

# Generalization in Multilingual Semantic Parsing

*Workshop on Ten Years of BabelNet and  
Multilingual Neurosymbolic  
Natural Language Understanding*  
5 July 2022

Daniel Hershcovich

UNIVERSITY OF COPENHAGEN



# Why Meaning Representation for NLP?



## Performance

Inductive bias

Access to structured data

Reasoning ability



## Understanding

Interpretability

Theoretical analysis

Fine-grained control



## Generalization

Languages

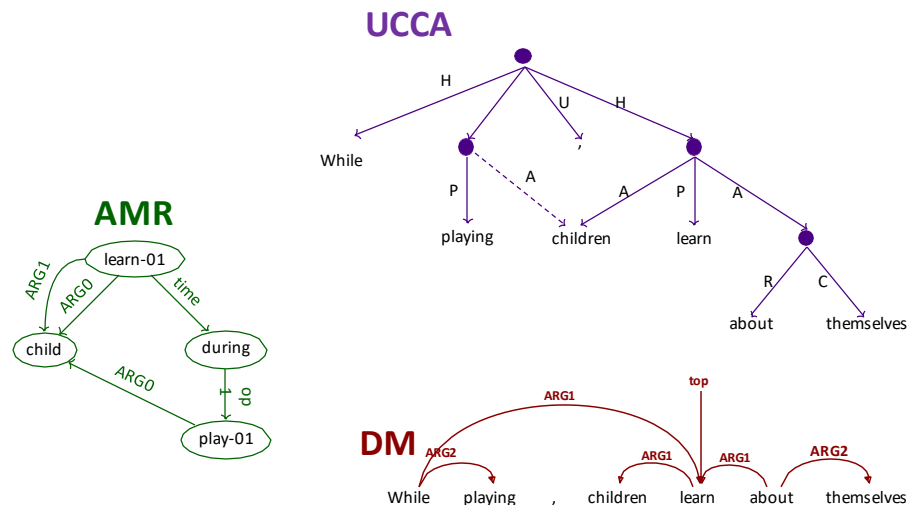
Domains

Tasks

# Graph-based Meaning Representations

Type of Information	DM	PSD	EDS	DMRS	UCCA	AMR	PMB
<b>Predicates–Arguments</b>	+	+	++	++	+	++	++
<b>Sense Differentiation</b>	+	++	+	+	–	++	++

Graph-Based Meaning Representations: Design and Processing (Koller et al., ACL 2019)



Multitask Parsing Across Semantic Representations (Hershcovich et al., ACL 2018)

# Universal Conceptual Cognitive Annotation (UCCA)

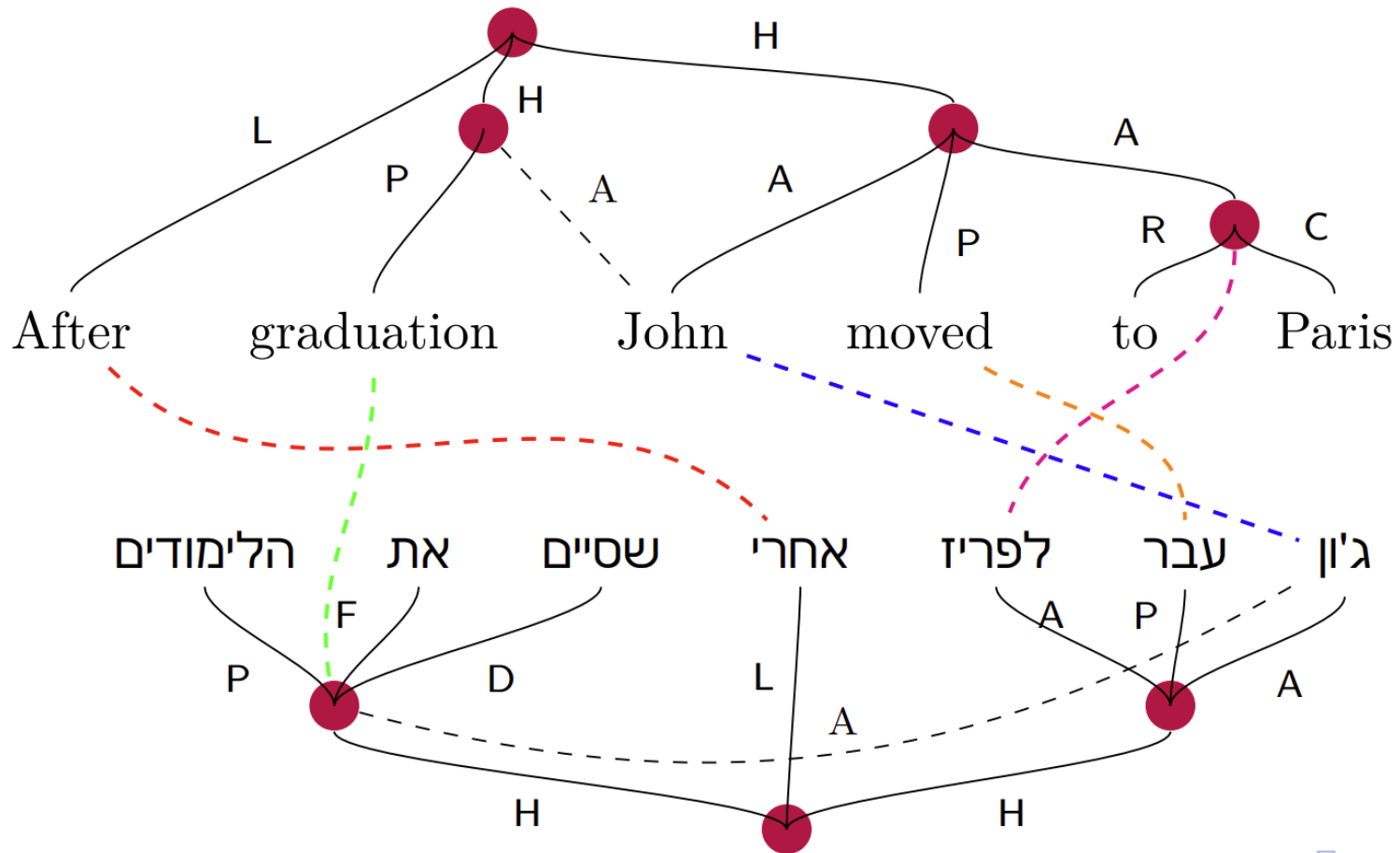
## Design principles

- Cross-linguistic portability and stability
- Accessibility to non-expert annotators
- Modularity of semantic components

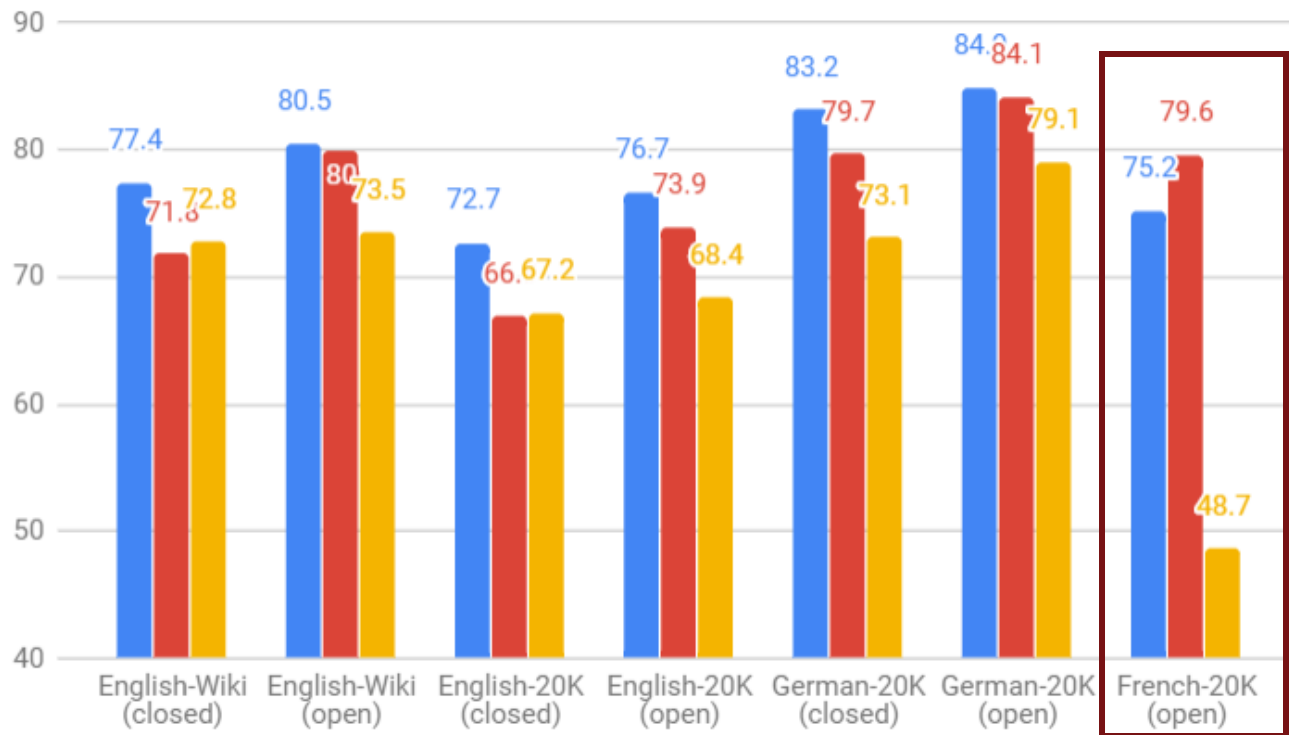
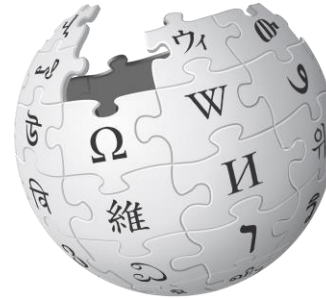
## Corpora

- English, German, French, Russian, Hebrew  
& now also Turkish

# UCCA



# UCCA Parsing

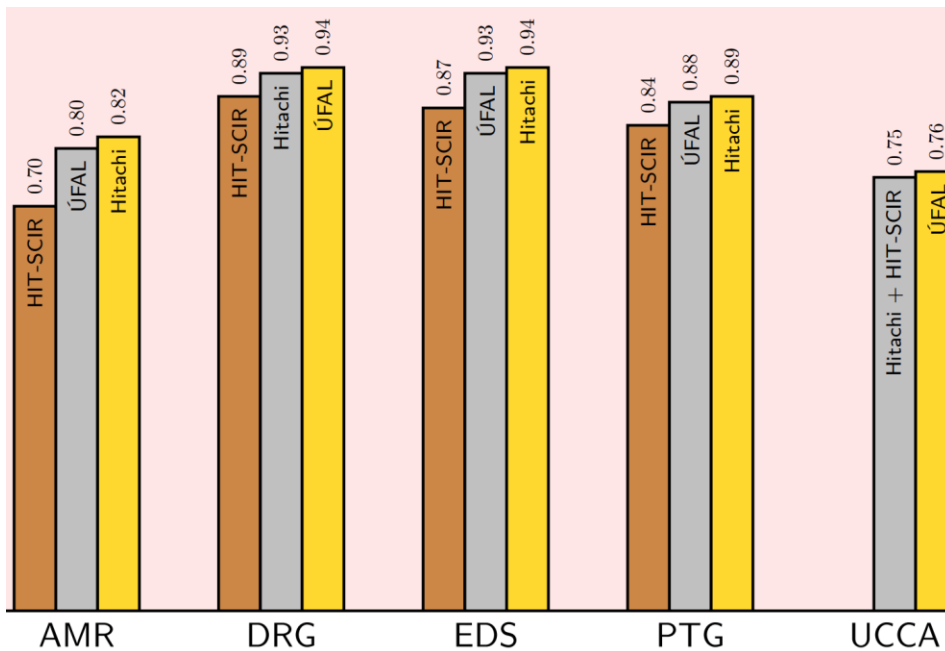


Successful cross-lingual transfer

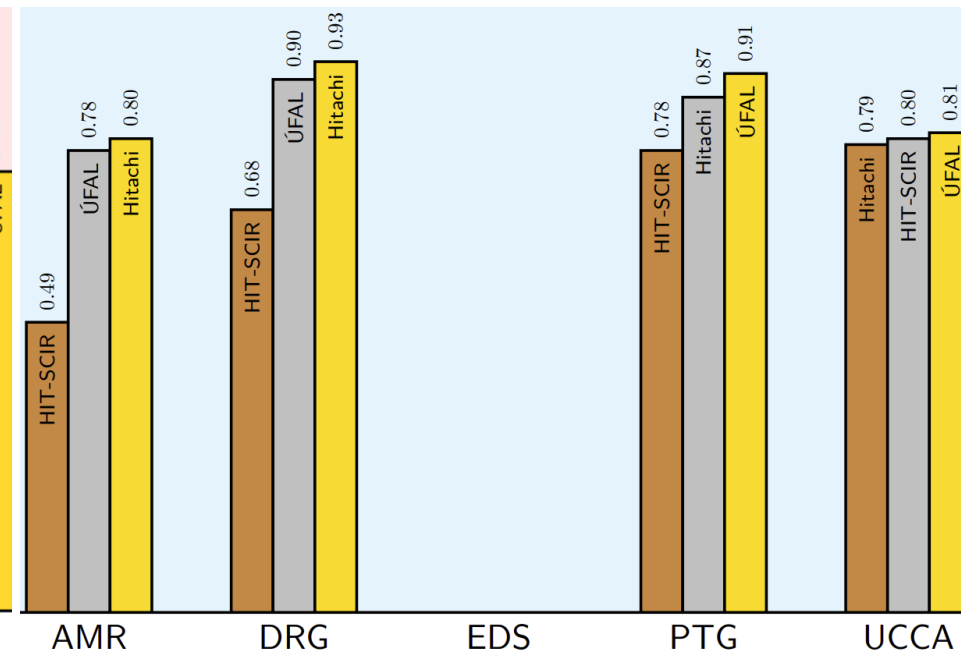
(Few-shot)

# Meaning Representation Parsing

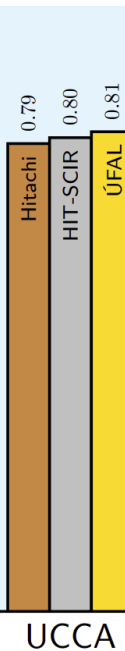
## English



## Chinese German



## Czech German

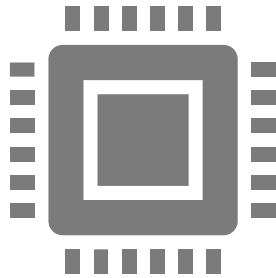


(Fully supervised)

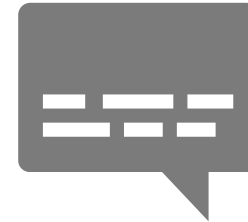
Similar results across languages

# Cross-lingual Generalization

Dominant approaches:



Multilingual models  
(mBERT, XLM-R etc.)



Annotation projection  
(manual/automatic translation)



# Compositional Generalization



"THE ABILITY TO SYSTEMATICALLY GENERALIZE TO **COMPOSED** TEST EXAMPLES OF A CERTAIN DISTRIBUTION AFTER BEING EXPOSED TO THE NECESSARY **COMPONENTS** DURING TRAINING ON A DIFFERENT DISTRIBUTION"

## Train set

Who directed Inception?

Did Greta Gerwig produce Goldfinger?

...

## Test set

Did Greta Gerwig direct Goldfinger?

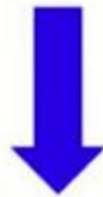
Who produced Inception?

...

# CFQ (Compositional Freebase Questions)

English

Did a film editor's parent direct M0?

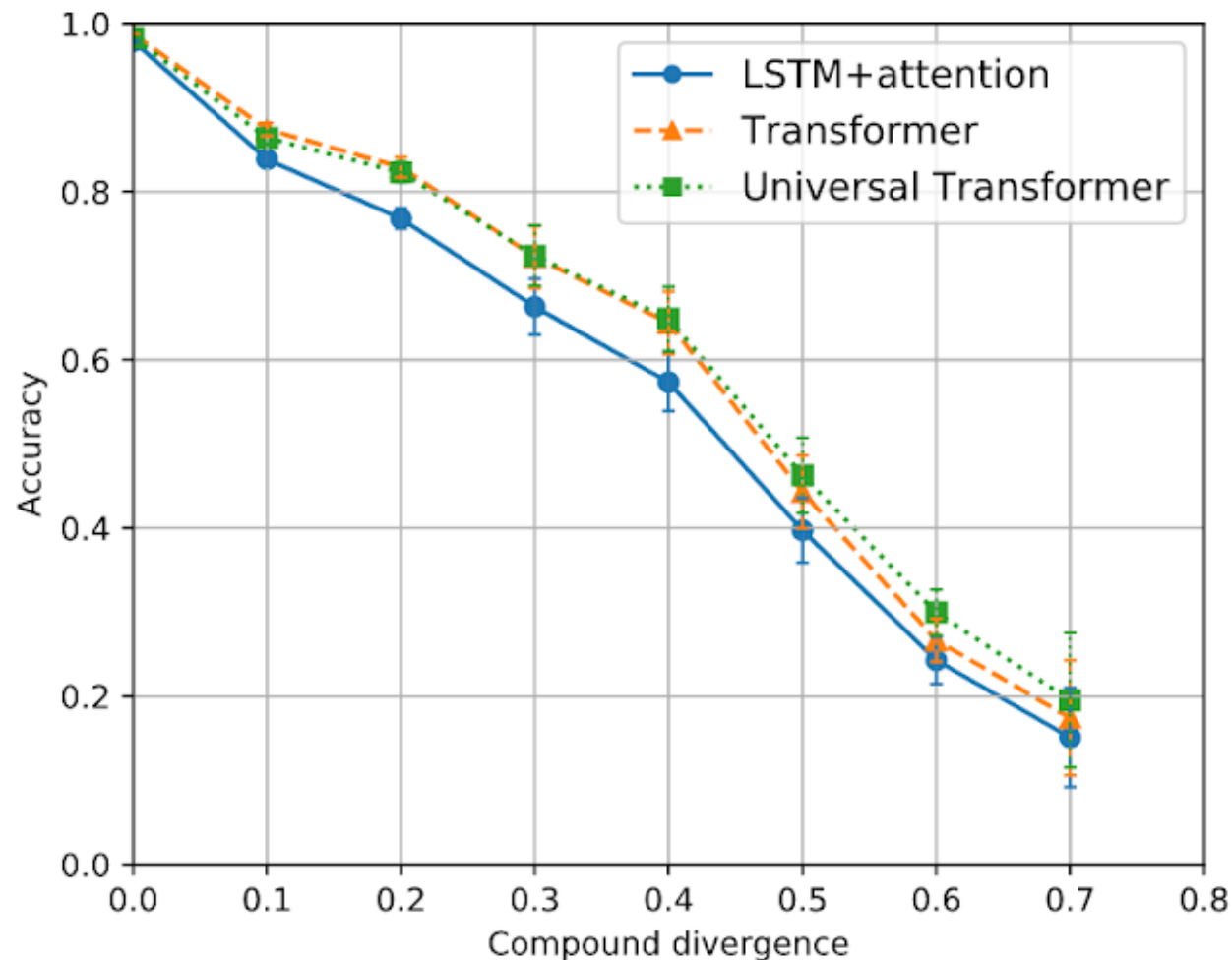


**Grammar Rules, Inference Rules,  
Resolution Rules, Knowledge**

SPARQL

```
SELECT count(*) WHERE {  
  ?x0 ns:film.director.film [M0] .  
  ?x0 ns:people.person.children ?x1 .  
  ?x1 a ns:film.editor  
}
```

# CFQ (Compositional Freebase Questions)



# Multilingual Compositional Wikidata Questions



Ruixiang Cui



Rahul Aralikatte



Heather Lent

CFQ

(239K instances)



Property mapping  
Entity substitution

English



Automatic & manual  
translation

MCWQ

(124K instances)



Hebrew, Kannada,  
Chinese & English

# Multilingual Compositional Wikidata Questions

Lang. Question

---

En	Did Lohengrin's male actor marry Margarete Joswig
He	האם הגבר הן השחקן הלוהנגרין עם האקטור הזכרני נשוי למרגרטה יוסוויג
Kn	ಲೋಹೆಂಗ್ರಿನ್ ಅವರ ಪುರುಷ ನಟ ವಿವಾಹವಾದರು ಮಾರ್ಗರೇಟ್ ಜೋಸ್ವಿಗ್
Zh	Lohengrin 的男演员嫁给了 Margarete Joswig 吗

---

SPARQL Query:

```
ASK WHERE { ?x0 wdt:P453 wd:Q50807639 . ?x0
wdt:P21 wd:Q6581097 . ?x0 wdt:P26 wd:Q1560129 .
FILTER ( ?x0 != wd:Q1560129 )}
```

# Multilingual Compositional Generalization

Within-language	Random split			
	En	He	Kn	Zh
Exact Match (%)				
LSTM+Attention	96.6	80.8	88.7	86.8
E. Transformer	99	90.4	93.7	92.2
mBERT	98.7	<b>91</b>	95.1	<b>93.3</b>
T5-base+RIR	98.5			
mT5-small+RIR	98.6	90	93.8	91.8
mT5-base+RIR	<b>99.1</b>	90.6	94.2	92.2

# Multilingual Compositional Generalization

Within-language Exact Match (%)	MCD <sub>mean</sub>				Random split			
	En	He	Kn	Zh	En	He	Kn	Zh
LSTM+Attention	19.4	15.5	17.6	16.2	96.6	80.8	88.7	86.8
E. Transformer	29.3	18.9	20.2	18.9	99	90.4	93.7	92.2
mBERT	26.6	22.7	21.6	23.4	98.7	<b>91</b>	95.1	<b>93.3</b>
T5-base+RIR	28.1				98.5			
mT5-small+RIR	<b>38.3</b>	29.3	31.5	<b>36.3</b>	98.6	90	93.8	91.8
mT5-base+RIR	33.8	<b>33.2</b>	<b>32.1</b>	29.6	<b>99.1</b>	90.6	94.2	92.2

# Multilingual Compositional Generalization

Within-language	MCD <sub>mean</sub>				Random split			
	En	He	Kn	Zh	En	He	Kn	Zh
LSTM+Attention	19.4	15.5	17.6	16.2	96.6	80.8	88.7	86.8
E. Transformer	29.3	18.9	20.2	18.9	99	90.4	93.7	92.2
mBERT	26.6	22.7	21.6	23.4	98.7	<b>91</b>	95.1	<b>93.3</b>
T5-base+RIR	28.1				98.5			
mT5-small+RIR	<b>38.3</b>	29.3	31.5	<b>36.3</b>	98.6	90	93.8	91.8
mT5-base+RIR	33.8	<b>33.2</b>	<b>32.1</b>	29.6	<b>99.1</b>	90.6	94.2	92.2

## Cross-lingual (from English)

mT5-small+RIR	0.2	0.3	0.2	0.5	0.4	1.1
mT5-base+RIR	0.4	0.7	1.5	1.1	0.9	7.2



# Multilingual Compositional Generalization

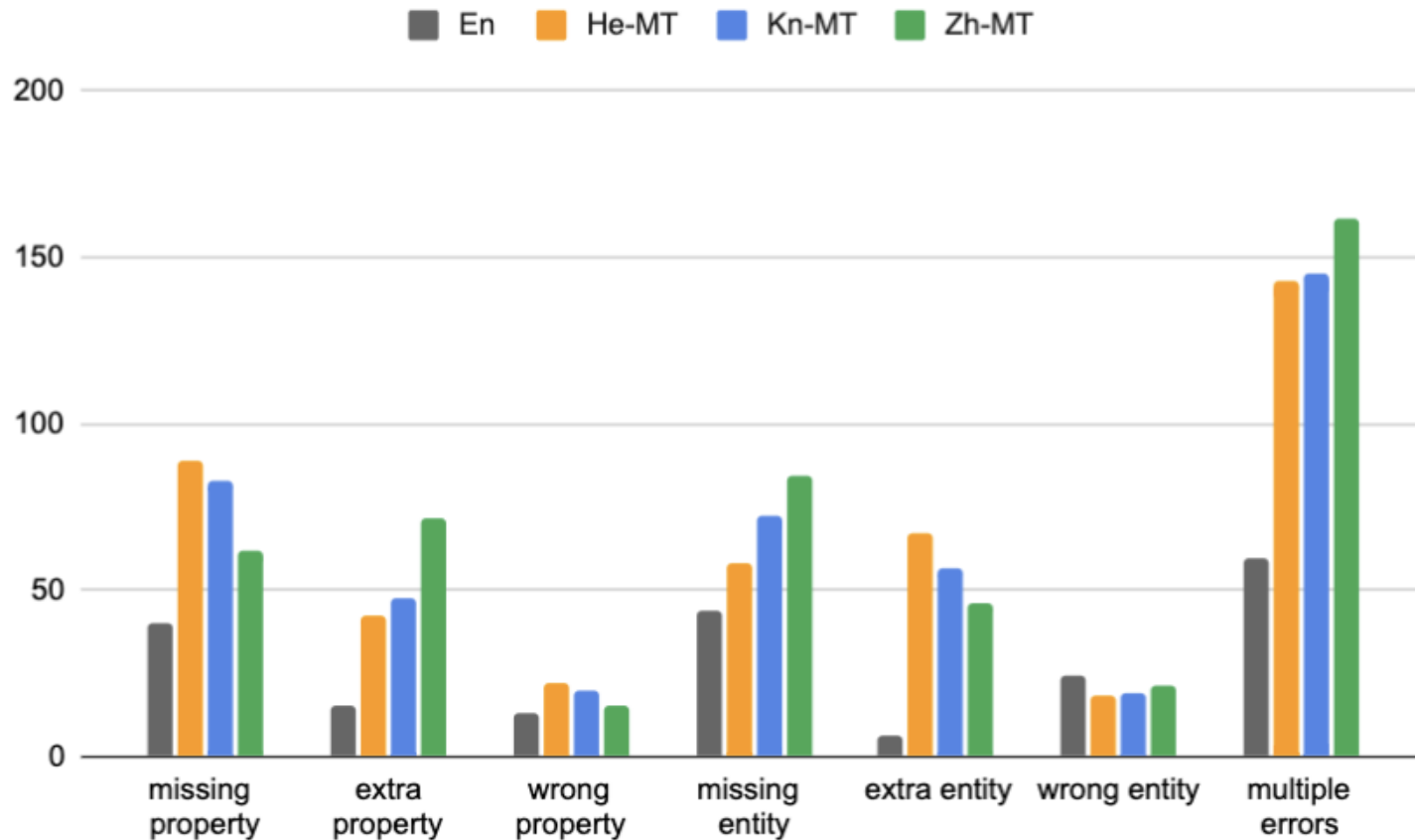
Within-language	MCD <sub>mean</sub>				Random split			
Exact Match (%)	En	He	Kn	Zh	En	He	Kn	Zh
LSTM+Attention	19.4	15.5	17.6	16.2	96.6	80.8	88.7	86.8
E. Transformer	29.3	18.9	20.2	18.9	99	90.4	93.7	92.2
mBERT	26.6	22.7	21.6	23.4	98.7	<b>91</b>	95.1	<b>93.3</b>
T5-base+RIR	28.1				98.5			
mT5-small+RIR	<b>38.3</b>	29.3	31.5	<b>36.3</b>	98.6	90	93.8	91.8
mT5-base+RIR	33.8	<b>33.2</b>	<b>32.1</b>	29.6	<b>99.1</b>	90.6	94.2	92.2

## Cross-lingual (from English)

SPARQL BLEU	En	He	Kn	Zh	En	He	Kn	Zh
mT5-small+RIR	87.5	53.8	53.2	59	99.9	60.4	59.9	63.8
mT5-base+RIR	86.4	46.4	46	52.7	99.9	63.2	63.5	70.6
Exact Match (%)								
mT5-small+RIR		0.2	0.3	0.2		0.5	0.4	1.1
mT5-base+RIR		0.4	0.7	1.5		1.1	0.9	7.2

# Error Analysis

## Cross-lingual prediction errors

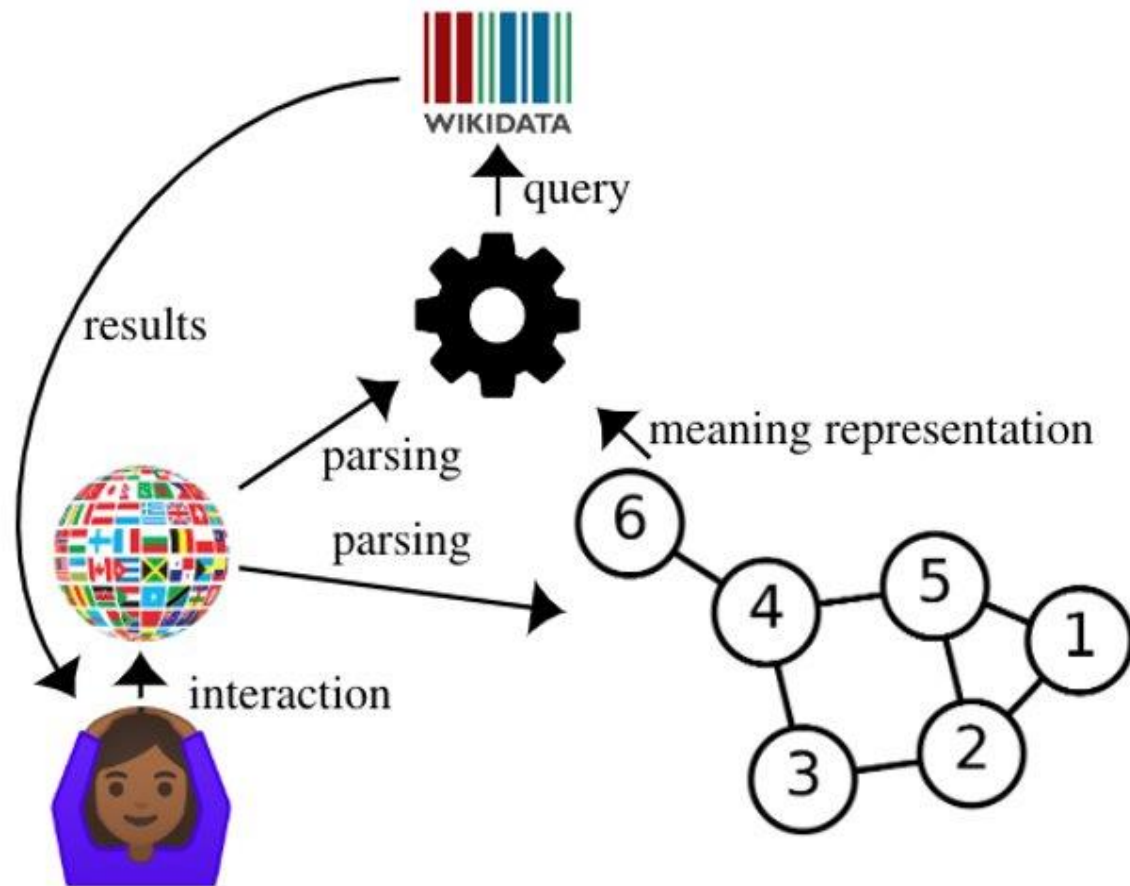
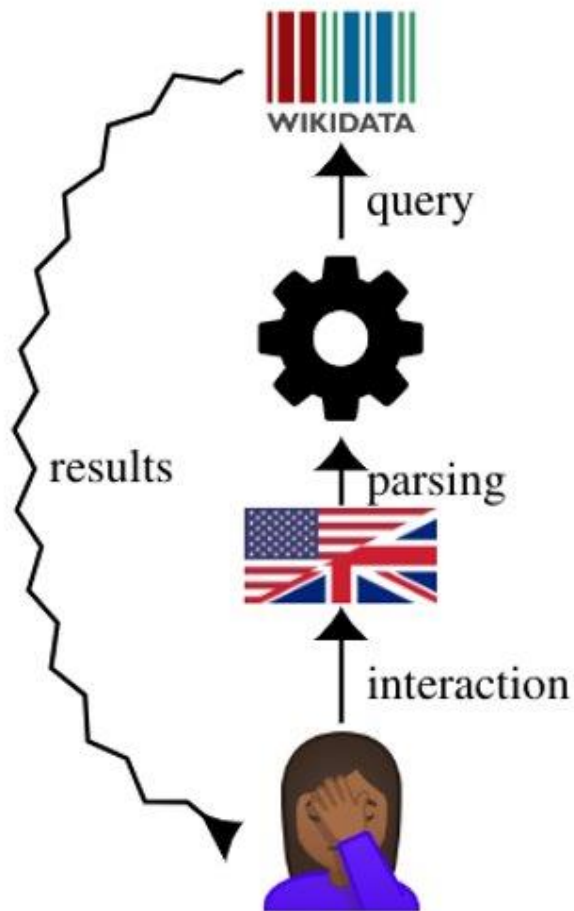


# Error Analysis

Question	Was M0 written by and directed by M1 , M2 , and M3
Gold	ASK WHERE { M0 wdt:P57 M1 . M0 wdt:P57 M2 . M0 wdt:P57 M3 . M0 wdt:P58 M1 . M0 wdt:P58 M2 . M0 wdt:P58 M3 }
Inferred	ASK WHERE { M0 wdt:P57 M1 . M1 wdt:P57 M2 . M0 wdt:P58 M3 } (director) (screenwriter)

Incorrect predicate-argument structure

# Can Multilingual Meaning Representation Help?



# Summary



Multilingual compositional generalization benchmark



Similar within-language generalization across languages



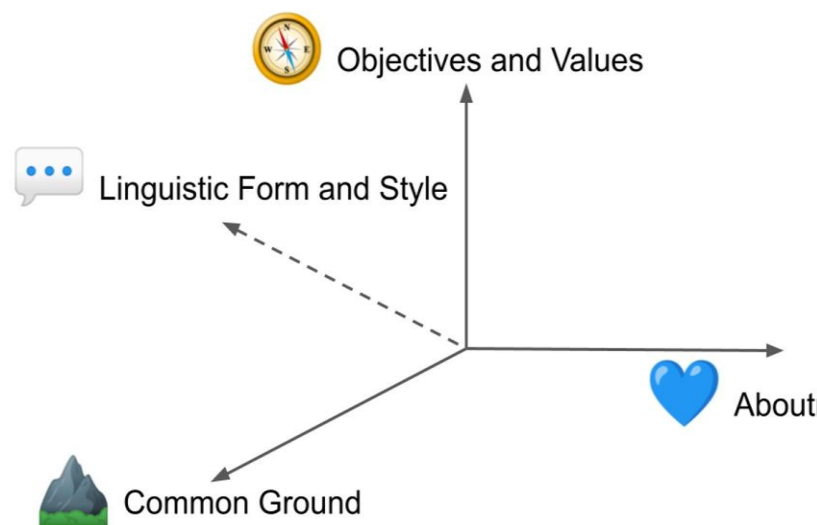
Zero-shot cross-lingual generalization fails

# Limitations & Future Work

- Synthetic & unnatural (inherited from CFQ)
  - ↳ Paraphrasing & expansion

- Mostly automatic translation
  - ↳ Improvement with RBMT

- No cultural adaptation
  - ↳ Native data collection



# Thanks!

danielhers.github.io

 @daniel\_hers

dh@di.ku.dk

# Semantics for Analysis

QA\_English

CONTEXT: *A piece of paper was later found on which he had written his last statements in **two** languages, Latin and German. Only **one** statement was in Latin and the rest in German.*

QUESTION: *In what language were **most** statements written?* ANSWER: German PREDICTED ANSWER: Latin and German

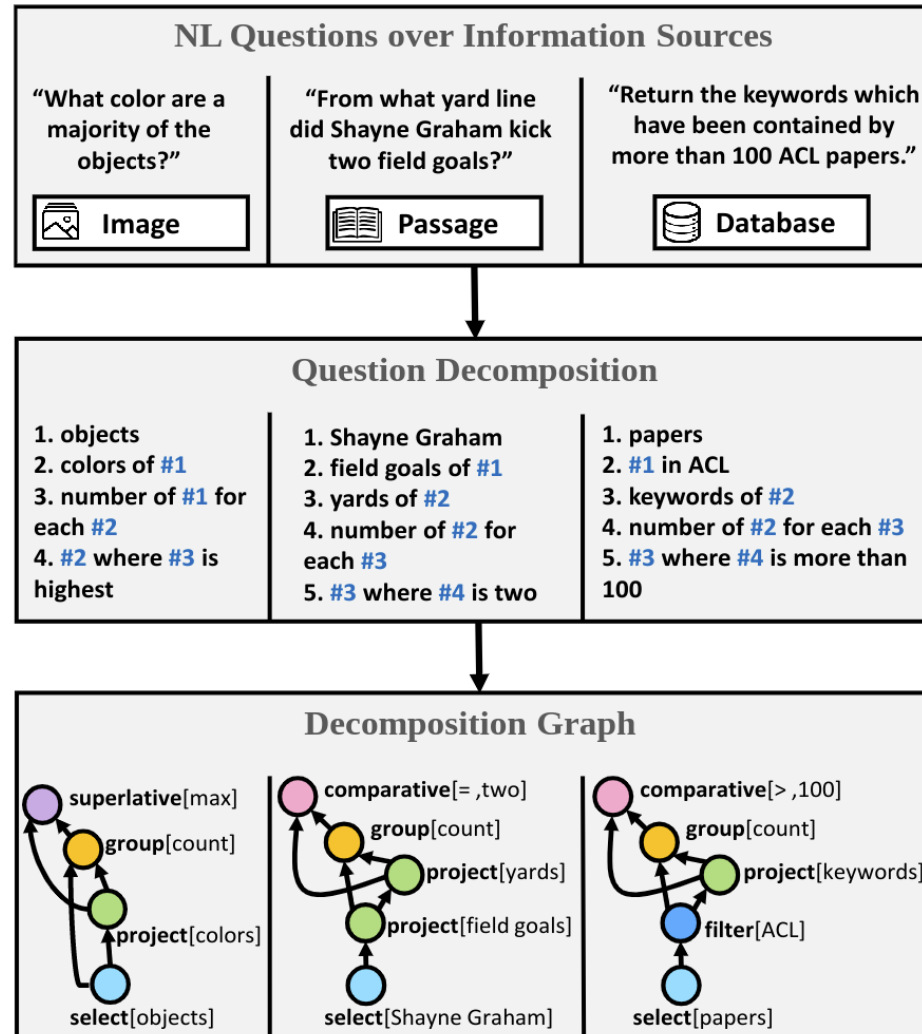
NLI\_Spanish

PREMISE: *Más de tres personas resultaron heridas en un accidente de **dos** vehículos el lunes por la noche. (translation: **More than three** people were injured in a **two-vehicle** crash Monday evening.)*

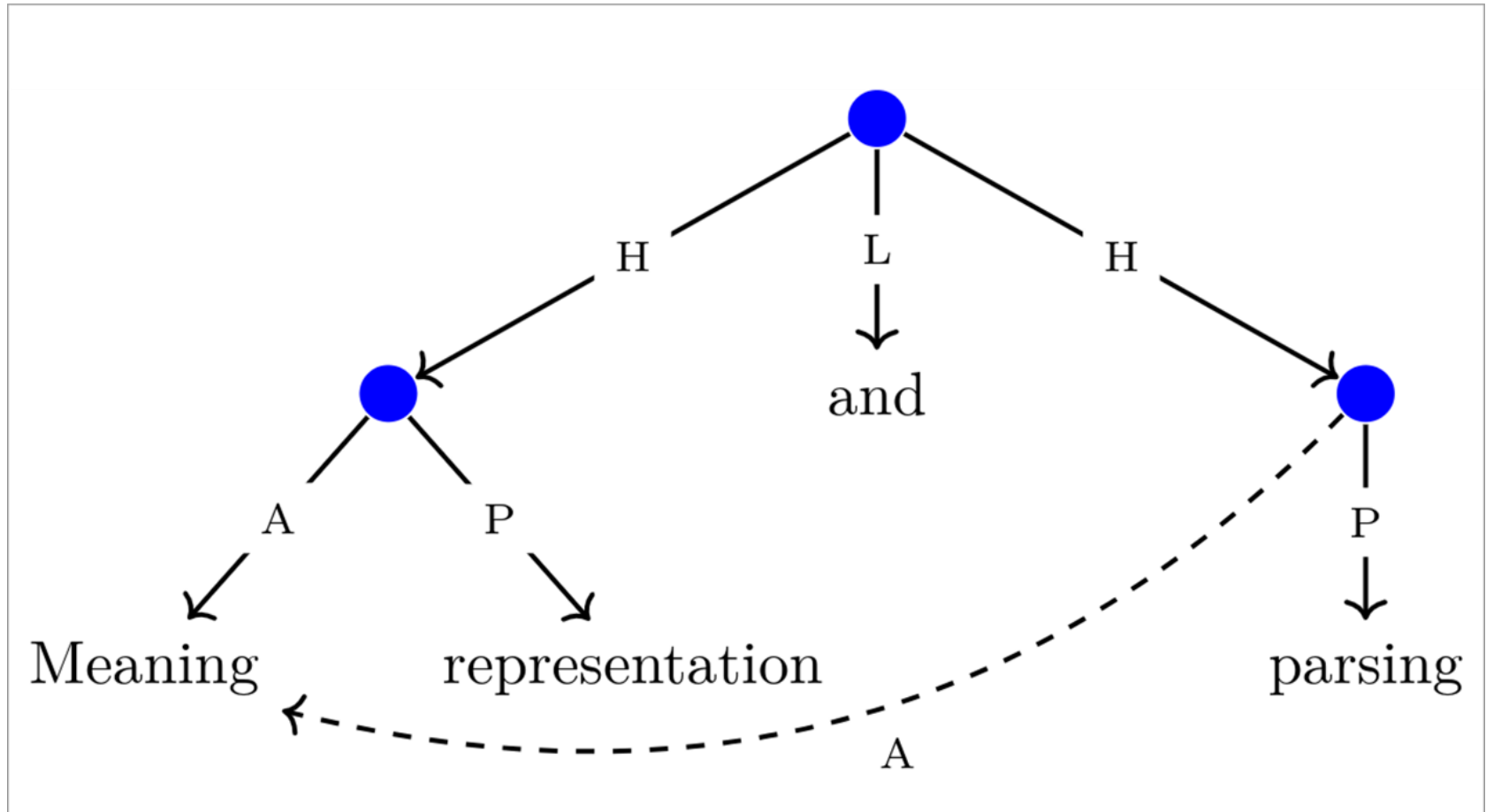
HYPOTHESIS: *Había **4** personas involucradas. (translation: There were **4** people involved.* LABEL: Neutral PREDICTED LABEL: Entailment



# Semantics for Interpretability



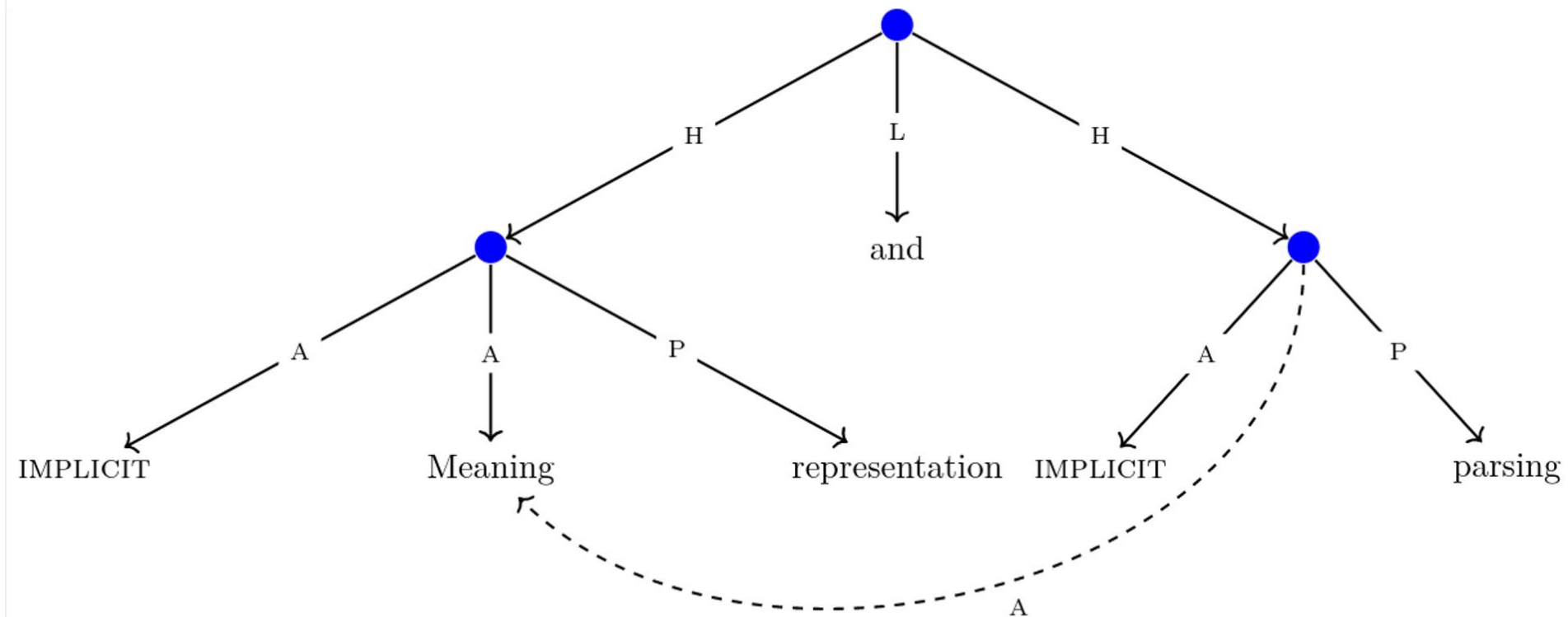
# UCCA



# Implicit Relations in UCCA



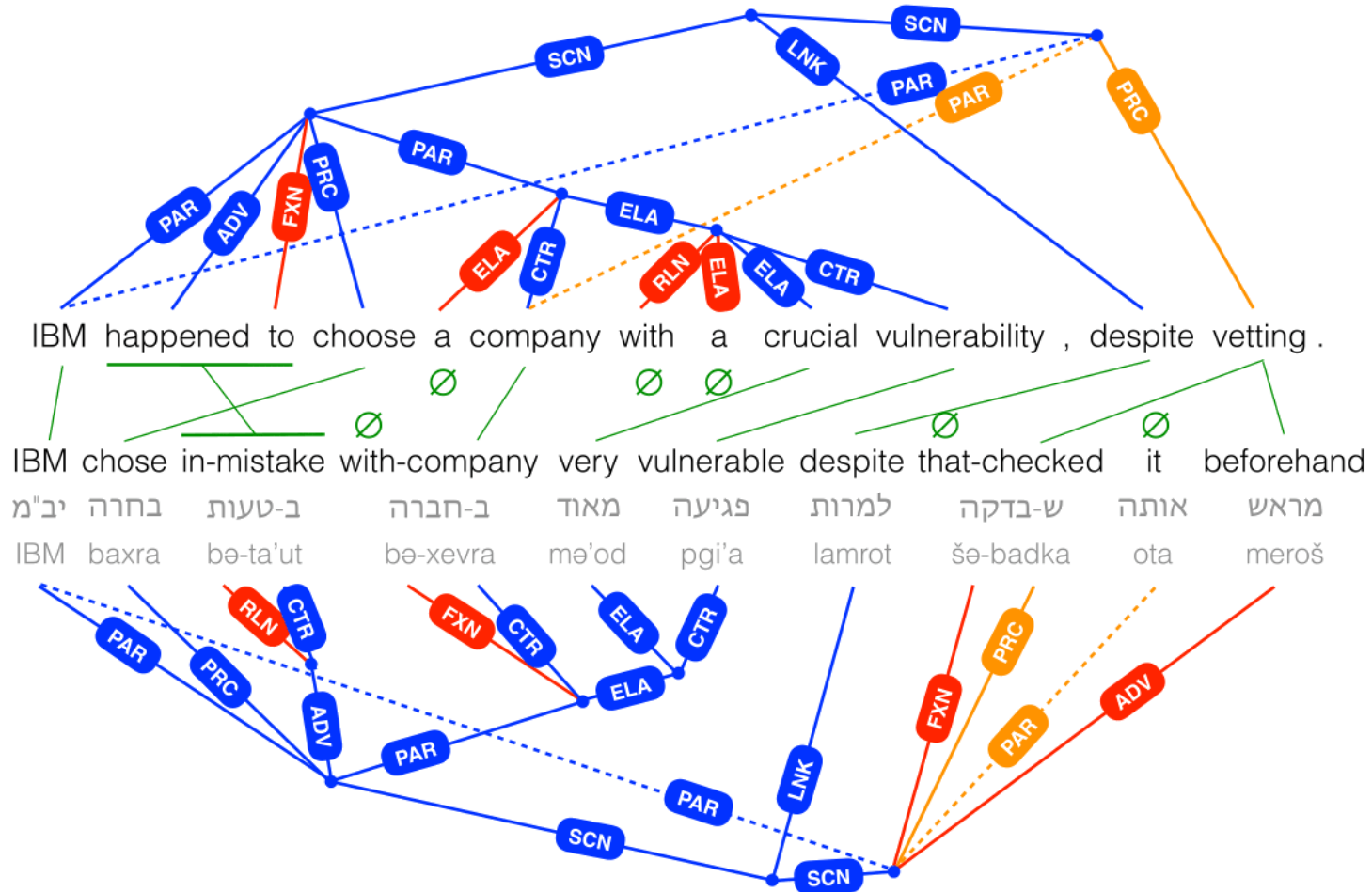
Ruixiang Cui



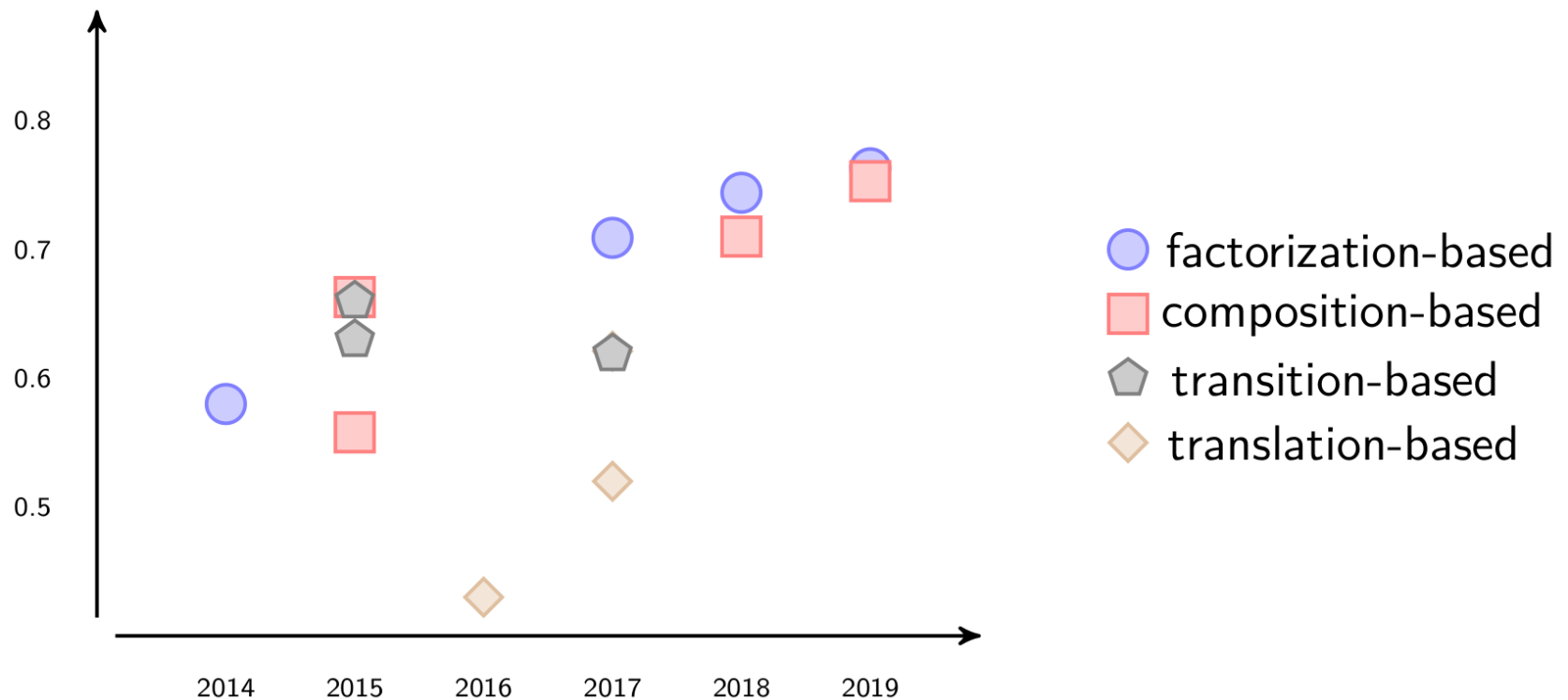
## Refining Implicit Argument Annotation for UCCA (Cui & Hershcovich, DMR 2020)

Great Service! Fine-grained Parsing of Implicit Arguments (Cui & Hershovich, IWPT 2021)

# UCCA

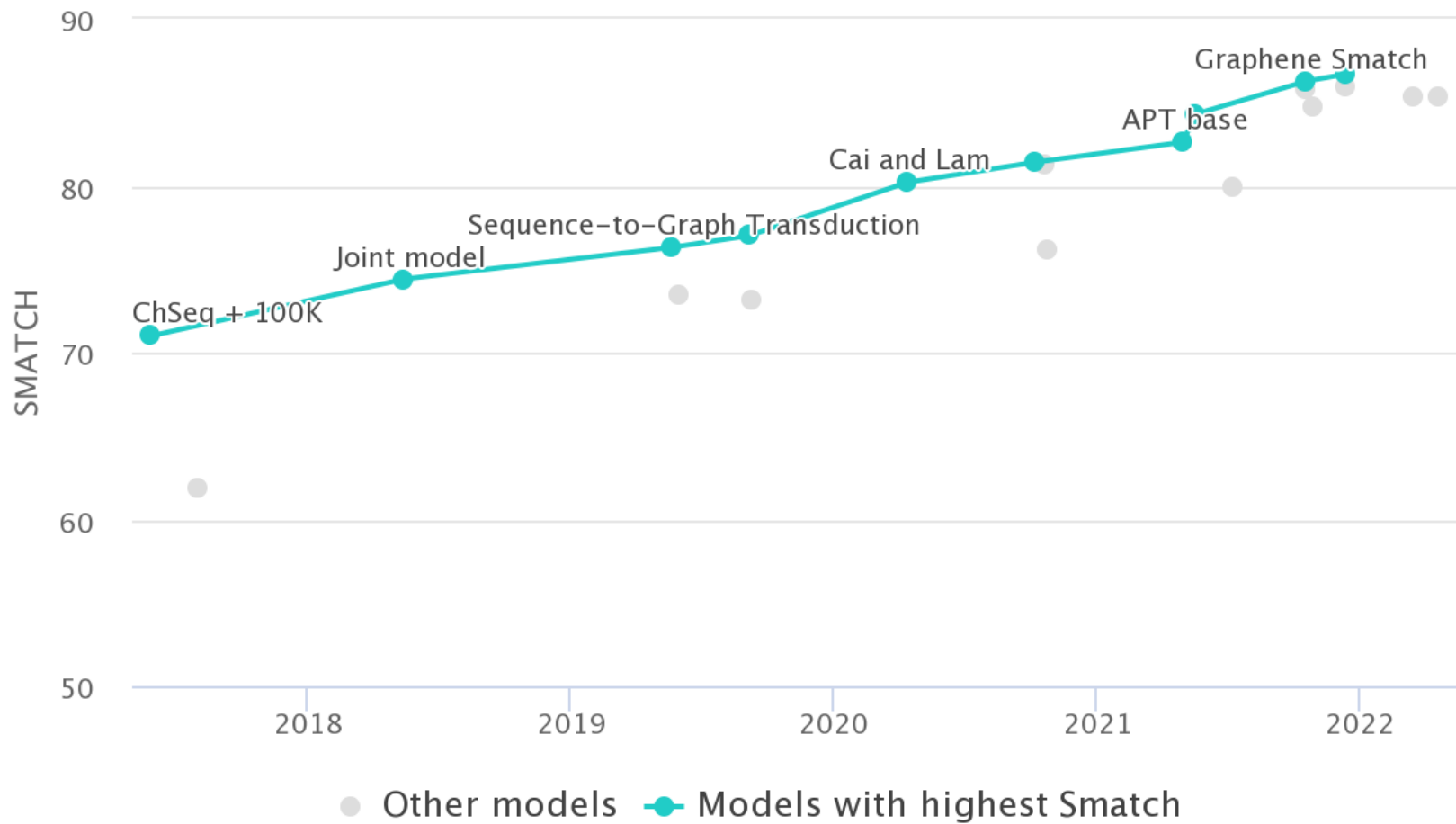


# Parsing Progress

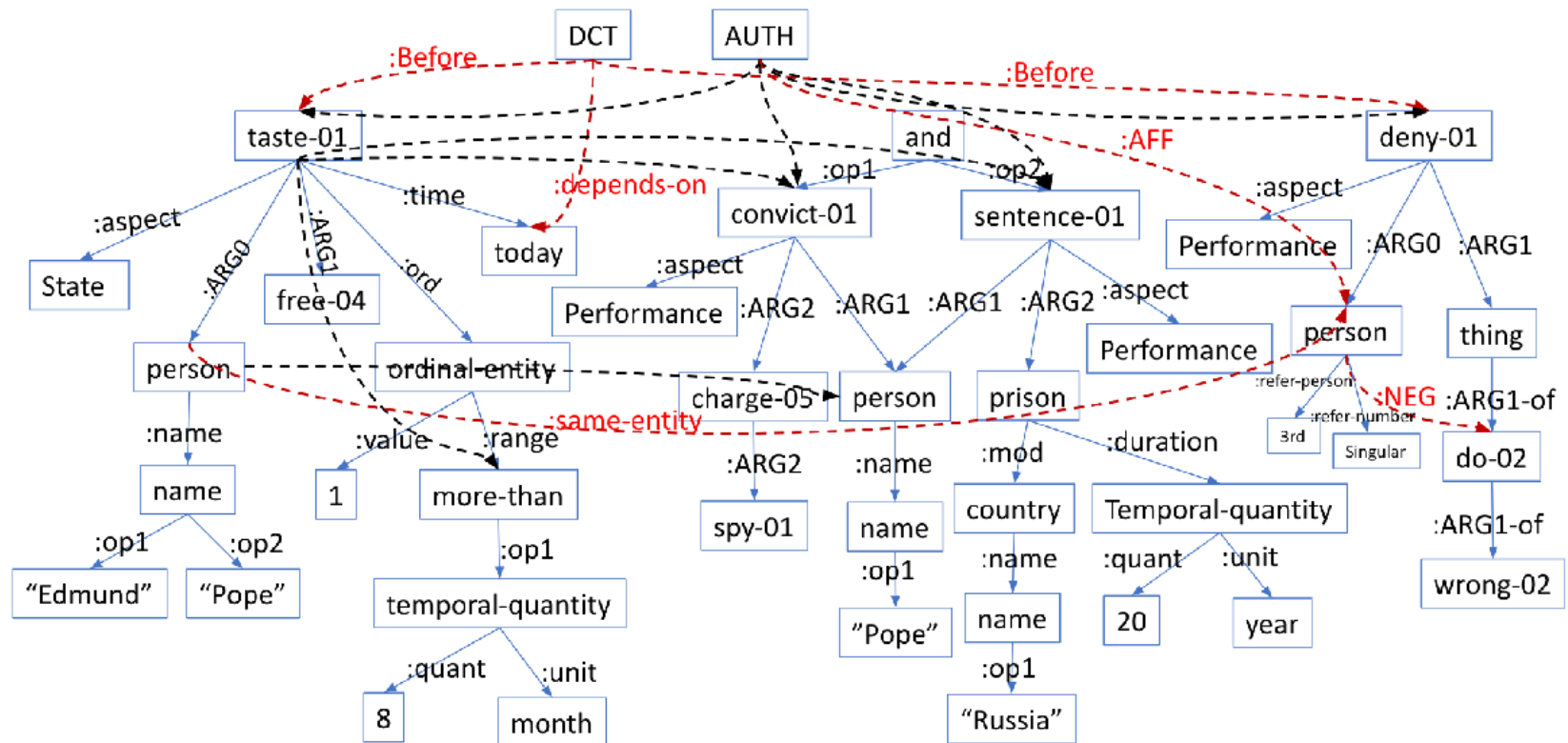


AMR 2015 Smatch F1

# Parsing Progress



# Uniform Meaning Representation



"Edmund Pope tasted freedom today for the first time in eight months."

"Pope was convicted on spying charges and sentenced to 20 years in a Russian prison."

"He denied any wrong-doing."

Designing a Uniform Meaning Representation for Natural Language Processing (Van Gysel et al., KI - Künstliche Intelligenz 2021)

UMR-Writer: A Web Application for Annotating Uniform Meaning Representations (Zhao et al., EMNLP 2021)