Domain-Independent Quality Measures for Crowd Truth

A swarm of locusts.Photo: Mitsuhiko Imamori/Minden (wired)

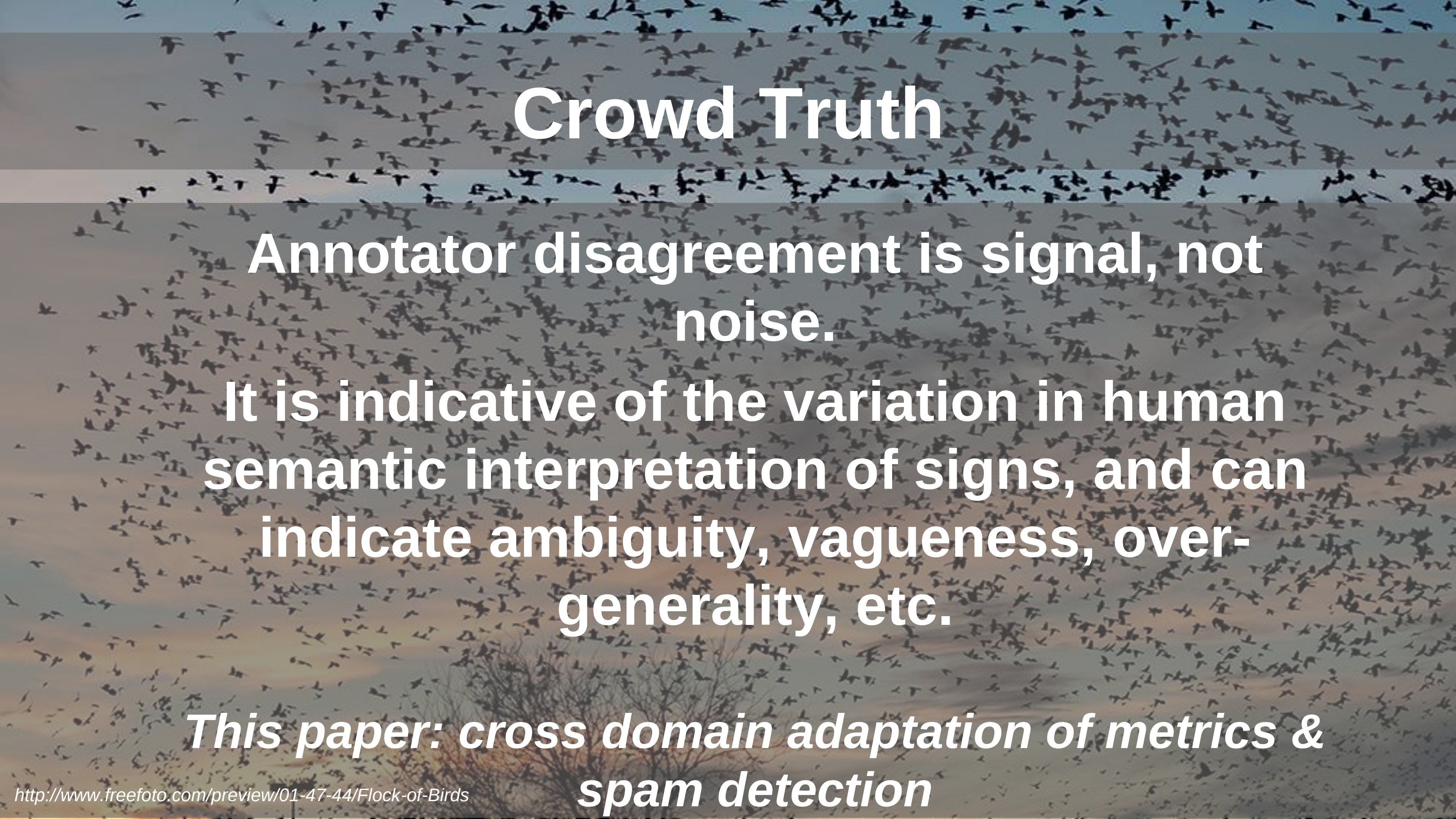
Oana Inel, Lora Aroyo, Chris Welty and Robert-Jan Sips

Crowd Truth

Annotator disagreement is signal, not noise.

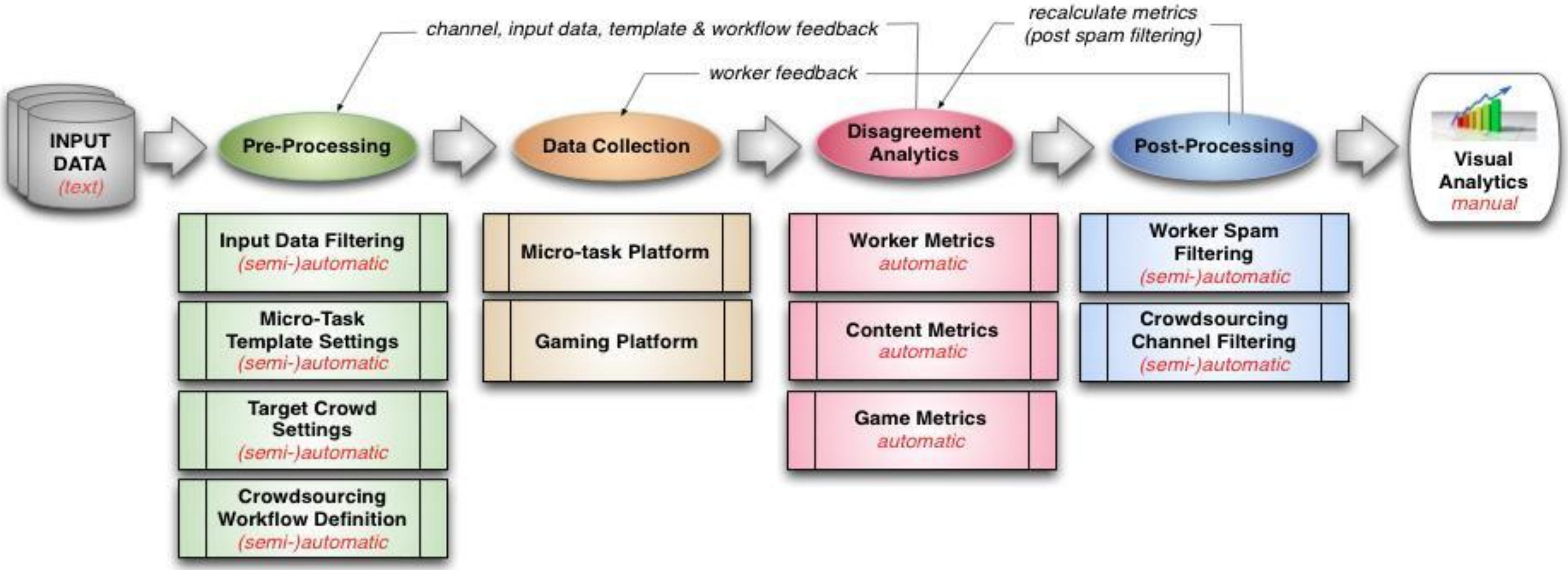
It is indicative of the variation in human semantic interpretation of signs, and can indicate ambiguity, vagueness, over-generality, etc.

This paper: cross domain adaptation of metrics & spam detection



http://www.freefoto.com/preview/01-47-44/Flock-of-Birds

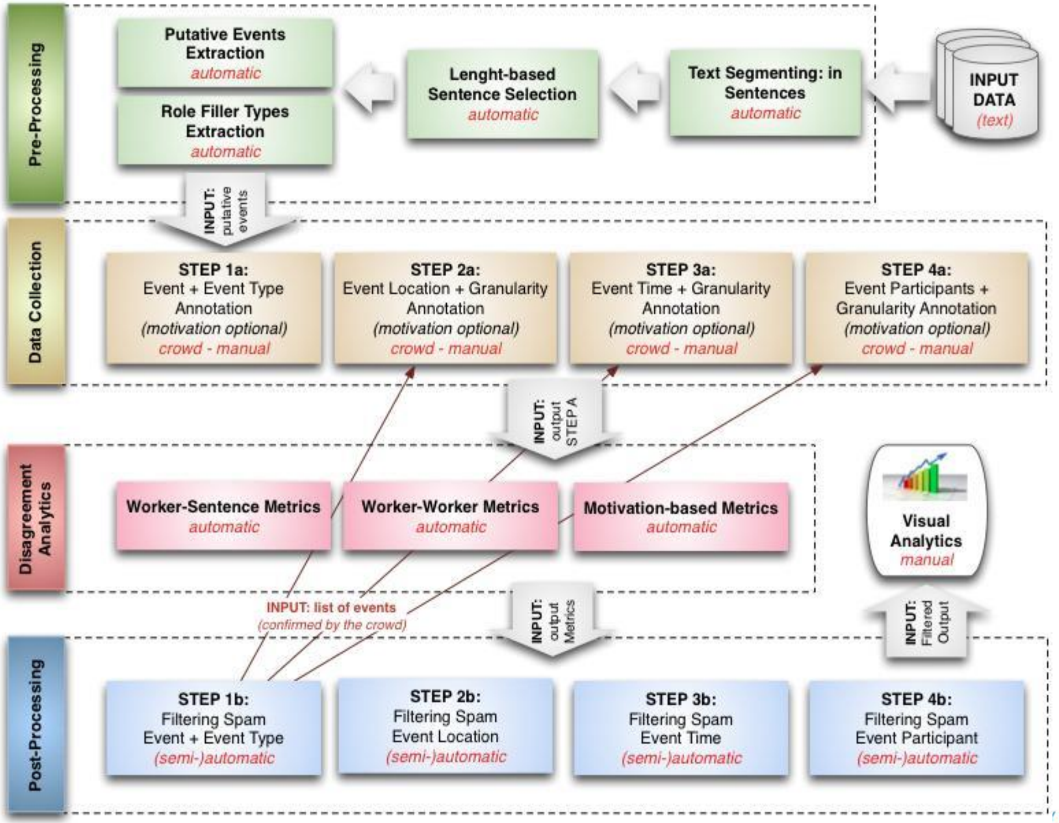
Background:
Crowd-Watson Framework for Medical Relation Extraction



Aroyo, L., Welty, C.: Measuring crowd truth for medical relation extraction. AAAI2013 Fall Symposium on Semantics for Big Data (in print), 2013
Aroyo, L., Welty, C.: Crowd Truth: Harnessing disagreement in crowdsourcing a relation extraction gold standard. *WebSci2013. ACM*, 2013

Crowd-Watson Adaptation to Newspapers Event Extraction

- newspapers corpus
- **identify events & role fillers** (e.g. type, location, time, participants)
- **understand the range of disagreements** by creating a space of possibilities with frequencies & similarities
- Crowd-Watson framework



Pre-Processing

INPUT DATA *(text)*

Text Segmenting: in Sentences *automatic*

Lenght-based Sentence Selection *automatic*

Putative Events Extraction *automatic*

Role Filler Types Extraction *automatic*

INPUT: putative events

Data Collection

STEP 1a: Event + Event Type Annotation *(motivation optional)* *crowd - manual*

STEP 2a: Event Location + Granularity Annotation *(motivation optional)* *crowd - manual*

STEP 3a: Event Time + Granularity Annotation *(motivation optional)* *crowd - manual*

STEP 4a: Event Participants + Granularity Annotation *(motivation optional)* *crowd - manual*

INPUT: output STEP A

Disagreement Analytics

Worker-Sentence Metrics *automatic*

Worker-Worker Metrics *automatic*

Motivation-based Metrics *automatic*

Visual Analytics *manual*

INPUT: list of events *(confirmed by the crowd)*

INPUT: output Metrics

INPUT: Filtered Output

Post-Processing

STEP 1b: Filtering Spam Event + Event Type *(semi-)automatic*

STEP 2b: Filtering Spam Event Location *(semi-)automatic*

STEP 3b: Filtering Spam Event Time *(semi-)automatic*

STEP 4b: Filtering Spam Event Participant *(semi-)automatic*

April 24, 2012, 7:27 pm | 6 Comments

Greenpeace Protests Apple's Energy Practices By Releasing Balloons

By C.J. HUGHES

C.J. Hughes for The New York Times

The police came to Apple's glass cube on Fifth Avenue on Tuesday to enforce order after activists released black balloons inside the cube to protest the company's environmental policies.

The skies over the glass cube that serves as the entrance to the Apple store on Fifth Avenue turned a bit dark on Tuesday afternoon, at least for some shoppers inside. And a passing storm had nothing to do with it.

The sudden shadows were caused by bunches of balloons released by Greenpeace, the environmental advocacy group, in order to call attention to Apple's use of coal power at its facilities.

Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators entered the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.

As they neared the bottom of the cube's stairs, they released the balloons, which hit the ceiling of the cube and lodged themselves in place. Other demonstrators came later with paper shopping bags filled with balloons, which they let trickle out as they entered the store, before being escorted out.

The stunt, which did not lead to any arrests, according to Greenpeace, was timed to coincide with last week's release of a report that ranked the eco-friendliness of top technology companies. Apple's "clean energy index" of 15 percent was below that of Google, with 39 percent, and Dell, with 56 percent.

That relatively low score, according to Greenpeace, is partly because much of the electricity Apple uses at a new data center in Maiden, N.C., where the company's iCloud storage is powered, comes from burning coal, a pollutant.

April 24, 2012, 7:27 pm | 6 Comments

Greenpeace Protests Apple's Energy Practices By Releasing Balloons

By C.J. HUGHES

C.J. Hughes for The New York Times

The police came to Apple's glass cube on Fifth Avenue on Tuesday to enforce order after activists released black balloons inside the cube to protest the company's environmental policies.

The skies over the glass cube that serves as the entrance to the Apple store on Fifth Avenue turned a bit dark on Tuesday afternoon, at least for some shoppers inside. ~~And a passing storm had nothing to do with it.~~

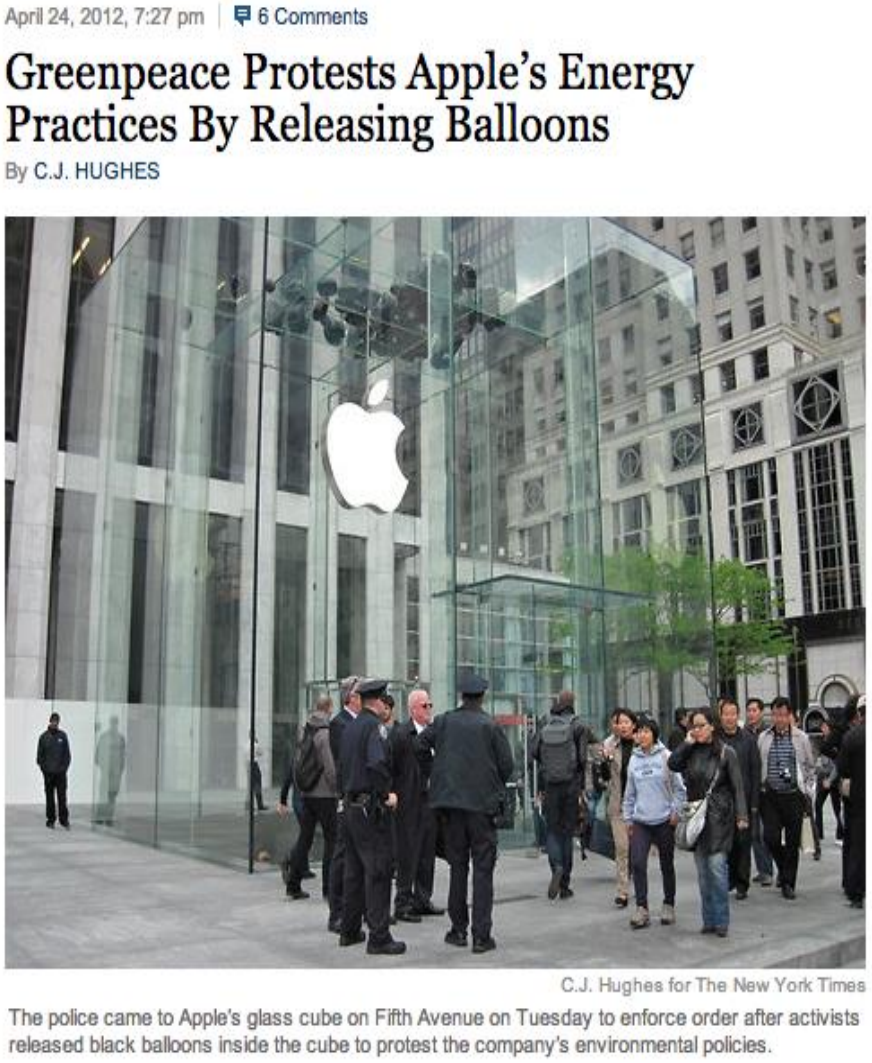
The sudden shadows were caused by bunches of balloons released by Greenpeace, the environmental advocacy group, in order to call attention to Apple's use of coal power at its facilities.

Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators entered the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.

As they neared the bottom of the cube's stairs, they released the balloons, which hit the ceiling of the cube and lodged themselves in place. Other demonstrators came later with paper shopping bags filled with balloons, which they let trickle out as they entered the store, before being escorted out.

The stunt, which did not lead to any arrests, according to Greenpeace, was timed to coincide with last week's release of a report that ranked the eco-friendliness of top technology companies. Apple's "clean energy index" of 15 percent was below that of Google, with 39 percent, and Dell, with 56 percent.

That relatively low score, according to Greenpeace, is partly because much of the electricity Apple uses at a new data center in Maiden, N.C., where the company's iCloud storage is powered, comes from burning coal, a pollutant.



April 24, 2012, 7:27 pm | 6 Comments

Greenpeace Protests Apple's Energy Practices By Releasing Balloons

By C.J. HUGHES

C.J. Hughes for The New York Times

The police came to Apple's glass cube on Fifth Avenue on Tuesday to enforce order after activists released black balloons inside the cube to protest the company's environmental policies.

The police came to Apple's glass cube on Fifth Avenue on Tuesday to enforce order after activists released black balloons inside the cube to **[protest]** the company's environmental policies.

The police came to Apple's glass cube on Fifth Avenue on Tuesday **[to enforce]** order after activists released black balloons inside the cube to protest the company's environmental policies.

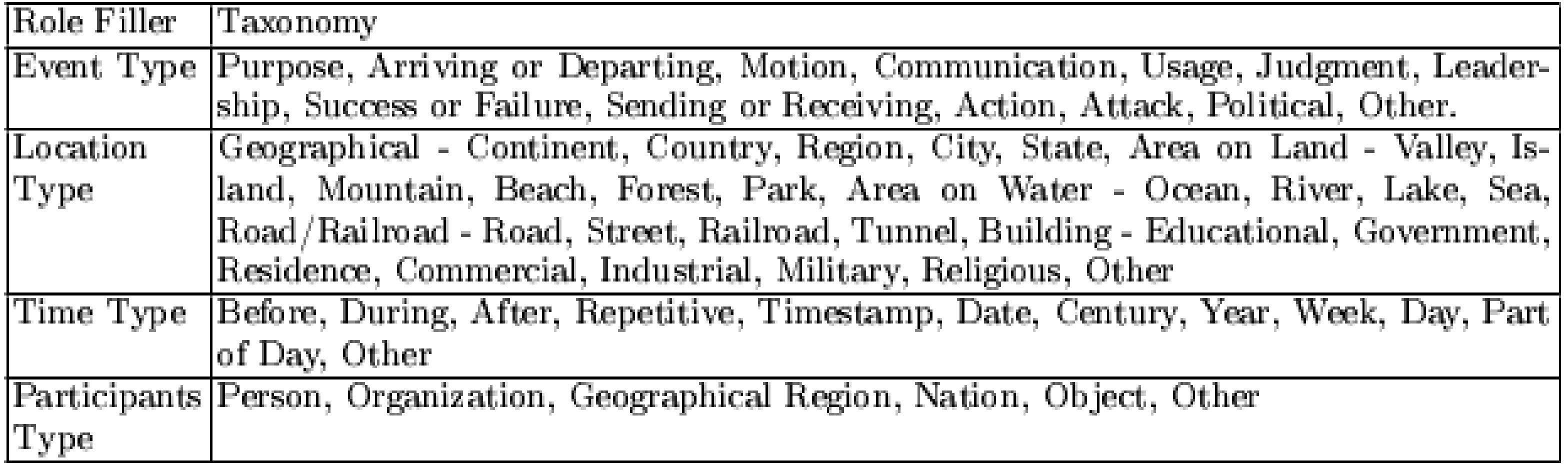
The police came to Apple's glass cube on Fifth Avenue on Tuesday **[to enforce order]** after activists released black balloons inside the cube to protest the company's environmental policies.

The police came to Apple's glass cube on Fifth Avenue on Tuesday to enforce order after activists **[released]** black **[balloons]** inside the cube to protest the company's environmental policies.

The police **[came]**to Apple's glass cube on Fifth Avenue on Tuesday **[to enforce]** order after activists released black balloons inside the cube to protest the company's environmental policies.

Role-Filler Taxonomies

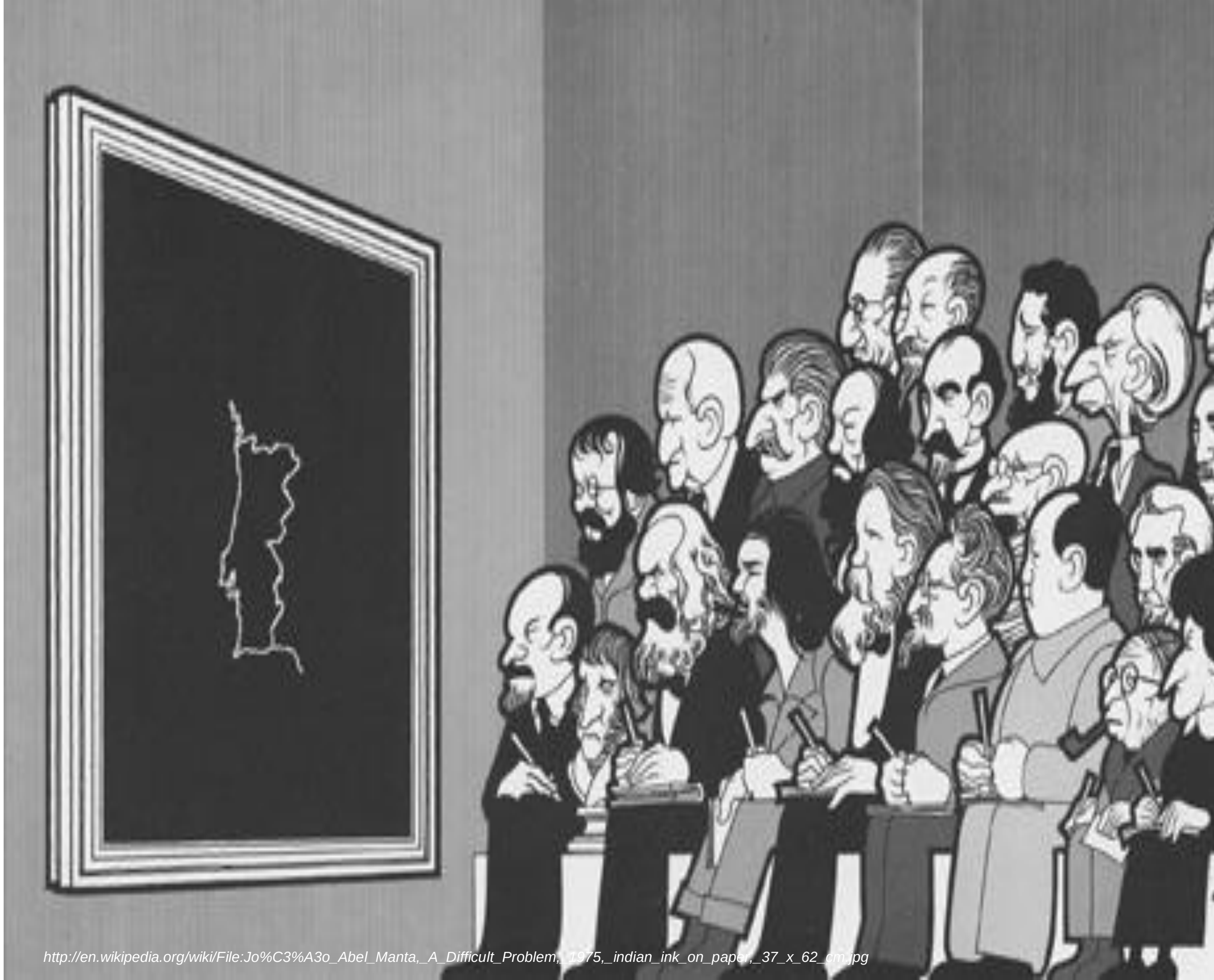
- **event type:** Semafor
- **event location type:** GeoNames
- **event time type:** Allen’s time theory, KSL time ontology
- **event participant type:** based on proper nouns classes



Role Filler	Taxonomy
Event Type	Purpose, Arriving or Departing, Motion, Communication, Usage, Judgment, Leadership, Success or Failure, Sending or Receiving, Action, Attack, Political, Other.
Location Type	Geographical - Continent, Country, Region, City, State, Area on Land - Valley, Island, Mountain, Beach, Forest, Park, Area on Water - Ocean, River, Lake, Sea, Road/Railroad - Road, Street, Railroad, Tunnel, Building - Educational, Government, Residence, Commercial, Industrial, Military, Religious, Other
Time Type	Before, During, After, Repetitive, Timestamp, Date, Century, Year, Week, Day, Part of Day, Other
Participants Type	Person, Organization, Geographical Region, Nation, Object, Other

How do we represent & measure disagreement in a way that it can be harnessed?

Events semantics are hard



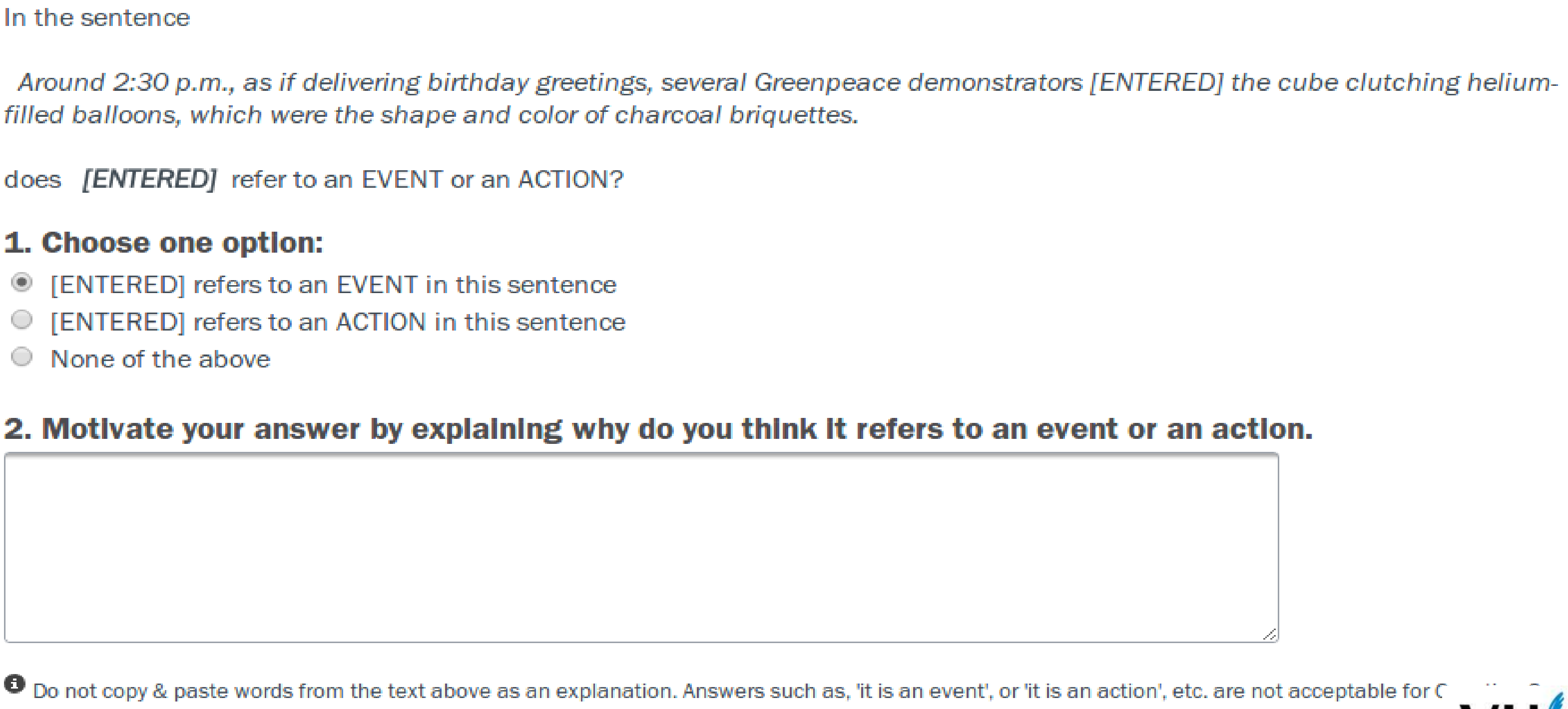
http://en.wikipedia.org/wiki/File:Jo%C3%A3o_Abel_Manta,_A_Difficult_Problem,_1975,_indian_ink_on_paper,_37_x_62_cm.jpg

Events have multiple dimensions

Micro-task Template

Events have multiple dimensions

Micro-task Template



In the sentence

Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [ENTERED] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.

does *[ENTERED]* refer to an EVENT or an ACTION?

1. Choose one option:

● [ENTERED] refers to an EVENT in this sentence

○ [ENTERED] refers to an ACTION in this sentence

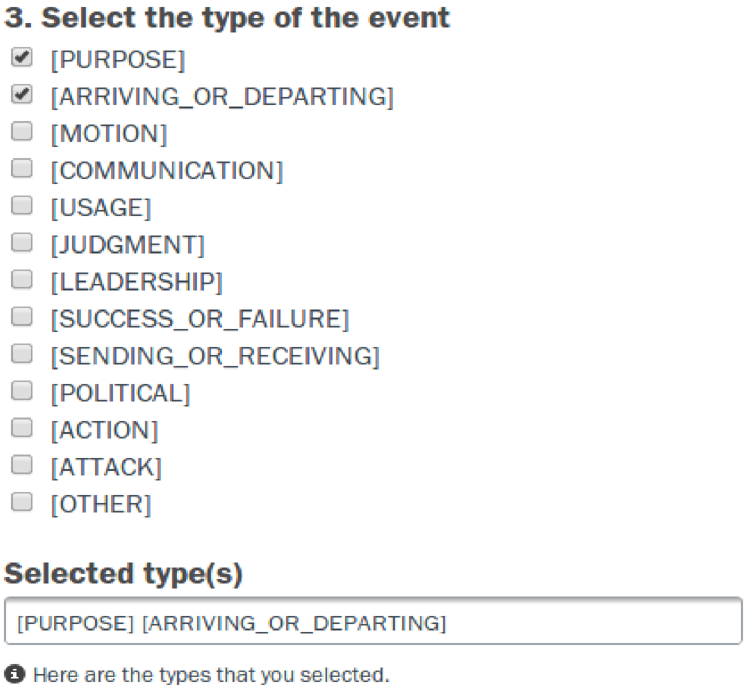
○ None of the above

2. Motivate your answer by explaining why do you think it refers to an event or an action.

ℹ Do not copy & paste words from the text above as an explanation. Answers such as, 'it is an event', or 'it is an action', etc. are not acceptable for C

Events have multiple dimensions

Micro-task Template



3. Select the type of the event

☑ [PURPOSE]
☑ [ARRIVING_OR_DEPARTING]
□ [MOTION]
□ [COMMUNICATION]
□ [USAGE]
□ [JUDGMENT]
□ [LEADERSHIP]
□ [SUCCESS_OR_FAILURE]
□ [SENDING_OR_RECEIVING]
□ [POLITICAL]
□ [ACTION]
□ [ATTACK]
□ [OTHER]

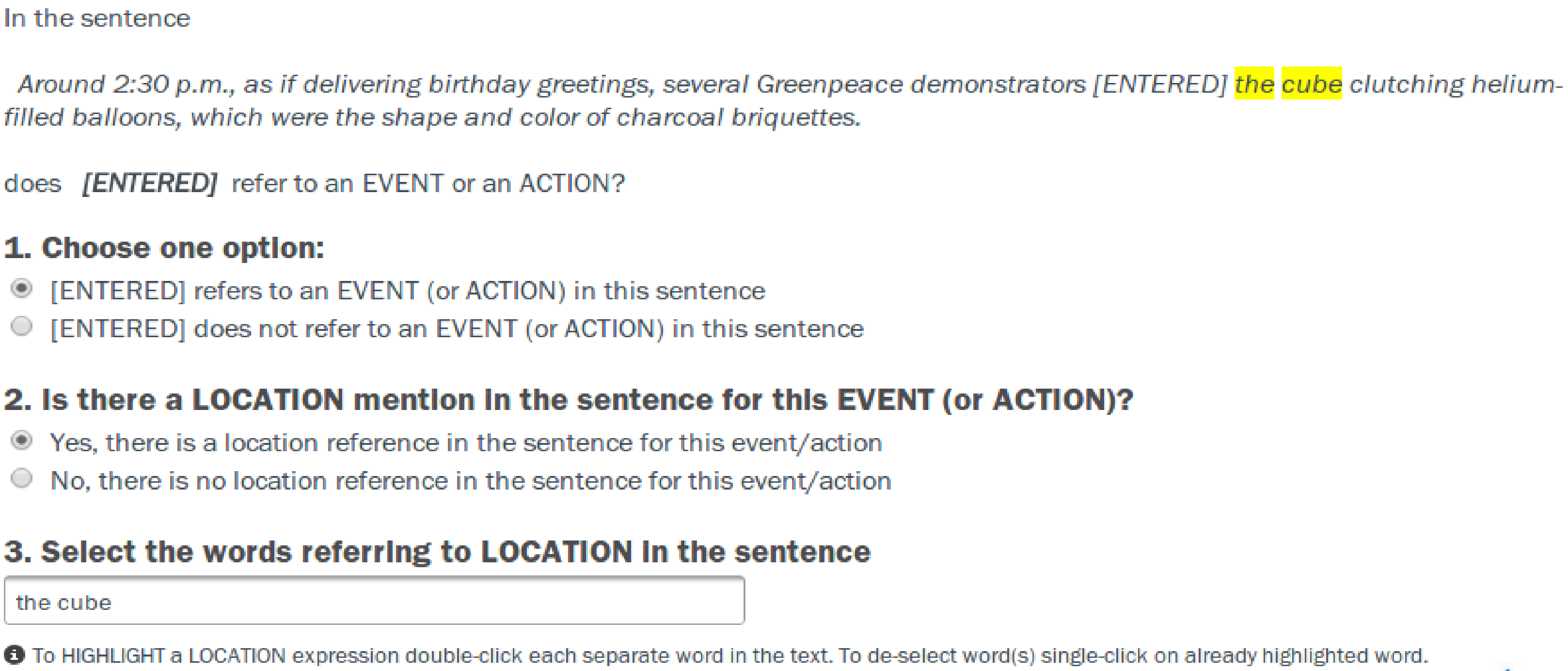
Selected type(s)

[PURPOSE] [ARRIVING_OR_DEPARTING]

ℹ Here are the types that you selected.

Events have multiple dimensions

Micro-task Template



In the sentence

Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [ENTERED] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.

does *[ENTERED]* refer to an EVENT or an ACTION?

1. Choose one option:

● [ENTERED] refers to an EVENT (or ACTION) in this sentence

○ [ENTERED] does not refer to an EVENT (or ACTION) in this sentence

2. Is there a LOCATION mention in the sentence for this EVENT (or ACTION)?

● Yes, there is a location reference in the sentence for this event/action

○ No, there is no location reference in the sentence for this event/action

3. Select the words referring to LOCATION in the sentence

the cube

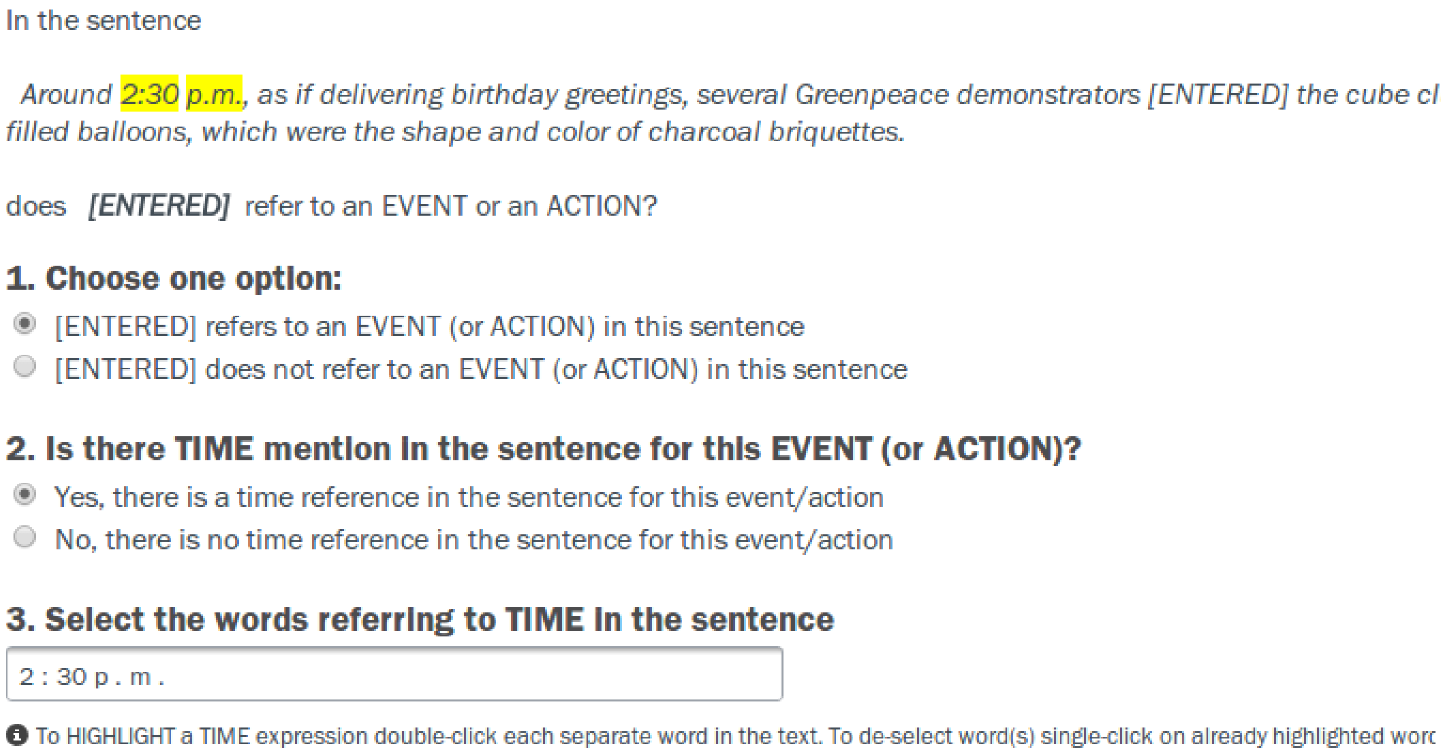
To HIGHLIGHT a LOCATION expression double-click each separate word in the text. To de-select word(s) single-click on already highlighted word.

Each dimension has different granularity

Micro-task Template

Each dimension has different granularity

Micro-task Template



In the sentence

Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [ENTERED] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.

does *[ENTERED]* refer to an EVENT or an ACTION?

1. Choose one option:

● [ENTERED] refers to an EVENT (or ACTION) in this sentence

○ [ENTERED] does not refer to an EVENT (or ACTION) in this sentence

2. Is there TIME mention in the sentence for this EVENT (or ACTION)?

● Yes, there is a time reference in the sentence for this event/action

○ No, there is no time reference in the sentence for this event/action

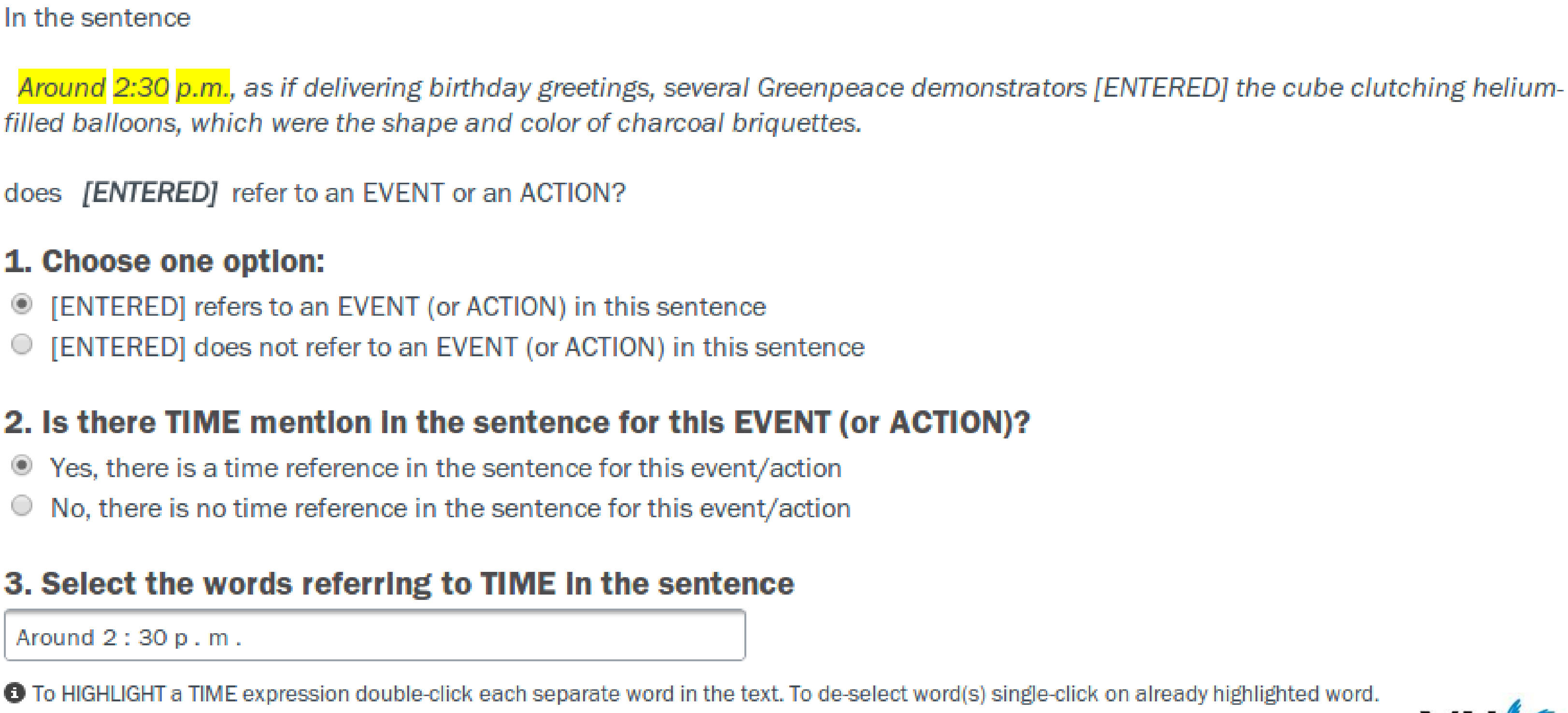
3. Select the words referring to TIME in the sentence

2 : 30 p . m .

ℹ To HIGHLIGHT a TIME expression double-click each separate word in the text. To de-select word(s) single-click on already highlighted worc

Each dimension has different granularity

Micro-task Template



In the sentence

Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [ENTERED] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.

does *[ENTERED]* refer to an EVENT or an ACTION?

1. Choose one option:

● [ENTERED] refers to an EVENT (or ACTION) in this sentence

○ [ENTERED] does not refer to an EVENT (or ACTION) in this sentence

2. Is there TIME mention in the sentence for this EVENT (or ACTION)?

● Yes, there is a time reference in the sentence for this event/action

○ No, there is no time reference in the sentence for this event/action

3. Select the words referring to TIME in the sentence

Around 2 : 30 p . m .

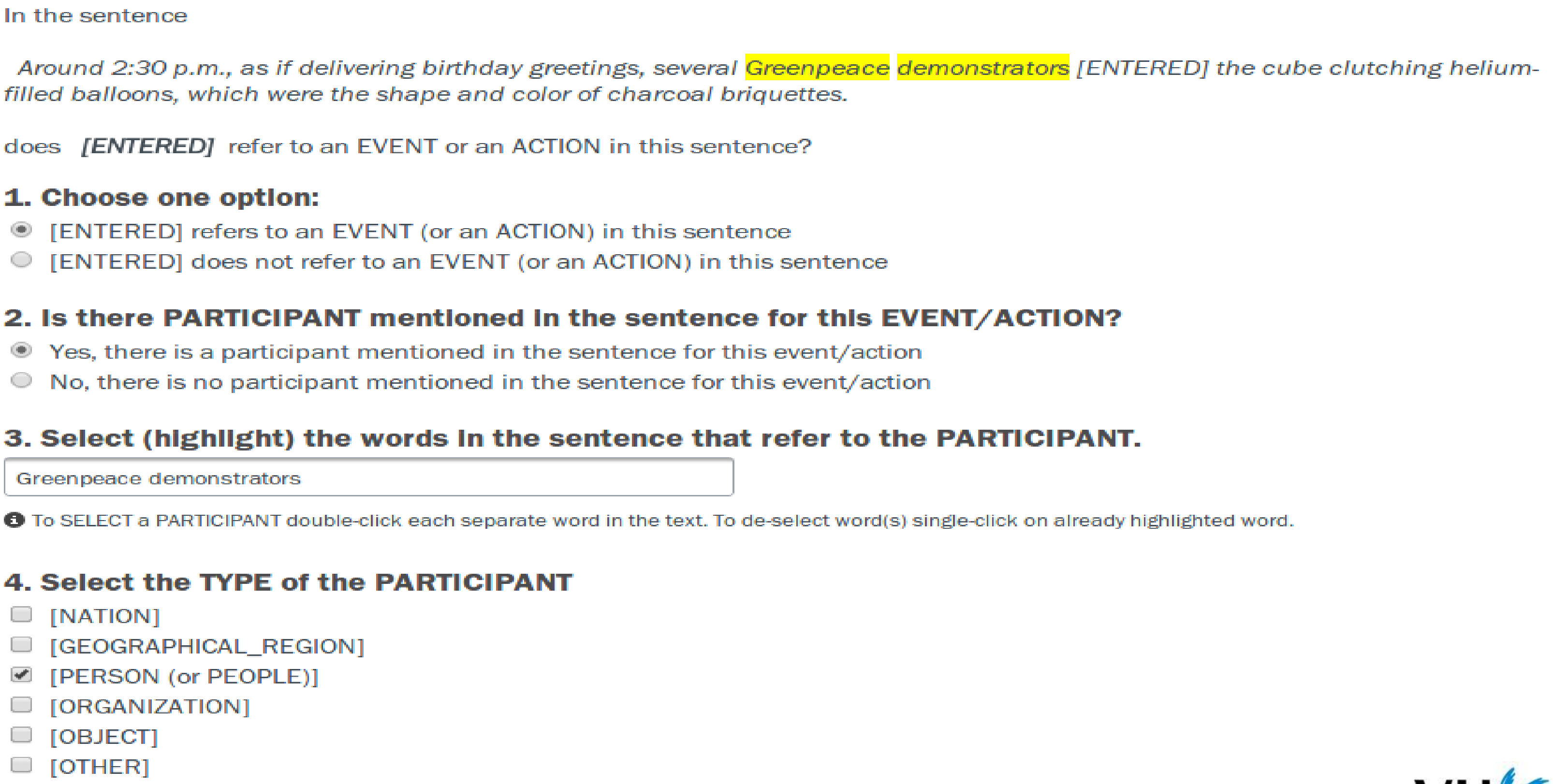
ℹ To HIGHLIGHT a TIME expression double-click each separate word in the text. To de-select word(s) single-click on already highlighted word.

People have different points of views

Micro-task Template

People have different points of views

Micro-task Template



In the sentence

Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [ENTERED] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.

does *[ENTERED]* refer to an EVENT or an ACTION in this sentence?

1. Choose one option:

● [ENTERED] refers to an EVENT (or an ACTION) in this sentence
○ [ENTERED] does not refer to an EVENT (or an ACTION) in this sentence

2. Is there PARTICIPANT mentioned in the sentence for this EVENT/ACTION?

● Yes, there is a participant mentioned in the sentence for this event/action
○ No, there is no participant mentioned in the sentence for this event/action

3. Select (highlight) the words in the sentence that refer to the PARTICIPANT.

Greenpeace demonstrators

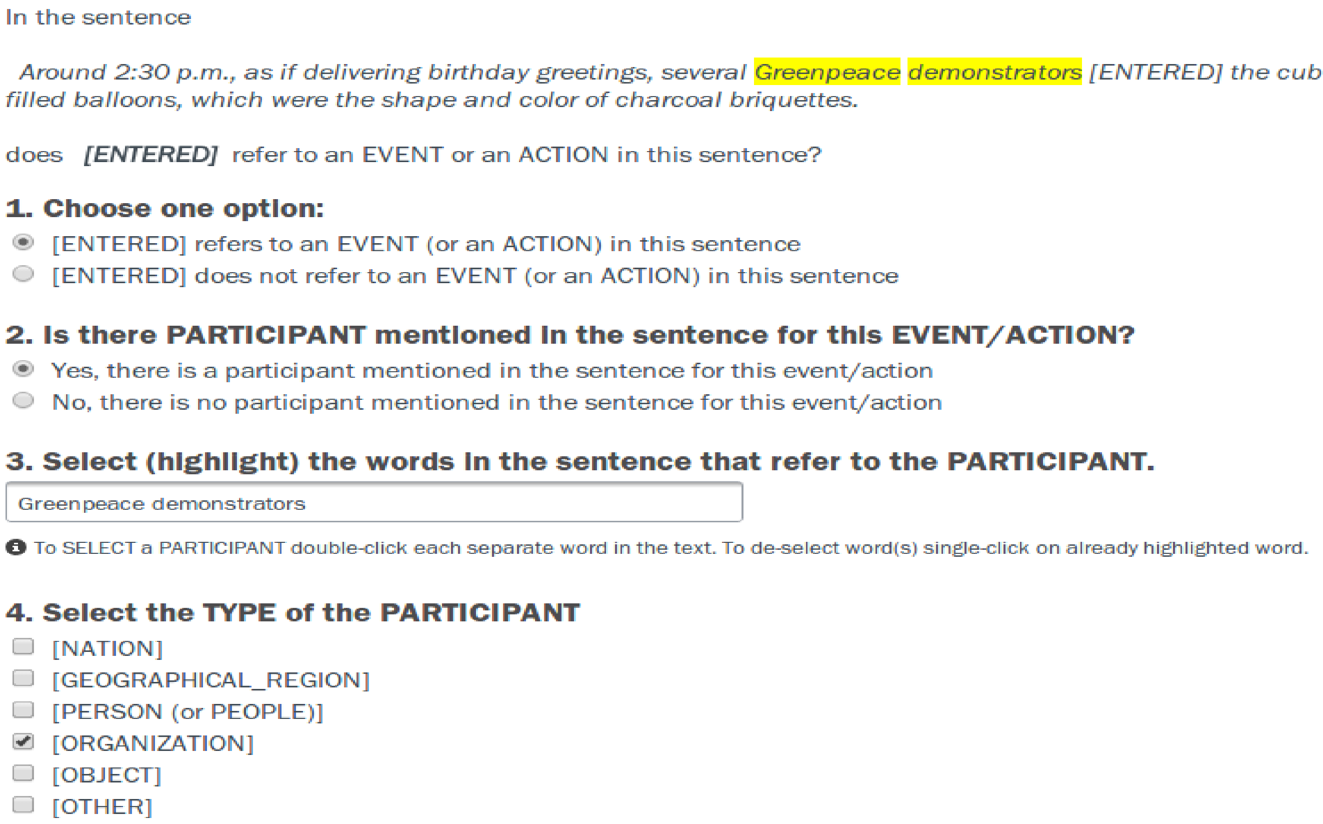
To SELECT a PARTICIPANT double-click each separate word in the text. To de-select word(s) single-click on already highlighted word.

4. Select the TYPE of the PARTICIPANT

☐ [NATION]
☐ [GEOGRAPHICAL_REGION]
☑ [PERSON (or PEOPLE)]
☐ [ORGANIZATION]
☐ [OBJECT]
☐ [OTHER]

People have different points of views

Micro-task Template



In the sentence

Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [ENTERED] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.

does *[ENTERED]* refer to an EVENT or an ACTION in this sentence?

1. Choose one option:

● [ENTERED] refers to an EVENT (or an ACTION) in this sentence

○ [ENTERED] does not refer to an EVENT (or an ACTION) in this sentence

2. Is there PARTICIPANT mentioned in the sentence for this EVENT/ACTION?

● Yes, there is a participant mentioned in the sentence for this event/action

○ No, there is no participant mentioned in the sentence for this event/action

3. Select (highlight) the words in the sentence that refer to the PARTICIPANT.

Greenpeace demonstrators

ℹ To SELECT a PARTICIPANT double-click each separate word in the text. To de-select word(s) single-click on already highlighted word.

4. Select the TYPE of the PARTICIPANT

☐ [NATION]

☐ [GEOGRAPHICAL_REGION]

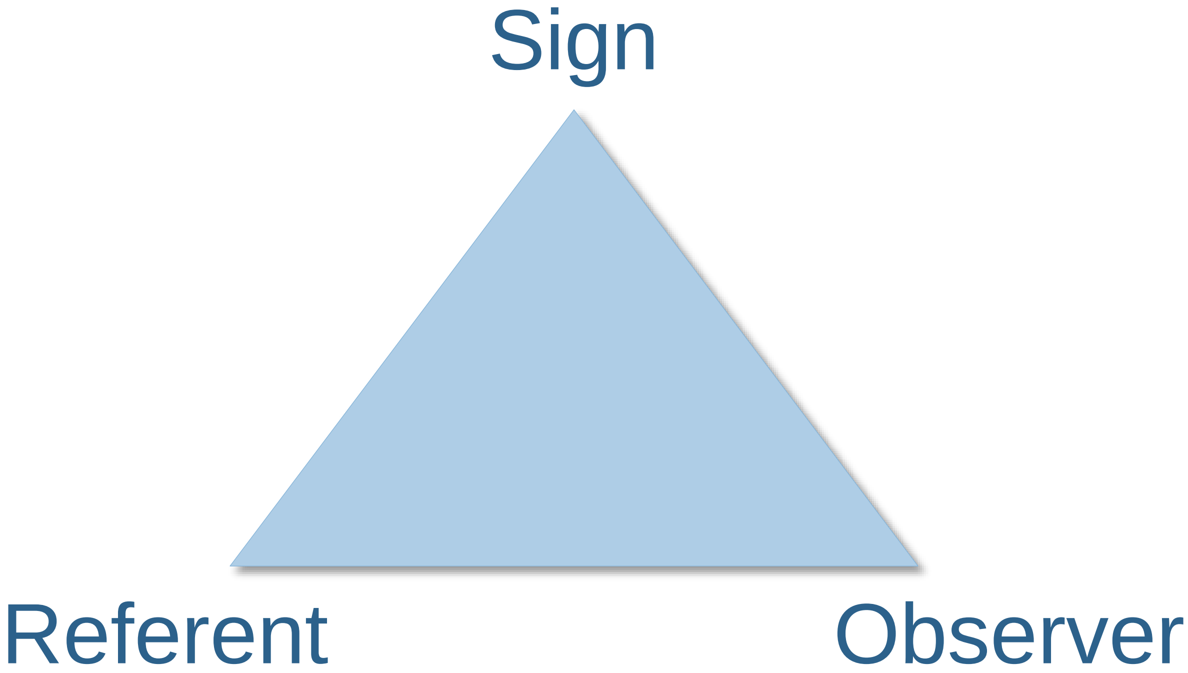
☐ [PERSON (or PEOPLE)]

☑ [ORGANIZATION]

☐ [OBJECT]

☐ [OTHER]

Why do people disagree?

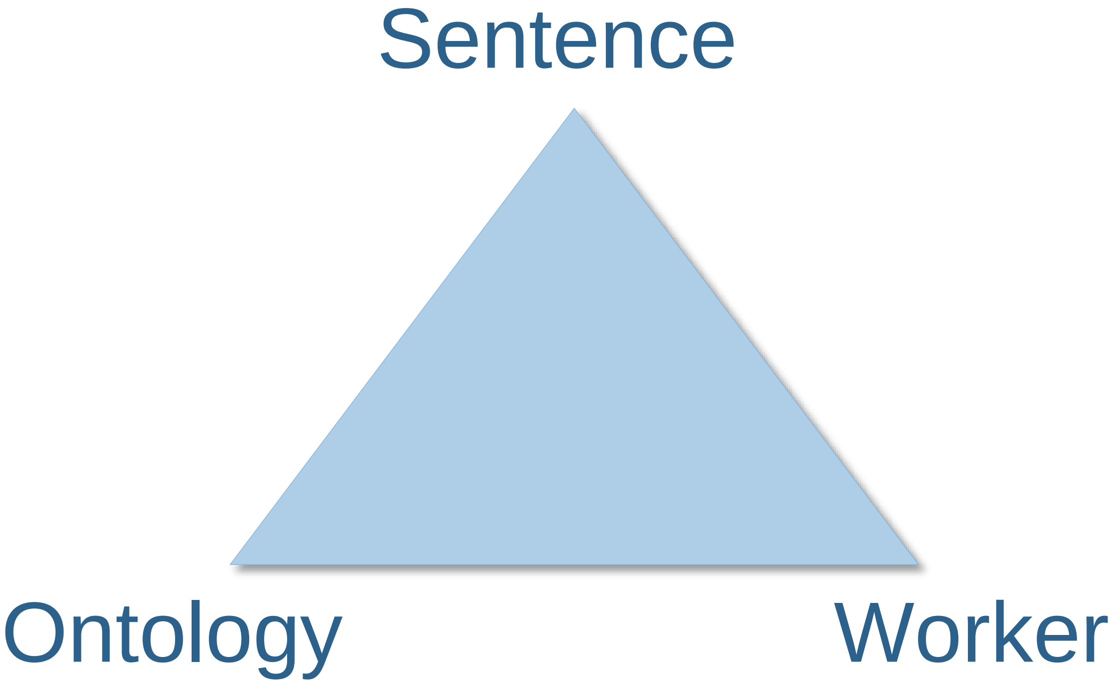


Sign

Referent

Observer

Why do people disagree?



Sentence

Ontology

Worker

Disagreement Analytics

- **sentence metrics:** sentence clarity
- **ontology metrics:** (future work)
- **worker content-based metrics:**
  - *worker-sentence disagreement*
  - *worker-worker disagreement*
  - *avg number of annotations per sentence*
- **worker explanation-based metrics:**
  - *valid words in explanation text*
  - *same explanation across contributions*
  - *“[OTHER]” + different type*

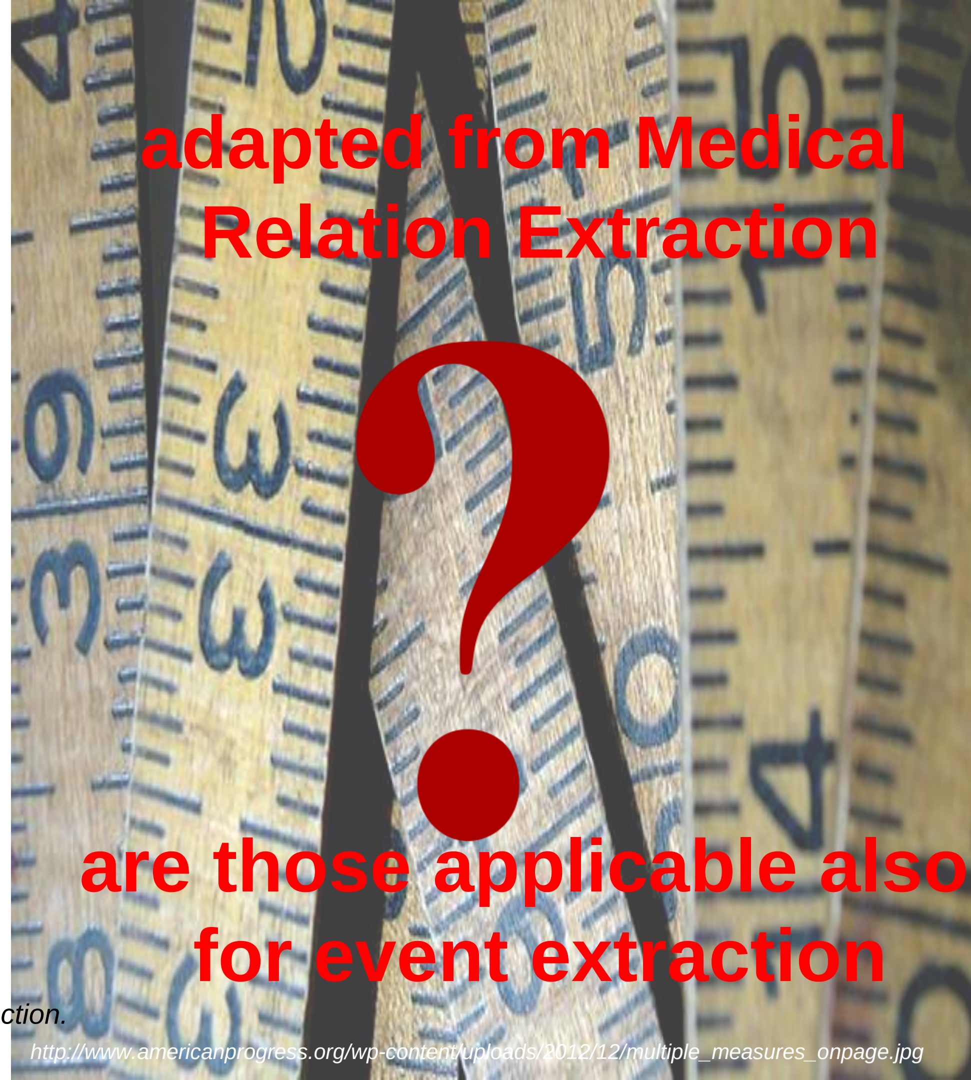
G. Soberón et al (2013): Crowd truth metrics. CrowdSem13 Workshop
L Aroyo, C. Welty (2013): Measuring crowd truth for medical relation extraction.
AAAI2013 Fall Symposium on Semantics for Big Data (in print) (2013)

http://www.americanprogress.org/wp-content/uploads/2012/12/multiple_measures_onpage.jpg

Disagreement Analytics

- **sentence metrics:** sentence clarity
- **ontology metrics:** (future work)
- **worker content-based metrics:**
  - *worker-sentence disagreement*
  - *worker-worker disagreement*
  - *avg number of annotations per sentence*
- **worker explanation-based metrics:**
  - *valid words in explanation text*
  - *same explanation across contributions*
  - *“[OTHER]” + different type*

G. Soberón et al (2013): Crowd truth metrics. CrowdSem13 Workshop
L Aroyo, C. Welty (2013): Measuring crowd truth for medical relation extraction. AAAI2013 Fall Symposium on Semantics for Big Data (in print) (2013)



adapted from Medical Relation Extraction

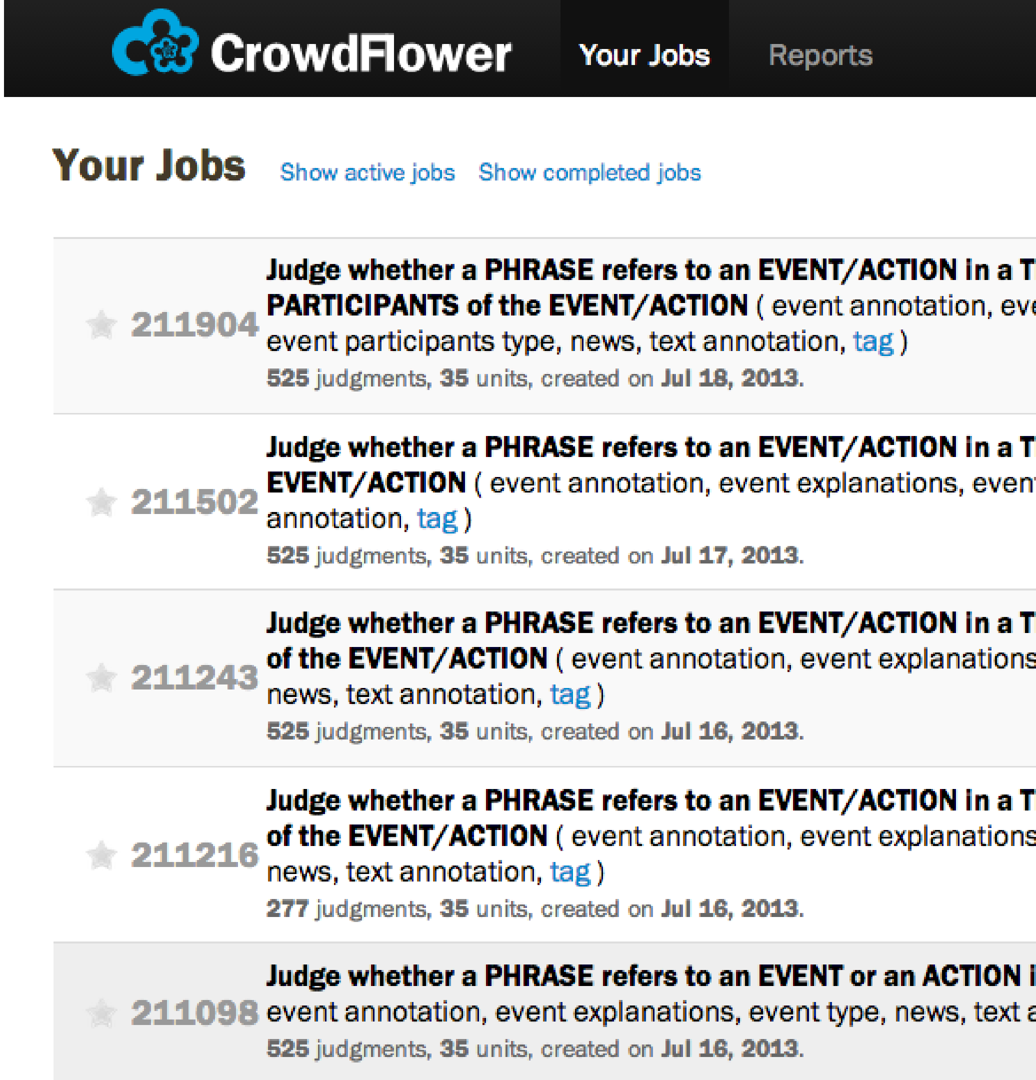
?

are those applicable also for event extraction

http://www.americanprogress.org/wp-content/uploads/2012/12/multiple_measures_onpage.jpg

Experimental Setting

- **2 batches of 35 putative events:** total of 70 putative events
- **8 experiments:** 2 for each event role filler
- **annotations:** 15 per each putative event
- **maximum annotations per worker:** 10
- **workers:** native English speakers on CF



211904 **Judge whether a PHRASE refers to an EVENT/ACTION in a T PARTICIPANTS of the EVENT/ACTION** (event annotation, eve event participants type, news, text annotation, tag)
525 judgments, 35 units, created on Jul 18, 2013.

211502 **Judge whether a PHRASE refers to an EVENT/ACTION in a T EVENT/ACTION** (event annotation, event explanations, even annotation, tag)
525 judgments, 35 units, created on Jul 17, 2013.

211243 **Judge whether a PHRASE refers to an EVENT/ACTION in a T of the EVENT/ACTION** (event annotation, event explanations news, text annotation, tag)
525 judgments, 35 units, created on Jul 16, 2013.

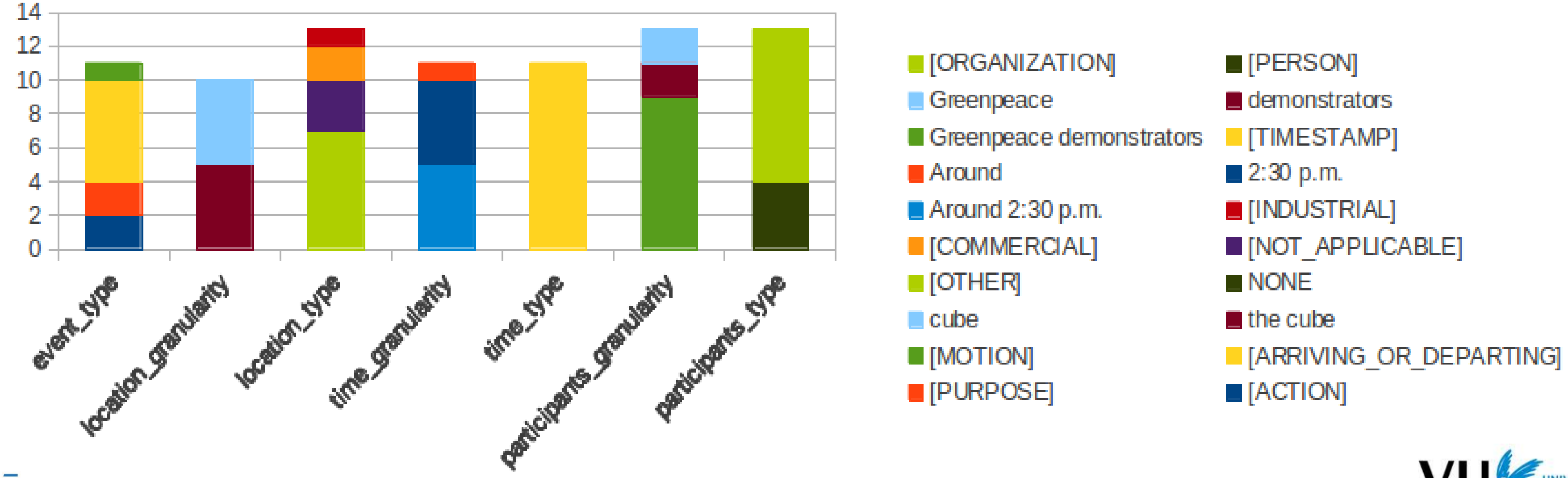
211216 **Judge whether a PHRASE refers to an EVENT/ACTION in a T of the EVENT/ACTION** (event annotation, event explanations news, text annotation, tag)
277 judgments, 35 units, created on Jul 16, 2013.

211098 **Judge whether a PHRASE refers to an EVENT or an ACTION i** event annotation, event explanations, event type, news, text a
525 judgments, 35 units, created on Jul 16, 2013.

Annotation Example

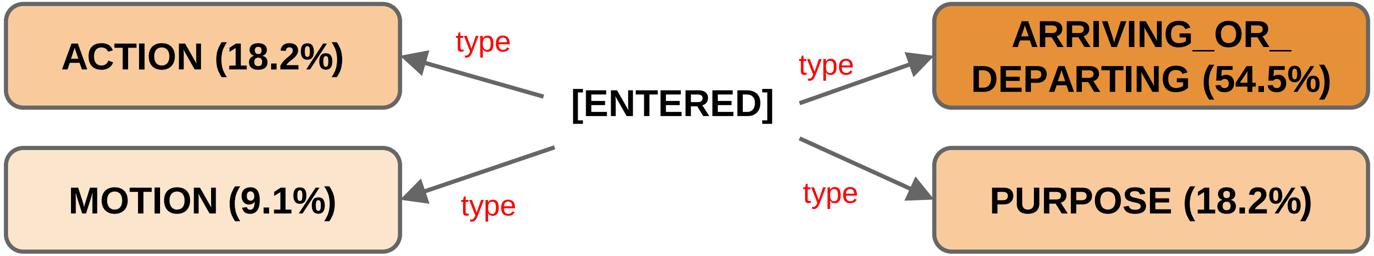
*Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators **[ENTERED]** the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.*

Overall annotation & granularity distribution:



Event Type Disagreement

*Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [**ENTERED**] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.*



[ENTERED]

type

ACTION (18.2%)

type

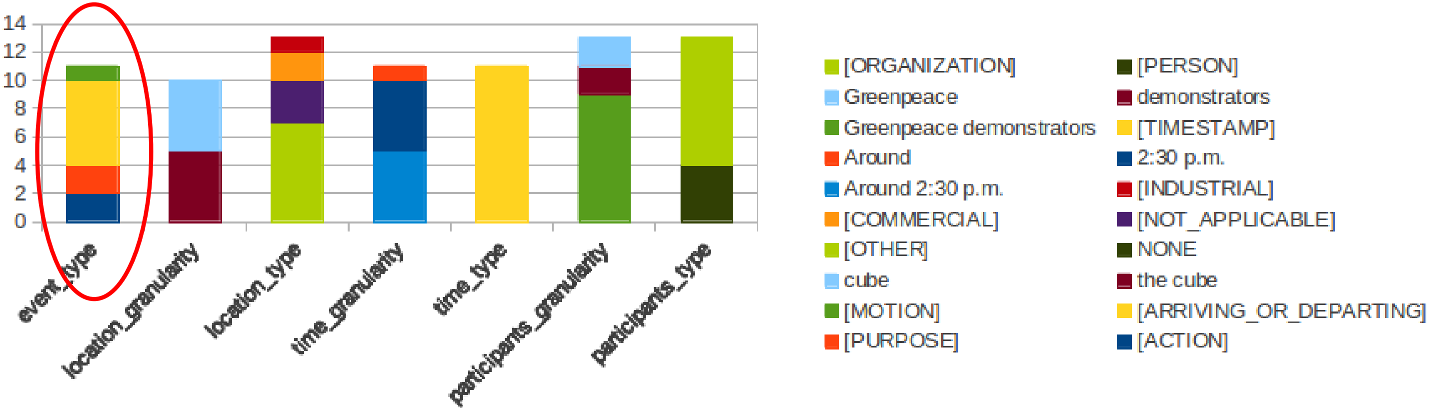
MOTION (9.1%)

type

ARRIVING_OR_ DEPARTING (54.5%)

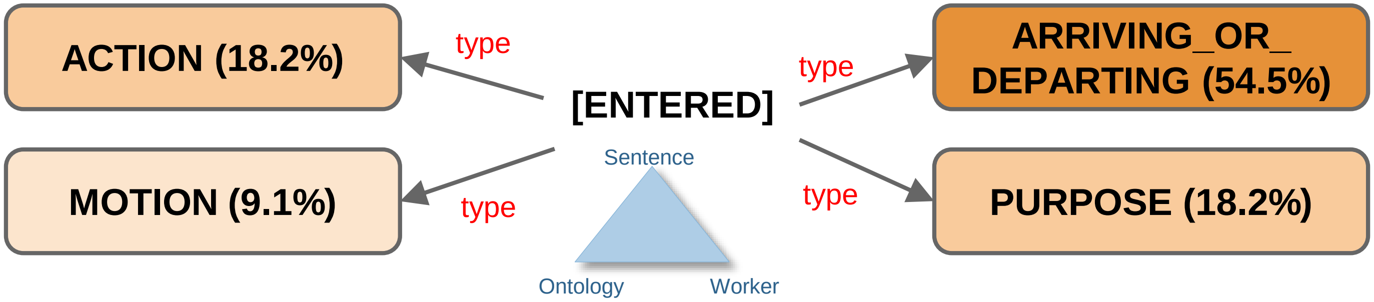
type

PURPOSE (18.2%)



Event Type Disagreement

*Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [**ENTERED**] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.*



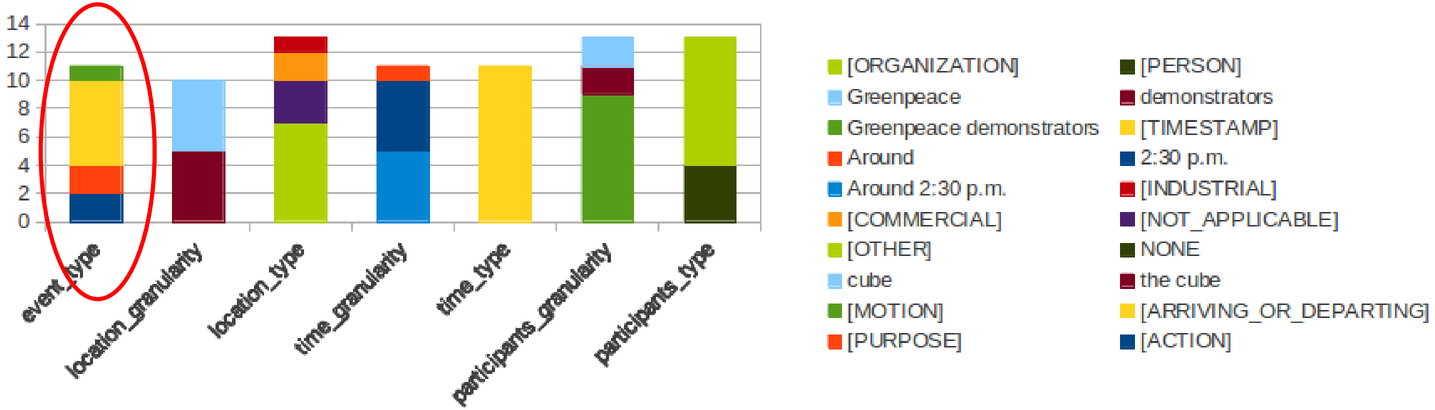
ACTION (18.2%)

MOTION (9.1%)

[ENTERED]

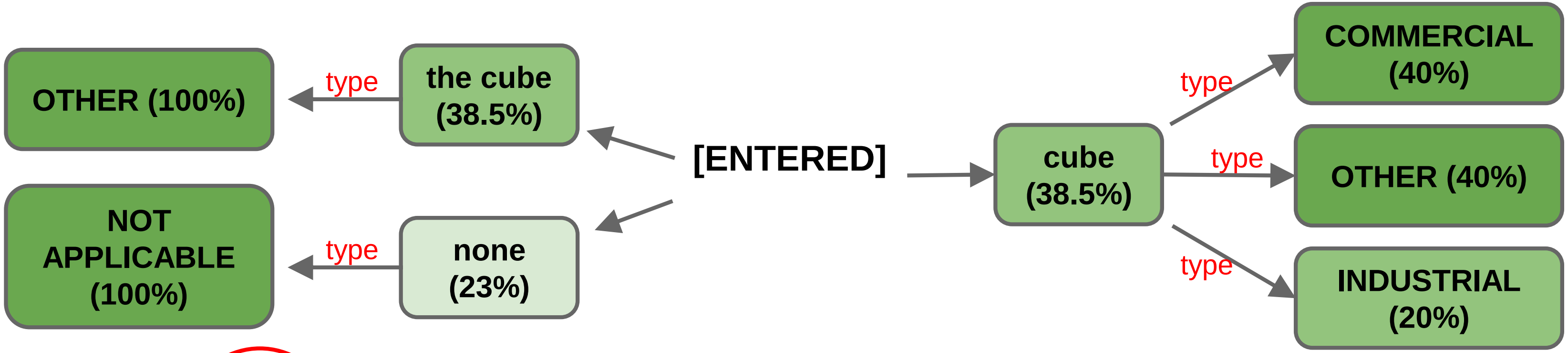
ARRIVING_OR_ DEPARTING (54.5%)

PURPOSE (18.2%)



Event Location Disagreement

*Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [**ENTERED**] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.*



OTHER (100%)

type

the cube (38.5%)

NOT APPLICABLE (100%)

type

none (23%)

[ENTERED]

cube (38.5%)

type

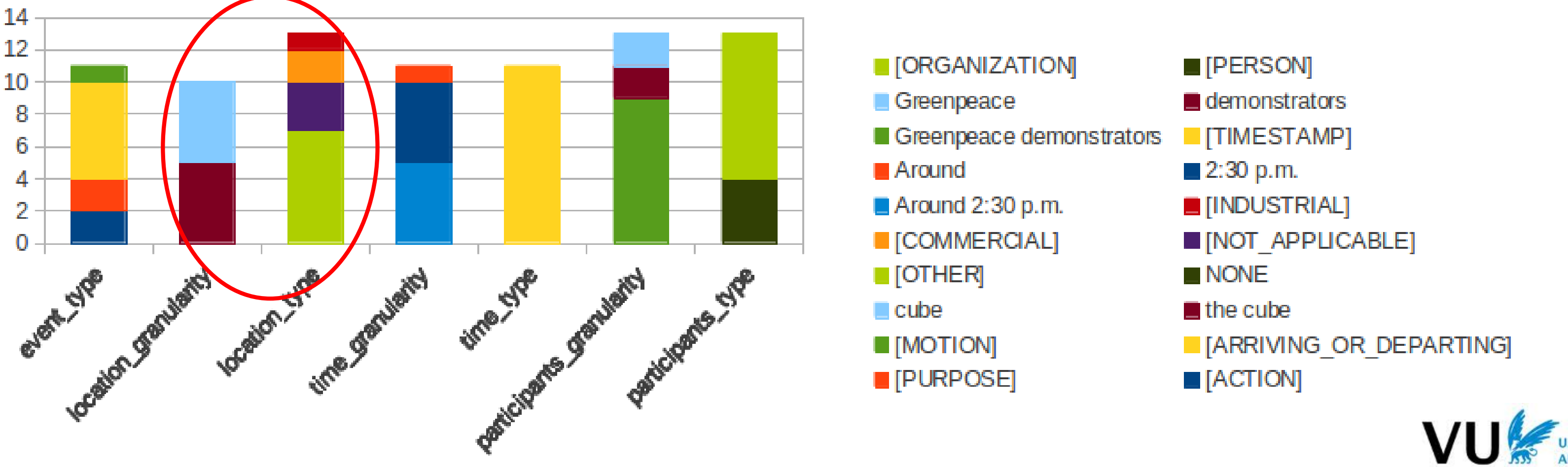
COMMERCIAL (40%)

type

OTHER (40%)

type

INDUSTRIAL (20%)

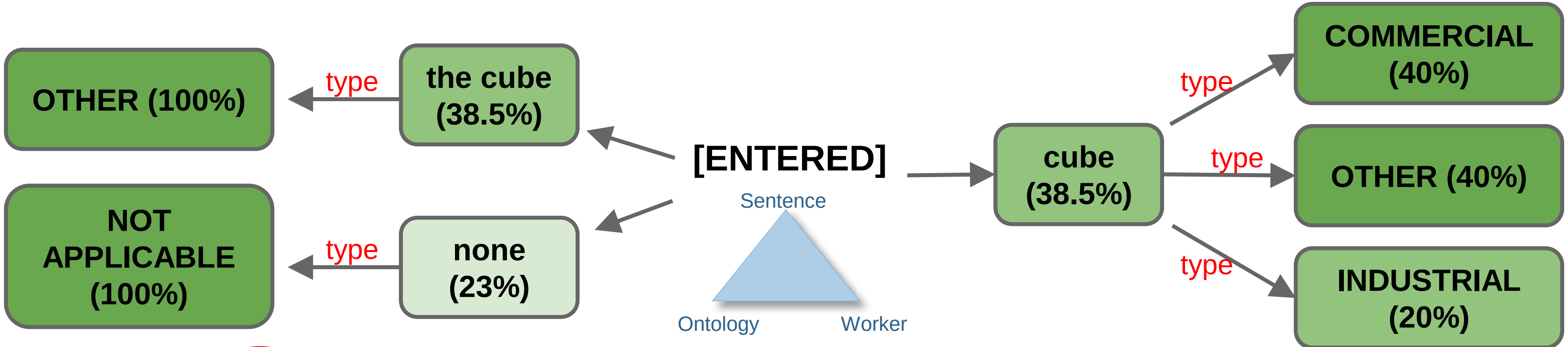


IBM

VU UNIVERSITY AMSTERDAM

Event Location Disagreement

*Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [**ENTERED**] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.*



OTHER (100%)

type

the cube (38.5%)

NOT APPLICABLE (100%)

type

none (23%)

[ENTERED]

Sentence

Ontology

Worker

cube (38.5%)

type

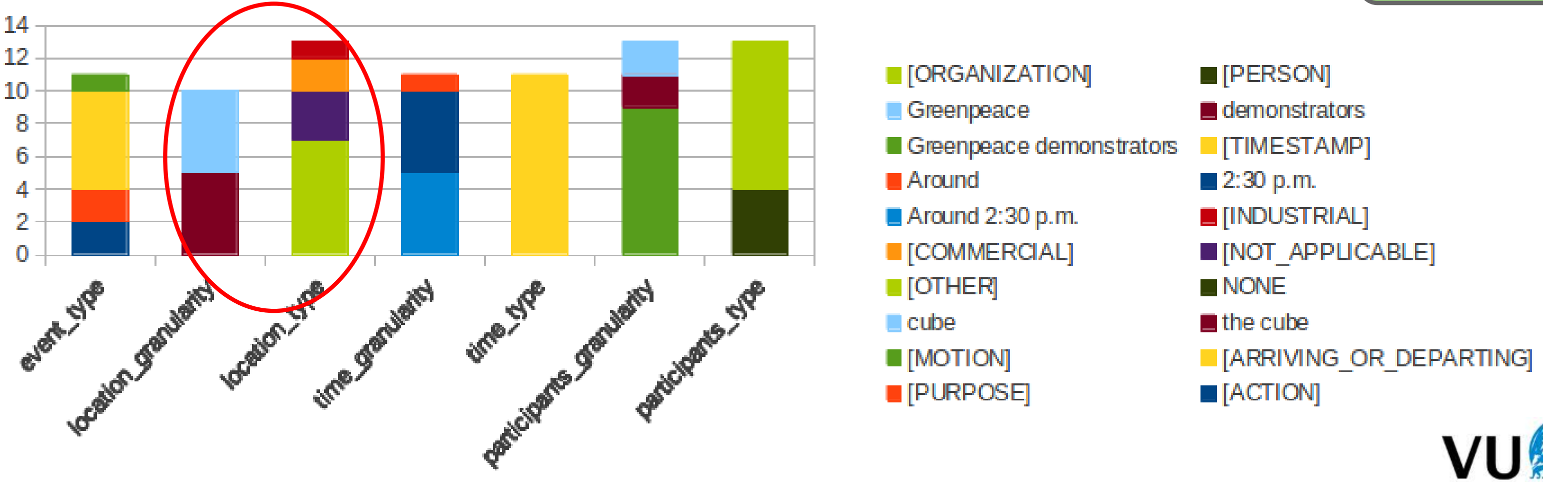
COMMERCIAL (40%)

type

OTHER (40%)

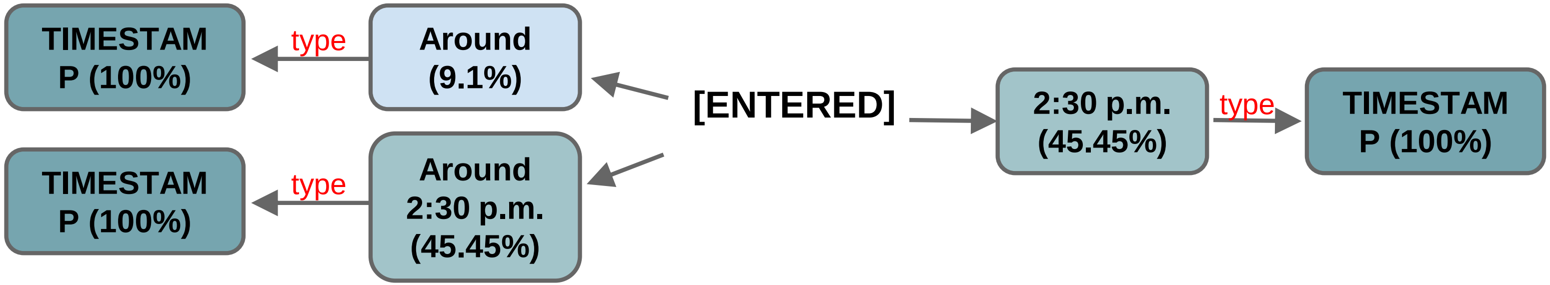
type

INDUSTRIAL (20%)



Event Time Disagreement

*Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [**ENTERED**] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.*



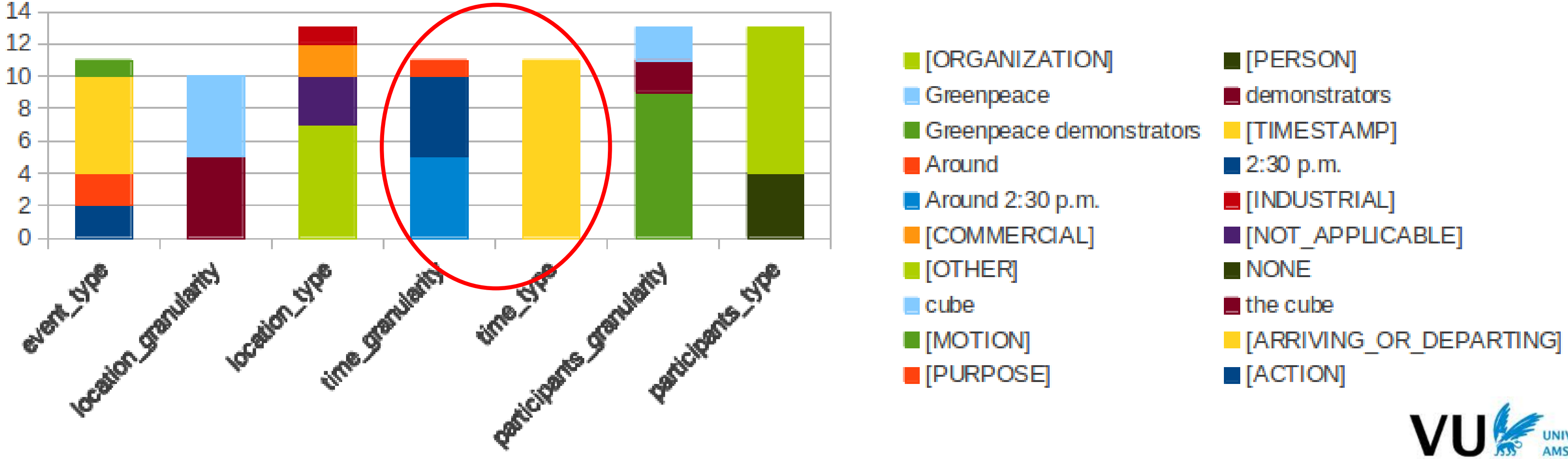
TIMESTAMP (100%) ← type — Around (9.1%)

TIMESTAMP (100%) ← type — Around 2:30 p.m. (45.45%)

[ENTERED] → Around (9.1%)

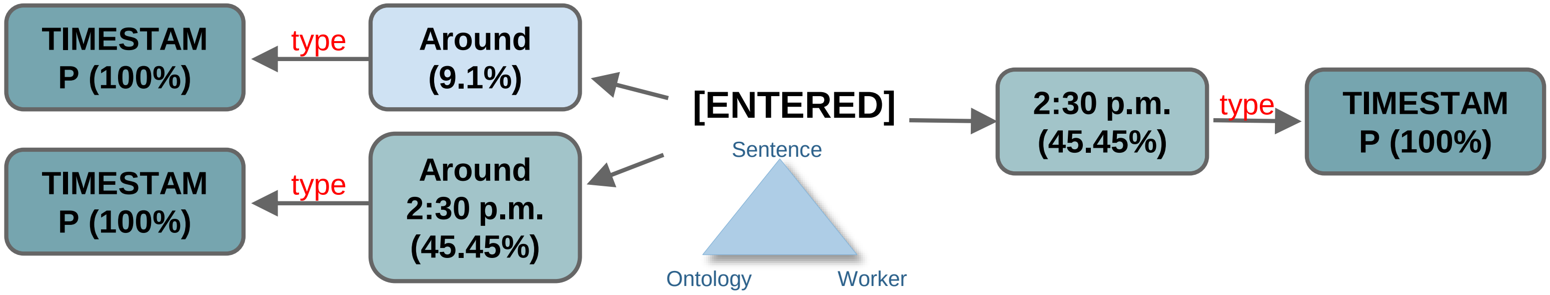
[ENTERED] → Around 2:30 p.m. (45.45%)

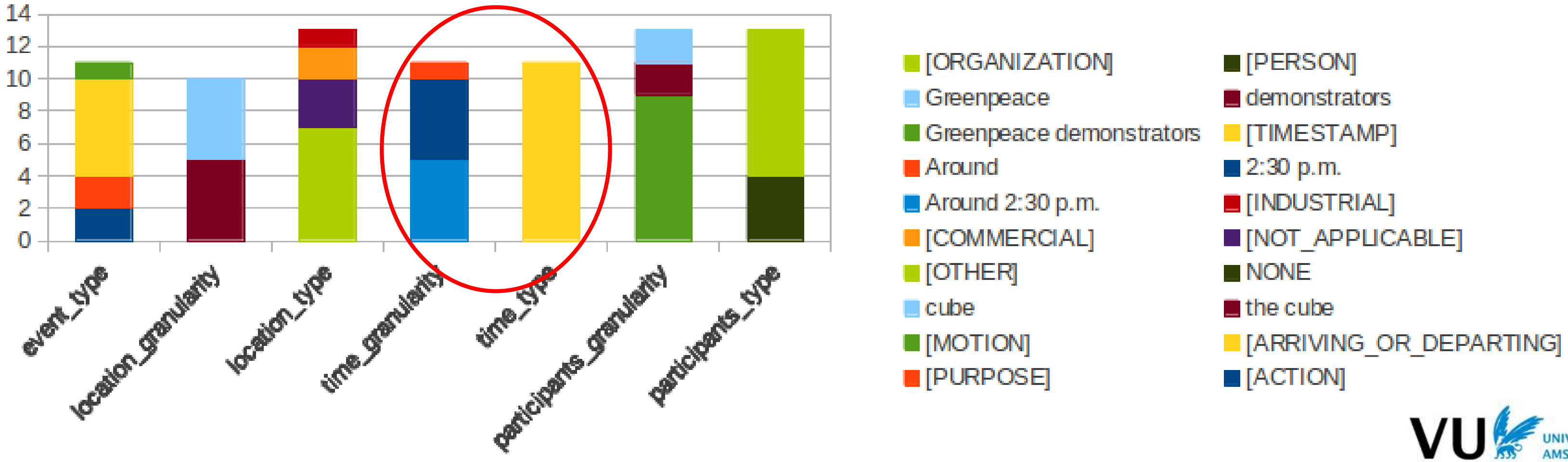
[ENTERED] → 2:30 p.m. (45.45%) — type → TIMESTAMP (100%)



Event Time Disagreement

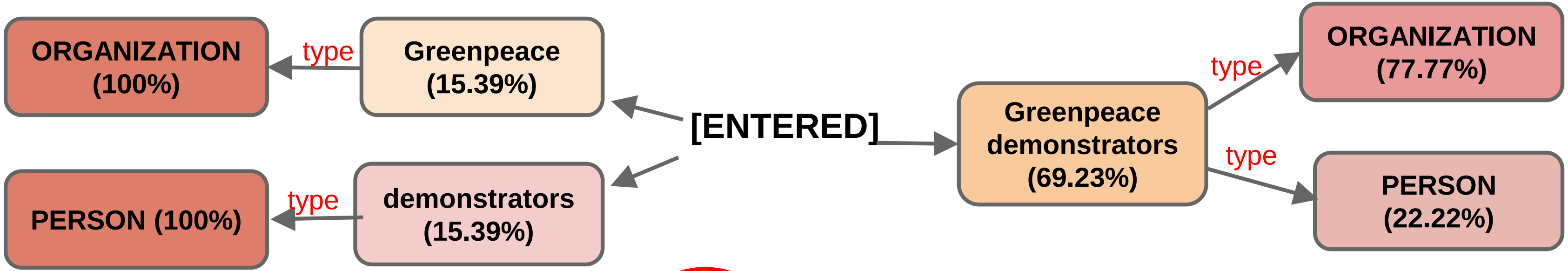
*Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [**ENTERED**] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.*





Event Participant Disagreement

*Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [**ENTERED**] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.*



ORGANIZATION (100%)

type

Greenpeace (15.39%)

PERSON (100%)

type

demonstrators (15.39%)

[ENTERED]

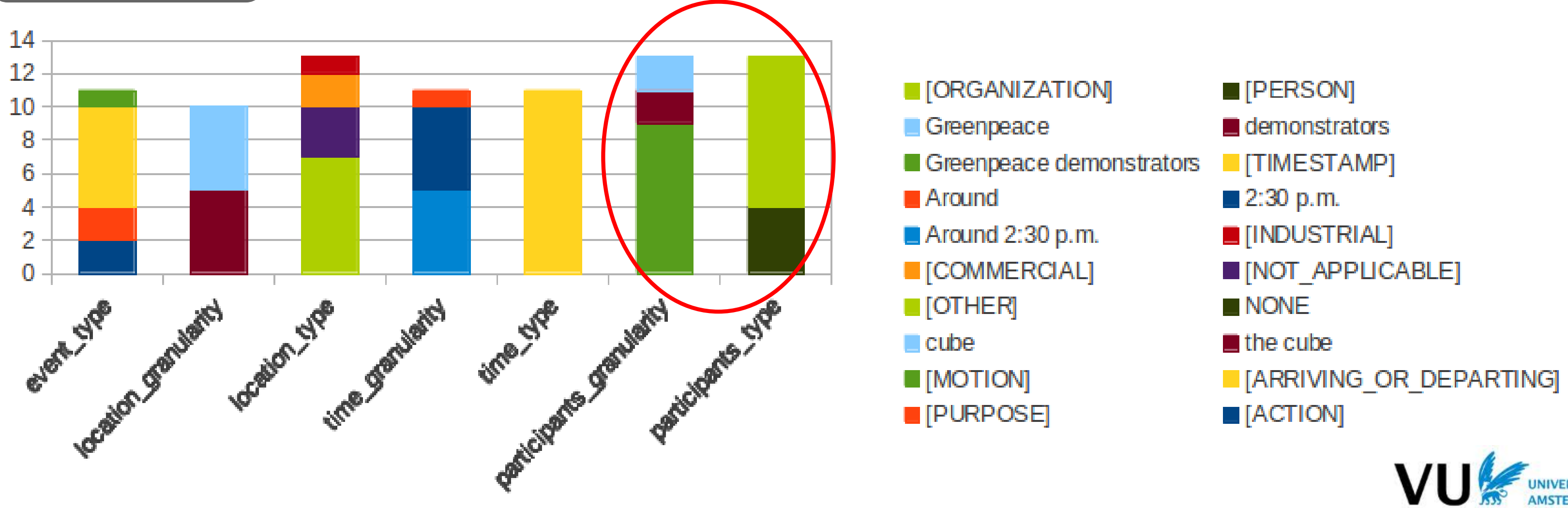
Greenpeace demonstrators (69.23%)

type

ORGANIZATION (77.77%)

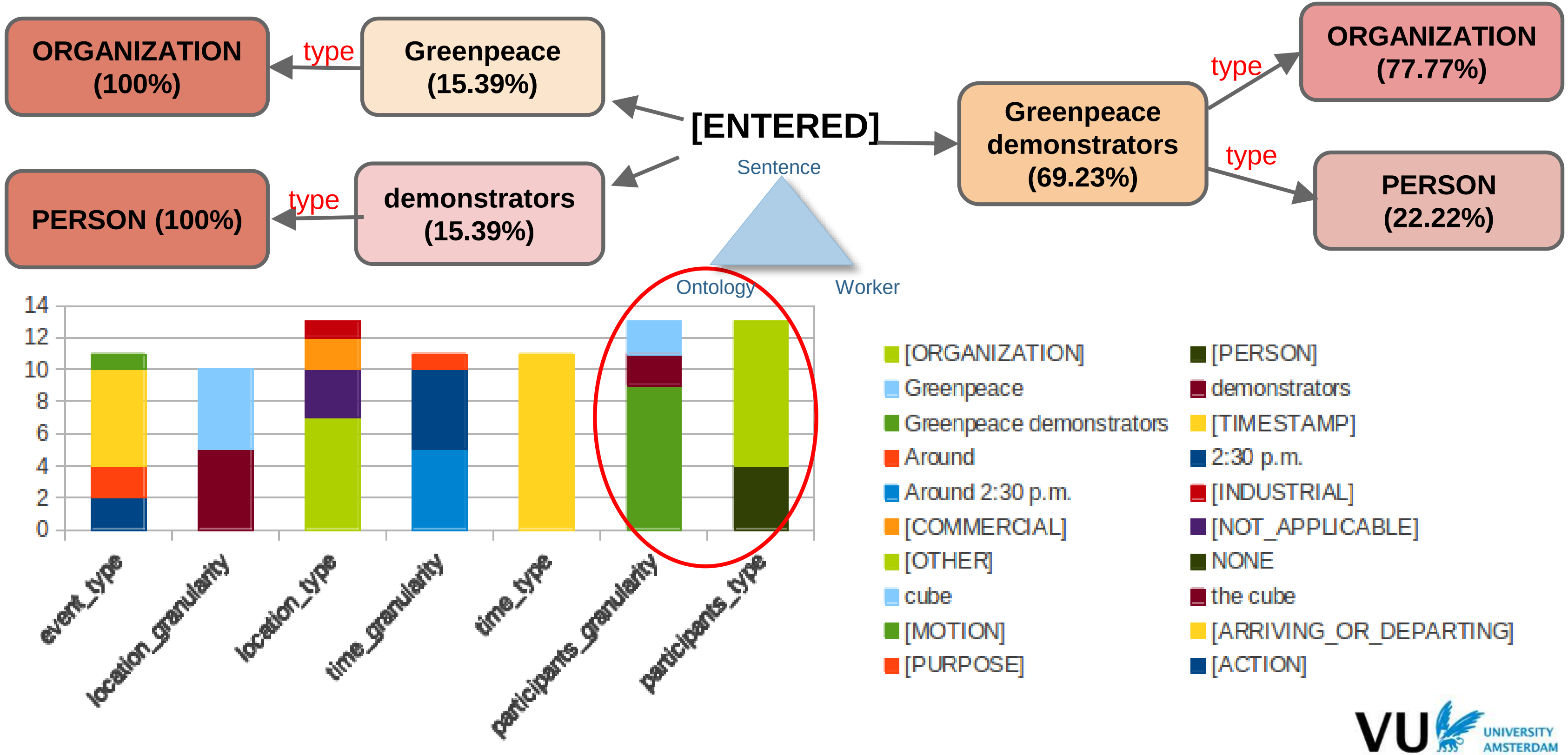
type

PERSON (22.22%)



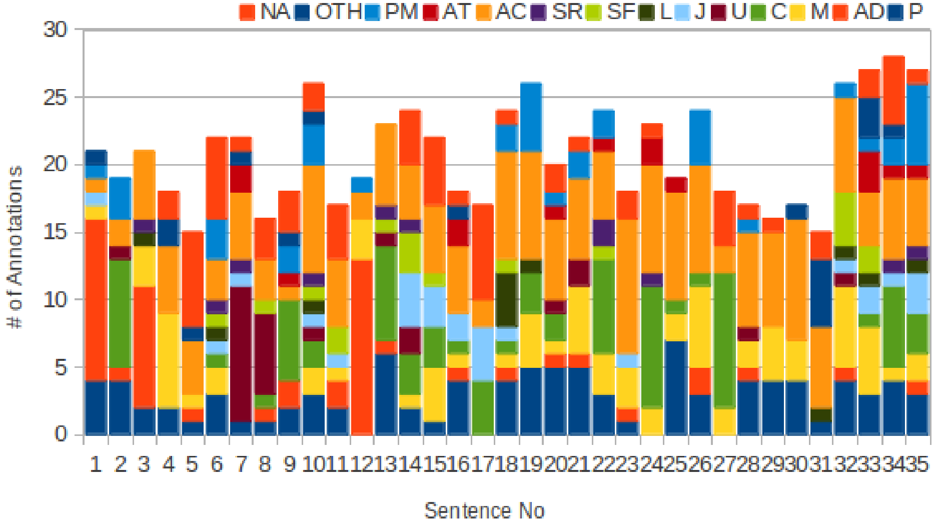
Event Participant Disagreement

*Around 2:30 p.m., as if delivering birthday greetings, several Greenpeace demonstrators [**ENTERED**] the cube clutching helium-filled balloons, which were the shape and color of charcoal briquettes.*

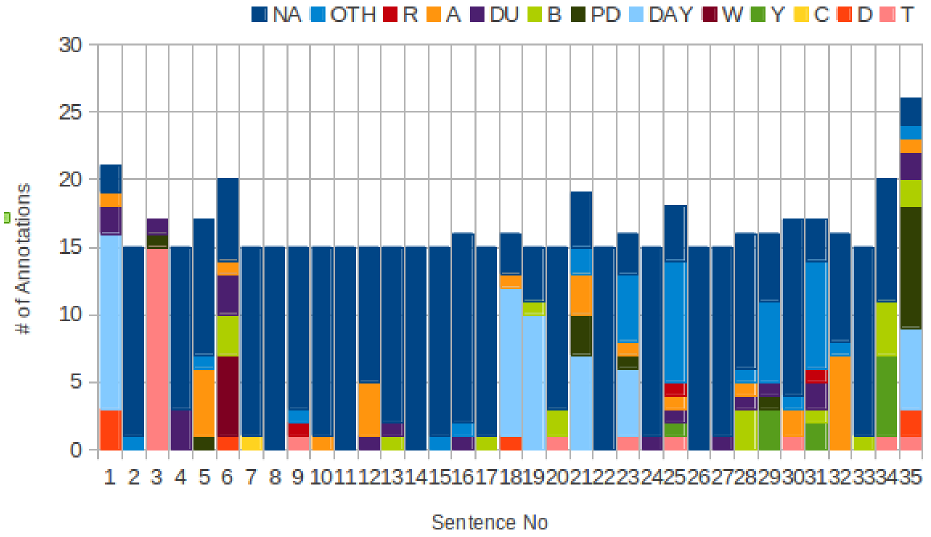


Comparative Annotation Distribution

Event Type Distribution



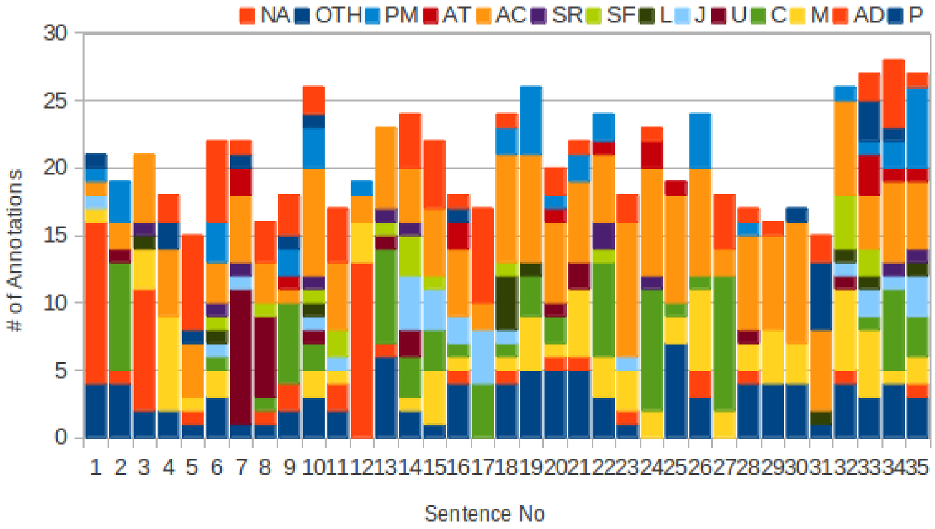
Time Type Distribution



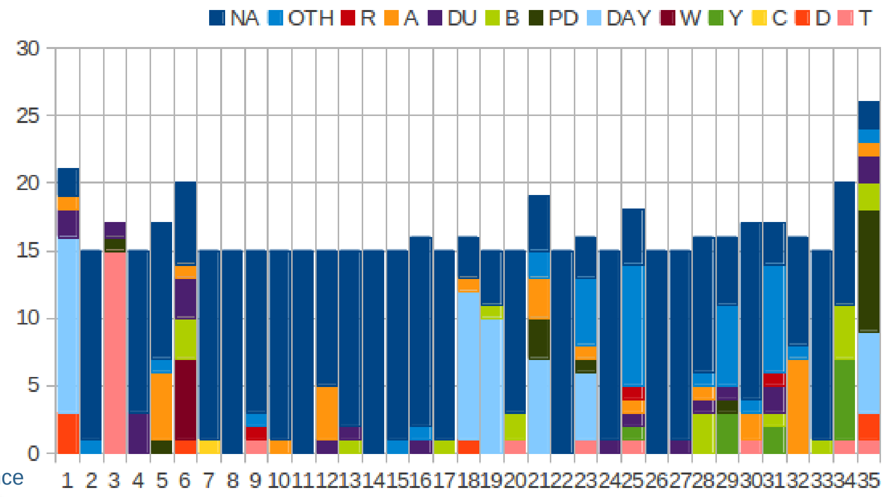
*The high disagreement for event type **across all sentences** likely indicates problems with the ontology. These event types are difficult to distinguish between. The event classes may overlap, be confusable, too vague, etc.*

Comparative Annotation Distribution

Event Type Distribution



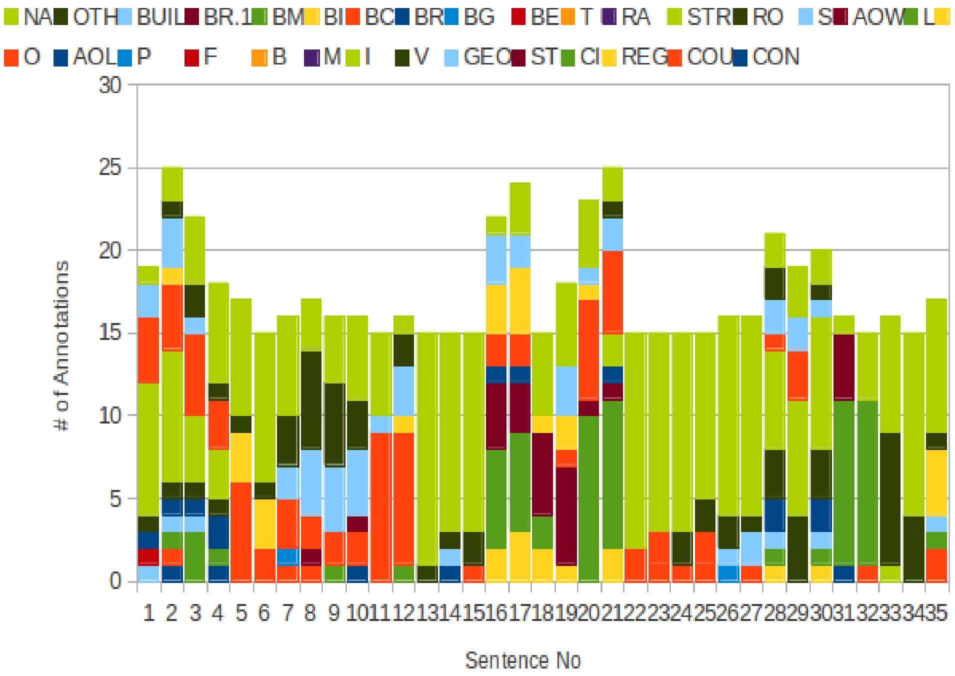
Time Type Distribution



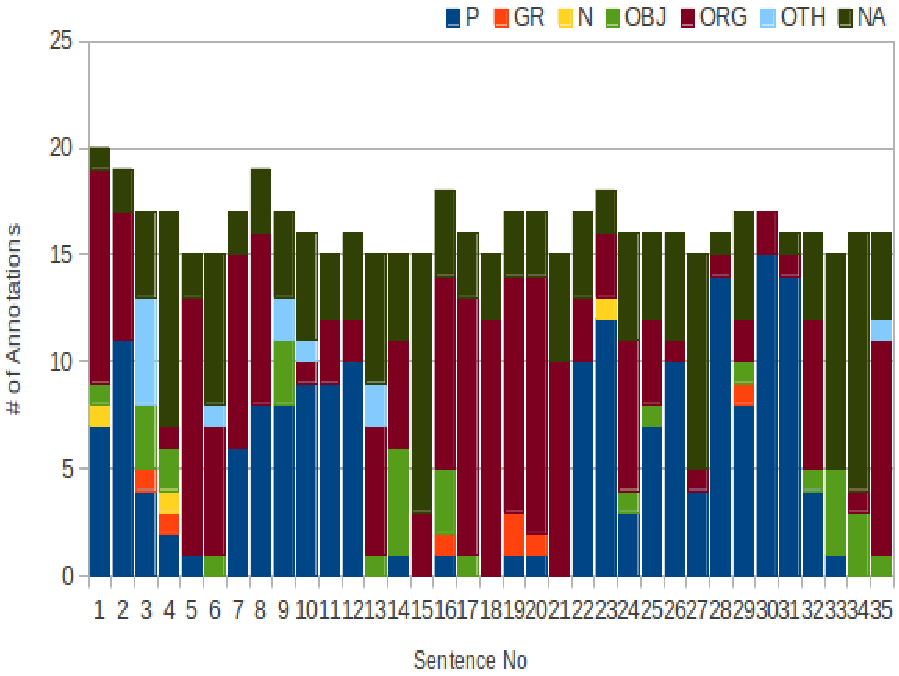
Sentence
Ontology
Worker

The high disagreement for event type ***across all sentences*** *likely indicates problems with the ontology. These event types are difficult to distinguish between. The event classes may overlap, be confusable, too vague, etc.*

Comparative Annotation Distribution



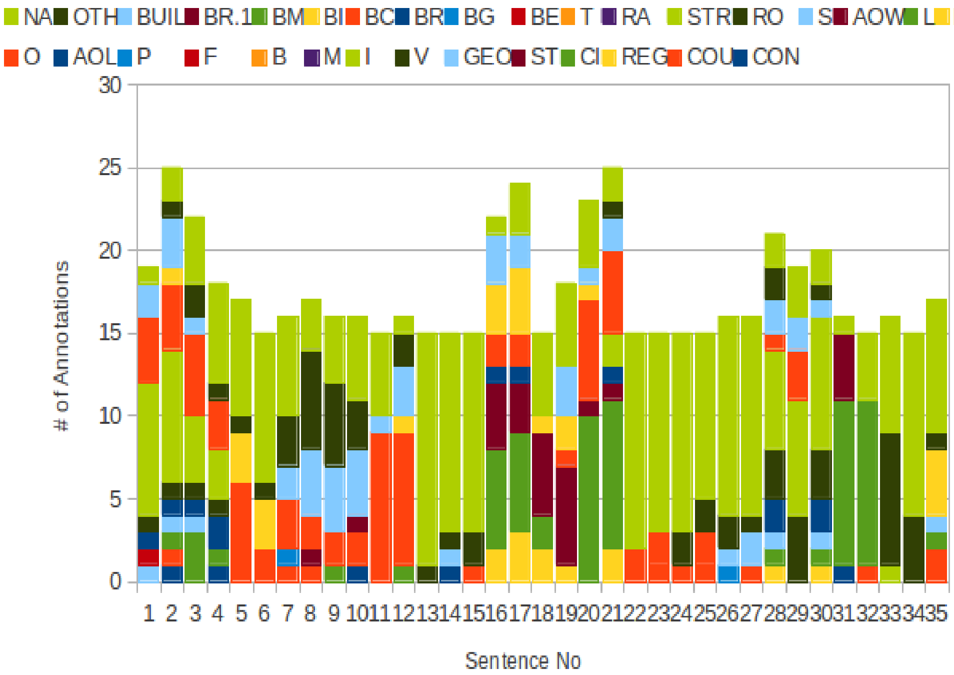
Location Type Distribution



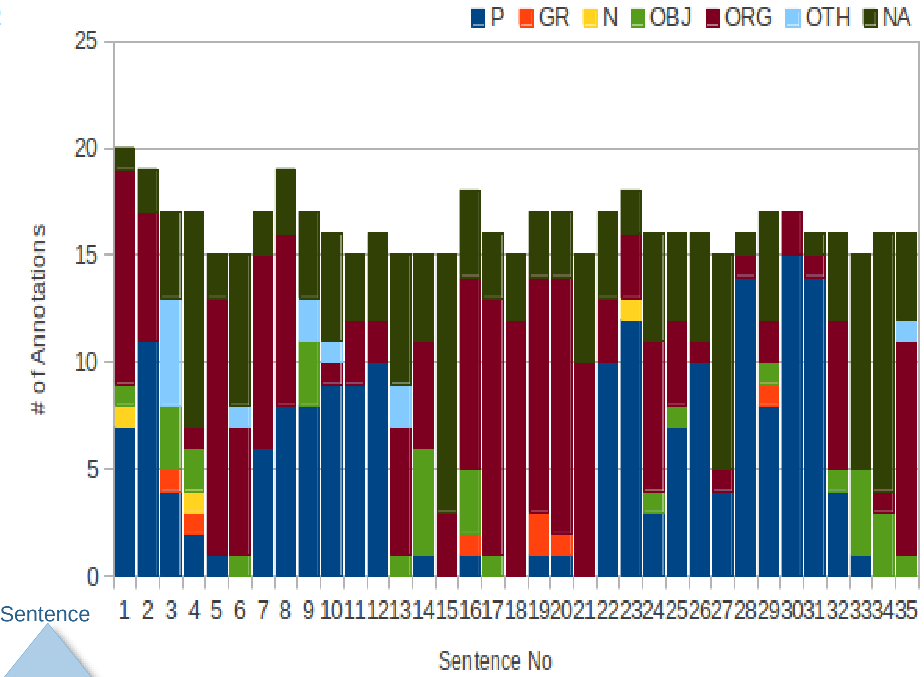
Participant Type Distribution

Comparative Annotation Distribution

Location Type Distribution



Participant Type Distribution



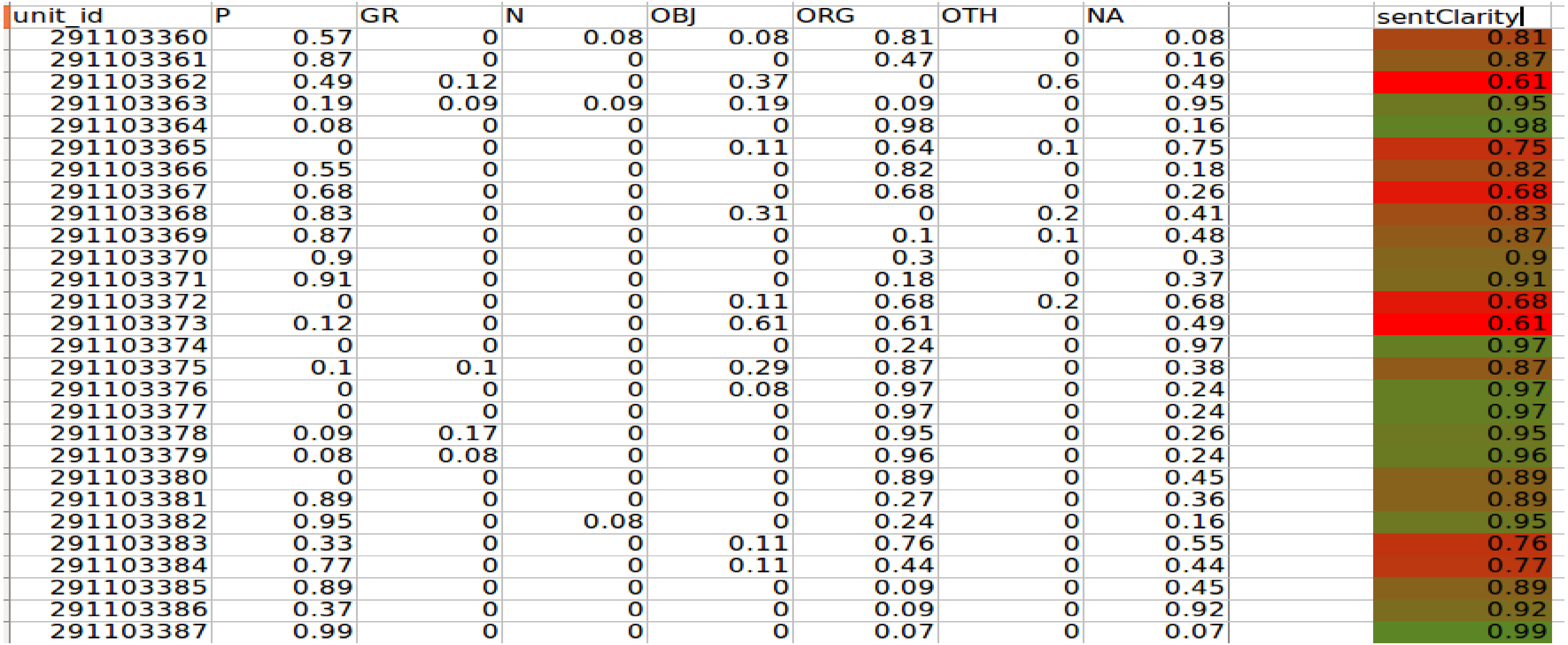
Sentence

Ontology

Worker

Sentence Clarity

Identifies sentences that are unclear or ambiguous based on the distribution of types



unit_id	P	GR	N	OBJ	ORG	OTH	NA		sentClarity
291103360	0.57	0	0.08	0.08	0.81	0	0.08		0.81
291103361	0.87	0	0	0	0.47	0	0.16		0.87
291103362	0.49	0.12	0	0.37	0	0.6	0.49		0.61
291103363	0.19	0.09	0.09	0.19	0.09	0	0.95		0.95
291103364	0.08	0	0	0	0.98	0	0.16		0.98
291103365	0	0	0	0.11	0.64	0.1	0.75		0.75
291103366	0.55	0	0	0	0.82	0	0.18		0.82
291103367	0.68	0	0	0	0.68	0	0.26		0.68
291103368	0.83	0	0	0.31	0	0.2	0.41		0.83
291103369	0.87	0	0	0	0.1	0.1	0.48		0.87
291103370	0.9	0	0	0	0.3	0	0.3		0.9
291103371	0.91	0	0	0	0.18	0	0.37		0.91
291103372	0	0	0	0.11	0.68	0.2	0.68		0.68
291103373	0.12	0	0	0.61	0.61	0	0.49		0.61
291103374	0	0	0	0	0.24	0	0.97		0.97
291103375	0.1	0.1	0	0.29	0.87	0	0.38		0.87
291103376	0	0	0	0.08	0.97	0	0.24		0.97
291103377	0	0	0	0	0.97	0	0.24		0.97
291103378	0.09	0.17	0	0	0.95	0	0.26		0.95
291103379	0.08	0.08	0	0	0.96	0	0.24		0.96
291103380	0	0	0	0	0.89	0	0.45		0.89
291103381	0.89	0	0	0	0.27	0	0.36		0.89
291103382	0.95	0	0.08	0	0.24	0	0.16		0.95
291103383	0.33	0	0	0.11	0.76	0	0.55		0.76
291103384	0.77	0	0	0.11	0.44	0	0.44		0.77
291103385	0.89	0	0	0	0.09	0	0.45		0.89
291103386	0.37	0	0	0	0.09	0	0.92		0.92
291103387	0.99	0	0	0	0.07	0	0.07		0.99

Spam Detection

aims to efficiently remove the spam and low-quality contributors

- ***filter sentences based on their clarity score*** in order to avoid penalizing workers for contributing on difficult or ambiguous sentences
- ***apply the worker metrics to analyze worker agreement*** on a specific putative event or across all the putative events that s(he) solved
- ***apply explanation-based filters*** in order to assess the overall quality of each worker rational

Spam Detection

Worker Metrics Evaluation

	accuracy	precision	recall	F-measure
event type	87	88	91	89
event location	81	75	92	82
event time	81	75	90	81
event participants	88	92	89	91

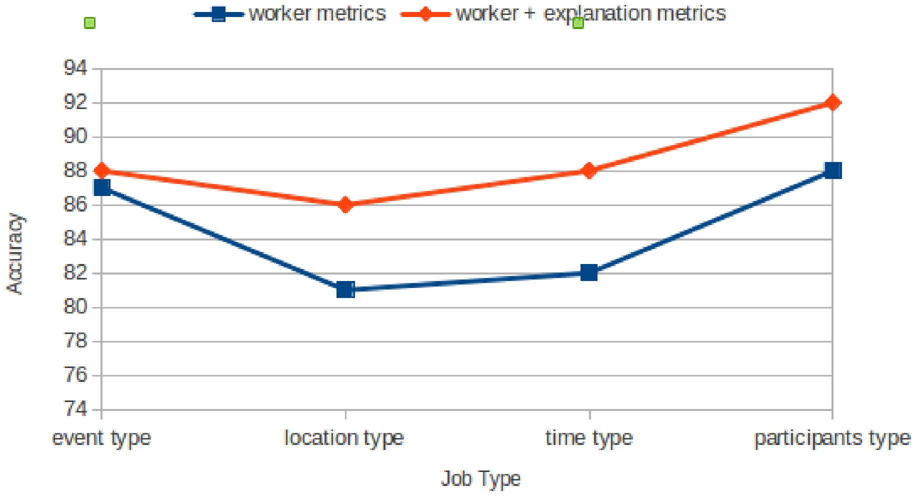
Worker Metrics and Evaluation-based Filters Evaluation

	accuracy	precision	recall	F-measure
event type	88	91	89	90
event location	86	82	97	88
event time	88	86	97	91
event participants	92	94	94	94

the explanation-based filters are able to increase the accuracy of detecting spam and low-quality workers with at least 5%, which leads to a better interpretation of the crowdsourced data and representation of the events

Spam Detection

Worker Metrics Evaluation

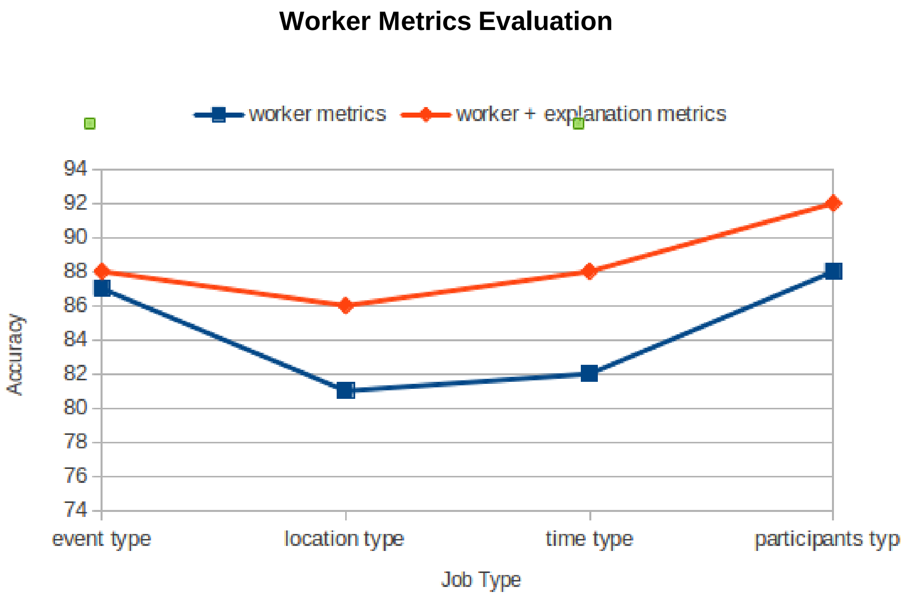


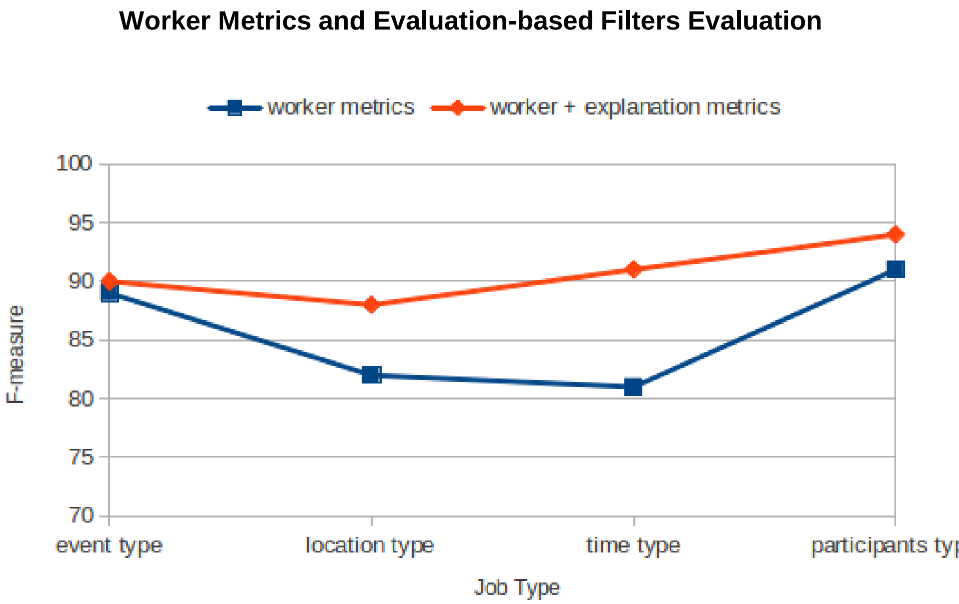
Worker Metrics and Evaluation-based Filters Evaluation

	accuracy	**precision**	**recall**	**F-measure**
event type	88	91	89	90
event location	86	82	97	88
event time	88	86	97	91
event participants	92	94	94	94

the explanation-based filters are able to increase the accuracy of detecting spam and low-quality workers with at least 5%, which leads to a better interpretation of the crowdsourced data and representation of the events

Spam Detection





the explanation-based filters are able to increase the accuracy of detecting spam and low-quality workers with at least 5%, which leads to a better interpretation of the crowdsourced data and representation of the events

What more …

Understand human disagreement on event extraction with focus on ***ambiguity***:

- Would *different classification (ontology)* of putative events perform better?
- Does the *overlapping of the types (ontology)* influence the results?
- Identify *the right role fillers (per event) for multiple putative events*.
- Would *event clustering* help with determining the most appropriate structure of the event and its role fillers?

Conclusions

- **disagreement metrics adaptable across domains** - helped us to understand a bit more the vagueness and the clarity of a sentence/putative event

- clarity or the vagueness of sentences help **select the good cases for automated training**

- **micro-task template design was most difficult process** as it aimed at harnessing diversity and disagreement, while making the task understandable and affordable for the crowdsourcers

- *understanding disagreement can help us understand event semantics*

http://crowd-watson.nl