

# *Generative Event Schema Induction with Entity Disambiguation*

Kiem-Hieu Nguyen<sup>1,2</sup> Xavier Tannier<sup>3,1</sup>  
Olivier Ferret<sup>2</sup> Romaric Besançon<sup>2</sup>

(1) LIMSI-CNRS

(2) CEA LIST

(3) Univ. Paris-Sud



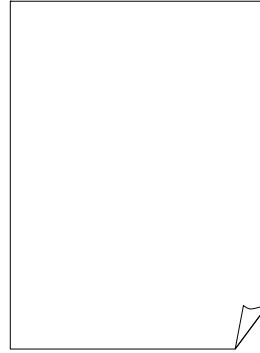
What we do?

**Task 1: Unsupervised event schema induction**

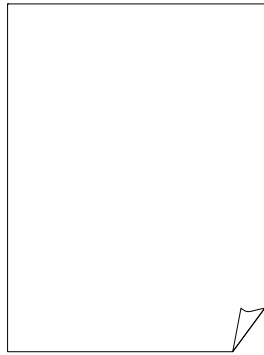
**Task 2: Slot filling**

# Task 1: Unsupervised event schema induction

## Task 2: Slot filling



Doc 1



Doc 2

...



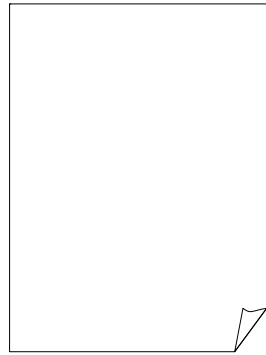
Doc 3

# Task 1: Unsupervised event schema induction

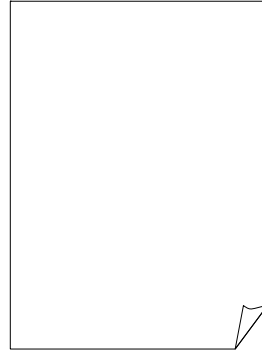
## Task 2: Slot filling

### Slot 2

*citizen  
woman  
police  
victim  
civilian*



Doc 2



Doc 1

...



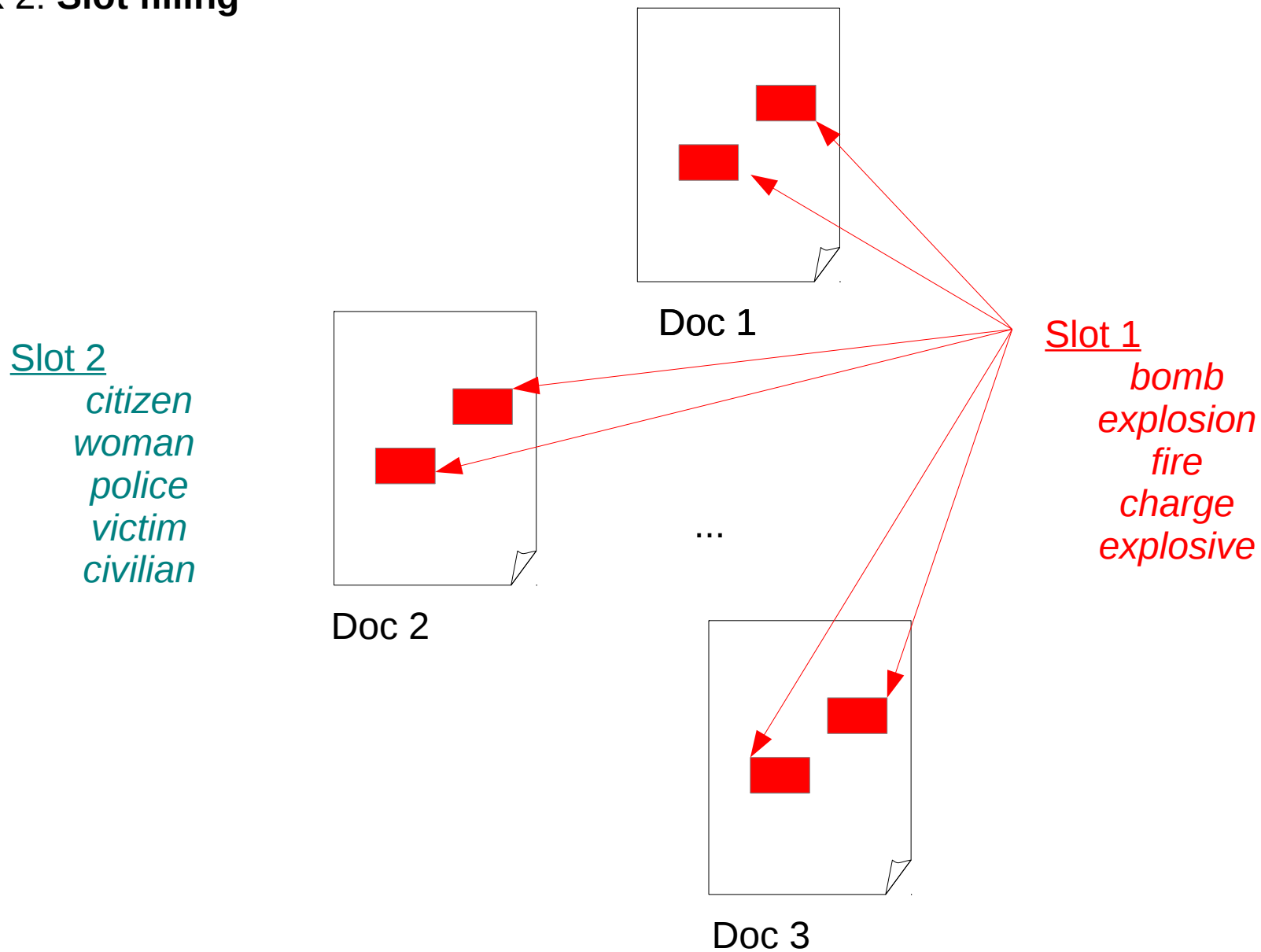
Doc 3

### Slot 1

*bomb  
explosion  
fire  
charge  
explosive*

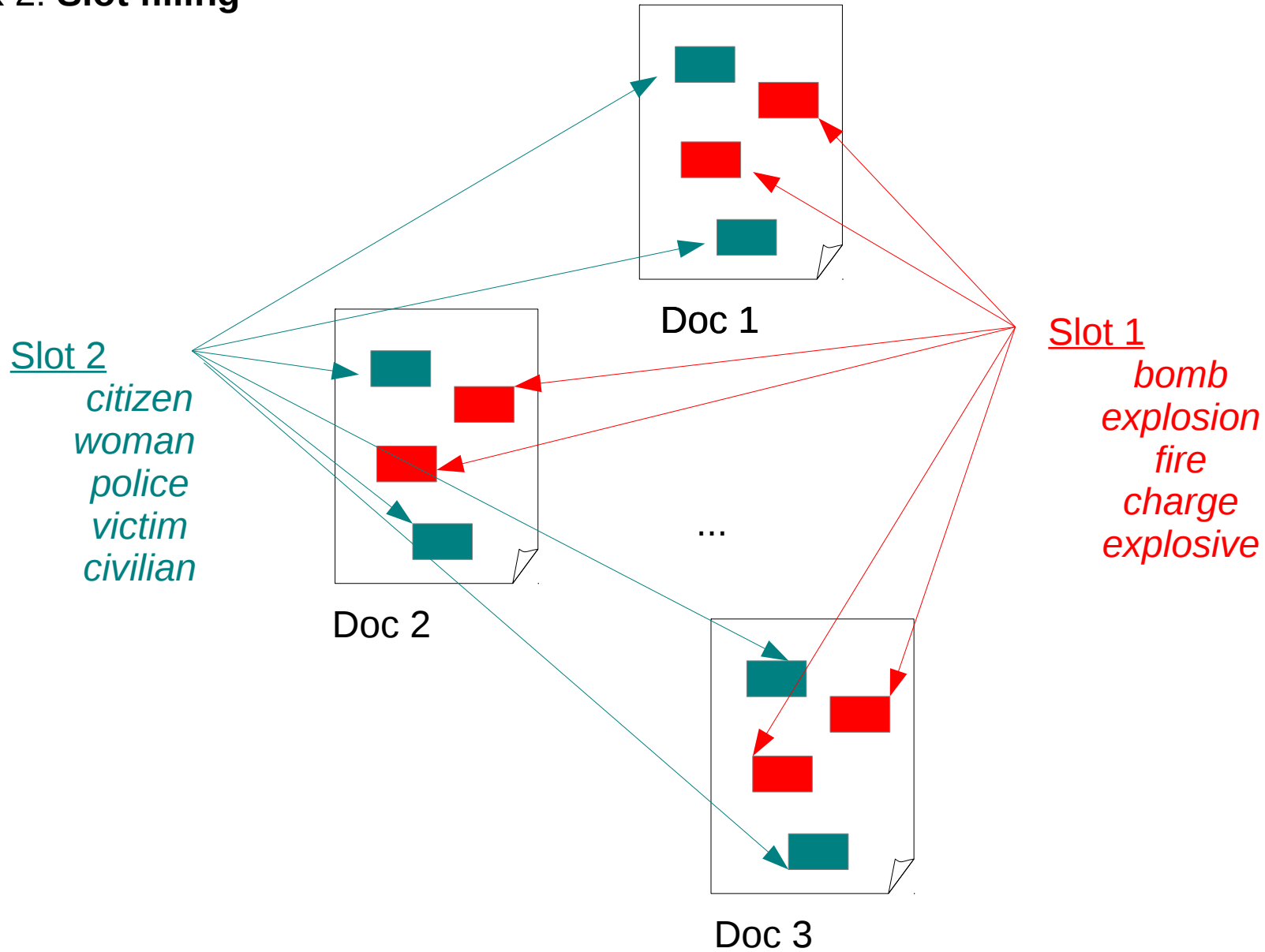
# Task 1: Unsupervised event schema induction

## Task 2: Slot filling



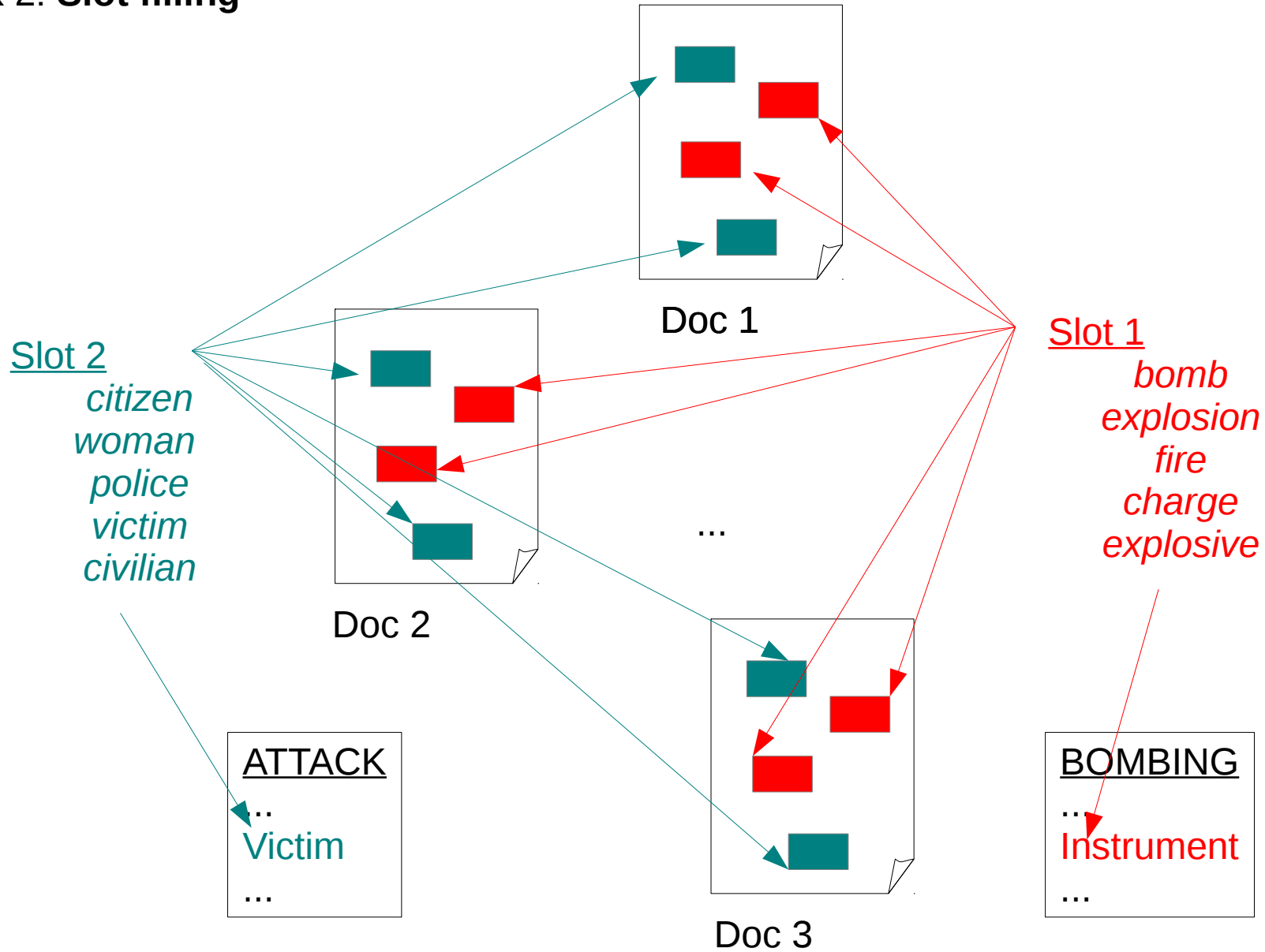
# Task 1: Unsupervised event schema induction

## Task 2: Slot filling



# Task 1: Unsupervised event schema induction

## Task 2: Slot filling





The group was traveling in a 4-wheel drive vehicle between Cucuta and the rural area known as Campanario when their vehicle was blown up by four explosive charges that exploded on the highway.

A UCR district headquarters in Buenos Aires province was "completely destroyed by a bomb explosion." The destroyed UCR headquarters is in the Moreno district of Buenos Aires.

Alleged guerrilla urban commandos launched two highpower bombs against a car dealership in downtown San Salvador this morning. A police report said that the attack set the building of fire, but did not result in any casualties although economic losses are heavy.

The terrorist attack was launched from a moving vehicle, from which unidentified individuals threw three dynamite charges that injured a passerby and damaged the front of the building, according to the police.

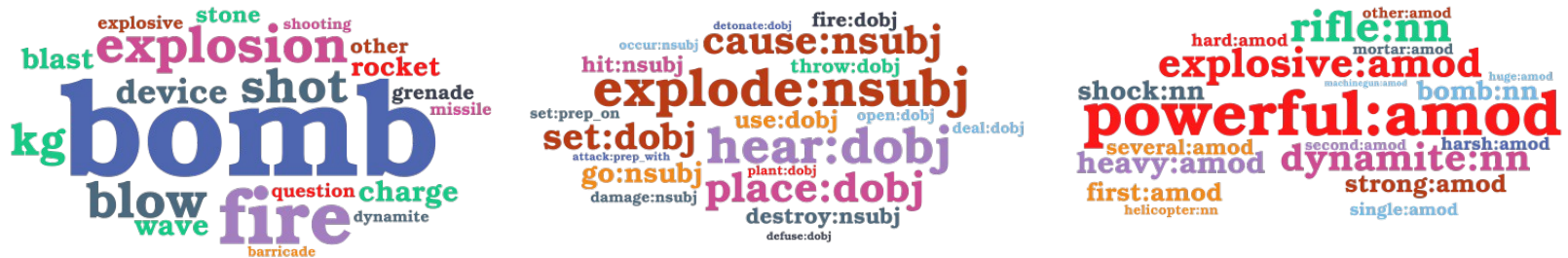
The group was traveling in a 4-wheel drive vehicle between Cucuta and the rural area known as Campanario when their vehicle was blown up by four explosive charges that exploded on the highway.

A UCR district headquarters in Buenos Aires province was "completely destroyed by a bomb explosion." The destroyed UCR headquarters is in the Moreno district of Buenos Aires.

Alleged guerrilla urban commandos launched two highpower bombs against a car dealership in downtown San Salvador this morning. A police report said that the attack set the building of fire, but did not result in any casualties although economic losses are heavy.

The terrorist attack was launched from a moving vehicle, from which unidentified individuals threw three dynamite charges that injured a passerby and damaged the front of the building, according to the police.

Slot 1



*BOMBING\_instrument*

The group was traveling in a 4-wheel drive vehicle between Cucuta and the rural area known as Campanario when their vehicle was blown up by four explosive **charges** that exploded on the highway.

A UCR district headquarters in Buenos Aires province was "completely destroyed by a bomb **explosion**." The destroyed UCR headquarters is in the Moreno district of Buenos Aires.

Alleged guerrilla urban commandos launched two highpower **bombs** against a car dealership in downtown San Salvador this morning. A police report said that the attack set the building of fire, but did not result in any casualties although economic losses are heavy.

The terrorist attack was launched from a moving vehicle, from which unidentified individuals threw three dynamite **charges** that injured a passerby and damaged the front of the building, according to the police.

Slot 1



*BOMBING\_instrument*

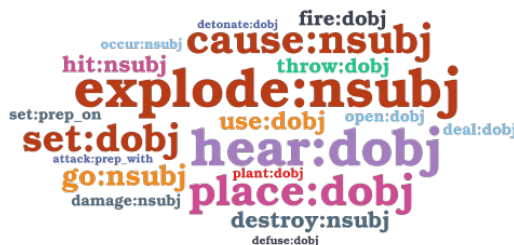
The group was traveling in a 4-wheel drive vehicle between Cucuta and the rural area known as Campanario when their vehicle was blown up by four **explosive charges** that exploded on the highway.

A UCR district headquarters in Buenos Aires province was "completely destroyed by a **bomb explosion**." The destroyed UCR headquarters is in the Moreno district of Buenos Aires.

Alleged guerrilla urban commandos launched two **highpower bombs** against a car dealership in downtown San Salvador this morning. A police report said that the attack set the building on fire, but did not result in any casualties although economic losses are heavy.

The terrorist attack was launched from a moving vehicle, from which unidentified individuals threw three **dynamite charges** that injured a passerby and damaged the front of the building, according to the police.

Slot 1



*BOMBING\_instrument*

The group was traveling in a 4-wheel drive vehicle between Cucuta and the rural area known as Campanario when their vehicle was **blown up** by four **explosive charges** that **exploded** on the highway.

A UCR district headquarters in Buenos Aires province was "completely **destroyed** by a **bomb explosion**." The destroyed UCR headquarters is in the Moreno district of Buenos Aires.

Alleged guerrilla urban commandos **launched** two **highpower bombs** against a car dealership in downtown San Salvador this morning. A police report said that the attack set the building on fire, but did not result in any casualties although economic losses are heavy.

The terrorist attack was launched from a moving vehicle, from which unidentified individuals **threw** three **dynamite charges** that **injured** a passerby and **damaged** the front of the building, according to the police.

Slot 1



*BOMBING\_instrument*

About 50 peasants of various ages have been kidnapped by FMLN in San Miguel department. According to that garrison, the mass kidnapping took place on 30 December in San Luis de la Reina.

Ricardo Alfonso Castellar, mayor of Achi, who was kidnapped on 5 January, apparently by ELN guerrillas, was found dead today.

The MNR reports the disappearance and kidnapping of MNR assistant secretary general Hector Oqueli Colindres in Guatemala city today, 12 January.

The Salvadoran government today deplored the disappearance of Social Democratic leader Hector Oqueli Colindres this morning in Guatemala.



**KIDNAPPING\_victim**

About 50 **peasants** of various ages have been kidnapped by FMLN in San Miguel department. According to that garrison, the mass kidnapping took place on 30 December in San Luis de la Reina.

**Ricardo Alfonso Castellar**, mayor of Achi, who was kidnapped on 5 January, apparently by ELN guerrillas, was found dead today.

The MNR reports the disappearance and kidnapping of MNR assistant secretary general **Hector Oqueli Colindres** in Guatemala city today, 12 January.

The Salvadoran government today deplored the disappearance of Social Democratic leader **Hector Oqueli Colindres** this morning in Guatemala.



**KIDNAPPING\_victim**

About 50 **peasants** of various ages have been kidnapped by **FMLN** in San Miguel department. According to that garrison, the mass kidnapping took place on 30 December in San Luis de la Reina.

**Ricardo Alfonso Castellar**, mayor of Achi, who was kidnapped on 5 January, apparently by ELN **guerrillas**, was found dead today.

The MNR reports the disappearance and kidnapping of MNR assistant secretary general **Hector Oqueli Colindres** in Guatemala city today, 12 January.

The Salvadoran government today deplored the disappearance of Social Democratic leader **Hector Oqueli Colindres** this morning in Guatemala.



**KIDNAPPING\_victim**



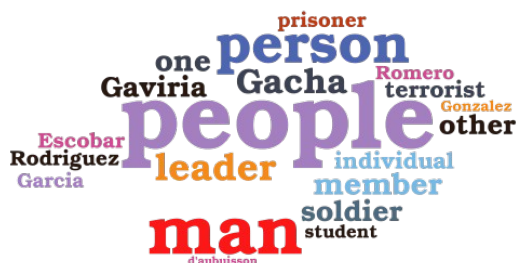
About 50 **peasants** of various ages have been kidnapped by **FMLN** in San Miguel **department**. According to that garrison, the mass kidnapping took place on 30 December in **San Luis de la Reina**.

**Ricardo Alfonso Castellar**, mayor of Achi, who was kidnapped on 5 January, apparently by ELN **guerrillas**, was found dead today.

The MNR reports the disappearance and kidnapping of MNR assistant secretary general **Hector Oqueli Colindres** in Guatemala **city** today, 12 January.

The Salvadoran government today deplored the disappearance of Social Democratic leader **Hector Oqueli Colindres** this morning in **Guatemala**.

**Slot 2**



**KIDNAPPING\_victim**

About 50 **peasants** of various ages have been kidnapped by **FMLN** in San Miguel **department**. According to that garrison, the mass kidnapping took place on **30 December** in **San Luis de la Reina**.

**Ricardo Alfonso Castellar**, mayor of Achi, who was kidnapped on **5 January**, apparently by ELN **guerrillas**, was found dead today.

The MNR reports the disappearance and kidnapping of MNR assistant secretary general **Hector Oqueli Colindres** in Guatemala **city** today, **12 January**.

The Salvadoran government today deplored the disappearance of Social Democratic leader **Hector Oqueli Colindres** this **morning** in **Guatemala**.



**KIDNAPPING\_victim**

How we do it ?  
(what **new** ?)

PREPROCESSING

Documents



Entity  
Representation

**PREPROCESSING**

Documents

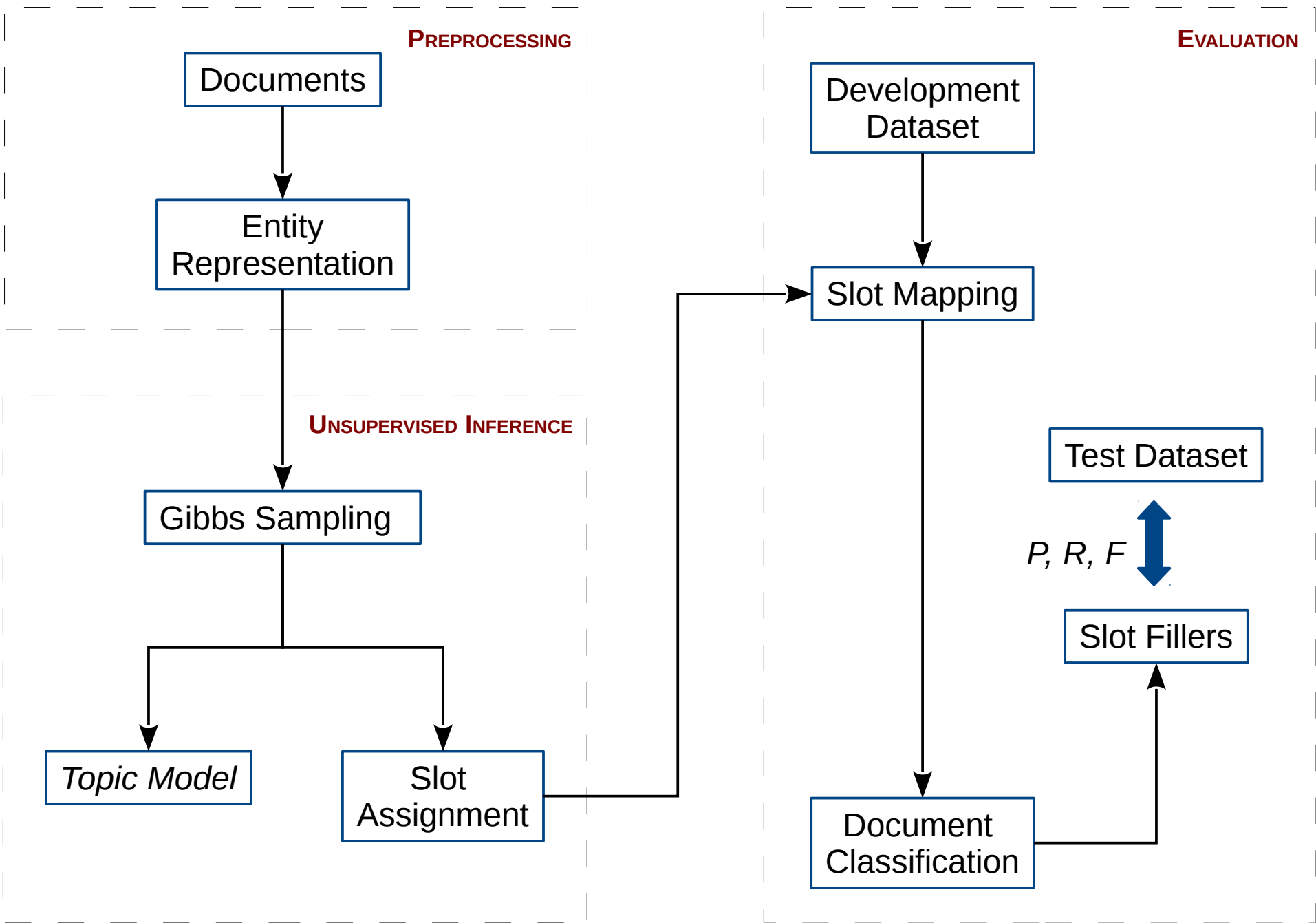
Entity  
Representation

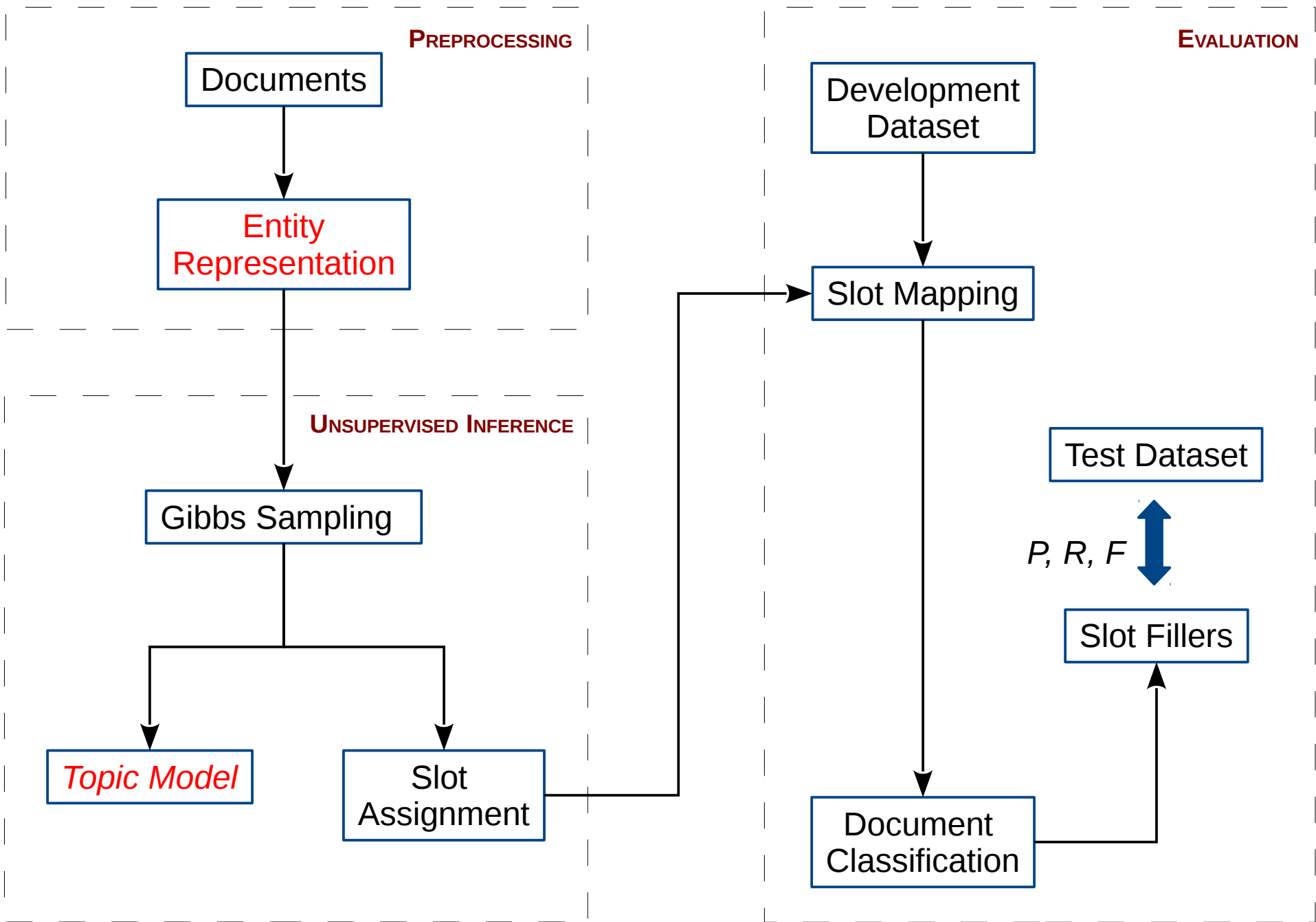
**UNSUPERVISED INFERENCE**

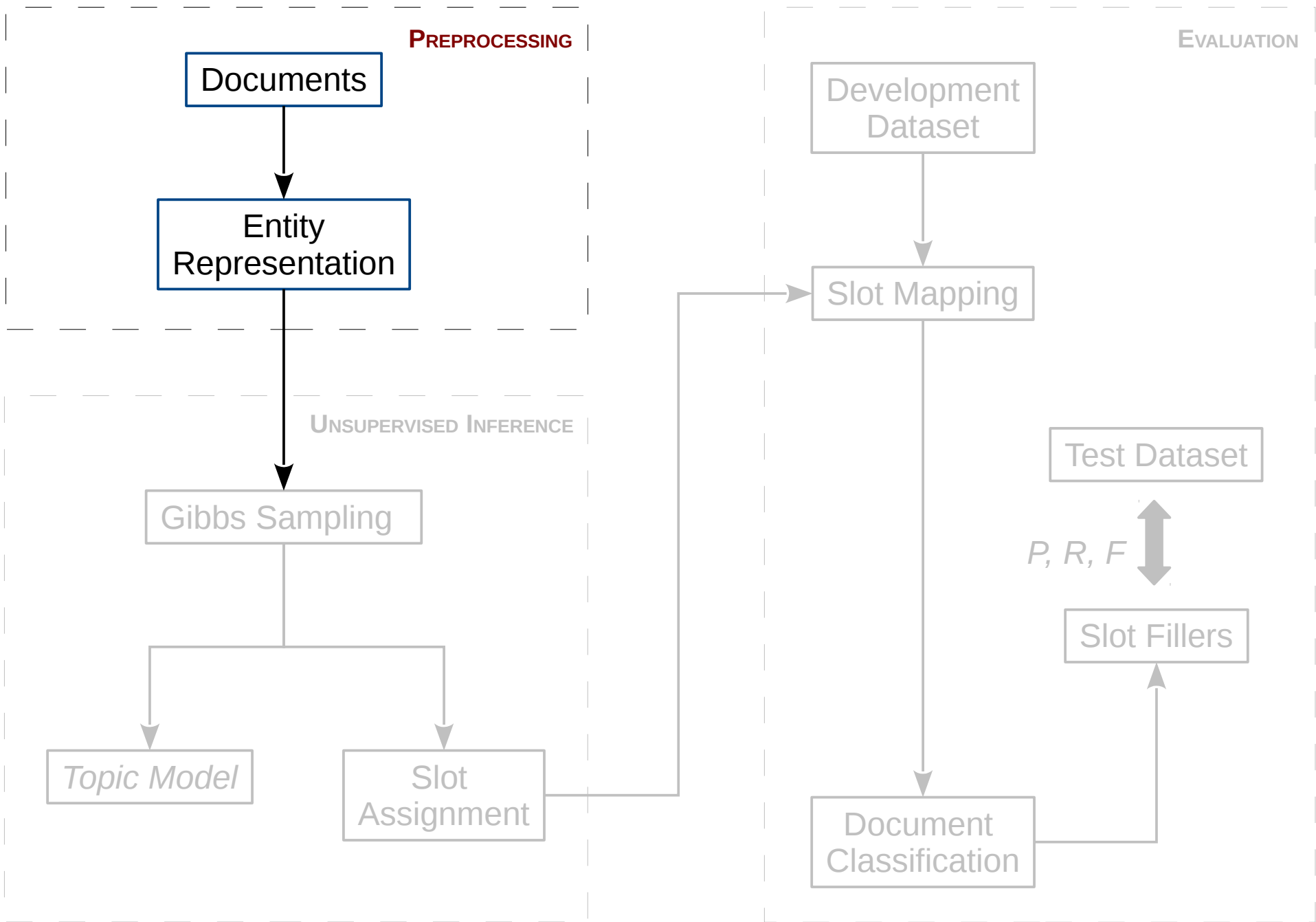
Gibbs Sampling

*Topic Model*

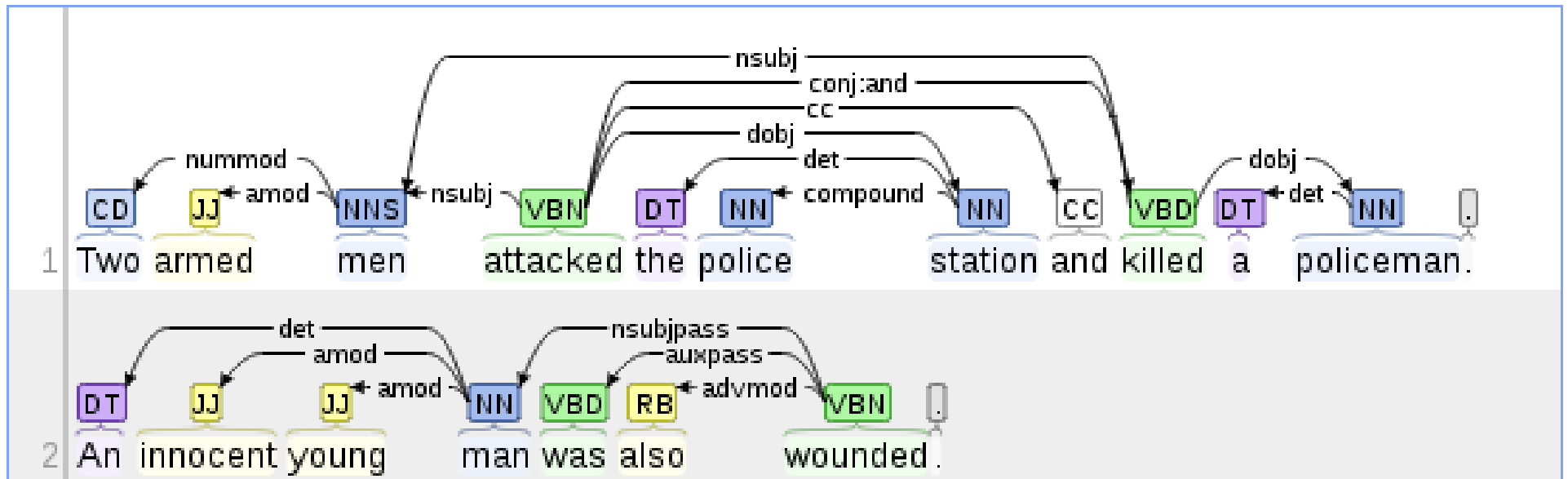
Slot  
Assignment

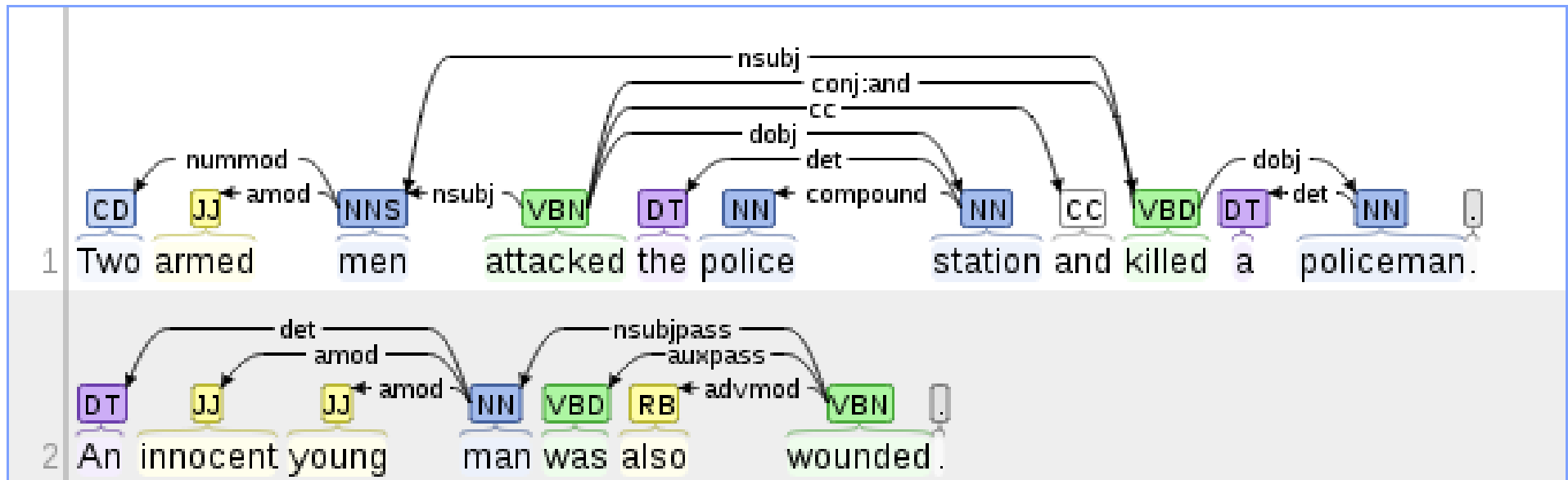








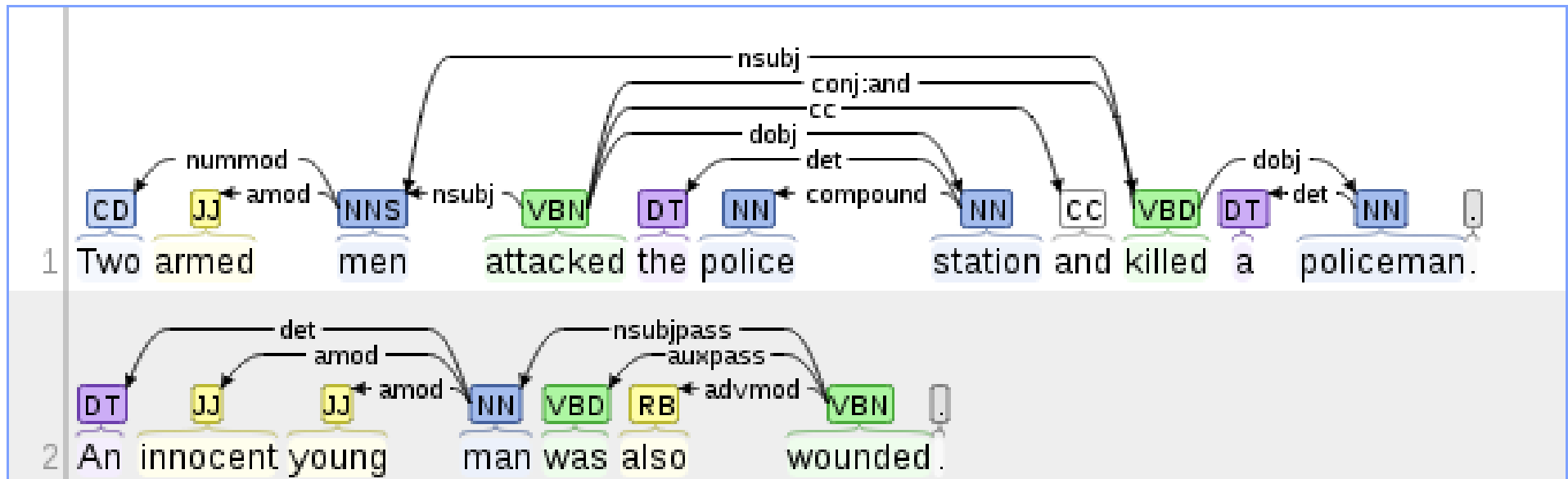




**Attributes**

**Head**

**Triggers**



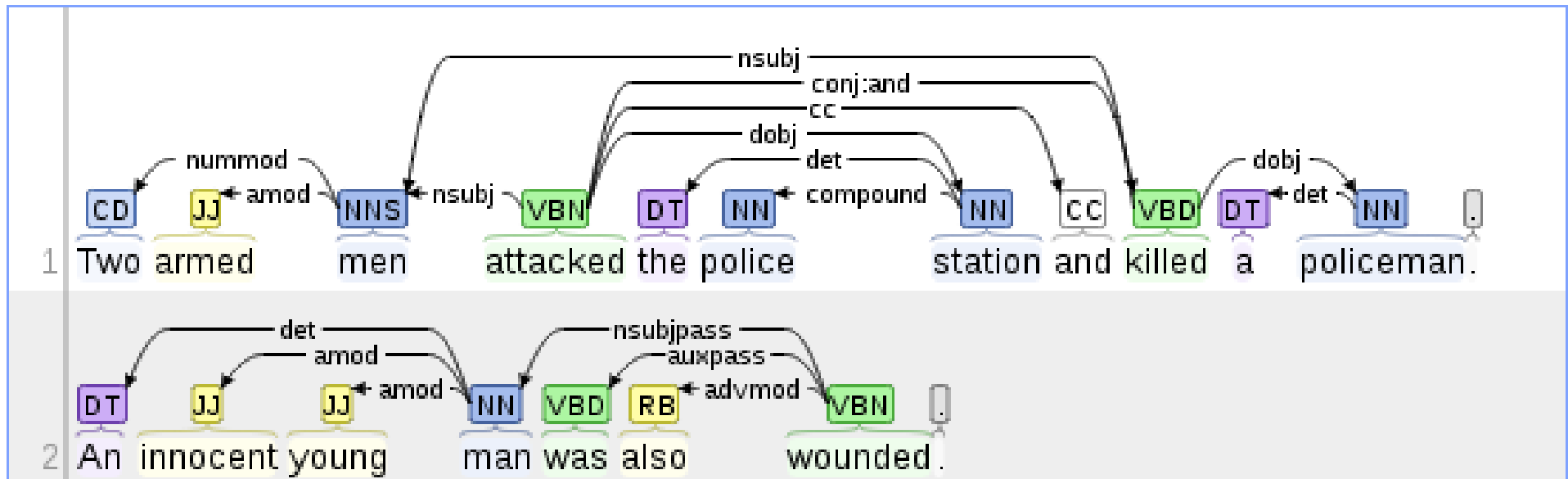
**Attributes**

#1

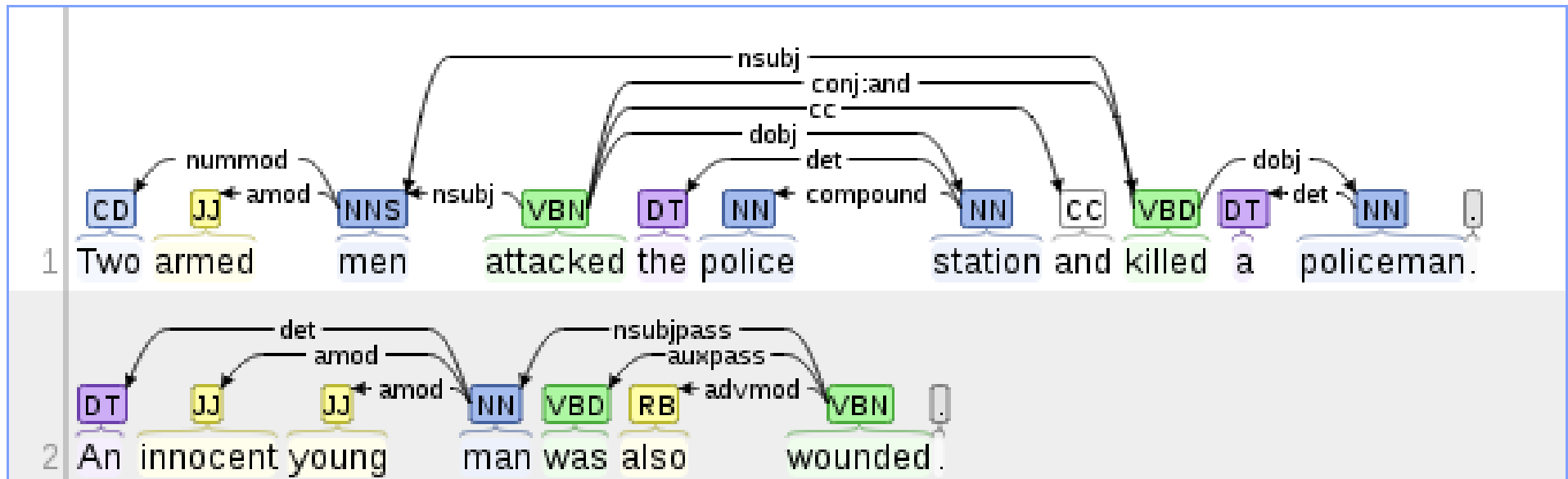
**Head**

man

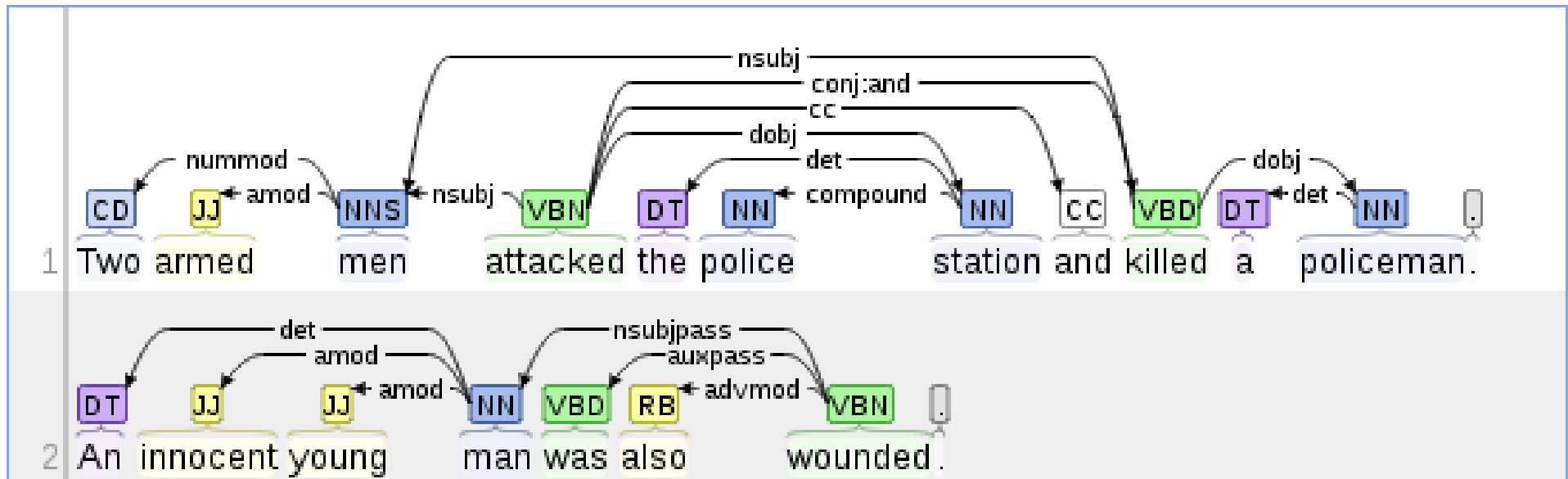
**Triggers**



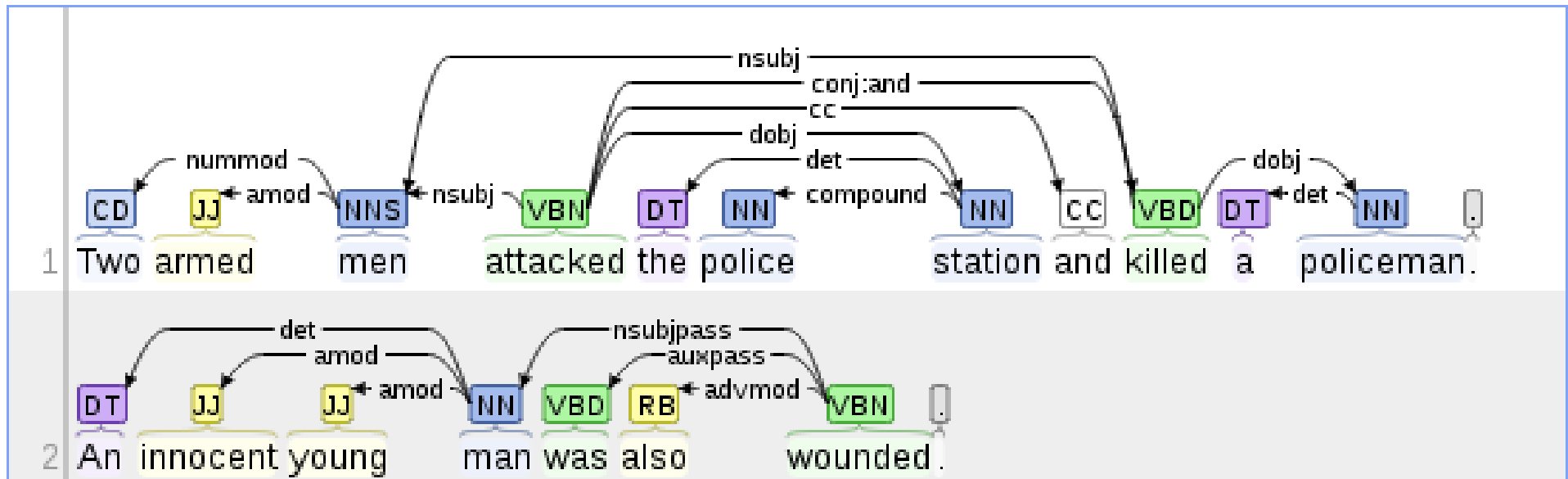
	Attributes	Head	Triggers
#1	[armed:amod]	man	



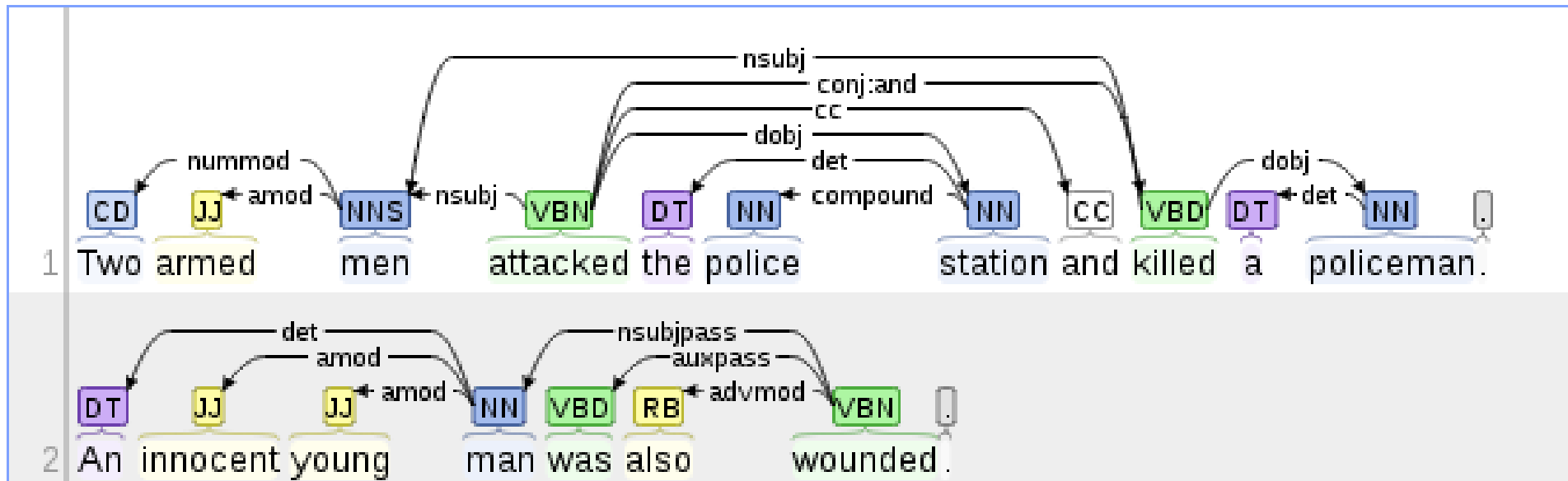
	Attributes	Head	Triggers
#1	[armed:amod]	man	[attack:nsbj,



	Attributes	Head	Triggers
#1	[armed:amod]	man	[attack:nsubj, kill:nsubj]

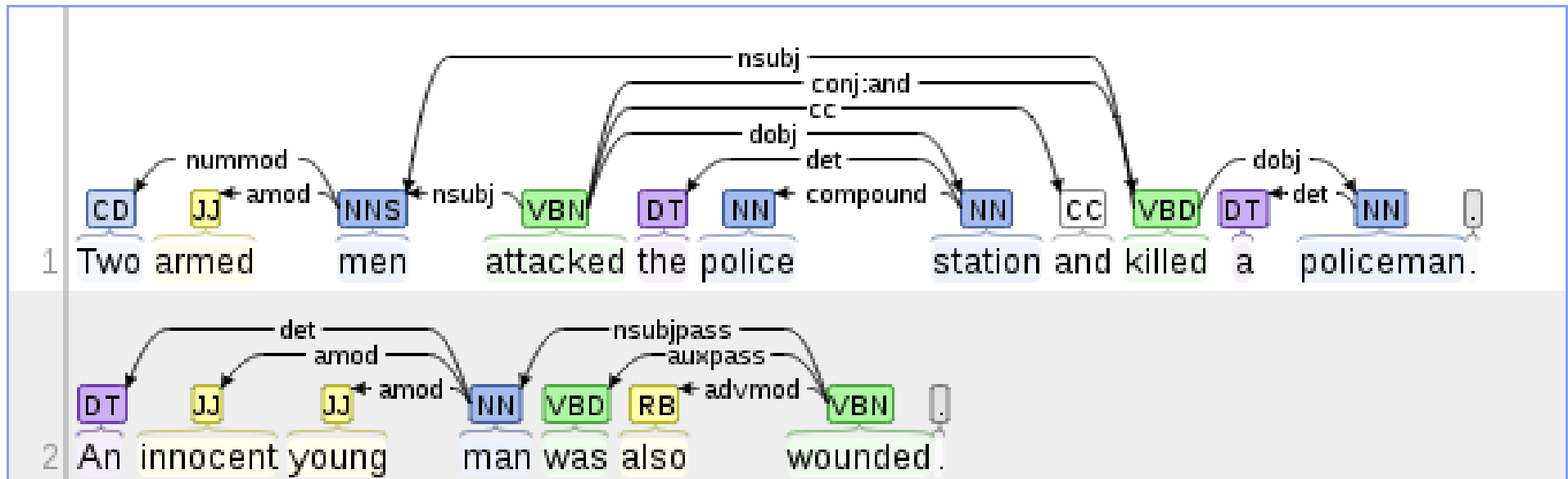


	Attributes	Head	Triggers
#1	[armed:amod]	man	[attack:nsbj, kill:nsbj]
#2	[police:nn]	station	[attack:dobj]

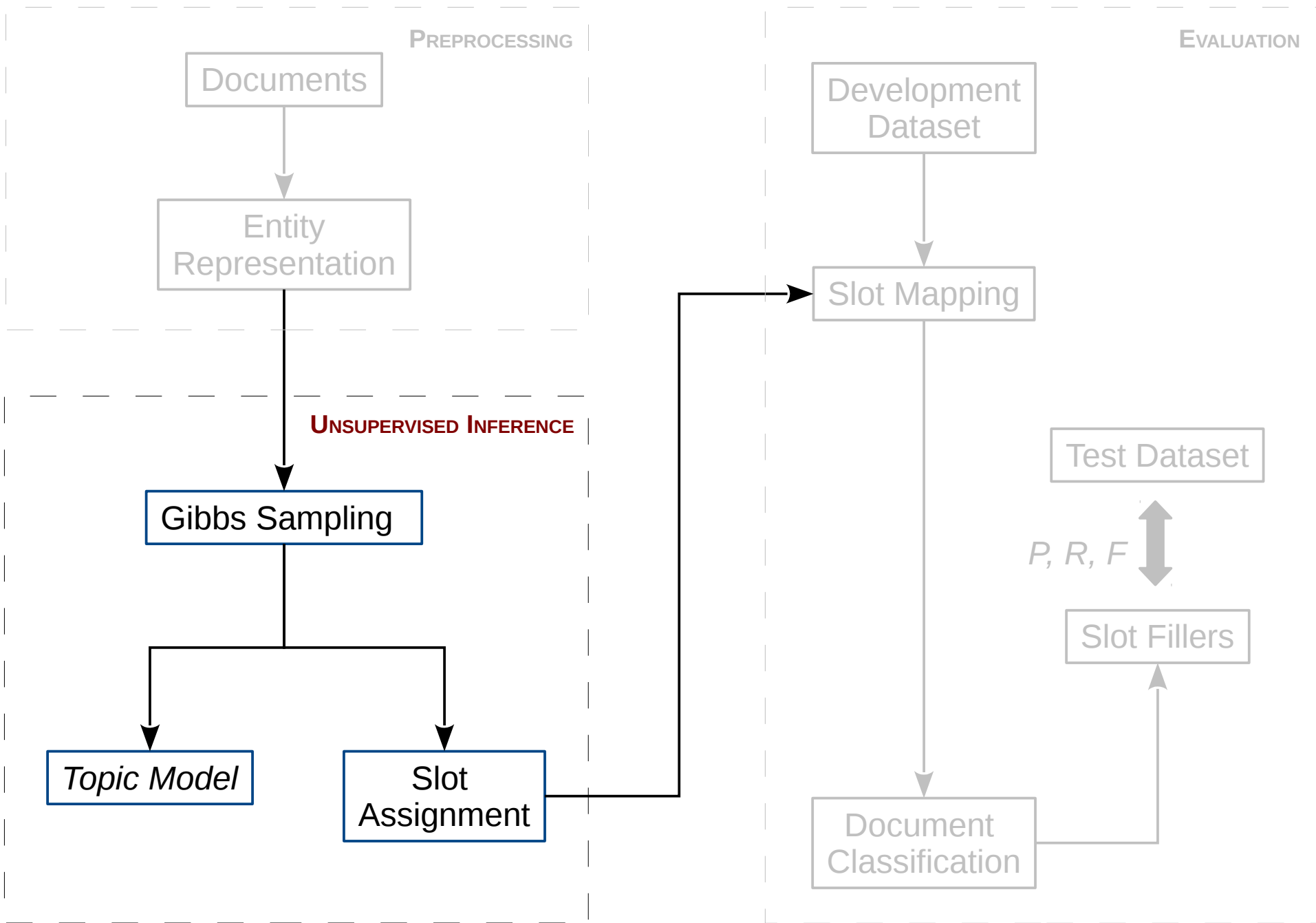


	Attributes	Head	Triggers
#1	[armed:amod]	man	[attack:nsubj, kill:nsubj]
#2	[police:nn]	station	[attack:dobj]
#3	[]	policeman	[kill:dobj]





	Attributes	Head	Triggers
#1	[armed:amod]	man	[attack:nsubj, kill:nsubj]
#2	[police:nn]	station	[attack:dobj]
#3	[]	policeman	[kill:dobj]
#4	[innocent:amod, young:amod]	man	[wound:dobj]



$\text{dir}(\gamma) \longrightarrow \varphi$

$\text{dir}(\beta) \longrightarrow \pi$

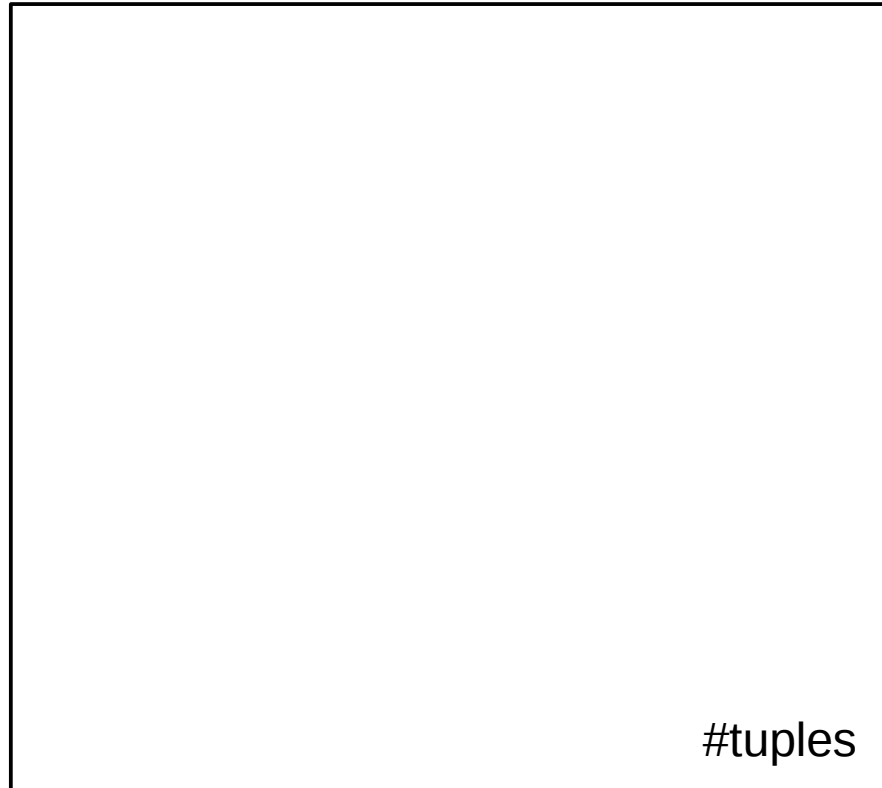
$\text{dir}(\alpha) \longrightarrow \theta$

$\text{uni}(1, K)$

$\text{dir}(\gamma) \longrightarrow \varphi$

$\text{dir}(\beta) \longrightarrow \pi$

$\text{dir}(\alpha) \longrightarrow \theta$



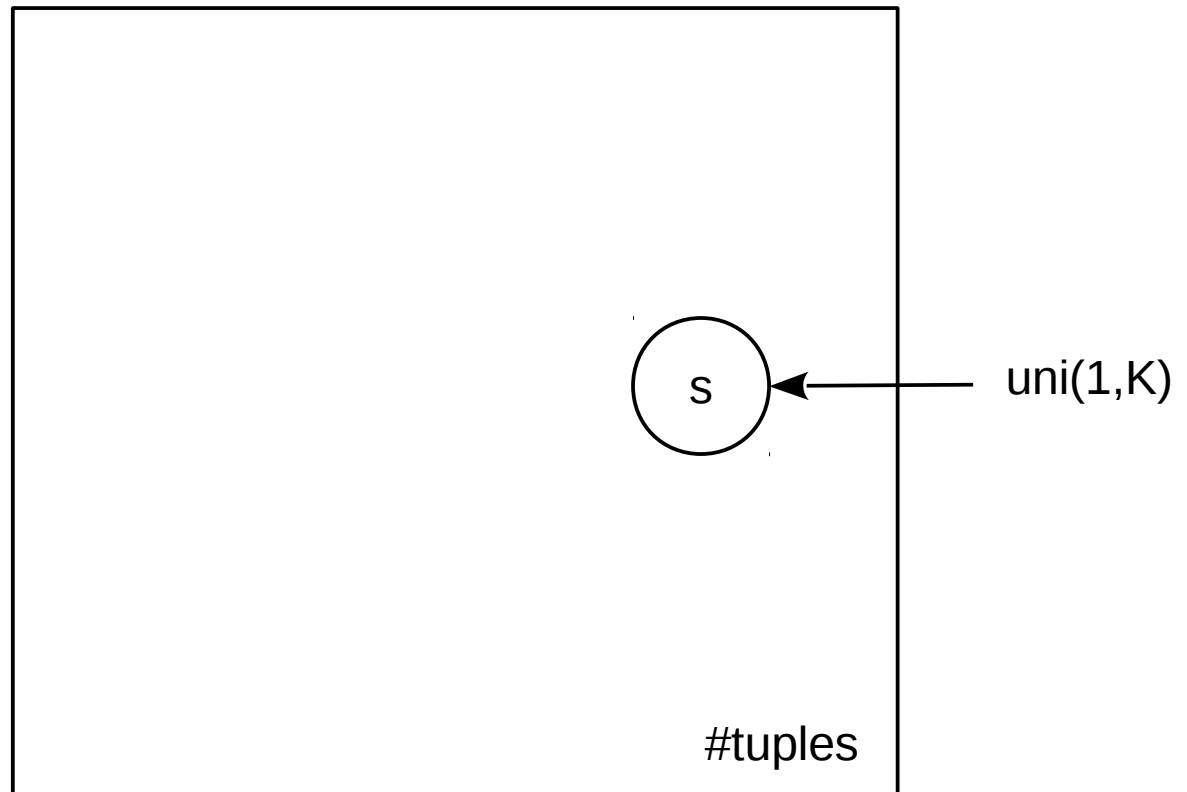
#tuples

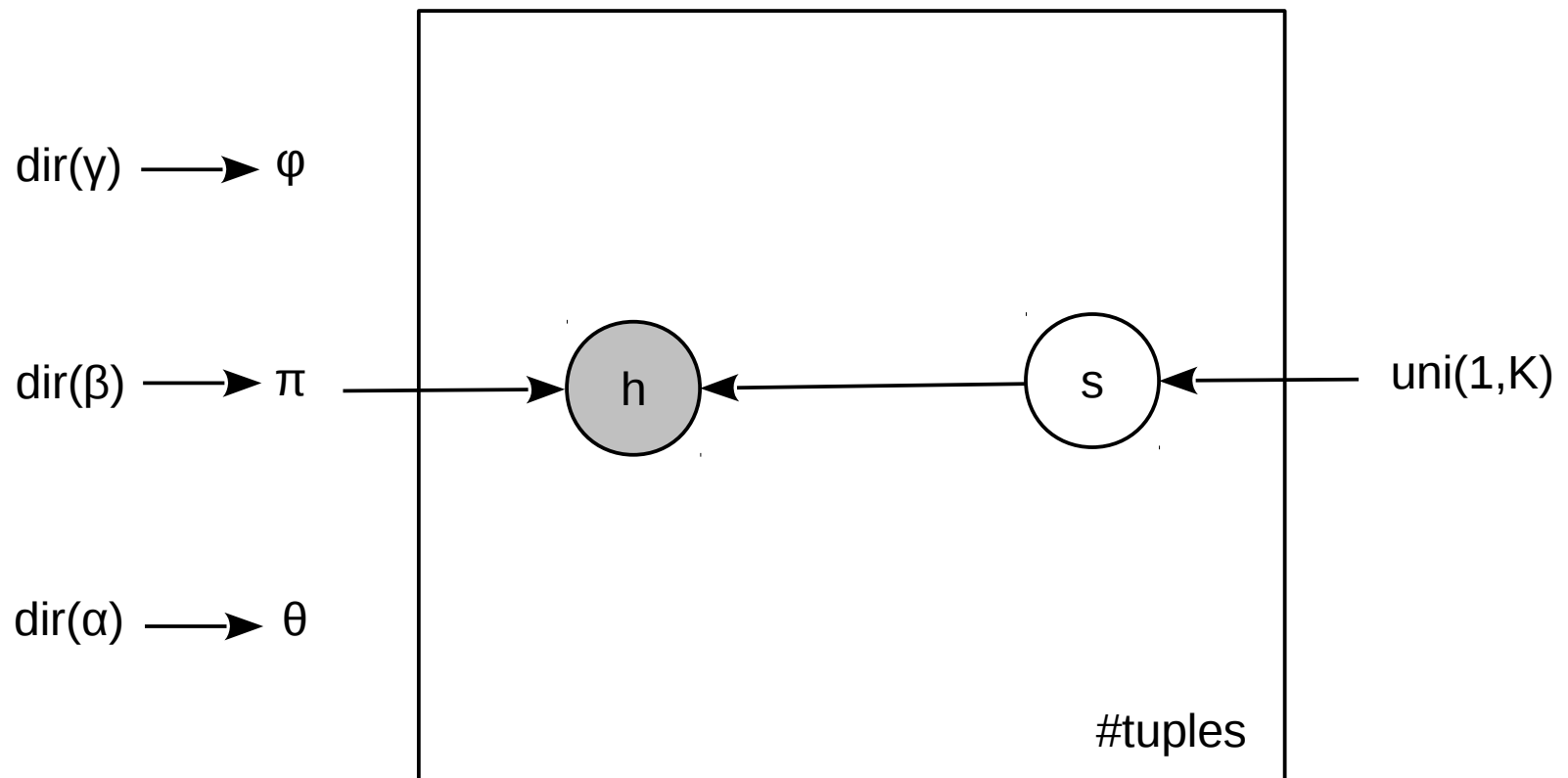
$\text{uni}(1, K)$

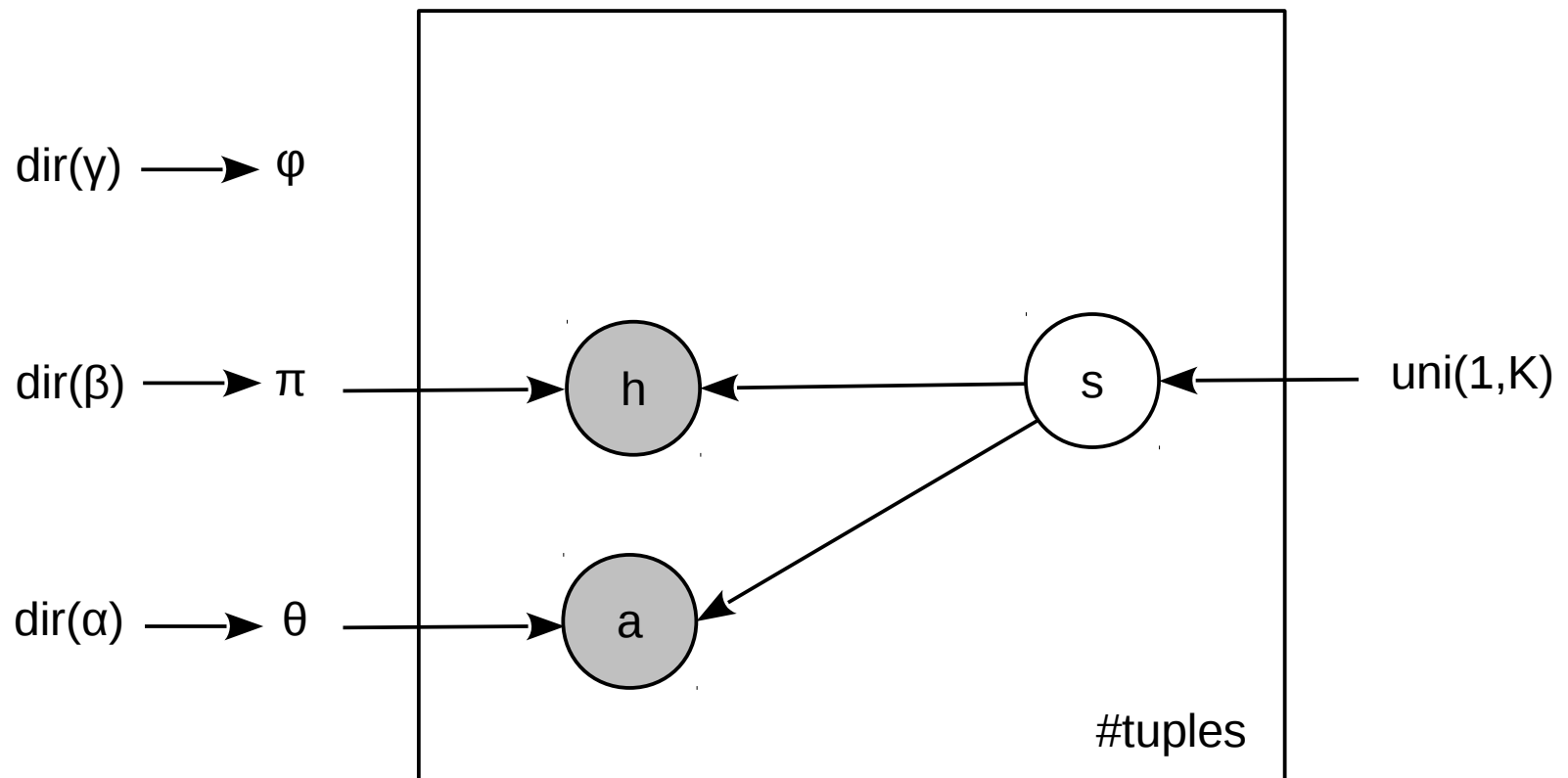
$\text{dir}(\gamma) \longrightarrow \varphi$

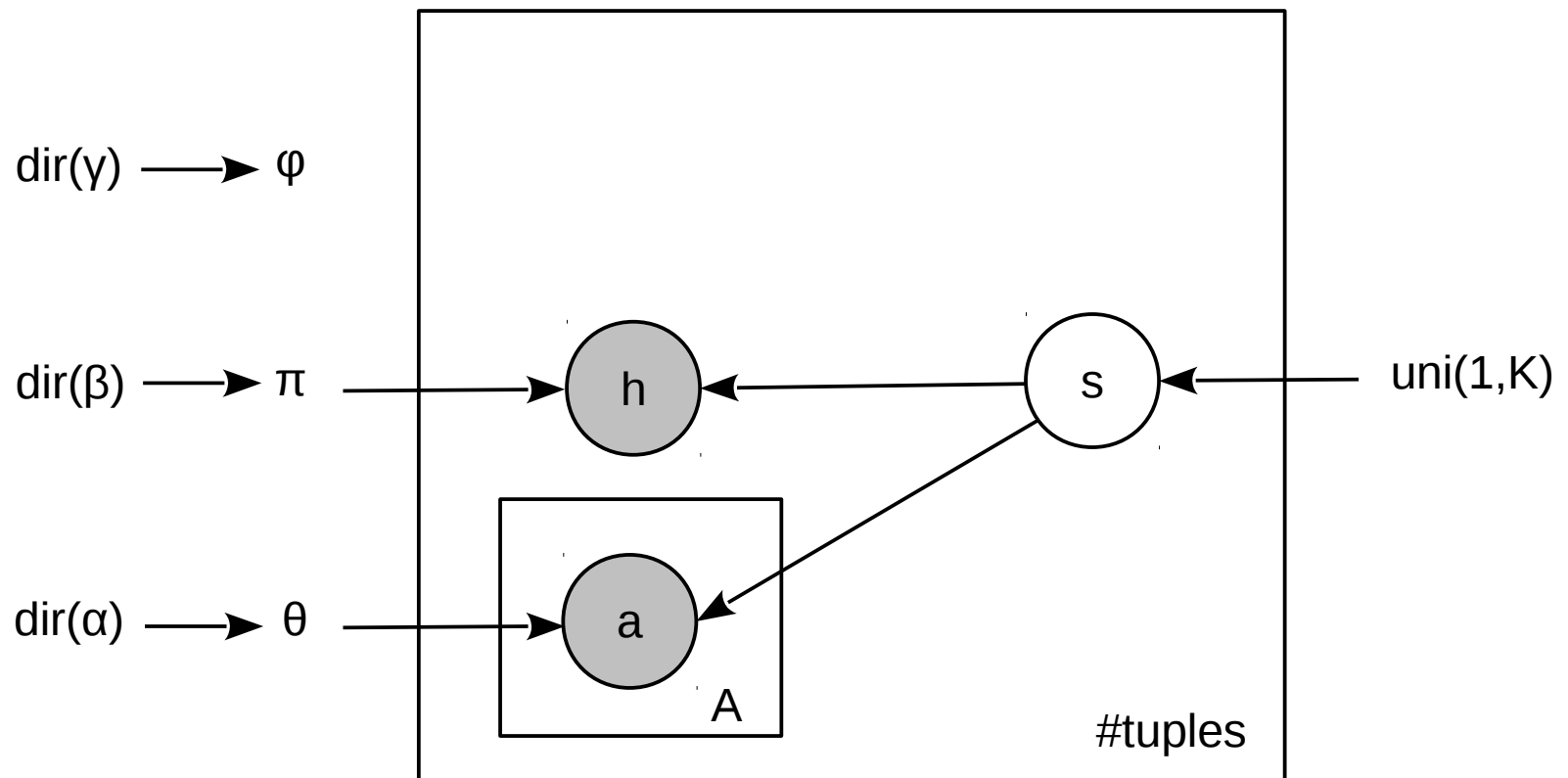
$\text{dir}(\beta) \longrightarrow \pi$

$\text{dir}(\alpha) \longrightarrow \theta$

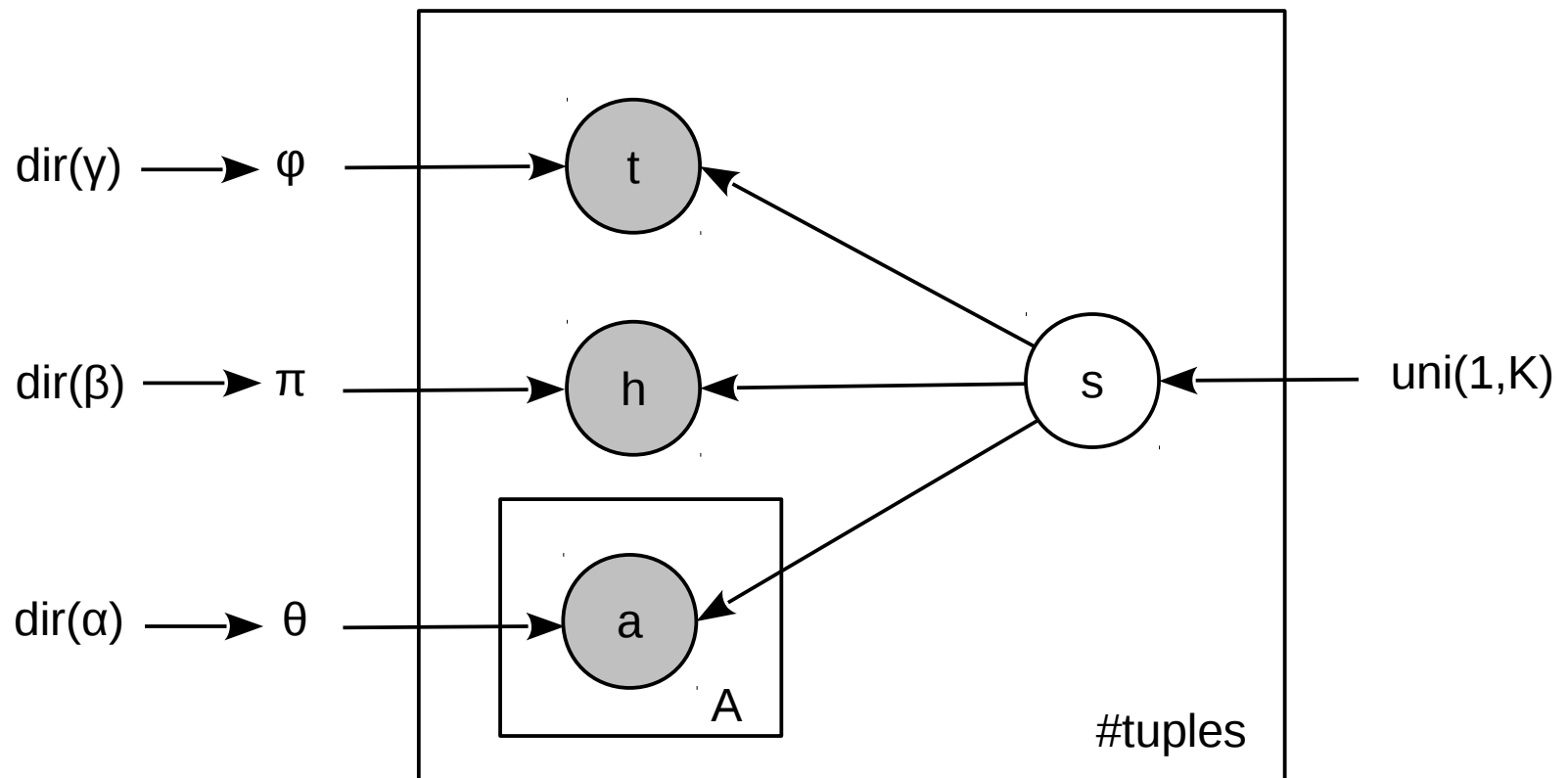


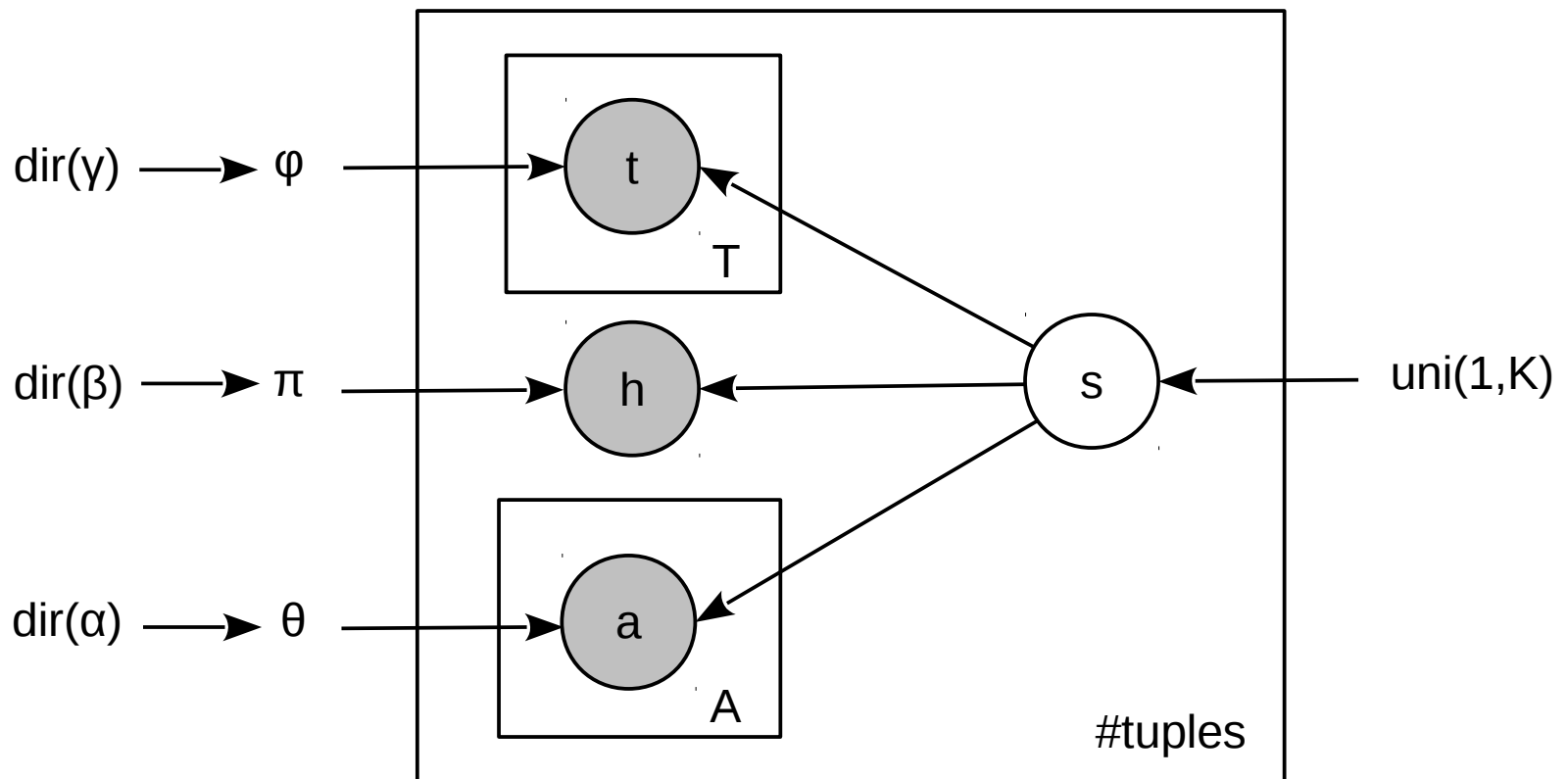






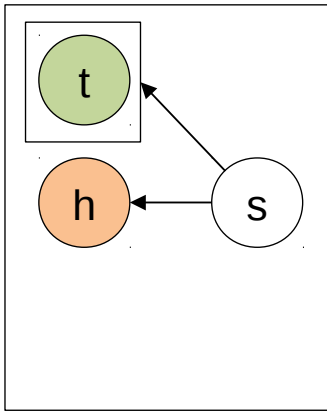




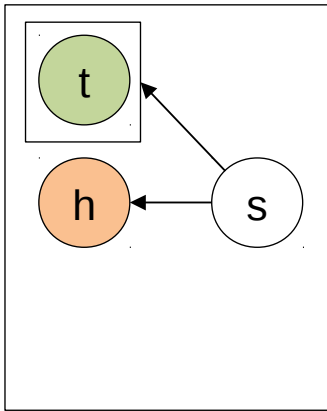


# Why ambiguity matters? & Our resolution

## Model



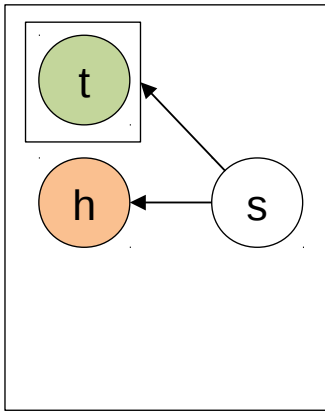
Model



Induced slot: **ATTACK\_victim**

<u>Head</u> $pr(h s)$	<u>Triggers</u> $pr(t s)$
<i>citizen</i>	<i>kill:dobj</i>
<i>woman</i>	<i>murder:dobj</i>
<i>man</i>	<i>die:nsubj</i>
<i>police</i>	<i>die:prep_of</i>
<i>terrorist</i>	<i>wound:dobj</i>

## Model



Induced slot: **ATTACK\_victim**

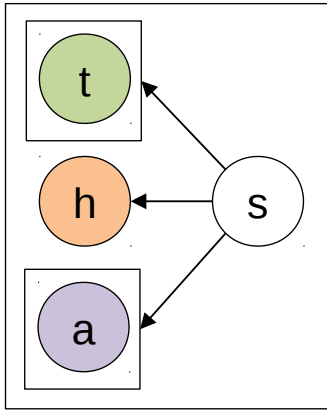
<u>Head</u> $pr(h s)$	<u>Triggers</u> $pr(t s)$
<i>citizen</i>	<i>kill:dobj</i>
<i>woman</i>	<i>murder:dobj</i>
<i>man</i>	<i>die:nsubj</i>
<i>police</i>	<i>die:prep_of</i>
<i>terrorist</i>	<i>wound:dobj</i>

## Slot Assignment

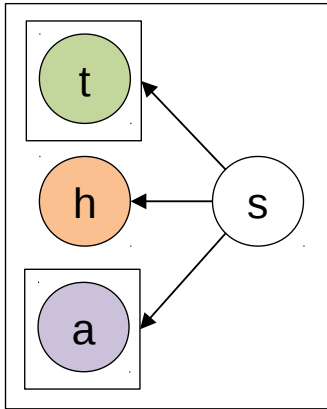
Municipal official **Sergio Horna** was seriously **wounded**.

Two extremist **terrorists** were reportedly **killed** by national officers.

## Model



Model

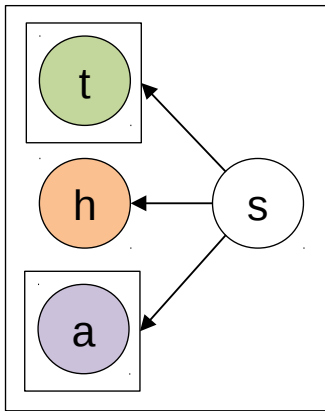


Induced slot: **ATTACK\_victim**

Head $pr(h s)$	Triggers $pr(t s)$
<i>citizen</i>	<i>murder:dobj (+)</i>
<i>woman</i>	<i>kill:dobj (-)</i>
<i>police</i>	<i>die:nsubj</i>
<i>victim (+)</i>	<i>die:prep_of</i>
<i>civilian (+)</i>	<i>assassinate:dobj (+)</i>



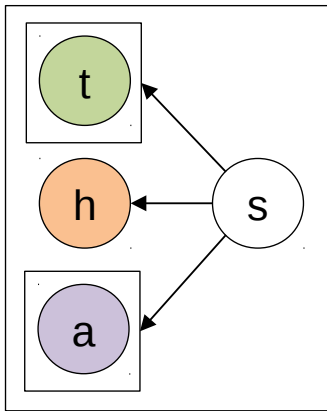
Model



Induced slot: **ATTACK\_victim**

Attributes $pr(a s)$	Head $pr(h s)$	Triggers $pr(t s)$
<i>innocent:amod</i>	<i>citizen</i>	<i>murder:dobj (+)</i>
<i>wounded:amod</i>	<i>woman</i>	<i>kill:dobj (-)</i>
<i>U.N:nn</i>	<i>police</i>	<i>die:nsubj</i>
<i>young:amod</i>	<i>victim (+)</i>	<i>die:prep_of</i>
<i>official:nn</i>	<i>civilian (+)</i>	<i>assassinate:dobj (+)</i>

## Model



## Induced slot: **ATTACK\_victim**

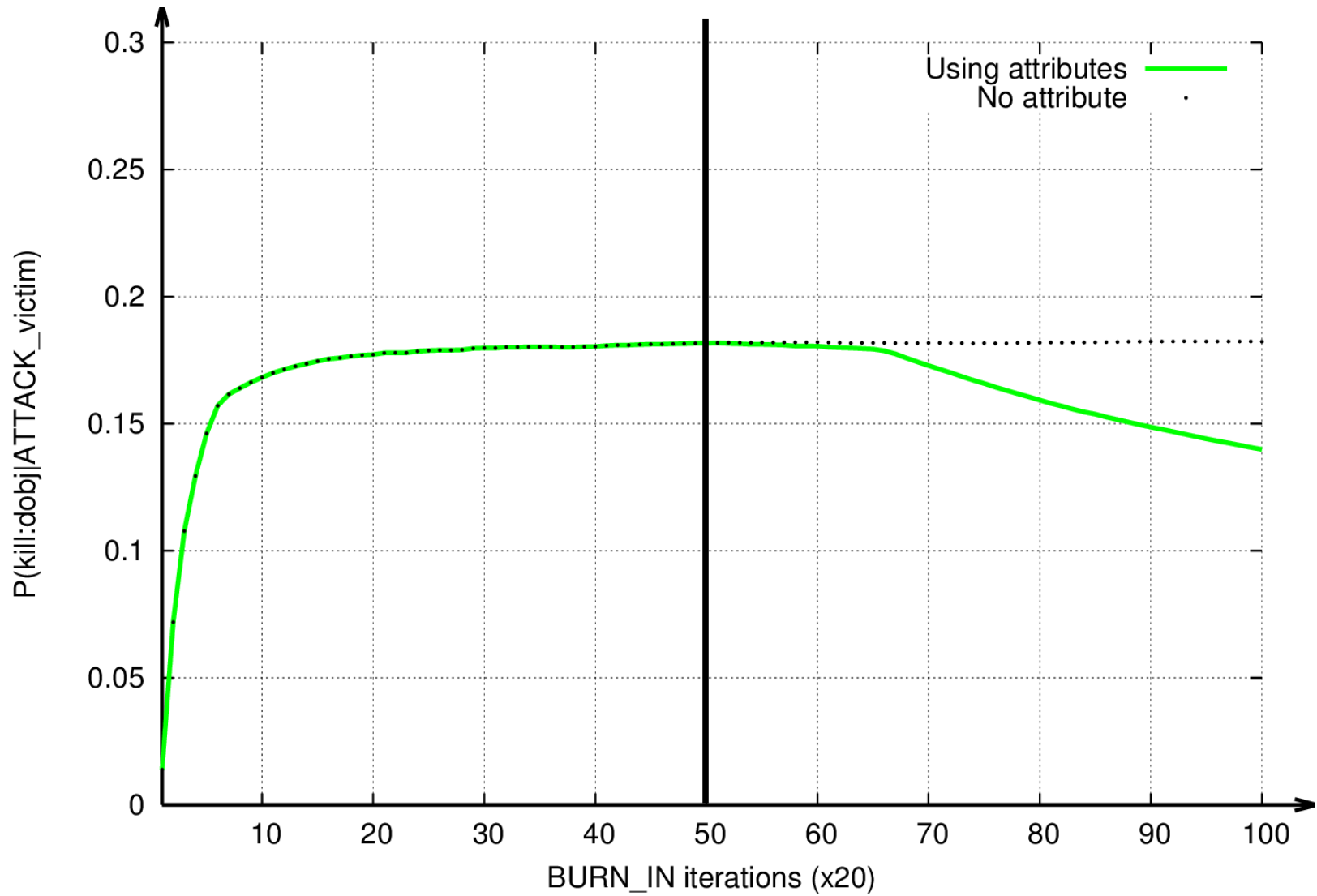
Attributes $pr(a s)$	Head $pr(h s)$	Triggers $pr(t s)$
<i>innocent:amod</i>	<i>citizen</i>	<i>murder:dobj (+)</i>
<i>wounded:amod</i>	<i>woman</i>	<i>kill:dobj (-)</i>
<i>U.N:nn</i>	<i>police</i>	<i>die:nsubj</i>
<i>young:amod</i>	<i>victim (+)</i>	<i>die:prep_of</i>
<i>official:nn</i>	<i>civilian (+)</i>	<i>assassinate:dobj (+)</i>

## Slot Assignment

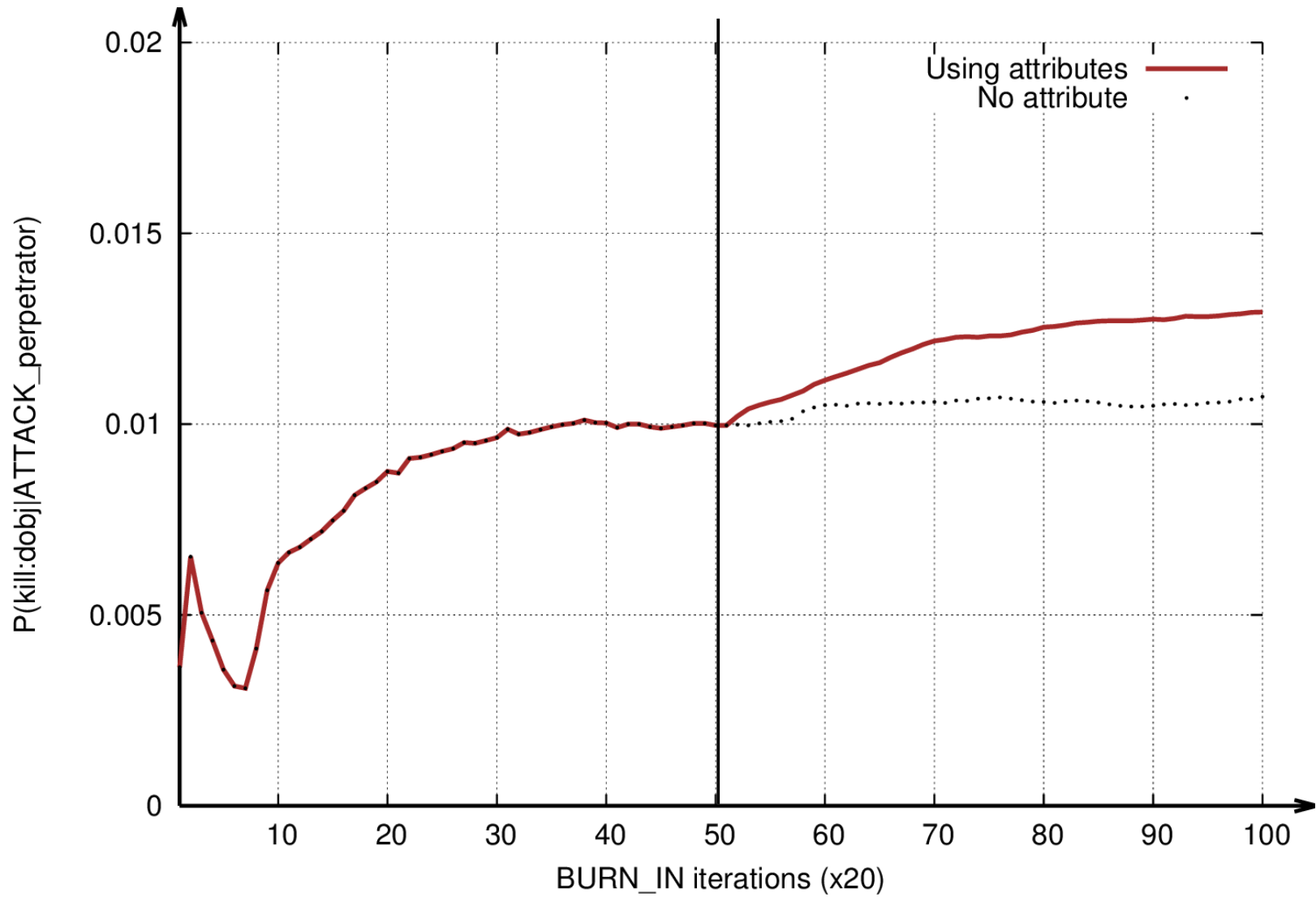
Municipal **official Sergio Horna** was seriously **wounded**.

Two **extremist terrorists** were reportedly **killed** by national officers.

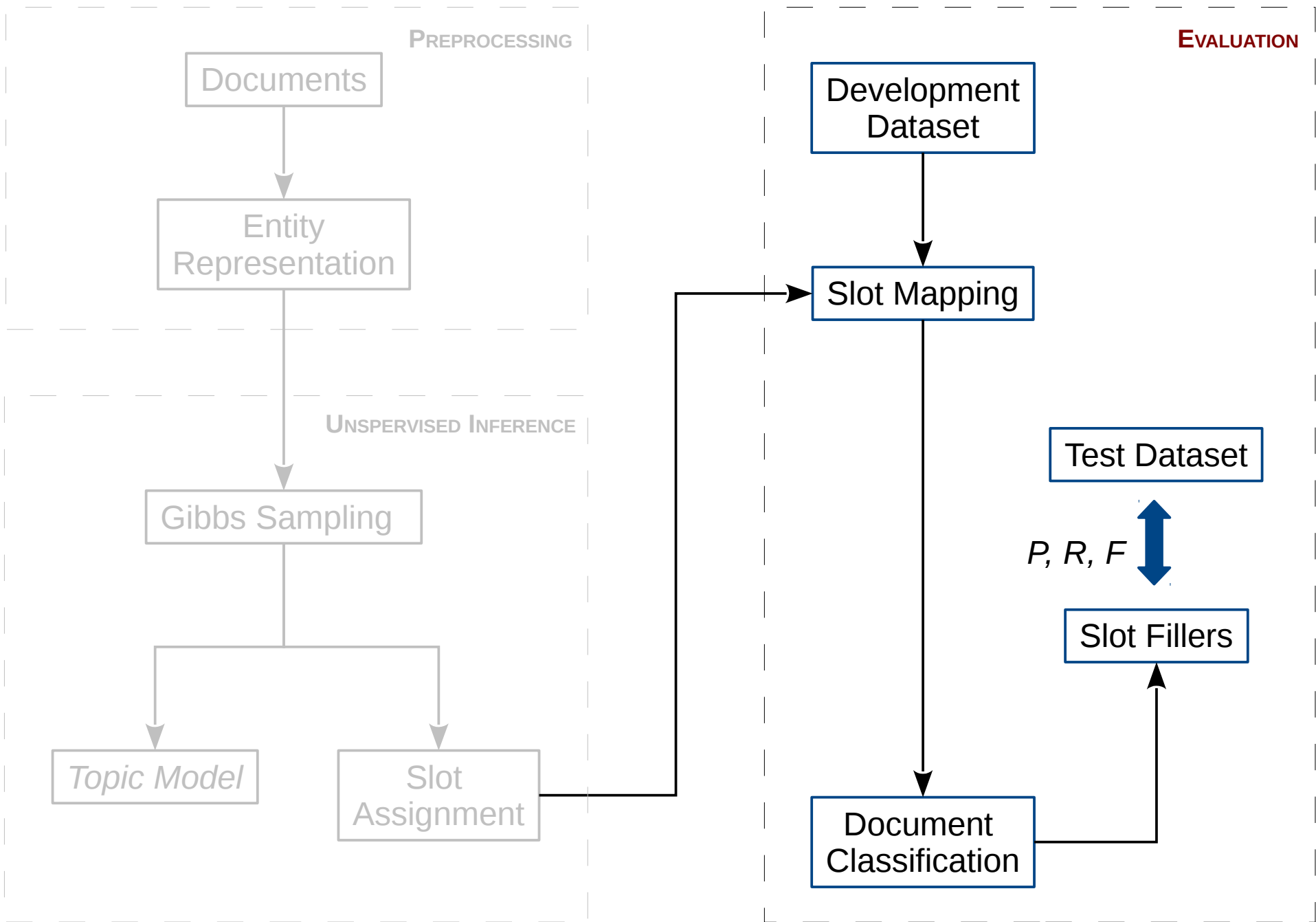
p(kill:dobj) in slot ATTACK\_victim



$p(\text{kill:dobj})$  in slot ATTACK\_perpetrator



# Evaluation



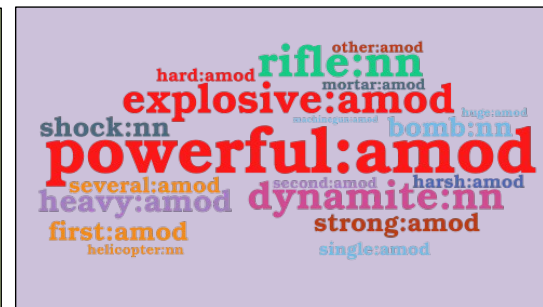
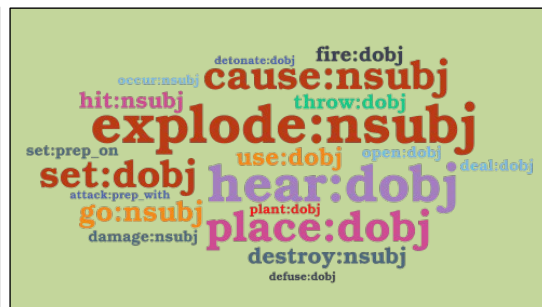
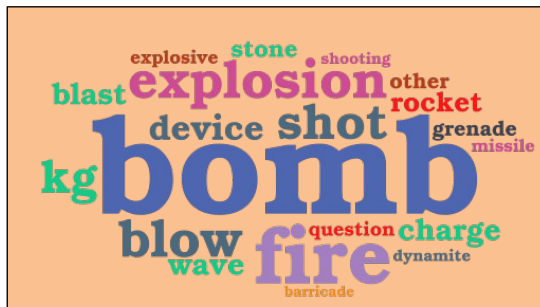
- Data: MUC-4 corpus
  - Terrorist events
  - 1,700 docs: 1,300 train / 200 dev (tst1-2)/ 200 test (tst3-4)
  - Uppercase restoration, linguistic analysis using Stanford NLP toolkit
- Evaluation task: Slot filling
  - Precision, recall, f-score
  - 'Right most word' phrase matching
  - Slots of different templates are exchangeable
  - #slots K is optimized on dev dataset (K=35)

# Slot mapping

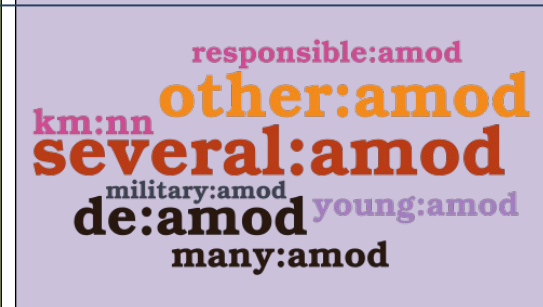
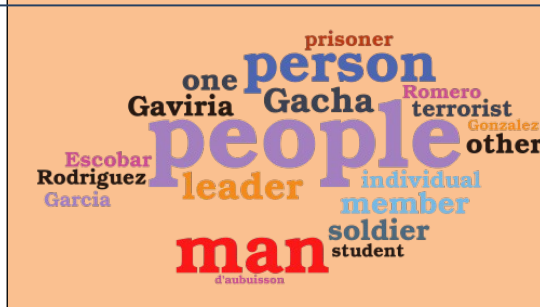
- $K$  ( $K = 35$ ) learned slots  $>$  reference slots (14)
- Map each reference slot to its best fit learned slot
- Require a dev dataset and a scoring function (F-score)



Slot 1



Slot 2



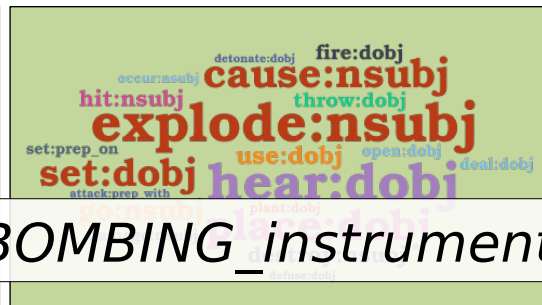
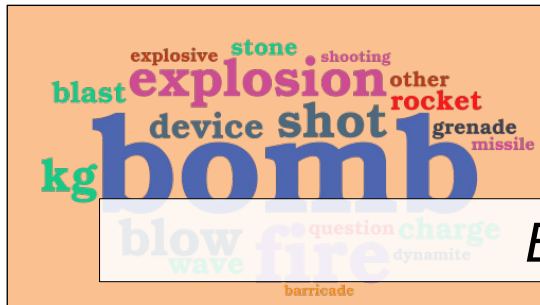
...

...

...

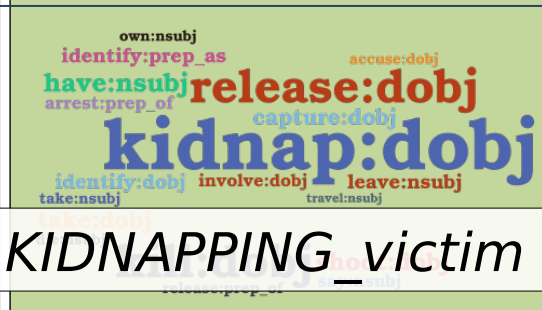
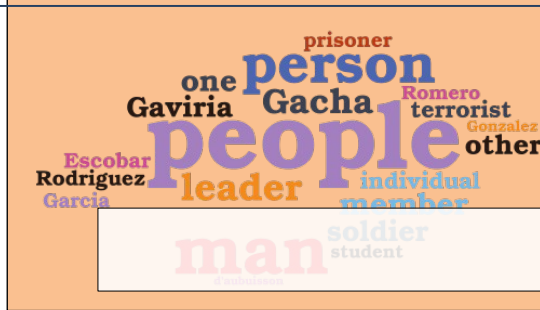
...

Slot 1



BOMBING\_instrument

Slot 2



KIDNAPPING\_victim

...

...

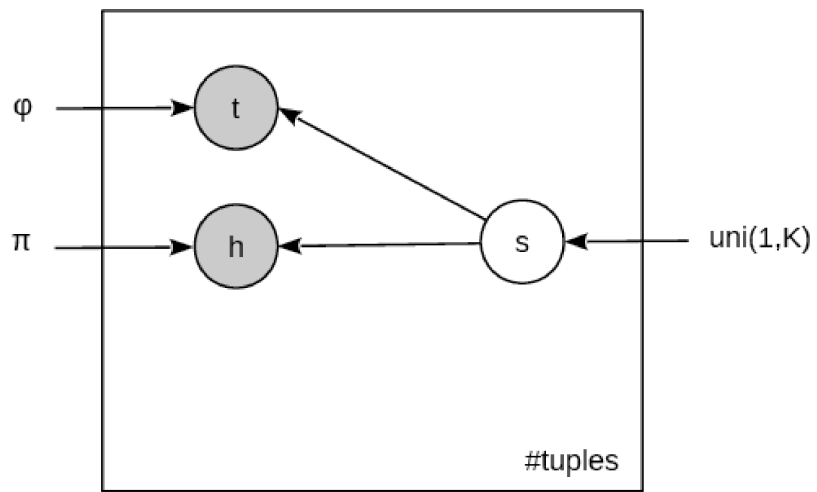
...

...

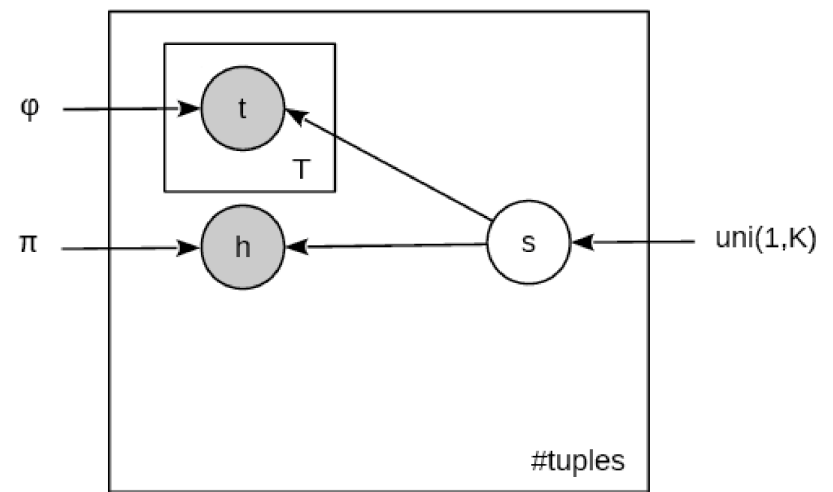
# Document classification

- Irrelevant documents (in MUC-4 dataset)
  - Containing *terrorist* entities
  - Not containing actual *terrorist* events
- Classifying relevance vs. Irrelevance

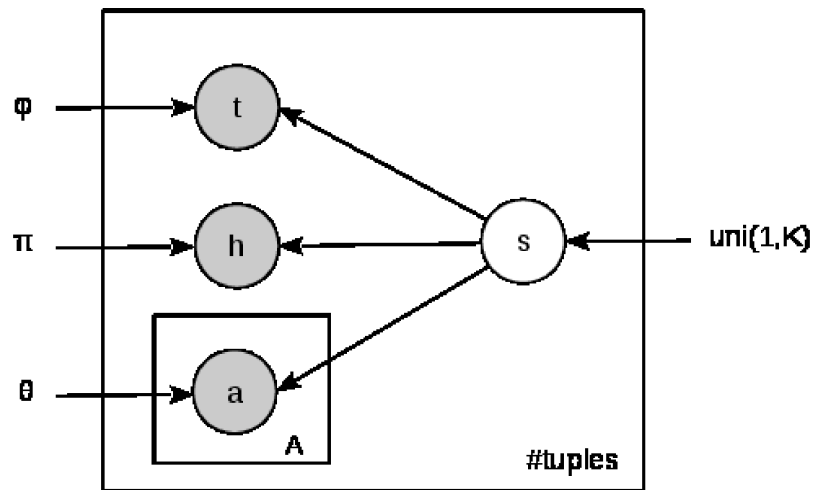
$$\mathit{relevance}(d) = \frac{\sum_{e \in d: s_e \in S_m} \sum_{t \in T_e} P(t|s_e)}{\sum_{e \in d} \sum_{t \in T_e} P(t|s_e)}$$



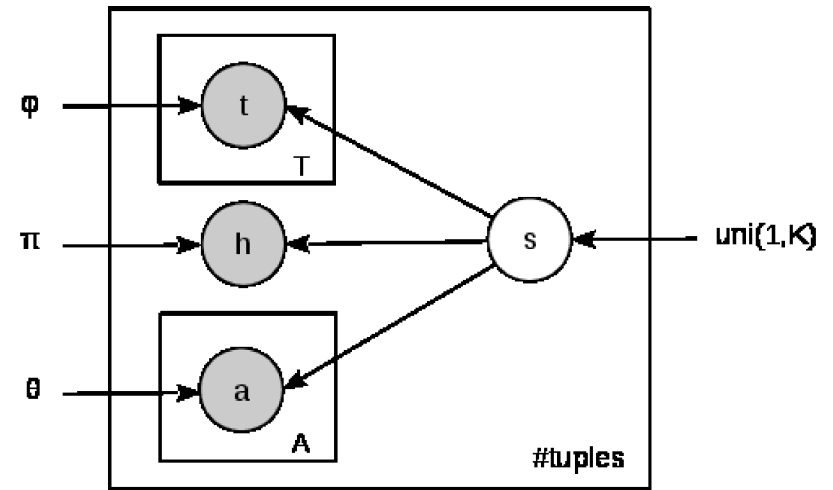
a) HT&Single



b) HT&Multi/Coref



c) HT+A&Single



d) HT+A&Multi/Coref

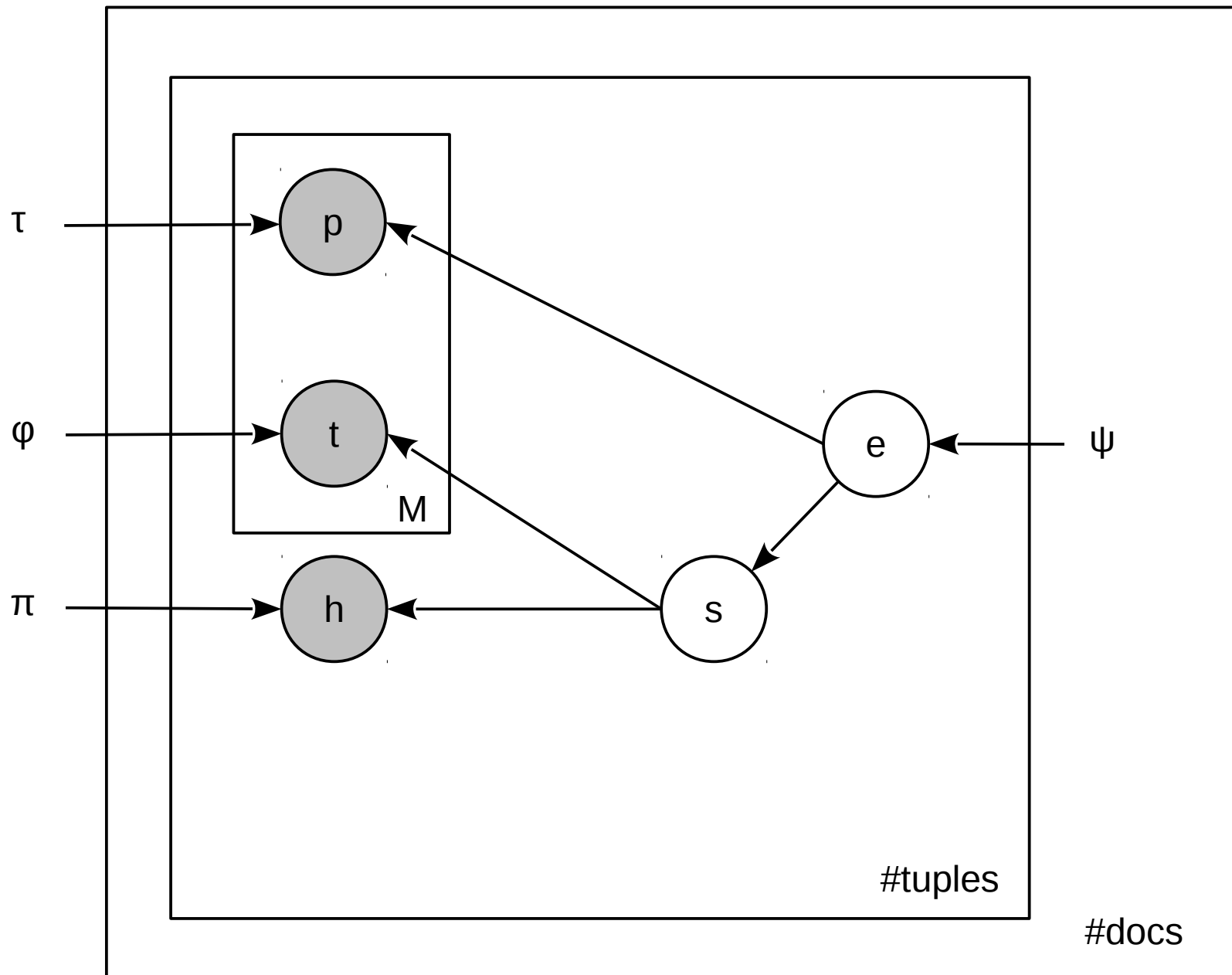


Figure: Chambers'13 model

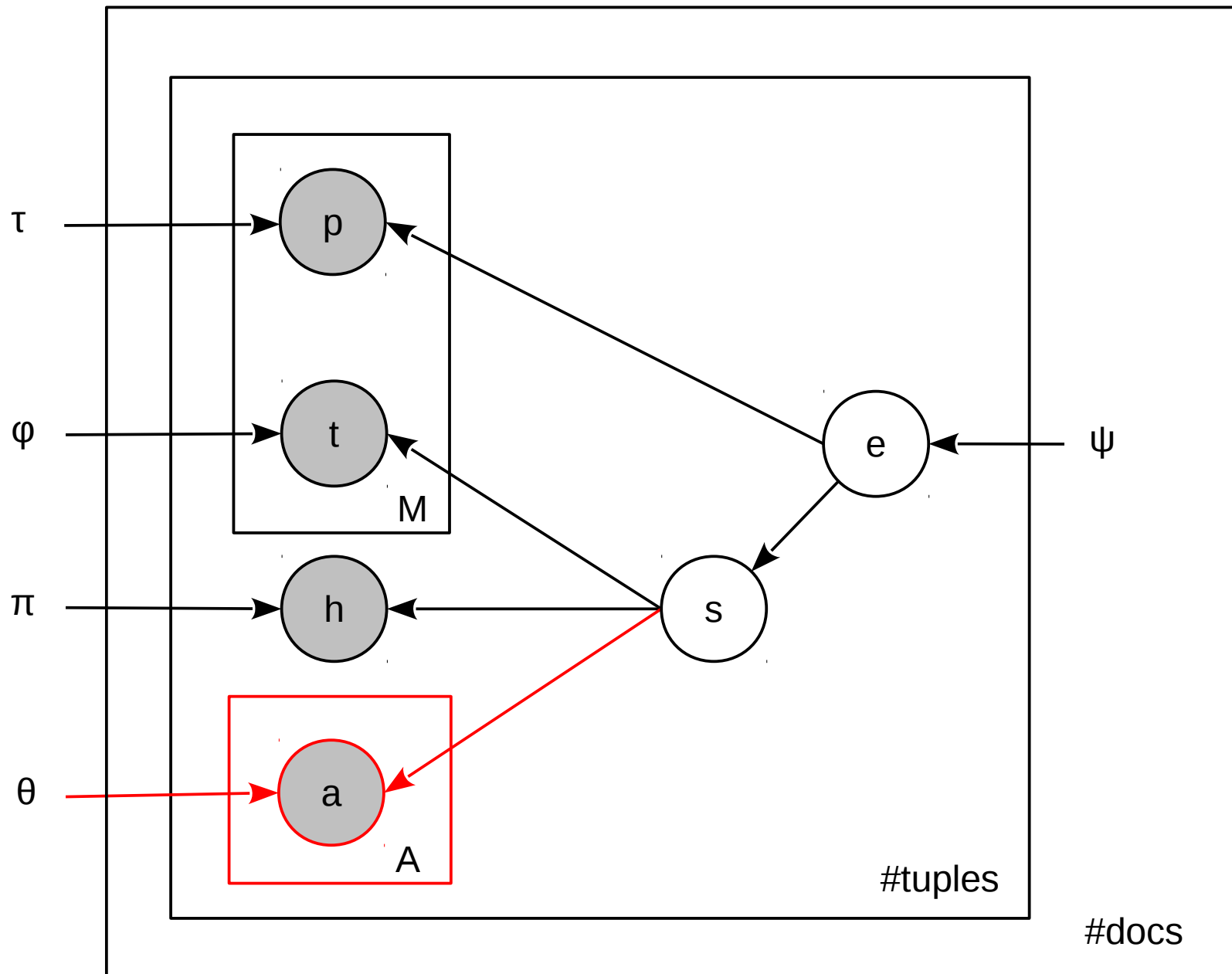


Figure: Chambers'13 model + Attribute distribution

---

	Precision	Recall	F-score
HT & Single	29.59	51.17	37.48
HT & Multi	29.32	52.21	37.52
HT & <b>Coref</b>	29.99	53.53	40.01

---

---

---

---

---

---

	Precision	Recall	F-score
HT & Single	29.59	51.17	37.48
HT & Multi	29.32	52.21	37.52
HT & <b>Coref</b>	29.99	53.53	40.01
HT + A & Single	30.22	52.41	38.33
HT + A & Multi	30.82	51.68	38.55
HT + A & <b>Coref</b>	32.42	<b>54.59</b>	40.62

---

---

---

---



	Precision	Recall	F-score
HT & Single	29.59	51.17	37.48
HT & Multi	29.32	52.21	37.52
HT & <b>Coref</b>	29.99	53.53	40.01
HT + A & Single	30.22	52.41	38.33
HT + A & Multi	30.82	51.68	38.55
HT + A & <b>Coref</b>	32.42	<b>54.59</b>	40.62
HT + A + Doc classification	35.57	53.89	<b>42.79</b>
HT + A + Oracle doc classification	44.58	54.59	49.08

	Precision	Recall	F-score
HT & Single	29.59	51.17	37.48
HT & Multi	29.32	52.21	37.52
HT & <b>Coref</b>	29.99	53.53	40.01
HT + A & Single	30.22	52.41	38.33
HT + A & Multi	30.82	51.68	38.55
HT + A & <b>Coref</b>	32.42	<b>54.59</b>	40.62
HT + A + Doc classification	35.57	53.89	<b>42.79</b>
HT + A + Oracle doc classification	44.58	54.59	49.08
Chambers'13(reimpl)	38.65	42.68	40.56
Chambers'13(reimpl) +Attr	<b>39.25</b>	43.68	41.31

---

	Precision	Recall	F-score
Cheung et al	32	37	34
Chambers'11	<b>48</b>	25	33
Chambers'13 (paper values)	41	41	41
HT+A+Doc classification	36	<b>54</b>	<b>43</b>

---

# Conclusions & Discussions

- We proposed
  - Joint model for role induction & filling
  - Entity disambiguation using syntactic information
- Future work
  - Evaluation on other domains, corpora
  - Integration of temporal-spatial information about events

謝謝你 *danke*  
German

*dziękuję* **gracias**

**obrigado** 고맙습니다

**thank you**

cảm ơn bạn *grazie*  
Italian

*salamat*

спасибо **merci**  
Russian French

- ありがとうございました  
Japanese