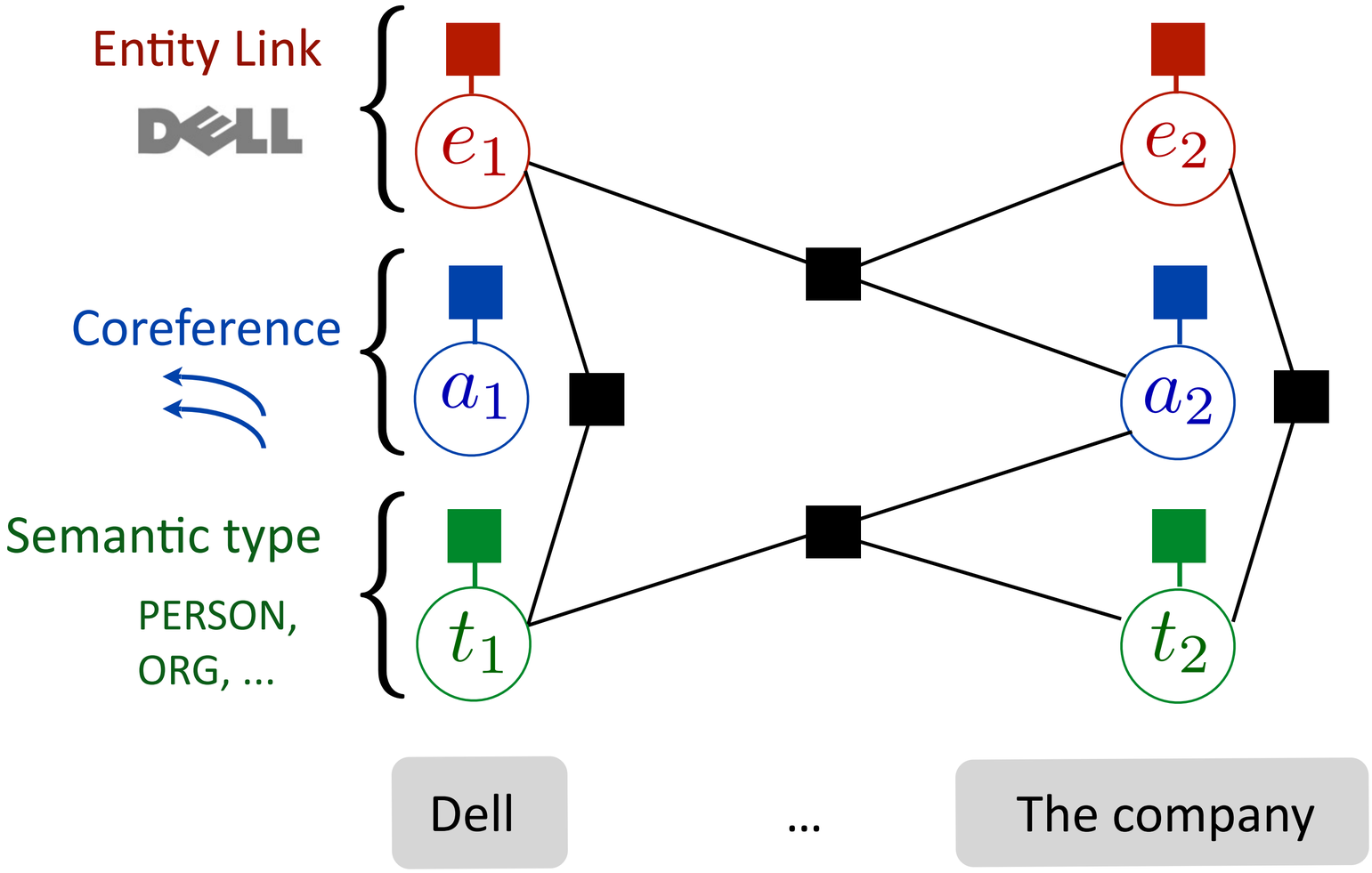
Typing — Linking



Typing — Linking

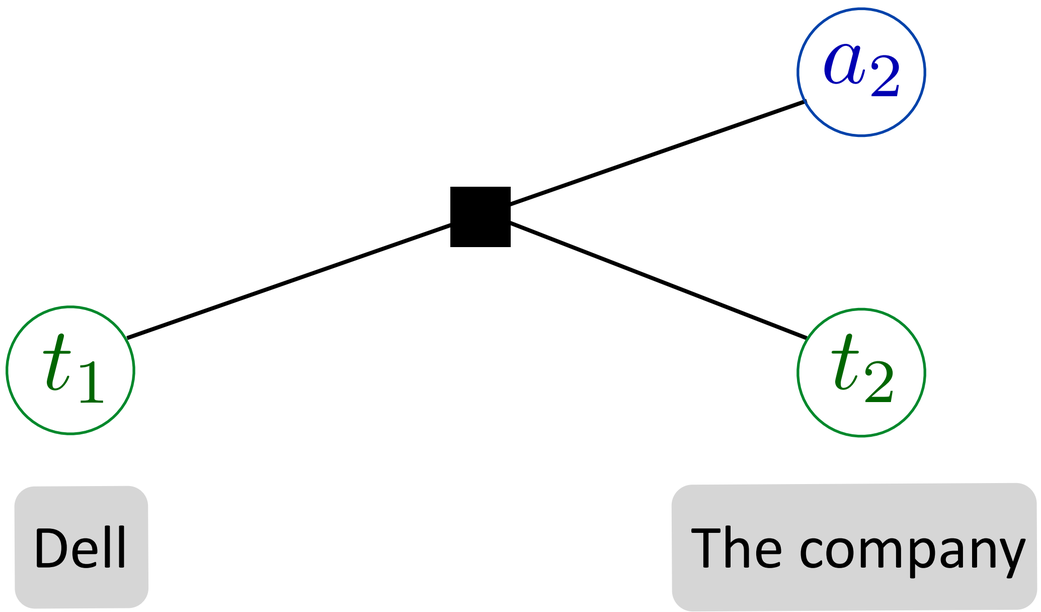


Joint Model

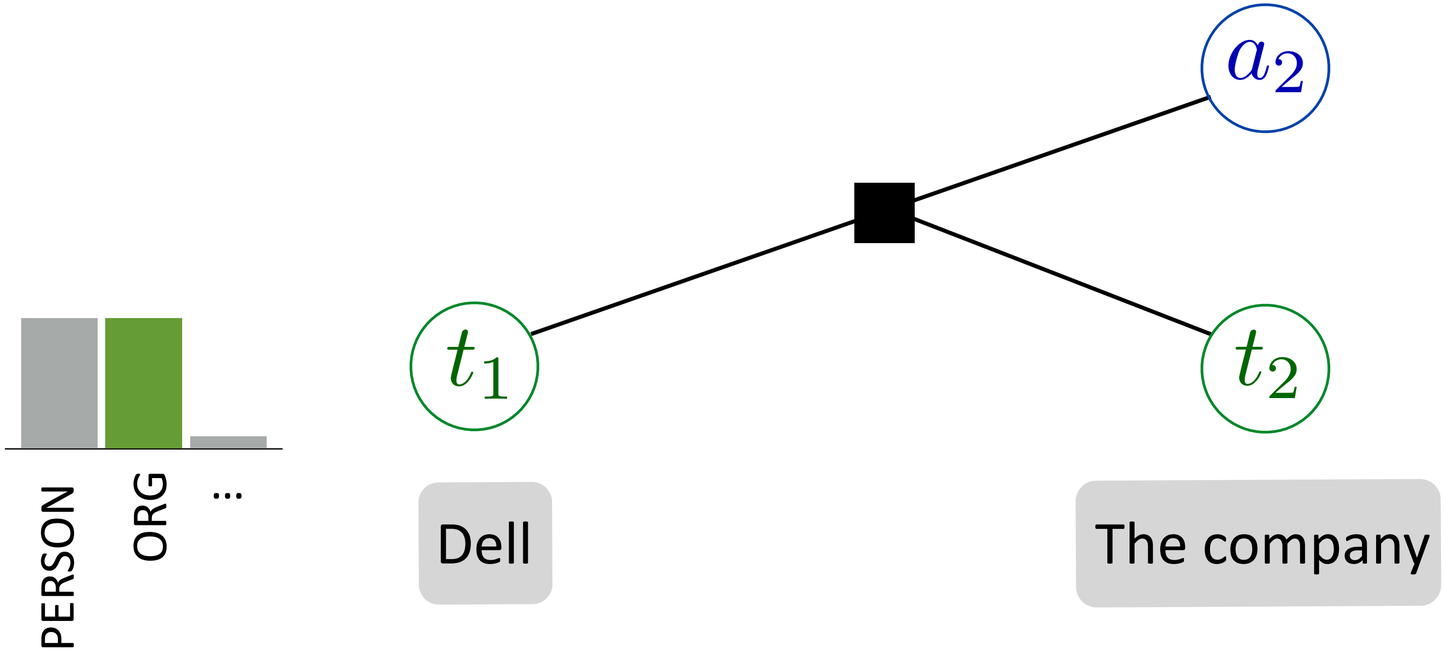

Entity Link
DELL
Coreference
Semantic type
PERSON,
ORG, ...
e_1
e_2
a_1
a_2
t_1
t_2
Dell
...
The company

Inference: Belief Propagation

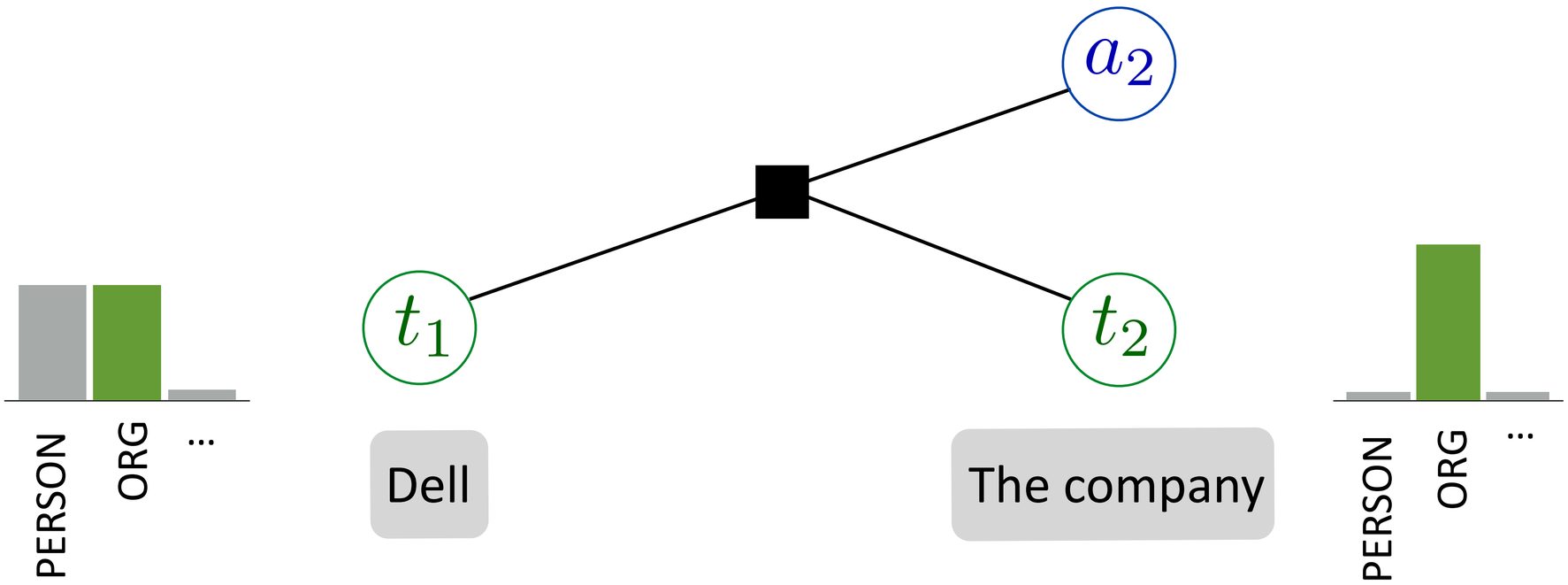
Inference: Belief Propagation



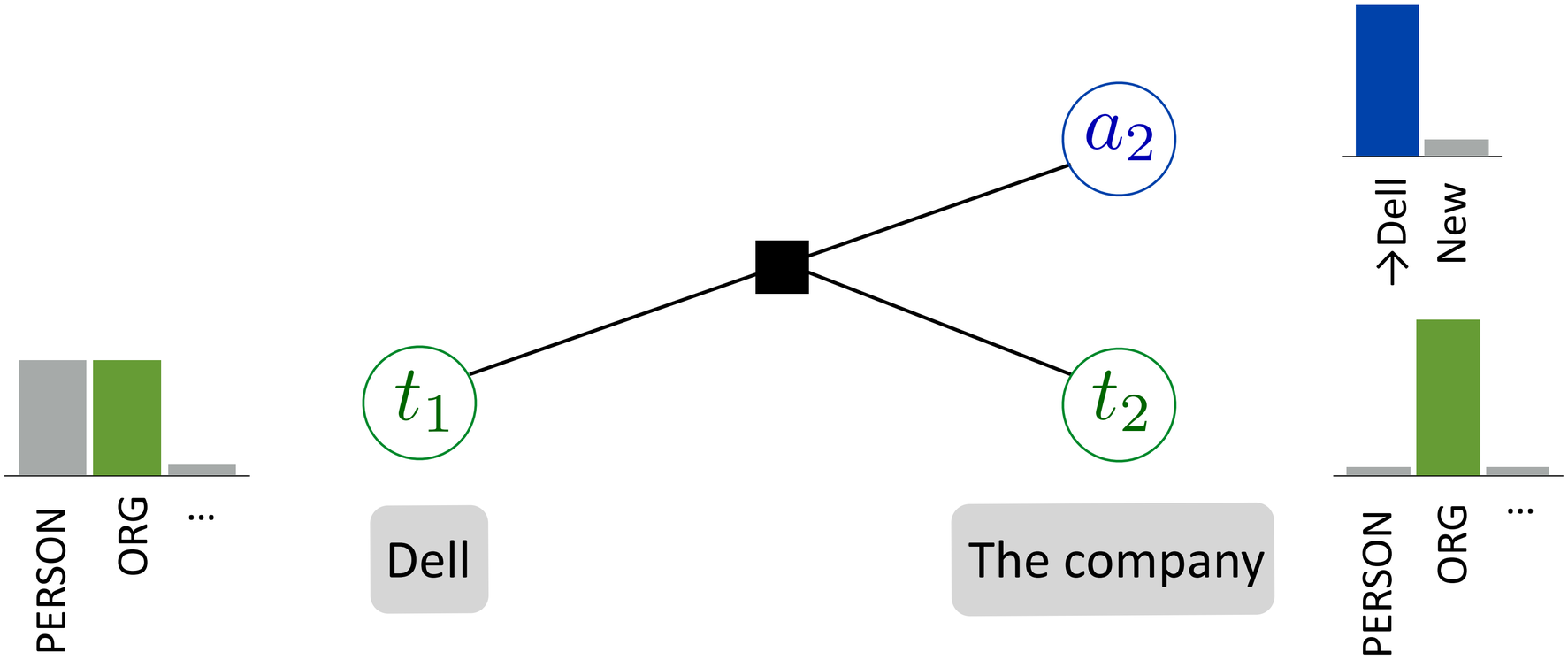
Inference: Belief Propagation



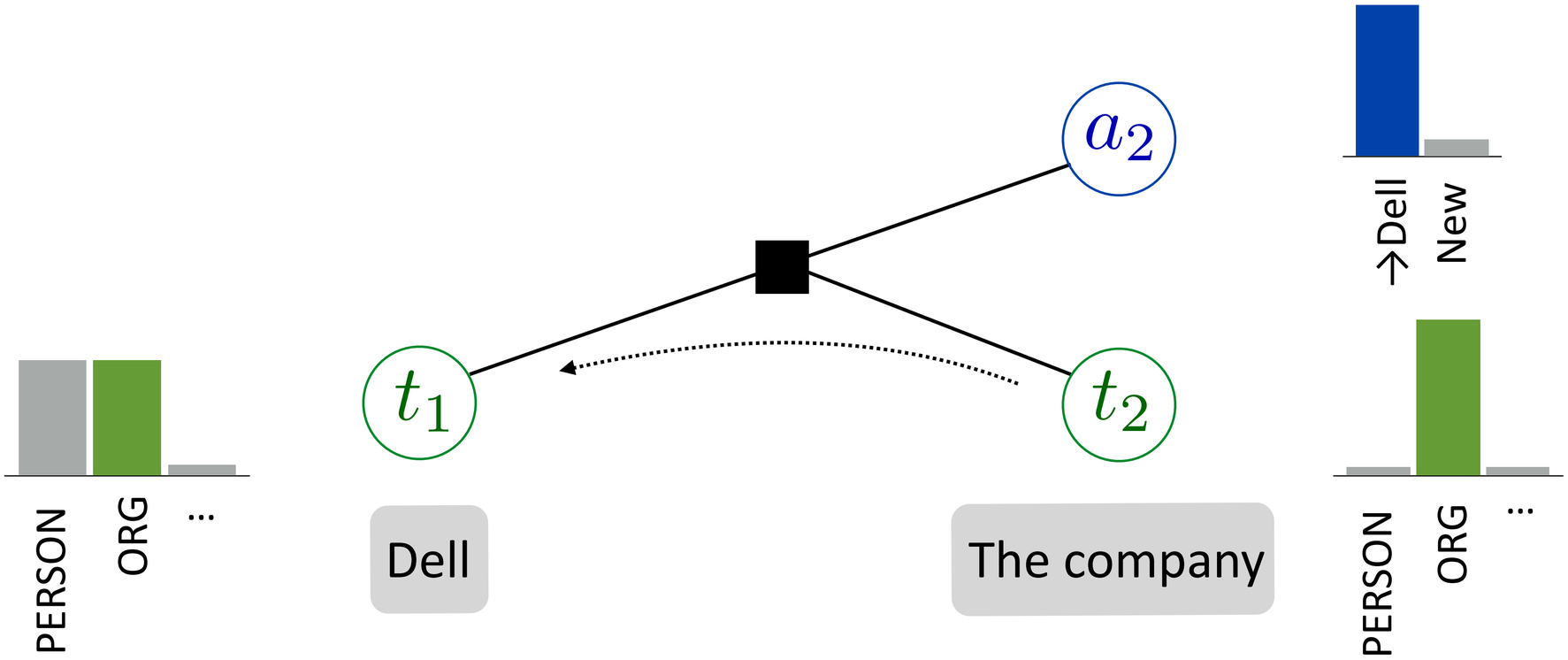
Inference: Belief Propagation



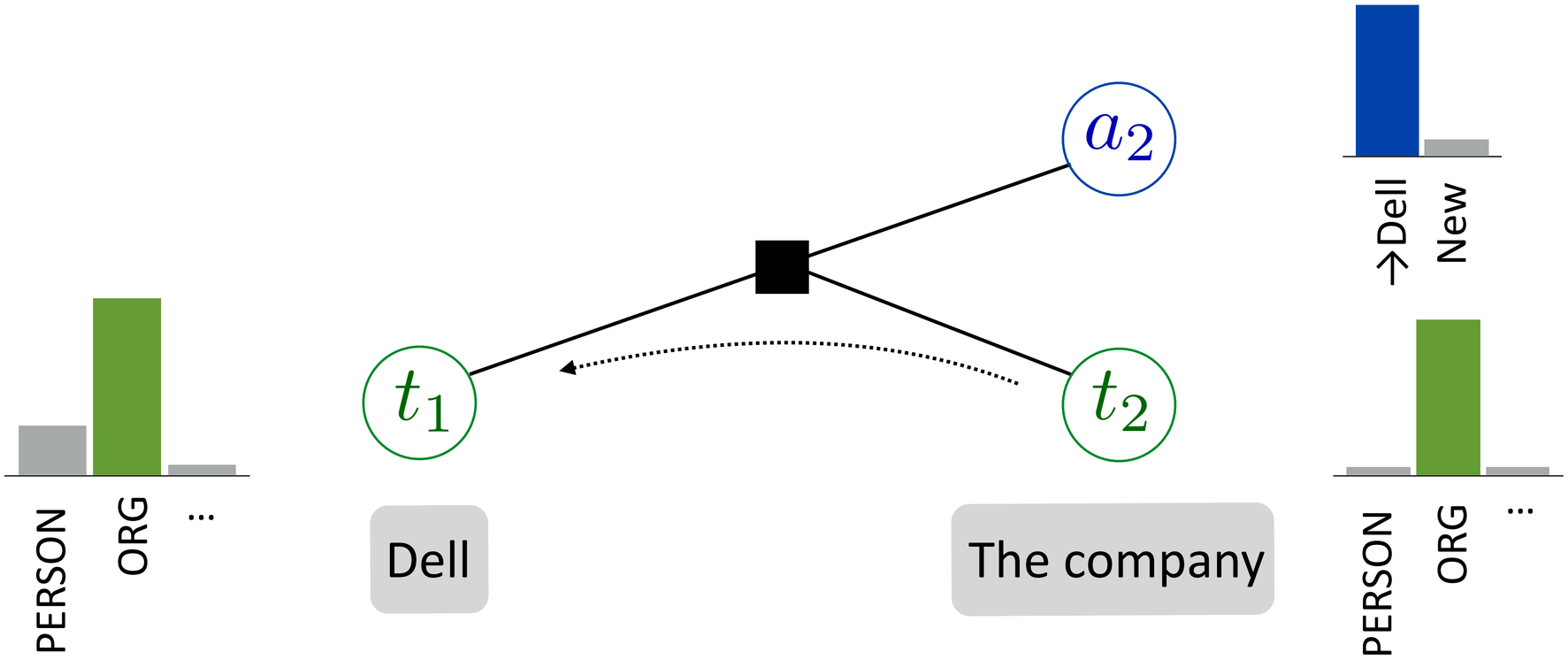
Inference: Belief Propagation



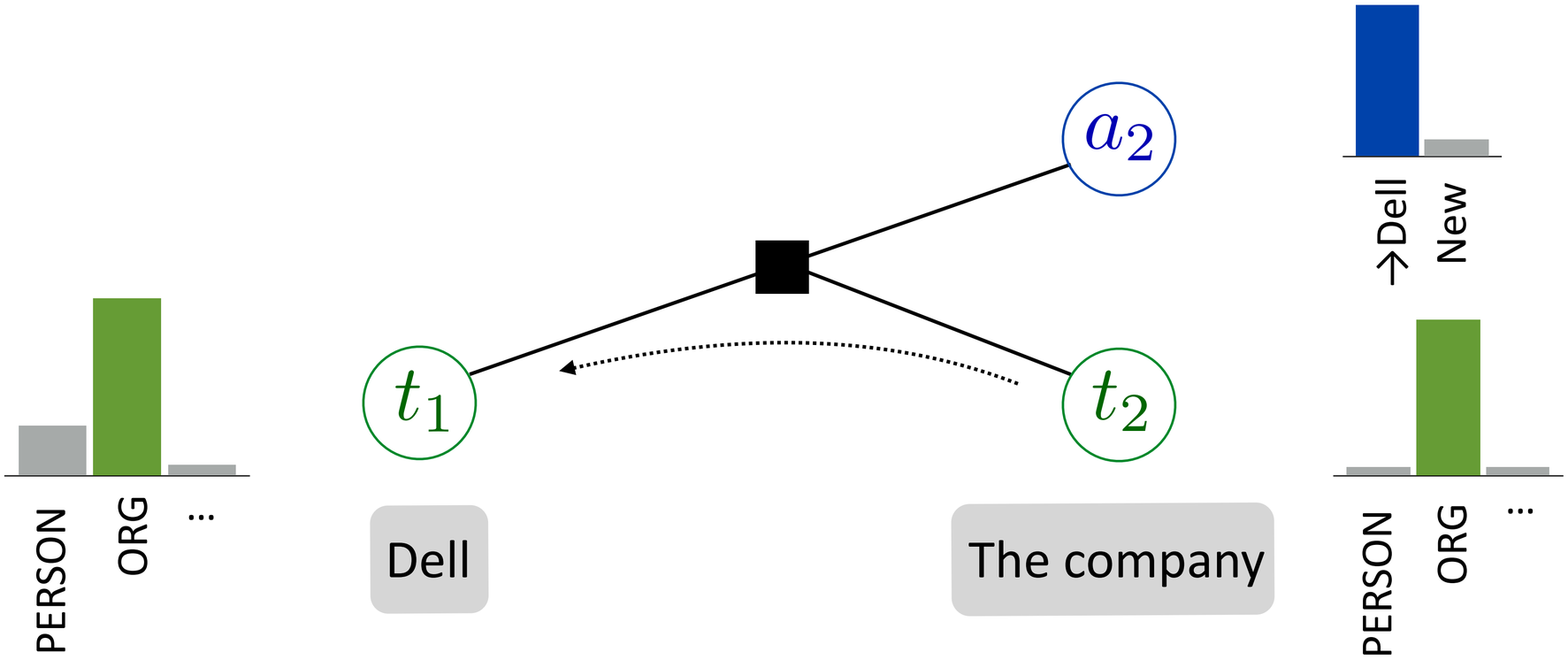
Inference: Belief Propagation



Inference: Belief Propagation

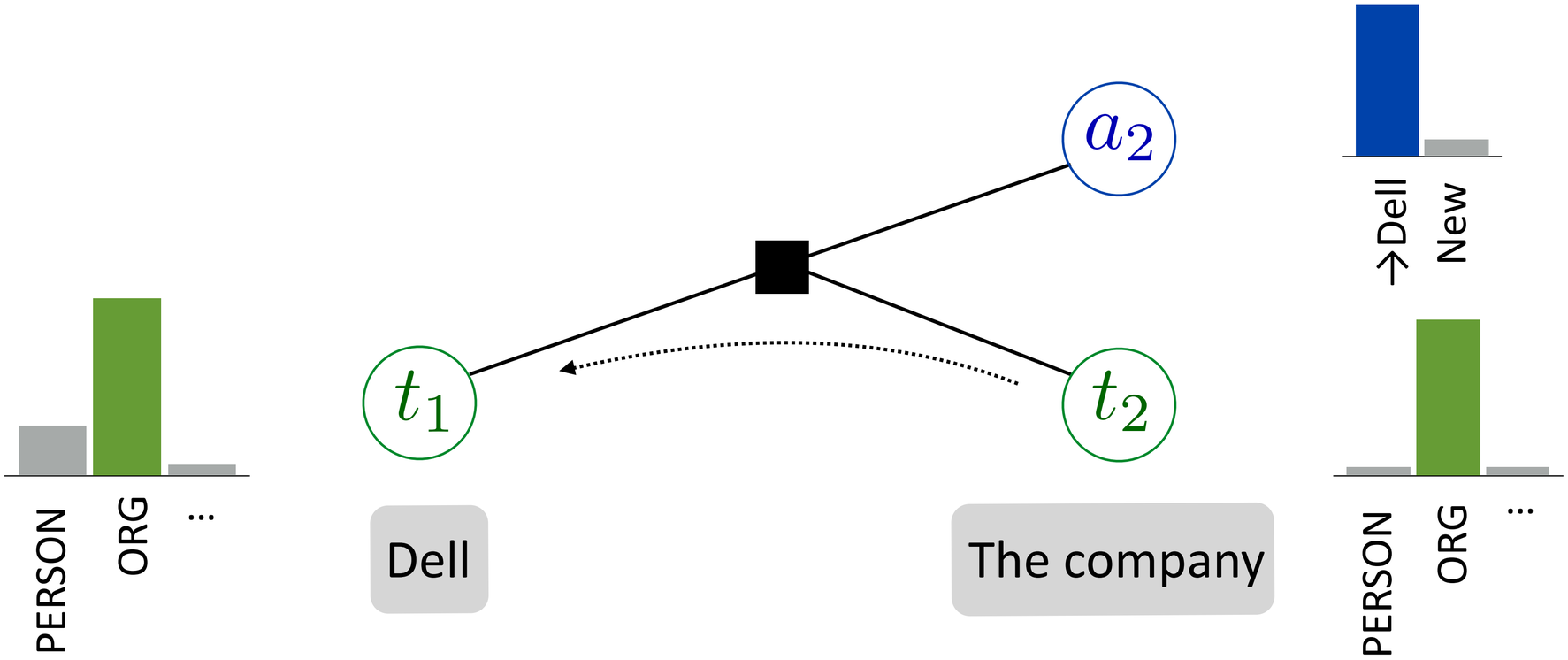


Inference: Belief Propagation



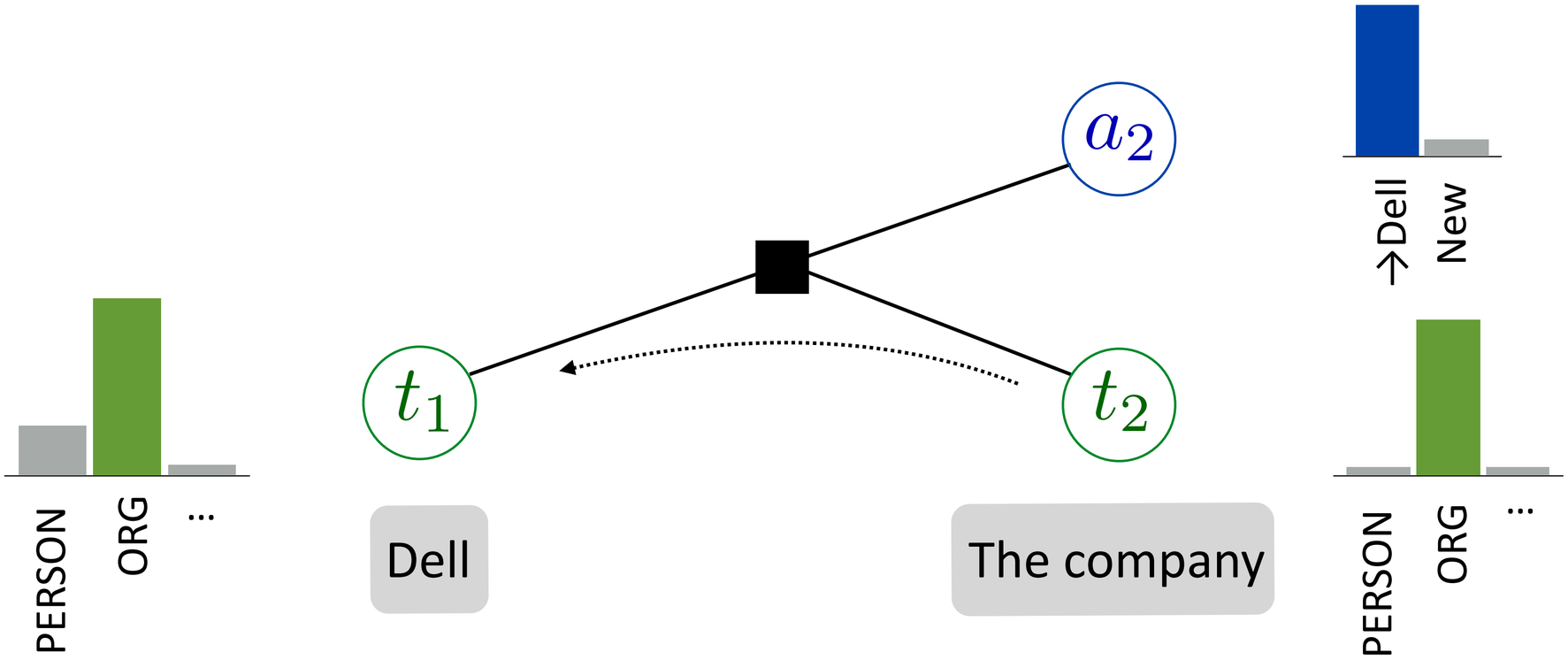
- BP encourages consensus among component models

Inference: Belief Propagation



- BP encourages consensus among component models
- Use coarse pruning to reduce the number of coreference links

Inference: Belief Propagation



- BP encourages consensus among component models
- Use coarse pruning to reduce the number of coreference links
- Only ~2x slower than individual models, ~5 seconds / document

Learning

Learning

Maximize $\mathcal{L}(\theta) = \sum_{i=1}^{d} \log \sum_{\substack{\text{latent} \\ \text{stuff}}} \mathbf{p}(\mathbf{t}_i^*, \mathbf{a}_i^*, \mathbf{e}_i^* | \mathbf{x}; \mathbf{w})$

Learning

Maximize $\mathcal{L}(\theta) = \sum_{i=1}^{d} \log \sum_{\substack{\text{latent} \\ \text{stuff}}} \mathbf{p}(\mathbf{t}_i^*, \mathbf{a}_i^*, \mathbf{e}_i^* | \mathbf{x}; \mathbf{w}) - \lambda \|\theta\|_1$

Learning

Maximize $\mathcal{L}(\theta) = \sum_{i=1}^{d} \log \sum_{\substack{\text{latent} \\ \text{stuff}}} \mathbf{p}(\mathbf{t}_i^*, \mathbf{a}_i^*, \mathbf{e}_i^* | \mathbf{x}; \mathbf{w}) - \lambda \|\theta\|_1$

$$\mathbf{p}'(\mathbf{t}_i, \mathbf{a}_i, \mathbf{e}_i | \mathbf{x}; \mathbf{w})) \propto \mathbf{p}(\mathbf{t}_i, \mathbf{a}_i, \mathbf{e}_i | \mathbf{x}; \mathbf{w}) \exp\left(\alpha_t \ell(\mathbf{t}) + \alpha_a \ell(\mathbf{a}) + \alpha_e \ell(\mathbf{e})\right)$$

task-specific loss functions

[Gimpel and Smith, 2010]

Related Work

Related Work

- Joint coreference and entity linking

Related Work

- Joint coreference and entity linking
  - Pipelined: Cheng and Roth (2013), Ponzetto and Strube (2006)

Related Work

- Joint coreference and entity linking
  - Pipelined: Cheng and Roth (2013), Ponzetto and Strube (2006)
  - Incremental inference: Hajishirzi et al. (2013), Zheng et al. (2013)

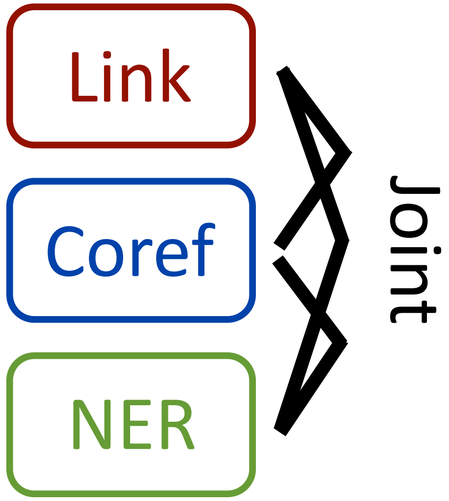
Related Work

- Joint coreference and entity linking
  - Pipelined: Cheng and Roth (2013), Ponzetto and Strube (2006)
  - Incremental inference: Hajishirzi et al. (2013), Zheng et al. (2013)

- Joint coreference, NER, and relation extraction: Singh et al (2013)

Outline

- Model
  - Independent models (unary, per-task factors)
  - Joint coupling (higher-arity, cross-task factors)
- Experiments



Data

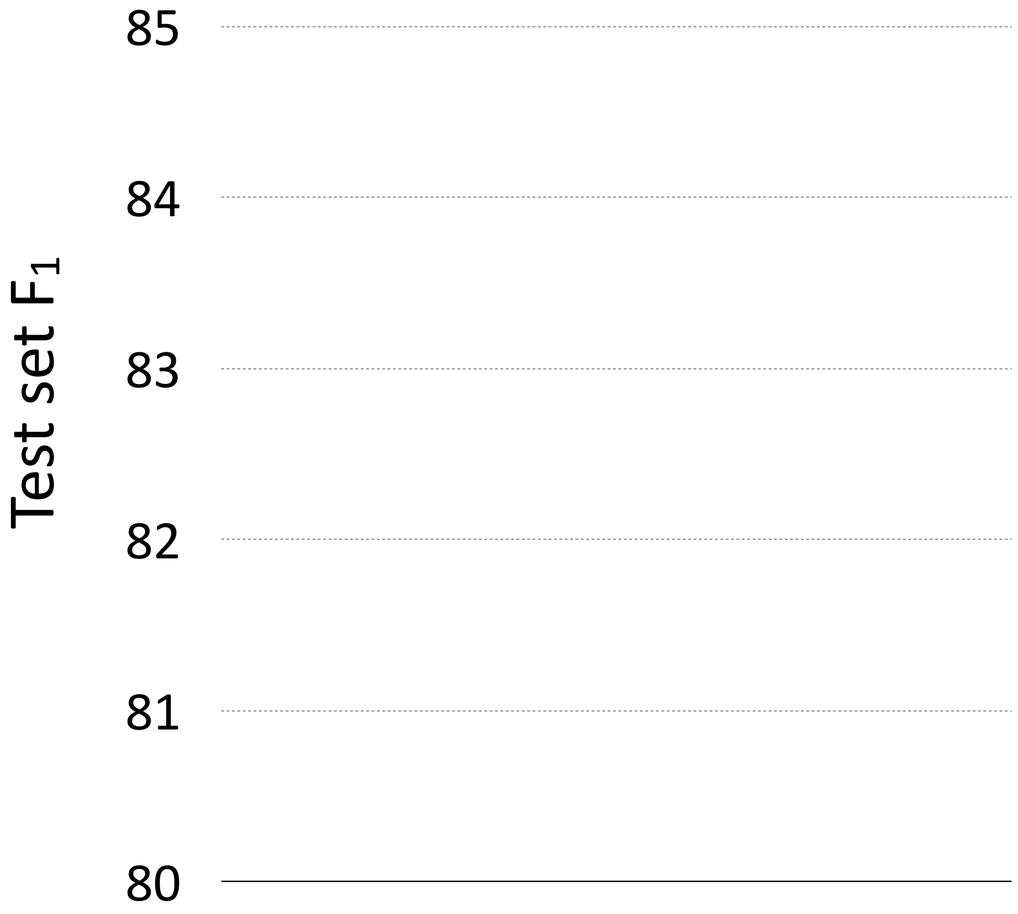
Data

- OntoNotes 5 (4000 docs): NER and coreference, no entity links

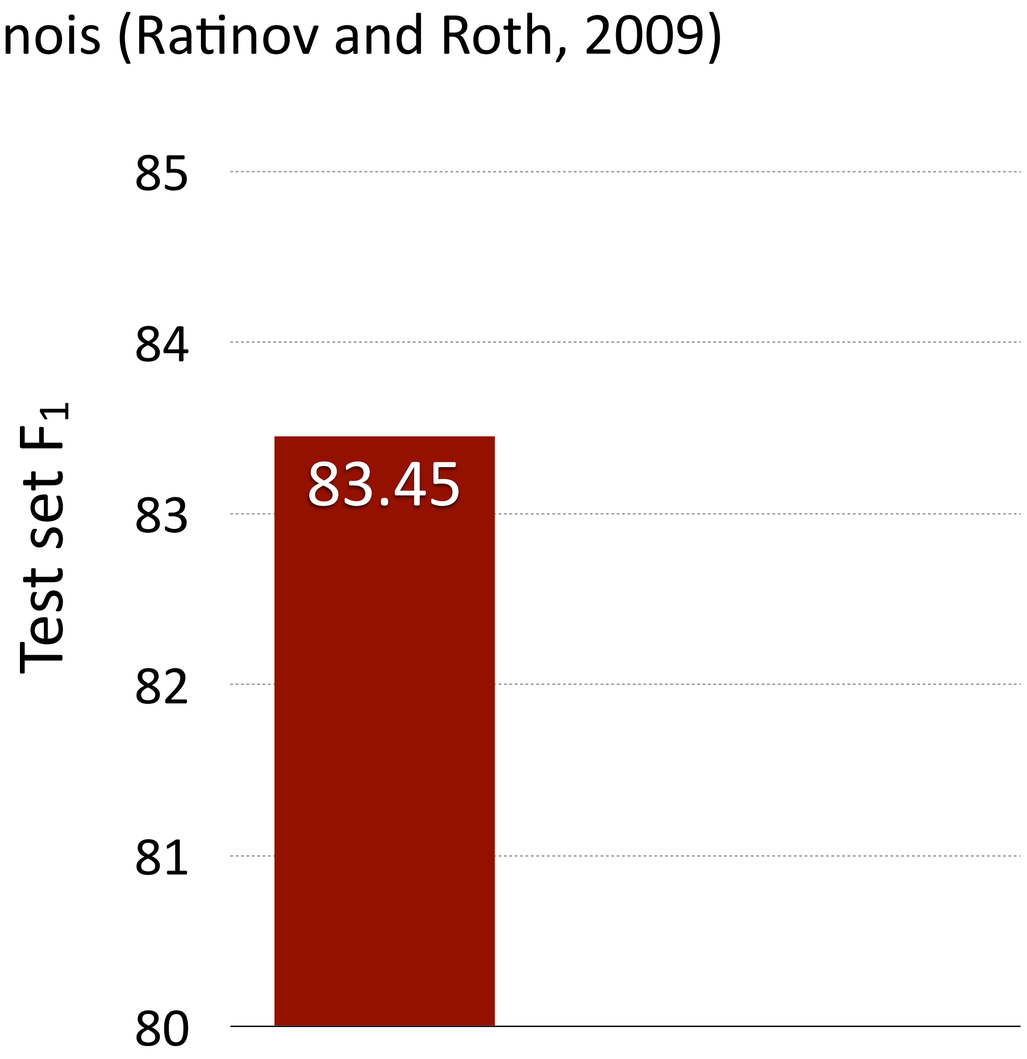
Data

- OntoNotes 5 (4000 docs): NER and coreference, no entity links
- ACE 2005 (600 docs): NER, coreference, entity links

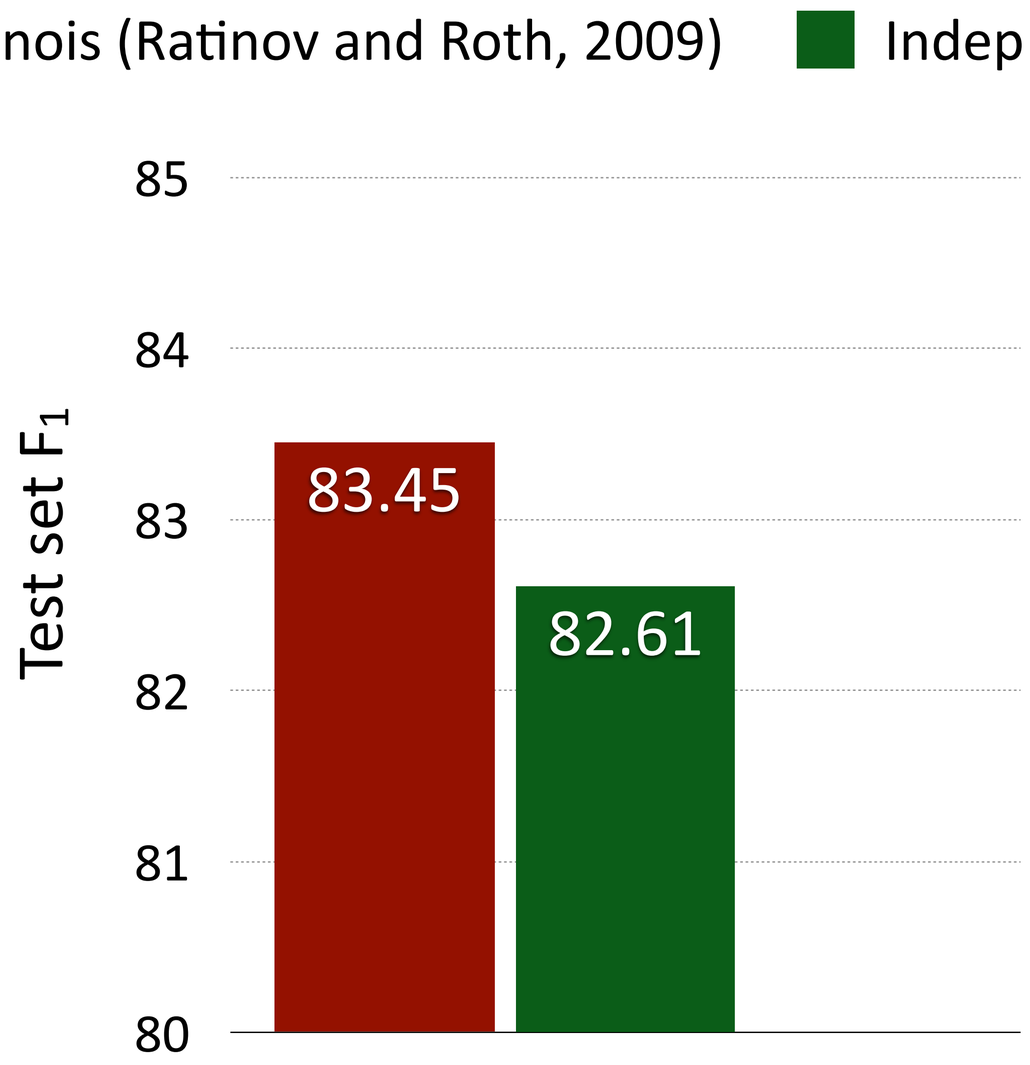
NER (OntoNotes)



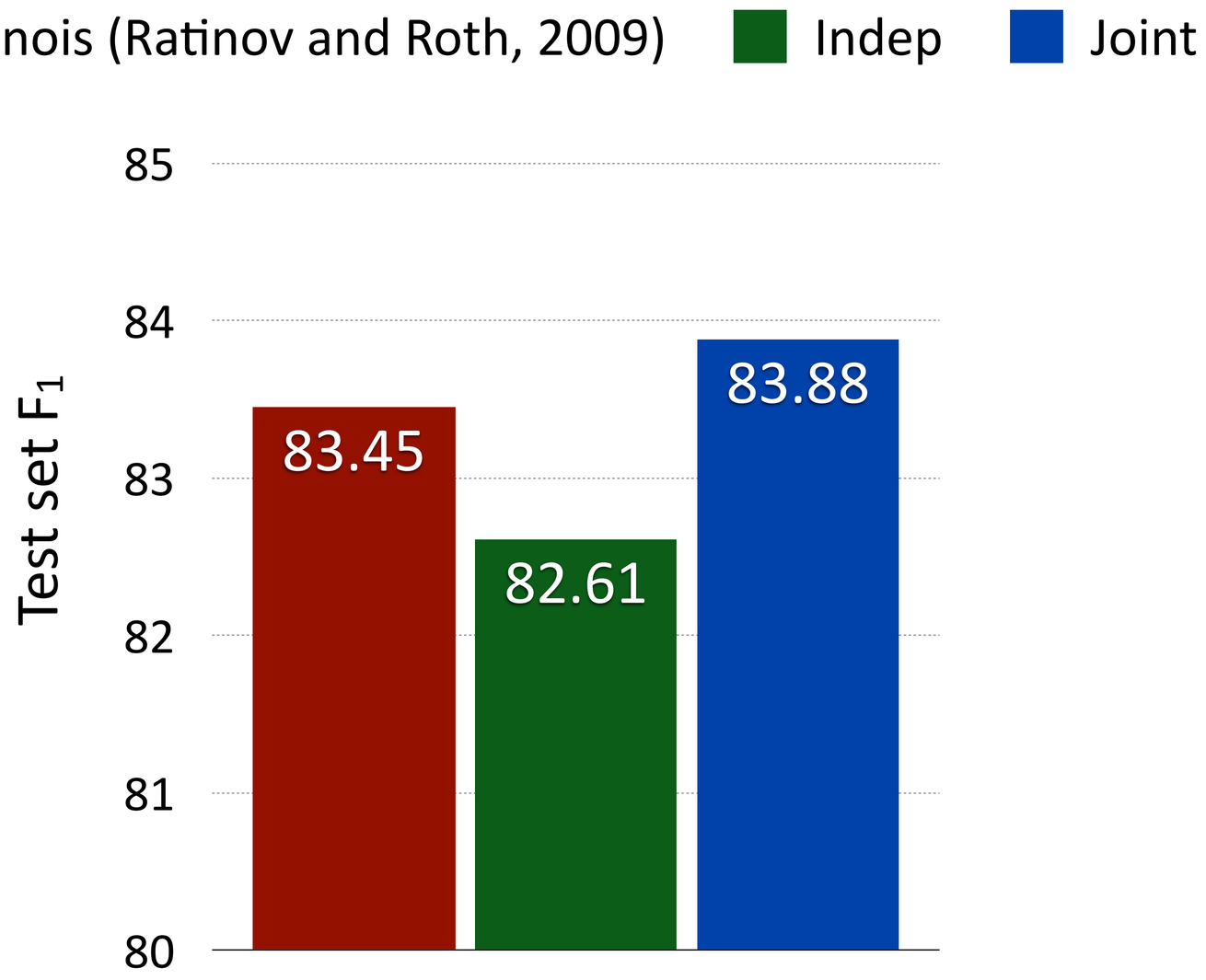
NER (OntoNotes)



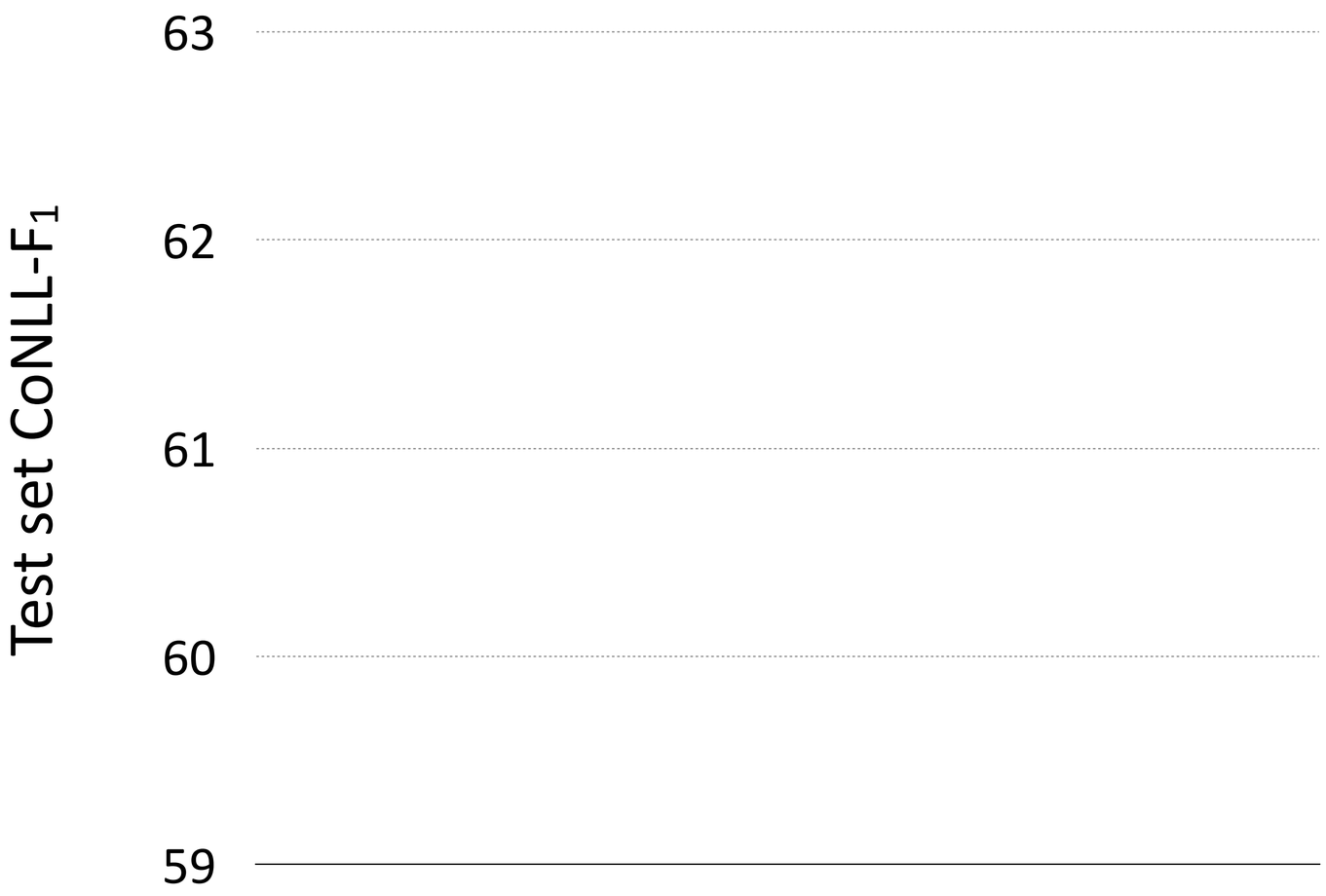
NER (OntoNotes)



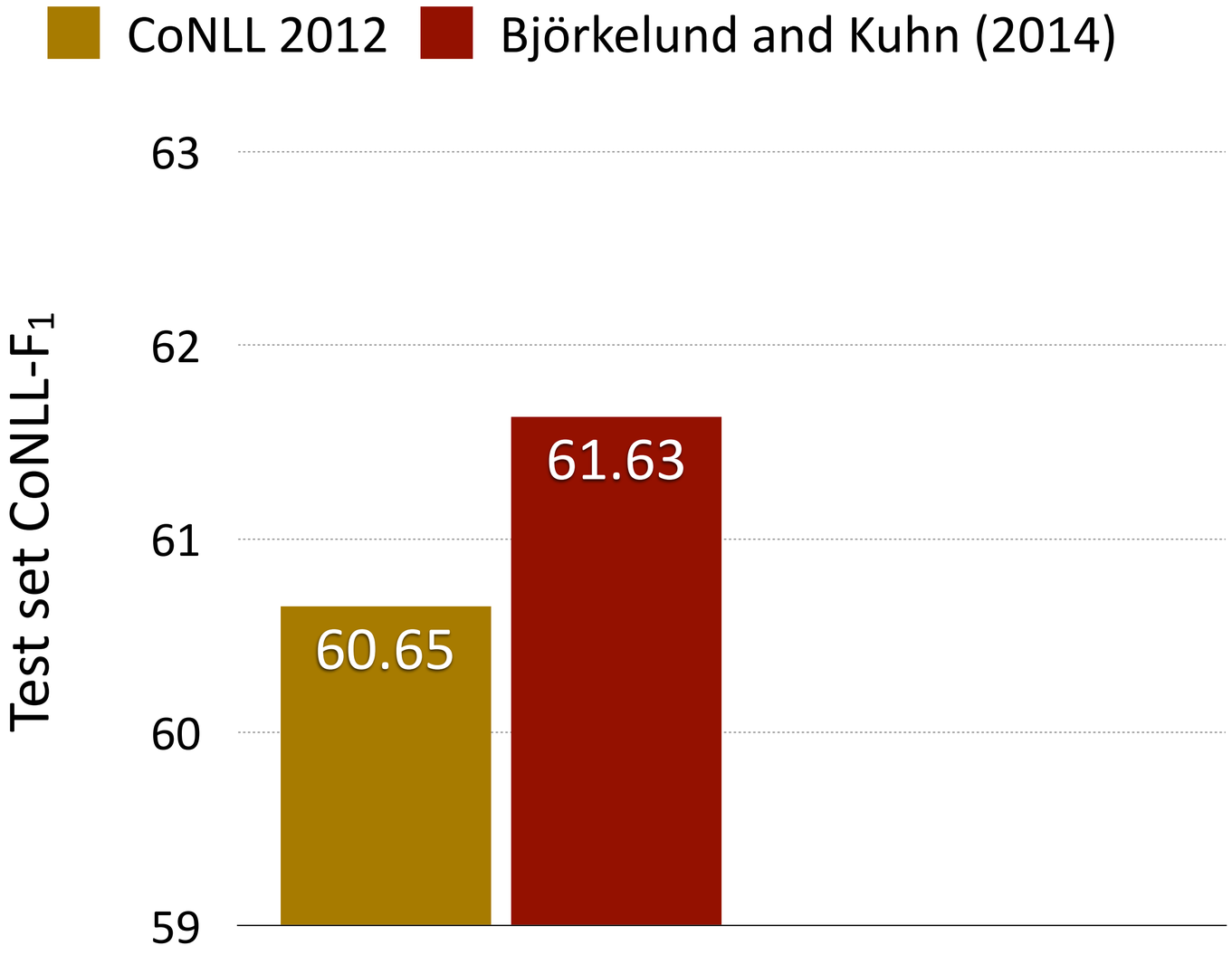
NER (OntoNotes)



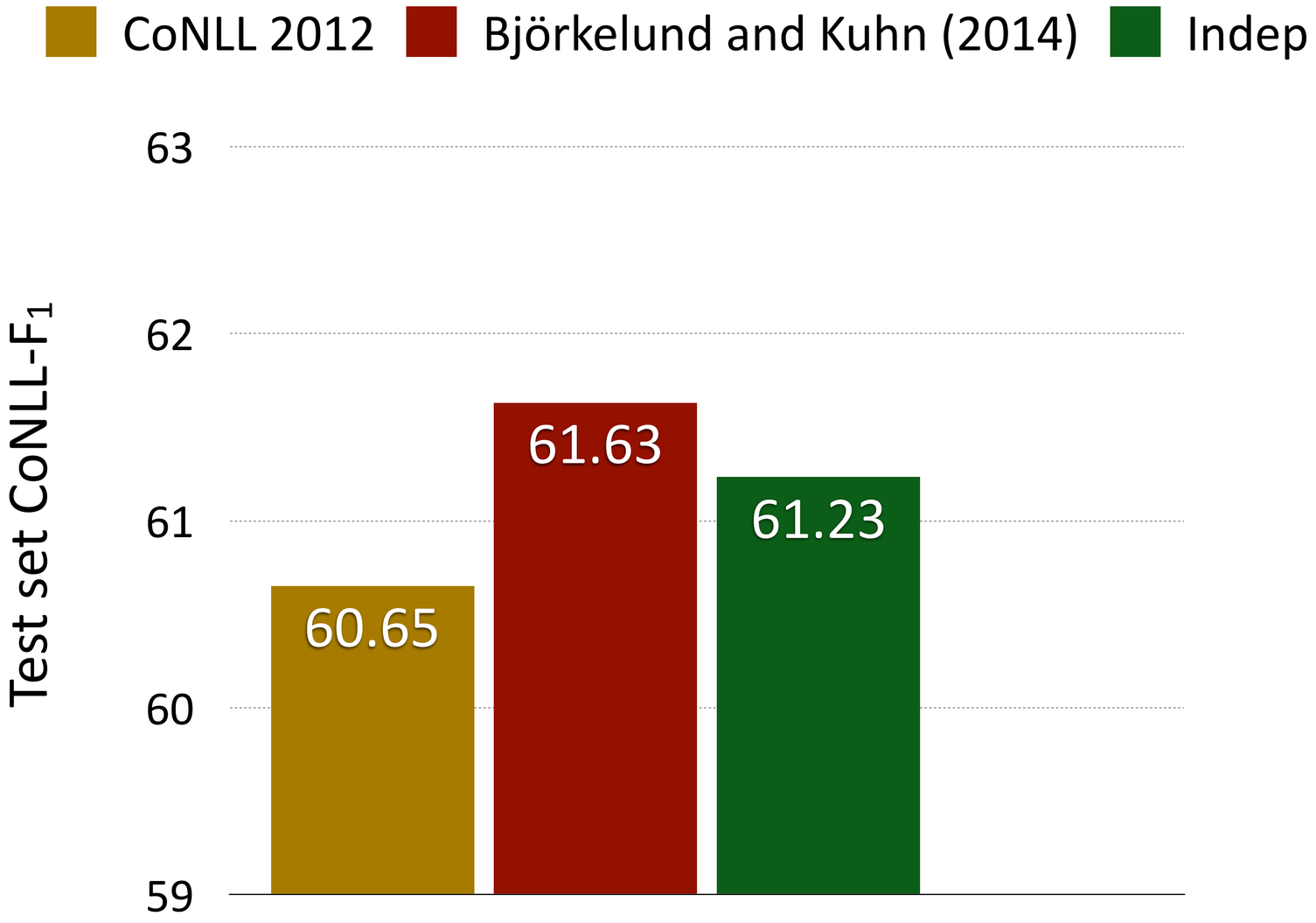
Coreference (OntoNotes)



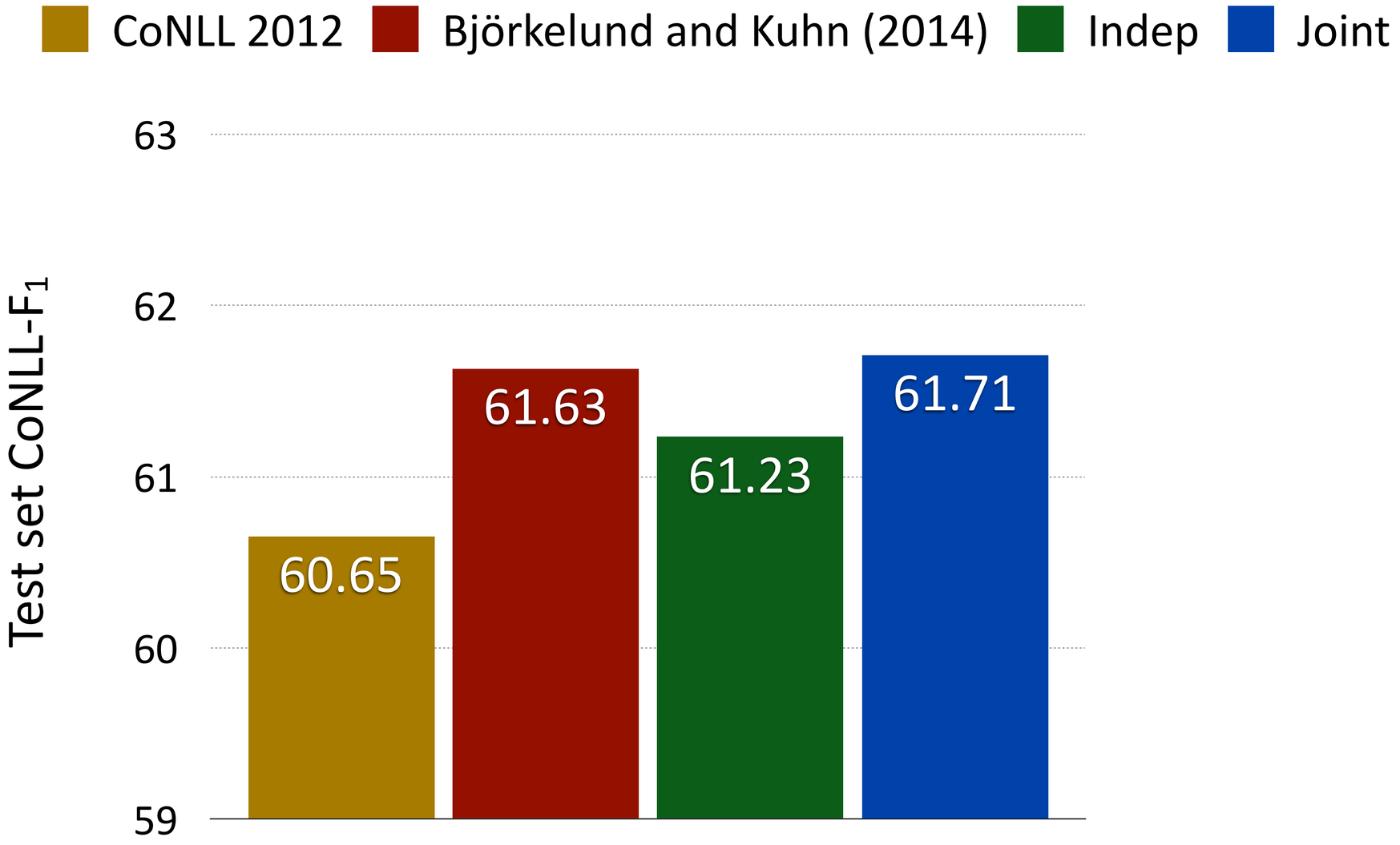
Coreference (OntoNotes)



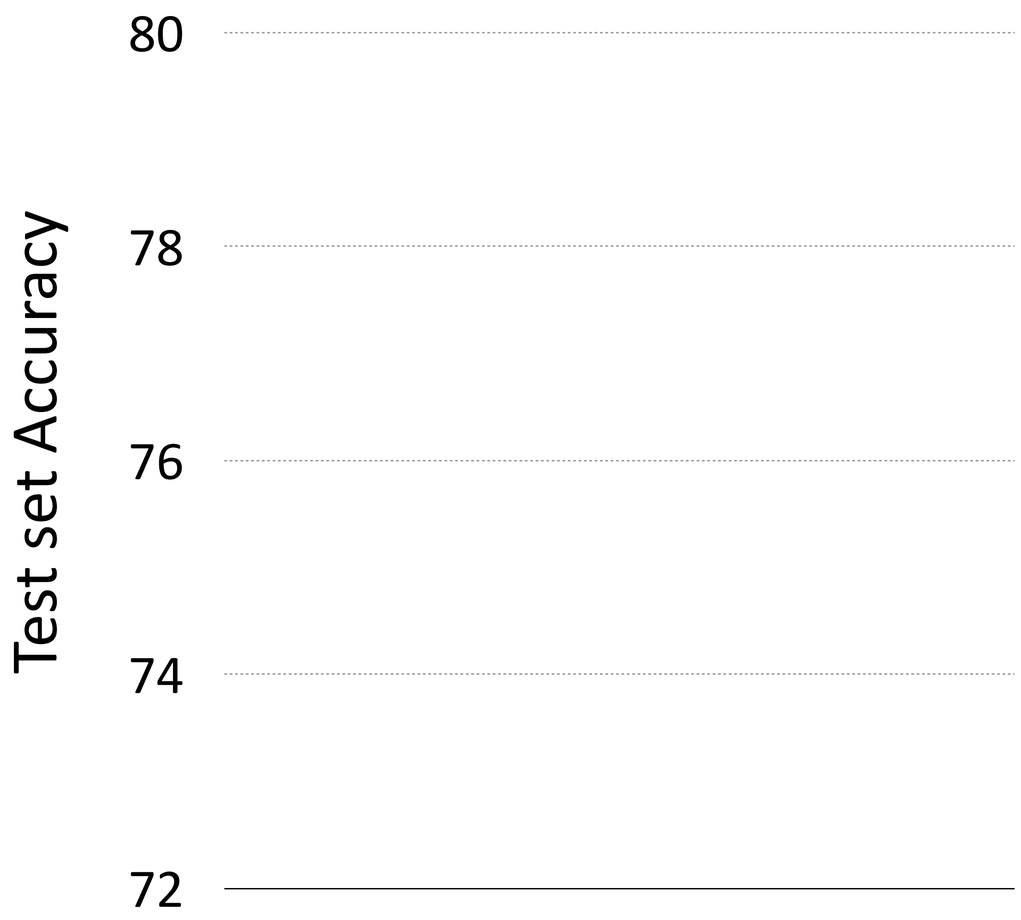
Coreference (OntoNotes)



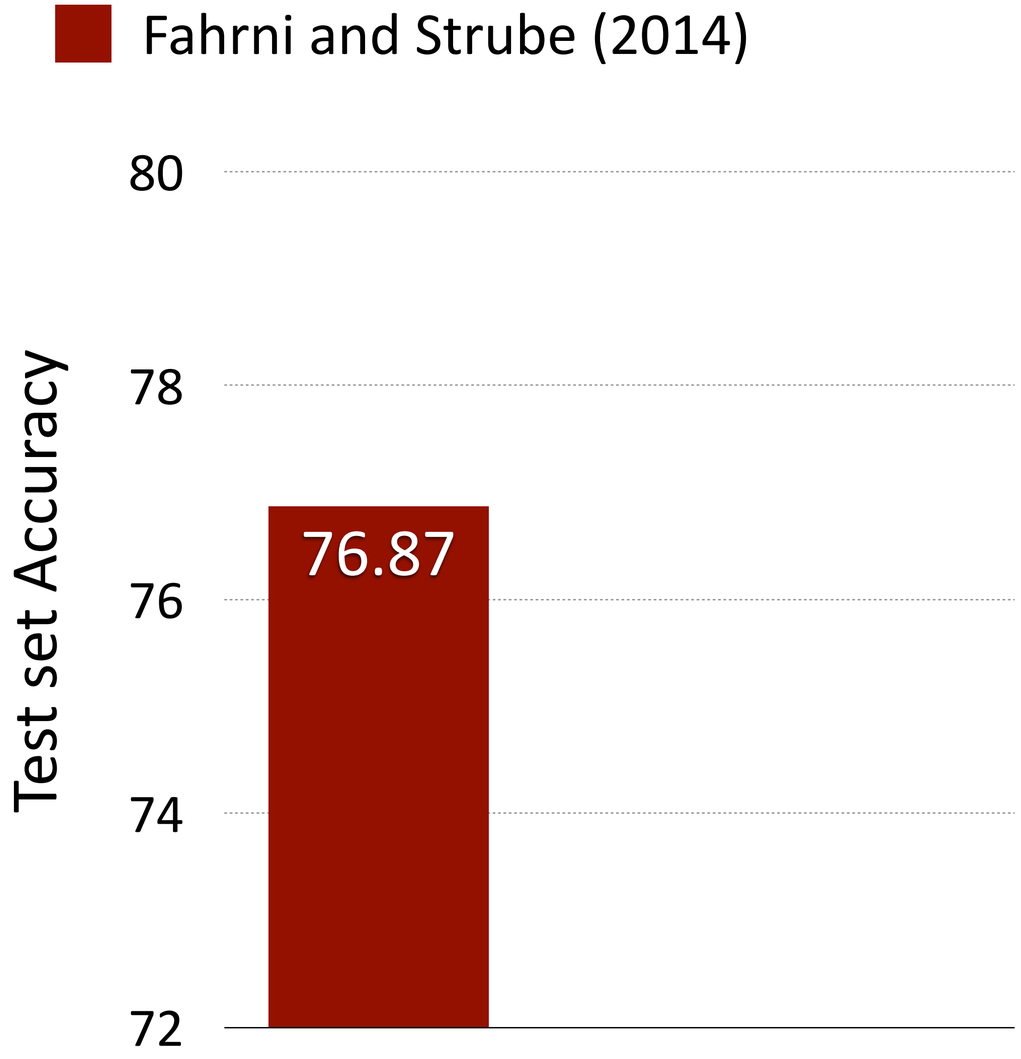
Coreference (OntoNotes)



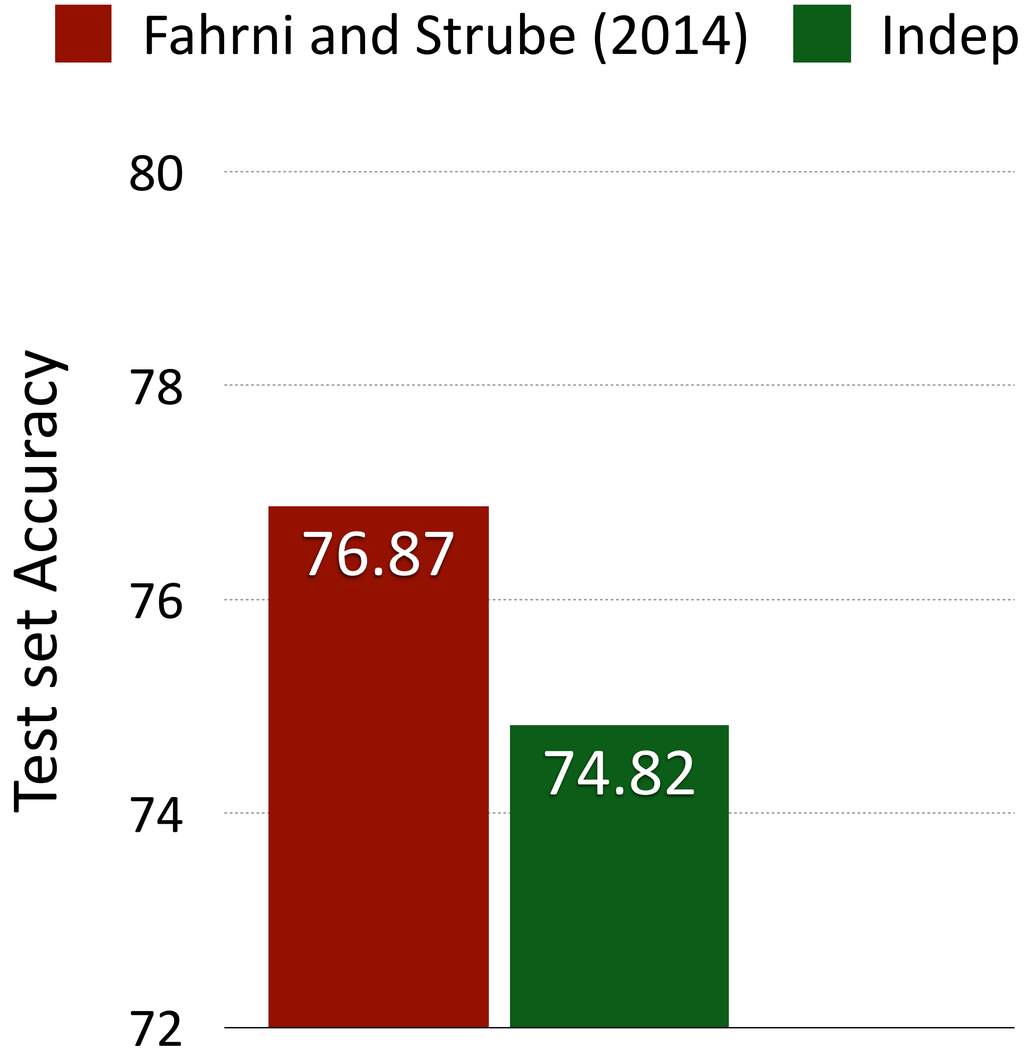
Entity Linking (ACE)



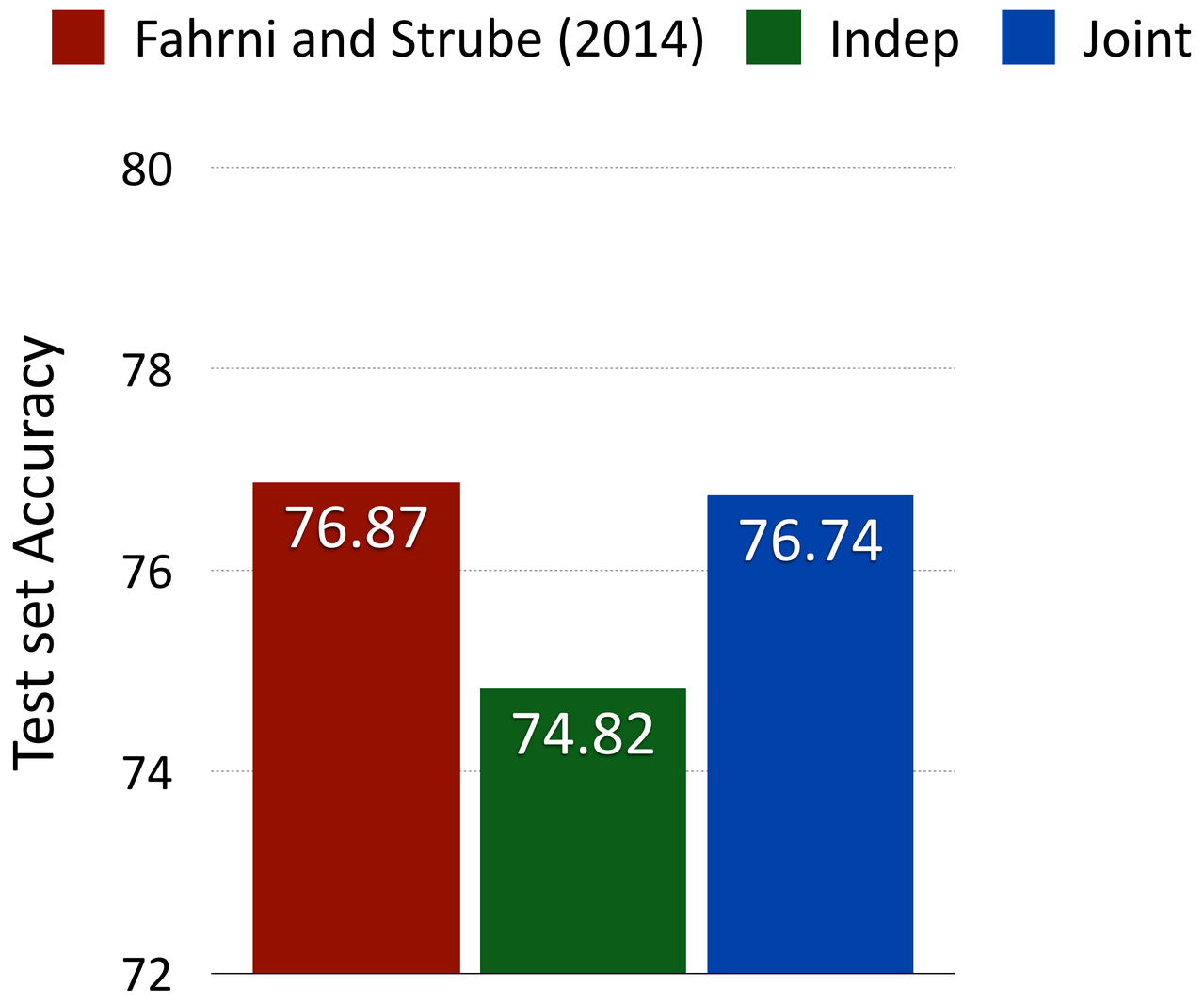
Entity Linking (ACE)



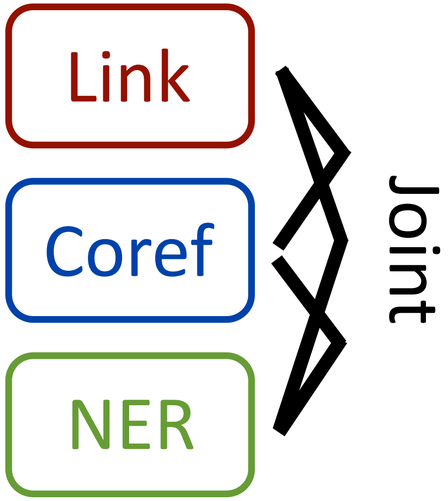
Entity Linking (ACE)



Entity Linking (ACE)

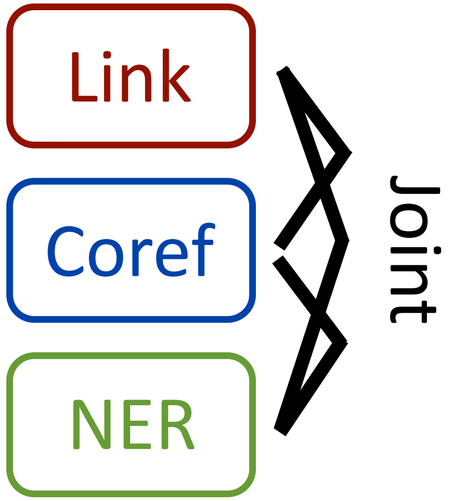


Modularity



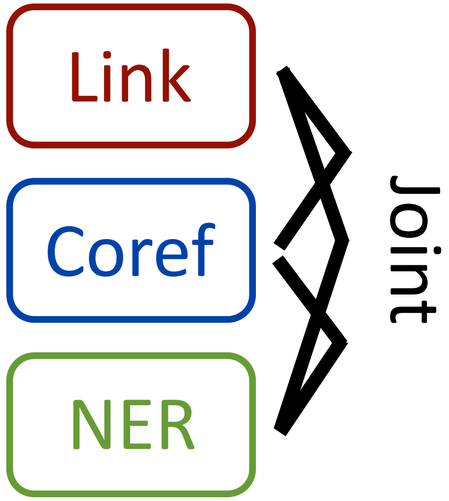
Modularity

- Component models are completely modularized, only need to expose marginals



Modularity

- Component models are completely modularized, only need to expose marginals
- Joint factors are also conceptually decoupled



Conclusion

Conclusion

- Semantic typing, coreference, and entity linking constitute three core entity analysis tasks

Conclusion

- Semantic typing, coreference, and entity linking constitute three core entity analysis tasks
- Joint modeling of the entity stack outperforms independent modeling on all three tasks

Conclusion

- Semantic typing, coreference, and entity linking constitute three core entity analysis tasks
- Joint modeling of the entity stack outperforms independent modeling on all three tasks
- Our joint model is structurally sophisticated yet modular

Conclusion

- Semantic typing, coreference, and entity linking constitute three core entity analysis tasks
- Joint modeling of the entity stack outperforms independent modeling on all three tasks
- Our joint model is structurally sophisticated yet modular
- Code available: `http://nlp.cs.berkeley.edu/projects/entity.shtml`

Conclusion

- Semantic typing, coreference, and entity linking constitute three core entity analysis tasks
- Joint modeling of the entity stack outperforms independent modeling on all three tasks
- Our joint model is structurally sophisticated yet modular
- Code available: `http://nlp.cs.berkeley.edu/projects/entity.shtml`

Thank you!

Ablations

	Coref	NER	Link
INDEP.	74.87	83.04	73.07
INDEP+LINKNER		+1.85	+2.41
INDEP+COREFNER	+0.56	+1.15	
INDEP+COREFLINK	+0.48		−0.16
JOINT−LINKNER	+0.79	+1.28	−0.06
JOINT−COREFNER	+0.56	+1.94	+2.07
JOINT−COREFLINK	+0.85	+2.68	+2.57
JOINT	+1.23	+2.90	+2.62

Chain-structured NER