



# Universal Information Extraction and Knowledge-driven Generation

Heng Ji (UIUC)

[hengji@illinois.edu](mailto:hengji@illinois.edu)



# Information Overload: Multimedia Multilingual Data



## STORY HIGHLIGHTS

Congress and Ukraine long sought anti-tank weapons in battle against separatists

(CNN) — The US is going to provide lethal anti-tank weapons to Ukraine to help it fight off separatists which have the backing of Russia, a senior State Department official told CNN.



Бійці батальйону «Дніпро»

Новообраний президент Володимир Зеленський заявив про необхідність повертати контроль над кождим миллиметром української границі

КИЇВ – Не порушає Мінських договоренностей, за год действия Операції Об'єднаних сил Україна вернула під свій контроль 24 квадратних кілометрів території на Донбасі. Об этом говориться в обращении командуючого Об'єднаними силами генерал-лейтенанта Сергєя Наєва по случаю первой годовщины проведения операции на Донбасі.

## Читайте также

На Донбасі завершилась АТО и началась операция Об'єднаних сил



Президент России Владимир Путин (архивное фото)

В предыдущем указе Владимира Путина был упрощен порядок предоставления гражданства жителями Донбасса

КИЇВ – Россия упрощает порядок получения гражданства для украинцев, временно проживающих на ее территории, украинским беженцам, жителям отдельных районов Донецкой и Луганской областей, а также другим категориям граждан из Украины и ряда стран.

## Читайте также

Украина протестует против выдачи паспортов России жителям Донбасса

Нуждаются ли жители Донбасса в российском

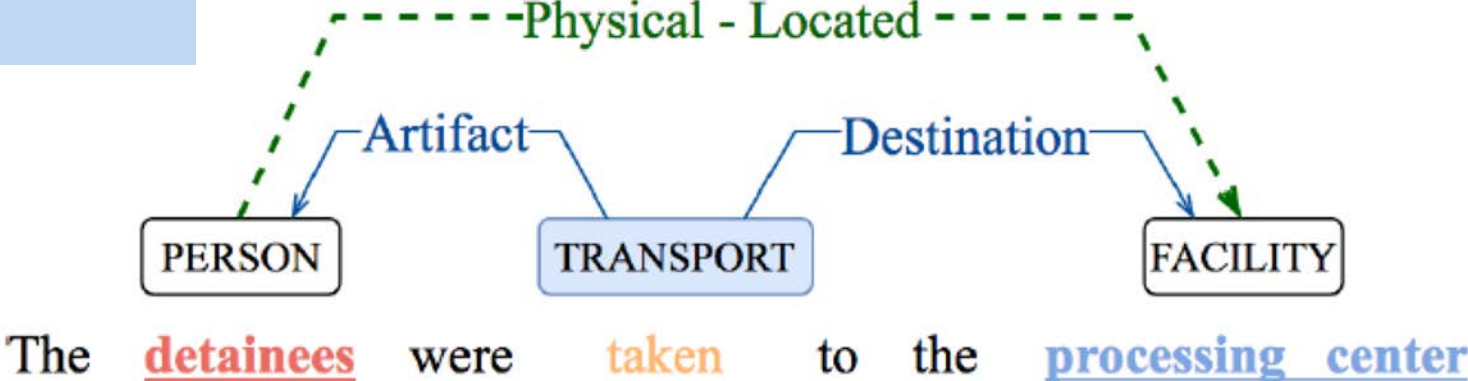


американський ракетний есминець Preble

американські військові повідомили, що два військові кораблі США пройшли біля островів, на які претендує Китай, у Південно-Китайському морі в неділок. Ці дії розгнівали Пекін у час напружених зв'язків між двома найбільшими економіками світу.

# Information Extraction: turn unstructured data to structured information

## Relation Extraction



## Event Extraction (Trigger Labeling + Argument Role Labeling)



# Applications: Disaster Relief

ELISA | Heat Map

**Context Sentence**

sentence\_18442  
Language: Somali  
Subagle wuxuu sheegay in kooxda Al-Shabaab ay isku dayday inay dib u qabsato magaalada Buulo-burte, oo toddobaad ka hor ay qabsadeen ciidanka xoogga **dalka Soomaaliya** oo gacan ka helaya kuwa dalka Jabuuti ee qeybta ka ah AMISOM.  
Translation: Subagle said al-Shabaab militants attempted to recapture the town of Buulo-burte, which had taken control of the Somali strongholds in Djibouti with the help of AMISOM troops.

**Spot (Language)**

**Icon (Topic)**

**Translation**

South Atlantic Ocean

Indian Ocean

**Style Selector**

**Topic Filter**

**Language Filter**

Map Style: Streets Light Dark Satellite

Topic: All

Language: All

Legend:

- Food Supply
- Water Supply
- Medical Assistance
- Terrorism or other Extreme Violence
- Utilities, Energy, or Sanitation
- Evacuation
- Shelter
- Search and Rescue
- Civil Inrest
- Infrastructure

Language Legend:

- Amharic
- Chinese
- Kurdish
- Lao
- Oromo
- Somali
- Spanish
- Tigrinya
- Ukrainian
- Uyghur
- Vietnamese


- Re-trainable Systems: [http://blender02.cs.rpi.edu:3300/elisa\\_ie/api](http://blender02.cs.rpi.edu:3300/elisa_ie/api)
- Data and Resources: <http://nlp.cs.rpi.edu/wikiann/>
- Demos: [http://blender02.cs.rpi.edu:3300/elisa\\_ie](http://blender02.cs.rpi.edu:3300/elisa_ie) [http://blender02.cs.rpi.edu:3300/elisa\\_ie/heatmap](http://blender02.cs.rpi.edu:3300/elisa_ie/heatmap) 4

# Applications: Event Recommendation

Home Back Query: Target: Януковича (Yanukovich) Number of Events: 2

Automated Summary:  
[Show Visual Knowledge Elements](#) [Hide Visual Knowledge Elements](#)

Столкновения 20 февраля стали одним из ключевых факторов, вынудивших Президента Украины Виктора Януковича пойти на подписание Соглашения об урегулировании политического кризиса на Украине, потере доверия к самому Януковичу и к реформатированию парламентского большинства, вынесшего вечером 20 февраля постановление о запрете применения силы властью (Translation: The clashes on February 20 became one of the key factors that forced President of Ukraine Viktor Yanukovich to sign the Agreement on the settlement of the political crisis in Ukraine, the loss of confidence in Yanukovich himself and the reformatting of the parliamentary majority, which issued a resolution on the evening of February 20 banning the use of force), что согласно всем имеющимся уликам те милиционеры и демонстранты, что стали жертвами снайперского огня, застрелены одними и теми же снайперами (Translation: that according to all available evidence, those policemen and demonstrators who were victims of sniper fire were shot by the same snipers)



Date: 201402  
 Location: Unknown  
 Attackers: Unknown  
 Target: Януковича (Yanukovich)  
 Instrument: Unknown  
 Type of Attack: ConflictAttack

Source Document [English Translation](#)

Meta-path based Ranking Incorporating User Feedback and Cognitive Fit Theory  
[Similar Events](#) [Dissimilar Events](#)

Suggested Events

Date	Location	Attackers	Target	Instrument	Type of Attack
201402	country	his	Unknown	Unknown	ConflictAttack

Event Extraction

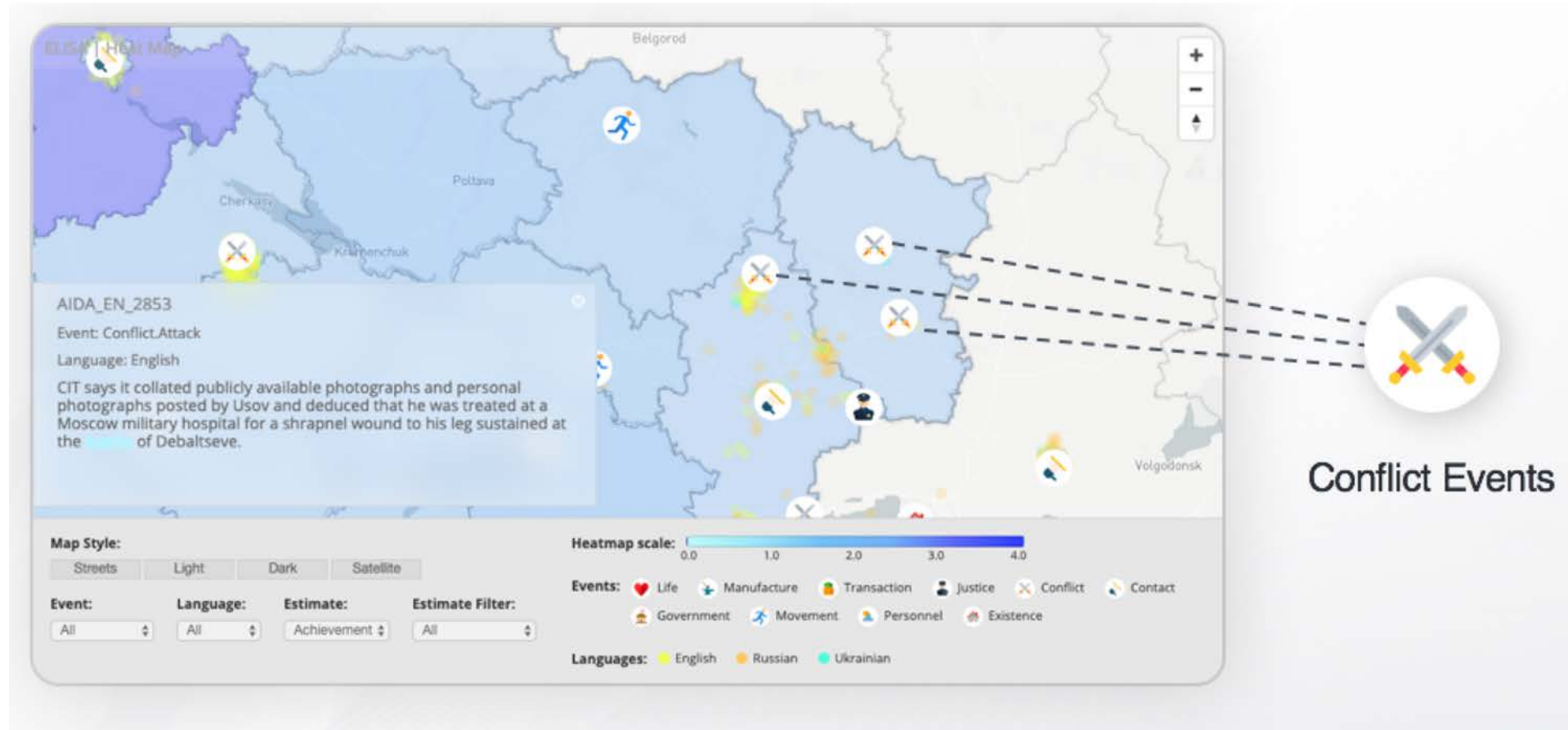
```

https://blender04.cs.rpi.edu/~lud2/video_recommendation_demo2019/an...
criminologists near the inlet on the side opposite the Maidan
5 The case of Kiev snipers question about the organizers and perpetrators of
sniper shooting at Euromaidan participants and at the same time law
enforcement officers in Kiev on February 20, 2014, which killed 53 people
(49 protesters and 4 law enforcement officers)
6 [one]
7 .
8 Protesters and new Ukrainian authorities accuse the Yanukovich regime of
these events
9 [2]
10 [3]
11 as well as Russia.
12 The Russian authorities and the then Ukrainian leaders, in turn, believe
that the shooting was organized by Euromaidan representatives to create
sacred sacrifices and exacerbate the revolutionary situation.
13 February 24th and. about. Interior Minister Arsen Avakov reported on his
page in one of the social networks about the initiation of a criminal case
into the massacres of civilians, in connection with which Yanukovich and
several other officials were wanted
14 [four]
15 .
16 The investigation also brought accusations to law enforcement officials
17 [five]
18 [6]
19 .
20 At the same time, in early March 2014, there were suggestions that other
  
```

Type of Attack: ConflictAttack

[https://blender04.cs.rpi.edu/~lud2/video\\_recommendation\\_demo2019/navigation\\_dark.html](https://blender04.cs.rpi.edu/~lud2/video_recommendation_demo2019/navigation_dark.html)

# Applications: Ukraine Event and Moral Value Map



- Achievement, Benevolence, Conformity, Hedonism, Power, Security, Self-direction, Simulation, Tradition and Universalism (Schwartz, 2012)

# English IE: Expensive but Generally Happy

Expert consumers  
define what to  
link/extract



Human Annotators  
annotate a large amount  
of documents



Researchers design  
features/rules, train  
supervised learning  
models



- *High Cost*: requires manual clean annotations for 500 documents
- *Poor Portability*: e.g., only covers 41 relation types and 33 event types
- Limited to a certain domain, genre, language, and data modality



# English doesn't deserve to be the center of language universe

- 6000+ living languages, 300+ languages have digital news data
- Certain information is often reported predominantly in local news in low-resource languages
  - e.g., the vast majority of Physical-Located relations and Meeting events involving Aung San Suu Kyi are only reported locally in Burmese news
  - e.g., language barrier was one of the main difficulties faced by humanitarian workers responding to the Ebola crisis in 2014
- Publicly available gold-standard annotations for IE exist for only a few languages
- Annotations for edge (relation and event) extraction are more expensive than node (entity) extraction because relations/events are structured and require a rich label space – not suitable for crowd-sourcing





# One Possible Solution: Share and Transfer



Enhance Quality with deep knowledge acquisition and reasoning



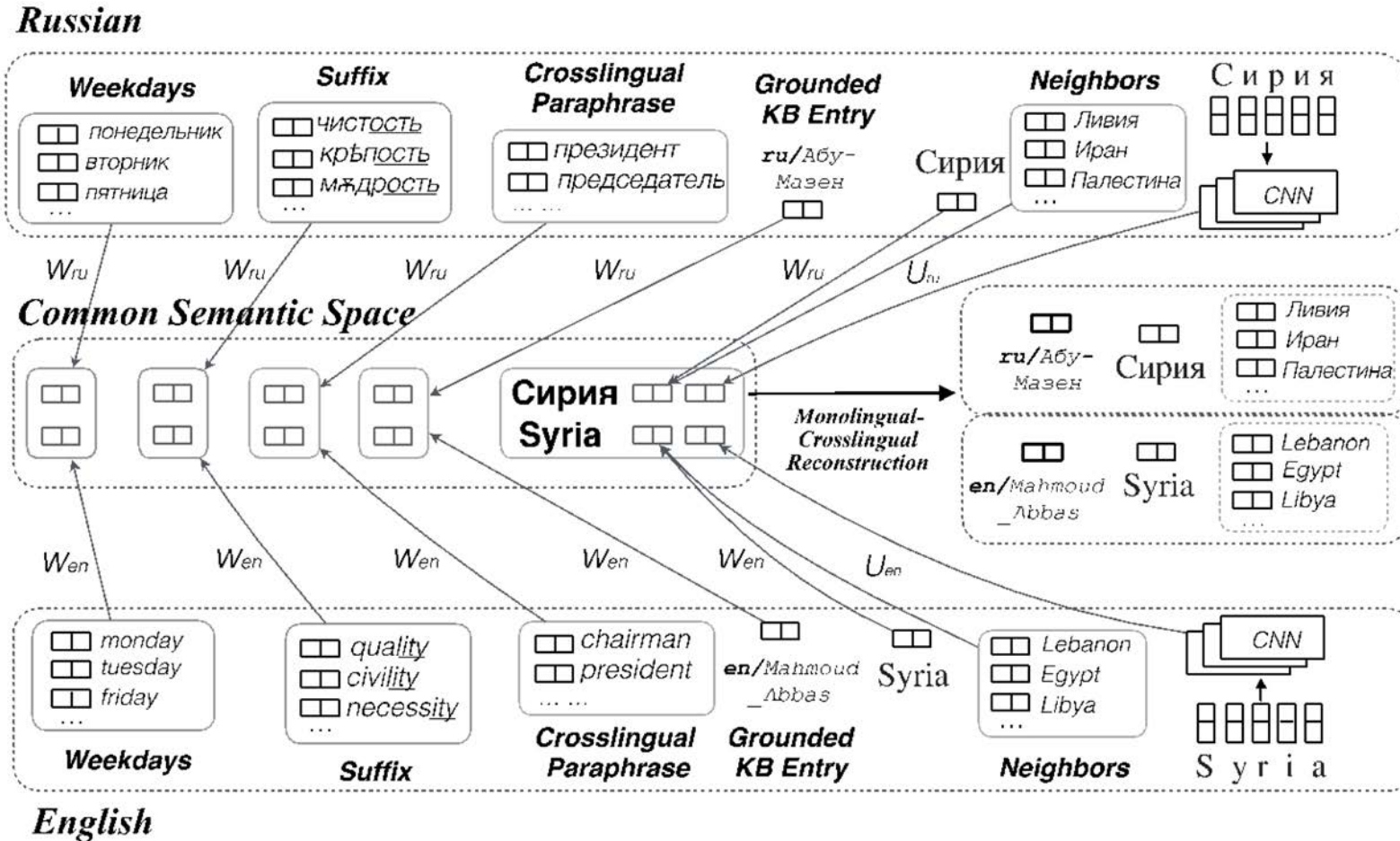
Transfer knowledge across domain/genre/language/data modality



# Cross-lingual Node (Entity) Transfer

- Leverage Language-Universal Non-Conventional Linguistic Resources
- Cross-lingual Embedding Representations
  - Cluster-consistent embedding: avoid using bi-lingual dictionaries or parallel corpora
  - Joint Entity and Word embedding
  - Cross-lingual language modeling for contextualized embedding
- Cross-lingual Transfer Learning
  - Multi-task Multi-lingual transfer learning
  - Adversarial learning to select language-universal resources and features
- Allow non-speakers to annotate any language

# Share: Multi-lingual Cluster-consistent Common Space Construction



- Our new hypothesis: Cluster distribution tends to be consistent across languages (Huang et al., EMNLP2018)

# Share: Cross-lingual Joint Entity and Word Embedding Learning

(Pan et al., DeepLo2019)

- Code-switch cross-lingual entity/word data generation

*Example Chinese Wikipedia Sentence:*

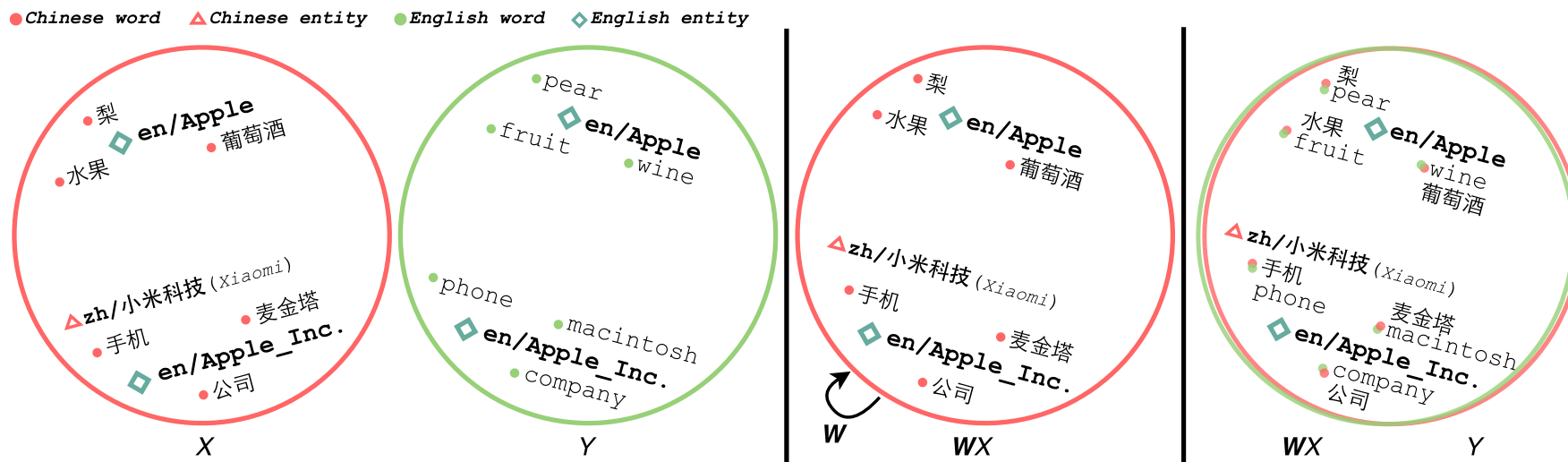
[[小米科技|小米]] 被誉为中国的 [[苹果公司|苹果]]。

link ↓      langlink      link ↓      langlink  
zh/小米科技 ~~→~~      zh/苹果公司 → en/Apple\_Inc.

*Our Approach:*

zh/小米科技 被 誉为 中国的 en/Apple\_Inc.。  
(Xiaomi)      (is) (known as) (Chinese)

- Use English entities as anchor points to learn a mapping (rotation matrix)  $W$

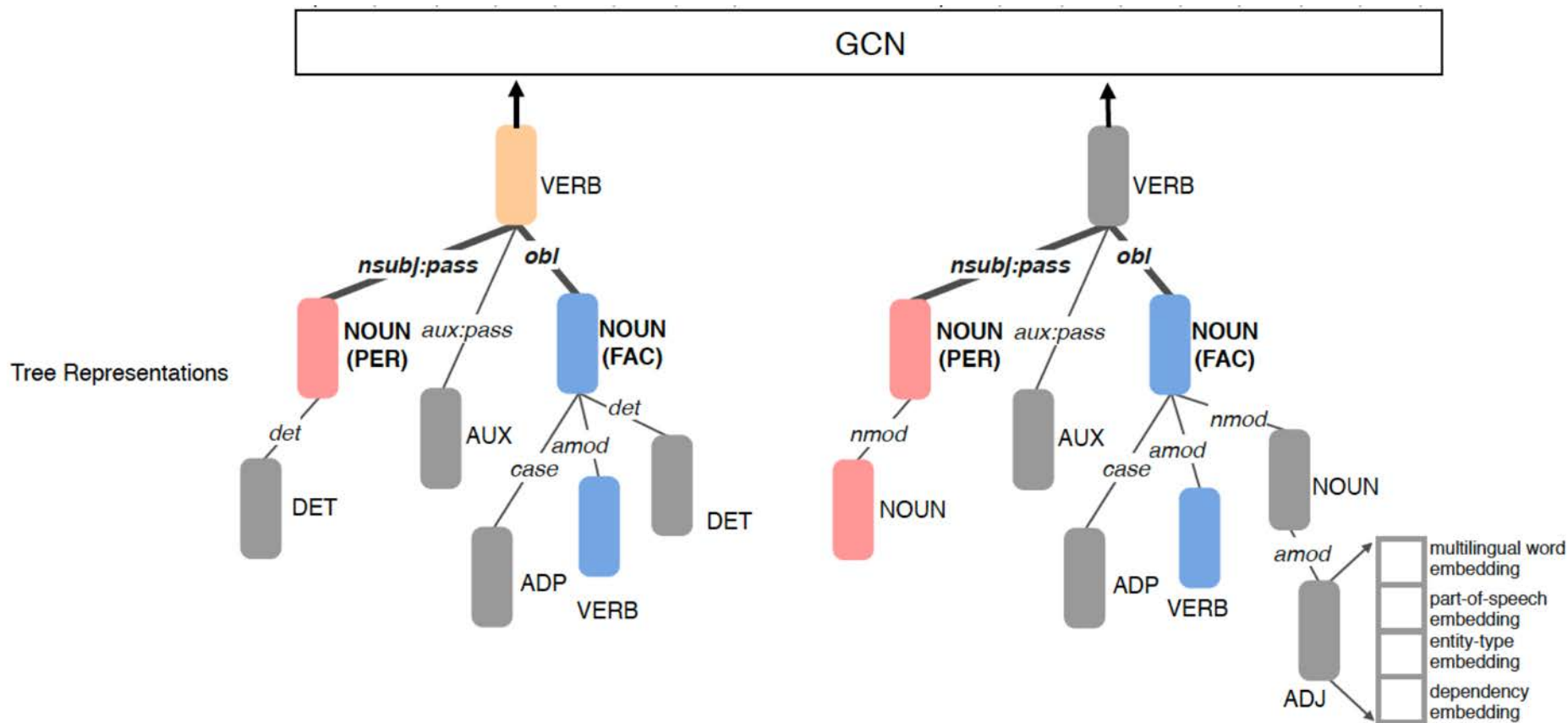




# Cross-lingual Edge (Relation and Event Argument) Transfer

- (Subburathinam et al., EMNLP2019)
- Hypothesis: Relational facts are typically expressed by identifiable patterns which are consistent across languages (Lin et al., 2017)
- Language-universal symbolic representations
  - POS tagging and dependency parsing are available for 84 languages (Nivre et al., 2018)
  - Entity extraction is available for 282 languages (Pan et al., 2017)
- Language-universal distributional representations
  - multi-lingual word embedding is available for 44 languages (Bojanowski et al., 2017; Joulin et al., 2018)

# Toward more Structured Representations



The **detainees** were **taken** to a **processing center**

**Команды врачей** были **замечены** в упакованных отделениях **скорой помощи**  
 (**teams of doctors** were seen in packed **emergency rooms**)

# Graph Convolutional Networks (GCN) Encoder

- Extend the monolingual design (Zhang et al., 2018) to cross-lingual
- Convert a sentence with N tokens into N\*N adjacency matrix A
- Node: token, each edge is a directed dependency edge
- Initialization of each node's representation

$$\mathbf{h}_i^{(0)} = \mathbf{x}_i^w \oplus \mathbf{x}_i^p \oplus \mathbf{x}_i^d \oplus \mathbf{x}_i^e$$

Word embedding   POS tag   Dependency relation   Entity type

- At the  $k^{\text{th}}$  layer, derive the hidden representation of each node from the representations of its neighbors at previous layer

$$\mathbf{h}_i^{(k)} = \text{ReLU} \left( \sum_{j=0}^N \frac{\mathbf{A}_{ij} \mathbf{W}^{(k)} \mathbf{h}_j^{(k-1)}}{d_i + b^{(k)}} \right)$$

# Application on Relation Extraction

- Task: Classify each pair of entity mentions into one of pre-defined relation types or NONE
- Max-pooling over the final node representations to obtain representations for sentence and two entity mentions, and concatenate them
- A softmax output layer for relation type classification

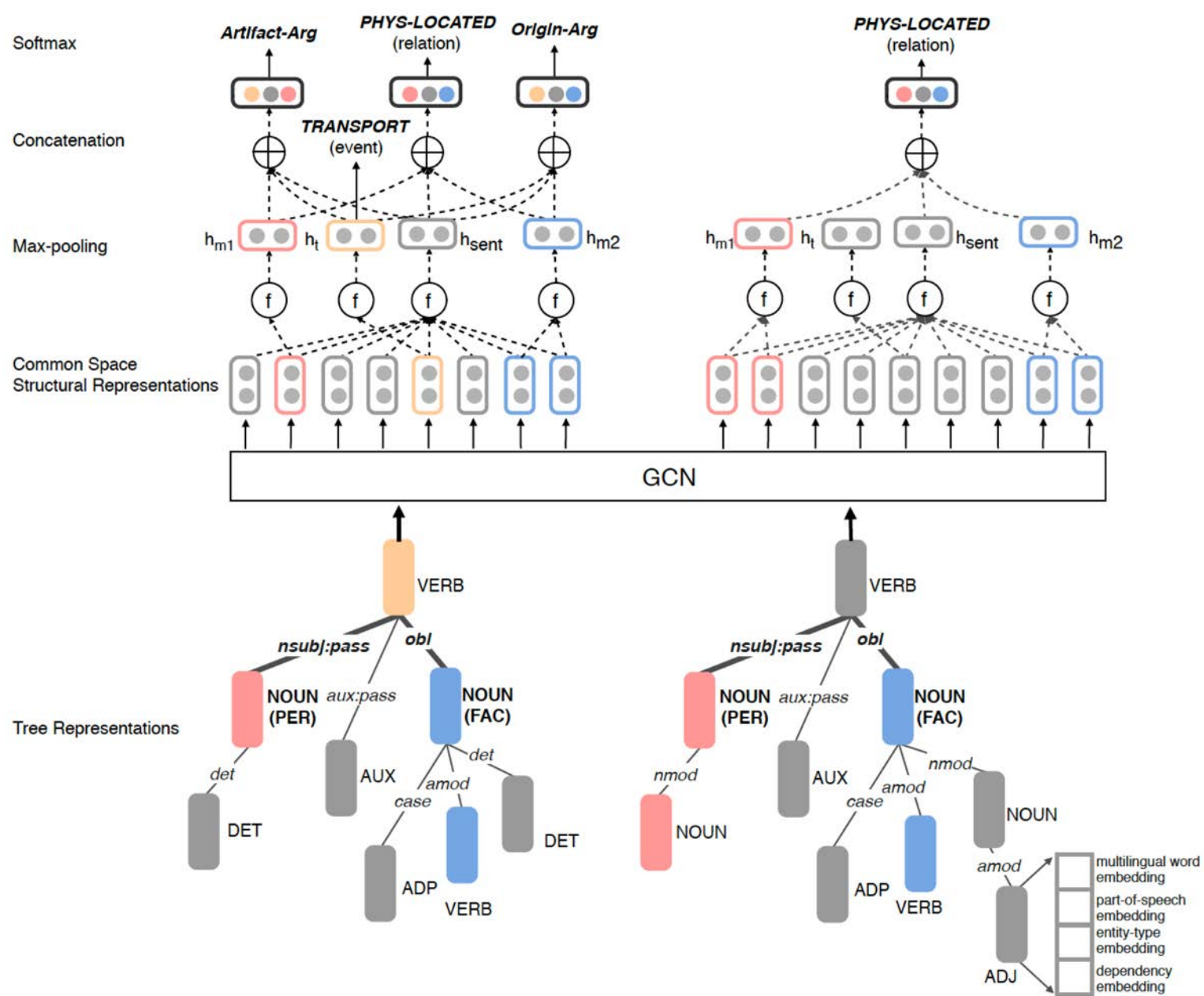
$$L^r = \sum_{i=1}^N \sum_{j=1}^{L_i} y_{ij} \log(\sigma(\mathbf{U}^r \cdot [\mathbf{h}_i^{m1}; \mathbf{h}_{ij}^s; \mathbf{h}_j^{m2}]))$$



# Application on Event Argument Extraction

- Task: Classify each pair of event trigger and entity mentions into one of pre-defined event argument roles or NONE
- Max-pooling over the final node representations to obtain representations for sentence, trigger and argument candidate, and concatenate them
- A softmax output layer for argument role labeling

$$L^a = \sum_{i=1}^N \sum_{j=1}^{L_i} y_{ij} \log(\sigma(\mathbf{U}^a \cdot [\mathbf{h}_i^t; \mathbf{h}_{ij}^s; \mathbf{h}_j^a]))$$



The **detainees** were **taken** to a **processing center**

**Команды врачей** были **замечены** в упакованных отделениях **скорой помощи**  
 (**teams of doctors** were seen in packed **emergency rooms**)

# Overall Performance

- Relation Extraction

Train \ Test	English	Chinese	Arabic
English	68.2	<b>42.5</b>	<b>58.7</b>
Chinese	<b>62.6</b>	69.4	54.0
Arabic	58.6	35.2	67.4

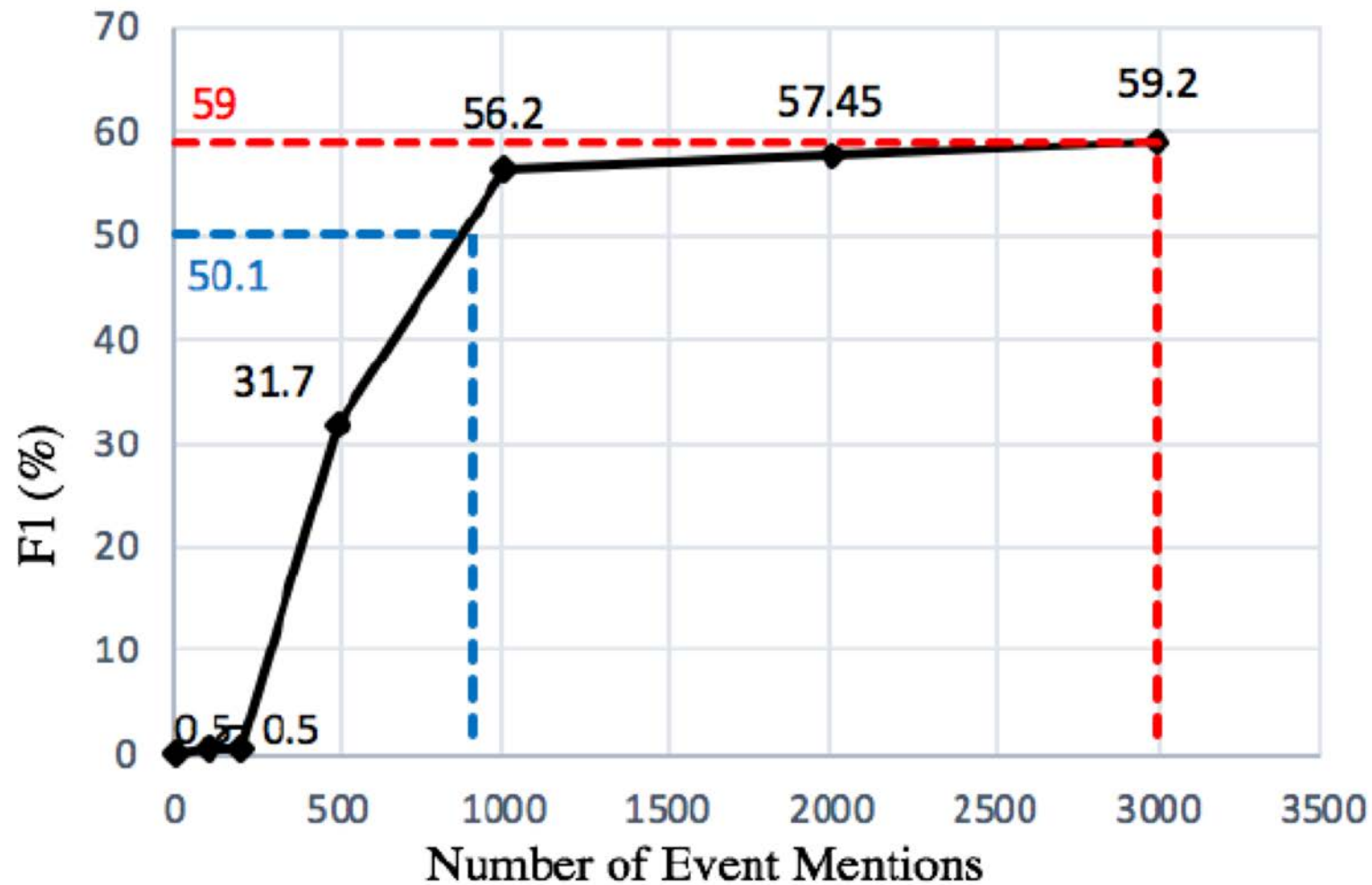
- Event Argument Extraction

Train \ Test	English	Chinese	Arabic
English	63.9	<b>59.0</b>	<b>61.8</b>
Chinese	<b>51.6</b>	59.3	60.6
Arabic	43.1	50.1	64.0
English + Chinese	—	—	<b>63.1</b>
English + Arabic	—	<b>60.1</b>	—
Chinese + Arabic	<b>51.9</b>	—	—

Much better than state-of-the-art on cross-lingual transfer (47.7%, Hsi et al., 2016)

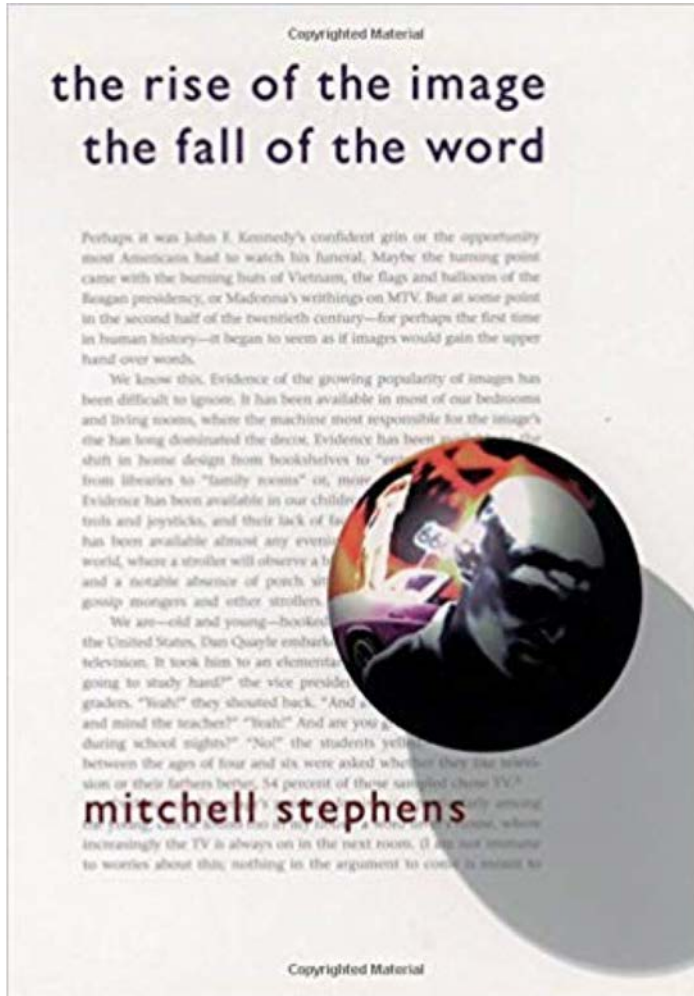
# Comparison with Supervised Approach

- Chinese Event Argument Extraction





# Ongoing Work: Cross-media Event Structure Transfer



- We produce and consume news content simultaneously through different media
- By randomly checking 100 multimedia documents from Voice of America, 34% images contain visual objects that serve as event arguments which are not mentioned in surrounding texts



*Vote\_? = soldier*



*TransportPerson\_Instrument = stretcher*

# Multimedia Common Space Construction

- Treat Image/Video as a foreign language

<b>Language</b>	<b>Vision</b>
Word Sense/Phrase	Image Region
Entity	Visual Concept
Relation	Visual Relation
Entity-Relation Graph	Scene Graph
Event	Visual Event
AMR and Dependency Graph	Visual Semantic Graph



# A New Task: Multimedia Event Extraction

- Input: News Article & Events

In March , Turkish forces escalated **attacks** [**Conflict.Attack**] on the YPG in northern Syria , forcing U.S. to **deploy** [**Movement.TransportPerson**] a small number of forces in and around the town of Manbij to the northwest of Raqqa to “deter” Turkish - SDF clashes and ensure the focus remains on Islamic State. Meanwhile, Raqqa is being pummeled by **airstrikes** [**Conflict.Attack**] mounted by U.S.-led coalition forces and Syrian warplanes. Local anti-IS activists say the air **raids** [**Conflict.Attack**] fail to distinguish between military and non-military targets ..

Input: Image & Visual Objects



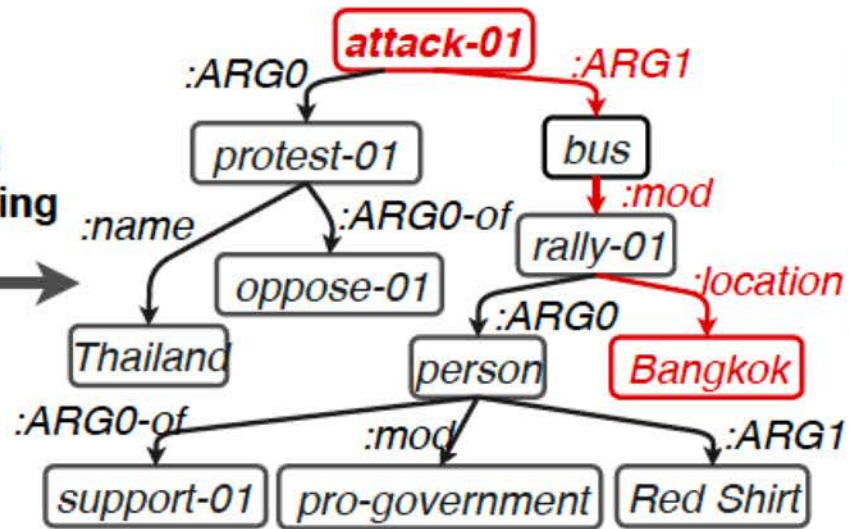
Output: Image-related Events & Visual Argument Roles

Visual Events	airstrikes [ <b>Conflict.Attack</b> ] raids [ <b>Conflict.Attack</b> ]	
Visual Arguments	Target	airplane
	Target	vehicle

# Multi-media Structured Common Space Construction

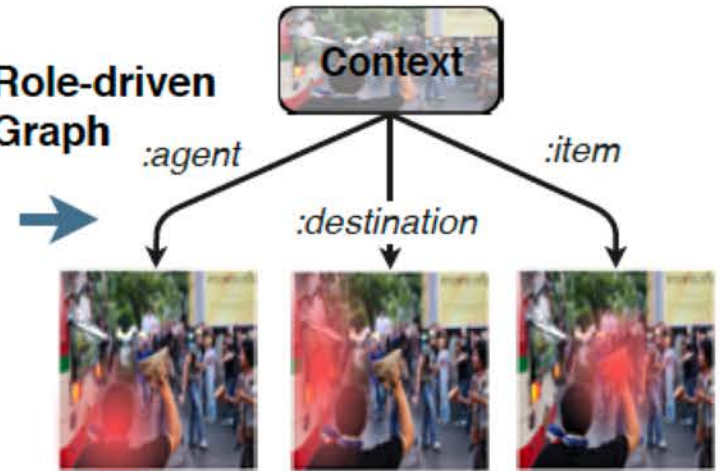
Thai opposition protesters **attack** a bus carrying pro-government Red Shirt supporters on their way to a rally at a stadium in **Bangkok**

AMR Parsing



ejecting	
agent	man
destination	bus
item	stone

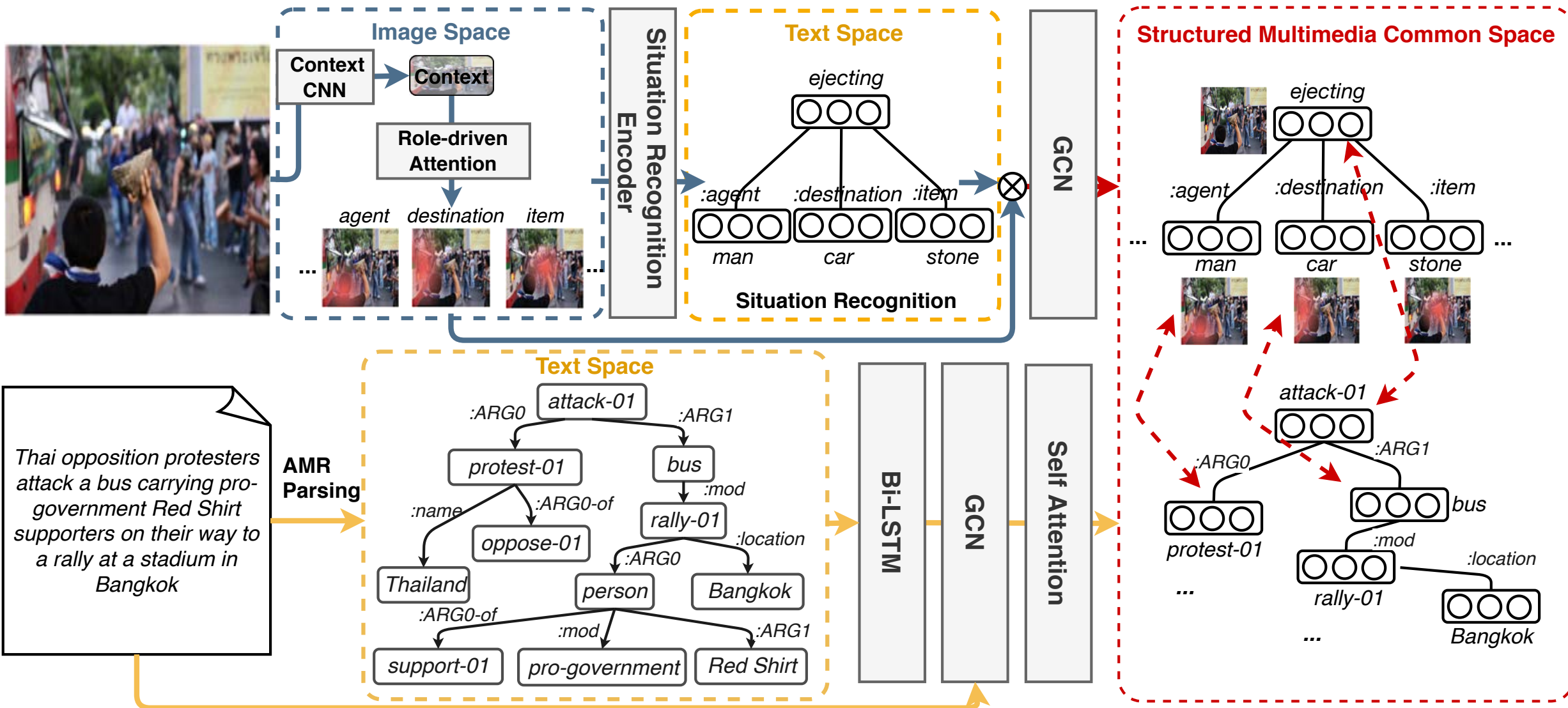
Role-driven Graph





# Weakly Aligned Structured Embedding (WASE)

-- Training Phase (Common Space Construction)



# Weakly Aligned Structured Embedding (WASE)

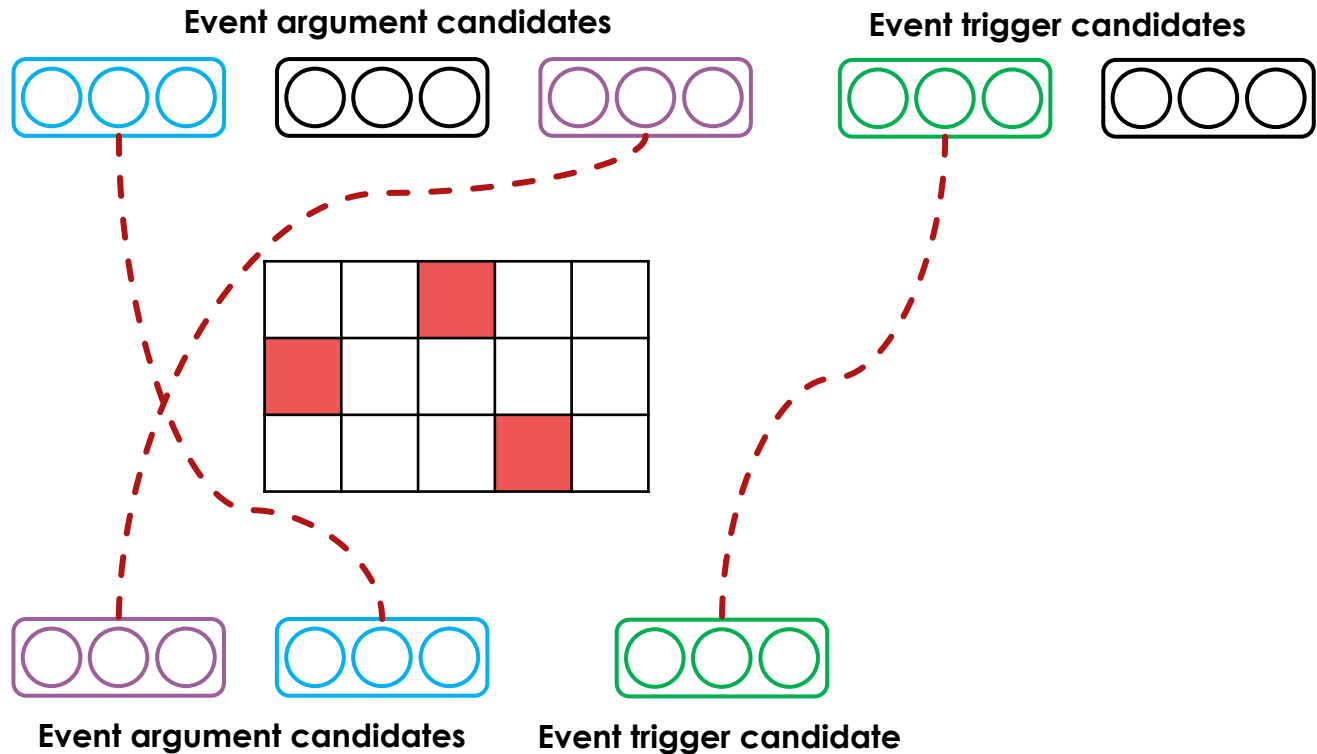
-- Training Phase (Common Space Construction)

Thai opposition protesters attack a bus carrying pro-government Red Shirt supporters on their way to a rally at a stadium in Bangkok



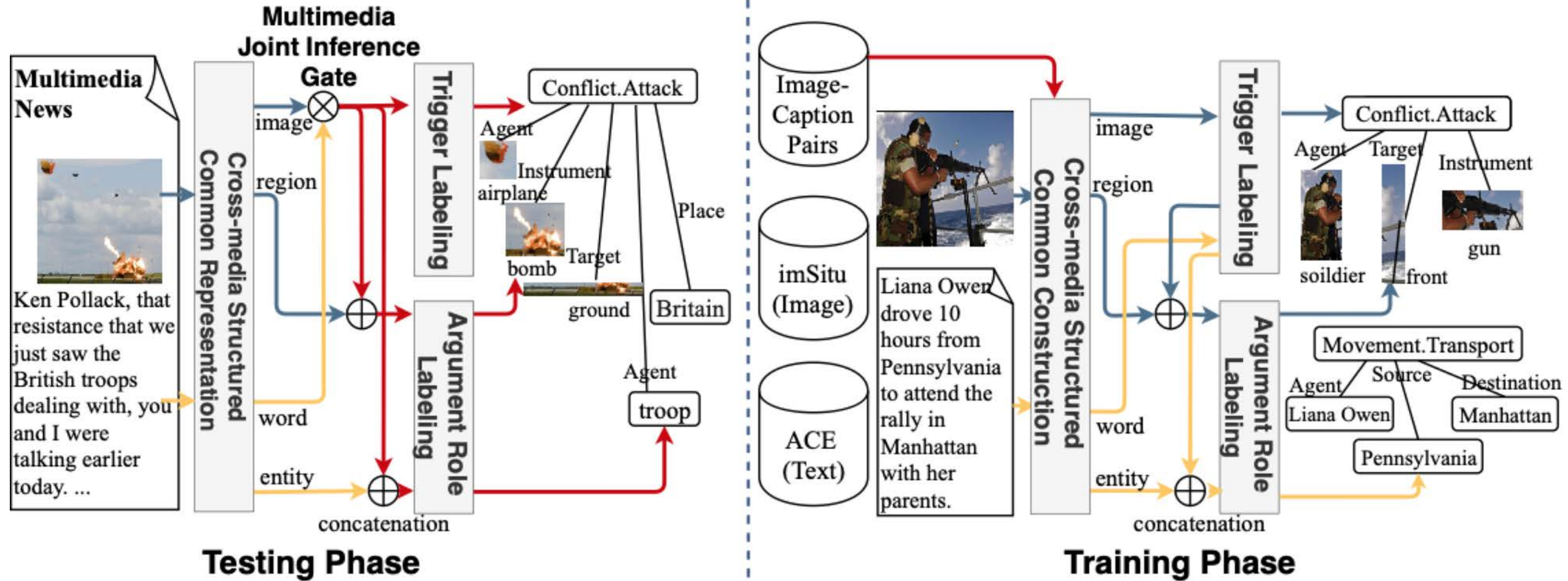
WASE (text)

WASE (img)



# Weakly Aligned Structured Embedding (WASE)

-- Testing & Training Phase



# Experimental Results

- Compare with text-only extraction (ACE) and image-only extraction (Situation Recognition)

Method	Text Event Trigger Labeling			Visual Argument Role Labeling		
	P(%)	R(%)	F <sub>1</sub> (%)	P(%)	R(%)	F <sub>1</sub> (%)
JMEE (Text-only) [1]	49.12	72.64	58.61	-	-	-
FC_Graph(Vision-only) [2]	-	-	-	19.15	19.15	19.15
Multimedia Flat Embedding	56.08	62.03	58.91	32.48	34.24	33.34
WASE	59.55	66.43	62.80	32.51	34.29	33.36

[1] Xiao Liu, Zhunchen Luo, and Heyan Huang. 2018a. Jointly multiple events extraction via attention-based graph information aggregation. EMNLP 2018

[2] Ruiyu Li, Makarand Tapaswi, Renjie Liao, Jiaya Jia, Raquel Urtasun, and Sanja Fidler. 2017. Situation recognition with graph neural networks. ICCV 2017



# Compare to State-of-the-art Cross-media Flat Representation



- ▶ Baseline:
  - ▶ Event: Justice:Arrest-Jail
  - ▶ Roles:
    - ▶ None
- ▶ Our Approach:
  - ▶ Event: Conflict.Attack
  - ▶ Roles:
    - ▶ Instrument = weapon



- ▶ Baseline:
  - ▶ Event: Justice:Arrest-Jail
  - ▶ Roles:
    - ▶ Entity = man
- ▶ Our Approach:
  - ▶ Event: Conflict:Demonstrate
  - ▶ Roles:
    - ▶ Entity = man



- ▶ Baseline:
  - ▶ Event: Justice:Arrest-Jail
  - ▶ Roles:
    - ▶ Agent = man
- ▶ Our Approach:
  - ▶ Event: Justice:Arrest-Jail
  - ▶ Roles:
    - ▶ Person = man

# Promising Progress

		2015	2019
Portability	# Languages for EDL and Event Extraction	1-3	300
	# Entity types	5	16,000
	# Relation types (English)	41	2,000
	# Event types (English)	33	1,000
Quality (Low-resource Languages without gold standard annotations)	Name Tagging	0%	Up to 76% F-score
	Cross-lingual Entity Linking	Up to 16% absolute improvement in accuracy	
Development Time		Half a year	1-10 hours
Cost		Supervised models based on 500 fully annotated documents	No manual annotation required for new language/domain

# Remaining Challenges: Acquire background knowledge

- The final Perahera of the **Ruhunu Kataragama Maha Devalaya** will be held today.
- In the communiqué the education ministry has cited as a cases in point several instances like the application by a doctor transferred to Bemmulla in Gampaha for admission of his child to the Colombo **D . S . Senanayake Vidyalaya**.
- The navy media unit stated that they suspect that the **Kerala Ganja Cannabis** was brought from India via the mainland
- When we come back media speculation run amuck over possible indictments at **sixteen hundred Pennsylvania** and the President 's scripted session with troops in Iraq .



D. S. Senanayake College



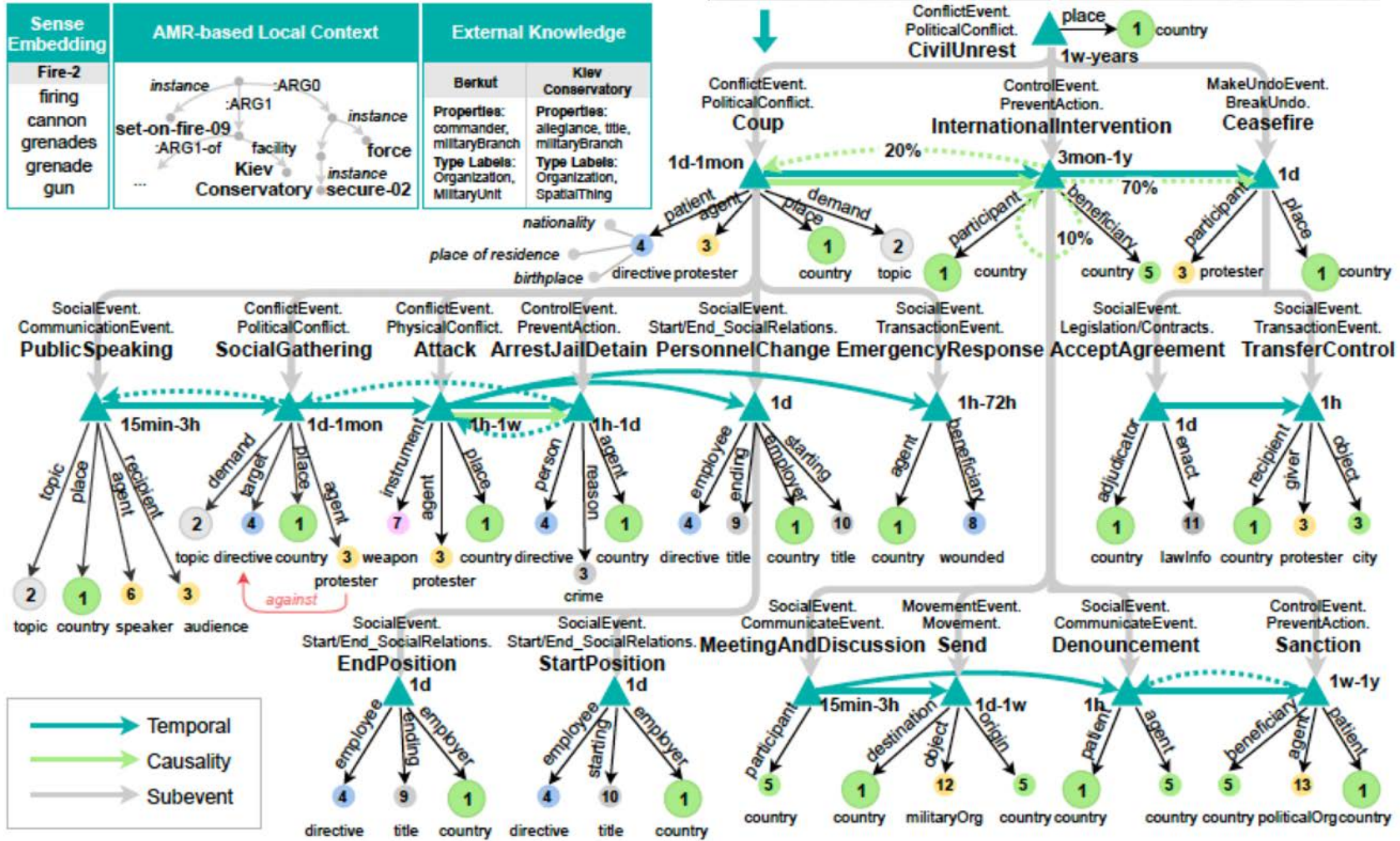


# Remaining Challenges: Cross-scenario Transfer through Complex Event Schema Induction

2014 Thai coup d'état: Однако протесты и блокада длятся уже почти 3 месяца, а военные так и не перешли к действиям  
 2013 Egyptian coup d'état: ... General Abdel Fattah el-Sisi announced that he there would be calling newpresidentialand Shura Council elections.  
 Ukrainian crisis: At 09:25, protesters pushed the Berkut back to the October Palace after security forces tried to set fire to Kiev Conservatory, which was being used as a field hospital for wounded protesters.



Sense Embedding	AMR-based Local Context	External Knowledge						
Fire-2 firing cannon grenades grenade gun	<pre> instance   :ARG0   :ARG1   set-on-fire-09   facility   Kiev   Conservatory   force   instance   :secure-02           </pre>	<table border="1"> <tr> <th>Berkut</th> <th>Kiev Conservatory</th> </tr> <tr> <td>Properties: commander, militaryBranch</td> <td>Properties: allegiance, title, militaryBranch</td> </tr> <tr> <td>Type Labels: Organization, MilitaryUnit</td> <td>Type Labels: Organization, SpatialThing</td> </tr> </table>	Berkut	Kiev Conservatory	Properties: commander, militaryBranch	Properties: allegiance, title, militaryBranch	Type Labels: Organization, MilitaryUnit	Type Labels: Organization, SpatialThing
Berkut	Kiev Conservatory							
Properties: commander, militaryBranch	Properties: allegiance, title, militaryBranch							
Type Labels: Organization, MilitaryUnit	Type Labels: Organization, SpatialThing							



# Ongoing Research Project: Let's Write a History Book!

- History is written by the victors
- Entity-centric knowledge base is much easier to construct than event-centric one
  - Who is Barack Obama's wife?
- Requires exhaustive search and information aggregation
  - How many people died in 911? Who are they?
- Multi-view verification
  - What happened in 1989 Tiananmen Square?
  - What happened in Hong Kong Protest 2019?
- It's international human right to know what happened in the history
- We aim to create a history book automatically so it can be more complete and authentic than human created ones



Attacker = protesters



Attacker = police

vs.



Attacker = protesters



Attacker = police

vs.



# What will such a History Book look like?

- Each chapter looks like a Wikipedia page
  - The description is organized by multimedia timeline with detailed source and evidence information, links to original news articles
  - Detailed participants (arguments) and their roles, and their connections and relations
  - Infobox shows event-event relations: temporal, causal and hierarchical
- Building blocks
  - Open-domain Multilingual Multimedia Event Extraction
  - Cross-document Event Coreference Resolution and Ranking
  - Event Schema Induction and Event-Event Relation Extraction
  - Cross-source Information Verification and Truth Finding
  - Abstract Summarization for Event Description
- Never-ending updated over time; put up to the wild for human editing and curation

**Protests against the government** [ edit ]

**4 December 2011** [ edit ]

On 4 November 2011, during the annual [Russian March](#) event, representatives of "[The Russians](#)" movement declared a protest action unapproved and took place on 4 December at 21:00 in Moscow. The statement of non-recognition of electoral results spread widely. Citizens said to have occurred during the elections. Alexander Belov declared the beginning of the "Putin, go away!" campaign.<sup>[37]</sup> The protest [Alexander Belov](#), [Dmitry Dyomushkin](#), [George Borovikov](#) were arrested along with dozens of other nationalists. The head of the banned observer. Mass detentions of other public organizations occurred in Moscow. According to police some 258 persons have been detained.

**5–7 December 2011** [ edit ]

On 5 December, around 5,000 opponents of the government began protesting in Moscow, denouncing [Vladimir Putin](#) and his government. Putin step down, whilst some demanded revolution.<sup>[15][39]</sup> [Alexey Navalny](#), a top blogger and anti-corruption activist who branded Putin's postings on his [LiveJournal](#) blog and Twitter account. Navalny's agitation was denounced by United Russia as "typical dirty self-promotion account."<sup>[40][41]</sup>

Many pro-government supporters, including the pro-Putin youth group [Nashi](#), were mobilized on 6 December at the site of the planned [Nashi on Manezhnaya Square](#)<sup>[43]</sup> and an 8,000-strong rally of the [Young Guard on Revolution Square](#).<sup>[44]</sup> About 500 pro-United Russia expected anti-government protests. It emerged that 300 protesters had been arrested in Moscow the night before, along with 120 in St. Petersburg. Slogans against Putin,<sup>[15]</sup> whilst anti-government protesters at Revolution Square clashed with riot police and [interior ministry troops](#). The [Bolotnaya Square](#) and dozens of arrests were reported, including [Boris Nemtsov](#), an opposition leader and former deputy prime minister,<sup>[48]</sup> and [Igor Yavlinsky](#) charged. At least one Russian journalist claimed he was beaten by police officers who stamped on him and hit his legs with batons.<sup>[50]</sup> After three and a half hours, the Moscow protest came to an end.<sup>[51]</sup>

Attempts to stage a large protest in Moscow on 7 December fizzled out due to a large police presence in the city.<sup>[13]</sup>

**10 December 2011** [ edit ]

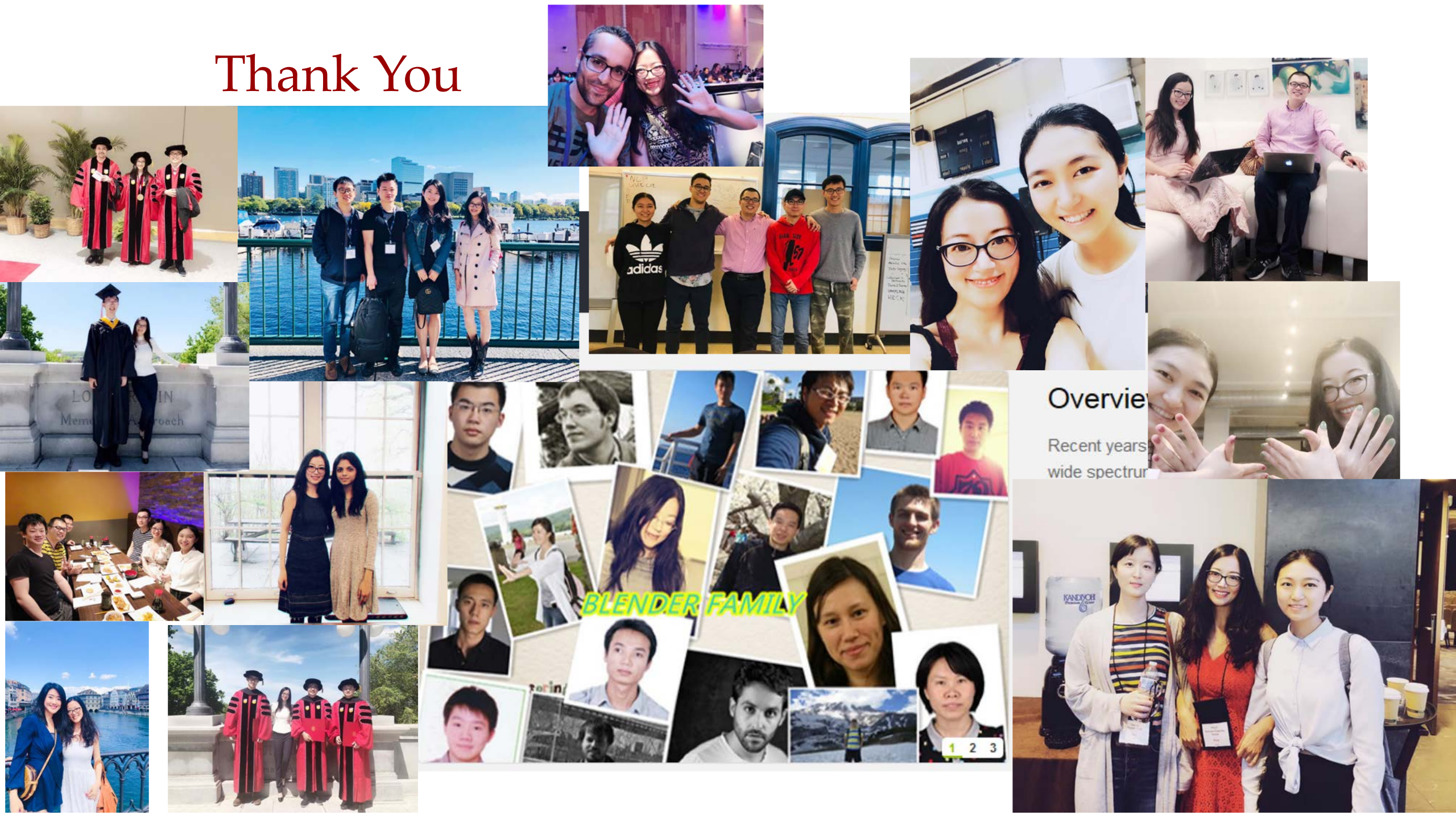


Rally in [Bolotnaya Square](#) in Moscow on 10 December 2011

Via a Facebook group "[Суббота на Болотной площади](#)" (Saturday at Bolotnaya Square),<sup>[52]</sup> [Bolotnaya Square](#) on 10 December.<sup>[53][54]</sup> Prior to the demonstration newspapers commented that tens of thousands of people gathered in Moscow,<sup>[55][56]</sup> and, similarly, over 5,000 in St. Petersburg.<sup>[57]</sup> A permit had originally been issued for the demonstration on Bolotnaya Square. By 8 December, more than 30,000<sup>[52]</sup> had accepted the Facebook invitation to attend. The demonstration was authorized by the Moscow government for the demonstration which took place on 10 December. It was made by Putin that police and security forces would be deployed to deal with anyone participating in the demonstration. The demonstration was peaceful and without attempts by the state to prevent or disrupt it.<sup>[59][60]</sup> Rapper [Noize MC](#) performed at the demonstration, specially from Paris for the occasion.<sup>[61]</sup> [Guerrilla theater](#) by [FEMEN](#) and the circulation of a poster with the slogan "Russia is a prison" were also part of the protests.<sup>[62][63]</sup>



# Thank You



# Some IE Demos



# Cross-lingual Entity Discovery and Linking for 287 Languages

Demo APIs Development Human In Loop Annotation Tool Heat Map Name Transliteration

## Incident Language IE Demo

Choose a language: Abkhazian

Language List (Pan et al., ACL2017)

Abkhazian	Achinese	Adyghe	Afrikaans	Akan	Albanian
Alemannisch	Amharic	Arabic	Aragonese	Aramaic	Armenian
Aromanian	Arpitan	Assamese	Asturian	Avaric	Aymara
Azerbaijani	Bambara	Bangla	Banjar	Basa Banyumasan	Bashkir
Basque	Bavarian	Belarusian	Belarusian (Tara...	Bihari	Bikol Central
Bishnupriya	Bislama	Bosnian	Breton	Buginese	Bulgarian
Burmese	Catalan	Cebuano	Central Kurdish	Chamorro	Chavacano de Za...
Chechen	Cherokee	Cheyenne	Chinese	Chinese (Min Nan)	Church Slavic

example 1 example 2 example 3 enter

```
# REDIRECT Атыкæицæда  
' ' ' Рашибуж ' ' ' ( ' ' ' ' ) Полша акалакъ .  
' ' ' Лос-Анџелес ' ' ' ( ' ' Los Angeles ' ' ) -  
Гæдоута араион .  
* Монтевидео ( проектируется )  
* Лыхны  
' ' ' Владимир ' ' ' - Урыстæвта акалакъ .  
# Стамчыра араион
```

System Missing Spurious

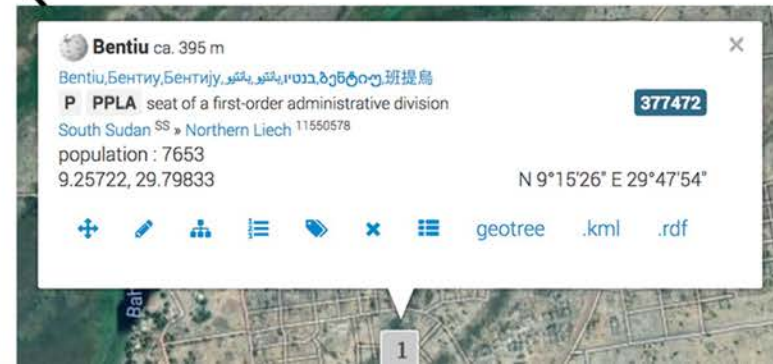
Mention ID: Boliang\_NE\_tagger\_00010  
Mention Str: Stephane Dujarric 601:617  
Reference KB: St%C3%A9phane\_Dujarric  
Entity Type: PER  
Mention Type: NAM  
Confidence Value: 1.0  
Images:

Extracted Name

Wikipedia Entry

Retrieved Image

GeoNames



# Cross-document Event Tracking

Timeline Home Examples System

Entity Name: All Event Type: All Event SubType: All Argument Role: All Start Date: End Date: Search

Legend: Argument Trigger Time Timeline Scale: Year Month

March 2011

## Life.Die

Independent groups say 113 men and 24 women have set themselves on fire since March 2011 and most have died of their burns.

- Victim= {their , themselves , most }

Source Doc: HC0007FAT

Timeline JS



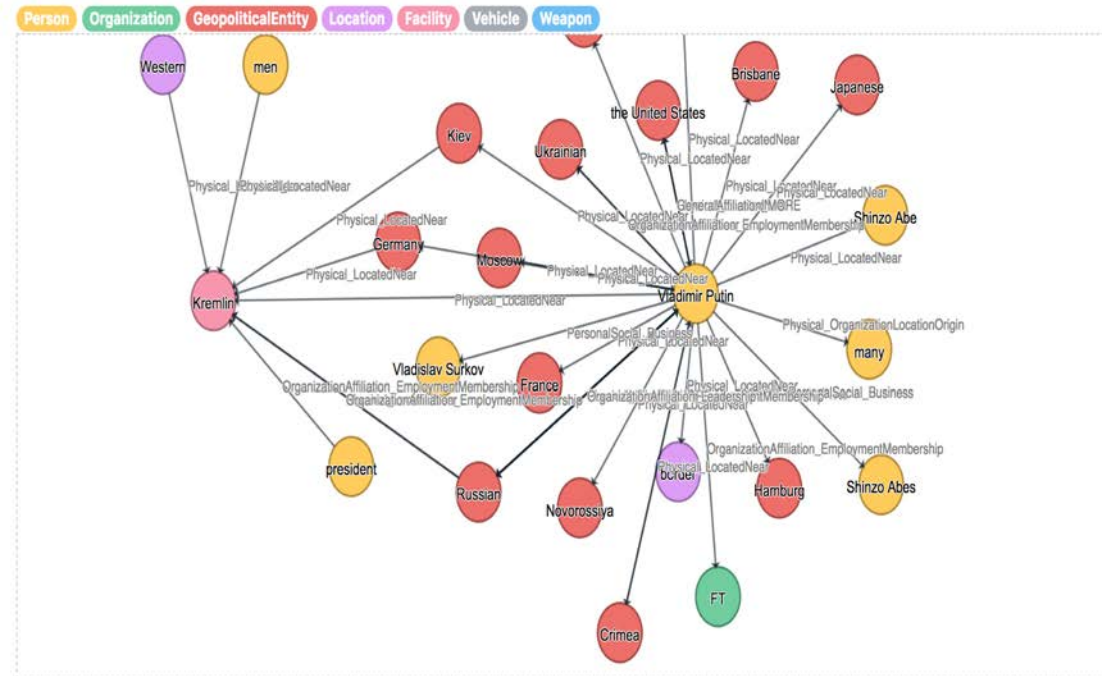
# Cross-document Entity Relation Tracking

Putin

Search

Name Search  Cypher Query(Example Cypher Queries)

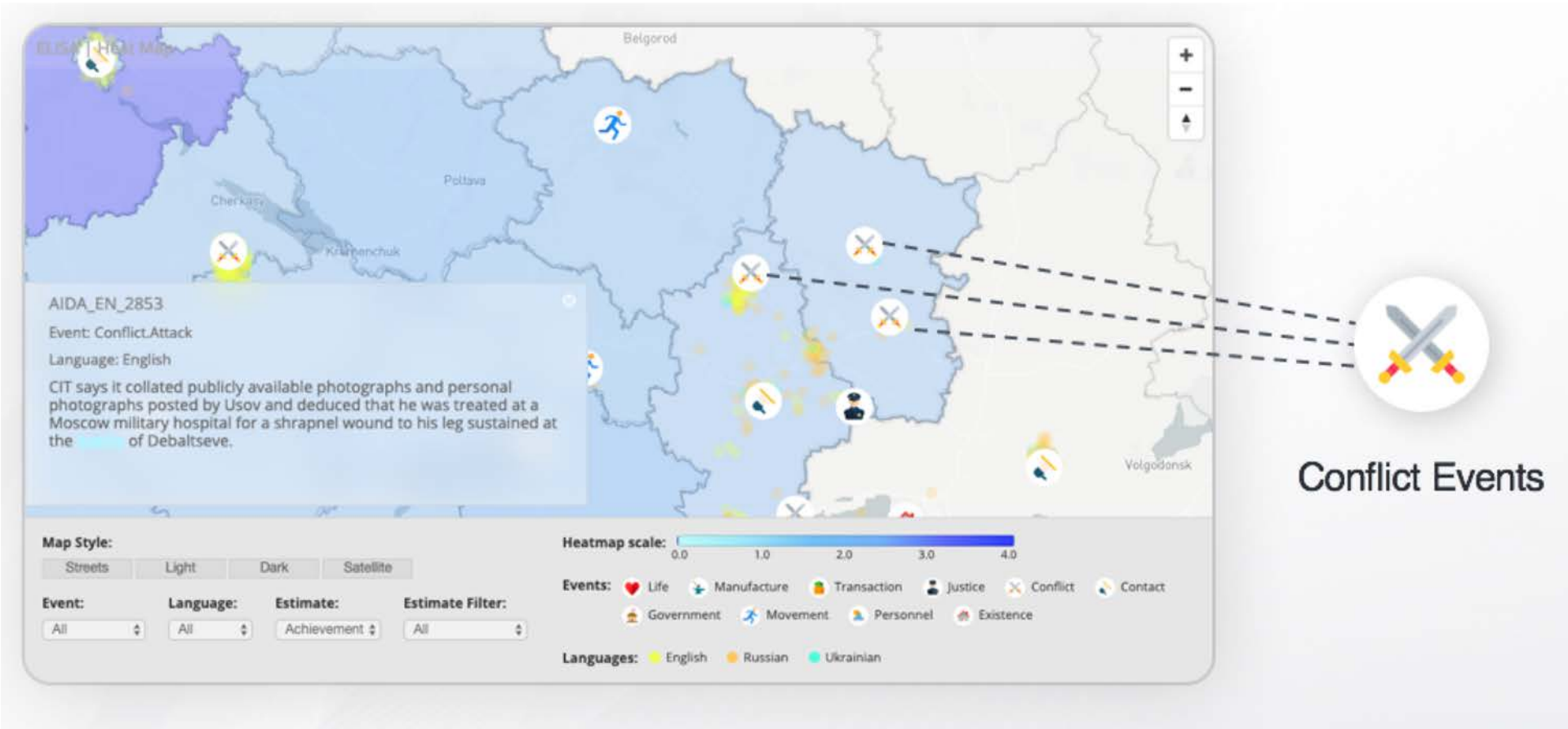
Graph Visualization: **Double Click** a node to see its relations, **Click** a node and then the CROSS icon to remove it from visualization. Zoom in and out by mouse wheel.



Properties: **Hover** over nodes and edges to see their properties here.

Type	ID	Name	Link	Mention
Person	:Entity_EDL_ENG_0153461	Vladimir Putin	LDC2015E42:m.08193	liberals,#Putin,Putin,own,President,gunmen,Mr,he,Vladimir Putin,Vladimir Putins,companies,murderers,victims,that,He,putin's,them,himself,group

# Ukraine Event and Moral Value Map



- Achievement, Benevolence, Conformity, Hedonism, Power, Security, Self-direction, Simulation, Tradition and Universalism (Schwartz, 2012)  
<http://162.243.120.148:8080/>

# Event Recommendation

Violence subsided in northern **Indian** cities on Tuesday as troops patrolled streets and a curfew remained in place after riots broke out a day earlier over the killing of a Sikh cleric in **Austria**.

Similar Events | Dissimilar Events | Suggest

Date	Location	Target of Protest	Protesters	Type of Protest	Demands of Protest
20090526	India	Unknown	Unknown	Riot	Religious

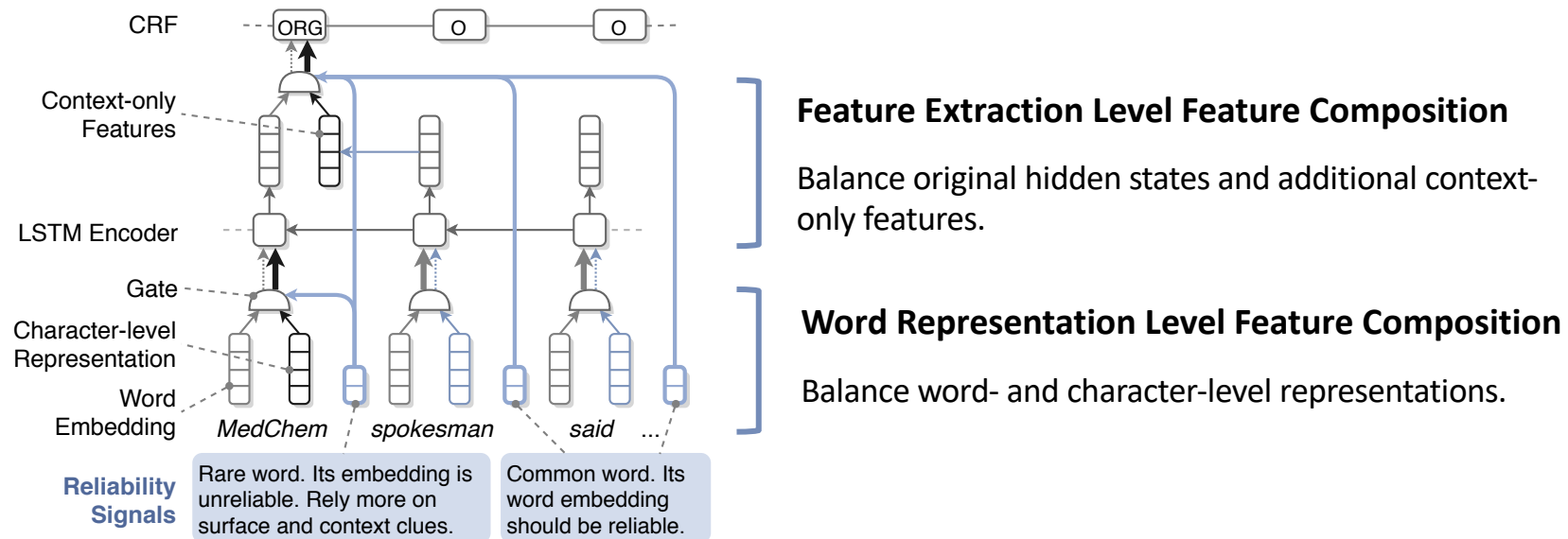
Date	Location	Target of Protest	Protesters	Type of Protest	Demands of Protest
20091231	Bauchi, Nigeria	Nigeria government	Unknown	Riot	Religious
20100911	Afghanistan	United States government	Unknown	Riot	Religious
20090706	Xinjiang, China	Unknown	Exiled Uighurs	Riot	Race
20100305	Athens, Greece				

explosion fire, anchorperson, demonstration or protest, person on street, fighting on street

**Cross-media coreference and inference**

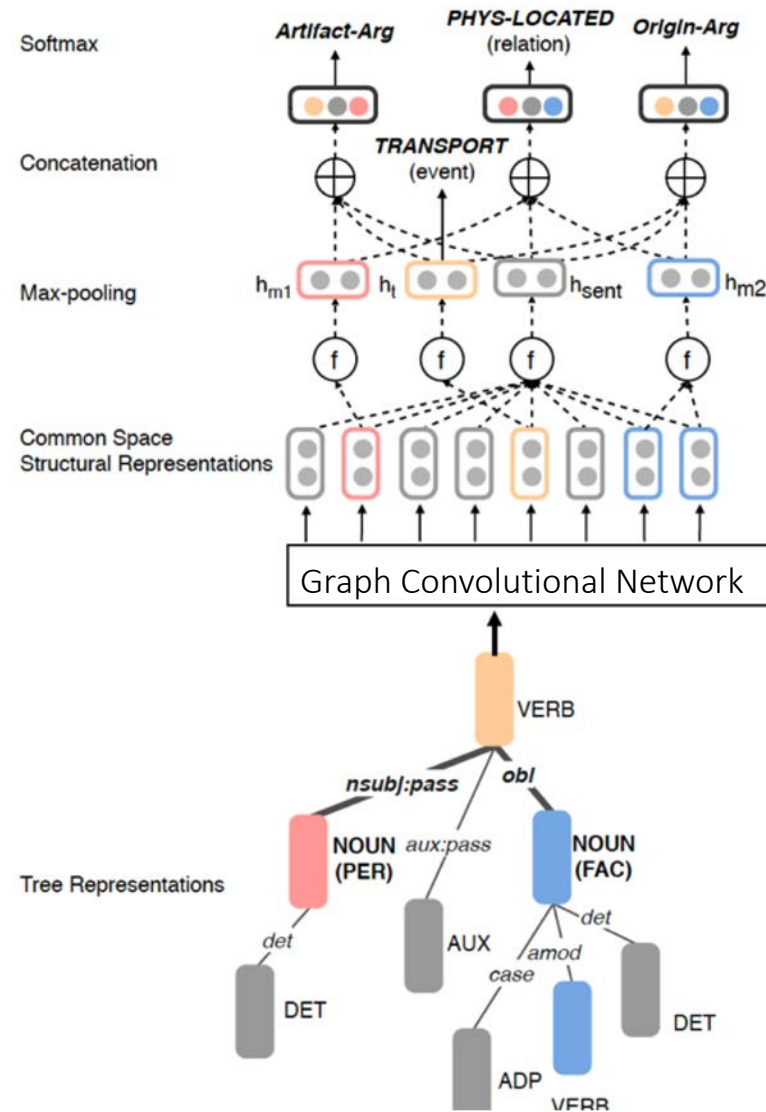
[https://blender04.cs.rpi.edu/~lud2/video\\_recommendation\\_demo2019/navigation\\_dark.html](https://blender04.cs.rpi.edu/~lud2/video_recommendation_demo2019/navigation_dark.html)

# Node (Entity) Extraction: Token Labeling with Sequence as Context



A **MedChem** spokesman said the products contribute about a third of **MedChem**'s sales and 10% to 20% of its earnings

# Edge (Relation and Event) Extraction: Node Pair Classification with Graph Context

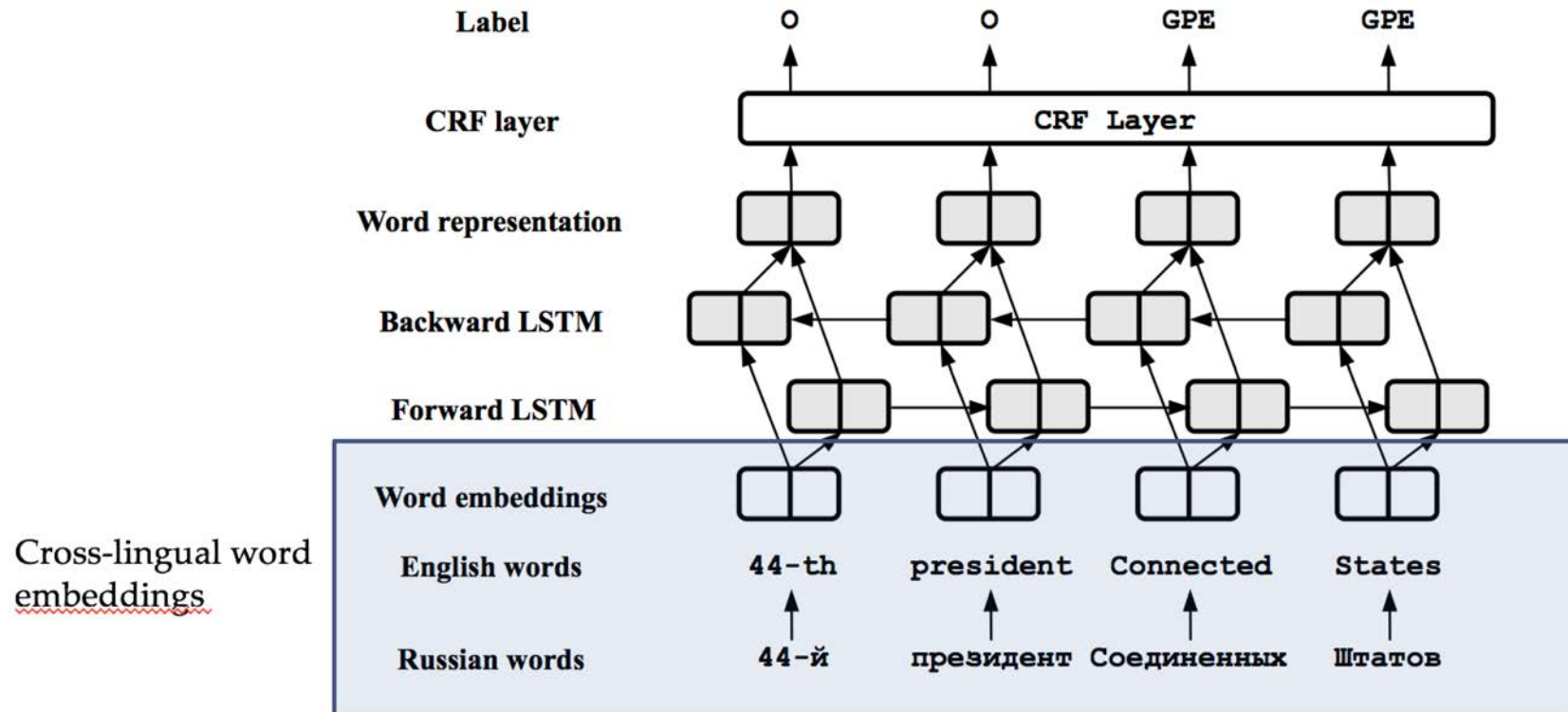


The detainees were taken to the processing center



# Previous Work

- Multilingual common space construction makes use of linear mappings (Mikolov et al., 2013; Rothe et al., 2016; Zhang et al., 2016; Baroni et al., 2015; Xing et al., 2015; Smith et al., 2017) or canonical correlation analysis (CCA) (Ammar et al., 2016; Faruqui and Dyer, 2014; Lu et al., 2015) to transfer surface features across languages
- Cross-lingual IE work mainly focuses on sequence labeling (name tagging) (Mayhew et al., 2017; Lin et al., 2018; Huang et al., 2019) which is not significantly influenced by word order



# Applications: Event Tracking

Timeline Home Examples System

Entity Name: All Event Type: All Event SubType: All Argument Role: All Start Date: End Date: Search

Legend: Argument Trigger Time Timeline Scale: Year Month

March 2011

## Life.Die

Independent groups say 113 men and 24 women have set themselves on fire since March 2011 and most have died of their burns.

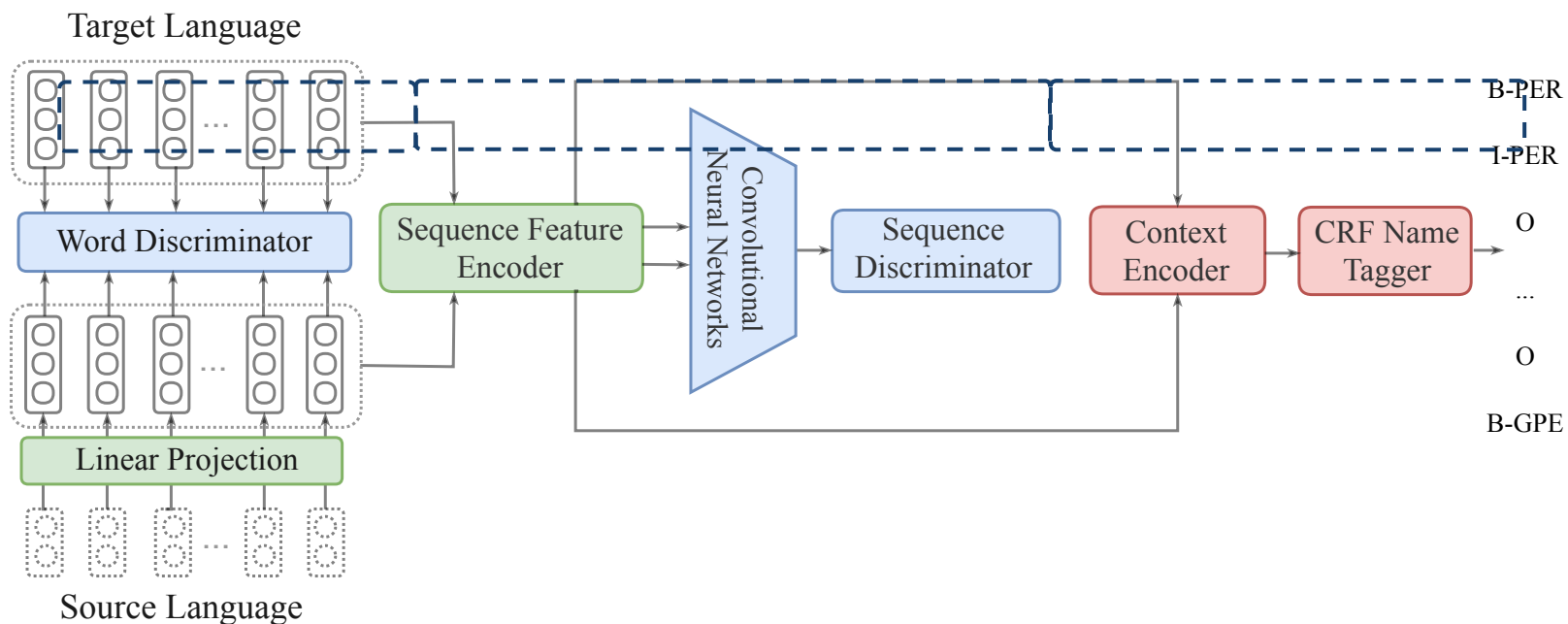
- Victim= {their , themselves , most }

Source Doc: HC0007FAT

Timeline JS

# Transfer: Unsupervised Adversarial Training

**Solution:** apply unsupervised adversarial training to extract language-agnostic features



Word-level  
adversarial transfer

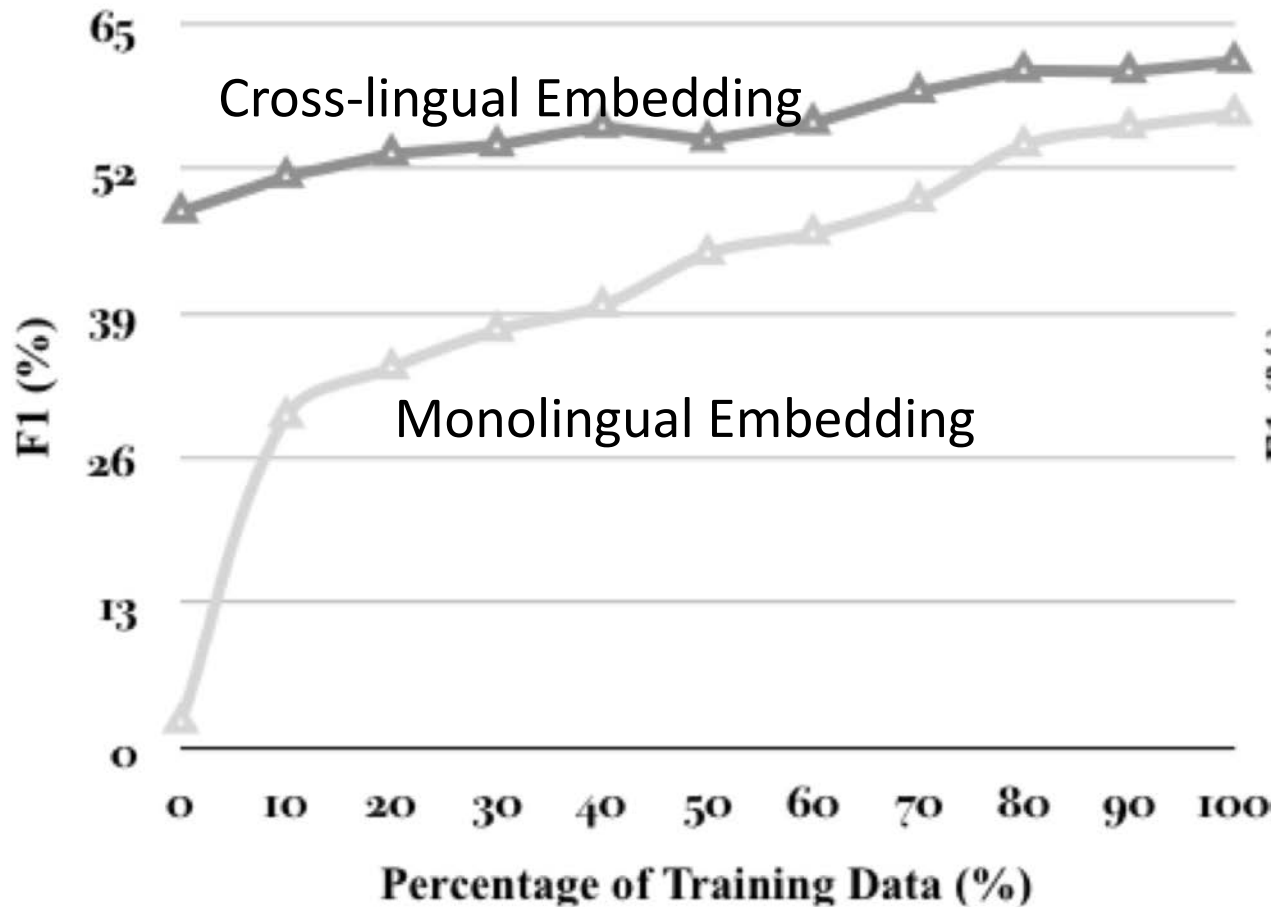
Sequence-level  
adversarial transfer

Bi-LSTM-CRF Name Tagger

- (Huang et al., NAACL2019)

# Name Tagging Performance

- Russian name tagging F-score; English to help Russian

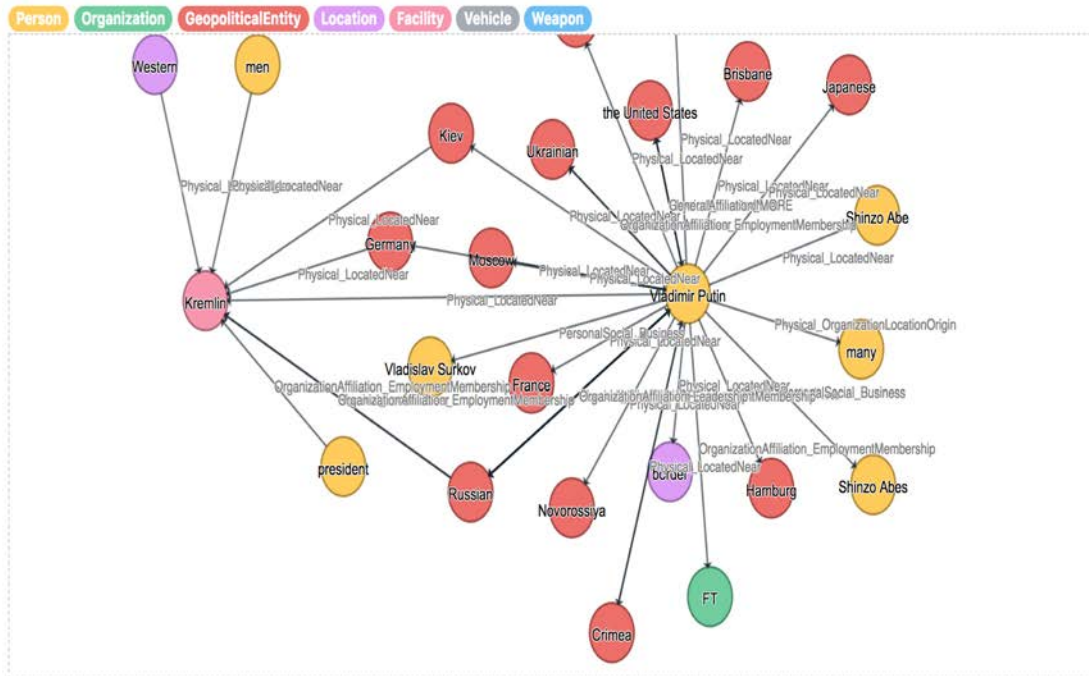


# Applications: Entity Relation Tracking



Name Search  Cypher Query(Example Cypher Queries)

Graph Visualization: **Double Click** a node to see its relations, **Click** a node and then the CROSS icon to remove it from visualization. Zoom in and out by mouse wheel.



Properties: **Hover** over nodes and edges to see their properties here.

Type	ID	Name	Link	Mention
Person	:Entity_EDL_ENG_0153461	Vladimir Putin	LDC2015E42:m.08193	liberals,#Putin,Putin,own,President,gunmen,Mr,he,Vladimir Putin,Vladimir Putins,companies,murderers,victims,that,He,putin's,them,himself,group



# Weakly Aligned Structured Embedding (WASE)

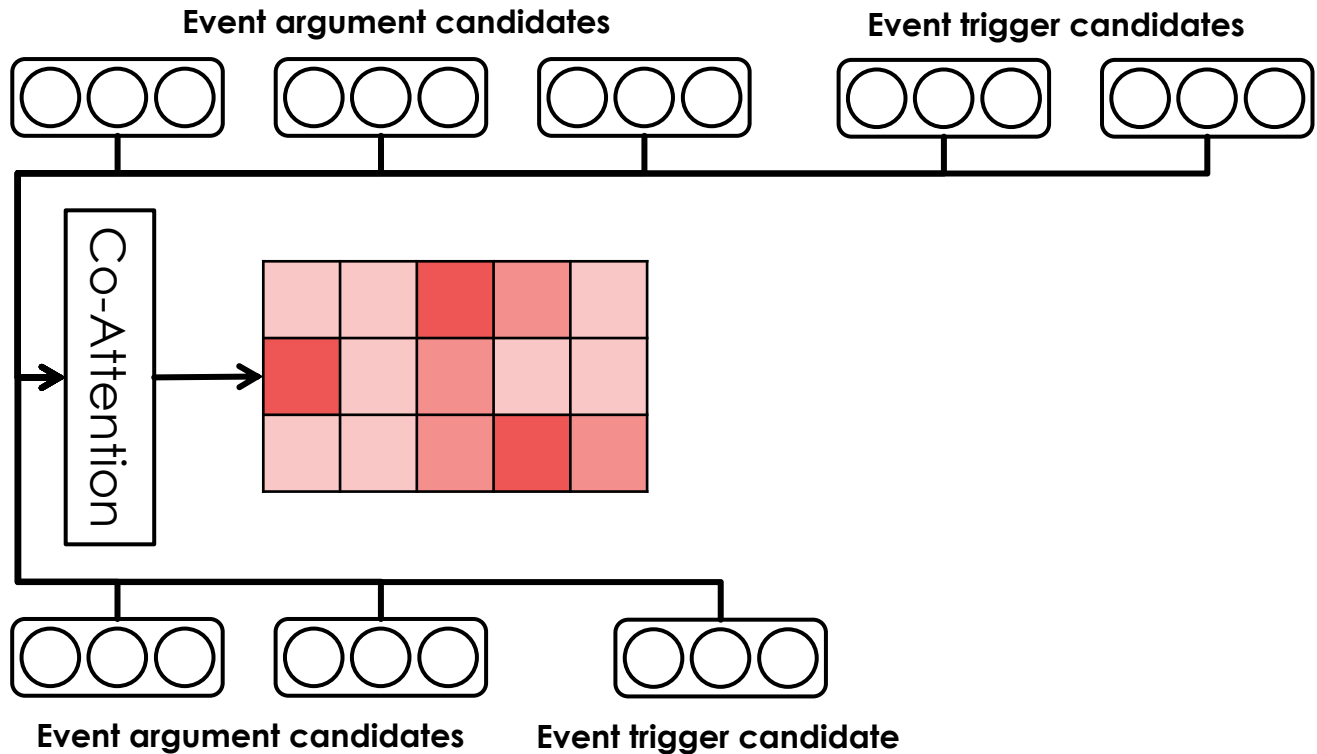
-- Training Phase (Common Space Construction)

Thai opposition protesters attack a bus carrying pro-government Red Shirt supporters on their way to a rally at a stadium in Bangkok



WASE (text)

WASE (img)



# Weakly Aligned Structured Embedding (WASE)

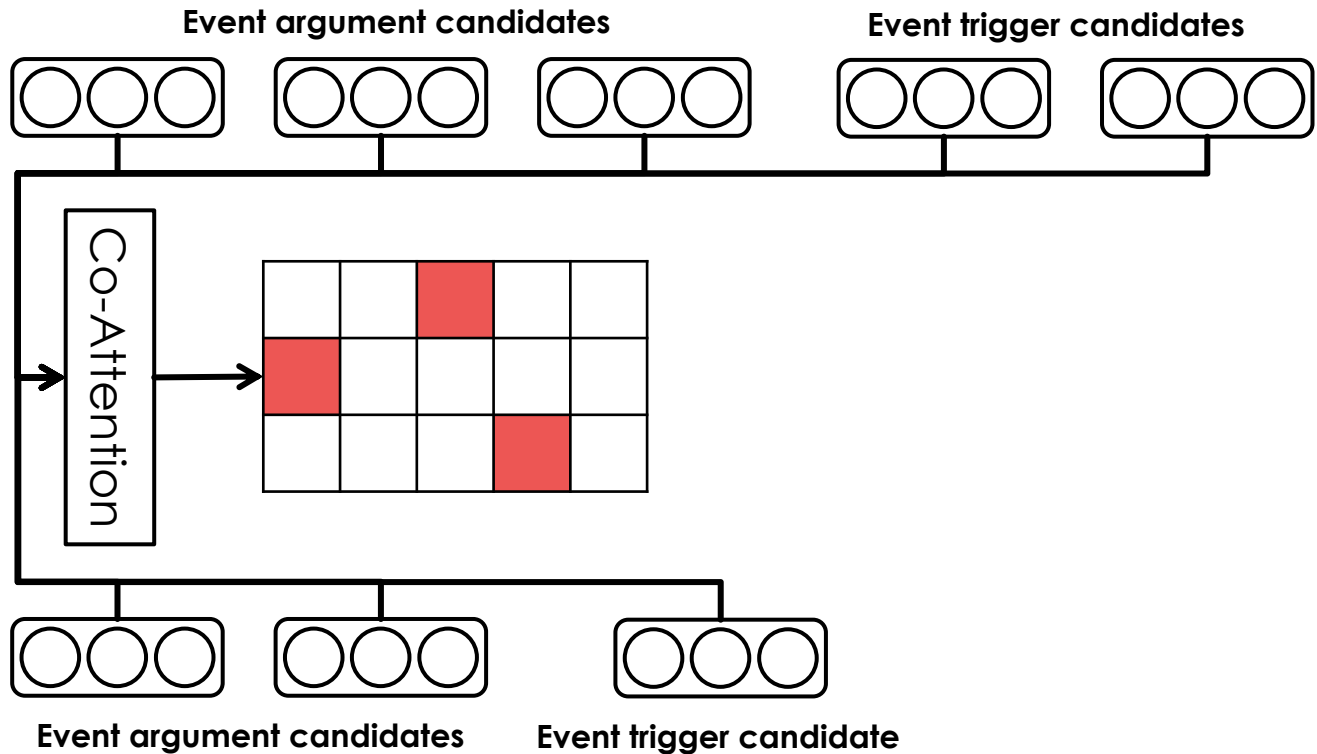
-- Training Phase (Common Space Construction)

Thai opposition protesters attack a bus carrying pro-government Red Shirt supporters on their way to a rally at a stadium in Bangkok



WASE (text)

WASE (img)



# Applications: County-Level Moral Concerns

- (Mooijman et al., Nature Human Behavior 2018, June Cover)

