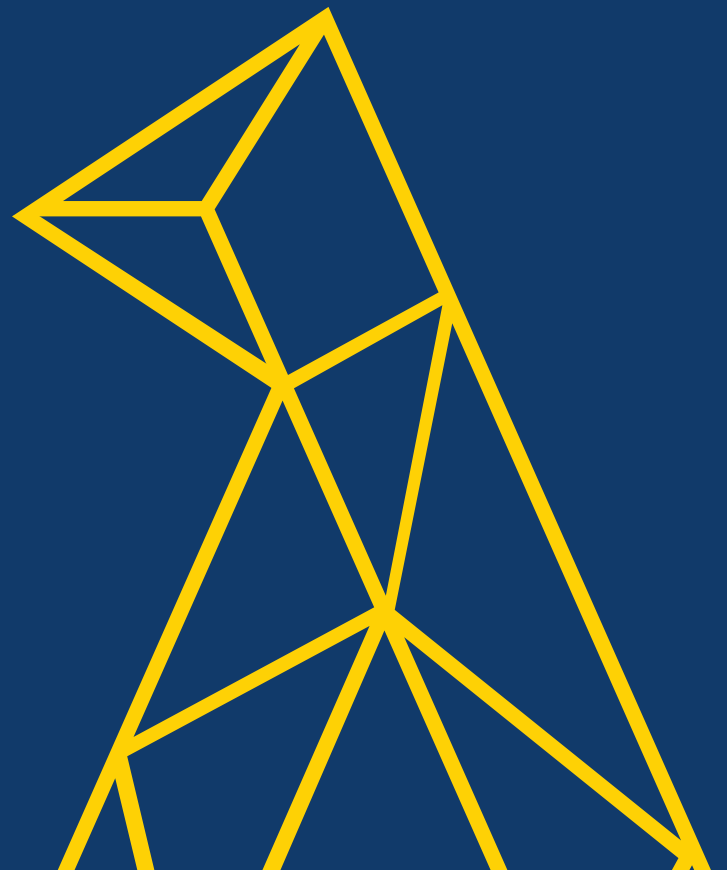
Understanding complex situation descriptions

Aaron Steven White
University of Rochester

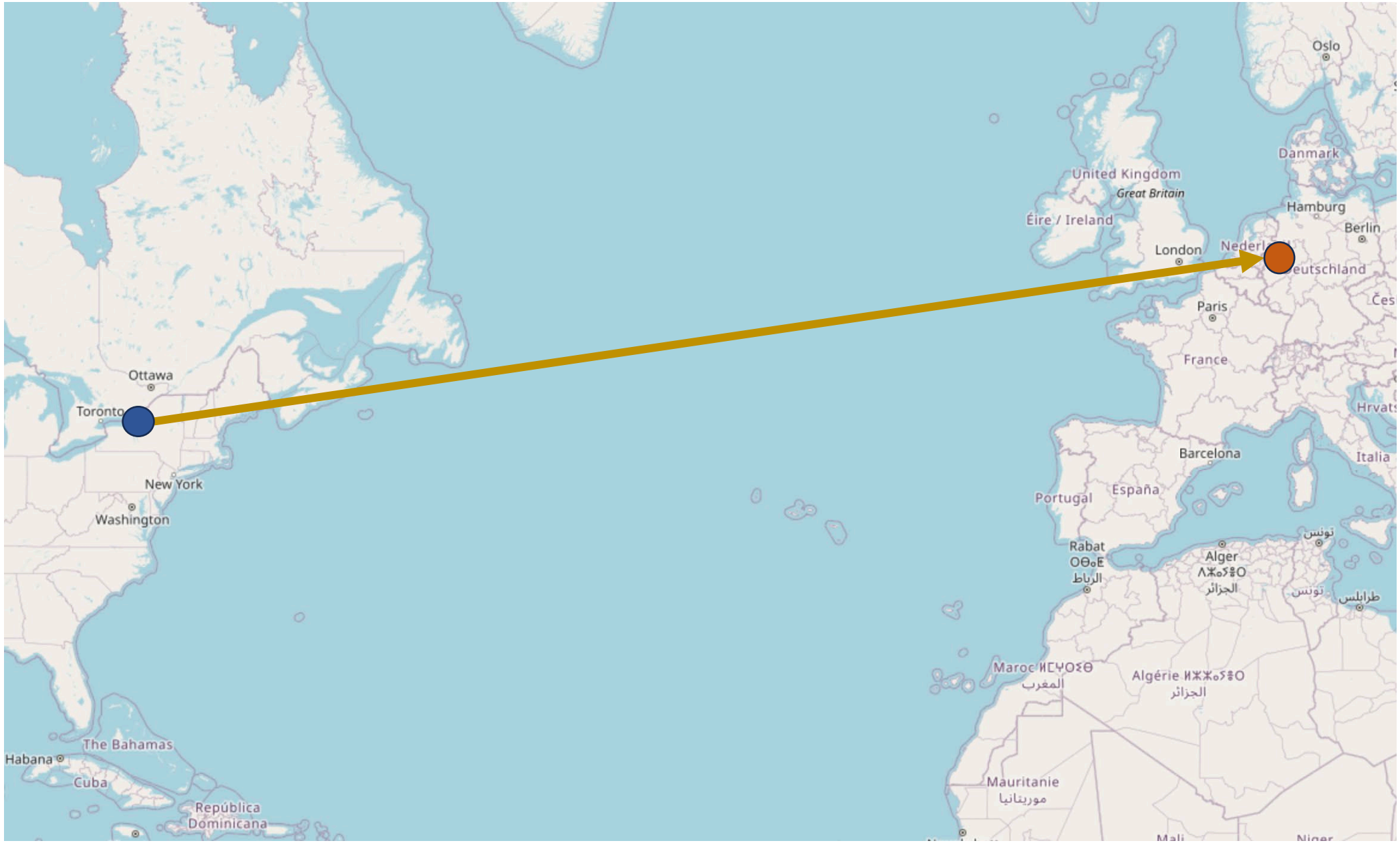
Keynote
NALOMA
4 August 2025



What we do

We use natural language to convey information about *situations*: things that happen or stuff that is true.

I traveled from Rochester to Bochum yesterday.



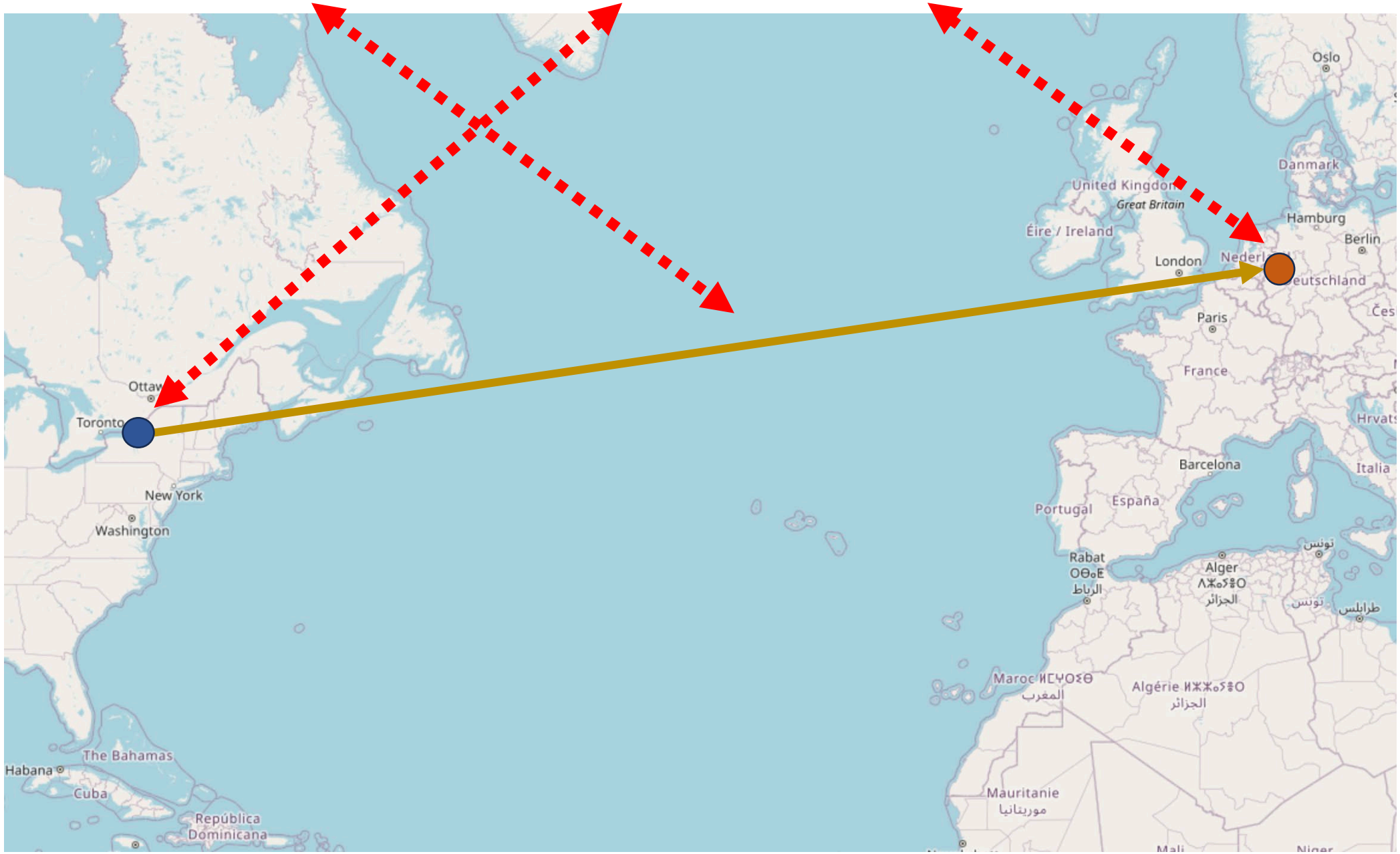
What we do

We use natural language to convey information about *situations*: things that happen or stuff that is true.

How we do this

Systematic relationships between the way we conceptualize situations and the way we describe them.

I traveled from Rochester to Bochum yesterday.



What we do

We use natural language to convey information about *situations*: things that happen or stuff that is true.

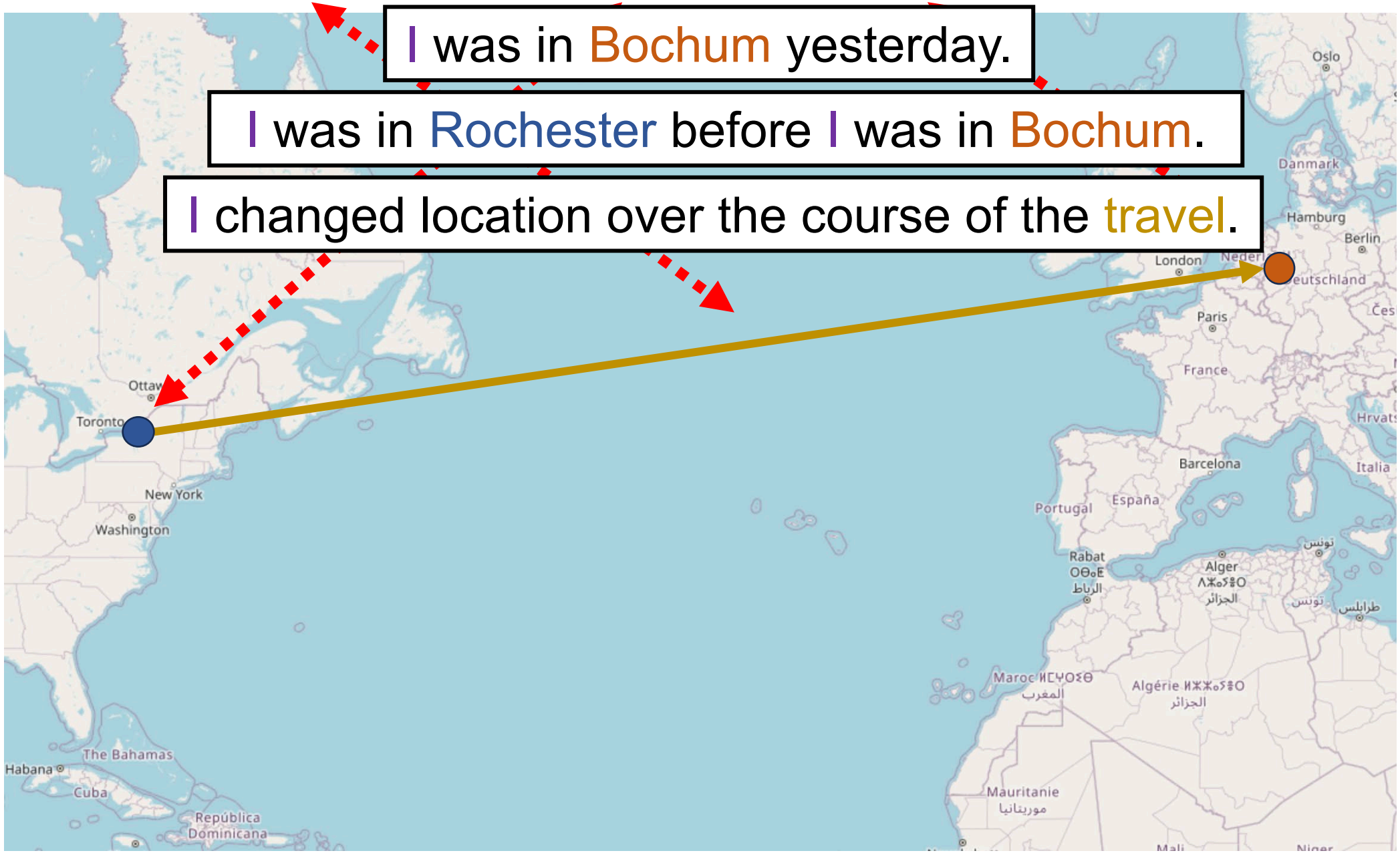
How we do this

Systematic relationships between the way we conceptualize situations and the way we describe them.

What this allows us to do

Draw inferences that go beyond what one strictly says in describing a situation.

I traveled from Rochester to Bochum yesterday.

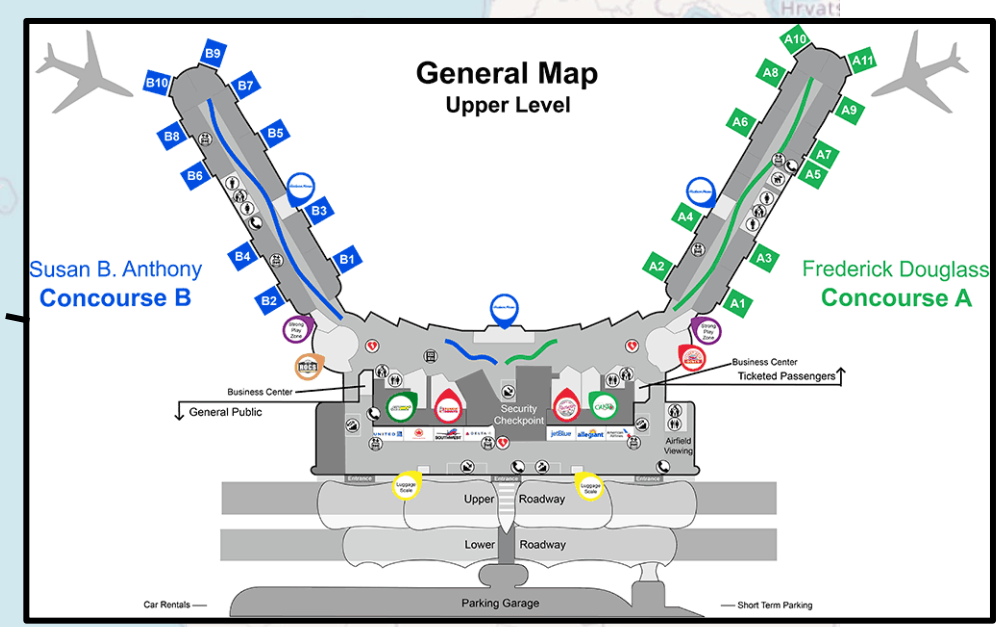


I traveled from Rochester to Bochum yesterday.

I took a flight yesterday.

I traveled to the Rochester airport yesterday.

I went through airport security at the Rochester airport yesterday.



Question

How do we design systems that capture the inferences we draw about situations based on their descriptions?

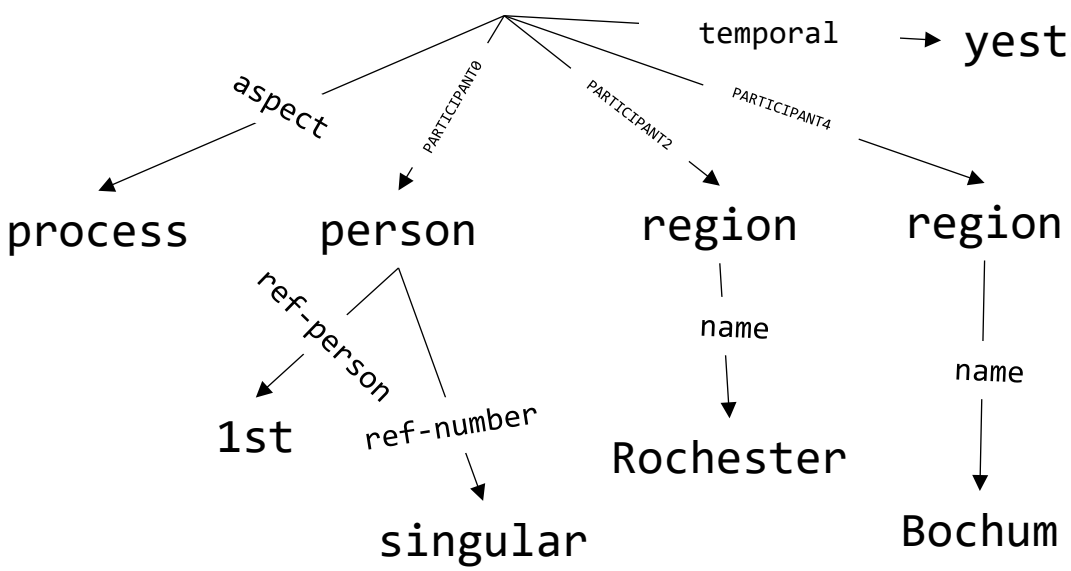
Ontology-factored approach

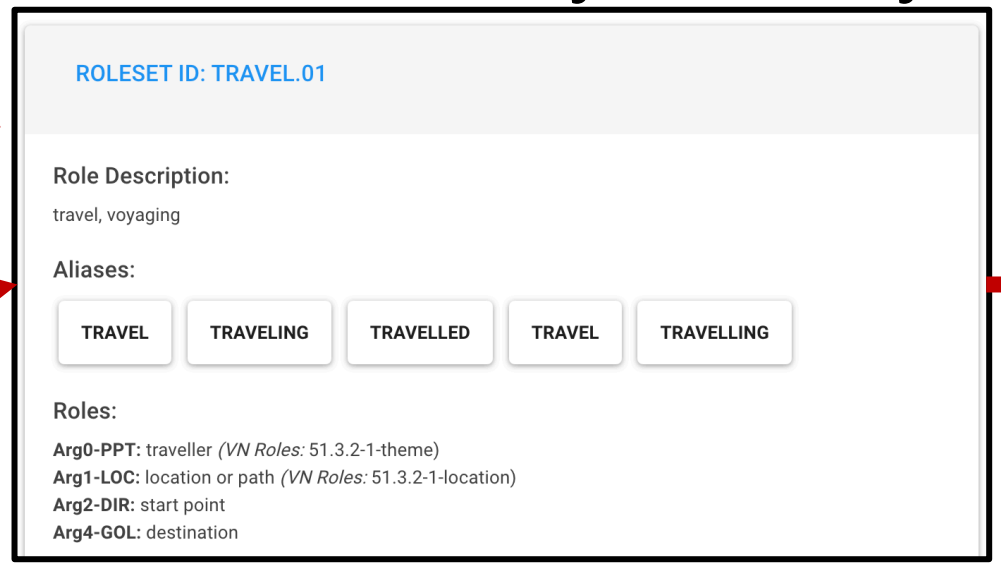
Map situation description to symbolic situation ontology and draw inferences using rules stated over that ontology.

I traveled from Rochester to Bochum yesterday.

Uniform Meaning Representation
Van Gysel et al. 2021

PropBank
Palmer et al. 2005





VerbNet
Kipper-Schuler 2005

Aaron was in Rochester yesterday.

Challenge #1: Expense

Annotated corpora are expensive to build and maintain because they require highly trained experts.

Challenge #2: Brittleness

Ontologies do not easily capture the ways in which context modulates the inferences that we draw.

An assassin in Colombia killed a federal judge on a Medellin street.

The assassin chose to be involved in killing the judge.

The antibody then kills the cell.

The antibody chose to be involved in killing the cell.

She was untrained and, in one botched job, killed a client.

She chose to be involved in killing the client.

ALIVE(e1 , Patient)

DO(e2 , ?Agent)

ient)

Definition:
Actor in an event who initiates and carries out the event intentionally or consciously, and who exists independently of the event.

Example:
THEY decided to go.

Reisinger et al. 2015

Question

How do we design systems that capture the inferences we draw about situations based on their descriptions?

Ontology-factored approach

Map situation description to symbolic situation ontology and draw inferences using rules stated over that ontology.

Ontology-free approach

Map situation descriptions to natural language strings expressing the inferences of interest.

QA-SRL/QAMR

UCD ***finished*** the 2006 championship as Dublin champions ,
by ***beating*** St Vincents in the final .

finished

- Who finished something? - UCD
- What did someone finish? - the 2006 championship
- What did someone finish something as? - Dublin champions
- How did someone finish something? - by beating St Vincents in the final

beating

- Who beat someone? - UCD
- When did someone beat someone? - in the final
- Who did someone beat? - St Vincents

He et al. 2015, Michael et al. 2018

Challenge #1: Lack of Abstraction

Inferences we can draw by templatizing a string don't provide information beyond that already in the string.

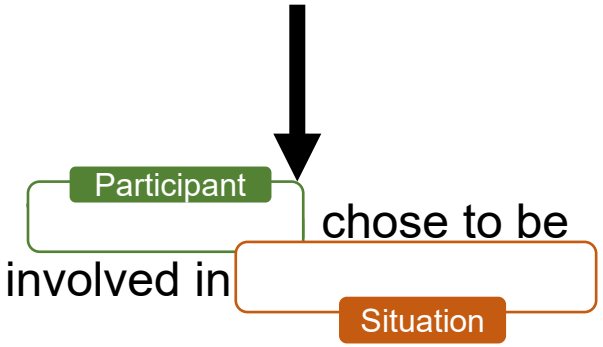
Key Idea: Representational Scaffolding

Ontologies provide guidance about what the interesting inferences are; that guidance undergirds multiple tasks.

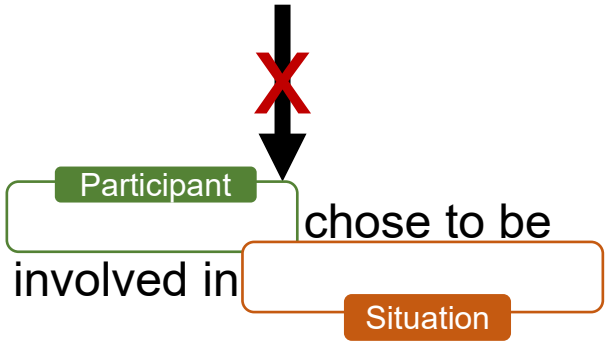
Part 1: Natural Language Inference

Leverage ontology descriptions to generate natural language inference templates filled by semantic parses.

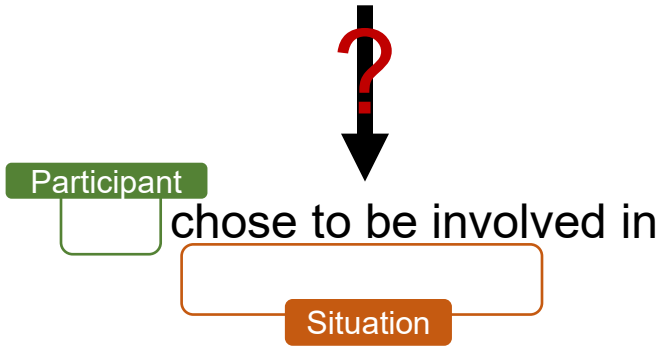
An assassin in Colombia killed a federal judge on a Medellin street.



The antibody then kills the cell.



She was untrained and, in one botched job, killed a client.



ALIVE(e1 , Patient)

DO(e2 , ?Agent)

ient)

Definition:
Actor in an event who initiates and carries out the event intentionally or consciously, and who exists independently of the event.

Example:
THEY decided to go.

Reisinger et al. 2015

Challenge #2: Too Much Abstraction

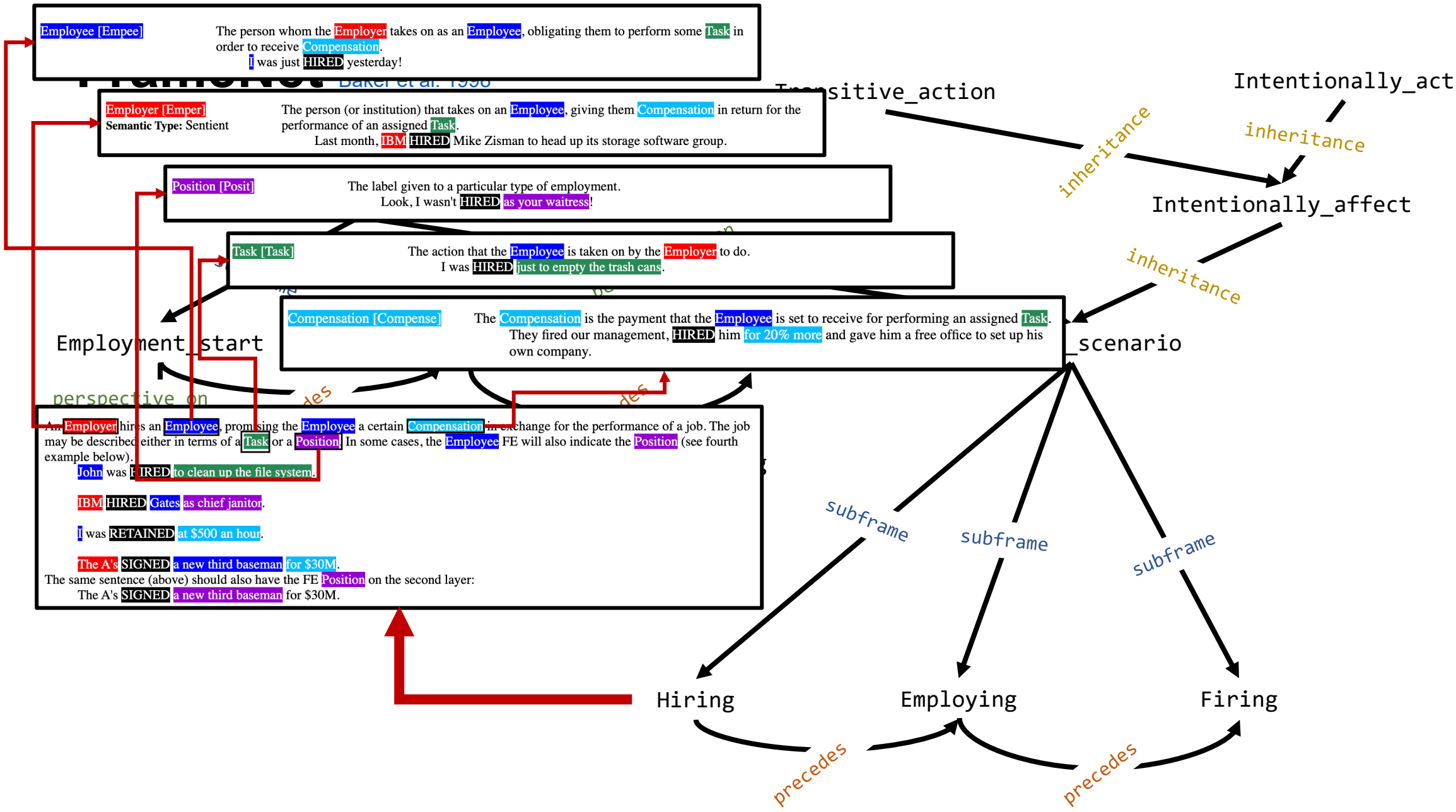
The inferences are interesting and broadly applicable but limited in their specificity.

Part 1: Natural Language Inference

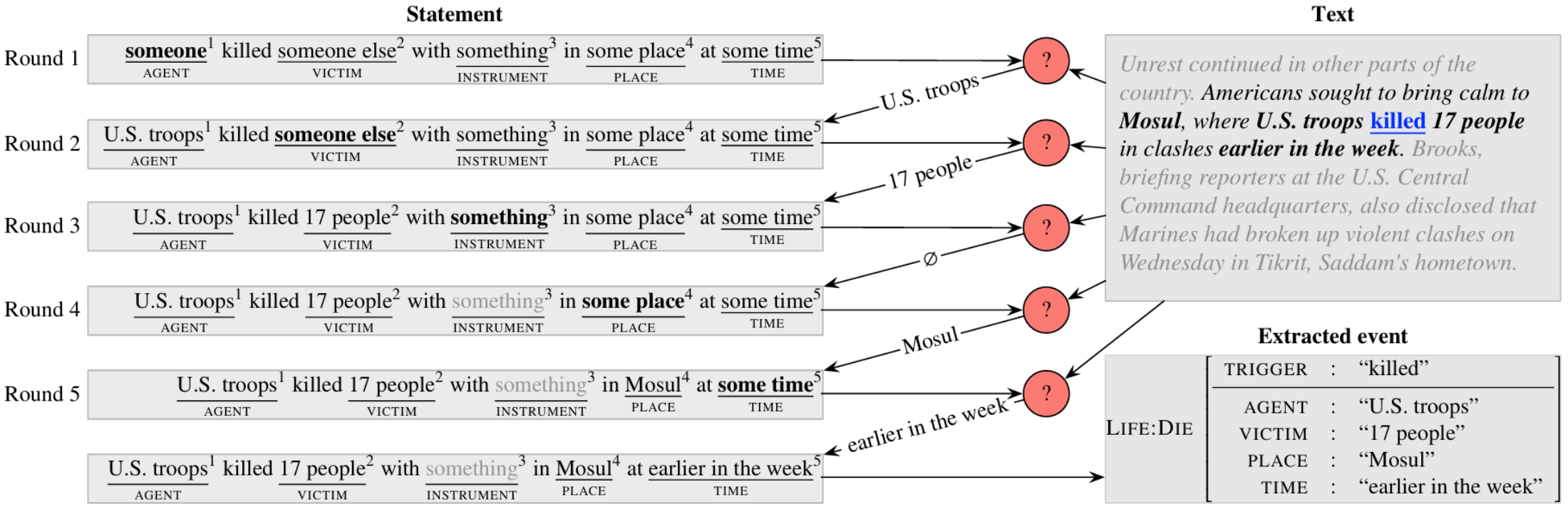
Leverage ontology descriptions to generate natural language inference templates filled by semantic parses.

Part 2: Template Filling

Leverage ontology structure to fill natural language templates across documents describing same event.


Employee [Empee]
The person whom the Employer takes on as an Employee, obligating them to perform some Task in order to receive Compensation.
I was just HIRED yesterday!
Employer [Emper]
Semantic Type: Sentient
The person (or institution) that takes on an Employee, giving them Compensation in return for the performance of an assigned Task.
Last month, IBM HIRED Mike Zisman to head up its storage software group.
Position [Posit]
The label given to a particular type of employment.
Look, I wasn't HIRED as your waitress!
Task [Task]
The action that the Employee is taken on by the Employer to do.
I was HIRED just to empty the trash cans.
Compensation [Compense]
The Compensation is the payment that the Employee is set to receive for performing an assigned Task.
They fired our management, HIRED him for 20% more and gave him a free office to set up his own company.
An Employer hires an Employee, promising the Employee a certain Compensation in exchange for the performance of a job. The job may be described either in terms of a Task or a Position. In some cases, the Employee FE will also indicate the Position (see fourth example below).
John was HIRED to clean up the file system.
IBM HIRED Gates as chief janitor.
I was RETAINED at $500 an hour.
The A's SIGNED a new third baseman for $30M.
The same sentence (above) should also have the FE Position on the second layer:
The A's SIGNED a new third baseman for $30M.
Transitive_action
Intentionally_act
Intentionally_affect
inheritance
inheritance
inheritance
Employment_start
_scenario
perspective_on
subframe
subframe
subframe
Hiring
Employing
Firing
precedes
precedes

“Reading the (annotation) manual”



Chen et al. 2020

Challenge #2: Too Much Abstraction

The inferences are interesting and broadly applicable but limited in their specificity.

Challenge #3: Not Enough Context-Sensitivity

Inferences are interesting and specific but lack sensitivity to the context of the situation description.

Part 1: Natural Language Inference

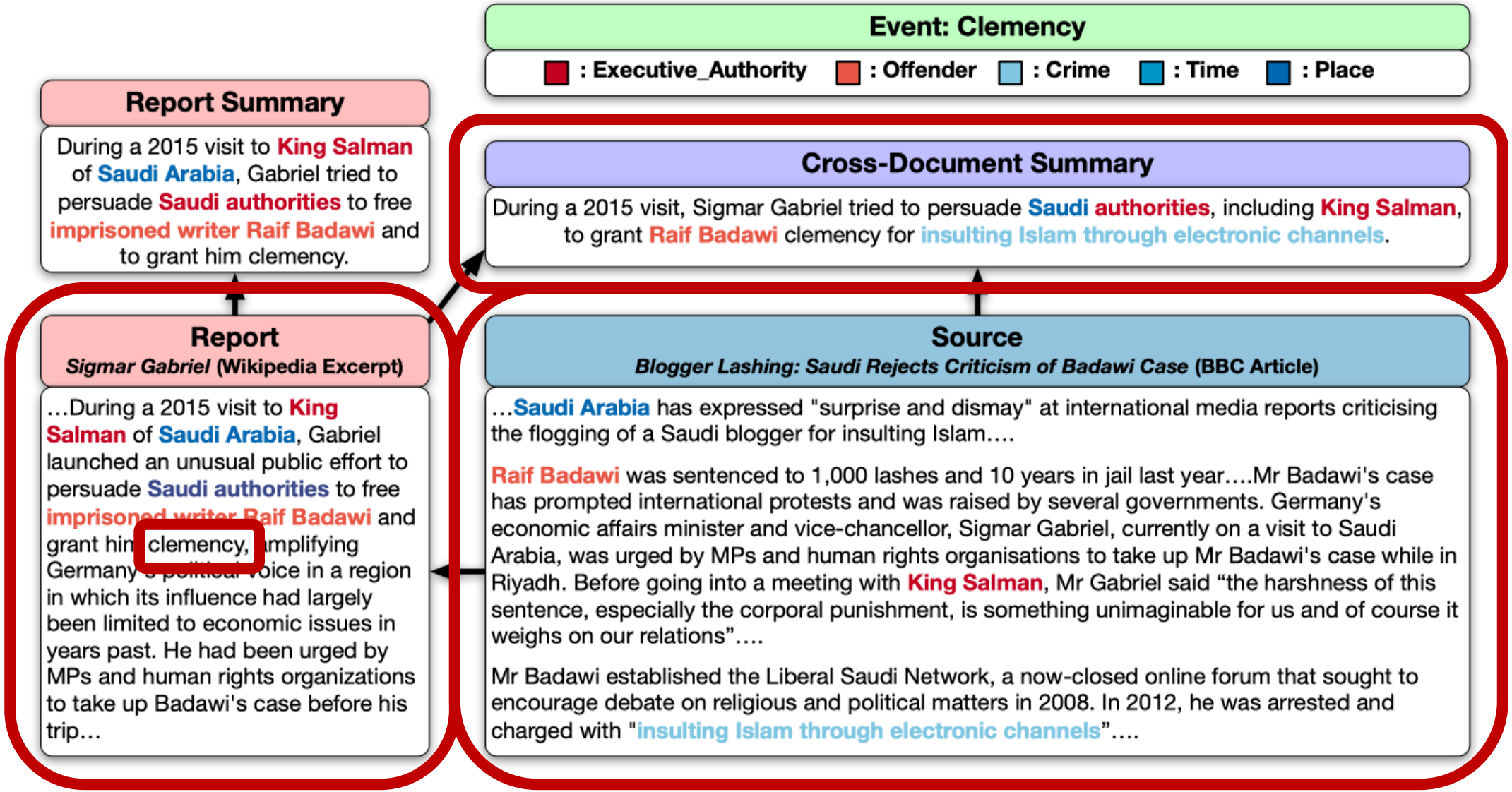
Leverage ontology descriptions to generate natural language inference templates filled by semantic parses.

Part 2: Template Filling

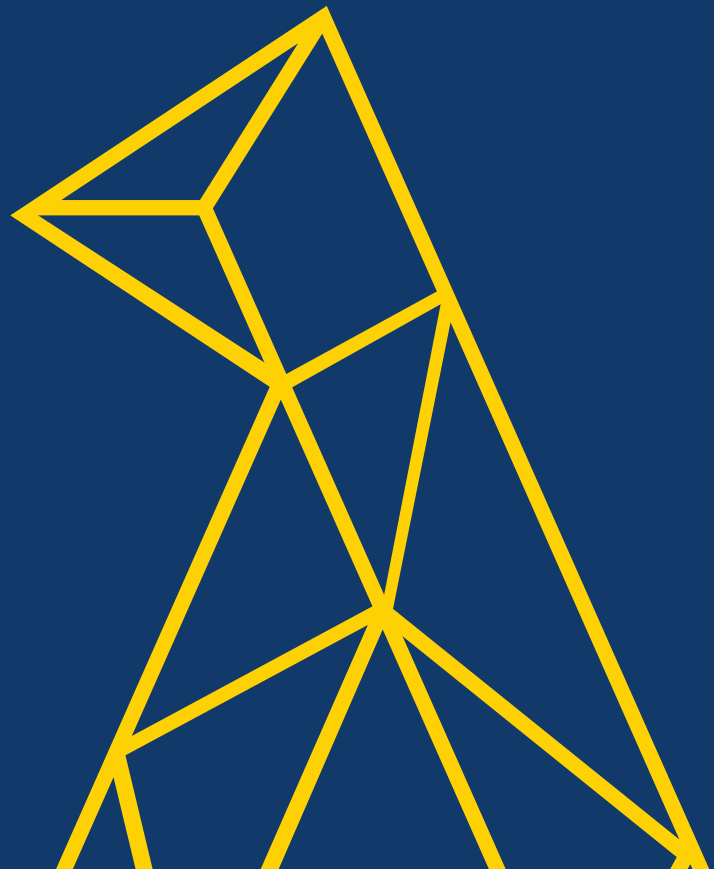
Leverage ontology structure to fill natural language templates across documents describing same event.

Part 3: Event-Keyed Summarization

Use semantic parse to propose events of interest and synthesize info about those events across documents.


Event: Clemency
: Executive_Authority
: Offender
: Crime
: Time
: Place
Report Summary
During a 2015 visit to King Salman of Saudi Arabia, Gabriel tried to persuade Saudi authorities to free imprisoned writer Raif Badawi and to grant him clemency.
Cross-Document Summary
During a 2015 visit, Sigmar Gabriel tried to persuade Saudi authorities, including King Salman, to grant Raif Badawi clemency for insulting Islam through electronic channels.
Report
Sigmar Gabriel (Wikipedia Excerpt)
…During a 2015 visit to King Salman of Saudi Arabia, Gabriel launched an unusual public effort to persuade Saudi authorities to free imprisoned writer Raif Badawi and grant him clemency, amplifying Germany's political voice in a region in which its influence had largely been limited to economic issues in years past. He had been urged by MPs and human rights organizations to take up Badawi's case before his trip…
Source
Blogger Lashing: Saudi Rejects Criticism of Badawi Case (BBC Article)
…Saudi Arabia has expressed "surprise and dismay" at international media reports criticising the flogging of a Saudi blogger for insulting Islam….
Raif Badawi was sentenced to 1,000 lashes and 10 years in jail last year….Mr Badawi's case has prompted international protests and was raised by several governments. Germany's economic affairs minister and vice-chancellor, Sigmar Gabriel, currently on a visit to Saudi Arabia, was urged by MPs and human rights organisations to take up Mr Badawi's case while in Riyadh. Before going into a meeting with King Salman, Mr Gabriel said “the harshness of this sentence, especially the corporal punishment, is something unimaginable for us and of course it weighs on our relations”….
Mr Badawi established the Liberal Saudi Network, a now-closed online forum that sought to encourage debate on religious and political matters in 2008. In 2012, he was arrested and charged with "insulting Islam through electronic channels"….

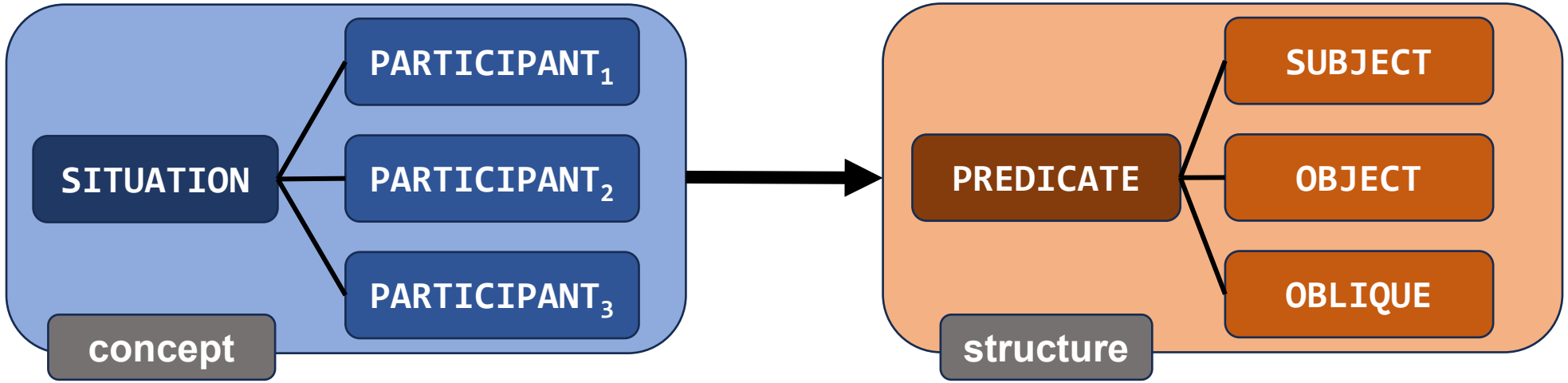
Natural Language Inference



Idea

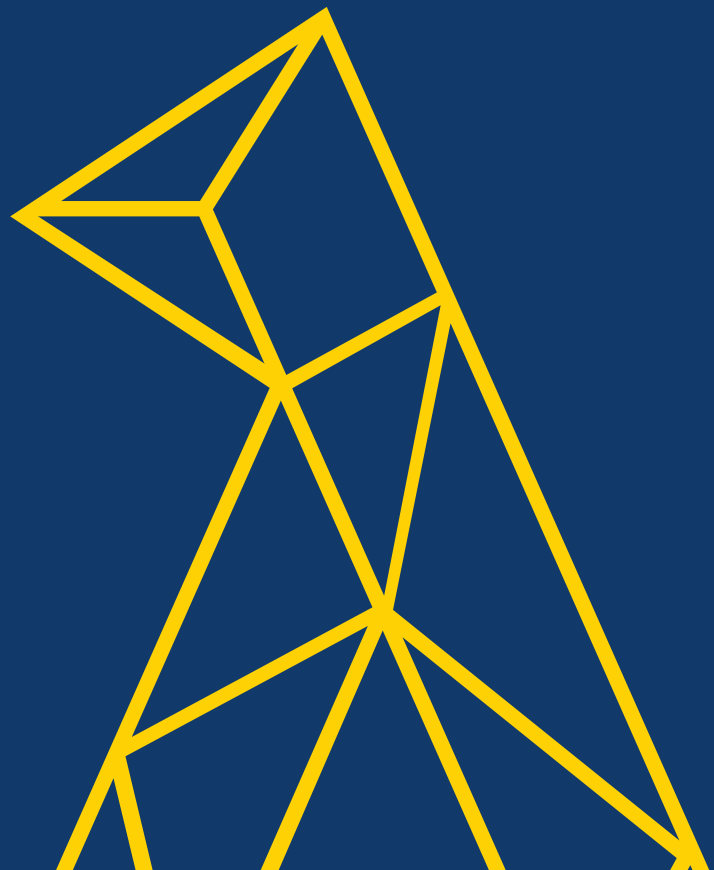
Inferences relevant to mapping from concepts to linguistic structure tend to be inferences humans really care about.

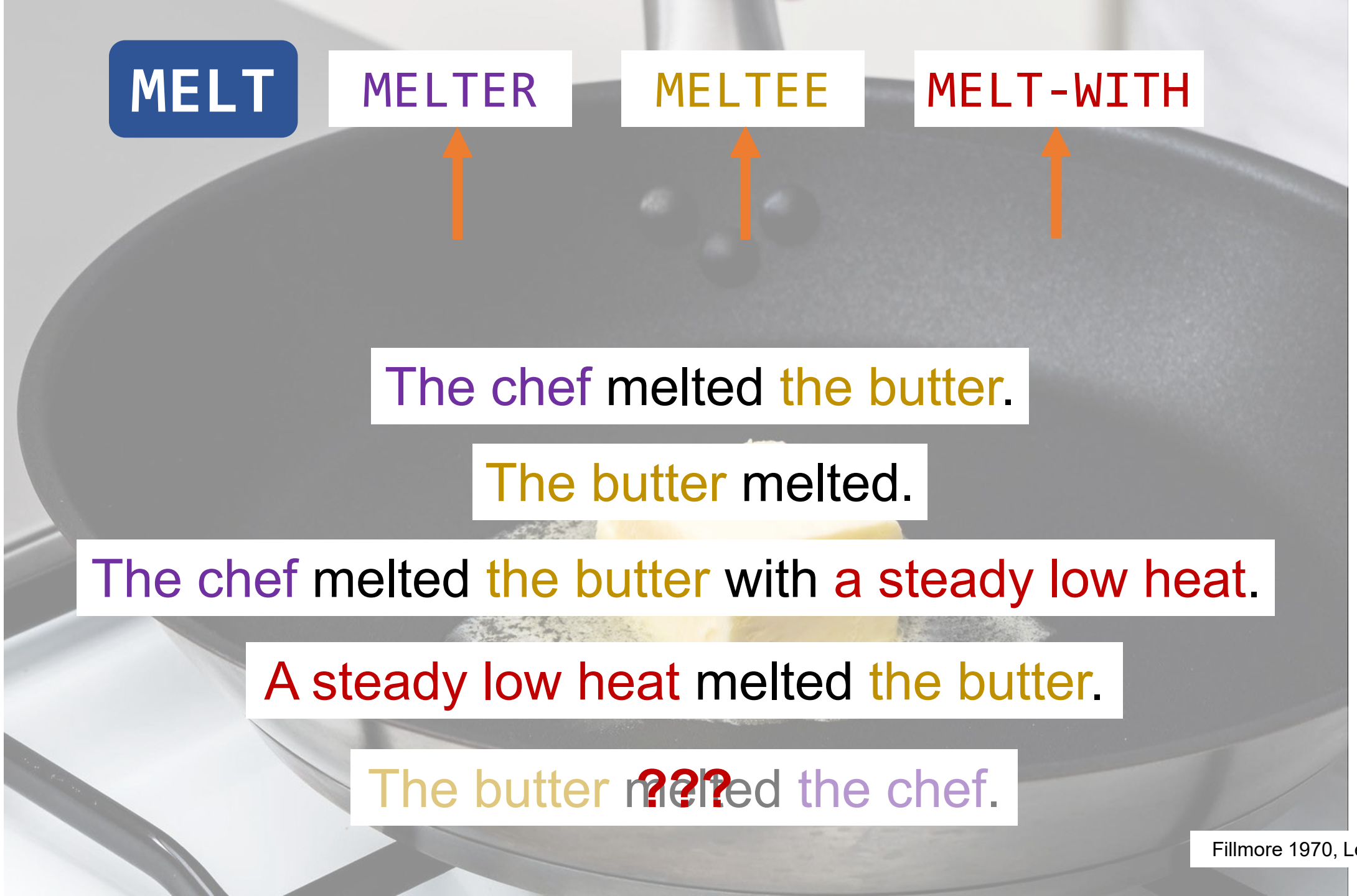
Example #1: Linking Theory



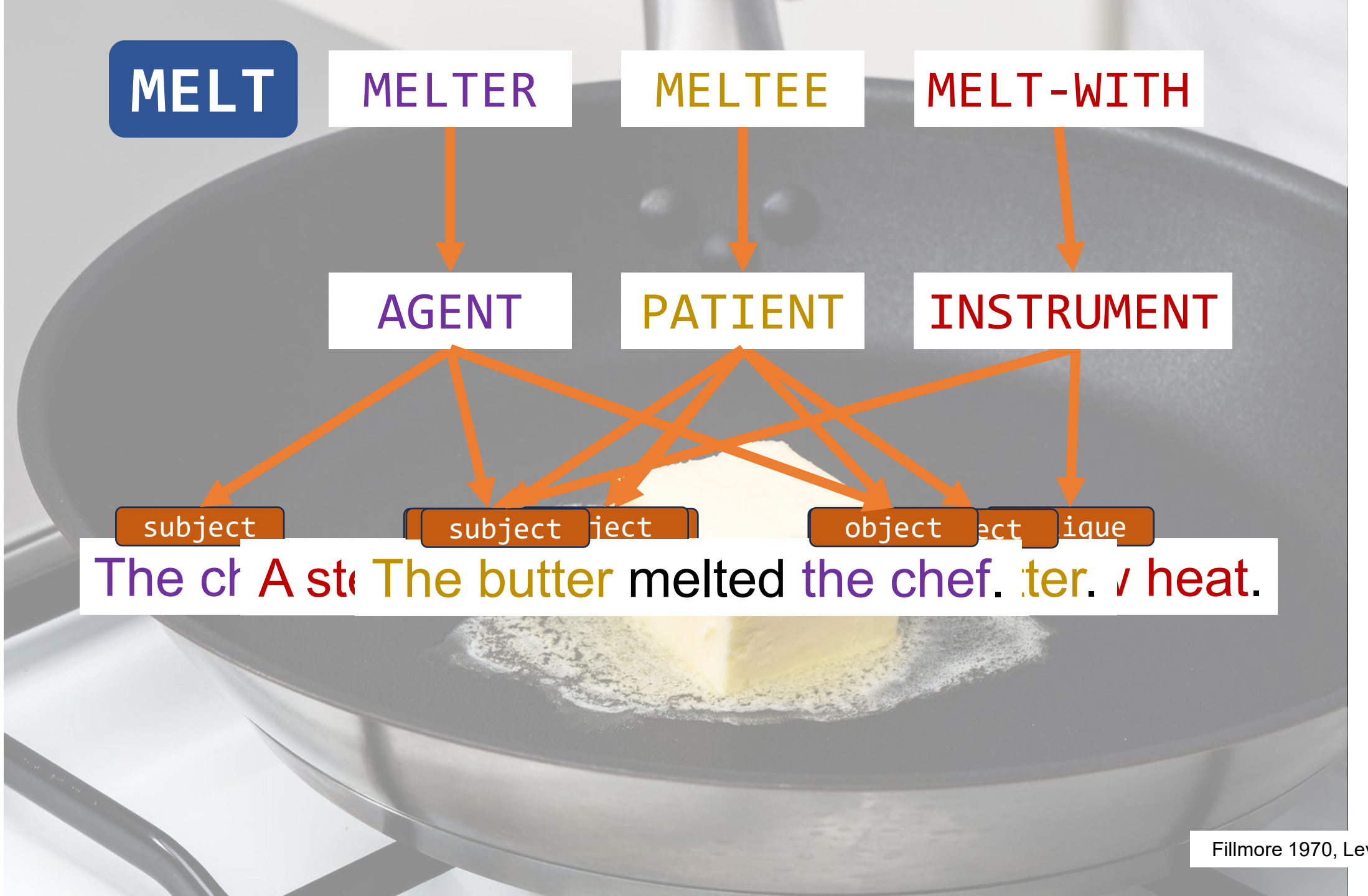
Natural Language Inference

Sentence-Level Descriptions

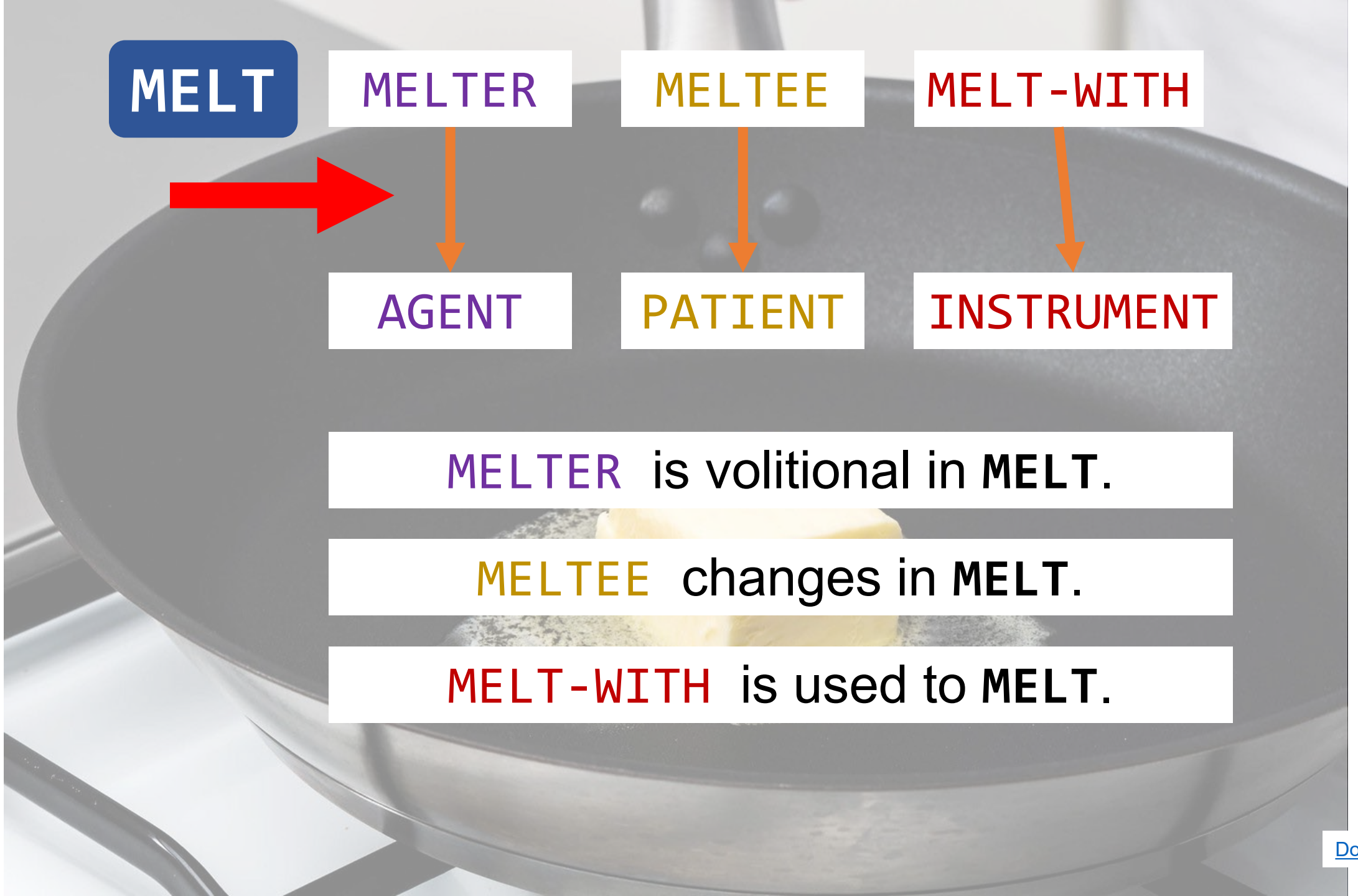




Fillmore 1970, Levin 1993



Fillmore 1970, Levin 1993


MELT
MELTER
MELTEE
MELT-WITH
AGENT
PATIENT
INSTRUMENT
MELTER is volitional in MELT.
MELTEE changes in MELT.
MELT-WITH is used to MELT.

Dowty 1991

Dowty's idea

Having more prototypically agentive properties compared to other roles makes a role more likely to map to `SUBJ`.

Upshot

Inferences relevant to linking theory are probably ones humans tend to care a lot about.

Protoagent

- PARTICIPANT chose to be involved in SITUATION.
- PARTICIPANT caused SITUATION to happen.
- PARTICIPANT was aware of being involved in SITUATION.
- PARTICIPANT changed location during SITUATION.
- PARTICIPANT existed before SITUATION began.

- PARTICIPANT existed during SITUATION.
- PARTICIPANT existed after SITUATION stopped.

Protopatient

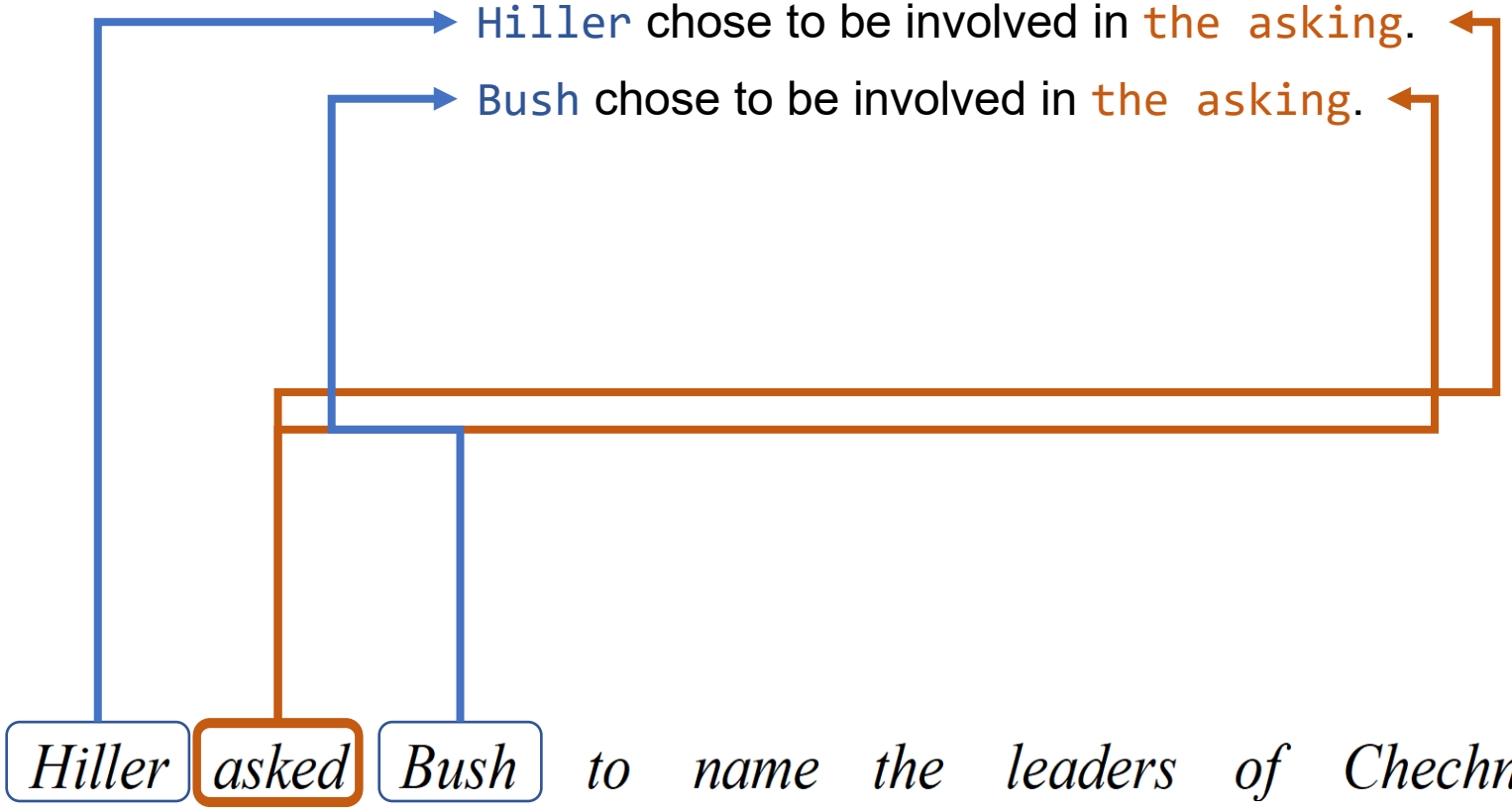
- PARTICIPANT changed possession during SITUATION.
- PARTICIPANT was used in carrying out SITUATION.
- PARTICIPANT was changed or somehow altered during or by the end of SITUATION.
- SITUATION happened for the benefit of PARTICIPANT.
- The change in PARTICIPANT happened throughout the SITUATION.
- Only a part or portion of PARTICIPANT was involved in SITUATION.

Dowty 1991, Reisinger et al. 2015, White et al. 2016

Two steps

1. Find relevant SITUATION-PARTICIPANT pairs.

PARTICIPANT chose to be involved in SITUATION.



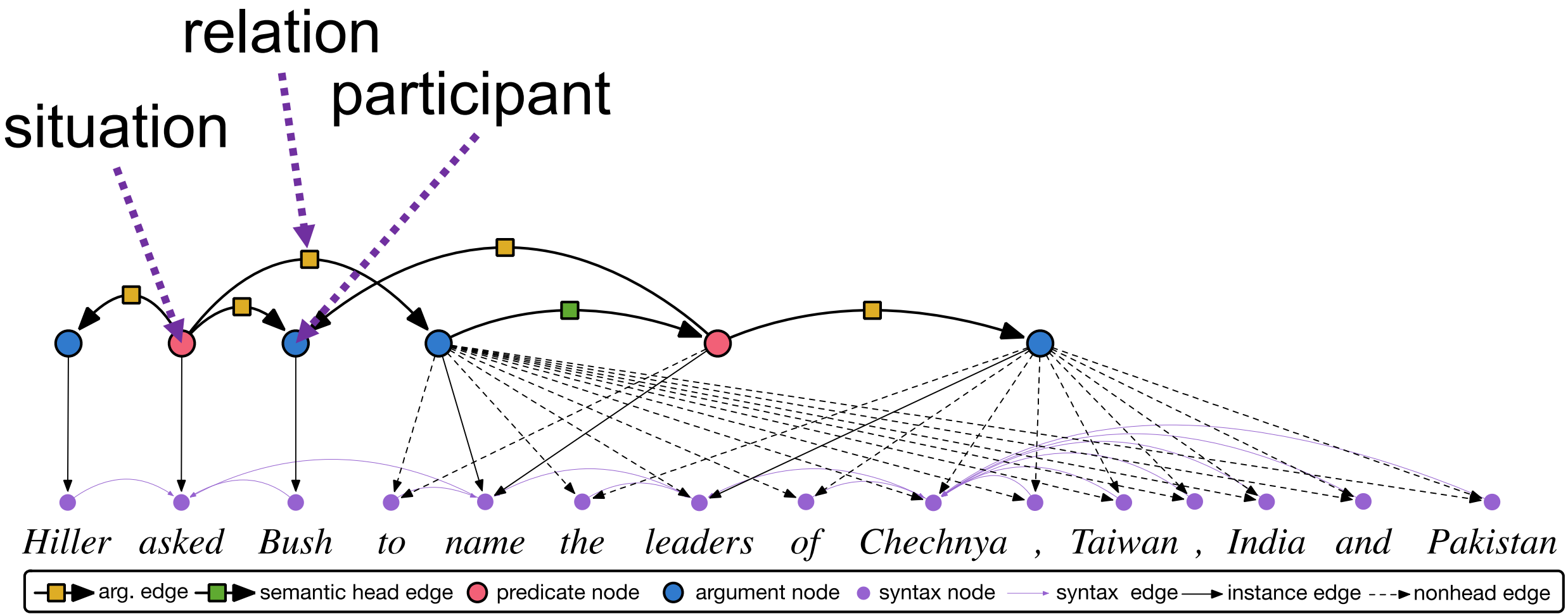
White et al. 2016, 2020, Zhang et al. 2017

Two steps

1. Find relevant `SITUATION`-`PARTICIPANT` pairs.

Requirement

Very light scaffolding ontology including at least `SITUATION`, `PARTICIPANT`, and `RELATION`.



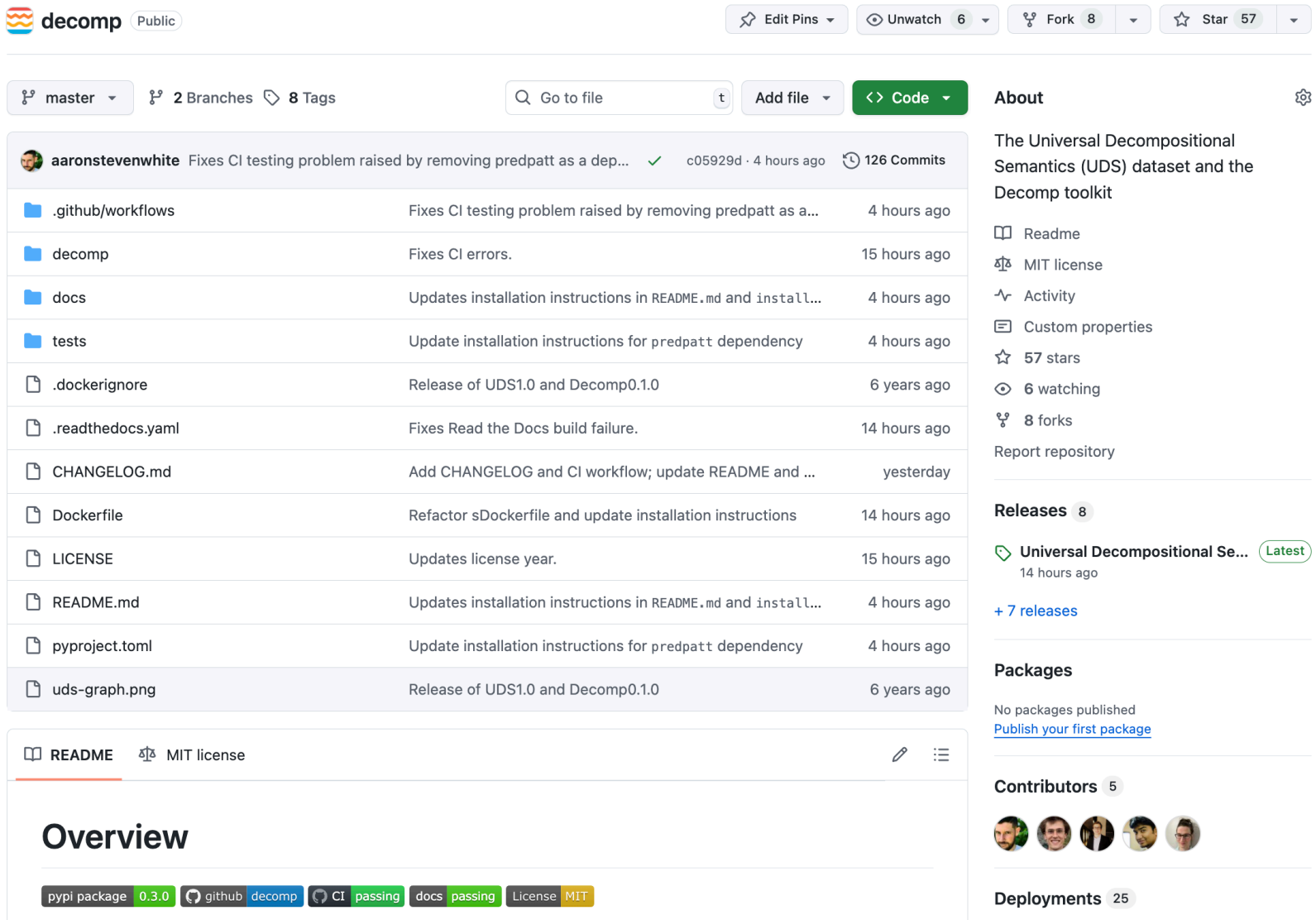
Similar to OpenIE [Banko et al. 2007]

[White et al. 2016], [2020], [Zhang et al. 2017]

Rule-based approach

White et al. 2016, Zhang et al. 2017
based on Rudinger and Van Durme 2014

A set of rules defined on top of Universal Dependencies syntactic annotations.



https://github.com/decompositional-semantics-initiative/decomp/

White et al. 2020

Rule-based approach

White et al. 2016, Zhang et al. 2017
based on Rudinger and Van Durme 2014

A set of rules defined on top of Universal Dependencies syntactic annotations.

Transductive parsing approach

Stengel-Eskin 2020, 2021
based on Zhang et al. 2019

Encoder-decoder networks that take raw sentence and produce semantic graph on the decoder side.

Two steps

1. Find relevant `SITUATION`-`PARTICIPANT` pairs.
2. Figure out which inferences hold of which pairs.

Requirement

Very light scaffolding ontology including at least `SITUATION`, `PARTICIPANT`, and `RELATION`.

PARTICIPANT chose to be involved in SITUATION.

Hiller chose to be involved in the asking.

Bush chose to be involved in the asking.

Hiller asked Bush to name the leaders of Chechnya , Taiwan , India and Pakistan

White et al. 2016, 2020, Zhang et al. 2017

Hiller asked Bush to name the leaders of Chechnya, Taiwan, India, and Pakistan.

How likely is it that Hiller chose to be involved in the asking?

very unlikely somewhat unlikely not enough information somewhat likely very likely

Reisinger et al. 2015, White et al. 2016, 2020, Zhang et al. 2017

Hiller asked Bush to name the leaders of Chechnya, Taiwan, India, and Pakistan.

How likely is it that Bush chose to be involved in the asking?

very unlikely somewhat unlikely not enough information somewhat likely very likely

Data White et al. 2016

~10k predicate-argument pairs from Penn TreeBank + ~6k predicate-argument pairs from Universal Dependencies English Web Treebank.

Challenge

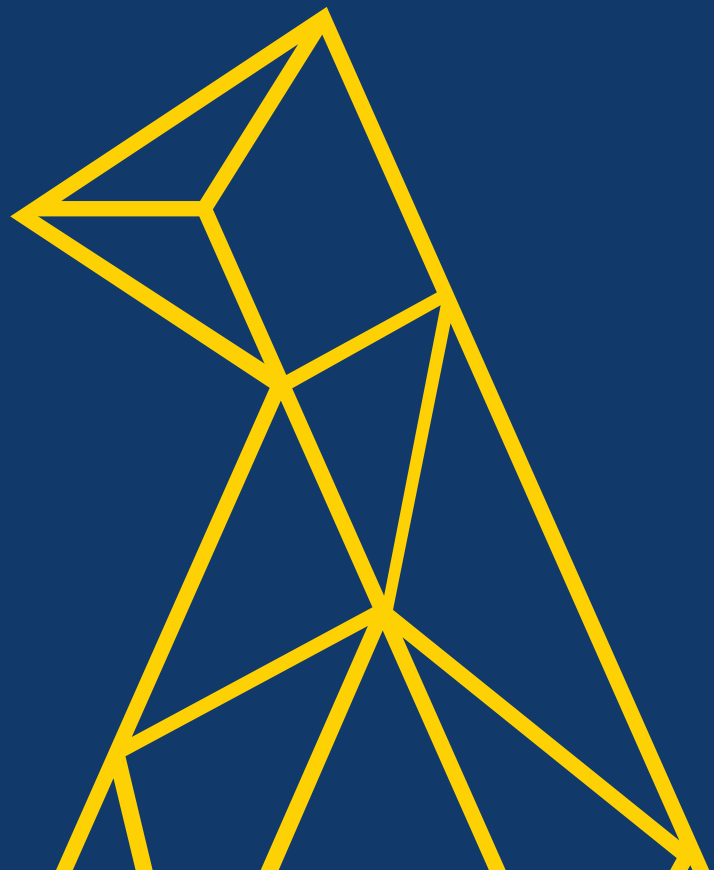
Semantic parses are sentence-bound, meaning it's hard to capture information about complex situations.

Approach Vashishtha et al. 2019, Gantt et al. 2022

Augment light scaffolding with cross-sentential relations and construct inference templates for those relations.

Natural Language Inference

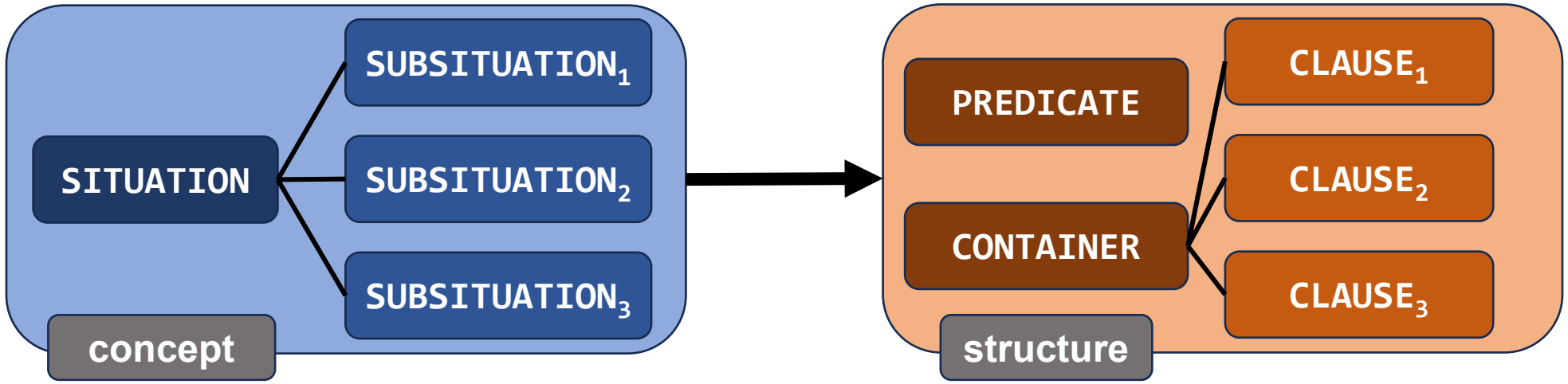
Document-Level Descriptions



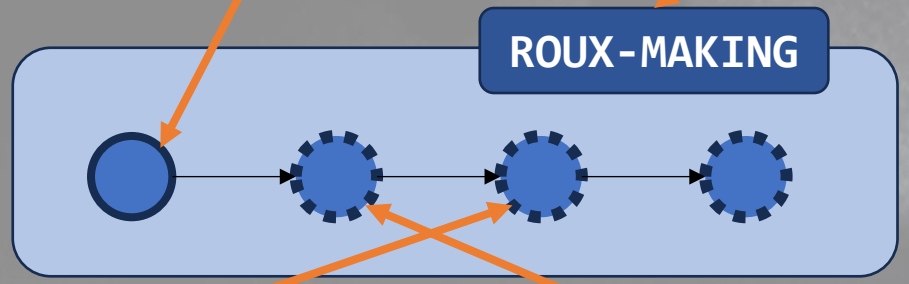
Idea

Inferences relevant to mapping from concepts to linguistic structure tend to be inferences humans really care about.

Example #2: Event Structure



The chef began to make the roux.



Before adding the flour, she melted the butter.

Making the roux took some amount of time.

Adding the flour was part of making the roux.

Melting the butter ended before adding the flour started.

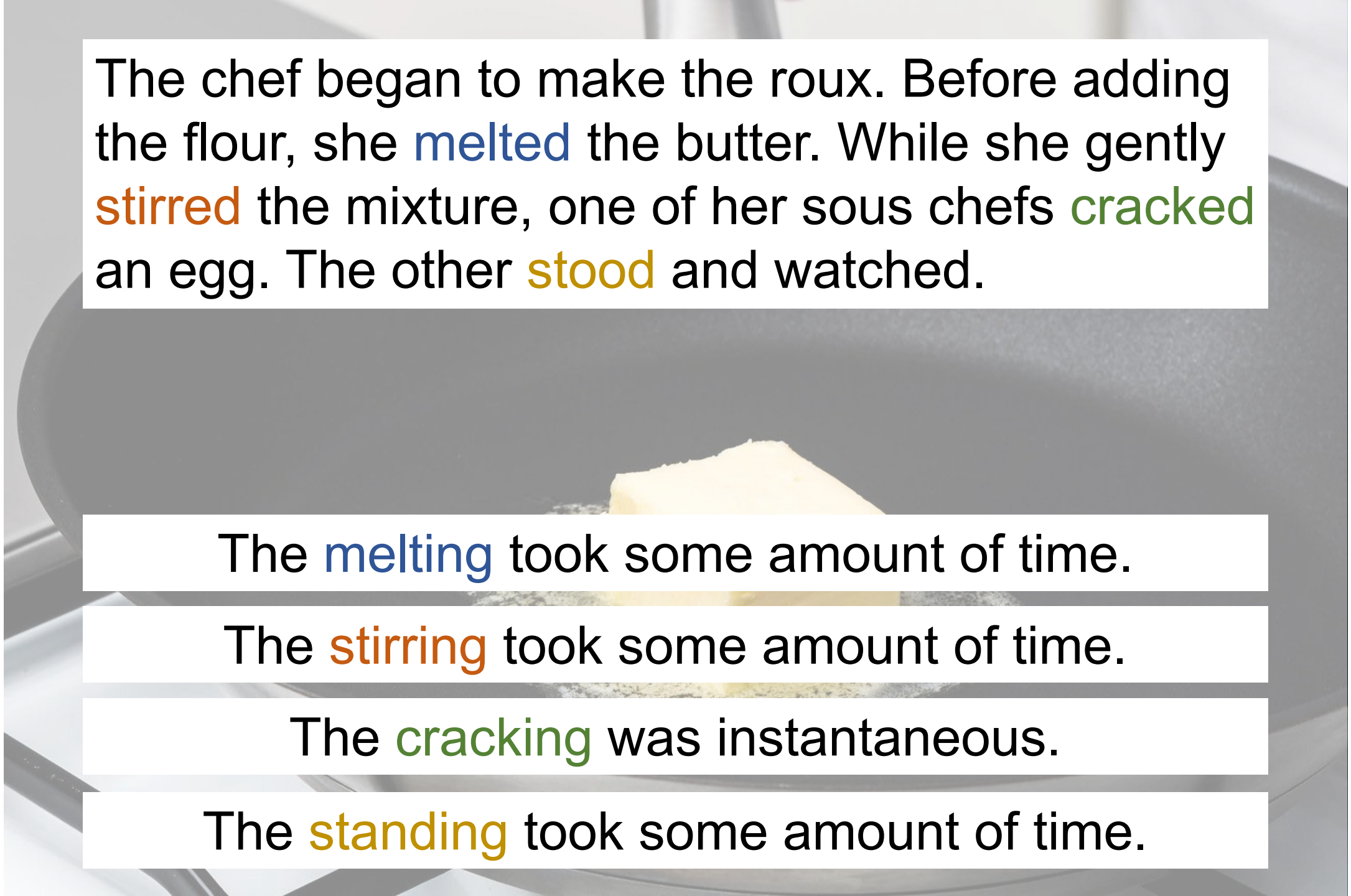
Idea #1 Vendler 1957

Situations fall into classes based on their temporal progression.

Idea #2

These classes can be decomposed in terms of the inferences they trigger about their temporal progression and the relations among their subevents.

Kenny 1963, Lakoff 1965, Verkuyl 1972, Bennett and Partee 1978, Mourelatos 1978, Dowty 1979, Jackendoff 1990, Pustejovsky 1995, i.a.



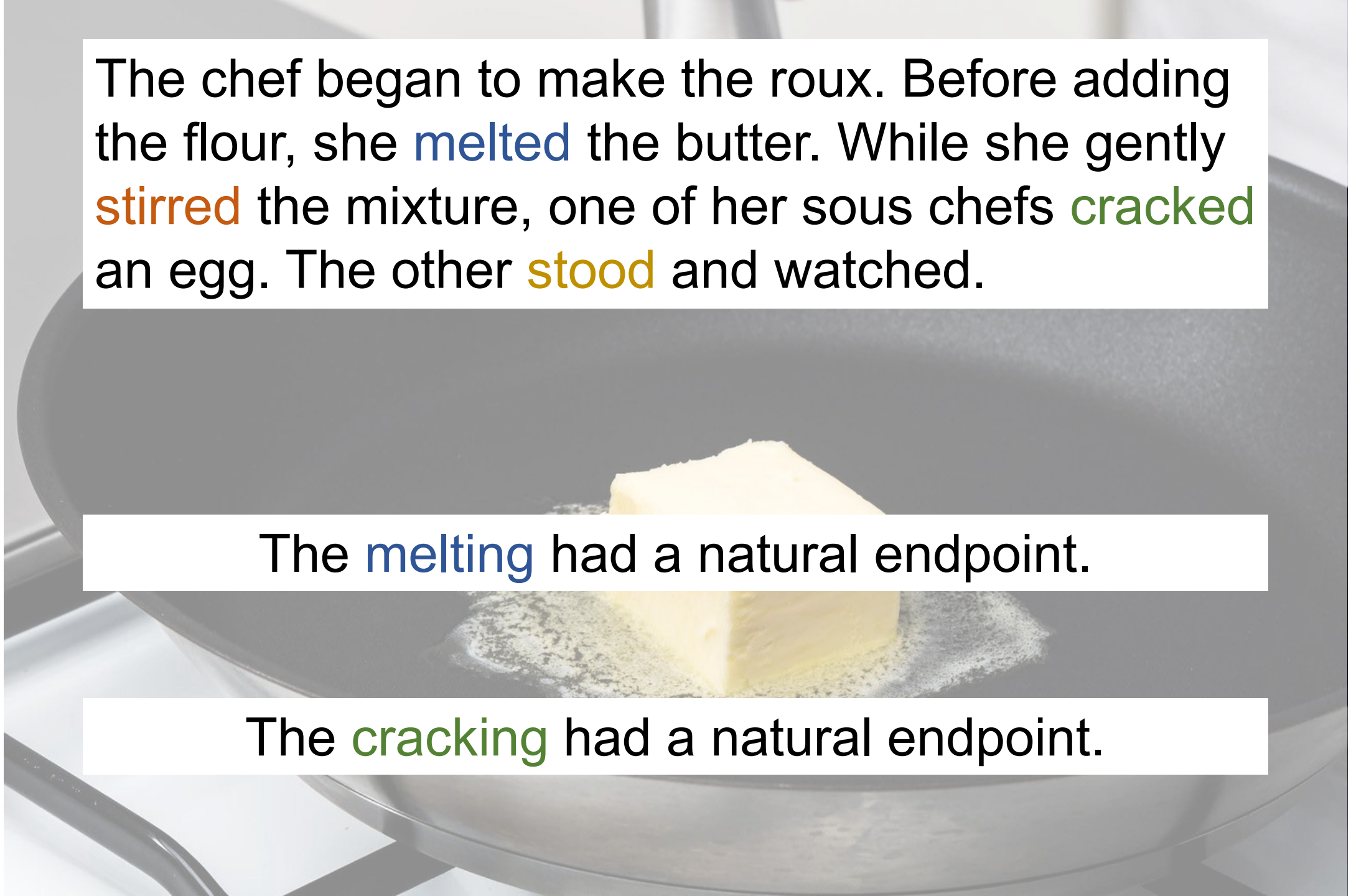
The chef began to make the roux. Before adding the flour, she melted the butter. While she gently stirred the mixture, one of her sous chefs cracked an egg. The other stood and watched.

The melting took some amount of time.

The stirring took some amount of time.

The cracking was instantaneous.

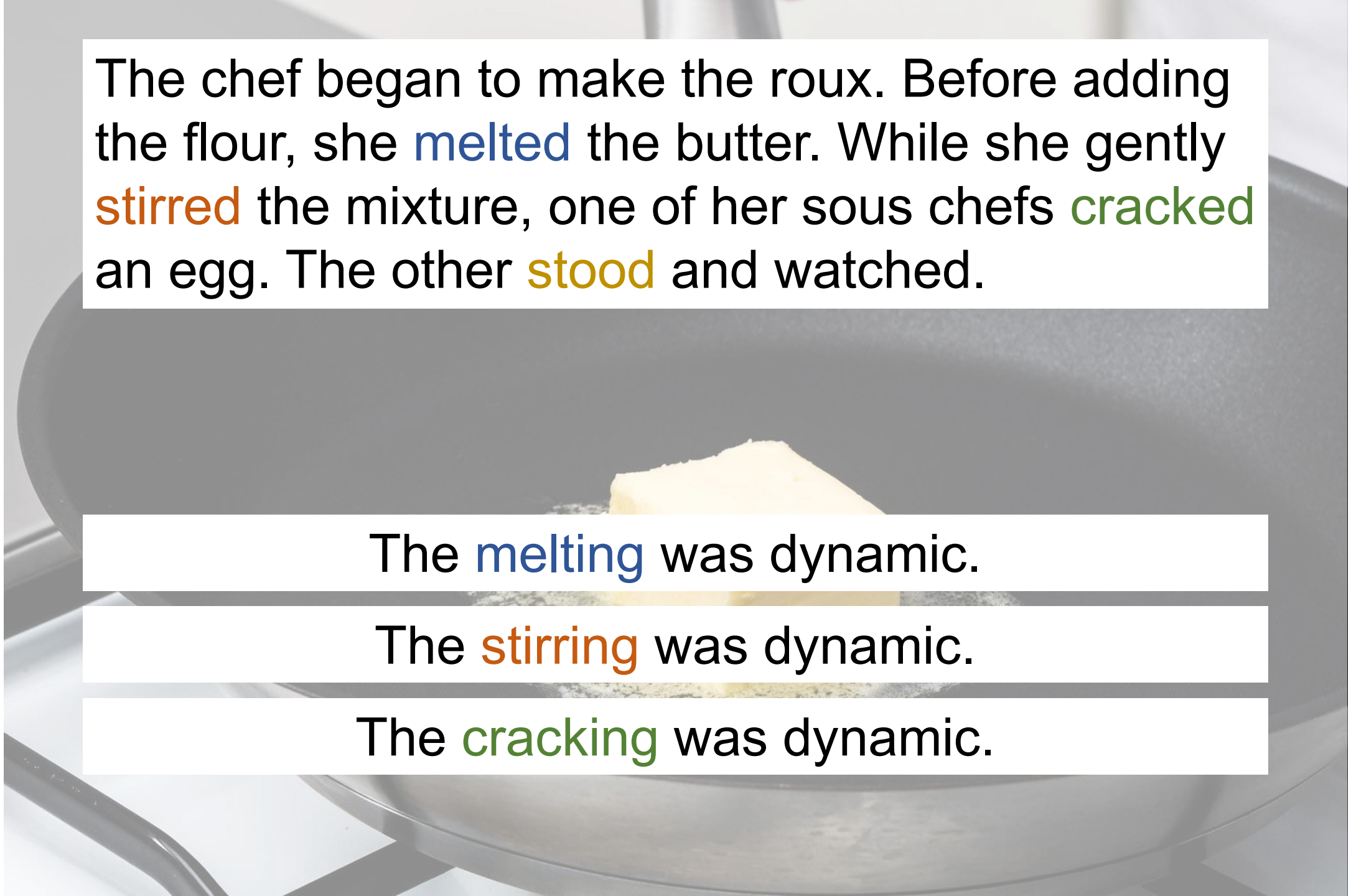
The standing took some amount of time.



The chef began to make the roux. Before adding the flour, she melted the butter. While she gently stirred the mixture, one of her sous chefs cracked an egg. The other stood and watched.

The melting had a natural endpoint.

The cracking had a natural endpoint.



The chef began to make the roux. Before adding the flour, she melted the butter. While she gently stirred the mixture, one of her sous chefs cracked an egg. The other stood and watched.

The melting was dynamic.

The stirring was dynamic.

The cracking was dynamic.

Event internal

- SITUATION lasted TIME-UNIT.
- SITUATION was dynamic.
- SITUATION had natural parts.
  - Parts of SITUATION are similar to each other.
  - Parts of SITUATION lasted TIME-UNIT on average.

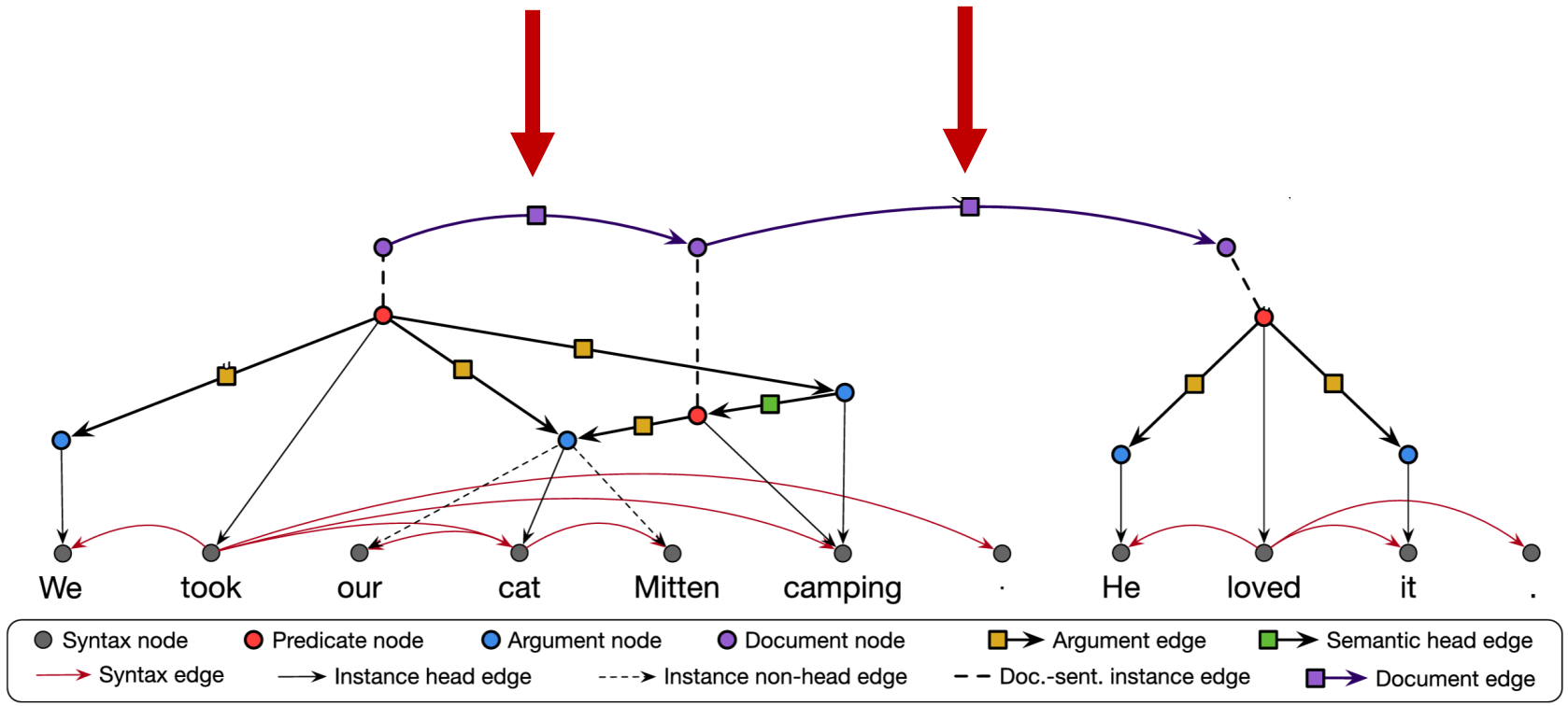
Event mereology

- SITUATION1 was part of SITUATION2.
- SITUATION2 was part of SITUATION1.

Temporal relations

- SITUATION1 started before SITUATION2 started.
- SITUATION1 ended before SITUATION2 started.
- SITUATION1 started before SITUATION2 ended.
- SITUATION1 ended before SITUATION2 ended.

Govindarajan et al. 2019, Vashishtha et al. 2019, 2020, Gantt et al. 2022

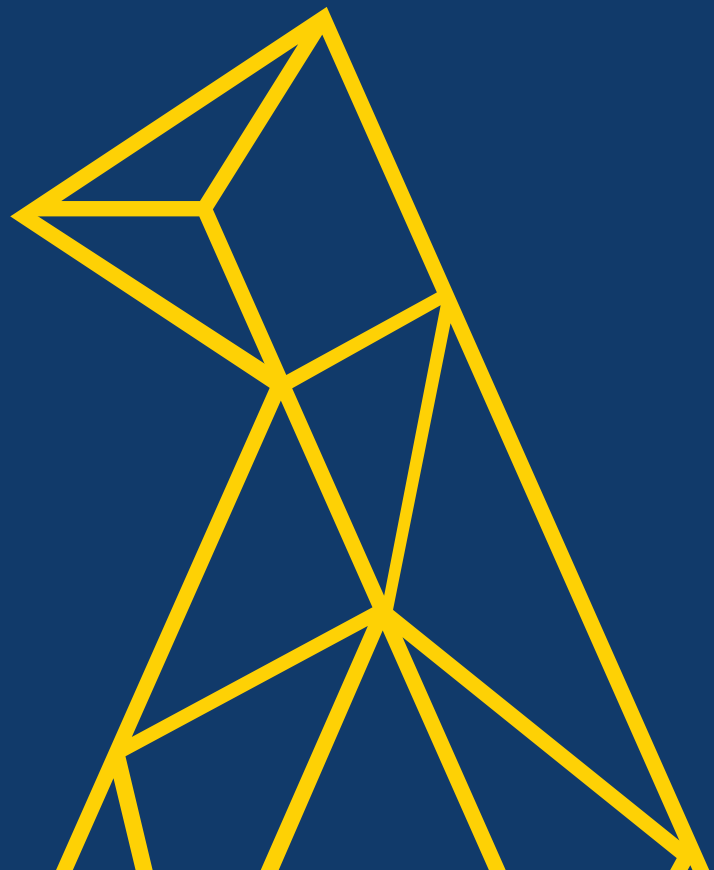

We
took
our
cat
Mitten
camping
.
He
loved
it
.
Syntax node
Predicate node
Argument node
Document node
Argument edge
Semantic head edge
Syntax edge
Instance head edge
Instance non-head edge
Doc.-sent. instance edge
Document edge

Data

~32k predicates from UD-EWT for
~70k predicate-predicate pairs from UD-EWT

Natural Language Inference

Interim Discussion



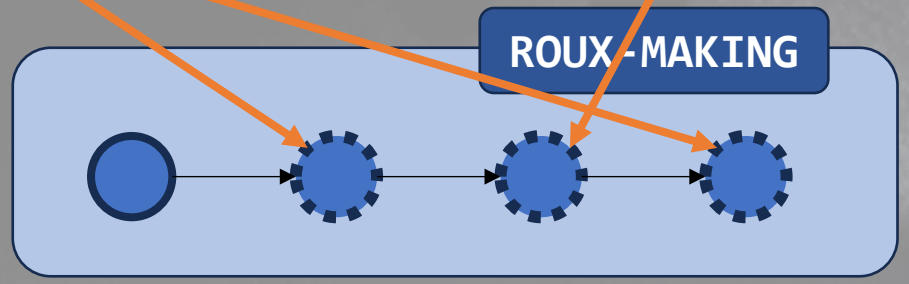
Goal

Find a set of inference templates that are both interesting and broadly applicable across many situations.

Challenge

Not all situation descriptions are anchored by a predicate; yet, we can draw inferences about that situation.

She melted the butter, then added the flour and stirred the mixture gently.



Challenge

Not all situation descriptions are anchored by a predicate; yet, we can draw inferences about that situation.

Approach

Lift the requirement that a situation needs to be anchored by a predicate.

Task

Generalized template extraction

Similar to

Event argument extraction, which extends semantic role labeling from the sentence level to the document level.

She melted the butter. Then, she added the flour. And finally, she stirred the mixture gently.

Template 1: MAKE-ROUX

MAKER: {"She"}
FAT: {"the butter"}
THICKENER: {"the flour"}
ROUX: {"the mixture"}
HEATING-SURFACE: NULL
HEATING-METHOD: NULL
STIRRING-IMPLEMENT: NULL

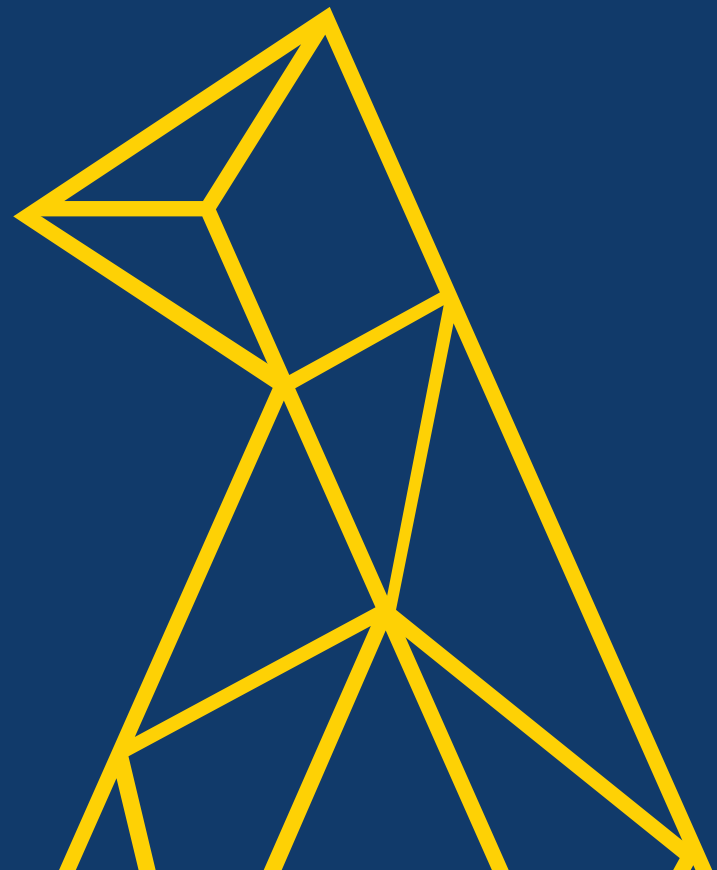
Template 2: MELT

MELTER: {"She"}
MELTEE: {"the butter"}
HEATING-SURFACE: NULL
HEATING-METHOD: NULL

Template 3: MIX

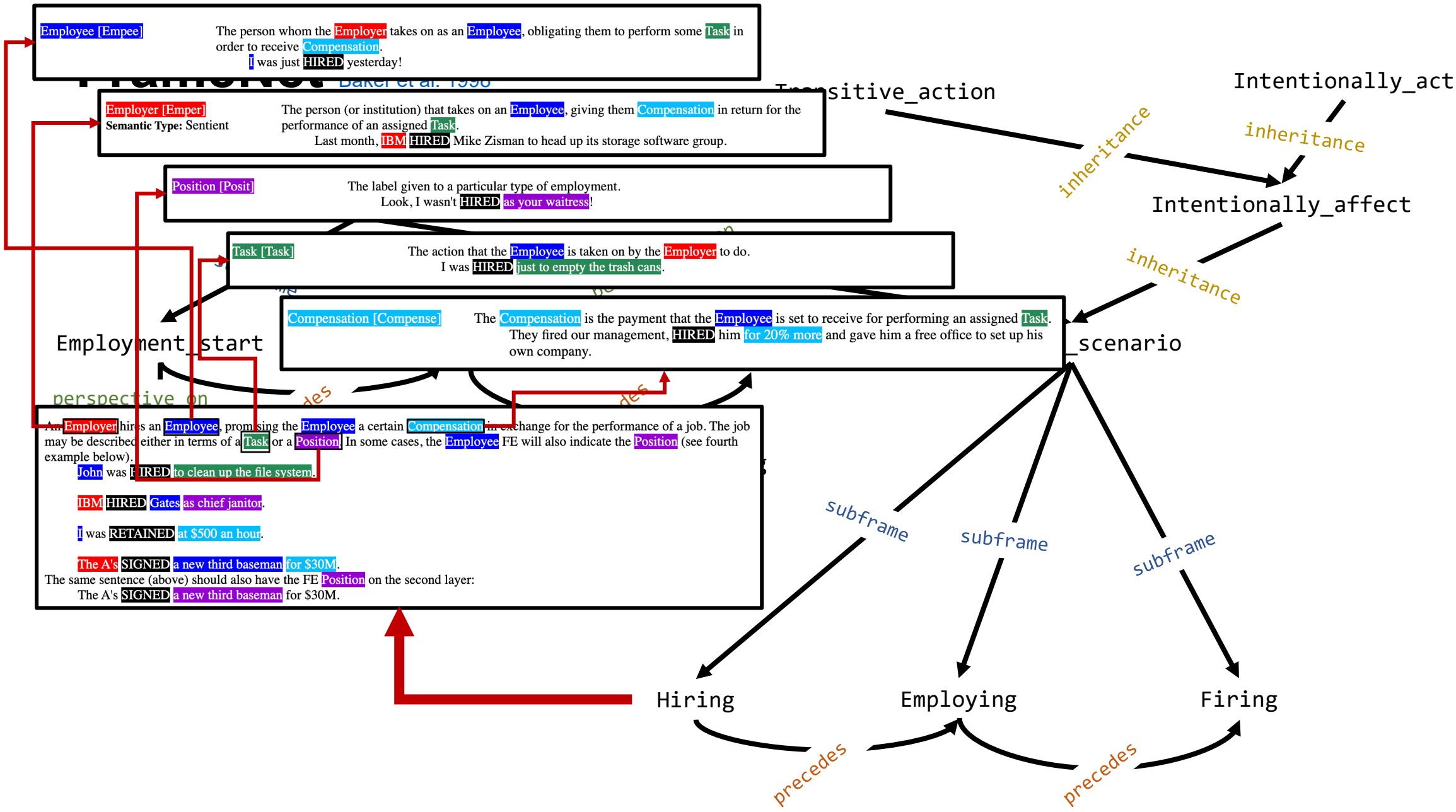
MIXER: {"She"}
MIXING-IMPLEMENT: NULL
INGREDIENTS: {"the butter", "the flour"}
RESULT: {"the mixture"}

Template Filling



Goal

Leverage ontology structure to fill natural language templates across documents describing same event.

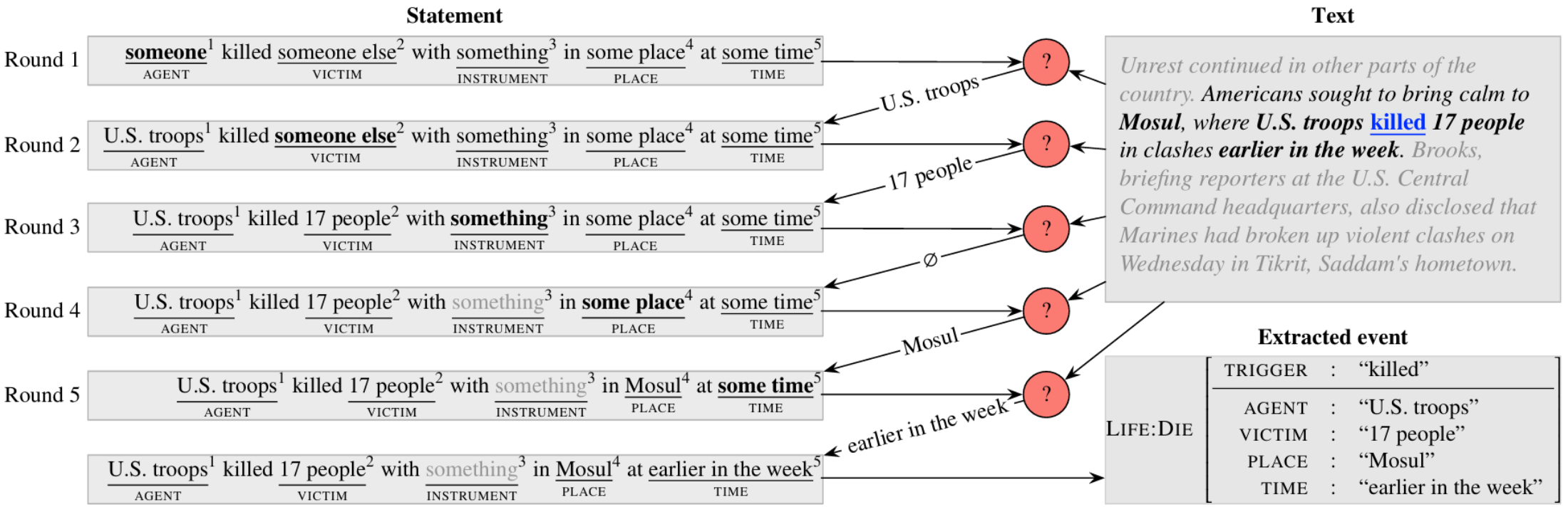

Employee [Empee]
The person whom the Employer takes on as an Employee, obligating them to perform some Task in order to receive Compensation.
I was just HIRED yesterday!
Employer [Emper]
Semantic Type: Sentient
The person (or institution) that takes on an Employee, giving them Compensation in return for the performance of an assigned Task.
Last month, IBM HIRED Mike Zisman to head up its storage software group.
Position [Posit]
The label given to a particular type of employment.
Look, I wasn't HIRED as your waitress!
Task [Task]
The action that the Employee is taken on by the Employer to do.
I was HIRED just to empty the trash cans.
Compensation [Compense]
The Compensation is the payment that the Employee is set to receive for performing an assigned Task.
They fired our management, HIRED him for 20% more and gave him a free office to set up his own company.
An Employer hires an Employee, promising the Employee a certain Compensation in exchange for the performance of a job. The job may be described either in terms of a Task or a Position. In some cases, the Employee FE will also indicate the Position (see fourth example below).
John was HIRED to clean up the file system.
IBM HIRED Gates as chief janitor.
I was RETAINED at $500 an hour.
The A's SIGNED a new third baseman for $30M.
The same sentence (above) should also have the FE Position on the second layer:
The A's SIGNED a new third baseman for $30M.
Transitive_action
Intentionally_act
Intentionally_affect
inheritance
inheritance
inheritance
Employment_start
_scenario
perspective_on
subframe
subframe
subframe
Hiring
Employing
Firing
precedes
precedes

Goal

Leverage ontology structure to fill natural language templates across documents describing same event.

Idea

Datasets annotated with broad-coverage ontologies provide guidance about candidate inferences.



Chen et al. 2020

Challenge #1

FrameNet (Baker et al. 1998), PropBank (Palmer et al. 2005)

Datasets annotated with broad-coverage ontologies* tend to be annotated at the sentence level.

*that anchor to the text and have the sorts of rich glosses we need: MS-AMR (O'Gorman et al., 2018), UMR (Van Gysel et al. 2021)

Challenge #2

RAMS (Ebner et al., 2020), WikiEvents (Li et al., 2021), DocEE (Tong et al., 2022)

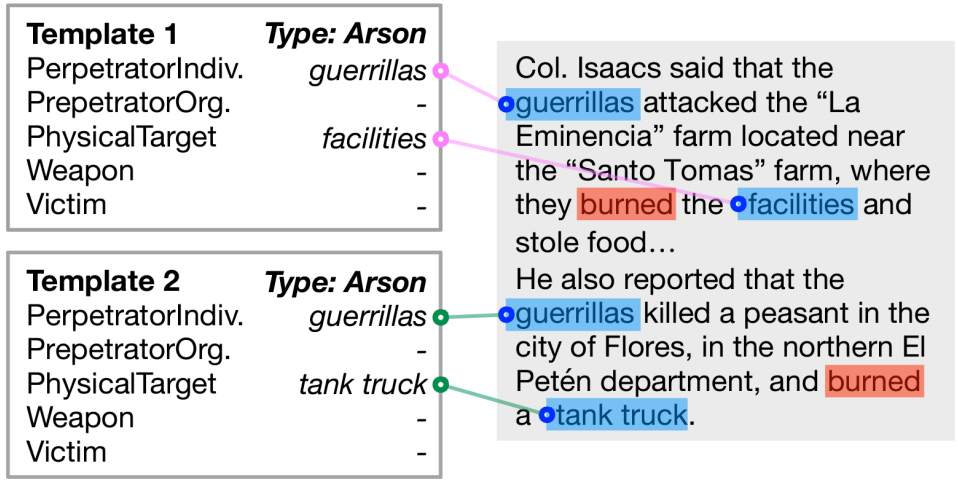
Datasets annotated at the document-level tend to be annotated with narrow ontologies.

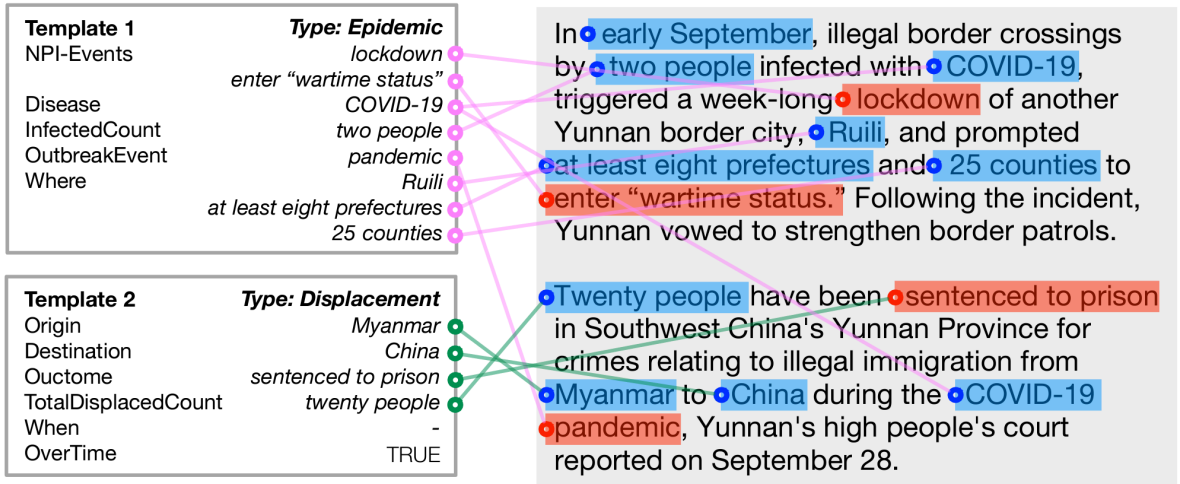
MUC-4 Ontology

- arson
- attack
- bombing
- kidnapping
- murder
- robbery

IARPA BETTER Granular Ontology

- incidents of corruption
- natural disasters
- human migration events
- disease outbreaks or epidemics
- protests or demonstrations
- acts of terrorism





Chen et al. 2023

Challenge #3

Even for datasets annotated with highly constrained ontologies, it is hard to predict templates.

Model (Encoder)	SciREX						MUC-4								
	CEAF-REE$_{def}$			CEAF-RME			CEAF-REE$_{def}$			CEAF-REE$_{impl}$			CEAF-RME		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
TempGen (BART$_{base}$)	8.											46.4	58.3	31.0	40.5
TempGen (BART$_{large}$)	19.											47.2	**61.3**	32.9	42.8
GTT (BERT$_{base}$)	-											50.2	55.0	36.8	44.1
IterX (BERT$_{base}$)	16.2	7.6	10.4	16.2	17.4	16.8	41.3	**27.9**	33.3	52.3	**51.1**	51.7	47.2	**45.0**	46.1
IterX (BART$^{enc}_{base}$)	15.0	**15.0**	15.0	14.3	35.4	20.3	39.2	24.8	30.4	49.8	45.7	47.6	44.8	40.1	42.3
IterX (T5$^{enc}_{large}$)	**26.4**	12.4	**16.9**	**25.0**	**40.6**	**31.0**	53.5	26.2	**35.2**	60.9	46.9	**53.0**	55.8	42.4	**48.2**

This is a hard task!

Chen et al. 2023a

metametric Public

Watch 2 | Fork 0 | Star 3

main | 3 Branches | 8 Tags | Go to file t | Add file | Code

ctongfei better typing (#37) ✓ a368b66 · 8 months ago | 87 Commits

.github/workflows	Update release.yml	9 months ago
docs	some doc touch-ups (#21)	2 years ago
src/metametric	better typing (#37)	8 months ago
tests	Changed the build backend and package management to ...	9 months ago
.gitignore	Added release action (#32)	9 months ago
.python-version	Changed the build backend and package management to ...	9 months ago
README.md	Updated documentation and authorships. (#15)	2 years ago
mkdocs.yml	Update mkdocs.yml (#22)	2 years ago
pyproject.toml	Bumped the version and added loading version from pypr...	8 months ago
uv.lock	Bumped the version and added loading version from pypr...	8 months ago

README

metametric

The `metametric` Python package offers a set of tools for quickly and easily defining and implementing evaluation metrics for a variety of structured prediction tasks in natural language processing (NLP) based on the framework presented in the following paper:

> A Unified View of Evaluation Metrics for Structured Prediction. Yunmo Chen, William Gantt, Tongfei Chen, Aaron Steven White, and Benjamin Van Durme. *EMNLP 2023*.

About

This is the codebase of the paper "A Unified View of Evaluation Metrics for Structured Prediction".

metametric.omnuy.me/

Readme
Activity
3 stars
2 watching
0 forks
Report repository

Releases 8

0.2.1 Latest
on Nov 4, 2024

+ 7 releases

Contributors 3

wanmok Yunmo Chen
ctongfei Tongfei Chen
wgantt Will Walden

Deployments 35

github-pages 8 months ago
release 8 months ago

https://github.com/wanmok/metametric

Chen et al. 2023b

Challenge #3

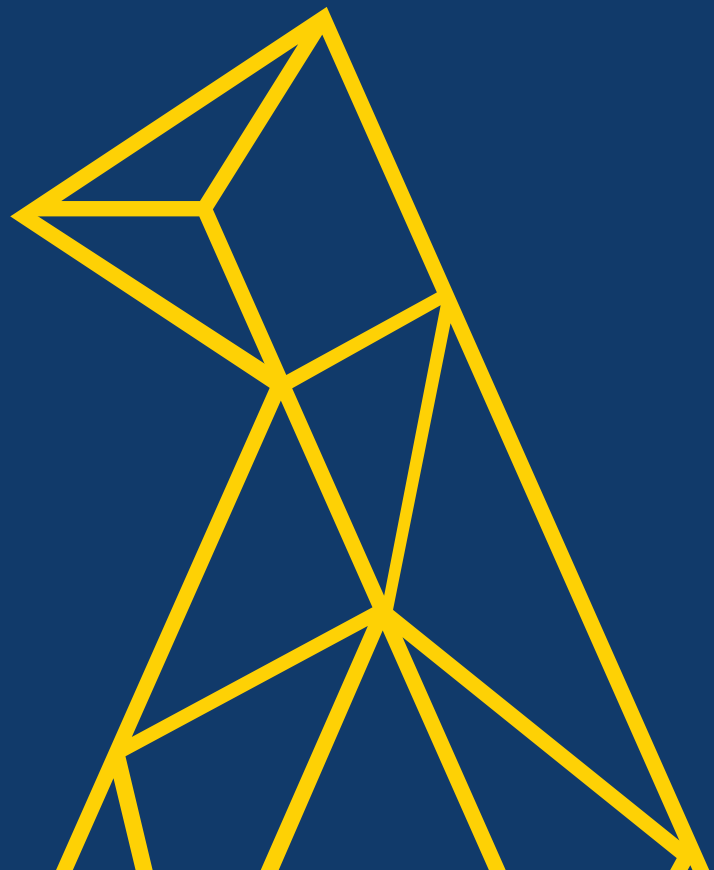
Even for datasets annotated with highly constrained ontologies, it is hard to predict templates.

Upshot

Need template extraction systems on broad-coverage ontologies to generate interesting candidate inferences.

Template Filling

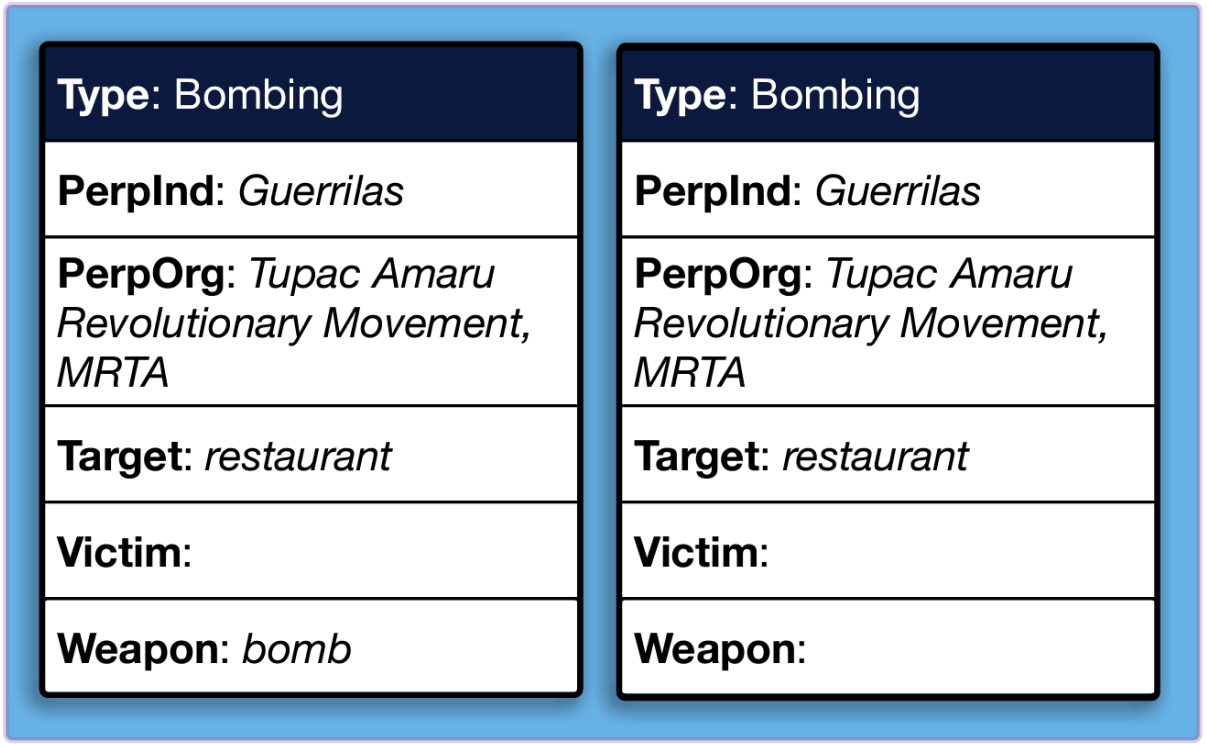
Event individuation



Challenge

Difficult for annotators to agree on how many instances of a particular complex situation type are described by text.

A bomb exploded today in a Lima restaurant, and a second device that had been placed in the same establishment was deactivated by the Peruvian National Police. There were no victims, and the explosion caused very little damage to the restaurant…Guerrillas of the Tupac Amaru Revolutionary Movement (MRTA) have claimed credit for the terrorist act….



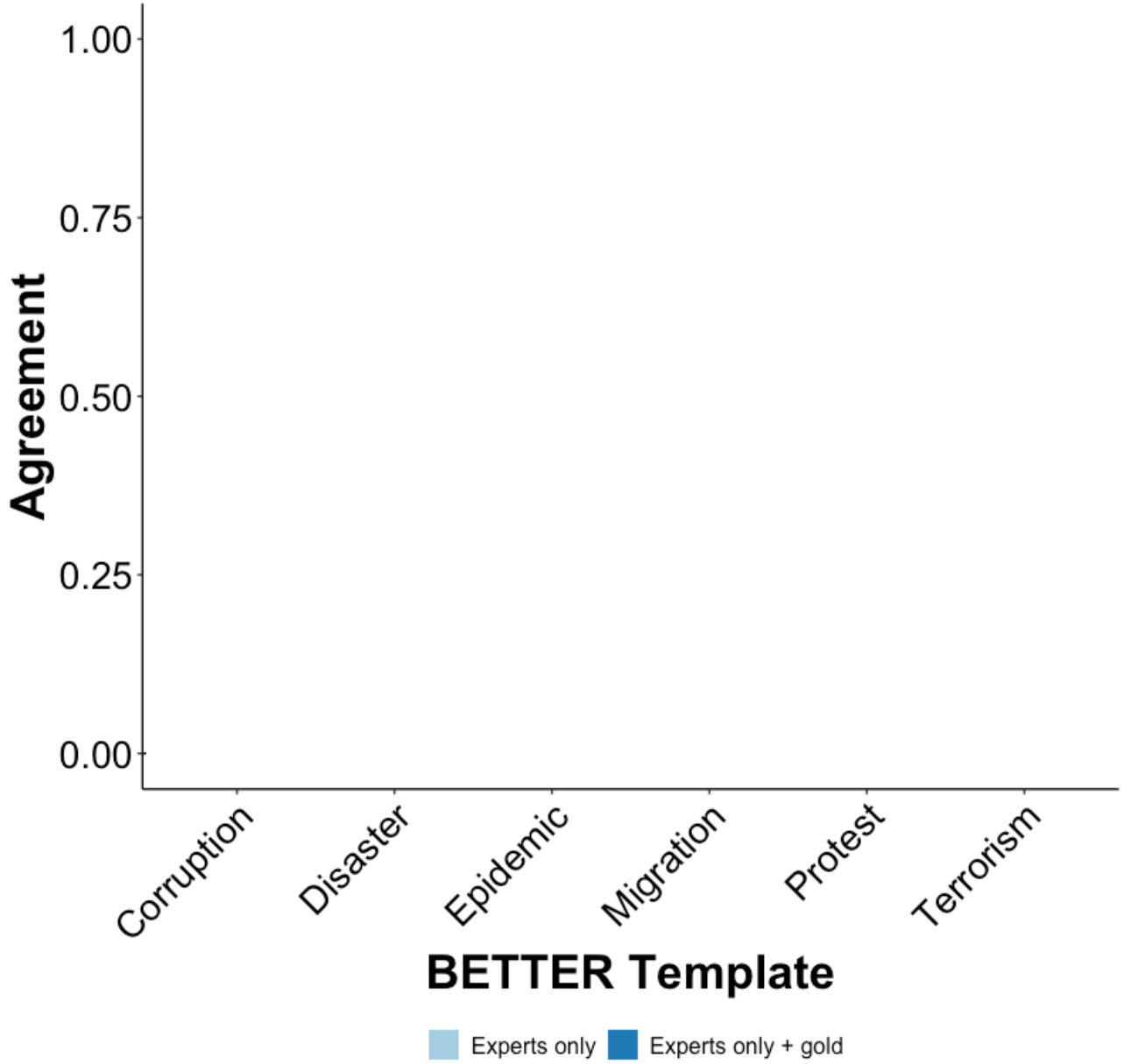
Gantt et al. 2023

Approach

Have three experts reannotate portions of the MUC-4 and BETTER data for number of instances of a template.

Evaluation

Compare agreement among experts annotations and with gold annotation.



Gantt et al. 2023

Challenge

Difficult for annotators to agree on how many instances of a particular complex situation type are described by text.

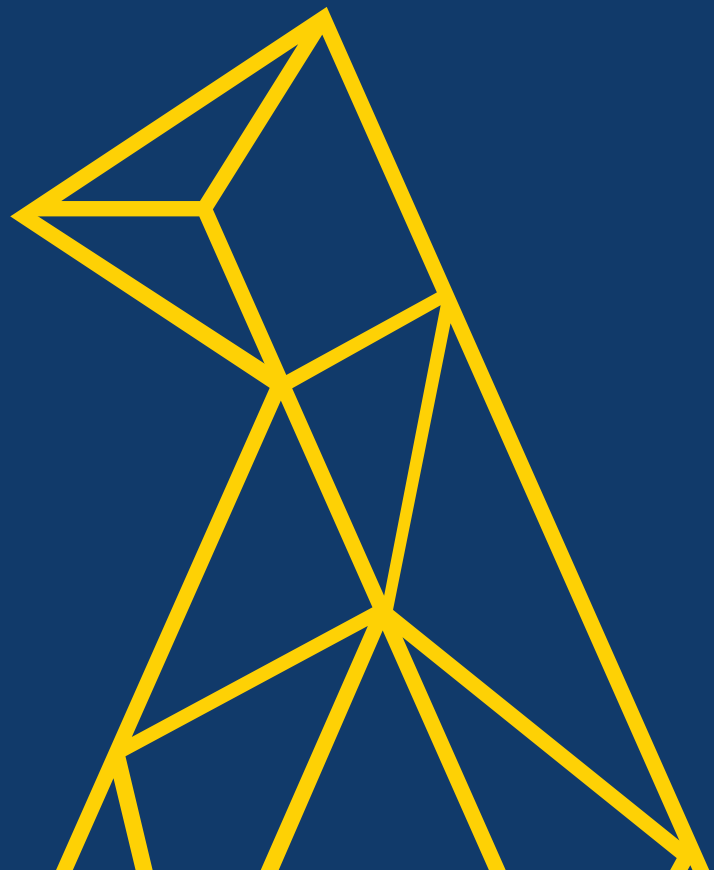
Approach: Cross-doc argument extraction*

Explicitly point to a complex situation description in one document and fill a template for it in another.

*Related to but distinct from similar tasks like Event Linking (Nothman 2012), Cross-Document Event Coreference (Bagga and Baldwin 1999, Cybulska and Vossen 2014, (Eirewetal. 2021,2022), and Predicate-Argument Alignment (Roth and Franke 2012, Wolfe et al. 2013, 2015).

Template Filling

Cross-document argument extraction



Report

On 16 July 2007 the inquiry was **adjourned** until 4 September with a final deadline for the submission of evidence of 14 August 2007 . On 11 September 2007 the inquiry was again adjourned until 19 September to allow the Highways Agency to review traffic evidence , with further adjournments until 18 December .

Source

THE public in
Tintwistle by
Highways Agen
wrong . All t
the inquiry ,
will now have

The agency told the inquiry , at Stalybridge Civic Hall , it had made a ` significant error ' in its estimation of how much traffic would use the road by underestimating journey times on the A616 trunk road ...

Inspector John Watson , chairing the inquiry , adjourned the hearing ...and said the new evidence should be heard and parties who have already given evidence be allowed to return .

Is the **highlighted event** in **Report** mentioned in the **Source**?

Source Validation

YES

Cross Document Argument
ion

Data available at:

https://github.com/FACTSlab/FAMuS

		Source
Activity	"the inquiry"	ublic inquiry into the controversial Mottram - Tintwistle bypass"
Agent	-	"John Watson"
Place	-	"Stalybridge Civic Hall"
Time	"16 July 2007"	"when the Highways Agency admitted it had got its figures wrong"

Vashishtha et al. 2023

Base Corpus [Barham et al. 2023]

MegaWika: >71 million source-report pairs over 50 diverse languages

Available at

https://huggingface.co/datasets/hltcoe/megawika

Longdendale Bypass

Add languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

Coordinates: 53°27′56″N 1°59′59″W

The **Longdendale Bypass** (also known as the **A57/A628 Mottram-in-Longdendale, Hollingworth & Tintwistle Bypass**) is a long-planned road scheme in England by the Highways Agency. The aim is to alleviate traffic congestion on the A57 road/A628 road/A616 road routes that presently pass through the villages. There is both support and opposition for this long-planned scheme which will pass through the valley of Longdendale and part of the Peak District National Park.

Public inquiry [edit]

Further information: Public inquiry

Details of the public inquiry were announced in April 2007.[24] It was to be run by Persona Associates with John Watson as Inspector.[24] An official public inquiry website[25] was launched offering news, transcripts and documentation for download.

The inquiry opened on 26 June 2007.[24] On 16 July 2007 the inquiry was adjourned until 4 September with a final deadline for the submission of evidence of 14 August 2007.[24] On 11 September 2007 the inquiry was again adjourned until 19 September to allow the Highways Agency to review traffic evidence,[26] with further adjournments until 18 December.[24][27]

On 4 December 2007 the Highways Agency published *Route Restraint Measures – Explanation of the Further Change in the Traffic Forecasts* and suggested that this data would not be available until Easter 2008.[28] The inquiry reviewed this new information on 18 December 2007 at which time it was adjourned to allow the Highways Agency to correct their traffic modelling. The inspector commented that "it was the fifth iteration of the traffic model since the original announcement in February 2006". The inquiry was then "Adjourned Indefinitely Pending the Publication of Revised Evidence by The Highways Agency and Tameside MBC".[24]

Manchester Evening News

NEWS IN YOUR AREA MAN UTD MAN CITY WHAT'S ON SPORT MORE

BUY A PAPER JOBS VOUCHER CODES ADVERTISE WITH US DIRECTORY FUNERAL NOTICES MARKETPLACE DATING BOOK AN AD PUBLIC NOTICES NEWSLETTER SIGNUP

News › Greater Manchester News

Inquiry into bypass halted

THE public inquiry into the controversial Mottram-Tintwistle bypass was dramatically halted when the Highways Agency admitted it had got its figures wrong. All the traffic evidence it has put before the inquiry, which has been running since June, will now have to be reviewed.

Barham et al. 2023

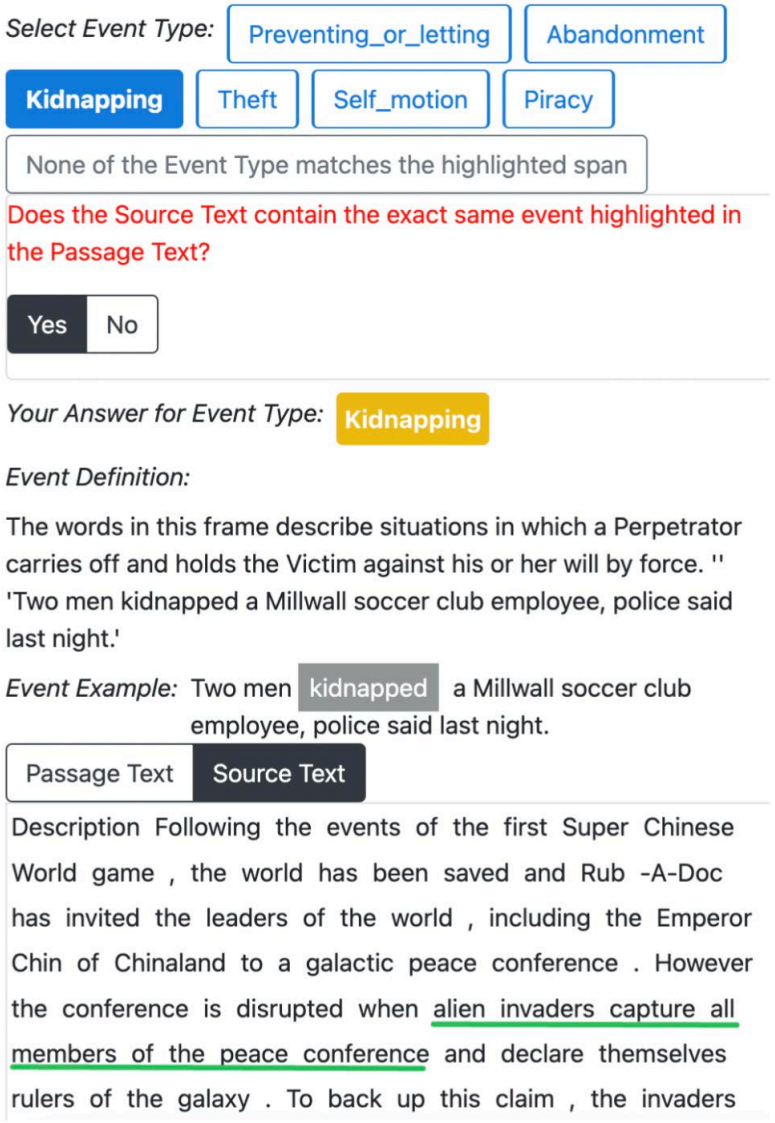
Corpus [Barham et al. 2023]

MegaWika: >71 million source-report pairs over 50 diverse languages

Ontology

FrameNet as the underlying event ontology for broad-coverage of situation types (events, states, processes).

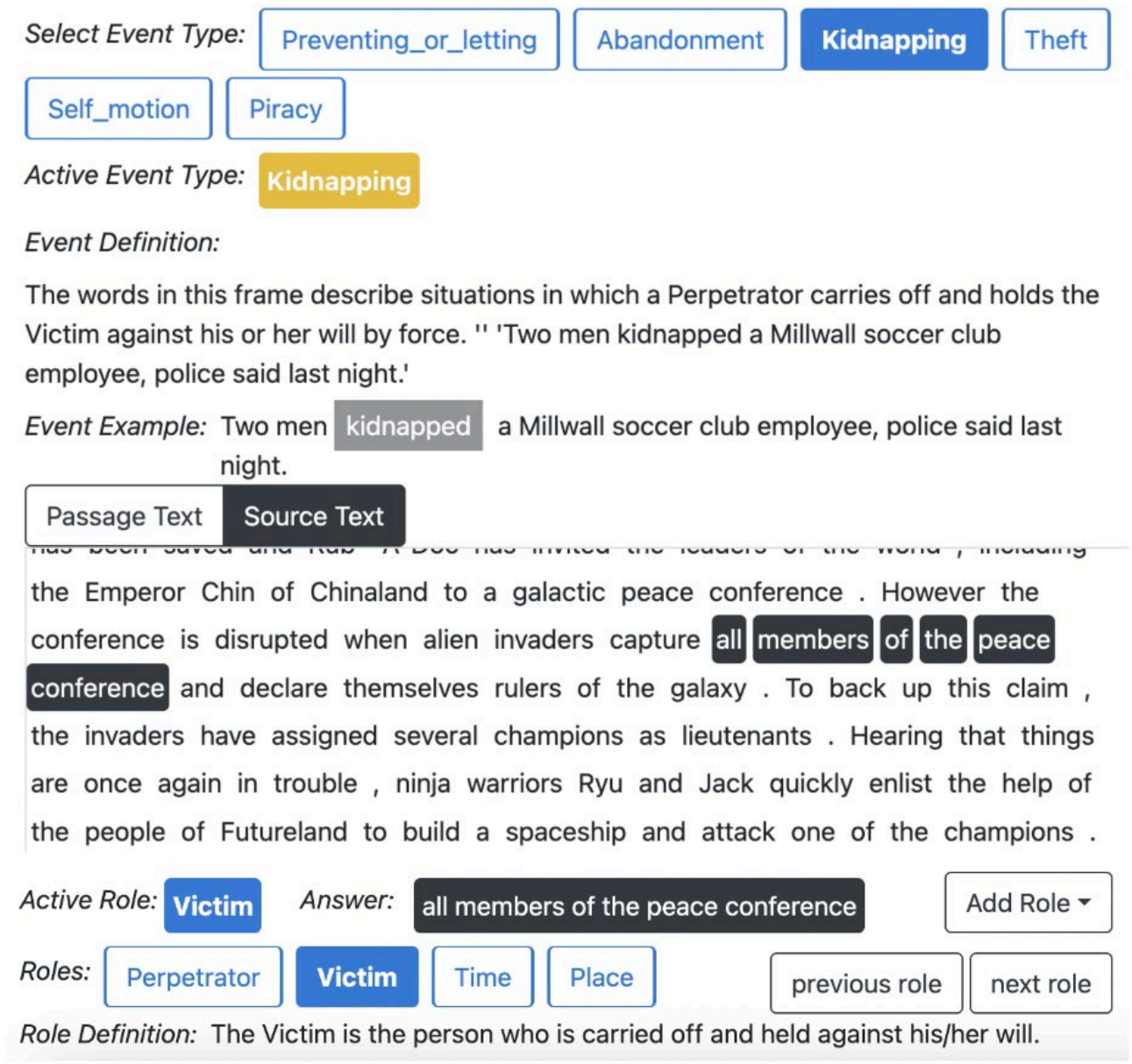
*Related to but distinct from similar tasks like Event Linking (Nothman 2012), Cross-Document Event Coreference (Bagga and Baldwin 1999, Cybulska and Vossen 2014, (Eirewetal.,2021,2022), and Predicate-Argument Alignment (Roth and Franke 2012, Wolfe et al. 2013, 2015).



Source Validation

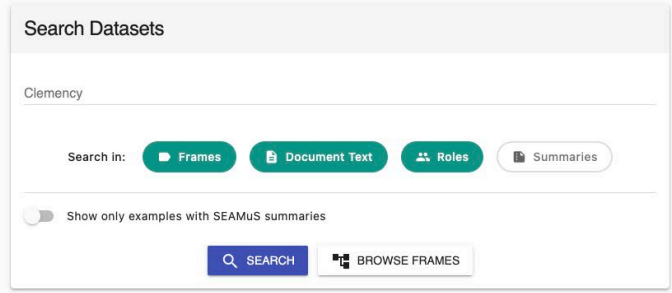
255 situation types from FrameNet

- **5 positive source validation (SV)** examples for each event type
- **5 negative source validation (SV)** examples for each event type

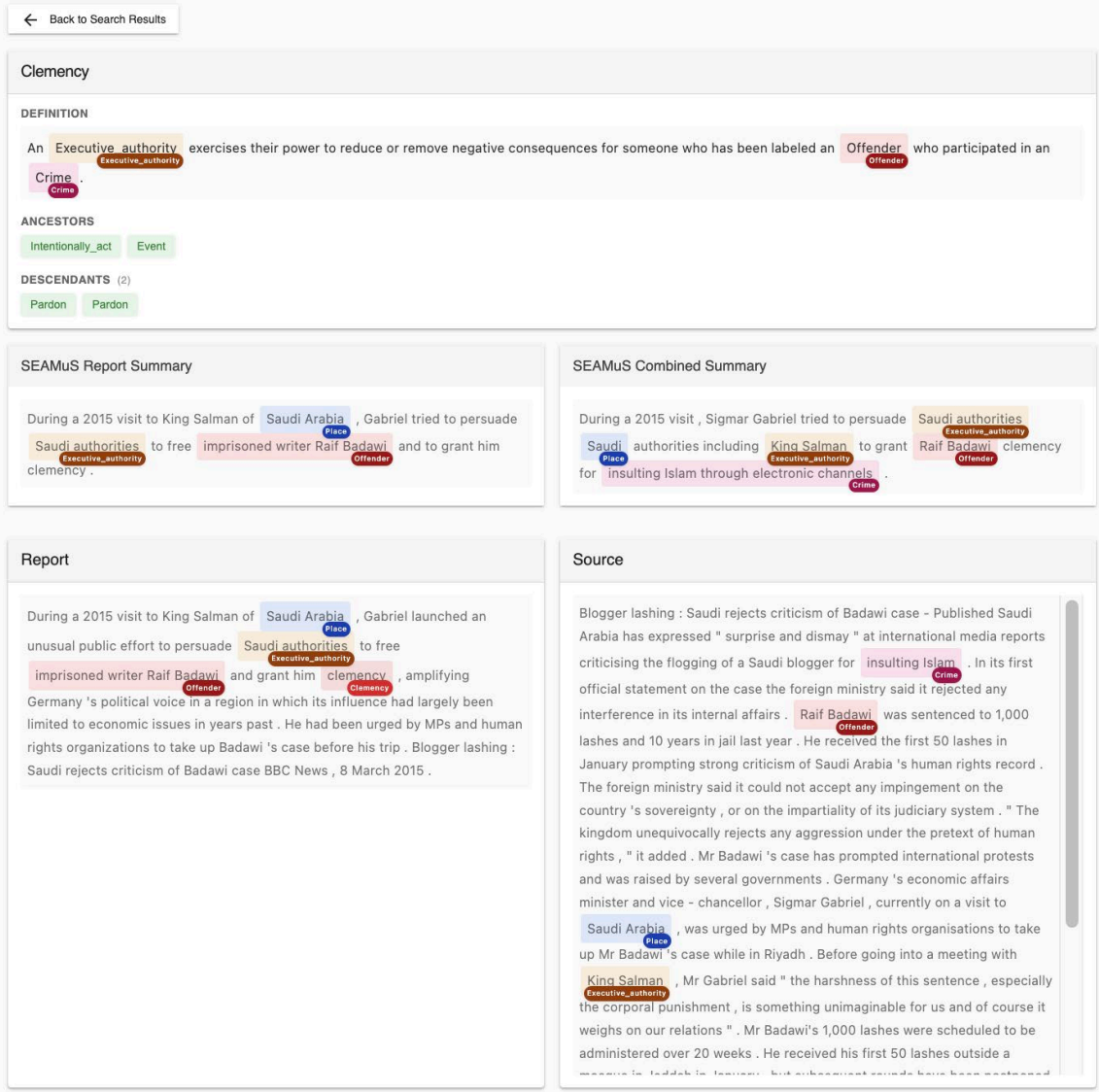


Cross-Document Argument Extraction

For each positive source validation, annotate roles in report and source.



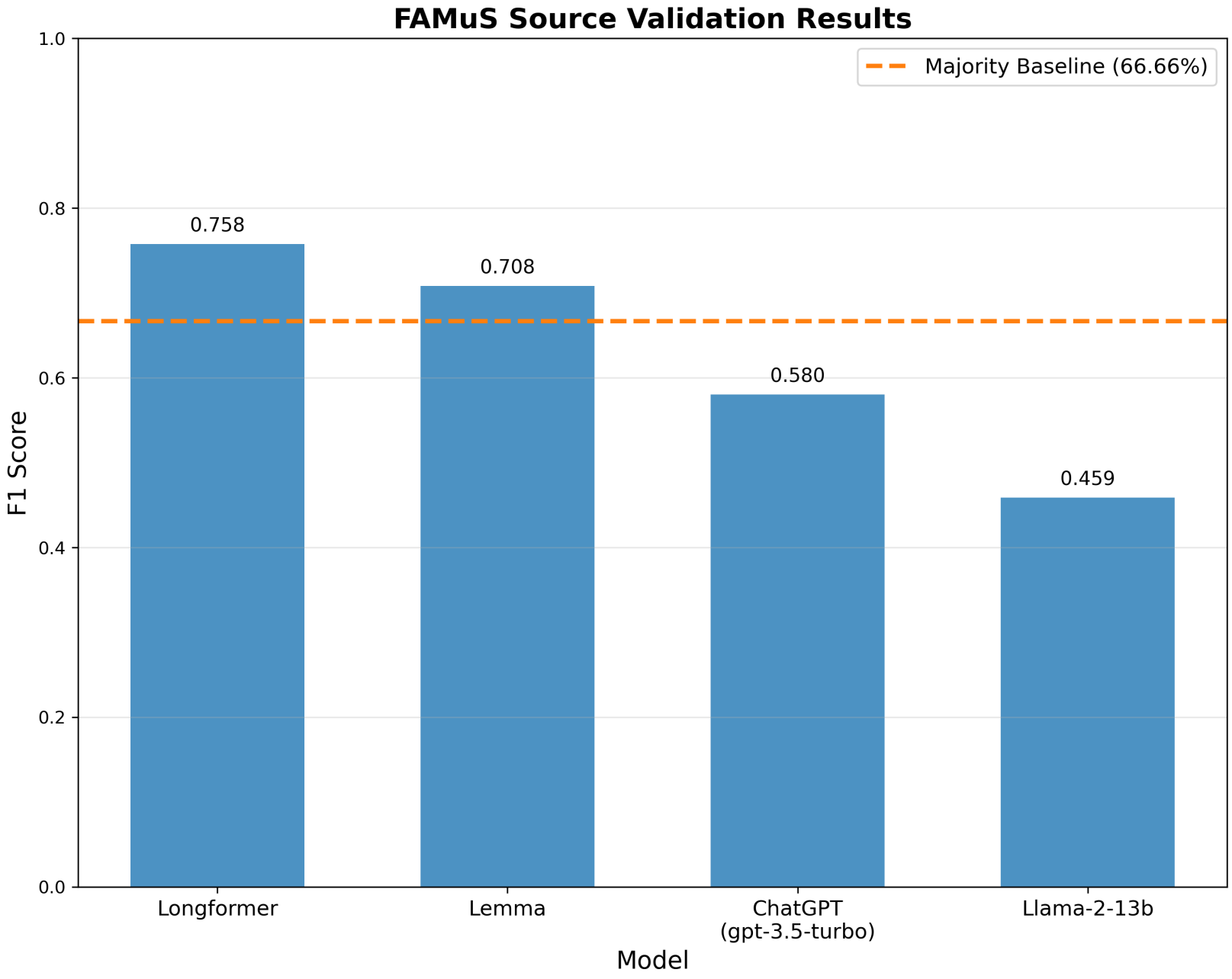
https://staramus.github.io/



Baselines: Source Validation

1. `Lemma`: target lemma is found in the source.
2. `Longformer`: document pair classifier
3. `ChatGPT`: prompt-based

Vashishtha et al. 2023


FAMuS Source Validation Results
Majority Baseline (66.66%)
F1 Score
1.0
0.8
0.6
0.4
0.2
0.0
0.758
0.708
0.580
0.459
Longformer
Lemma
ChatGPT
(gpt-3.5-turbo)
Llama-2-13b
Model

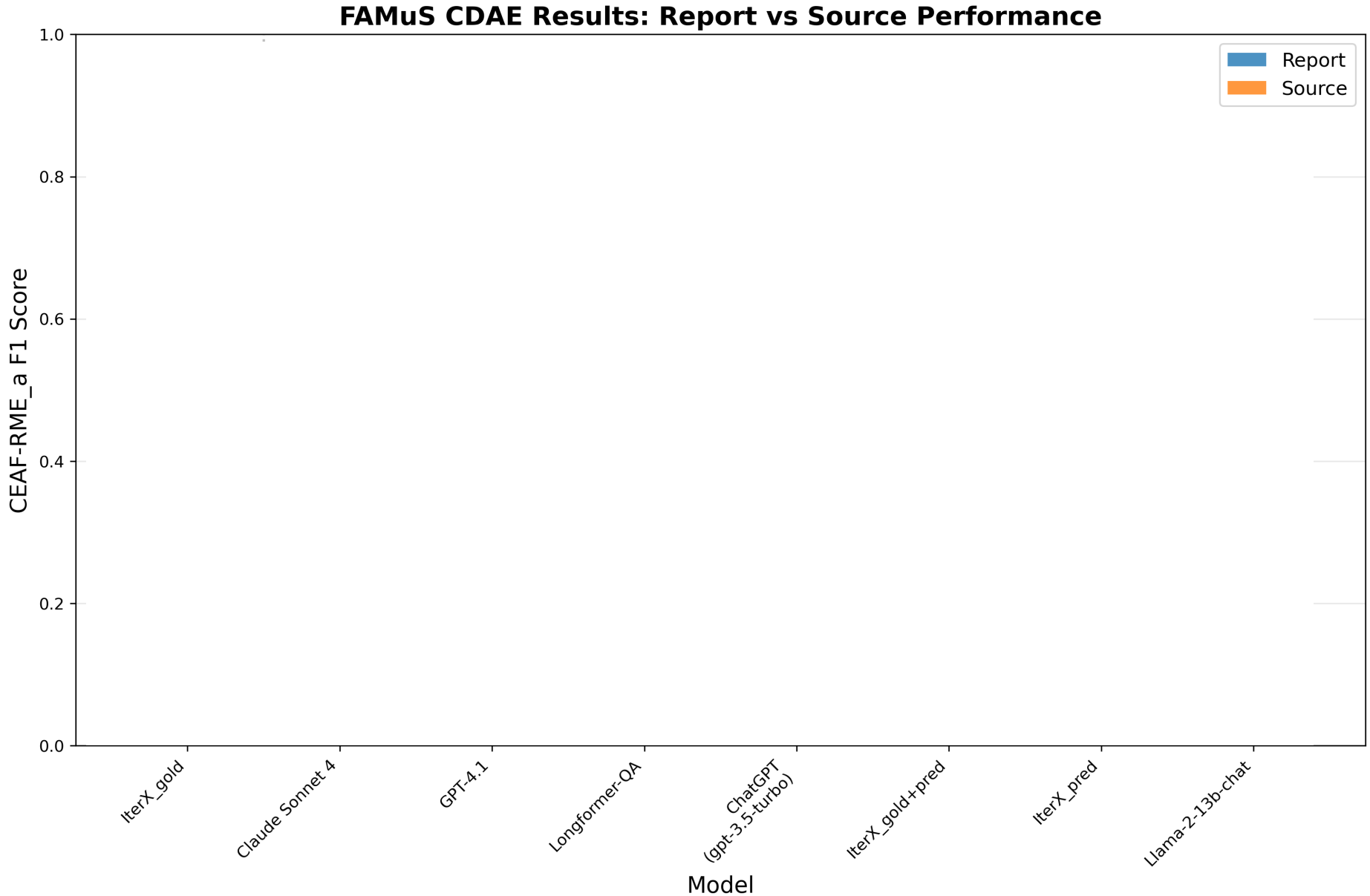
Baselines: Source Validation

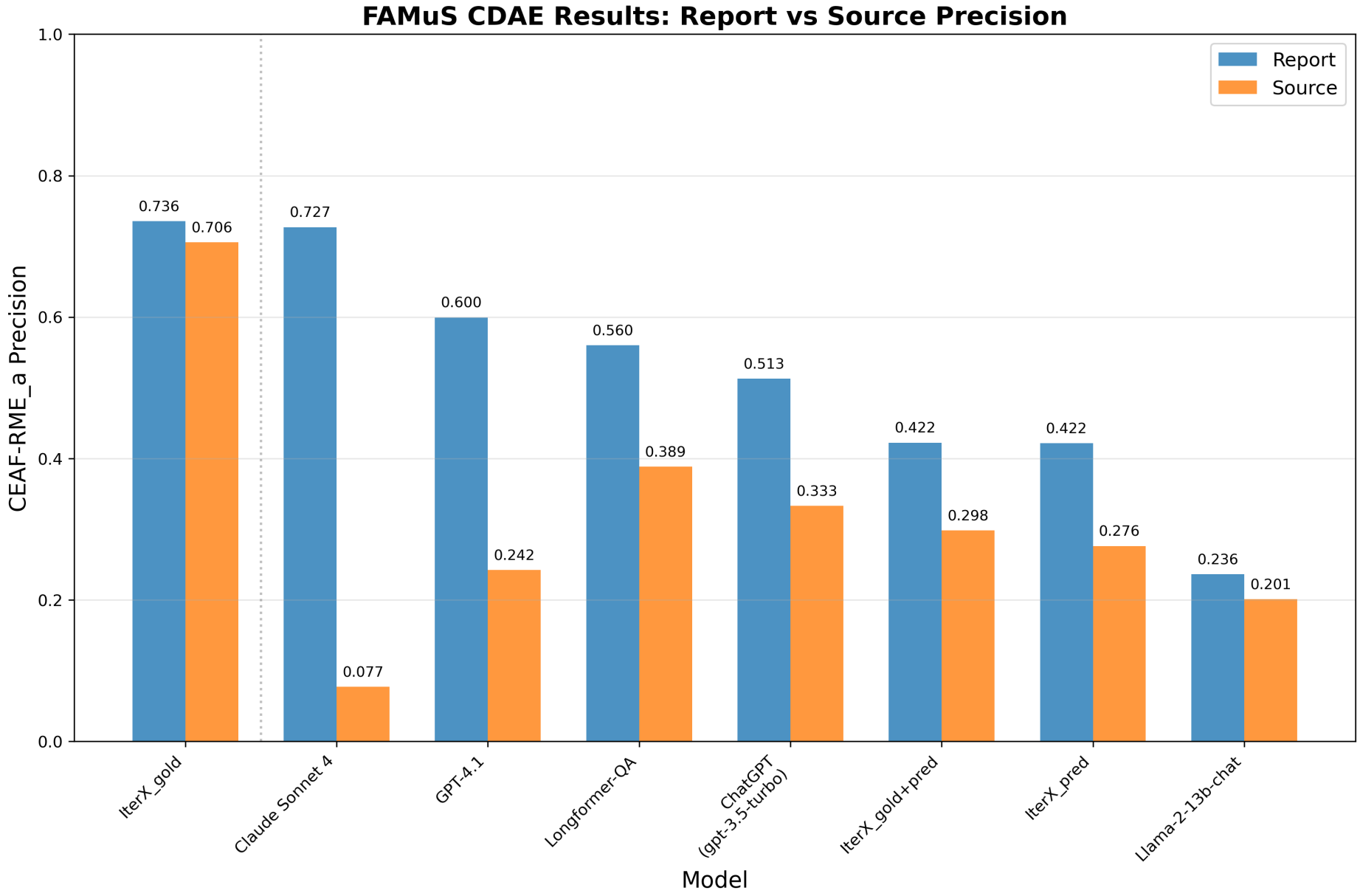
1. `Lemma`: target lemma is found in the source.
2. `Longformer`: document pair classifier
3. `ChatGPT`: prompt-based

Baselines: Cross-Doc Argument Extraction

1. `Report:` ignore the source and use spans from report
2. `IterX:` structured prediction model
3. `Longformer-QA:` fine-tuned QA model
4. `ChatGPT` and `Llama2`: few-shot prompting models

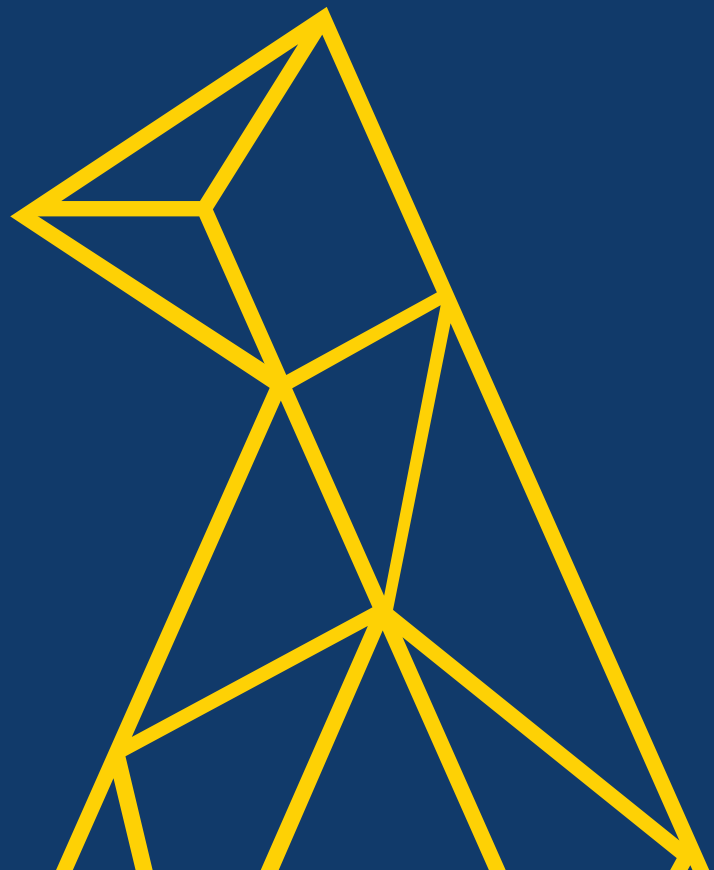
Vashishtha et al. 2023


FAMuS CDAE Results: Report vs Source Performance
Report
Source
CEAF-RME_a F1 Score
1.0
0.8
0.6
0.4
0.2
0.0
IterX_gold
Claude Sonnet 4
GPT-4.1
Longformer-QA
ChatGPT (gpt-3.5-turbo)
IterX_gold+pred
IterX_pred
Llama-2-13b-chat
Model


FAMuS CDAE Results: Report vs Source Precision
CEAF-RME_a Precision
Model
Report
Source
1.0
0.8
0.6
0.4
0.2
0.0
0.736
0.706
0.727
0.077
0.600
0.242
0.560
0.389
0.513
0.333
0.422
0.298
0.422
0.276
0.236
0.201
IterX_gold
Claude Sonnet 4
GPT-4.1
Longformer-QA
ChatGPT (gpt-3.5-turbo)
IterX_gold+pred
IterX_pred
Llama-2-13b-chat

Template Filling

Interim Discussion



Finding #1

We can do broad-coverage template extraction on short documents (e.g. Wikipedia snippets) reasonably well.

Finding #2

Long-document, broad-coverage template extraction is still very hard and the better you are at short documents, the worse you are on long ones.

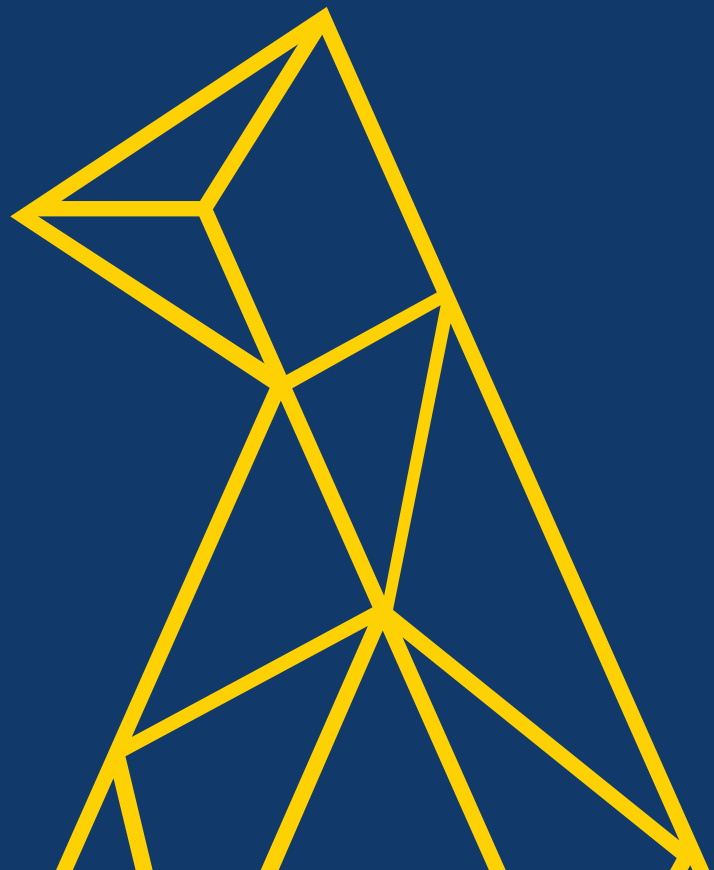
Challenge

Inferences are interesting and specific but lack sensitivity to the context of the situation description.

Idea

Generate an inference that captures the information captured by the template in a more context-sensitive way.

Event-Keyed Summarization



Goal

Leverage source-report relations to draw an inference capturing all available information relative to a template.

Idea

The template–in this case, a complex event type–is a complex query that guides inference.

Instructions

Select Event Type: Emotions_success_or_failure

Active Event Type: Emotions_success_or_failure

Event Definition:

A Topic or Stimulus fails or succeeds at causing an emotion in an Experiencer, which is conveyed by an Expressor or manifested in a State or description of an Event.

Event Example: They were dissatisfied with the arrangements.

Click here to know more details about the "Emotions_success_or_failure" Event Type

Valid Candidates in Report

(1) October 26

(2) On October 26

No candidate found for the current role

Report Summary

On October 26 , an immigration tribunal stated that it was " satisfied that the appellant has engaged in the unacceptable behaviour of fostering hatred , " which it said could lead to intercommunity violence in the UK .

Combined Summary

Enter your combined source-report summary here!

Unique ID: EN-4992-441-frame-Emotions_success_or_failure

Report Text | Source Text

On October 26 , an immigration tribunal concluded that May had been justified in her position . The tribunal stated that it is " satisfied that the appellant has engaged in the unacceptable behaviour of fostering hatred which might lead to intercommunity violence in the UK . "

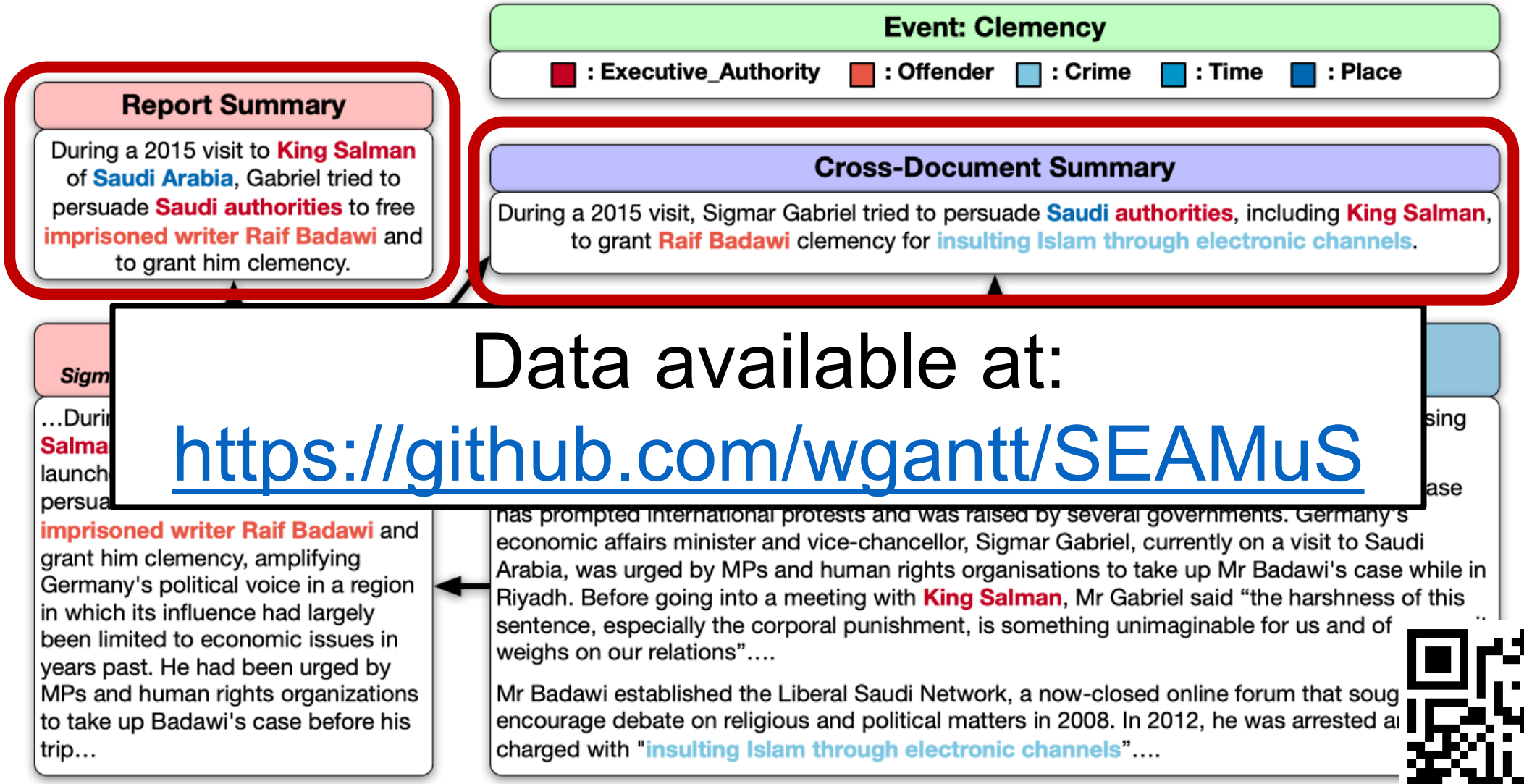
Active Role: Time *Answer:* October 26

Add Role

Roles: Experiencer | Topic | Stimulus | Expressor | State | Event | Time

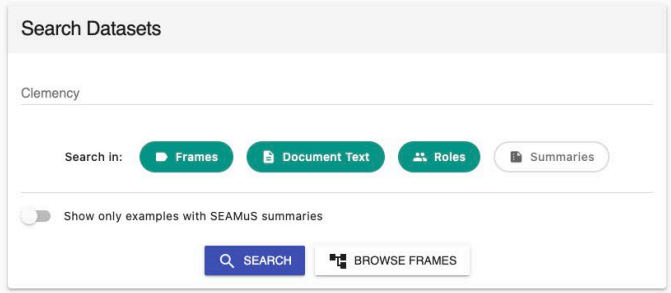
previous role | next role

Role Definition: The Time when the Experiencer, Expressor, Event, or State can be described as having said emotion.

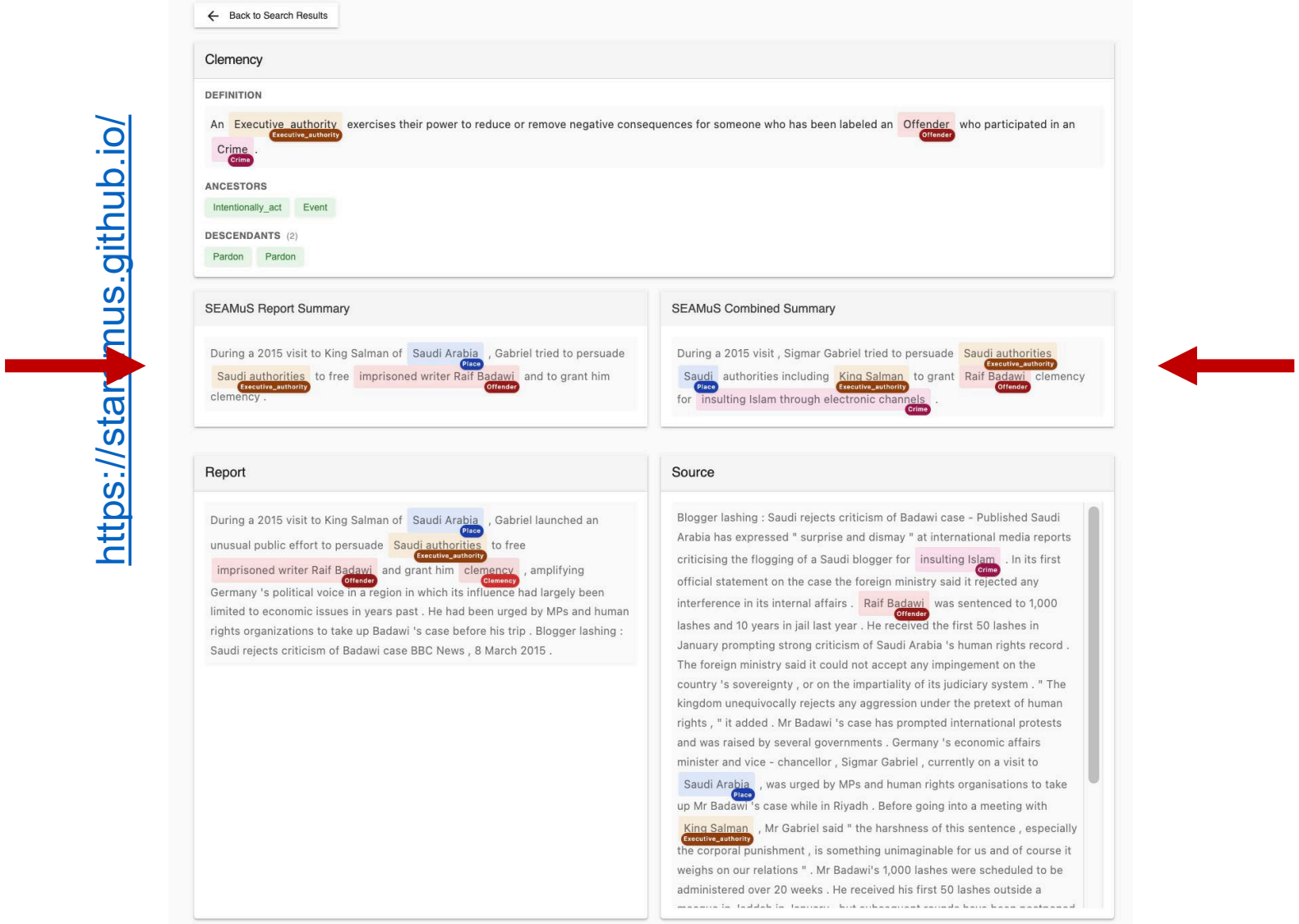


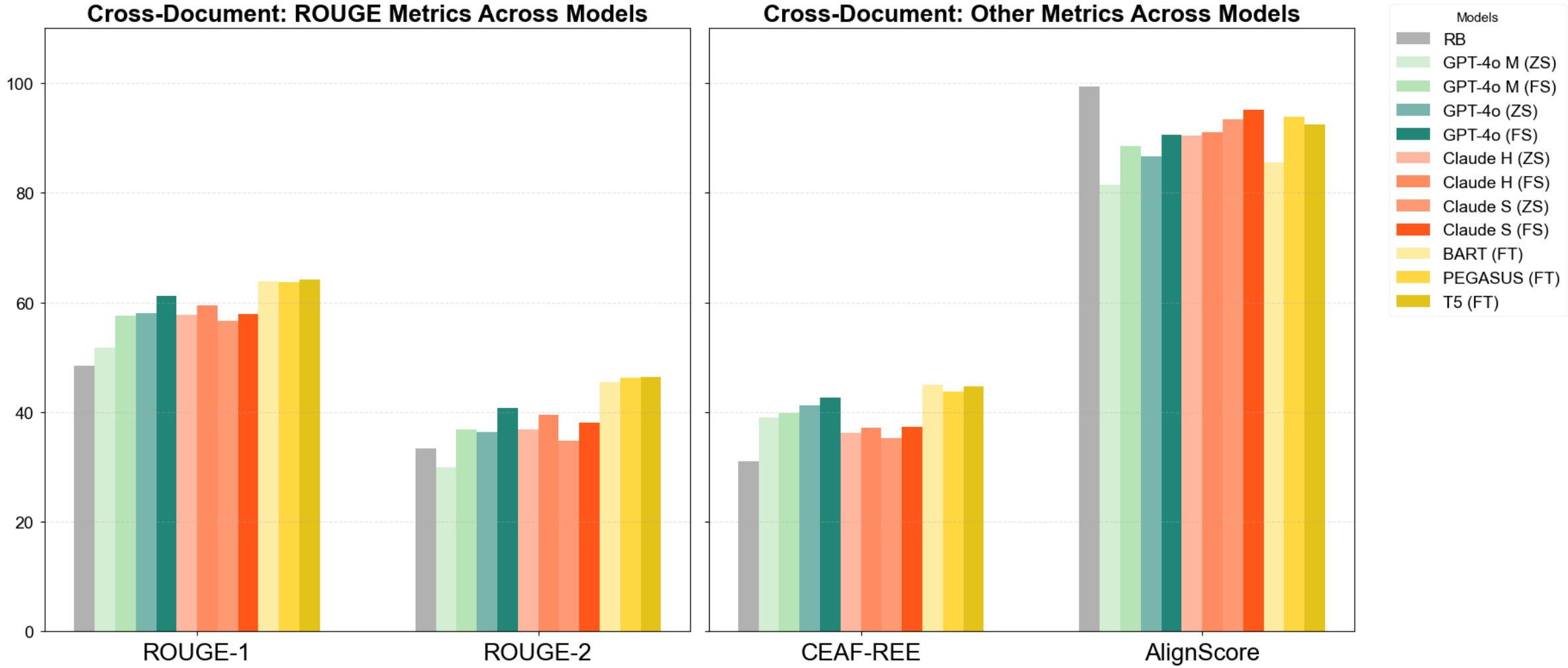
Data available at:

https://github.com/wgantt/SEAMuS

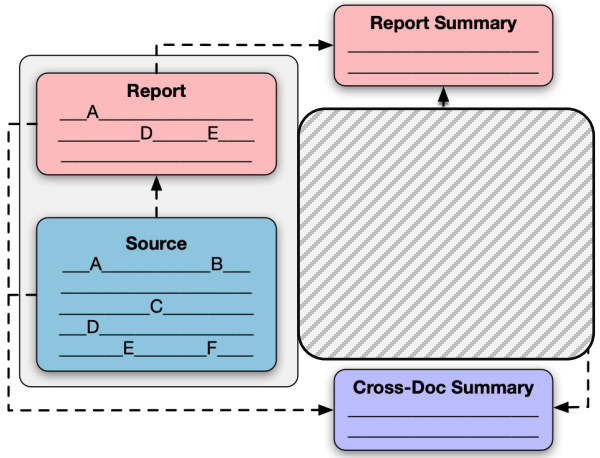


https://starmus.github.io/

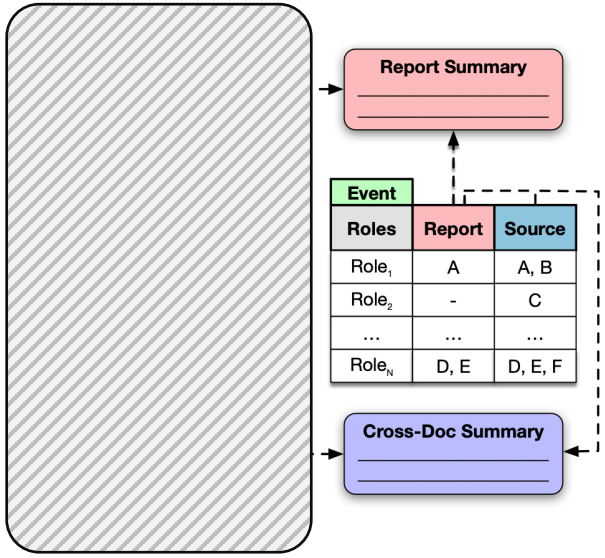



Cross-Document: ROUGE Metrics Across Models
Cross-Document: Other Metrics Across Models
100
80
60
40
20
0
ROUGE-1
ROUGE-2
CEAF-REE
AlignScore
Models
RB
GPT-4o M (ZS)
GPT-4o M (FS)
GPT-4o (ZS)
GPT-4o (FS)
Claude H (ZS)
Claude H (FS)
Claude S (ZS)
Claude S (FS)
BART (FT)
PEGASUS (FT)
T5 (FT)

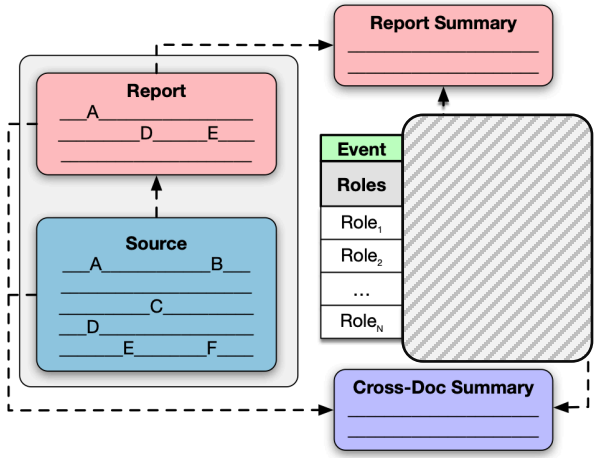
Text only: *only texts*



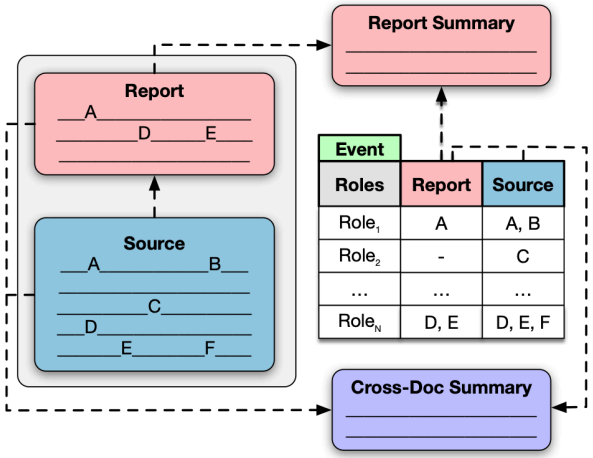
Event only: *only event argument annotations*

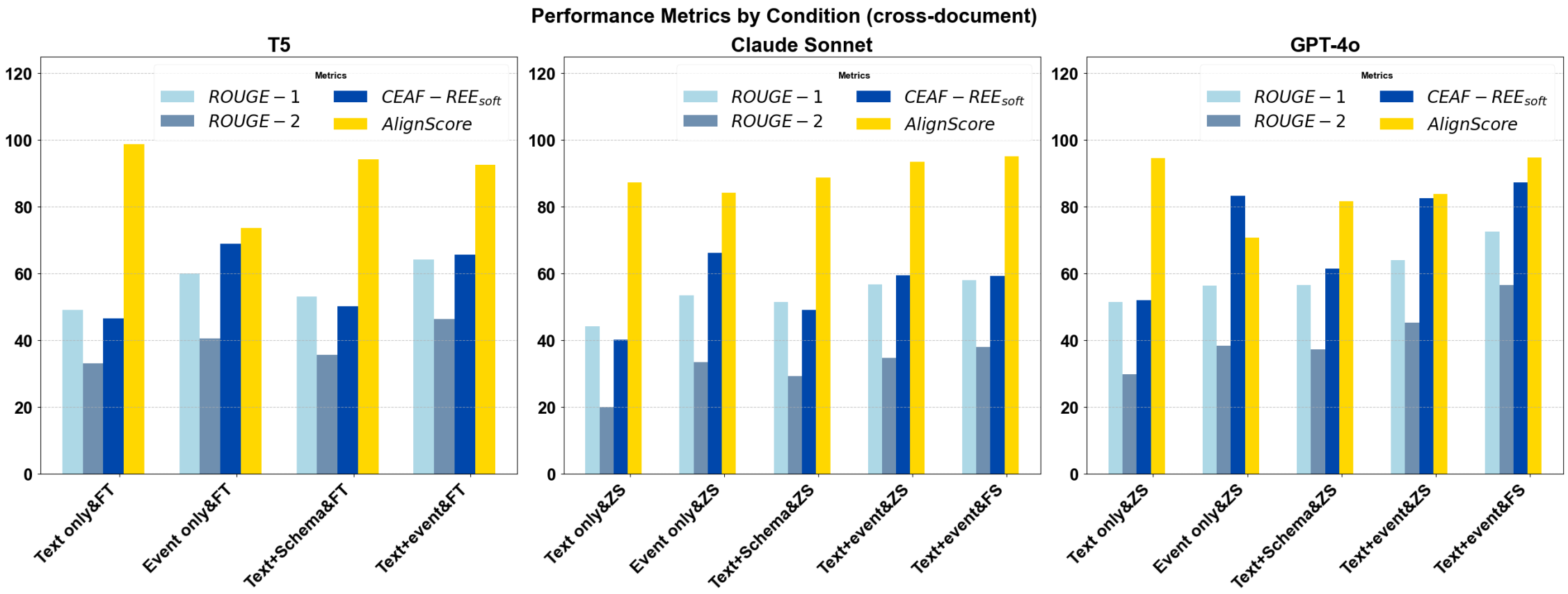


Text + Schema: *only texts, name of the event per text, and corresponding roles*

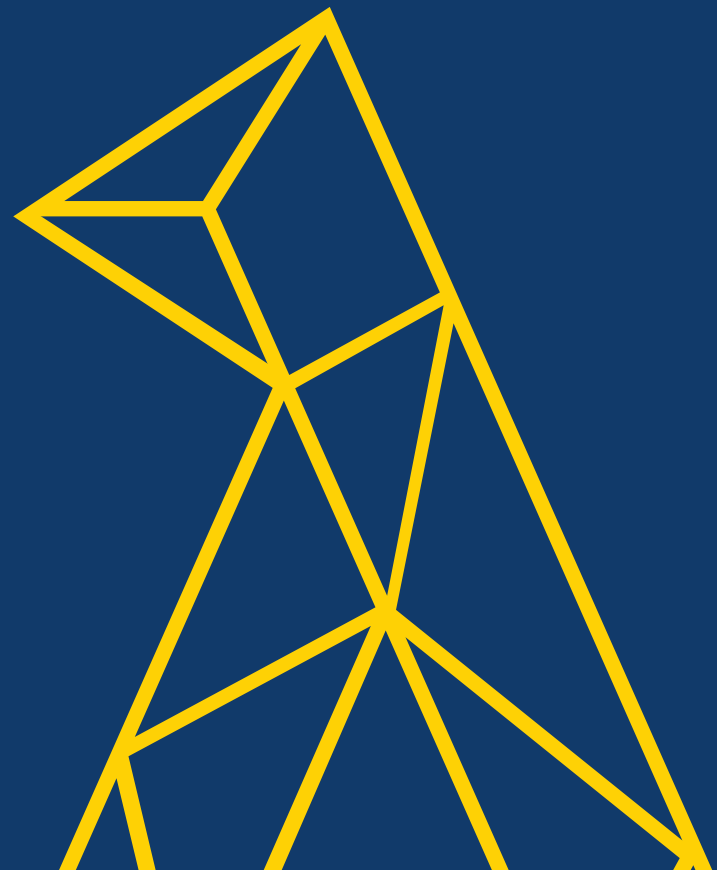


Text + Event: *unablated setting*




Performance Metrics by Condition (cross-document)
T5
Claude Sonnet
GPT-4o
Metrics
ROUGE – 1
ROUGE – 2
CEAF – REE$_{soft}$
AlignScore
120
100
80
60
40
20
0
Text only&FT
Event only&FT
Text+Schema&FT
Text+event&FT
Text only&ZS
Event only&ZS
Text+Schema&ZS
Text+event&ZS
Text+event&FS

Conclusion



Question

How do we design systems that capture the inferences we draw about situations based on their descriptions?

Ontology-factored approach

Map situation description to symbolic situation ontology and draw inferences using rules stated over that ontology.

Challenge #1: Expense

Ontologies and annotated corpora are expensive to build and maintain because they require highly trained experts.

Challenge #2: Brittleness

Ontologies do not easily capture the ways in which context modulates the inferences that we draw.

Question

How do we design systems that capture the inferences we draw about situations based on their descriptions?

Ontology-factored approach

Map situation description to symbolic situation ontology and draw inferences using rules stated over that ontology.

Ontology-free approach

Map situation descriptions to natural language strings expressing the inferences of interest.

Ontologies as Representational Scaffolding

1. Ontologies provide guidance about what the interesting inferences are.
2. These inferences are directly associated with a text, as in ontology-free approaches.

Thanks!

Supported by NSF-BCS-2040831 (*Computational Modeling of the Internal Structure of Events*), University of Rochester, JHU HLTCOE, DARPA AIDA, DARPA KAIROS, and IARPA BETTER.

Kyle Rawlins
JHU

Ben Van Durme
JHU

Jordan Boyd-Graber
UMD College Park

Rachel Rudinger
UMD College Park

Lelia Glass
Georgia Tech

Anqi Liu
JHU

Kevn Duh
JHU

Sid Vashishtha
DeepMind

Will Walden
JHU

Yunmo Chen
JHU

Sam Barham
JHU

Mahsa Yarmohammadi
JHU

Kenton Murray
JHU

Reno Kriz
JHU

Arsal Imtiaz
Everlaw

Chandrashekar Nair
Everlaw

Alex Martin
JHU

Weiwei Gu
ASU

Tongfei Chen
Microsoft

Venkat Govindarajan
Ithaca College

Dee Ann Reisinger
Bryn Mawr

Pushprendre Rastogi
DeepMind

Elias Stengel-Eskin
UT Austin

Sheng Zhang
Microsoft

Zhengping Jiang
JHU

Orion Weller
JHU

Michelle Yuan
UMD College Park

Adam Poliak
Barnard