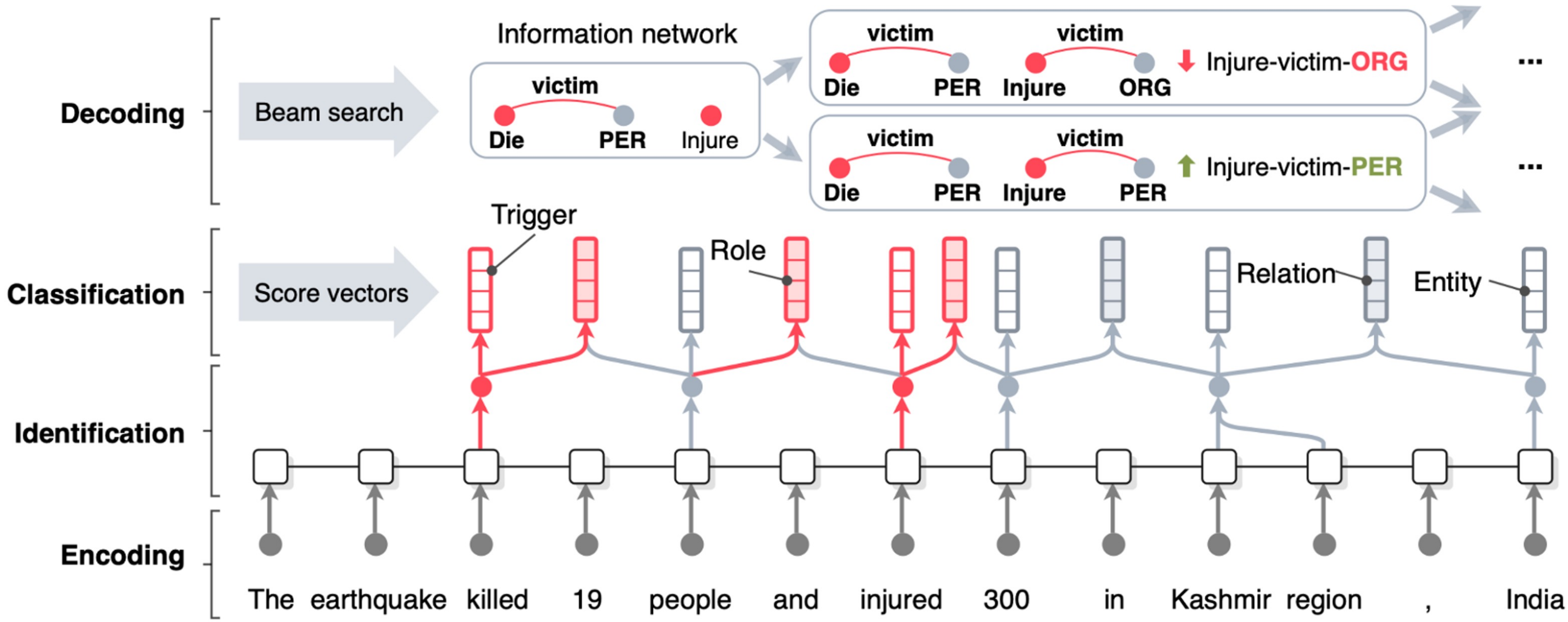
Meaning Representations for Natural Languages Tutorial Part 4

Applying Meaning Representations

Jeffrey Flanigan, Tim O’Gorman, Ishan Jindal, **Yunyao Li**, Nianwen Xue, Julia Bonn

Information Extraction



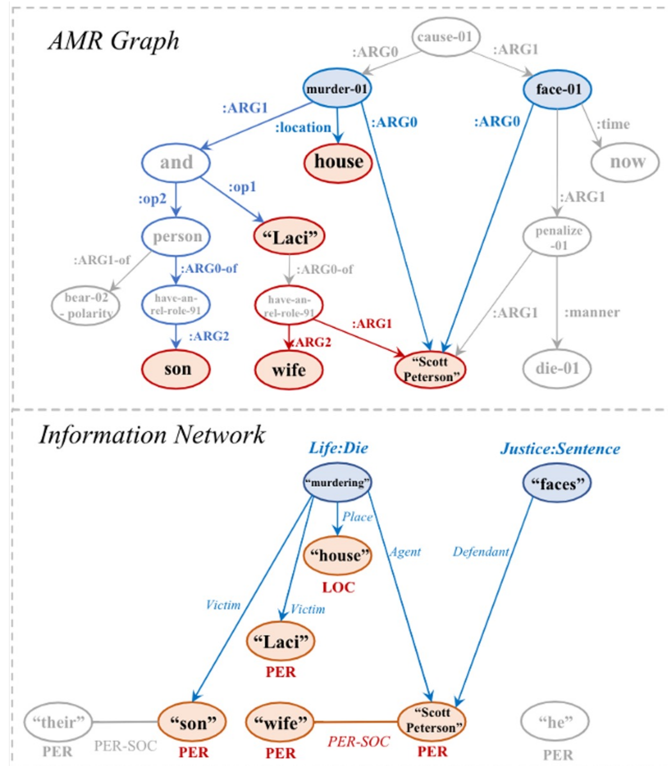
•OneIE [Lin et al., ACL2020] framework extracts the information graph from a given sentence in four steps: encoding, identification, classification, and decoding

Moving from Seq-to-Graph to Graph-to-Graph

- AMR converts input sentence into a directed and acyclic graph structure with fine-grained node and edge type labels

- AMR parsing shares inherent similarities with information network (IE output)
  - Similar node and edge semantics
  - Similar graph topology

- Semantic graphs can better capture non-local context in a sentence



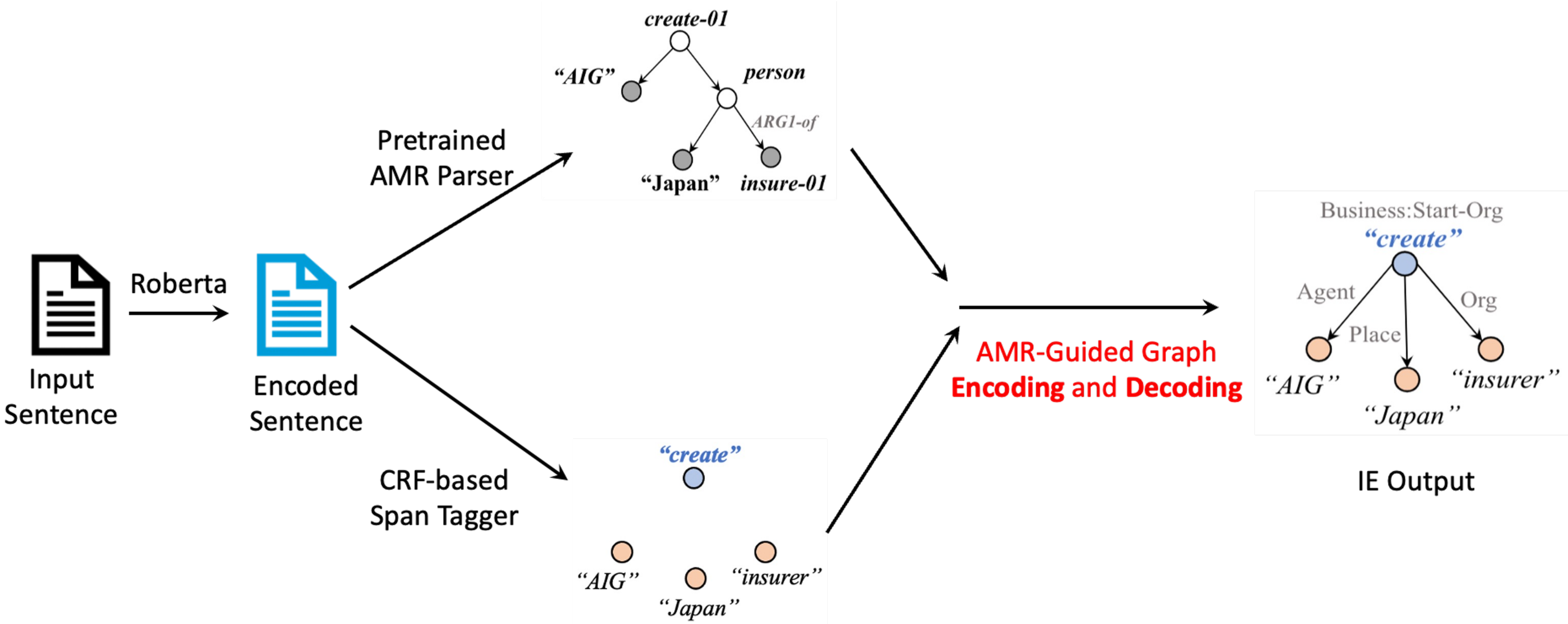
Key Idea:

Exploit the similarity between AMR and IE to for joint information extraction

Zixuan Zhang, Heng Ji. **AMR-IE: An AMR-guided encoding and decoding framework for IE.** NAACL'2021

Slide credit: Heng Ji

AMR-IE



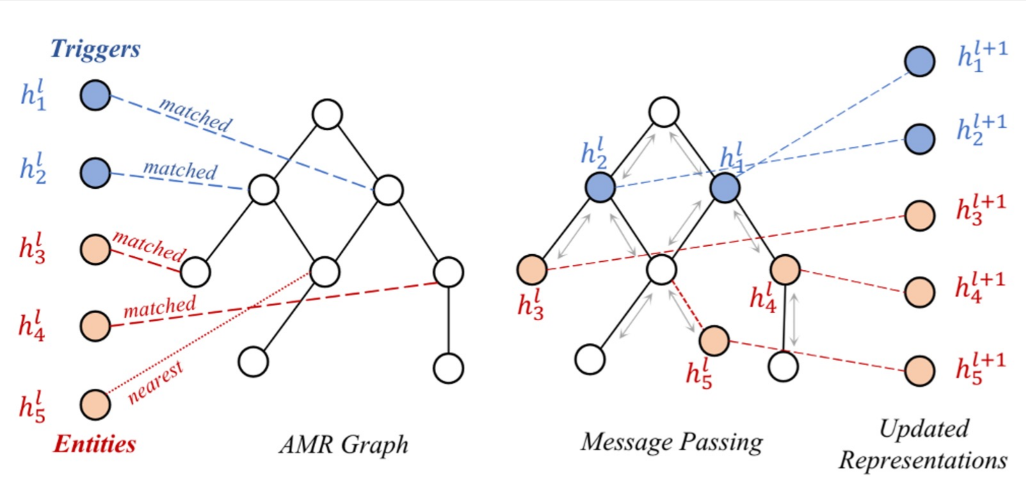
Zixuan Zhang, Heng Ji. **AMR-IE: An AMR-guided encoding and decoding framework for IE.** NAACL'2021

Slide credit: Heng Ji

AMR Guided Graph Encoding: Using an Edge-Conditioned GAT

- Map each candidate entity and event to AMR nodes.

- Update entity and event representations using an edge-conditioned GAT to incorporate information from AMR neighbors.



$$\alpha^l_{i,j} = \frac{\exp\left(\sigma\left(f^l[\mathbf{W}\boldsymbol{h}^l_i : \mathbf{W}_e \boldsymbol{e}_{i,j} : \mathbf{W}\boldsymbol{h}^l_j]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\sigma\left(f^l[\mathbf{W}\boldsymbol{h}^l_i : \mathbf{W}_e\boldsymbol{e}_{i,k} : \mathbf{W}\boldsymbol{h}^l_k]\right)\right)}$$

$$\boldsymbol{h}^* = \sum_{j\in\mathcal{N}_i} \alpha^l_{i,j}\boldsymbol{h}^l_i$$

$$\boldsymbol{h}^{l+1} = \boldsymbol{h}^l + \gamma \cdot \mathbf{W}^*\boldsymbol{h}^*$$

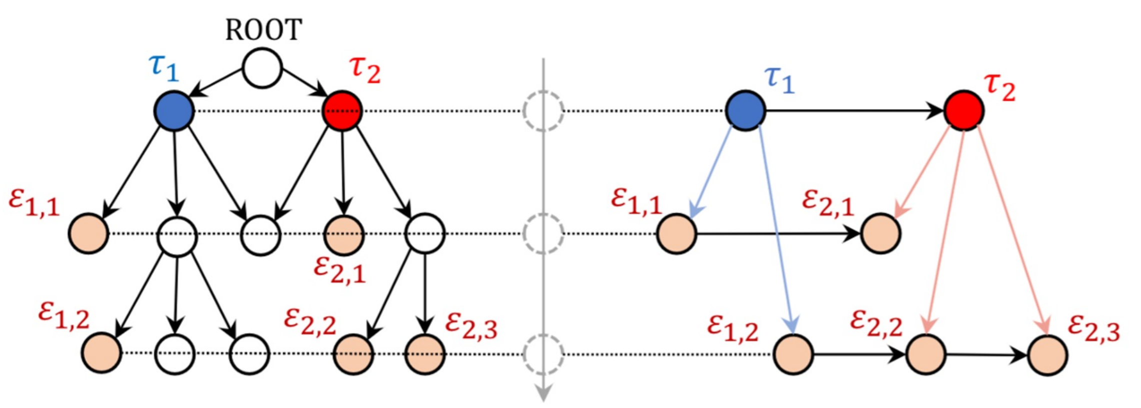
Zixuan Zhang, Heng Ji. **AMR-IE: An AMR-guided encoding and decoding framework for IE.** NAACL'2021

Slide credit: Heng Ji

AMR Guided Graph Decoding: Ordered decoding guided by AMR

- Beam search based decoding as in *OneIE* (Lin et al. 2020).
- The decoding order of candidate nodes are determined by the hierarchy in AMR in a top-to-down manner.
- E.g., the correct ordered decoding in the following graph is:

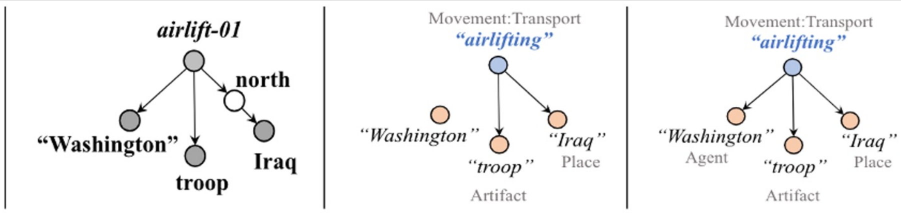
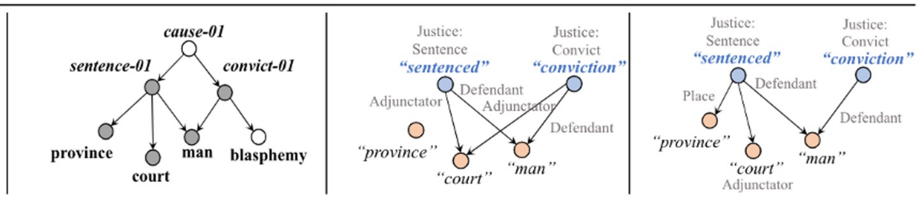
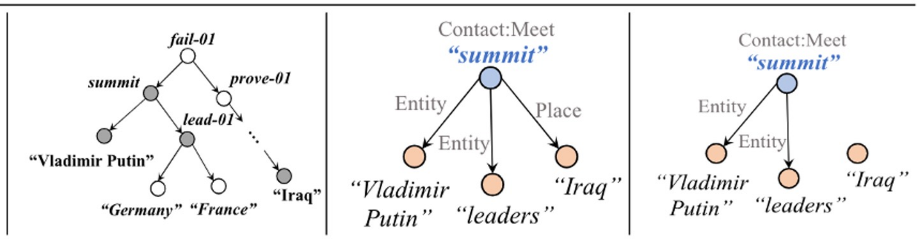
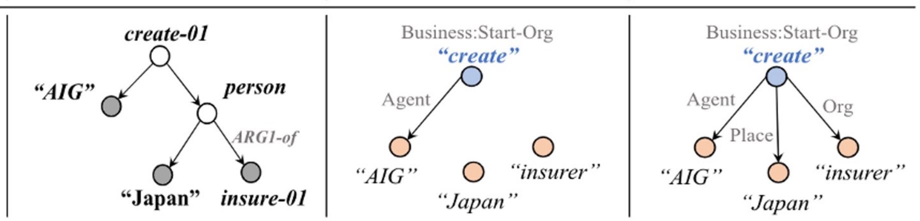
$$\tau_1, \tau_2, \varepsilon_{1,1}, \varepsilon_{2,1}, \varepsilon_{1,2}, \varepsilon_{2,2}, \varepsilon_{2,3}$$



Zixuan Zhang, Heng Ji. **AMR-IE: An AMR-guided encoding and decoding framework for IE.** NAACL'2021

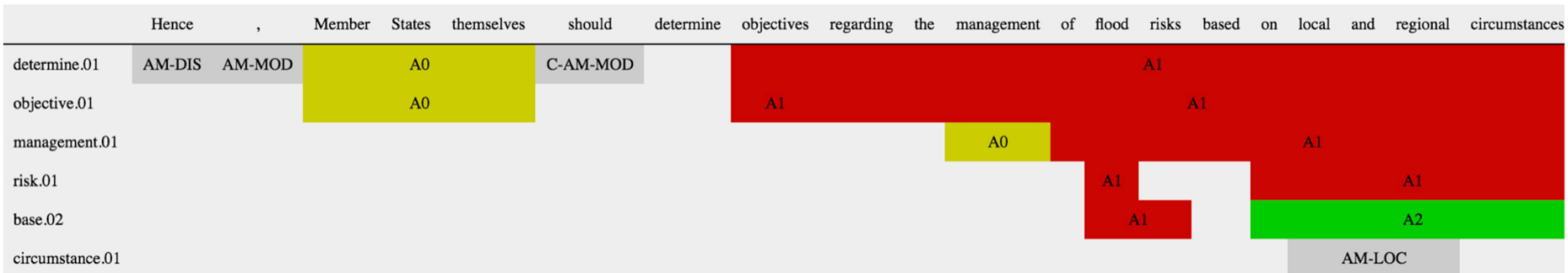
Slide credit: Heng Ji

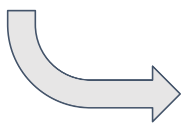
Examples on how AMR graphs help

Sentence	AMR Parsing	*OneIE outputs*	*AMR-IE outputs*
If the resolution is not passed, **Washington** would likely want to use the airspace for strikes against Iraq and for **airlifting** **troops** to northern **Iraq**.			
A Pakistani **court** in central Punjab **province** has **sentenced** a Christian **man** to life imprisonment for a blasphemy **conviction**, police said Sunday.			
Russian President **Vladimir Putin**'s **summit** with the **leaders** of Germany and France may have been a failure that proves there can be no long-term "peace camp" alliance following the end of war in **Iraq**.			
Major US insurance group **AIG** is in the final stage of talks to take over General Electric's Japanese life insurance arm in a deal to **create** **Japan**'s sixth largest life **insurer**, reports said Wednesday.			

Slide credit: Heng Ji

Leverage Meaning Representation for High-quality Rule-based IE





extraction rules

```
<Norm>
  <NormType>Obligation</NormType>
  <ActiveRole>Member States themselves
    </ActiveRole>
 <Action>objectives regarding the
 management of flood risk should be
 determined and should be based on local
 and regional circumstances</Action>
</Norm>
```

Llio Humphreys et al. **Populating Legal Ontologies using Semantic Role Labeling** LREC’20

Machine Translation

- MT methods using Transformers can make semantic errors
- Repeating words with same meaning

Src: It was noteworthy because of personal reasons , too .
Ref: Sie war auch aus persönlichen Gründen bemerkenswert .
Vanilla Transformer: Auch weil es aus persönlichen Gründen bemerkenswert war , war sie beachtenswert .

- Hallucinate information not contained in the source

Src: And these numbers hold up in early states .
Ref: Und diese Zahlen halten sich in frühen Staaten .
Vanilla Transformer: Und diese Zahlen sind in frühen Bundesstaaten verteilt .

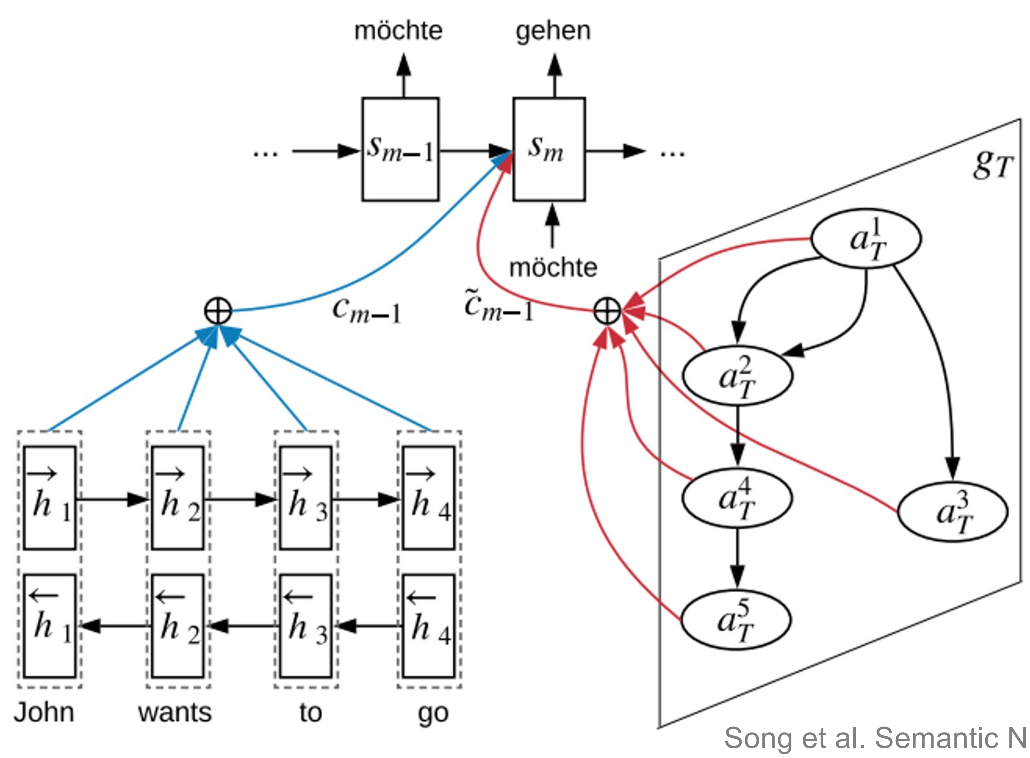
Machine Translation

This is mostly due to

Failing to accurately capture the semantics of the source in some cases.

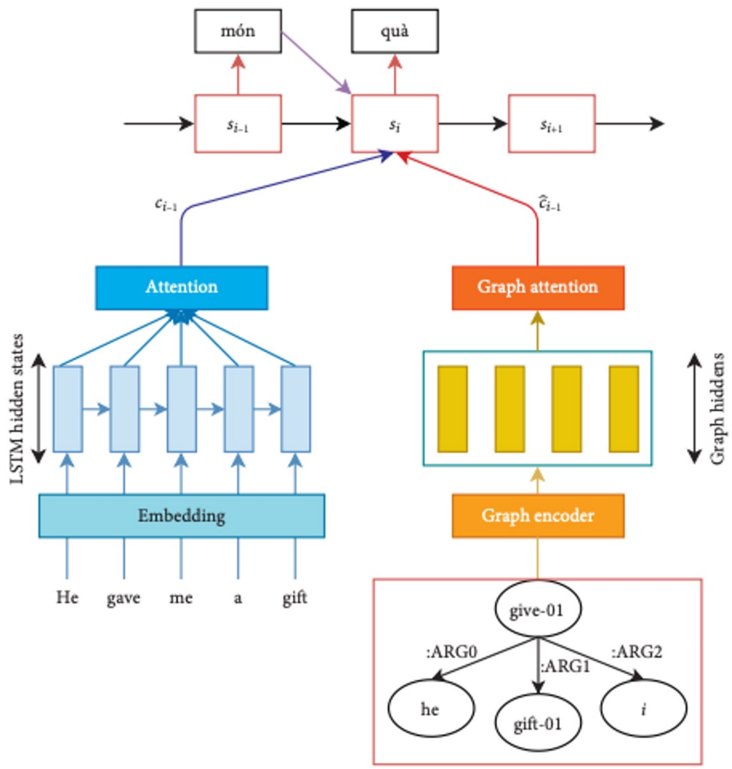
Goal: inject semantic information into Machine translation

Machine Translation



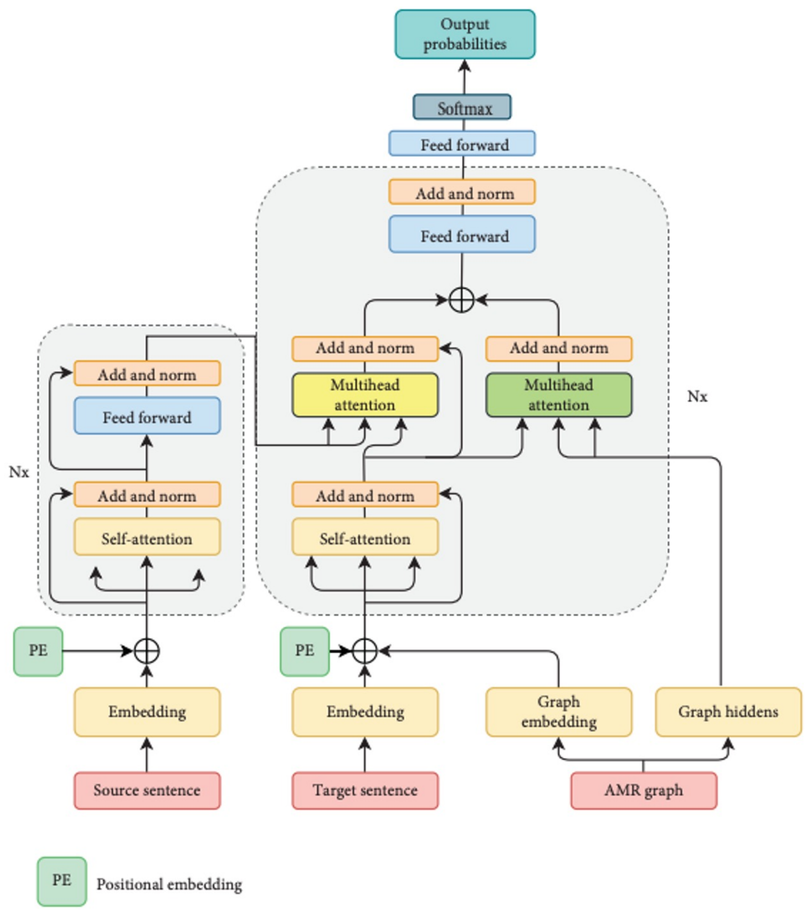
Song et al. Semantic Neural Machine Translation using AMR. TACL 2019.

Machine Translation



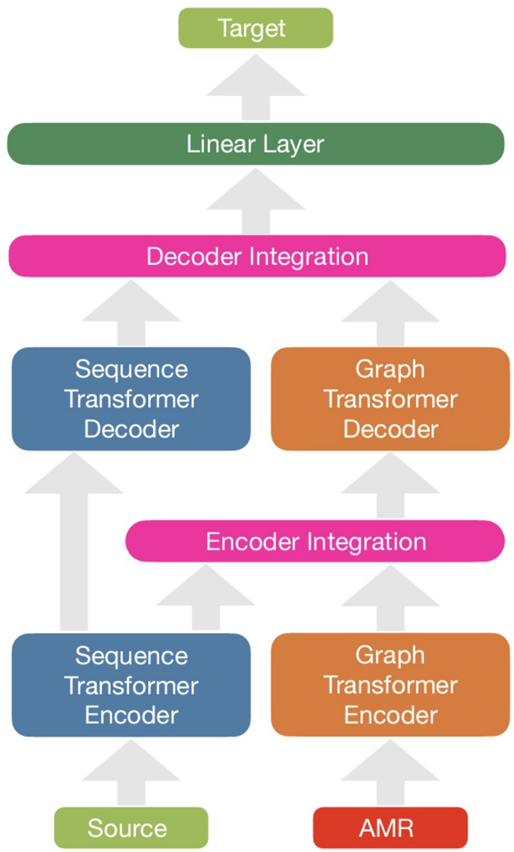
Nguyen et al. Improving Neural Machine Translation with AMR Semantic Graphs. Hindawi Mathematical Problems in Engineering 2021.

Machine Translation



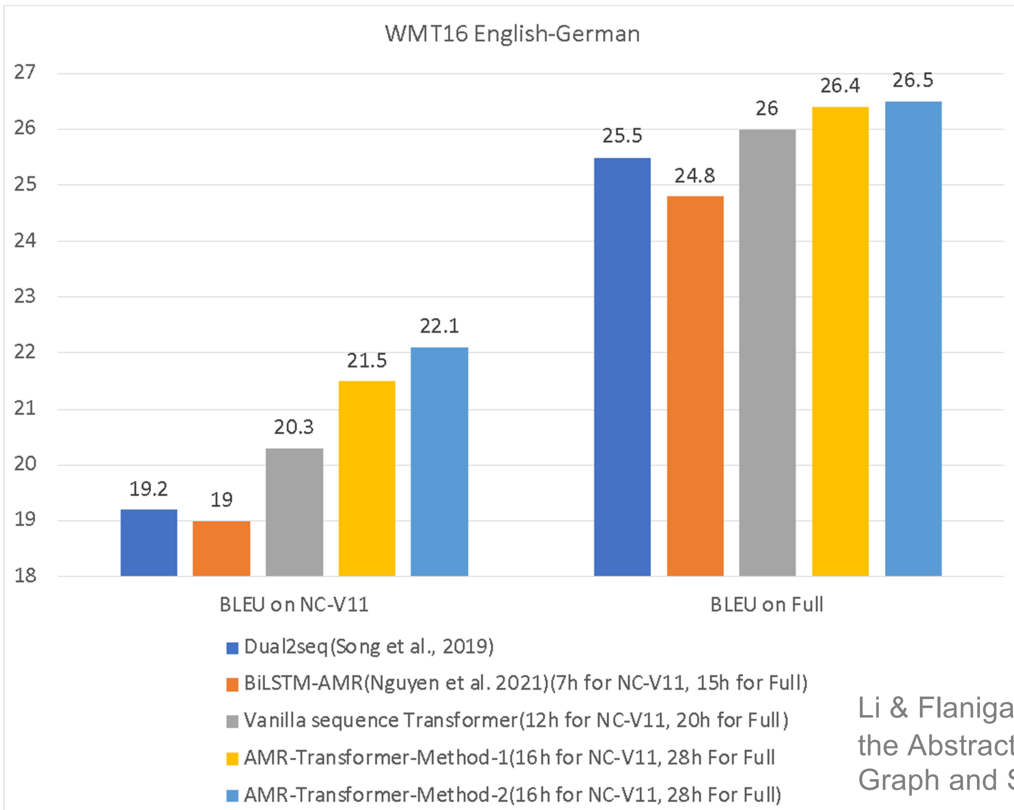
Nguyen et al. Improving Neural Machine Translation with AMR Semantic Graphs. Hindawi Mathematical Problems in Engineering 2021.

Machine Translation



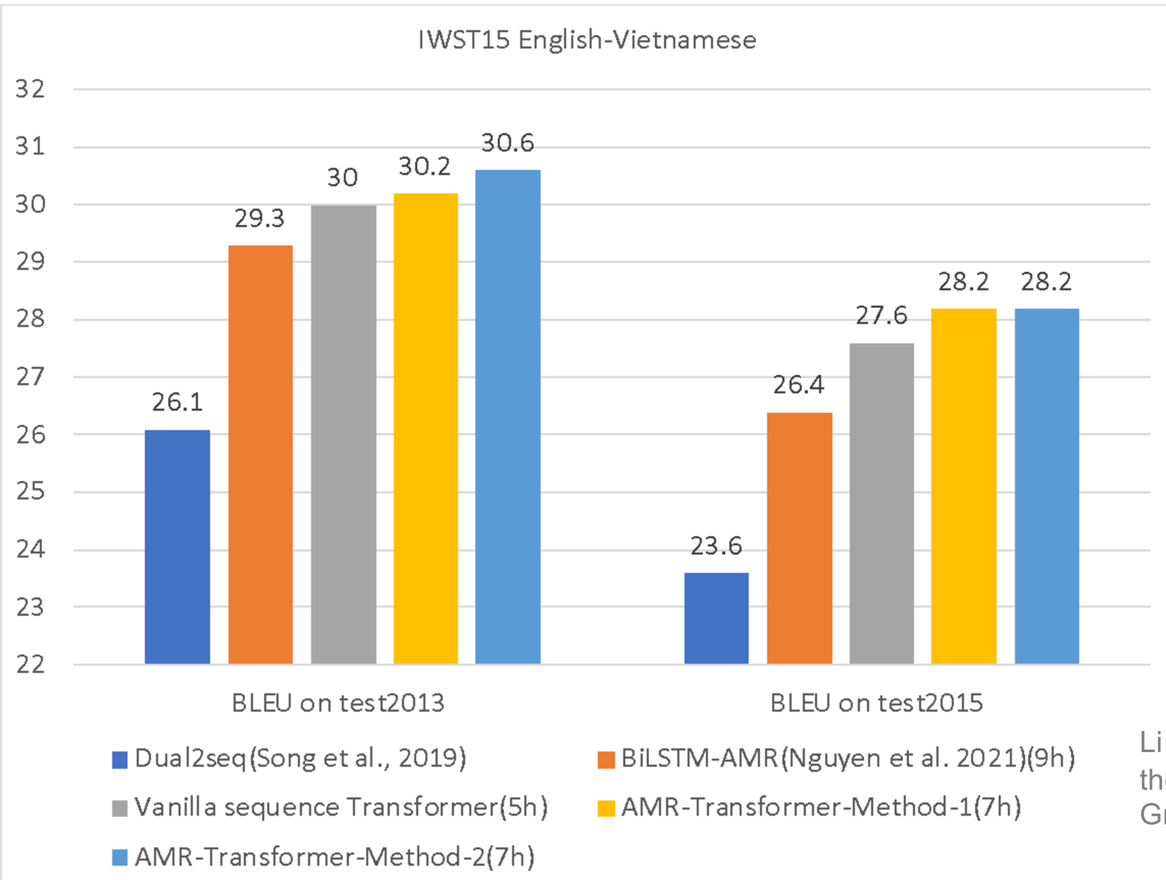
Li & Flanigan. Improving Neural Machine Translation with the Abstract Meaning Representation by Combining Graph and Sequence Transformers. DLG4NLP 2022.

Machine Translation



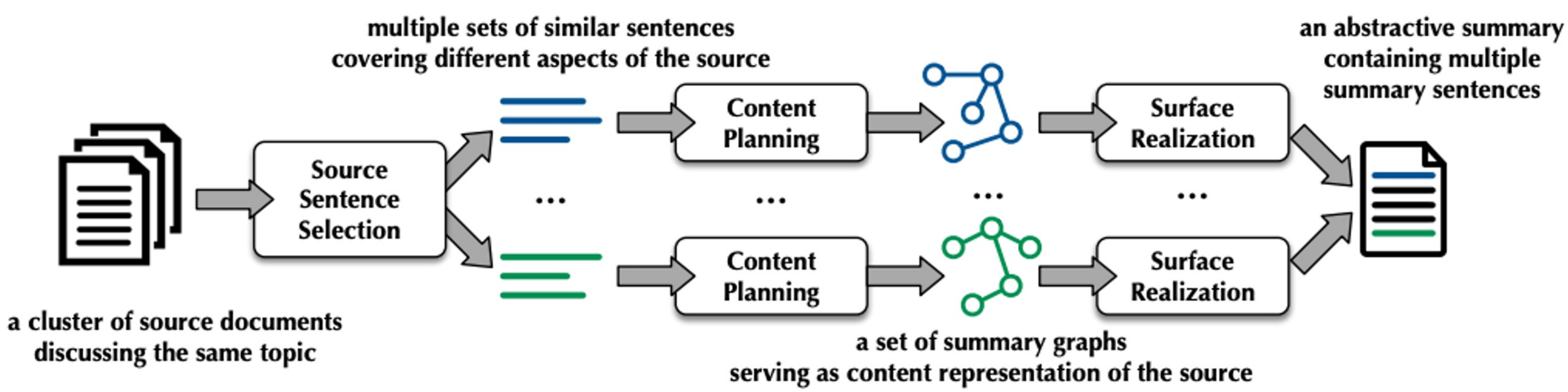
Li & Flanigan. Improving Neural Machine Translation with the Abstract Meaning Representation by Combining Graph and Sequence Transformers. DLG4NLP 2022.

Machine Translation



Li & Flanigan. Improving Neural Machine Translation with the Abstract Meaning Representation by Combining Graph and Sequence Transformers. DLG4NLP 2022.

Summarization



Liao et al. Abstract Meaning Representation for Multi-Document Summarization. ICCL 2018

Summarization

	System	ROUGE-1			ROUGE-2			ROUGE-SU4		
		P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
DUC 2004	*ext*-SumBasic	37.5	24.9	29.5	5.3	3.6	4.3	11.1	7.3	8.6
	ext-KL-Sum	31.1	31.1	31.0	6.0	6.1	6.0	10.2	10.3	10.2
	ext-LexRank	34.3	34.6	34.4	7.1	7.2	7.1	11.1	11.2	11.2
	abs-Opinosis	36.5	23.7	27.5	7.2	4.3	5.1	11.7	7.4	8.6
	abs-Pointer-Gen-all	37.5	20.9	26.5	8.0	4.4	**5.6**	12.3	6.7	8.5
	abs-Pointer-Gen	33.2	21.5	25.6	5.8	3.8	4.5	10.3	6.6	7.9
	abs-AMRSumm-Clst	29.9	30.5	**30.2**	4.1	4.2	4.1	8.7	8.9	**8.8**
	abs-AMRSumm-VSM	36.7	39.0	**37.8**	6.5	6.9	**6.6**	11.4	12.2	**11.8**
TAC 2011	*ext*-SumBasic	37.3	28.2	31.6	6.9	5.5	6.1	11.8	9.0	10.1
	ext-KL-Sum	31.2	31.4	31.2	7.1	7.1	7.1	10.5	10.6	10.6
	ext-LexRank	32.9	33.3	33.1	7.4	7.6	7.5	11.1	11.2	11.1
	abs-Opinosis	38.0	20.4	25.2	8.6	4.0	5.1	12.9	6.5	8.1
	abs-Pointer-Gen-all	37.3	22.2	27.6	7.8	4.6	**5.8**	12.2	7.1	8.9
	abs-Pointer-Gen	34.4	21.6	26.2	6.9	4.4	5.3	10.9	6.8	8.2
	abs-AMRSumm-Clst	32.2	31.7	**31.9**	4.7	4.7	4.7	9.8	9.7	**9.7**
	abs-AMRSumm-VSM	40.1	42.3	**41.1**	8.1	**8.5**	**8.3**	13.1	13.9	**13.5**

Table 2: Summarization results on DUC-04 and TAC-11 datasets. We compare the AMR summarization framework (AMRSumm-*) with both extractive (*ext*-*) and abstractive (*abs*-*) summarization systems.

Liao et al. Abstract Meaning Representation for Multi-Document Summarization. ICCL 2018

Natural Language Inference

Does premise *P* justify an inference to hypothesis *H*?

P: The information from the actor stopped the banker.
H: The banker stopped the actor.

Natural Language Inference

Does premise *P* justify an inference to hypothesis *H*?

P: The information from the actor stopped the banker.
H: The banker stopped the actor.

shallow heuristics
due to dataset biases
(e.g. lexicon overlap)

low generalization
on out-of-distribution
evaluation sets.

The HANS challenge dataset [McCoy et al., 2019] showed that NLI models trained on MNLI or SNLI datasets get fooled easily by heuristics when the input sentence pairs have high lexical similarity.

How Can Meaning Representation Help?

Semantic information(SRL)

- Improve the semantic knowledge of the NLI models
- Less prone to dataset biases.

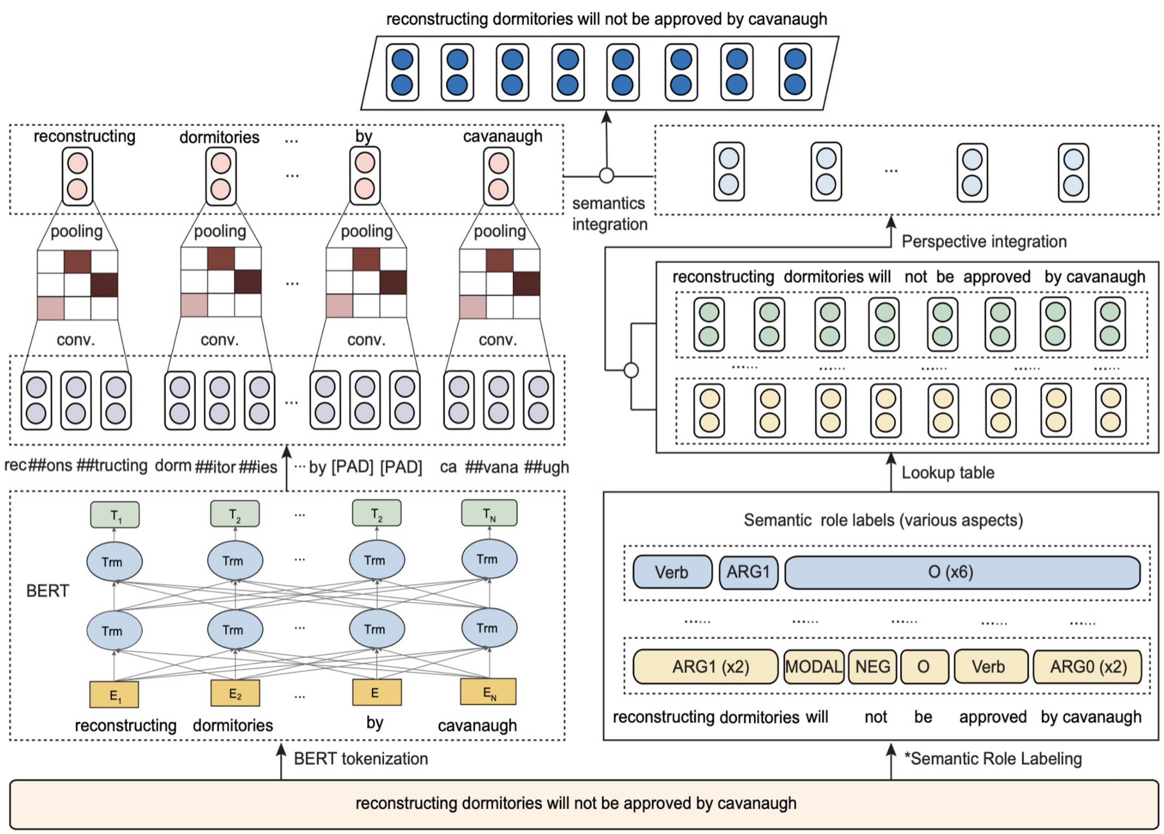
P: The information from the actor stopped the banker.

ARG0 VERB ARG1

H: The banker stopped the actor.

ARG0 VERB ARG1

SemBERT: Semantic Aware BERT



Zhuosheng Zhang, Yuwei Wu, Hai Zhao, Zuchao Li, Shuailiang Zhang, Xi Zhou, Xiang Zhou: **Semantics-Aware BERT for Language Understanding**. AAAI 2020

Incorporate SRL information with BERT representations.

SemBERT: Semantic Aware BERT

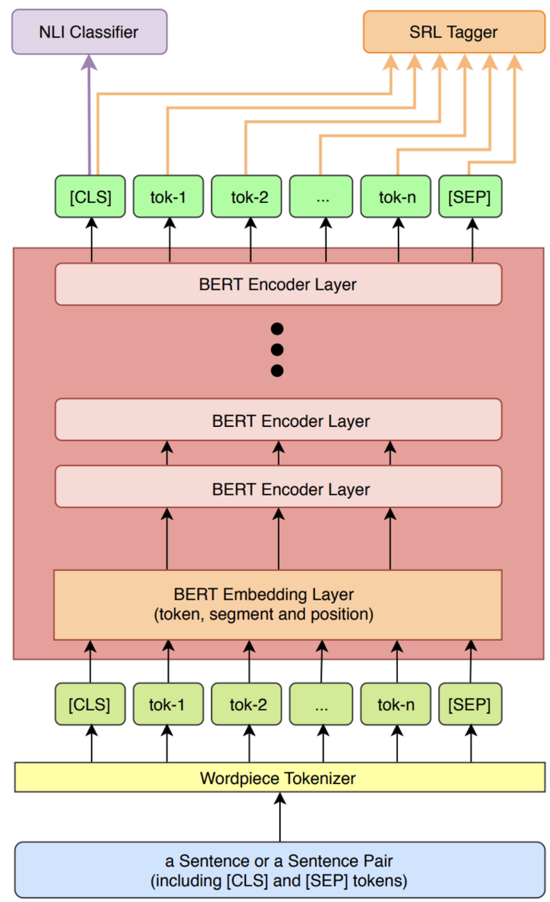
Method	Classification		Natural Language Inference			Semantic Similarity			Score
	CoLA (mc)	SST-2 (acc)	MNLI m/mm(acc)	QNLI (acc)	RTE (acc)	MRPC (F1)	QQP (F1)	STS-B (pc)	-
Leaderboard (September, 2019)									
ALBERT	69.1	97.1	91.3/91.0	99.2	89.2	93.4	74.2	92.5	89.4
RoBERTa	67.8	96.7	90.8/90.2	98.9	88.2	92.1	90.2	92.2	88.5
XLNET	67.8	96.8	90.2/89.8	98.6	86.3	93.0	90.3	91.6	88.4
In literature (April, 2019)									
BiLSTM+ELMo+Attn	36.0	90.4	76.4/76.1	79.9	56.8	84.9	64.8	75.1	70.5
GPT	45.4	91.3	82.1/81.4	88.1	56.0	82.3	70.3	82.0	72.8
GPT on STILTs	47.2	93.1	80.8/80.6	87.2	69.1	87.7	70.1	85.3	76.9
MT-DNN	61.5	95.6	86.7/86.0	-	75.5	90.0	72.4	88.3	82.2
$\text{BERT}_{\text{BASE}}$	52.1	93.5	84.6/83.4	-	66.4	88.9	71.2	87.1	78.3
$\text{BERT}_{\text{LARGE}}$	60.5	94.9	86.7/85.9	92.7	70.1	89.3	72.1	87.6	80.5
Our implementation									
$\text{SemBERT}_{\text{BASE}}$	57.8	93.5	84.4/84.0	90.9	69.3	88.2	71.8	87.3	80.9
$\text{SemBERT}_{\text{LARGE}}$	$\boxed{62.3}$	94.6	87.6/86.3	94.6	$\boxed{84.5}$	$\boxed{91.2}$	72.8	87.8	82.9

Zhuosheng Zhang, Yuwei Wu, Hai Zhao, Zuchao Li, Shuailiang Zhang, Xi Zhou, Xiang Zhou:
Semantics-Aware BERT for Language Understanding. AAAI 2020

Results on GLUE benchmark
Works particularly well for smaller dataset

Joint Training with SRL improves NLI generalization

Main idea: Improve sentence understanding (hence out-of-distribution generalization) with joint learning of explicit semantics



Cemil Cengiz, Deniz Yuret. **Joint Training with Semantic Role Labeling for Better Generalization in Natural Language Inference.** Rep4NLP'2020

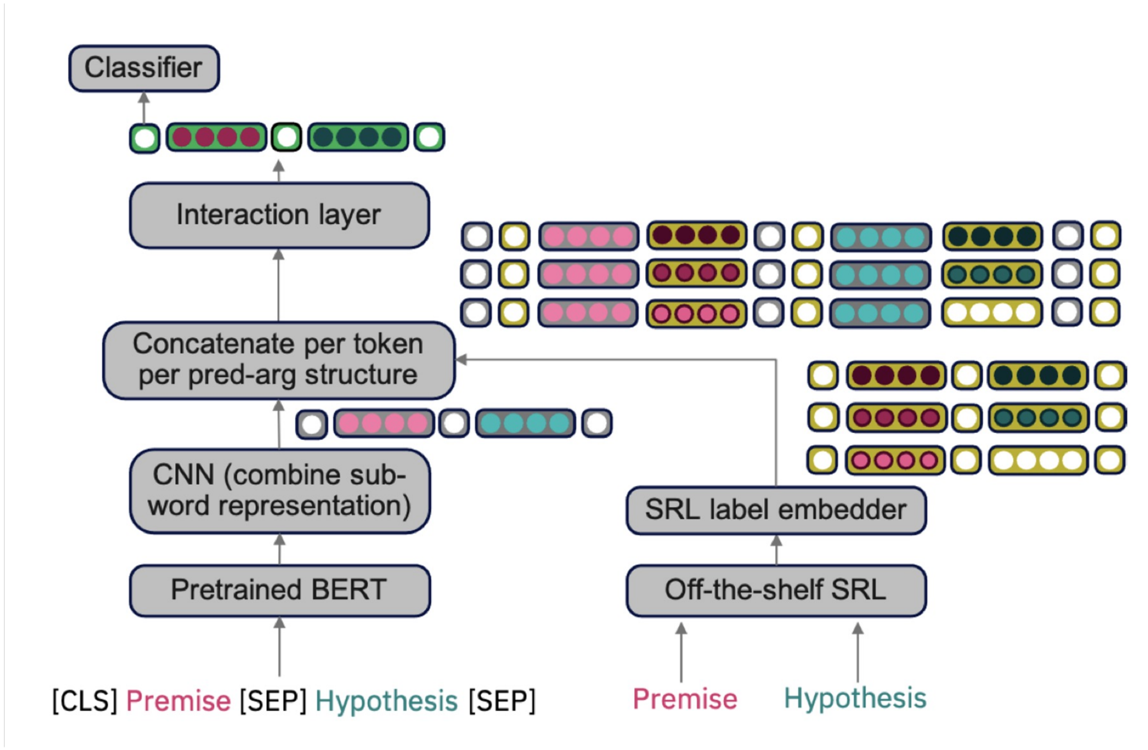
Joint Training with SRL improves NLI generalization

Main idea: Improve sentence understanding (hence out-of-distribution generalization) with joint learning of explicit semantics

	Training set: SNLI				Training set: MultiNLI			
BERT Model	**same**	**more/less**	**not**	**Avg.**	**same**	**more/less**	**not**	**Avg.**
Single-task	85.3	47.9	44.5	59.2	74.1	88.3	74.3	78.9
Multi-task	80.5	47.9	51.3	59.9	63.3	97.3	91.9	84.2

Cemil Cengiz, Deniz Yuret. **Joint Training with Semantic Role Labeling for Better Generalization in Natural Language Inference.**
Rep4NLP'2020

Is Semantic-Aware BERT More Linguistically Aware?



Infuse semantic knowledge via predicate-wise concatenation with BERT

Ling Liu, Ishan Jindal, Yunyao Li. **Is Semantic-aware BERT more Linguistically Aware? A Case Study on Natural Language Inference.** SUKI'2022

Is Semantic-Aware BERT More Linguistically Aware

Model	External knowledge	SNLI test	HANS	Breaking NLI
SNLI fine-tuned				
$\text{BERT}_{\text{Base}}$	-	90.30	58.83	93.84
(Pang et al., 2019)	SynParse	90.50	53.20	-
(Zhang et al., 2019a)	Semantic	89.60	-	-
(Kapanipathi et al., 2020)	KG	85.97	-	-
$\text{SemBERT}_{\text{Base}}$	Semantic	90.59*	57.89	93.16
$\text{LingBERT}_{\text{Base}}$	Semantic	**90.92**	**59.96**	**94.04**

Ling Liu, Ishan Jindal, Yunyao Li. **Is Semantic-aware BERT more Linguistically Aware? A Case Study on Natural Language Inference.** SUKI'2022

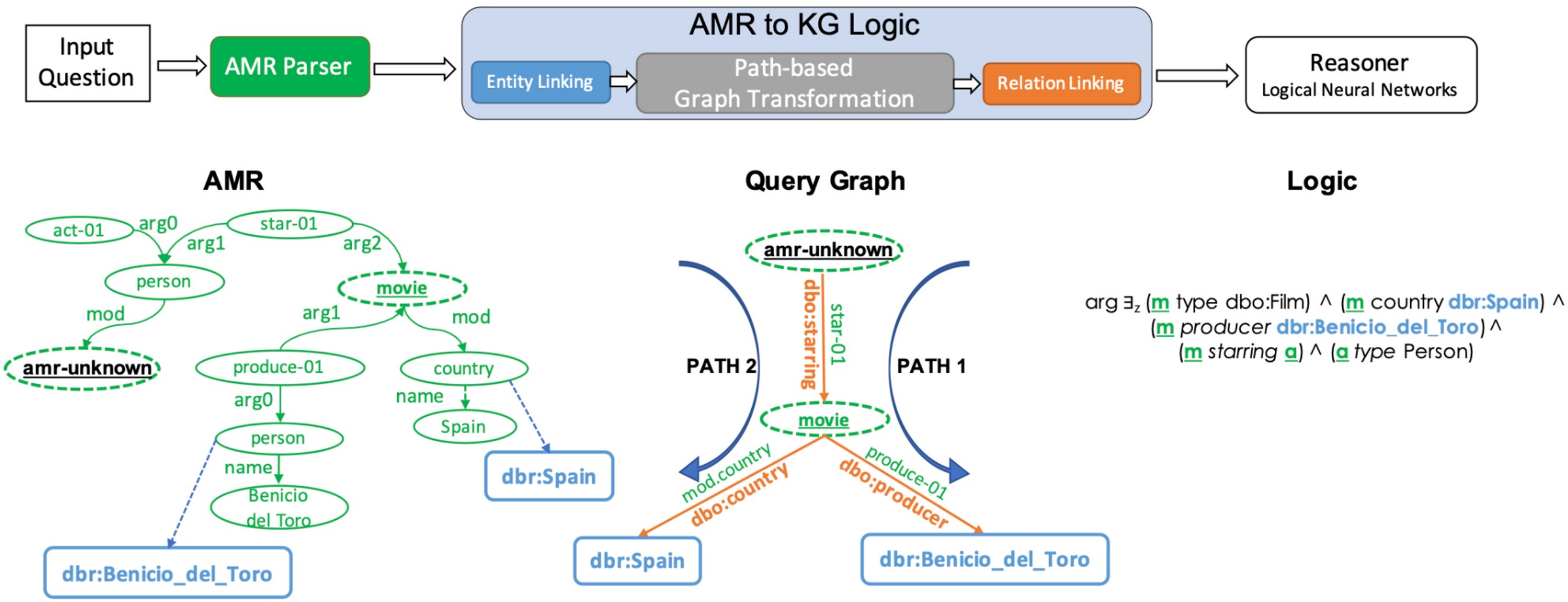
HANS Heuristics	*non-entailment* Examples	BERT	SemBERT	LingBERT
Lexical Overlap Heuristic		46.33	43.02	**54.40**
ln_conjunction	P_1: The authors recognized the president and the judges . H_1: The judges recognized the president .	40.93	33.57	**50.63**
ln_passive	P_2: The lawyers were recommended by the doctor . H_2: The lawyers recommended the doctor .	17.90	30.77	**33.00**
ln_preposition	P_3: The senators behind the lawyer contacted the student . H_3: The student contacted the senators .	58.37	49.20	**60.37**
ln_relative_clause	P_4: The student who the senators thanked stopped the scientist . H_4: The scientist stopped the student .	46.67	40.20	**52.63**
ln_subject/object_swap	P_5: The student saw the managers . H_5: The managers saw the student .	67.77	61.37	**75.37**
Subsequence Heuristic		**4.92**	3.69	4.01
sn_NP/S	*P_1: The author heard the presidents recommended the secretary .* *H_1: The author heard the presidents .*	**0.70**	0.03	0.53
sn_NP/Z	*P_2: Although the managers hid the actors saw the athlete .* *H_2: The managers hid the actors .*	**9.67**	6.43	5.27
sn_PP_on_subject	*P_3: The student near the secretaries supported the judges .* *H_3: The secretaries supported the judges .*	**9.03**	6.9	7.83
sn_past_participle	P_4: The artist avoided the author paid in the laboratory . H_4: The author paid in the laboratory .	**0.80**	0.27	**0.80**
sn_relative_clause_on_subject	P_5: The scientists that introduced the senator avoided the actor . H_5: The senator avoided the actor .	4.40	4.83	**5.63**
Constituent Heuristic		**5.2**	2.44	3.03
cn_adverb	*P_1: **Hopefully** the presidents introduced the doctors .* *H_1: The presidents introduced the doctors .*	**0.20**	0.00	0.00
cn_after_if_clause	*P_2: **Unless** the professor slept , the tourist saw the scientist.* *H_2: The tourist saw the scientist .*	0.00	0.00	0.00
cn_disjunction	*P_3: The actor recommended the lawyers , **or** the managers stopped the author .* *H_3: The actor recommended the lawyers .*	**0.33**	0.03	0.00
cn_embedded_under_if	*P_4: **If** the doctors mentioned the judge , the president thanked the student .* *H_4: The doctors mentioned the judge .*	**25.3**	12.2	15.1
cn_embedded_under_verb	*P_5: The lawyers **believed** that the tourists shouted .* *H_5: The tourists shouted .*	**0.13**	0.00	0.00

Performance on HANS *non-entailment* examples by models fine-tuned on *SNLI*. Examples in black and normal font are where BERT made wrong predictions and LingBERT made correct predictions. Examples in *blue and italics* are where none of the three models made the correct prediction. The last three columns are the accuracy in % on the *non-entailment* examples by BERT, SemBERT, and LingBERT respectively.

Better differentiate lexical similarity from world knowledge

Fails to help with subsequence /constituent heuristics

NSQA: AMR for Neural-Symbolic Question Answering over Knowledge Graph



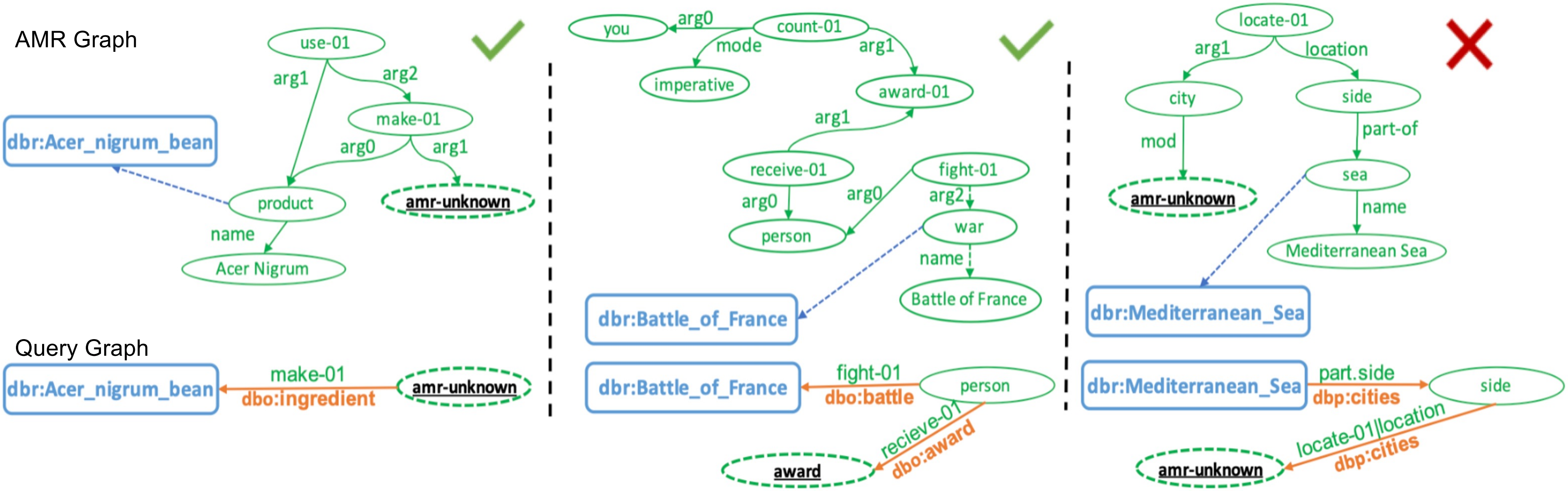
SPARQL

```
SELECT DISTINCT ?actor WHERE {
  ?movie  rdf:type      dbo:Film .
  ?movie  dbo:country   dbr:Spain .
  ?movie  dbo:producer  dbr:Benicio_del_Toro .
  ?movie  dbo:starring  ?actor .
}
```

Which **actors** starred in Spanish **movies** produced by Benicio del Toro?

Pavan Kapanipathi et al∗ **Leveraging Abstract Meaning Representation for Knowledge Base Question Answering**, ACL '2021

AMR Graph → Query Graph



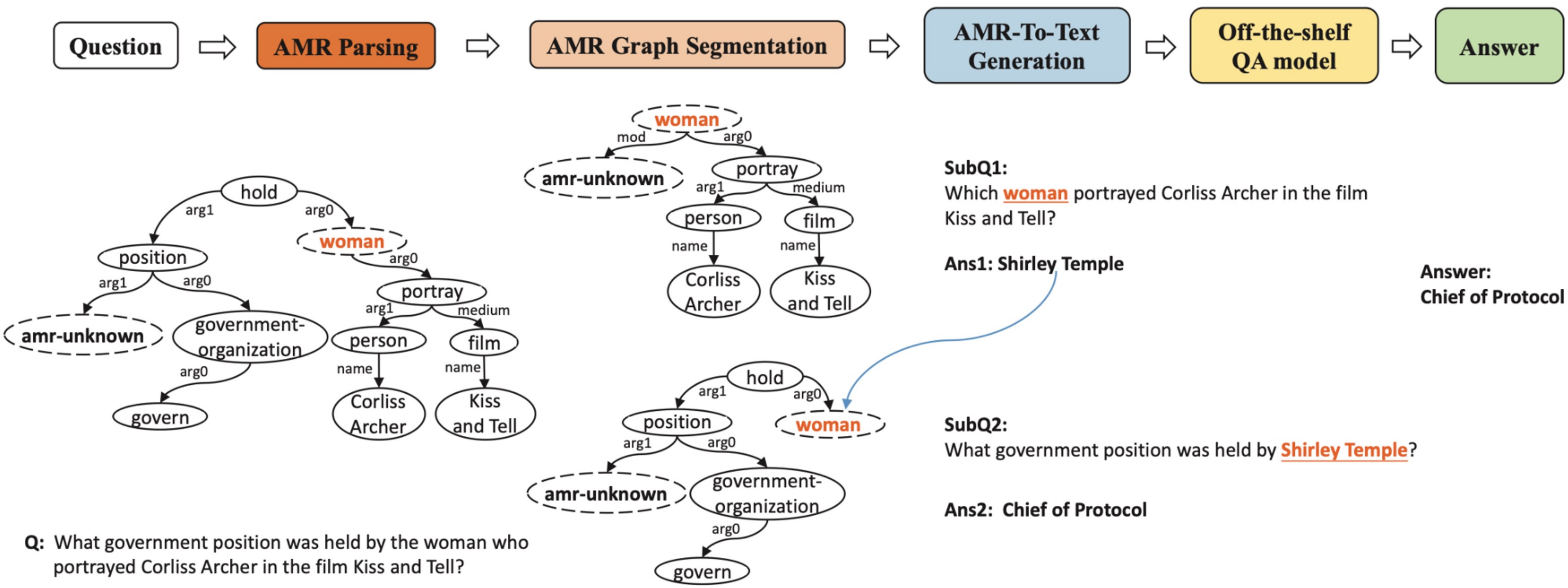
Acer nigrum is used in making what?

Count the awards received by the ones who fought the battle of france?"

What cities are located on the sides of mediterranean sea?

Pavan Kapanipathi et al* **Leveraging Abstract Meaning Representation for Knowledge Base Question Answering** ACL '2021

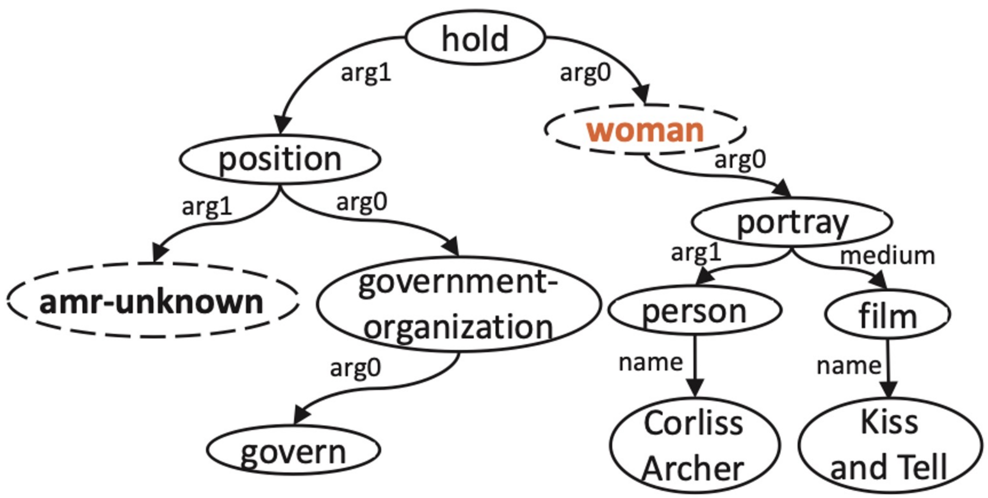
AMR-Based Question Decomposition



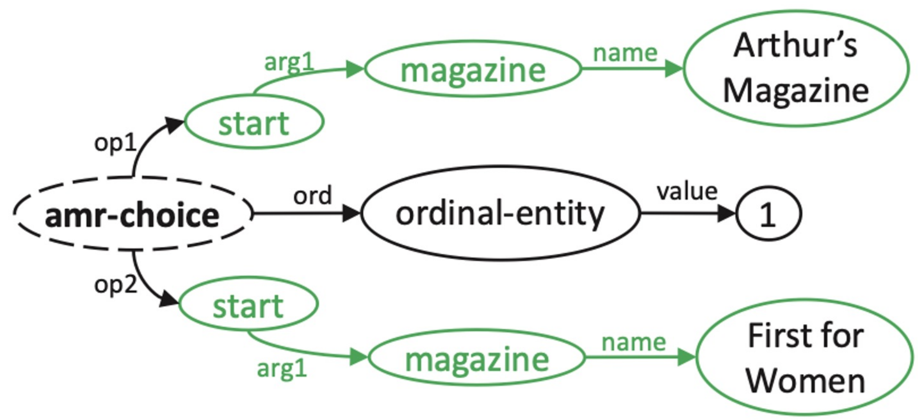
Zhenyun Deng et al. **Interpretable AMR-Based Question Decomposition for Multi-hop**

AMR-Based Question Decomposition

Q1: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell?



Q2: Which magazine was started first Arthur's Magazine or First for Women?



Zhenyun Deng et al. **Interpretable AMR-Based Question Decomposition for Multi-hop**

AMR-Based Question Decomposition

Comparison Question: Who is older, Annie Morton or Terry Richardson? **Answer: Terry Richardson**			
Methods	DecompRC	OUNS	QDAMR
subQ1	Who is older?	Who is Annie Morton?	How old is Terry Richardson?
Ans1	(Annie Morton)	(American model)	(August 14, 1965)
subQ2	Who is Annie Morton or Terry Richardson?	When was Terry Richardson born?	What was Annie Moore's age?
Ans2	(Annie Morton)	(26 July 1999)	(October 8, 1970)
subQ3	Which is smaller (Ans1)(Ans2)?	–	Which is smaller (Ans1)(Ans2)?
Final Ans	Annie Morton	26 July 1999	**Terry Richardson**

Intersection Question: Are both Coldplay and Pierre Bouvier from the same country? **Answer: No**			
Methods	DecompRC	OUNS	QDAMR
subQ1	Are both coldplay?	Where are Coldplay and Coldplay from?	From what country is ColdPlay?
Ans1	(British rock band)	(British)	(British)
subQ2	Are pierre bouvier from the same country?	What country is Pierre Bouvier from?	Where is Pierre Bouvier from?
Ans2	(Canadian)	(Canadian)	(Canadian)
operation	$Intersection$(Ans1,Ans2)	–	$Intersection$(Ans1,Ans2)
Final Ans	**No**	British	**No**

Better accuracy of the final answer and the quality of sub-questions

Zhenyun Deng et al. **Interpretable AMR-Based Question Decomposition for Multi-hop**

AMR-Based Question Decomposition

Decomp Method	bridge		intersec		comparison	
	EM	F1	EM	F1	EM	F1
DecomRC	55.24	71.53	54.55	69.29	52.81	63.44
OUNS	66.41	80.84	66.93	81.07	65.62	79.43
QDAMR	**69.45**	**82.35**	**66.98**	**81.15**	**66.02**	**80.24**

$$\mathbf{B} : \mathrm{Q}\rightarrow\mathrm{SubQ1}\rightarrow\mathrm{Ans1}\rightarrow\mathrm{SubQ2}\rightarrow\mathrm{Ans}$$

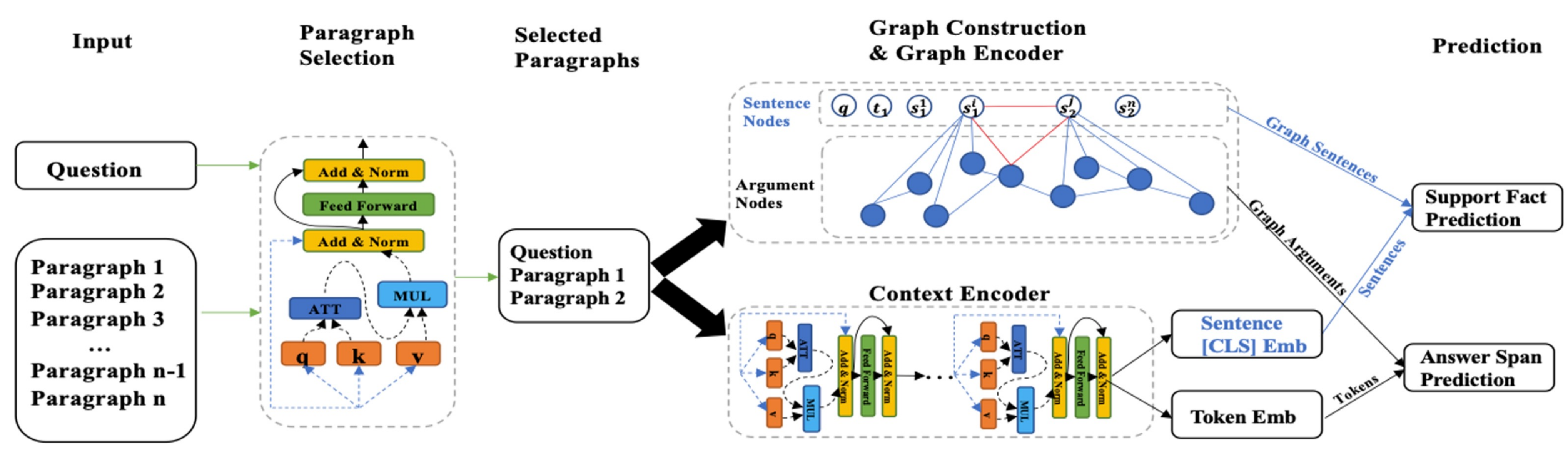
$$\mathbf{I} : \mathrm{Q}\rightarrow(\mathrm{SubQ1}, \mathrm{SubQ2})\rightarrow\mathrm{intersec}(\mathrm{Ans1}, \mathrm{Ans2})\rightarrow\mathrm{Ans}$$

$$\mathbf{C} : \mathrm{Q}\rightarrow(\mathrm{SubQ1}, \mathrm{SubQ2})\rightarrow(\mathrm{Ans1}, \mathrm{Ans2})\rightarrow\mathrm{SubQ3}\rightarrow\mathrm{Ans}.$$

Outperforming existing question-decomposition-based multi-hop QA approaches.

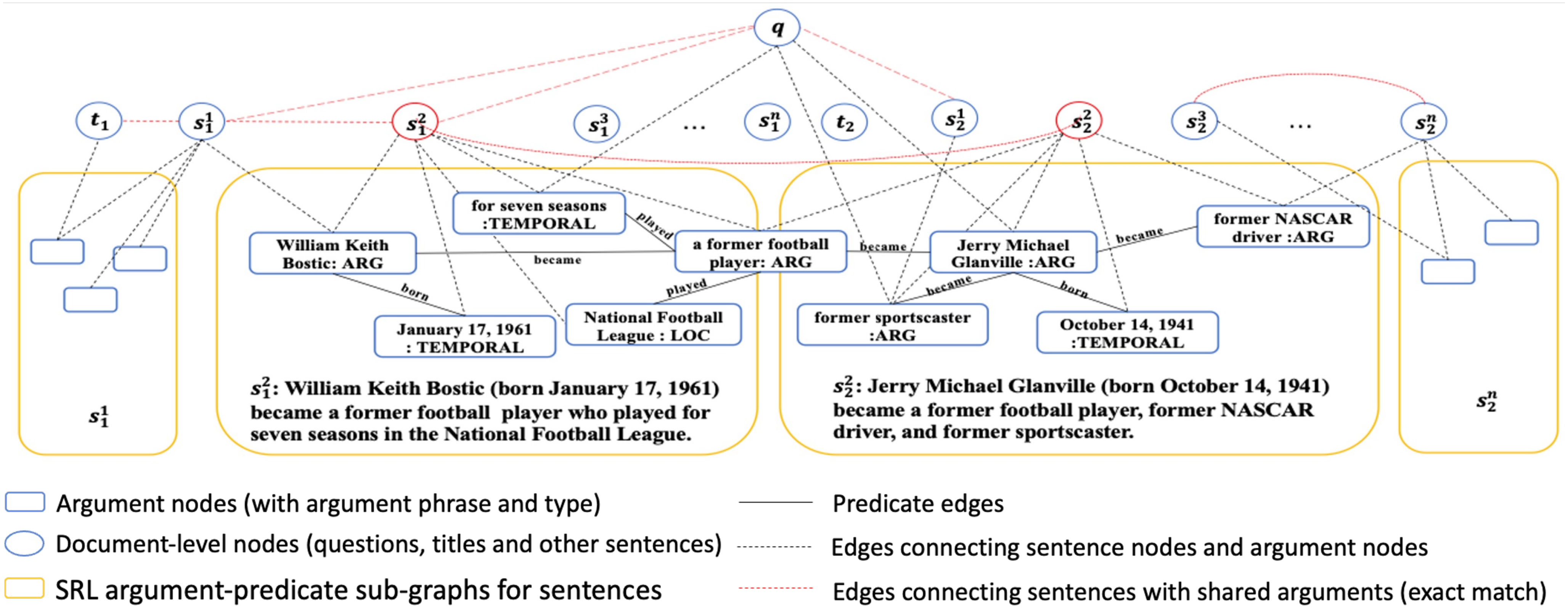
Zhenyun Deng et al. **Interpretable AMR-Based Question Decomposition for Multi-hop Question Answering.** IJCAI'2022

Cross-Document Multi-hop Reading Comprehension



Zheng and Kordjamshidi. **SRLGRN: Semantic Role Labeling Graph Reasoning Network**. EMNLP'2020.

Heterogeneous SRL Graph



Zheng and Kordjamshidi. **SRLGRN: Semantic Role Labeling Graph Reasoning Network**. EMNLP'2020.

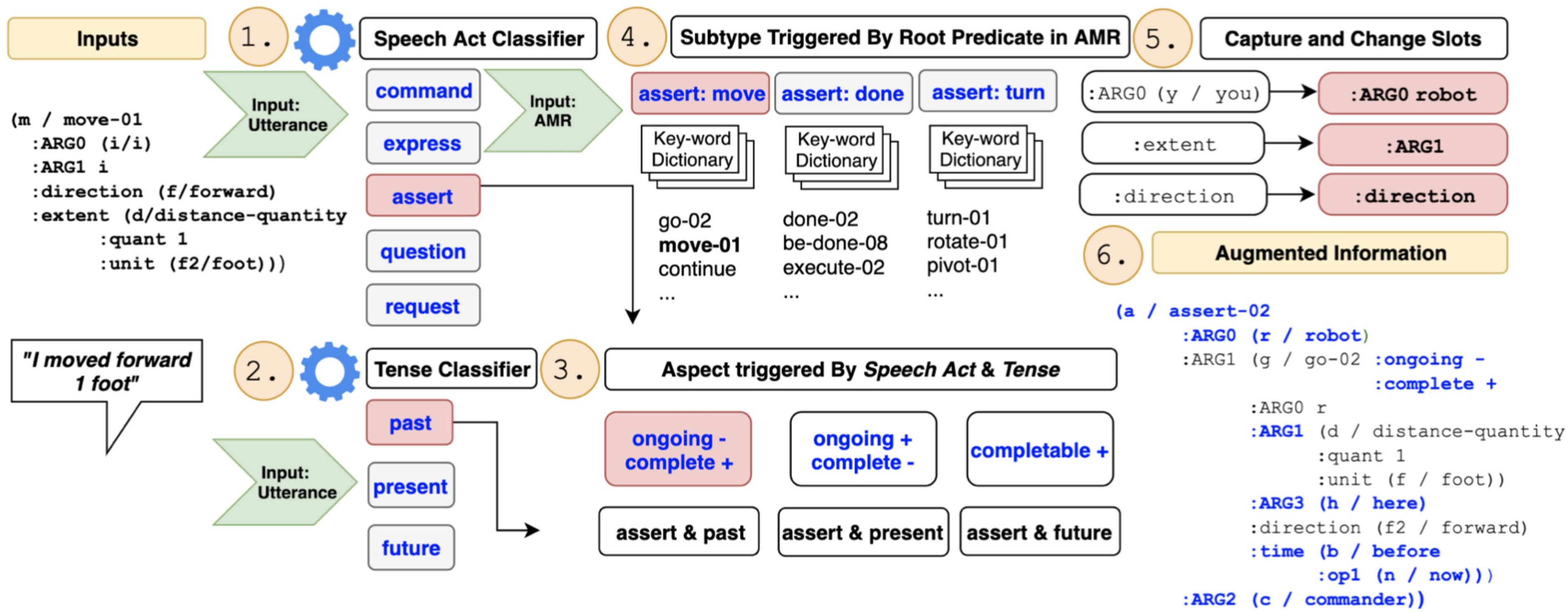
HotpotQA Result

Model	Ans(%)		Sup(%)		Joint(%)	
	EM	F1	EM	F1	EM	F1
Baseline Model (Yang et al., 2018)	45.60	59.02	20.32	64.49	10.83	40.16
KGNN (Ye et al., 2019)	50.81	65.75	38.74	76.79	22.40	52.82
QFE (Nishida et al., 2019)	53.86	68.06	57.75	84.49	34.63	59.61
DecompRC (Min et al., 2019)	55.20	69.63	-	-	-	-
DFGN (Xiao et al., 2019)	56.31	69.69	51.50	81.62	33.62	59.82
TAP	58.63	71.48	46.84	82.98	32.03	61.90
SAE-base (Tu et al., 2019)	60.36	73.58	56.93	84.63	38.81	64.96
ChainEx (Chen et al., 2019)	61.20	74.11	-	-	-	-
HGN-base (Fang et al., 2019)	-	74.76	-	**86.61**	-	**66.90**
SRLGRN-base	**62.65**	**76.14**	**57.30**	85.83	**39.41**	66.37

SRL graph improves the completeness of the graph network over NER graph

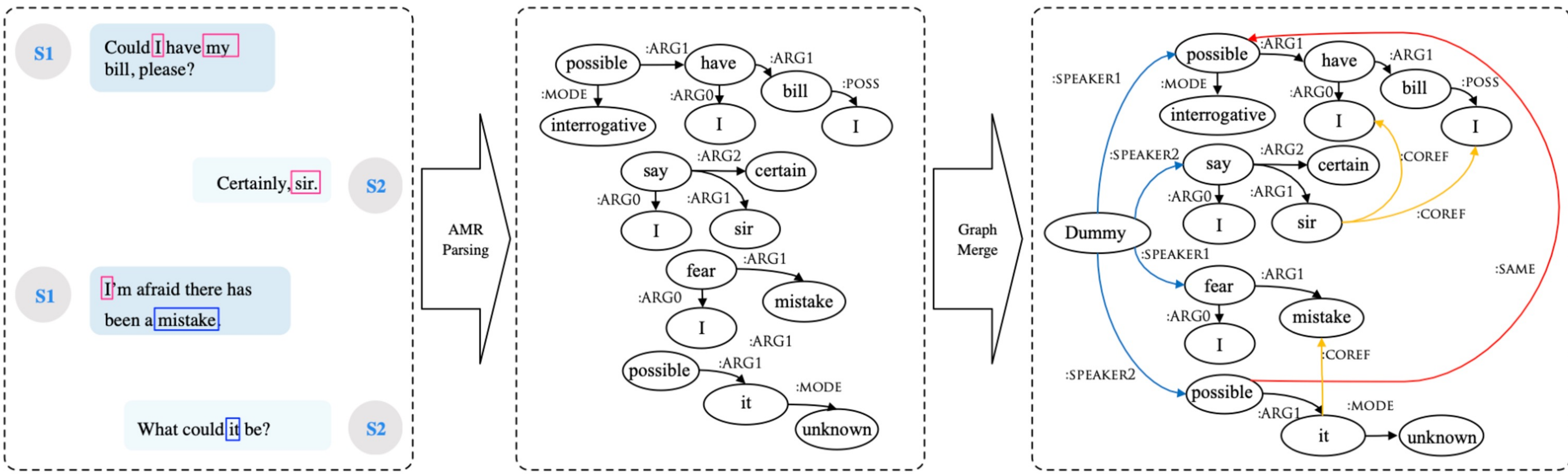
Zheng and Kordjamshidi. **SRLGRN: Semantic Role Labeling Graph Reasoning Network**. EMNLP'2020.

Dialog Modeling via AMR Transformation & Augmentation



Mitchell Abrams, Claire Bonial, L. Donatelli. **Graph-to-graph meaning representation transformations for human-robot dialogue.** SCIL. 2020
Claire Bonial et al. **Augmenting Abstract Meaning Representation for Human-Robot Dialogue.** ACL-DMR. 2019

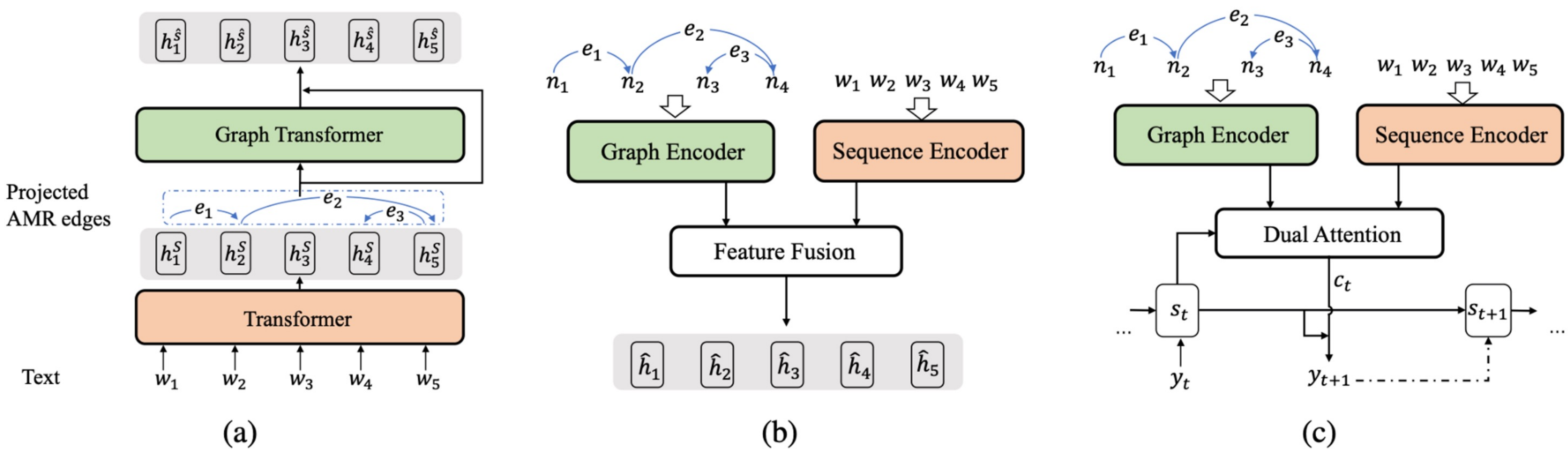
Dialog Modeling via AMR Transformation & Augmentation



(a) Raw Utterance Texts (b) Utterance AMR Graphs (c) Dialogue AMR Graph

Xuefeng Bai, Yulong Chen, Linfeng Song, Yue Zhang. **Semantic Representation for Dialogue Modeling.** ACL 2021

Dialog Modeling via AMR Transformation & Augmentation



(a) Using AMR to enrich text representation. (b,c) Using AMR independently.

Xuefeng Bai, Yulong Chen, Linfeng Song, Yue Zhang. **Semantic Representation for Dialogue Modeling.** ACL 2021

Dialog Modeling via AMR Transformation & Augmentation

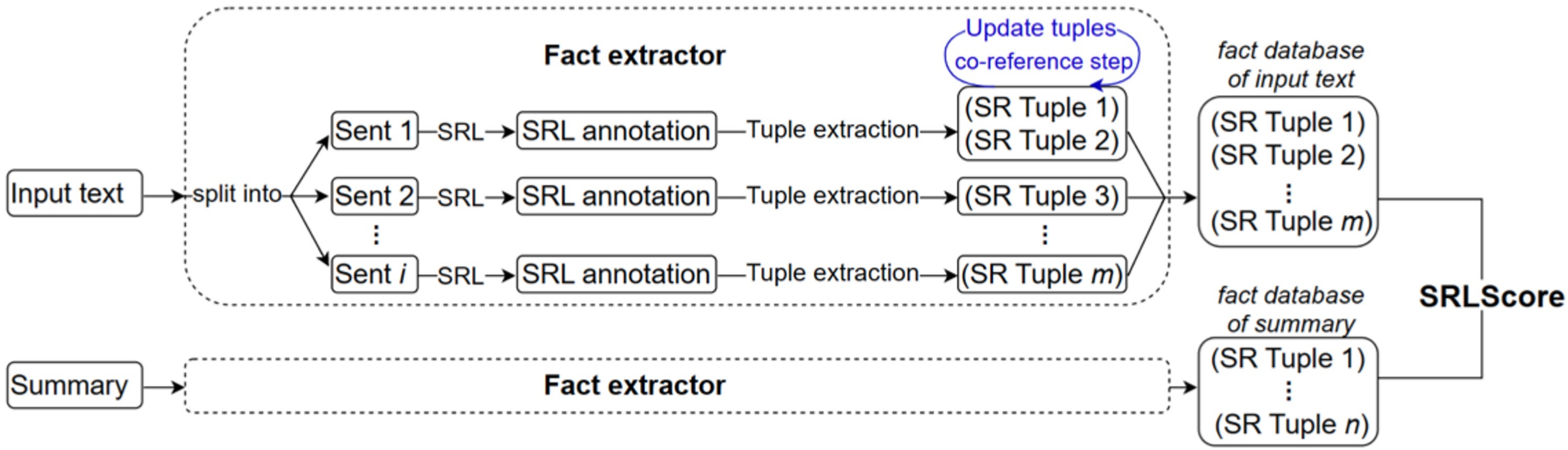
Setting	DialogRE (v2)	DailyDialog
Dialog-AMR(`Dual`)	**68.2**	**38.2/5.9**
-Speaker	67.5	37.7/5.7
-Ident. concept	68.0	37.9/5.8
-Coref	67.8	37.4/5.6
Utter-AMR	67.4	36.9/5.6
Text	66.2	35.4/5.5

manually added relations are useful in dialog relation extraction and dialog generation

semantic knowledge in formal AMR is helpful for dialogue modeling

Xuefeng Bai, Yulong Chen, Linfeng Song, Yue Zhang. **Semantic Representation for Dialogue Modeling.** ACL 2021

SRLScore for Factual Consistency in Text Summarization



- Reference free: Requiring no gold summary
- Adjustable weights for tuple comparison
- Extensible: coreference resolution, alternative similarity functions

Jing Fan, Dennis Aumiller, Michael Gertz. **Evaluating Factual Consistency of Texts with Semantic Role Labeling**. *SEM 2023

SRLScore for Factual Consistency in Text Summarization

Metrics	**QAGS-CNN/DM**		**QAGS-XSUM**		**SummEval**		**Avg.**
	ρ	s	ρ	s	ρ	s	ρ
ROUGE-1 (F1)	0.34	0.32	−0.01	−0.05	0.13	0.14	0.15
BLEU	0.13	0.33	0.08	0.03	0.09	0.14	0.10
METEOR	0.33	0.36	0.06	0.01	0.12	0.14	0.17
BARTScore	0.65	0.57	0.00	0.02	0.27	0.26	0.31
$\text{BARTScore}_{\text{cnn}}$	**0.73**	**0.68**	0.19	0.18	0.35	0.32	0.42
$\text{BARTScore}_{\text{cnn+para}}$	0.69	0.62	0.07	0.07	0.42	**0.37**	0.39
$\text{CoCo}_{\text{span}}$	0.64	0.55	0.22	0.20	0.40	0.35	0.42
$\text{CoCo}_{\text{sent}}$	0.68	0.59	0.16	0.14	0.39	0.35	0.41
$\text{ClozE-R}_{\text{en\_core\_web\_trf}}{}^{*}$	0.66	-	0.32	-	0.47	-	**0.48**
$\text{ClozE-R}_{\text{confidence}}{}^{*}$	0.65	-	0.29	-	**0.48**	-	0.47
$\text{SRLScore}_{\text{base}}$	0.67	0.59	0.20	0.18	0.43	0.33	0.43
$\text{SRLScore}_{\text{coref}}$	0.65	0.58	0.27	0.26	0.43	0.32	0.45
$\text{SRLScore}_{\text{coref-optimized}}$	-	-	**0.33**	**0.33**	-	-	-

- Pearson (ρ) and Spearman (s) correlation of metrics with human ratings on the evaluated datasets.
- No significant differences between any of the factuality-specific metrics (SRLScore, BARTScore, and CoCo)

Jing Fan, Dennis Aumiller, Michael Gertz. **Evaluating Factual Consistency of Texts with Semantic Role Labeling**. *SEM 2023

SRLScore for Factual Consistency in Text Summarization

simplified triplet representation

Metrics		QCNNDM		QXSUM		SummE	
		ρ	s	ρ	s	ρ	s
$\text{SRLScore}_{\text{openie}}$	Exact	0.59	0.51	0.09	0.09	0.34	0.28
	ROUGE	0.62	0.56	0.07	0.07	0.41	0.32
	SpaCy	0.59	0.53	0.13	0.10	0.37	0.32
$\text{SRLScore}_{\text{base}}$	Exact	0.61	0.54	0.14	0.15	$0.37^{\dagger}$	$0.31^{\ddagger}$
	ROUGE	**0.67**	**0.59**	$0.15^{\dagger}$	0.13	$\mathbf{0.43}^{\dagger}$	0.33
	SpaCy	0.63	0.55	**0.20**	**0.18**	$0.40^{\dagger}$	$\mathbf{0.34}^{\dagger}$

- SRL-based semantic representations enable better scoring function than (*agent*, *relation*, *patient*) triplets

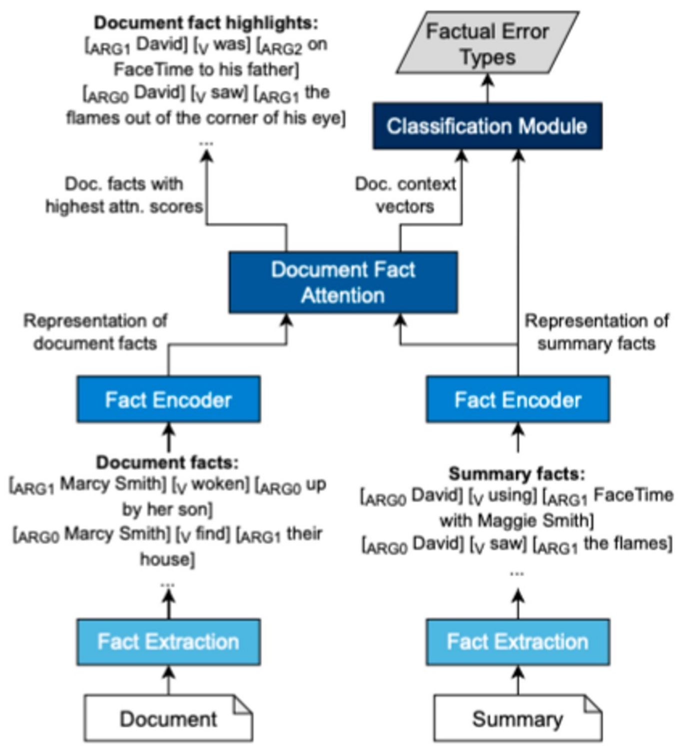
Jing Fan, Dennis Aumiller, Michael Gertz. **Evaluating Factual Consistency of Texts with Semantic Role Labeling**. *SEM 2023

Interpretable Automatic Fine-grained Inconsistency Detection

Source text	
Marcy Smith was woken up by her son David to find their house in Glovertown, Newfoundland and Labrador, completely engulfed in flames ... Mrs Smith said if it wasn't for her son, she and her daughter probably wouldn't have survived. **David was on FaceTime to his father at the time**, so was the only one awake and **saw the flames out of the corner of his eye** ...	
Error type	**Example summary**
Extrinsic noun phrase error: Errors that add new object(s), subject(s), or prepositional object(s) that cannot be inferred from the source article.	David was using FaceTime with ***Maggie Smith*** and saw the flames.
Intrinsic noun phrase error: Errors that misrepresent object(s), subject(s), or prepositional object(s) from the source article.	David was using FaceTime with ***Marcy Smith*** and saw the flames.
Extrinsic predicate error: Errors that add new main verb(s) or adverb(s) that cannot be inferred from the source article.	David was ***eating*** and saw the flames.
Intrinsic predicate error: Errors that misrepresent main verb(s) or adverb(s) from the source article.	David was ***engulfed*** and saw the flames.

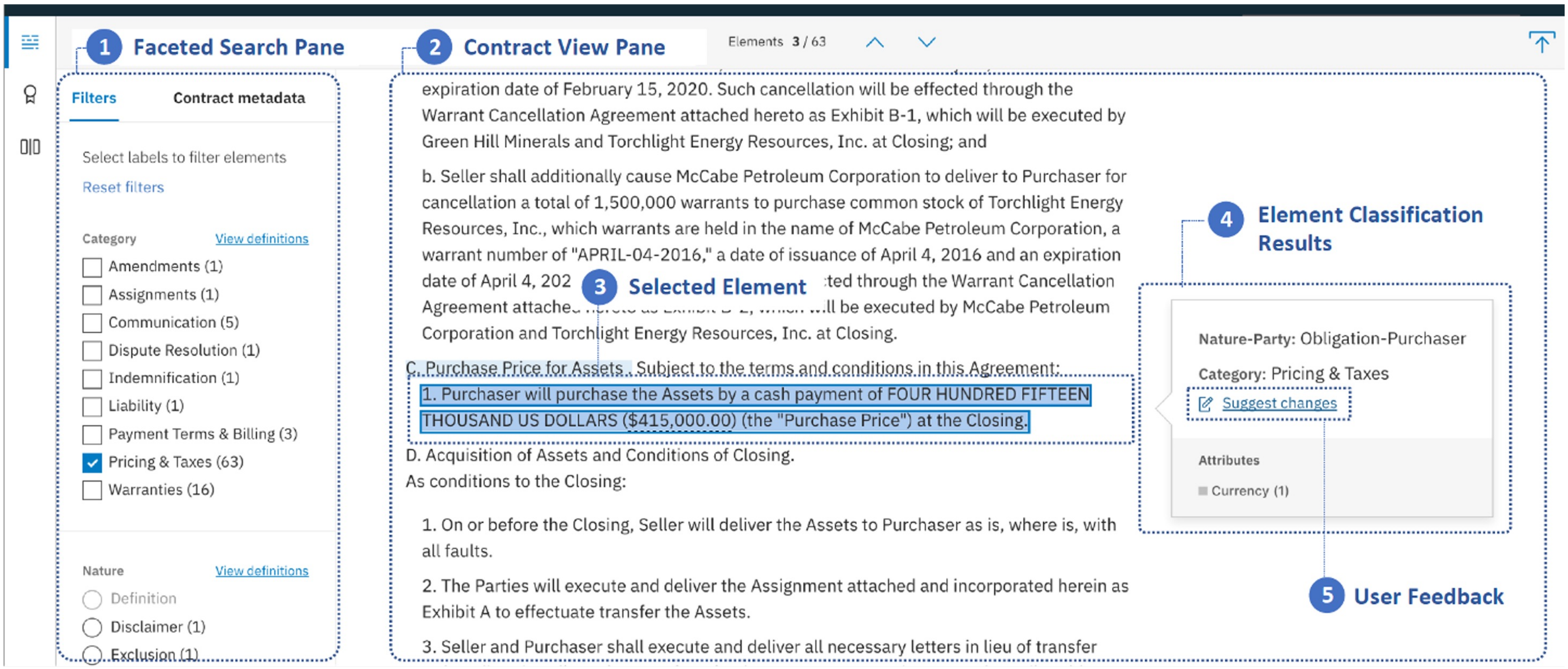
Hou Pong Chan1 Qi Zeng2 Heng Ji. **Interpretable Automatic Fine-grained Inconsistency Detection in Text Summarization**. ACL (findings) 2023

Interpretable Automatic Fine-grained Inconsistency Detection



Hou Pong Chan1 Qi Zeng2 Heng Ji. **Interpretable Automatic Fine-grained Inconsistency Detection in Text Summarization**. ACL (findings) 2023

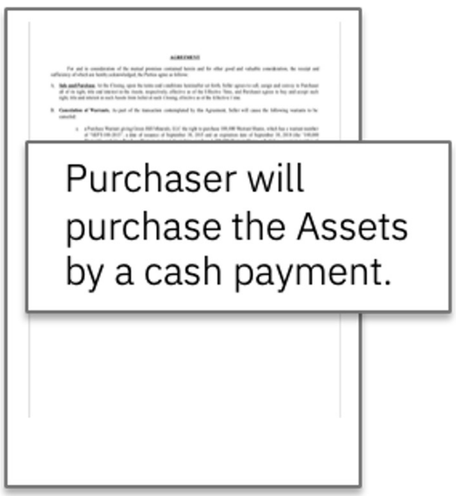
Case Study - Watson Discover Content Intelligence



A. Agarwal et al.**Development of an Enterprise-Grade Contract Understanding System.** NAACL (industry) 2021

Case Study - Watson Discover Content Intelligence

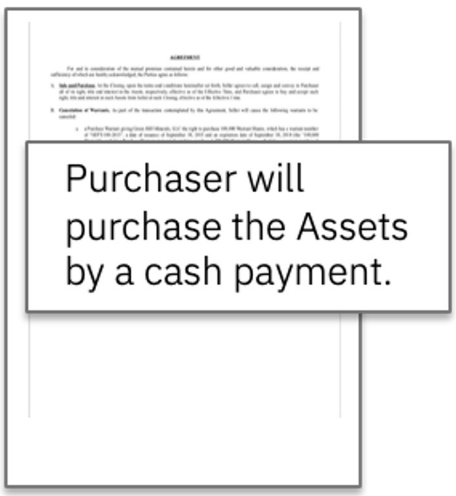
Element



A. Agarwal et al.**Development of an Enterprise-Grade Contract Understanding System.** NAACL (industry) 2021

Case Study - Watson Discover Content Intelligence

Element



Expanded SRL as
Semantic NLP Primitives

[Purchaser]$_{A0}$
[will]$_{TENSE\text{-}FUTURE}$
purchase
[the Assets]$_{A1}$
[by a cash payment]$_{ARGM\text{-}MNR}$

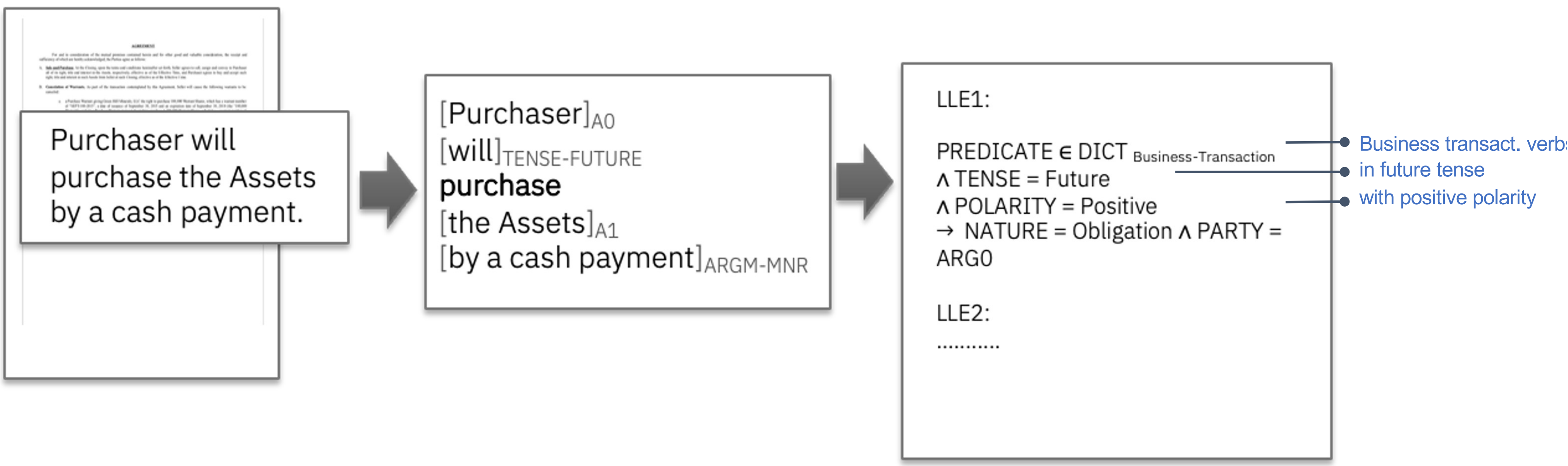
Provided by SystemT
[ACL '10, NAACL '18]

A. Agarwal et al**. Development of an Enterprise-Grade Contract Understanding System.** NAACL (industry) 2021

Case Study - Watson Discover Content Intelligence

Element

Expanded SRL as
Semantic NLP Primitives

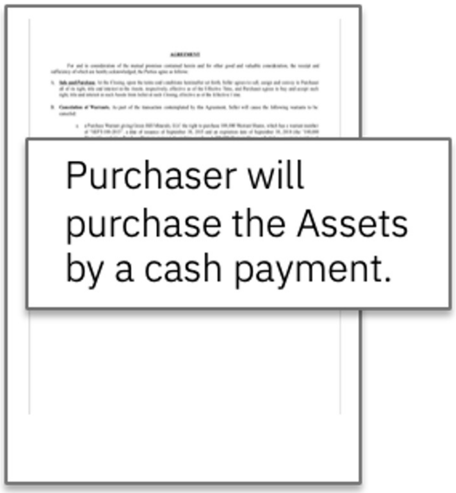


A. Agarwal et al**. Development of an Enterprise-Grade Contract Understanding System.** NAACL (industry) 2021

Case Study - Watson Discover Content Intelligence

Core NLP Primitives & Operators

Element



Semantic NLP Primitives

[Purchaser]$_{\text{A0}}$
[will]$_{\text{TENSE-FUTURE}}$
purchase
[the Assets]$_{\text{A1}}$
[by a cash payment]$_{\text{ARGM-MNR}}$

Provided by SystemT
[ACL '10, NAACL '18]

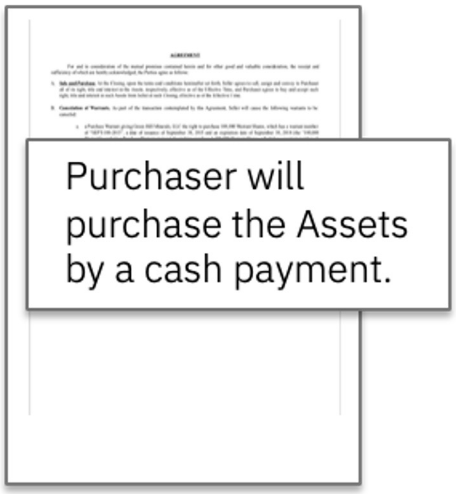
A. Agarwal et al**. Development of an Enterprise-Grade Contract Understanding System.** NAACL (industry) 2021

Case Study - Watson Discover Content Intelligence

Core NLP Primitives & Operators

Domain Specific Concepts

Element



Semantic NLP Primitives

[Purchaser]$_{\text{A0}}$
[will]$_{\text{TENSE-FUTURE}}$
purchase
[the Assets]$_{\text{A1}}$
[by a cash payment]$_{\text{ARGM-MNR}}$

Legal Domain LLEs

LLE1:

PREDICATE ∈ DICT $_{\text{Business-Transaction}}$
∧ TENSE = Future
∧ POLARITY = Positive
→ NATURE = Obligation ∧ PARTY = ARG0

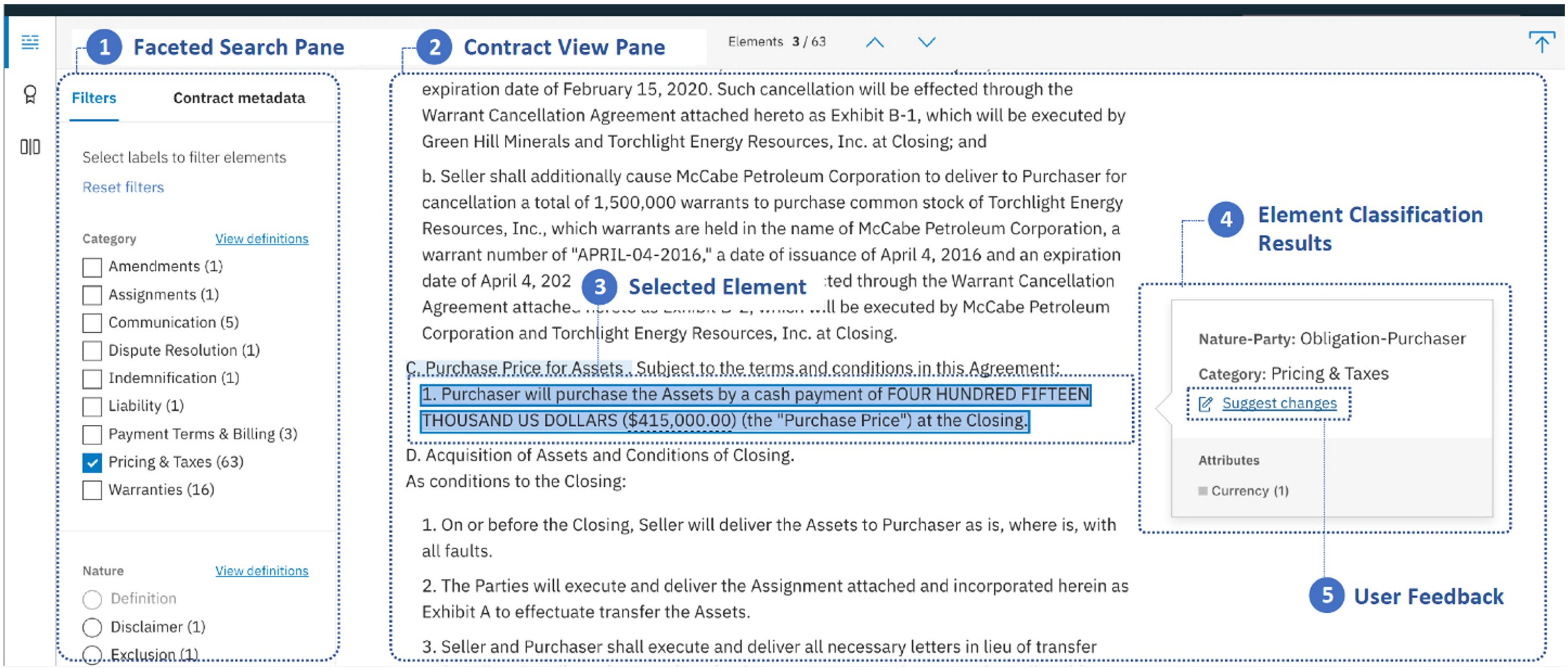
LLE2:
....

Business transact. verbs
in future tense
with positive polarity

Provided by SystemT
[ACL '10, NAACL '18]

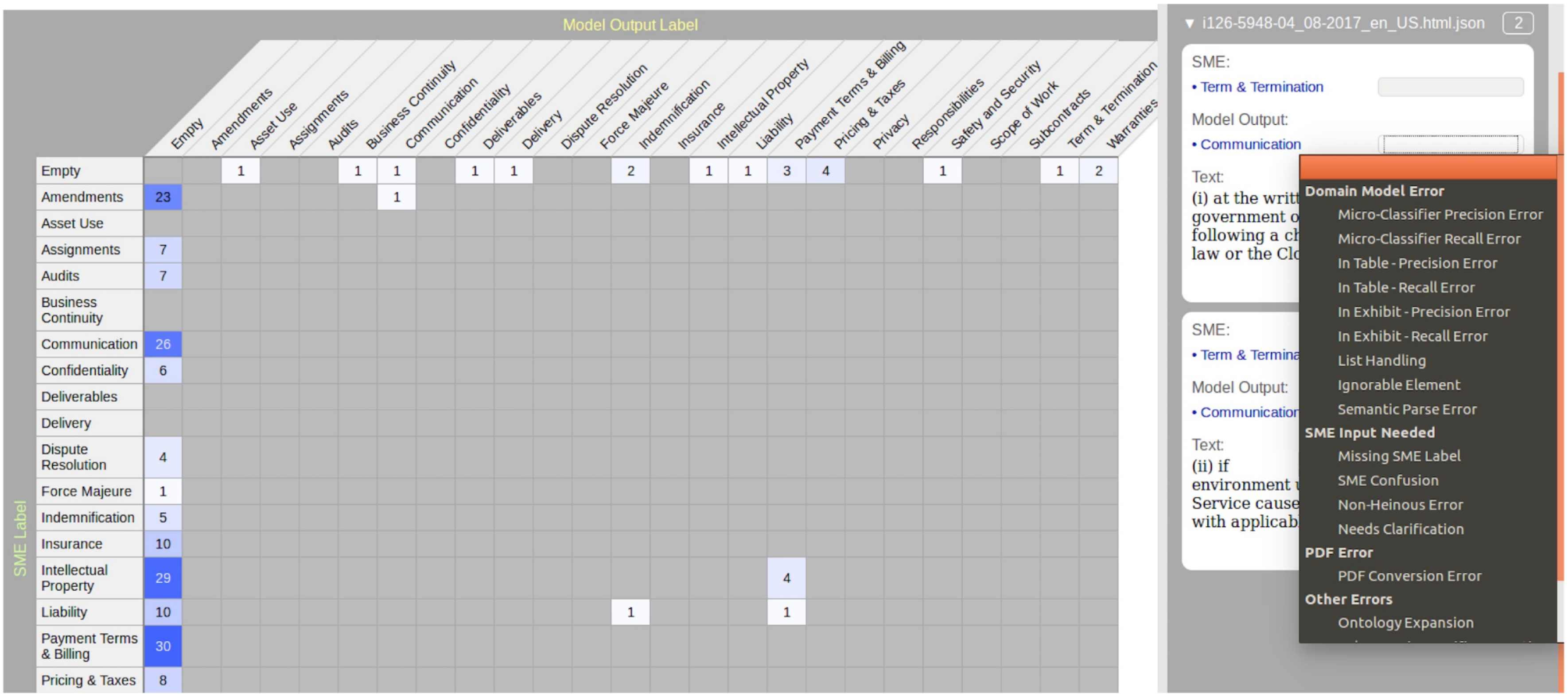
A. Agarwal et al**. Development of an Enterprise-Grade Contract Understanding System.** NAACL (industry) 2021

Case Study - Watson Discover Content Intelligence



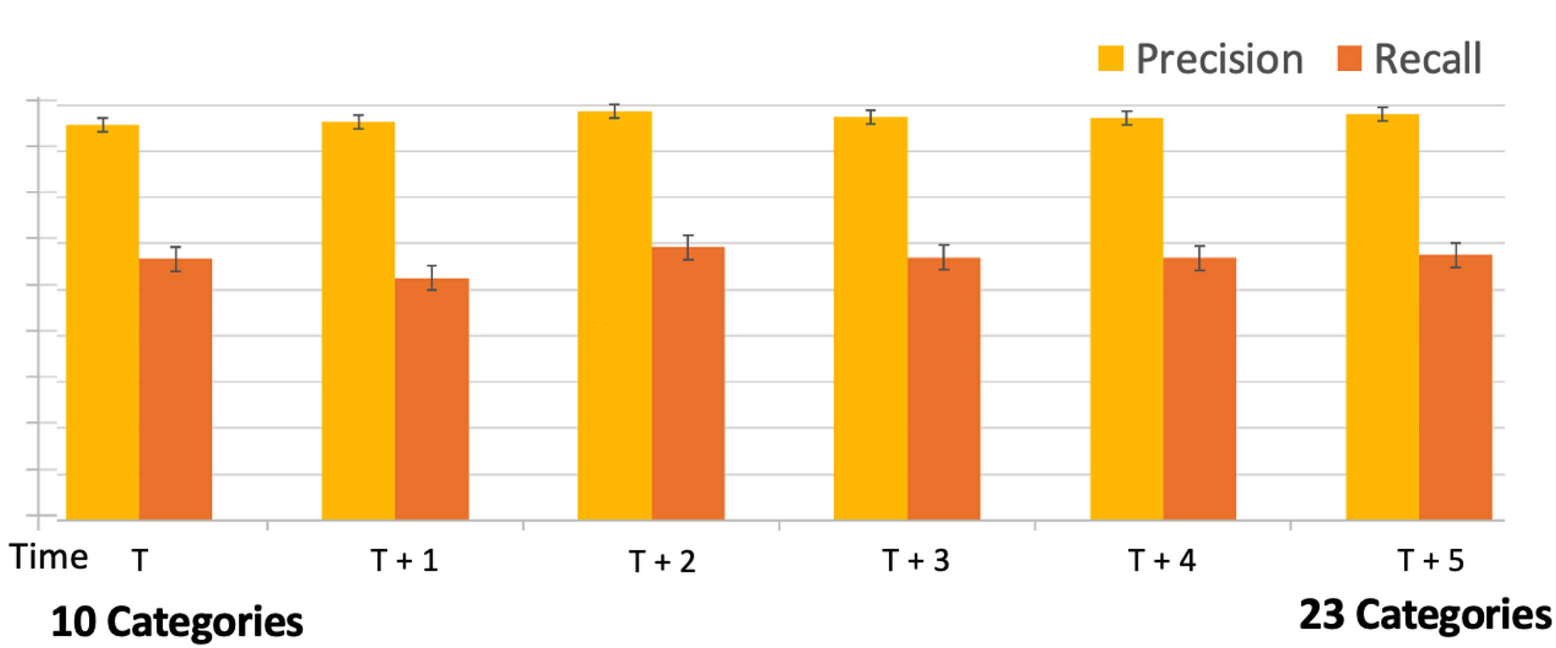
A. Agarwal et al. **Development of an Enterprise-Grade Contract Understanding System.** NAACL (industry) 2021

Explainability + Tooling → Better Root Cause Analysis



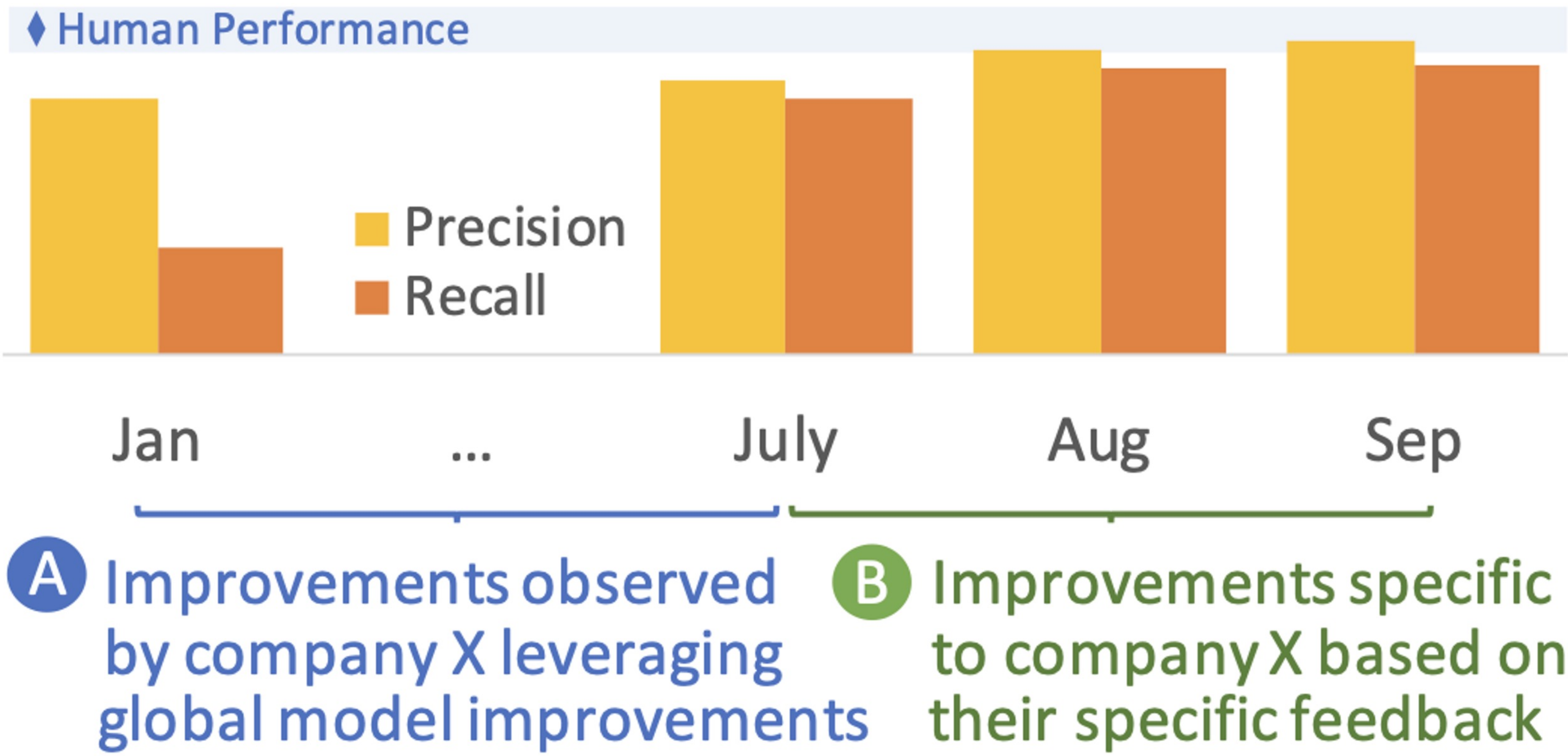
A. Agarwal et al**. Development of an Enterprise-Grade Contract Understanding System.** NAACL (industry) 2021
Yannis Katsis and Christine T. Wolf. **ModelLens: An Interactive System to Support the Model Improvement Practices of Data Science Teams.** CSCW 2019

Model Stability with Increasing Complexity



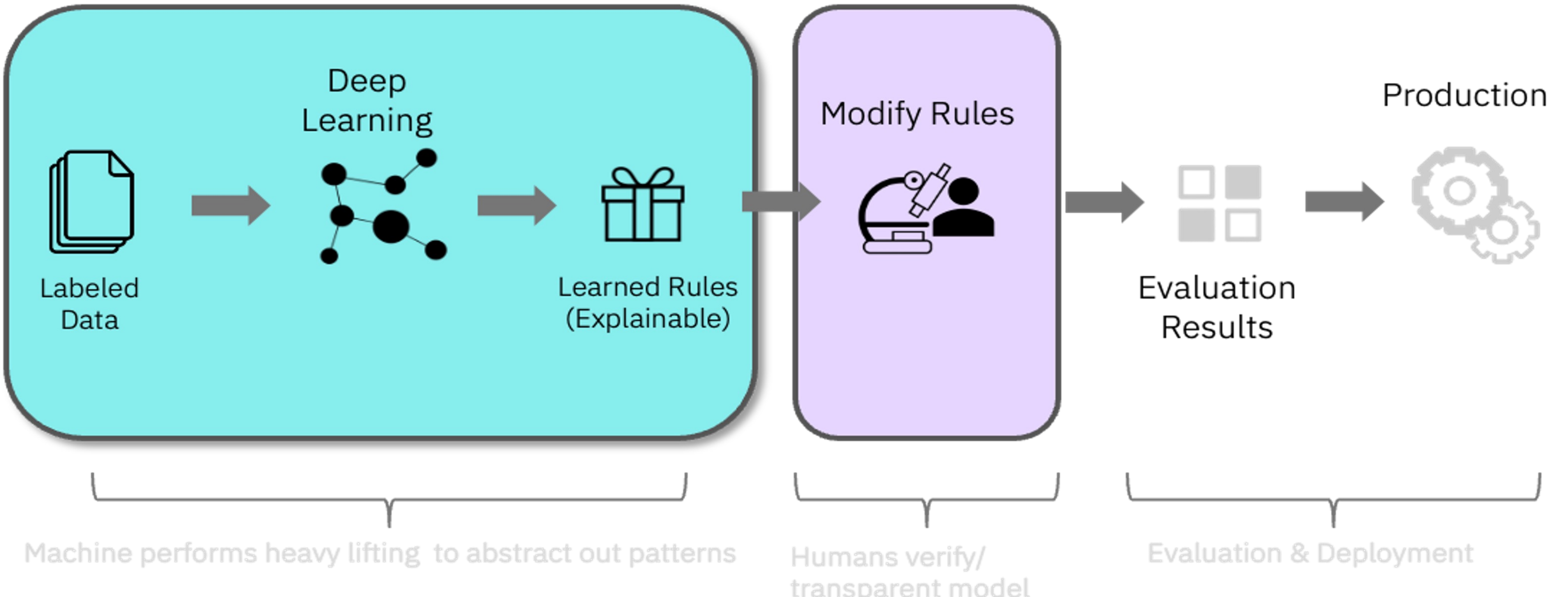
A. Agarwal et al**. Development of an Enterprise-Grade Contract Understanding System.** NAACL (industry) 2021

Effectiveness of Feedback Incorporation



A. Agarwal et al**. Development of an Enterprise-Grade Contract Understanding System.** NAACL (industry) 2021

Human & Machine Co-Creation



Prithvi Sen. et al. **HEIDL: Learning Linguistic Expressions with Deep Learning and Human-in-the-Loop**. ACL'2019

Prithvi Sen. et al. **Learning Explainable Linguistic Expressions with Neural Inductive Logic Programming for Sentence Classification.** EMNLP'2020

User Study: Human & Machine Co-Creation

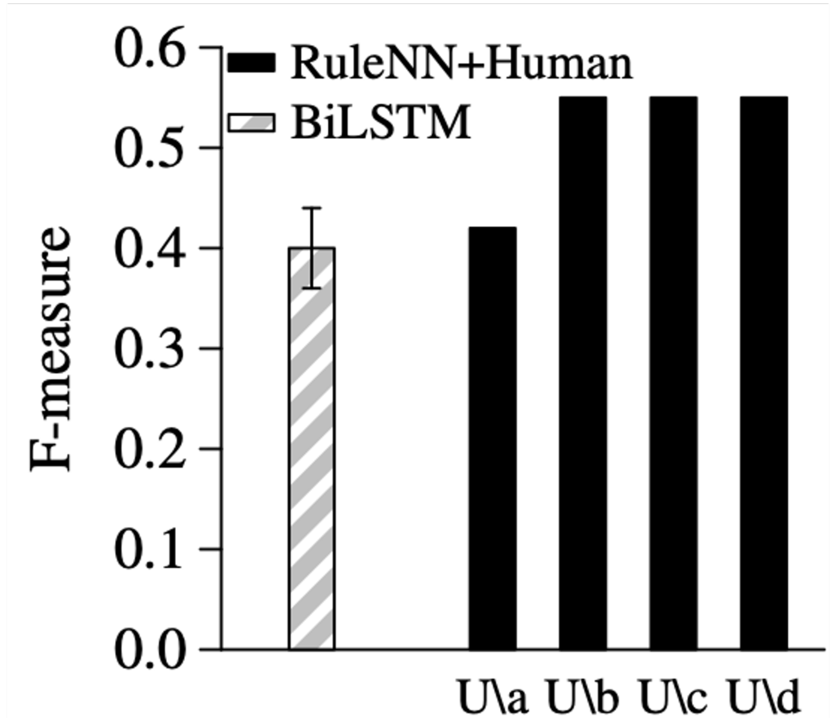
User study

–4 NLP Engineers with 1-2 year experience

–2 NLP experts with 10+ years experience

Key Takeaways

- **Explanation of learned rules**: Visualization tool is very effective
- **Reduction in human labor**: Co-created model created within 1.5 person-hrs outperforms black-box sentence classifier
- **Lower requirement on human expertise**: Co-created model is at par with the model created by Super-Experts



Prithvi Sen. et al. **HEIDL: Learning Linguistic Expressions with Deep Learning and Human-in-the-Loop**. ACL'2019

Prithvi Sen. et al. **Learning Explainable Linguistic Expressions with Neural Inductive Logic Programming for Sentence Classification.** EMNLP'2020

Summary: Value of Meaning Representation

SRL
AMR

	Work Out-of-box	Deeper understanding of text		
	Overcome Low-resource Challenges	Robustness against linguistics variants & complexity	Better model generalization	Explainability & Interpretability
Information Extraction	✓	✓	✓	
Text Classification	✓	✓	✓	✓
Natural Language Inference			✓	
Question Answering		✓		✓
Dialog			✓	
Machine Translation	✓	✓		
Factual Consistency	✓		✓	✓