Minimally Supervised Information Extraction

New Frontiers of Information Extraction (Part I)

Ben Zhou

Computer and Information Sciences

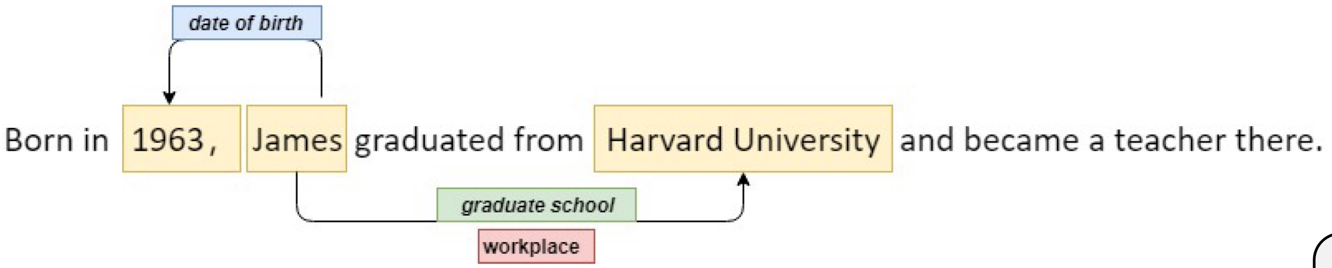
University of Pennsylvania

July 2022

NAACL Tutorials

New Frontiers of Information Extraction

Information Extraction



- From the texts:

- 1 . Identify the concepts
  - Entities, events, terms, etc.

- 2. Identify the relations and other properties
  - Entity-entity / event-event
  - Temporal properties
  - etc.

Lexical IE:
- Named entity recognition
- Entity/event typing
- Entity/event linking

Relational IE:
- Relation extraction
  - Entity / events
  - Sentence/Document
  - Temporal
- Coreference Resolution

How IE is mostly done?

- Direct supervision
  - CoNLL 2003: 20K+ entity mention annotations for NER
  - Ontonotes5.0: exhaustive NER annotation on 2.9M tokens
- But
  - Closed label set
  - Poor transferability
  - Annotation artifacts
  - “Worse” on “rare” items
  - High cost for new tasks

A multi-head classification model drops from **96** to **74** (F1) on CoNLL, when it’s trained on Ontonotes on the *same* types.

Zhou et al. (2021) reports $1.0 per instance to annotate clean temporal relations on obscure texts.

In this part of the tutorial:

› We explore the the central question

How do we find alternative supervision sources for IE tasks?

- Cheaper to generalize to new / rare / hard tasks
- Alleviate in-distribution artifacts

Alternative Supervision Sources

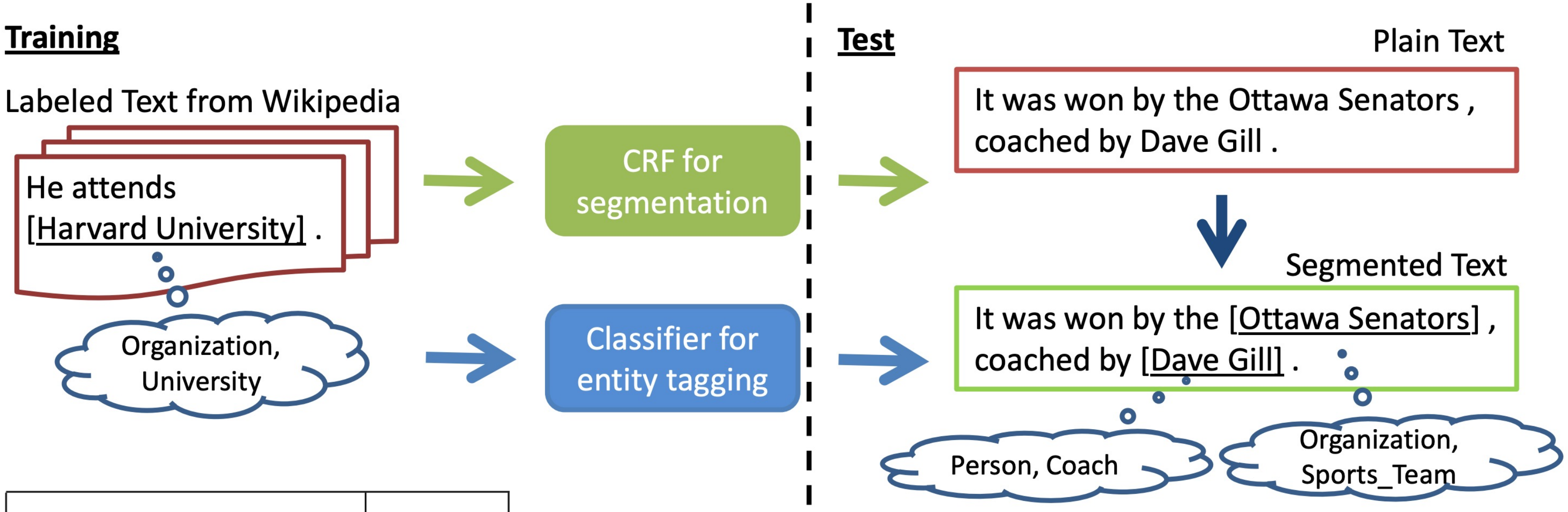
- **Weak supervision**: from task-related distant signals
  - Easy to acquire
  - Task-specific
  - May be noisy
- **Indirect supervision**: from other tasks
  - Human annotations from other popular tasks such as NLI and QA
  - Non-task-specific
  - Needs clever ways to be applied

Weak Supervision

- Mark joined Amazon a month ago.
  - What is the entity type?
- Weak Supervision:
  - From knowledge bases
    - Amazon.com, Inc is an American multinational technology company.

Weak Supervision – Knowledge Bases

- One of the earliest attempts: entity and entity relations
- Ling and Weld (2012): NER from KB supervision



Measure	Strict
NEL	0.220
Stanford (CoNLL)	0.425
FIGER	0.471
FIGER (GOLD)	0.532

Note: StanfordNER only predicts a subset of the taxonomy

Ling and Weld. Fine-Grained Entity Recognition. AAAI 2012.

Weak Supervision – Knowledge Bases

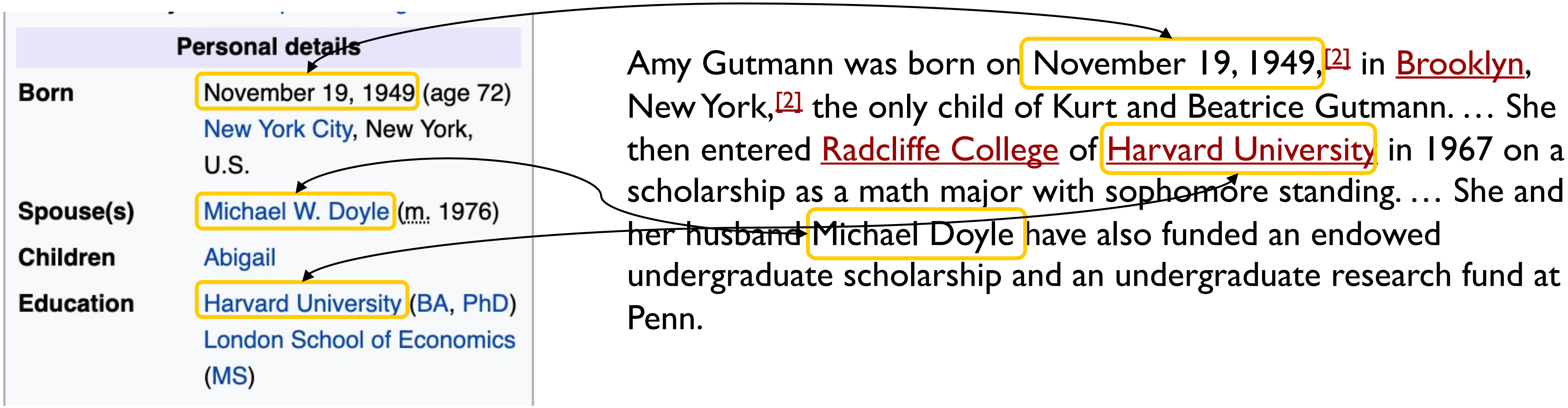
- One of the earliest attempts: entity and entity relations
- Mintz et al. (2009)
  - Assumes Freebase relations exist sentences that contain the same entity pairs.
  - Learns a large set of relations (102), but noisy

Freebase Relation	Entity Pairs	Sentence with same EPs
/location/location/contains	Paris, Montmartre	**Montmartre** is a large hill in **Paris**'s 18th arrondissement.
/film/director/film	Michael Mann, Collateral	***Collateral*** is a 2004 American neo-noir action thriller film directed and produced by **Michael Mann**.
/people/person/profession	Barak Obama, President	**Obama** announced his run for the **president**.

Mintz et al. Distant supervision for relation extraction without labeled data. ACL 2009.

Weak Supervision – Knowledge Bases

- One of the earliest attempts: entity and entity relations
- Hoffmann et al. (2010): Learn from Wikipedia infoboxes



  - Matching info box entities with context, to learn context-dependent relation extraction.
    - 5000+ relations
  - Many follow-up work on de-noising, but with similar weak signals

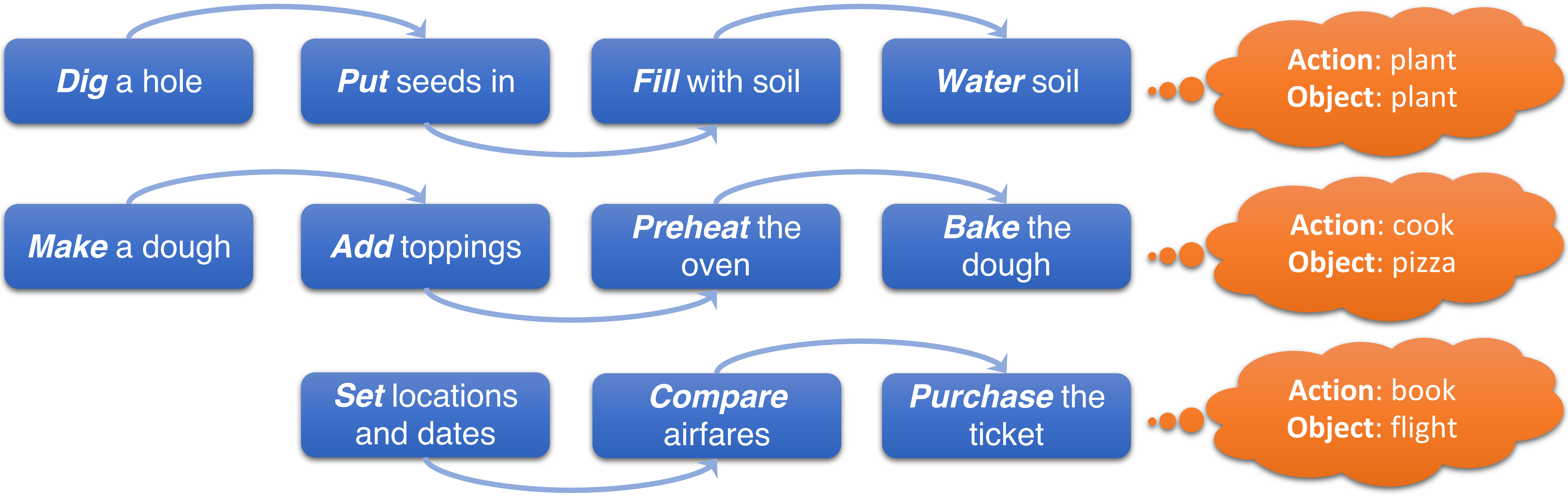
Hoffmann et al. Learning 5000 Relational Extractors. ACL 2010

Weak Supervision

- Mark joined Amazon a month ago.
  - What is the entity type?
- Weak Supervision:
  - From knowledge bases
    - Amazon.com, Inc is an American multinational technology company.
  - From weak but richer label representations
    - Word-embedding(company) is close to Word-embedding(Amazon)

Weak Supervision from Label Representations

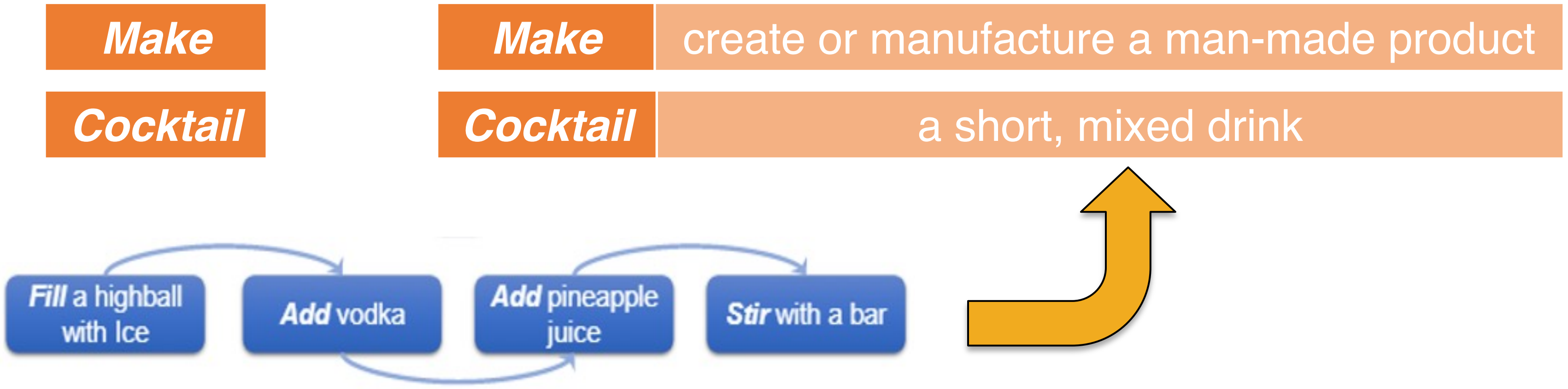
- Chen et al. (2020): Event Process Typing



Chen et al. "What Are You Trying to Do?" Semantic Typing of Event Processes. CoNLL 2020

Weak Supervision from Label Representations

- Chen et al. (2020): Event Process Typing
- Direct label understanding is difficult
  - Add glossary definition as a “weak” label defintion



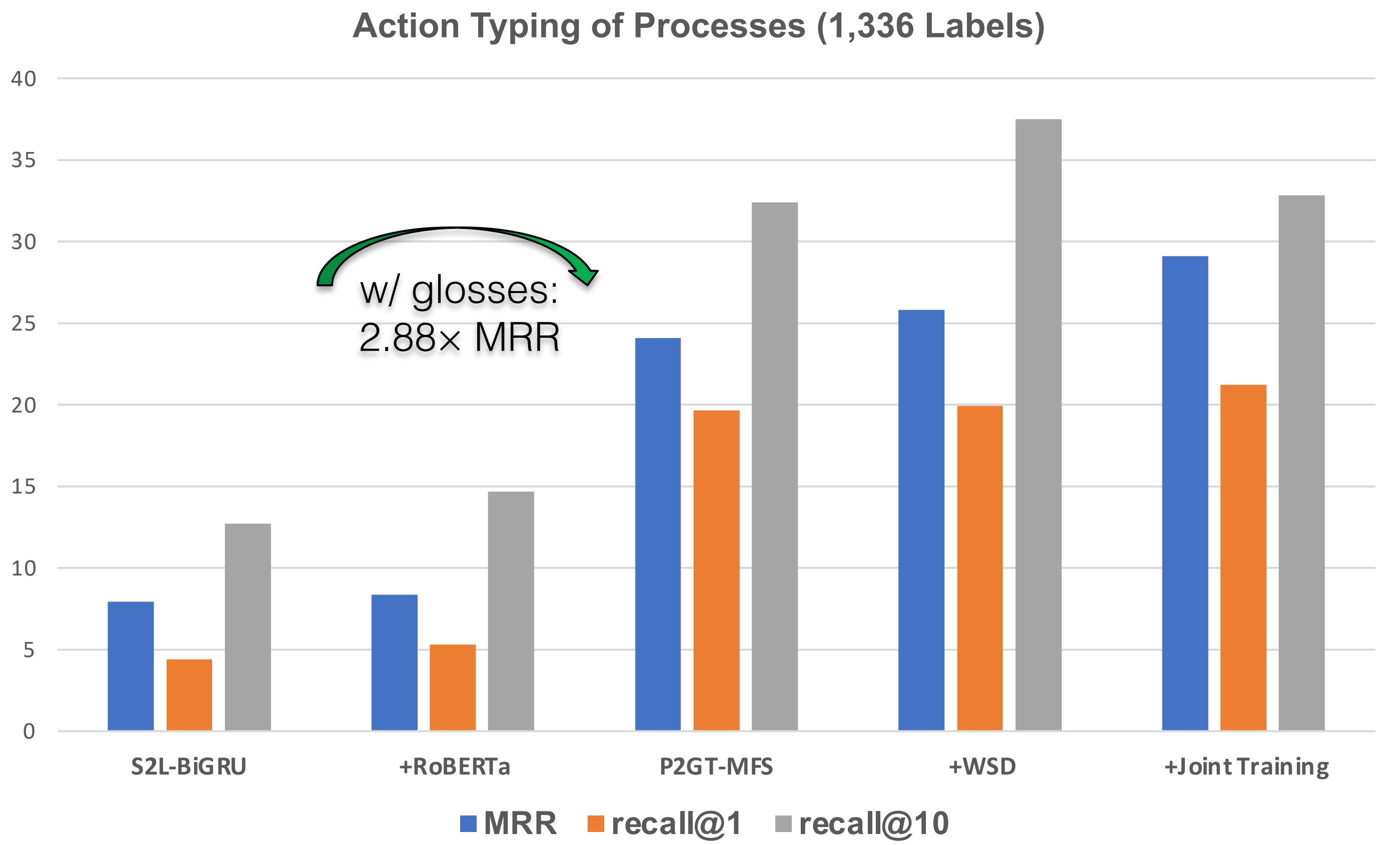
Why using label glosses?

- Semantically richer than labels themselves
- Capturing the association of a process-gloss pair (two sequences) is much easier
- Jump-starting few-shot label representations (and benefiting with fairer prediction)

Chen et al. “What Are You Trying to Do?” Semantic Typing of Event Processes. CoNLL 2020

Weak Supervision from Label Representations

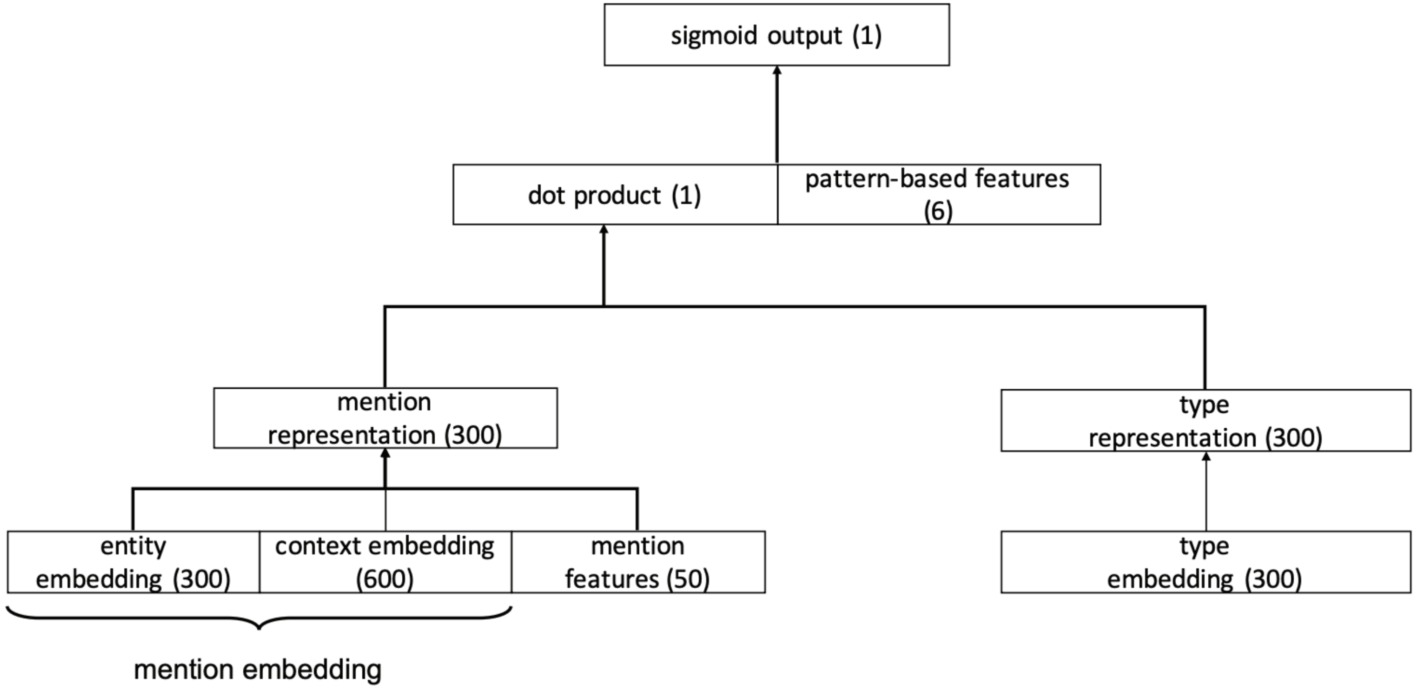
- Chen et al. (2020)
- Gloss knowledge brings the most improvement



Chen et al. "What Are You Trying to Do?" Semantic Typing of Event Processes. CoNLL 2020

Weak Supervision from Label Representations

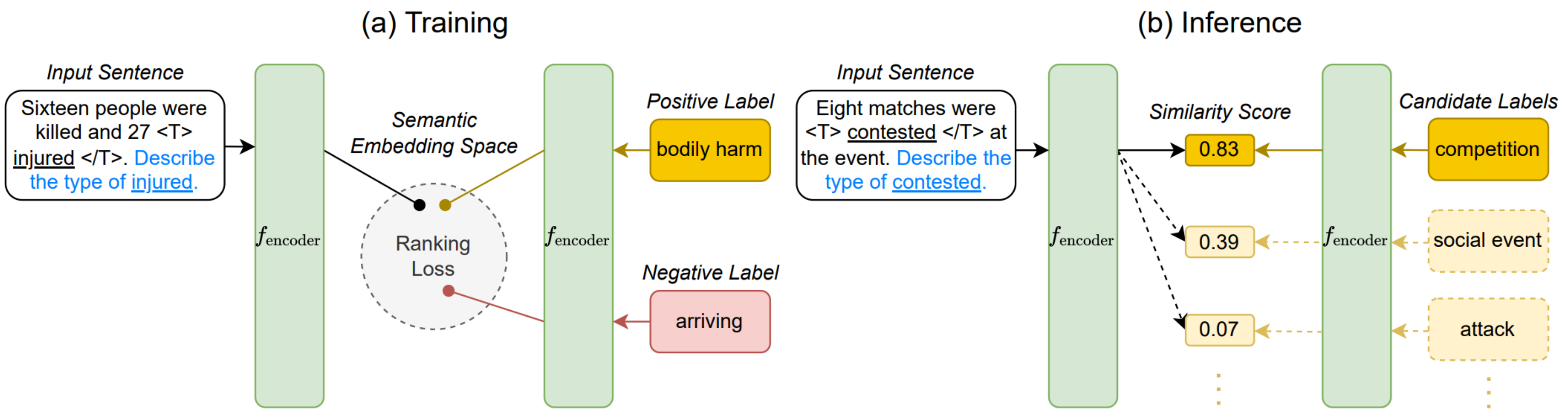
› Yuan and Downey (2018): Open entity typing from label embeddings
  » Labels as meaningless indices -> labels as word embeddings (carries information)
  » Optimizes gold "label embedding" to be closer to the mention embedding.



Type embedding: GloVe

Weak Supervision from Label Representations

- Huang et al. (2022): Similar idea but with modern LMs



Model	P	R	F1
UFET-biLSTM† (Choi et al., 2018)	48.1	23.3	31.3
LabelGCN† (Xiong et al., 2019)	50.3	29.2	36.9
LDET† (Onoe and Durrett, 2019)	51.5	33.0	40.1
Box4Types*† (Onoe et al., 2021)	52.8	38.8	44.8
LRN (Liu et al., 2021)	**54.5**	38.9	45.4
MLMET† (Dai et al., 2021)	53.6	45.3	49.1
UniST$_{\text{BASE}}$	49.2	49.4	49.3
UniST$_{\text{LARGE}}$*	50.2	**49.6**	**49.9**

Experiments on ultrafine dataset (Choi et al. 2018)

Huang et al.: Unified Semantic Typing with Meaningful Label Inference. NAACL 2022

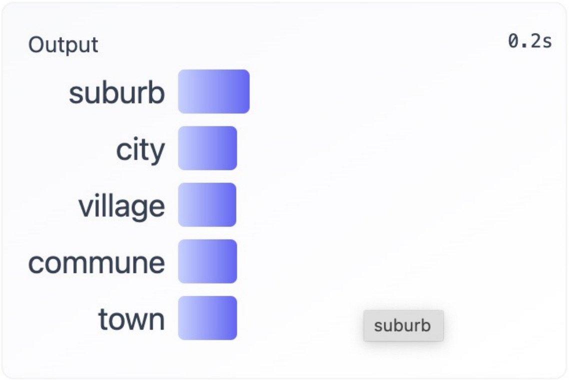
Weak Supervision

- Mark joined Amazon a month ago.
  - What is the entity type?
- Weak Supervision:
  - From knowledge bases
    - Amazon.com, Inc is an American multinational technology company.
  - From weak but richer label representations
    - Word-embedding(company) is close to Word-embedding(Amazon)
  - From pre-trained LMs
    - Amazon is a [MASK] <- [MASK] = company

Weak Supervision from PLMs

- Pre-trained language models can also be used as weak supervision
  - It did not use additional annotations
  - It is not task-specific
  - It contains inductive biases (weak signals)
- PLMs are applied for IE in many creative ways
  - Contextual embeddings to replace word embeddings
  - Direct probing
  - Direct probing + task-specific finetuning
  - Task-specific finetuning (not covered)

Paris is a [MASK].

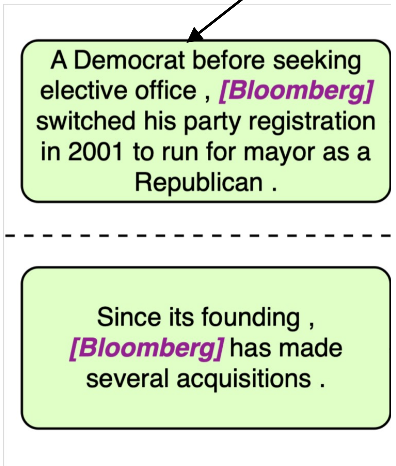


Using pre-trained LM Representations

- Zhou et al. (2018): One of the earliest attempts
  - Entity typing with pre-trained contextualized LM representations

Mentions with high contextual similarities (ELMo embeddings)

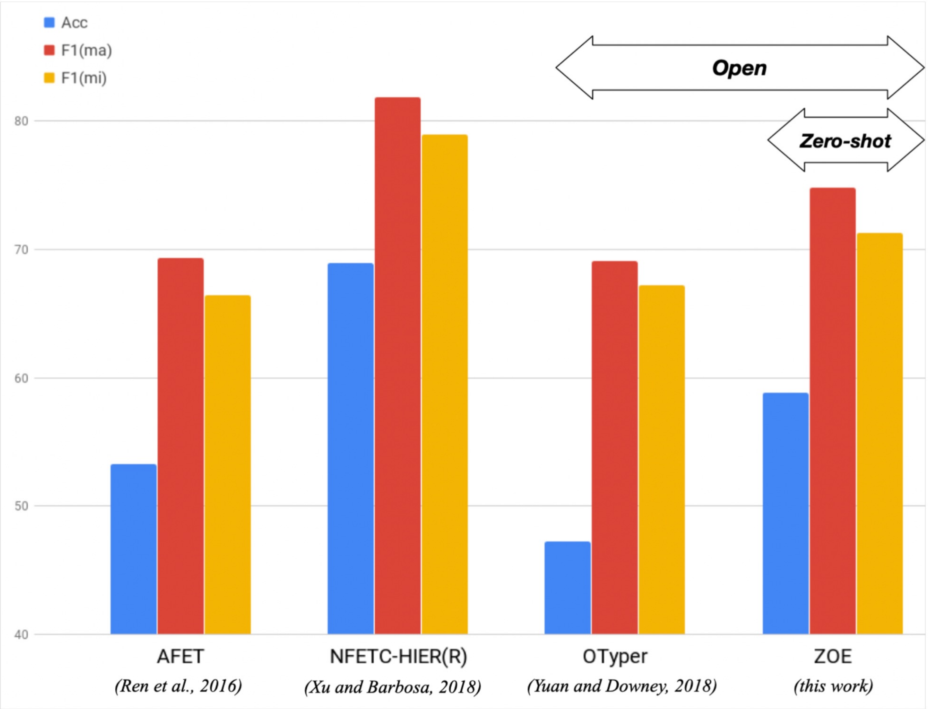
WikiLinks (free-to-use public resource):
- Bill Clinton is the Democratic nominee for president in the 2016 presidential election
- …Scott Brown, a little-known Republican Massachusetts state senator…

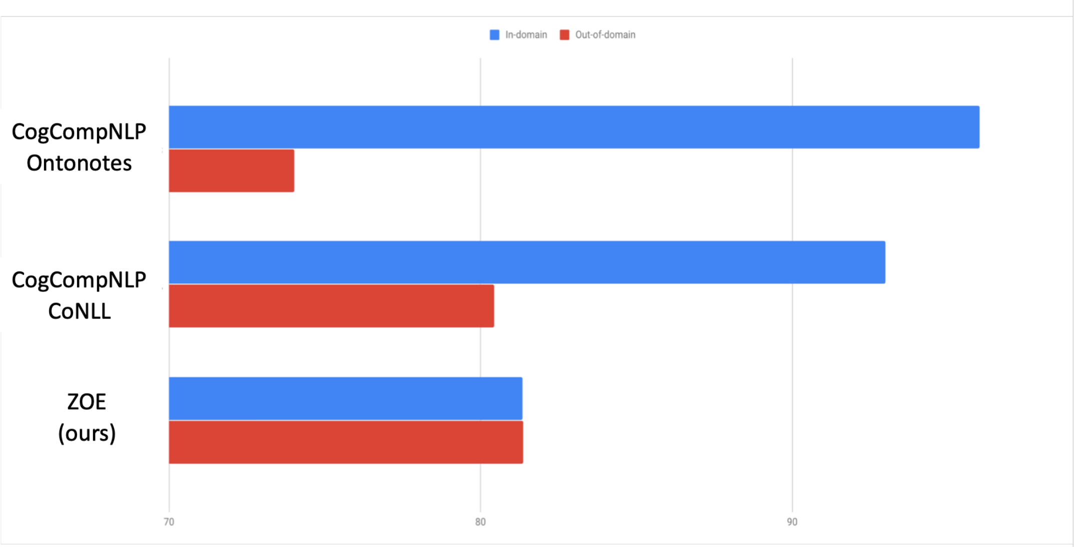


Zhou et al. Zero-Shot Open Entity Typing as Type-Compatible Grounding. EMNLP 2018

Using pre-trained LM Representations

› Zhou et al. (2018): entity typing with LM representation + Wikipedia

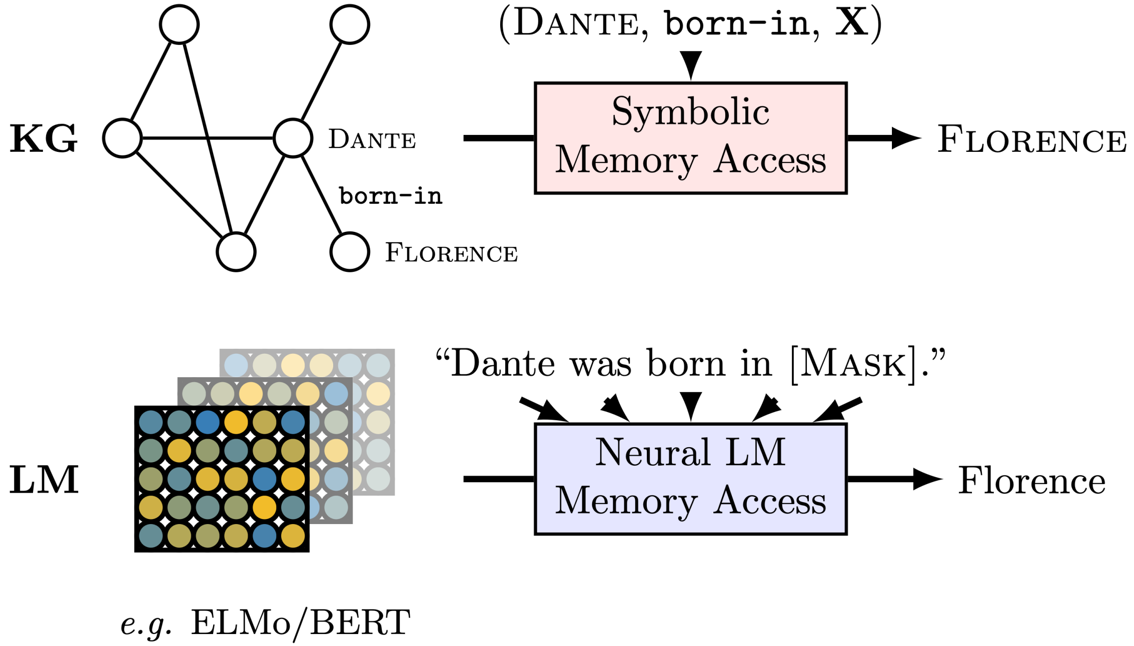




Zhou et al. Zero-Shot Open Entity Typing as Type-Compatible Grounding. EMNLP 2018

Probing pre-trained LMs

- Comparing to ELMo, BERT made direct probing easier
- Petroni et al. (2019): Language models as knowledge bases
  - Google-RE
  - 16.1% birth-place
  - 1.4% birth-date



Probing pre-trained LMs

› Petroni et al. (2019): Language models as knowledge bases

	Relation	Query	Answer	Generation
T-Rex	P19	Francesco Bartolomeo Conti was born in ____.	Florence	Rome [-1.8] , **Florence [-1.8]** , Naples [-1.9] , Milan [-2.4] , Bologna [-2.5]
	P20	Adolphe Adam died in ____.	Paris	**Paris [-0.5]** , London [-3.5] , Vienna [-3.6] , Berlin [-3.8] , Brussels [-4.0]
	P279	English bulldog is a subclass of ____.	dog	dogs [-0.3] , breeds [-2.2] , **dog [-2.4]** , cattle [-4.3] , sheep [-4.5]
	P37	The official language of Mauritius is ____.	English	**English [-0.6]** , French [-0.9] , Arabic [-6.2] , Tamil [-6.7] , Malayalam [-7.0]
	P413	Patrick Oboya plays in ____ position.	midfielder	centre [-2.0] , center [-2.2] , **midfielder [-2.4]** , forward [-2.4] , midfield [-2.7]
	P138	Hamburg Airport is named after ____.	Hamburg	Hess [-7.0] , Hermann [-7.1] , Schmidt [-7.1] , **Hamburg [-7.5]** , Ludwig [-7.5]
	P364	The original language of Mon oncle Benjamin is ____.	French	**French [-0.2]** , Breton [-3.3] , English [-3.8] , Dutch [-4.2] , German [-4.9]
	P54	Dani Alves plays with ____ .	Barcelona	Santos [-2.4] , Porto [-2.5] , Sporting [-3.1] , Brazil [-3.3] , Portugal [-3.7]
	P106	Paul Toungui is a ____ by profession .	politician	lawyer [-1.1] , journalist [-2.4] , teacher [-2.7] , doctor [-3.0] , physician [-3.7]
	P527	Sodium sulfide consists of ____.	sodium	water [-1.2] , sulfur [-1.7] , **sodium [-2.5]** , zinc [-2.8] , salt [-2.9]
	P102	Gordon Scholes is a member of the ____ political party.	Labor	Labour [-1.3] , Conservative [-1.6] , Green [-2.4] , Liberal [-2.9] , **Labor [-2.9]**
	P530	Kenya maintains diplomatic relations with ____.	Uganda	India [-3.0] , **Uganda [-3.2]** , Tanzania [-3.5] , China [-3.6] , Pakistan [-3.6]
	P176	iPod Touch is produced by ____.	Apple	**Apple [-1.6]** , Nokia [-1.7] , Sony [-2.0] , Samsung [-2.6] , Intel [-3.1]
	P30	Bailey Peninsula is located in ____.	Antarctica	**Antarctica [-1.4]** , Bermuda [-2.2] , Newfoundland [-2.5] , Alaska [-2.7] , Canada [-3.1]
	P178	JDK is developed by ____.	Oracle	IBM [-2.0] , Intel [-2.3] , Microsoft [-2.5] , HP [-3.4] , Nokia [-3.5]
	P1412	Carl III used to communicate in ____.	Swedish	German [-1.6] , Latin [-1.9] , French [-2.4] , English [-3.0] , Spanish [-3.0]
	P17	Sunshine Coast, British Columbia is located in ____.	Canada	**Canada [-1.2]** , Alberta [-2.8] , Yukon [-2.9] , Labrador [-3.4] , Victoria [-3.4]
	P39	Pope Clement VII has the position of ____ .	pope	cardinal [-2.4] , Pope [-2.5] , **pope [-2.6]** , President [-3.1] , Chancellor [-3.2]
	P264	Joe Cocker is represented by music label ____.	Capitol	EMI [-2.6] , BMG [-2.6] , Universal [-2.8] , **Capitol [-3.2]** , Columbia [-3.3]
	P276	London Jazz Festival is located in ____.	London	**London [-0.3]** , Greenwich [-3.2] , Chelsea [-4.0] , Camden [-4.6] , Stratford [-4.8]
	P127	Border TV is owned by ____.	ITV	Sky [-3.1] , **ITV [-3.3]** , Global [-3.4] , Frontier [-4.1] , Disney [-4.3]
	P103	The native language of Mammootty is ____.	Malayalam	**Malayalam [-0.2]** , Tamil [-2.1] , Telugu [-4.8] , English [-5.2] , Hindi [-5.6]
	P495	The Sharon Cuneta Show was created in ____.	Philippines	Manila [-3.2] , **Philippines [-3.6]** , February [-3.7] , December [-3.8] , Argentina [-4.0]
ConceptNet	AtLocation	You are likely to find a overflow in a ____.	drain	sewer [-3.1] , canal [-3.2] , toilet [-3.3] , stream [-3.6] , **drain [-3.6]**
	CapableOf	Ravens can ____.	fly	**fly [-1.5]** , fight [-1.8] , kill [-2.2] , die [-3.2] , hunt [-3.4]
	CausesDesire	Joke would make you want to ____.	laugh	cry [-1.7] , die [-1.7] , **laugh [-2.0]** , vomit [-2.6] , scream [-2.6]
	Causes	Sometimes virus causes ____.	infection	disease [-1.2] , cancer [-2.0] , **infection [-2.6]** , plague [-3.3] , fever [-3.4]
	HasA	Birds have ____.	feathers	wings [-1.8] , nests [-3.1] , **feathers [-3.2]** , died [-3.7] , eggs [-3.9]
	HasPrerequisite	Typing requires ____.	speed	patience [-3.5] , precision [-3.6] , registration [-3.8] , accuracy [-4.0] , **speed [-4.1]**
	HasProperty	Time is ____.	finite	short [-1.7] , passing [-1.8] , precious [-2.9] , irrelevant [-3.2] , gone [-4.0]
	MotivatedByGoal	You would celebrate because you are ____.	alive	happy [-2.4] , human [-3.3] , **alive [-3.3]** , young [-3.6] , free [-3.9]
	ReceivesAction	Skills can be ____.	taught	acquired [-2.5] , useful [-2.5] , learned [-2.8] , combined [-3.9] , varied [-3.9]
	UsedFor	A pond is for ____.	fish	swimming [-1.3] , fishing [-1.4] , bathing [-2.0] , **fish [-2.8]** , recreation [-3.1]

These predictions are highly relevant to typing and relation extraction

Petroni et al. Language Models as Knowledge Bases?. EMNLP 2019

Probing pre-trained LMs for IE

- Dai et al. (2021)
- Use templates + [MASK] to retrieve entity types

Pattern	F1
M and any other H	25.3
M and some other H	24.8
H such as M	20.7
such H as M	18.1
H including M	17.4
H especially M	11.5

Input	Top Words for [MASK]
In late 2015, [MASK] such as **Leonardo DiCaprio** starred in The Revenant.	actors, stars, actor, directors, filmmakers
At some clinics, **they** and some other [MASK] are told the doctors don't know how to deal with AIDS, and to go someplace else.	patients, people, doctors, kids, children
Finkelstein says he expects the company to "benefit from some of the disruption faced by **our competitors** and any other [MASK] ."	company, business, companies, group, investors

Dai et al. Ultra-Fine Entity Typing with Weak Supervision from a Masked Language Model. ACL 2021

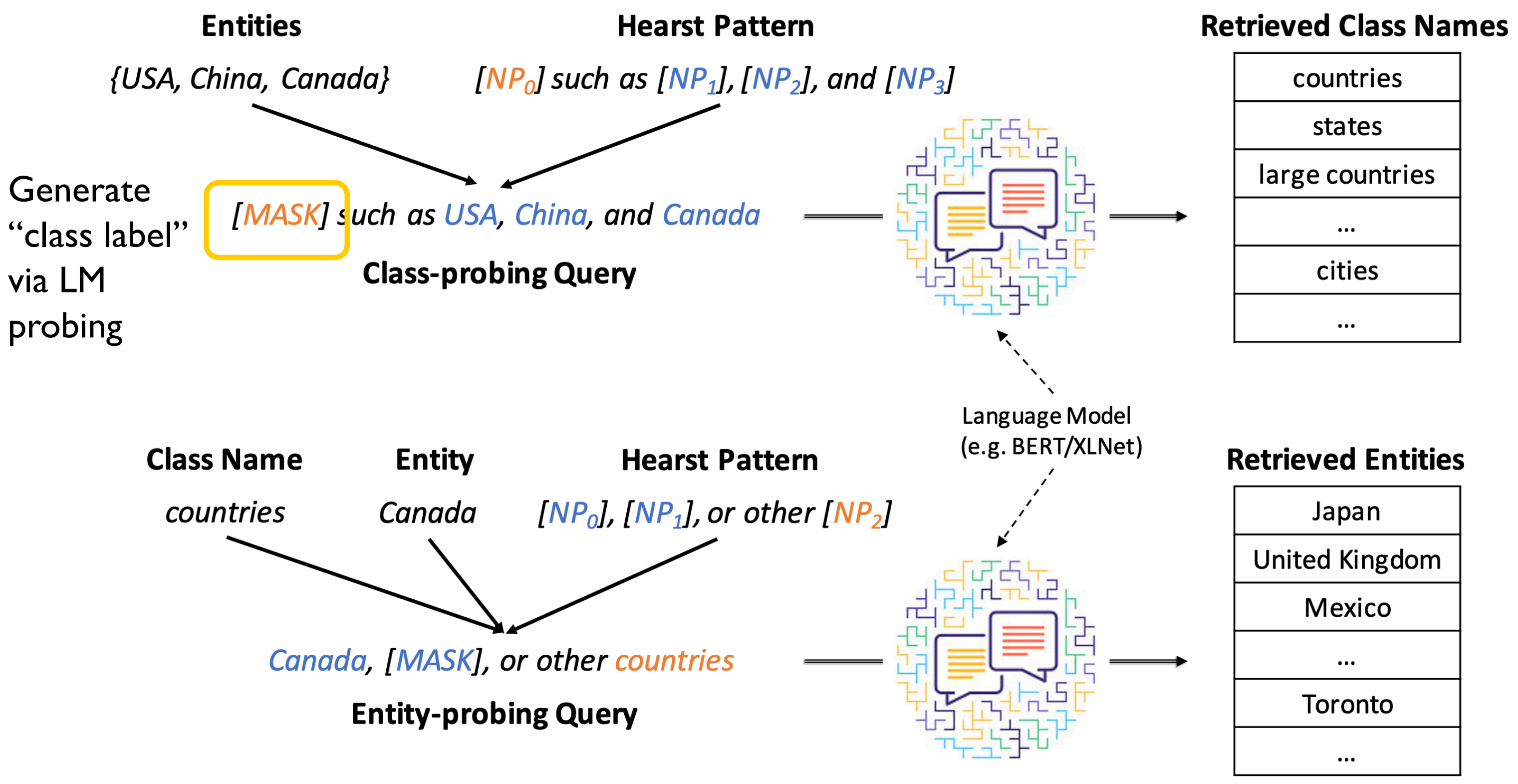
Probing pre-trained LMs for IE

- Zhang et al. (2020)
- Entity Set Expansion: expand a small set of entities with new ones belonging to the same semantic class.
  - {"United States", "China", "Canada} –>{"Japan", "Mexico"}
  - A "entity set" fine-grained typing

Zhang et al. Empower Entity Set Expansion via Language Model Probing. ACL 2020

Probing pre-trained LMs for IE

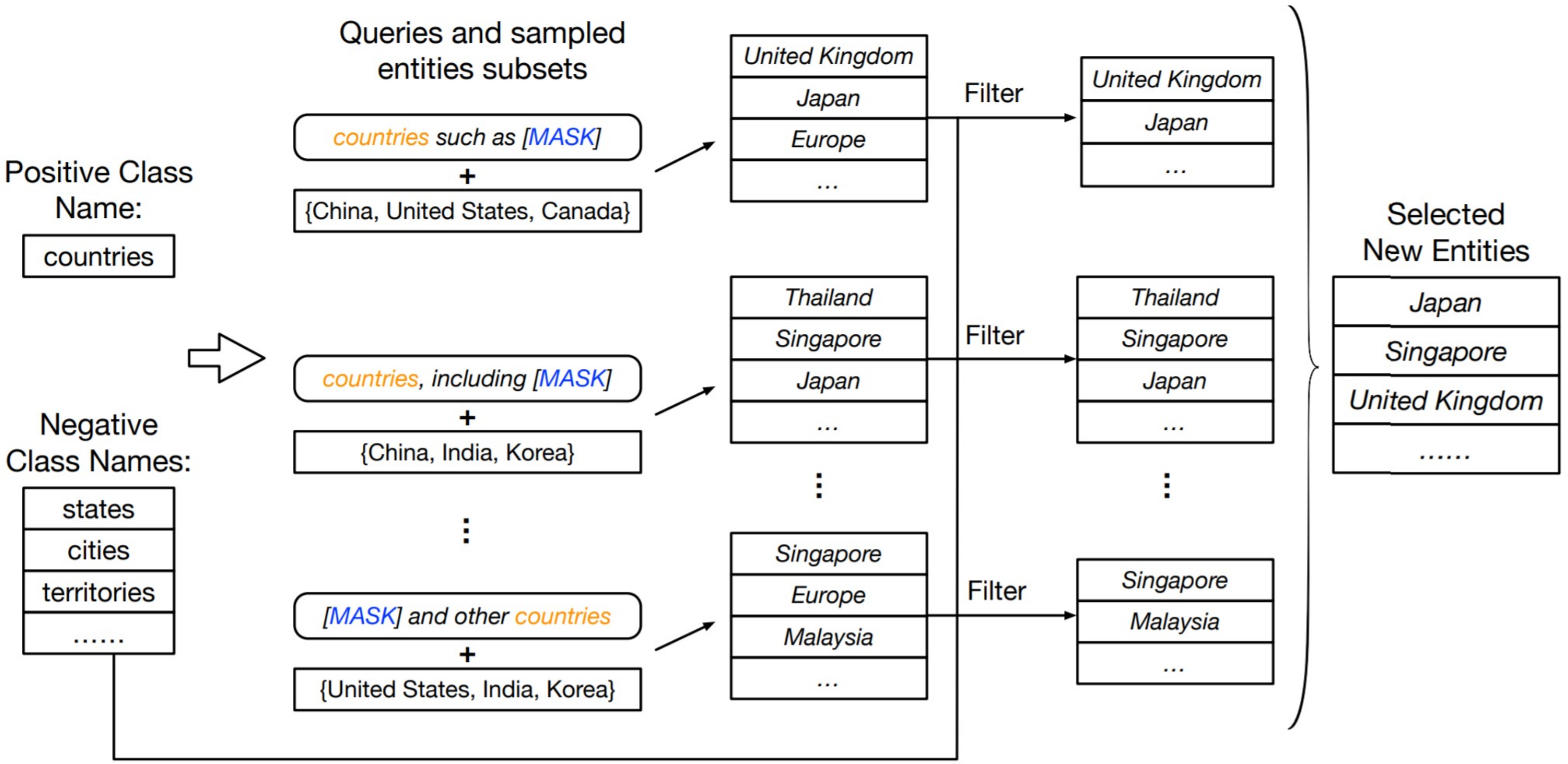
- Zhang et al. (2020)
- Entity set expansion by probing pre-trained LMs



Zhang et al. Empower Entity Set Expansion via Language Model Probing. ACL 2020

Probing pre-trained LMs for IE

- Zhang et al. (2020)
- Entity set expansion by probing pre-trained LMs



Zhang et al. Empower Entity Set Expansion via Language Model Probing. ACL 2020

Probing pre-trained LMs for IE

- Zhang et al. (2020)
- Entity set expansion by probing pre-trained LMs

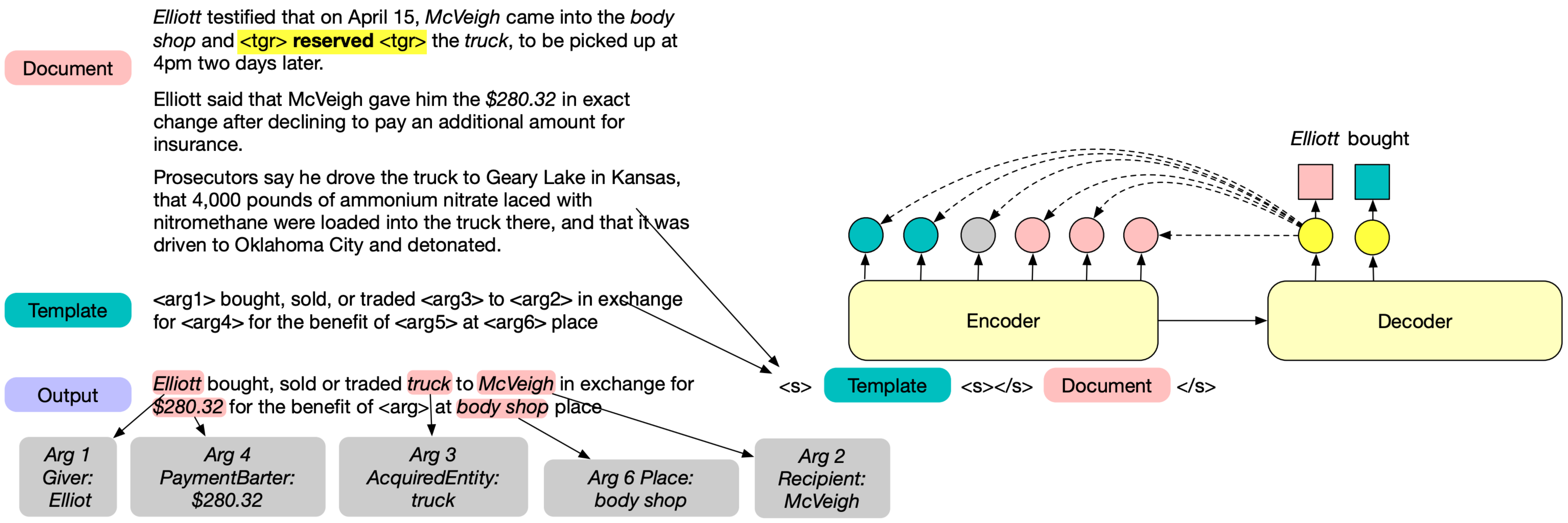
Methods	Wiki MAP@10	Wiki MAP@20	Wiki MAP@50	APR MAP@10	APR MAP@20	APR MAP@50
Egoset (Rong et al., 2016)	0.904	0.877	0.745	0.758	0.710	0.570
SetExpan (Shen et al., 2017)	0.944	0.921	0.720	0.789	0.763	0.639
SetExpander (Mamou et al., 2018)	0.499	0.439	0.321	0.287	0.208	0.120
CaSE (Yu et al., 2019b)	0.897	0.806	0.588	0.619	0.494	0.330
MCTS (Yan et al., 2019)	0.980^{∇}	0.930^{∇}	0.790^{∇}	0.960^{∇}	0.900^{∇}	0.810^{∇}
CGExpan-NoCN	0.968	0.945	0.859	0.909	0.902	0.787
CGExpan-NoFilter	0.990	0.975	0.890	0.979	0.962	0.892
CGExpan	**0.998**	**0.981**	**0.893**	**0.992**	**0.990**	**0.955**

Table 2: Mean Average Precision on Wiki and APR. "∇" means the number is directly from the original paper.

Zhang et al. Empower Entity Set Expansion via Language Model Probing. ACL 2020

Probing pre-trained LMs through Generation

- LMs that are trained with autoregressive structures can be probed through conditional generation (with some supervision)
- Li et al. (2021): Event argument extraction via BART + conditional generation



Li et al. Document-Level Event Argument Extraction by Conditional Generation. NAACL 2021

Weak Supervision

- Mark joined Amazon a month ago.
  - What is the entity type?
- Weak Supervision:
  - From knowledge bases
    - Amazon.com, Inc is an American multinational technology company.
  - From weak but richer label representations
    - Word-embedding(company) is close to Word-embedding(Amazon)
  - From pre-trained LMs
    - Amazon is a [MASK] <- [MASK] = company.
  - From linguistic patterns
    - PER *join* company

Weak Supervision – Linguistic Patterns

- We can use simple linguistic patterns to mine many relations from text

[event] at [time]

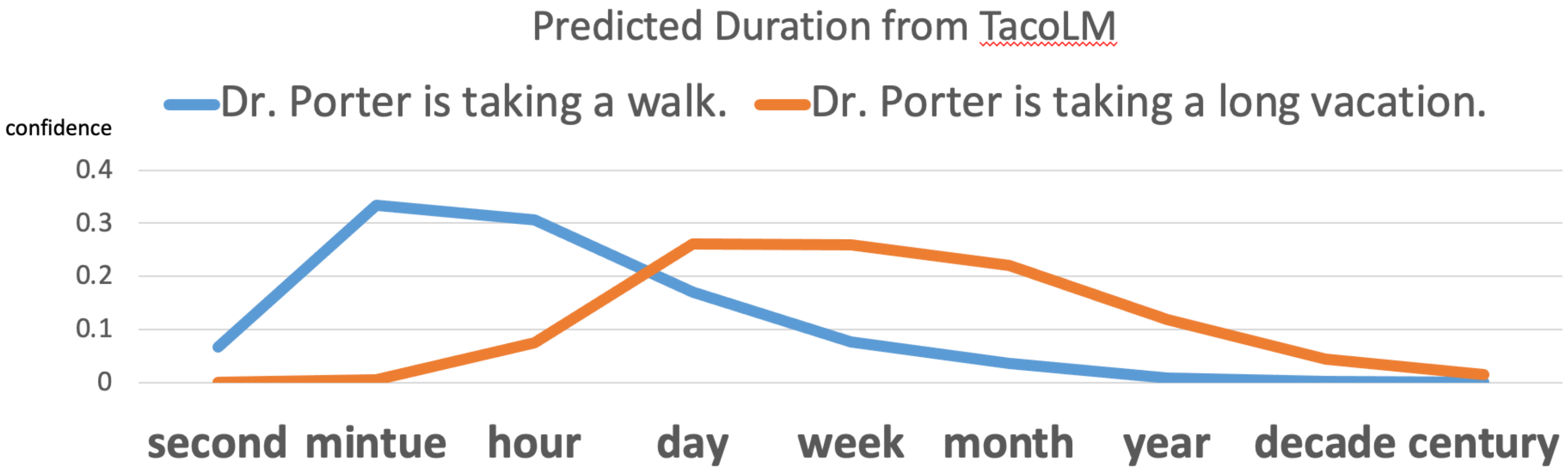
[event] because [event]

[entity] is [entity]

Works better with events, as entities would lose contextual information.

Weak Supervision – Linguistic Patterns

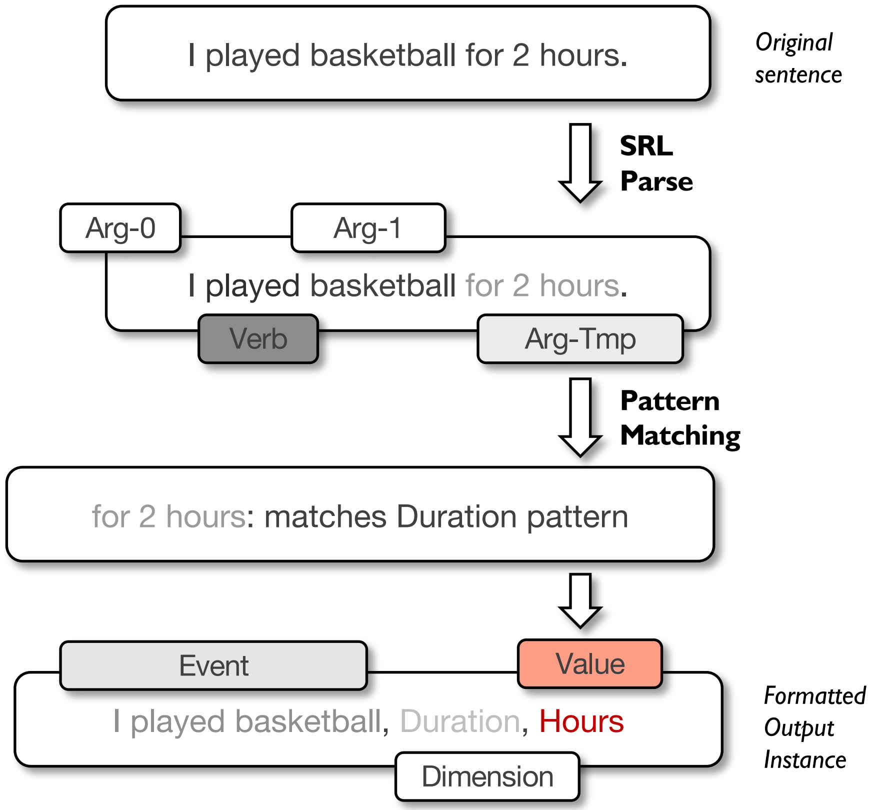
- Zhou et al. (2020): Temporal Information Extraction from Patterns
- Goal: model events' temporal property distributions
  - Duration, Frequency, Typical Time



Zhou et al. Temporal Common Sense Acquisition with Minimal Supervision. ACL 2020

Weak Supervision – Linguistic Patterns

› Zhou et al. (2020): Temporal Information Extraction from Patterns
  » Step 1: Extract distant signals of contextualized events and their duration, frequency etc. via linguistic patterns



Zhou et al. Temporal Common Sense Acquisition with Minimal Supervision. ACL 2020

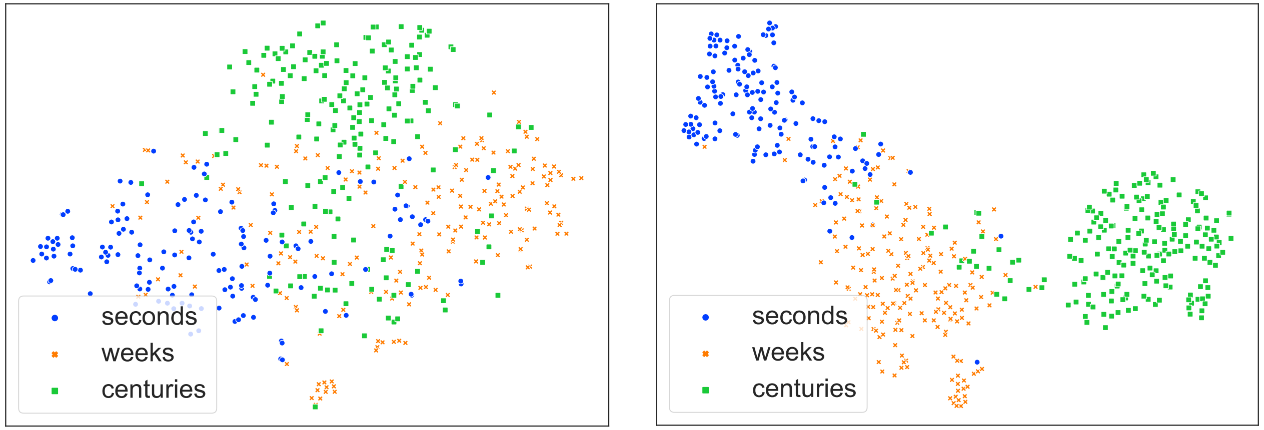
Weak Supervision – Linguistic Patterns

- Zhou et al. (2020): Temporal Information Extraction from Patterns
  - Step 1: Extract distant signals of contextualized events and their duration, frequency etc. via linguistic patterns
  - Step 2: further pre-train a language model with extracted instances

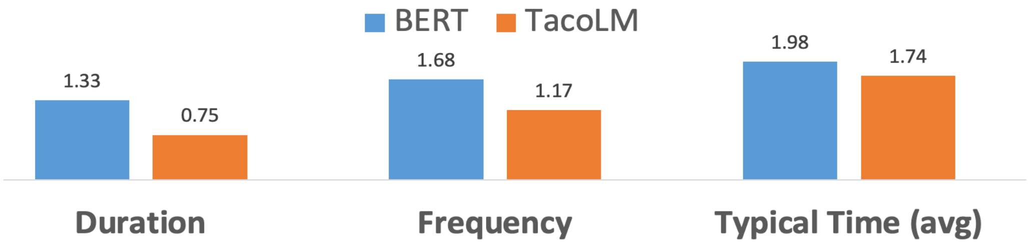
Zhou et al. Temporal Common Sense Acquisition with Minimal Supervision. ACL 2020

Weak Supervision – Linguistic Patterns

› Zhou et al. (2020): Temporal Information Extraction from Patterns



BERT TacoLM



Distance to gold value (lower the better)

Zhou et al. Temporal Common Sense Acquisition with Minimal Supervision. ACL 2020

Weak Supervision – Linguistic Patterns

- Zhou et al. (2021): Temporal relation extraction from patterns
  - event-event before/after relation

- Within-sentence extraction
  - Not enough:
    - LMs may know this already
    - Does not tell how far the two start times are

text

I went to the park on January 1st. I was very hungry after some hiking. Luckily, I purchased a lot of food before I went to the park. I enjoyed the trip and wrote an online review about the trip on the 10th.

within-sentence

[I purchased food, I went to the park.]: **before**

cross-sentence

[I went to the park, I wrote a review]: **before**, weeks

Zhou et al. Temporal Common Sense Acquisition with Minimal Supervision. ACL 2020

Weak Supervision – Linguistic Patterns

- Zhou et al. (2021): Temporal relation extraction from patterns
  - Automatically extracts weak supervision instances from unannotated texts

- Cross-sentence extraction
  - Based on explicit temporal expressions
  - Independent of event locations
  - Produces relative distance between start times

text

I went to the park on January 1st. I was very hungry after some hiking. Luckily, I purchased a lot of food before I went to the park. I enjoyed the trip and wrote an online review about the trip on the 10th.

within-sentence

[I purchased food, I went to the park.]: **before**

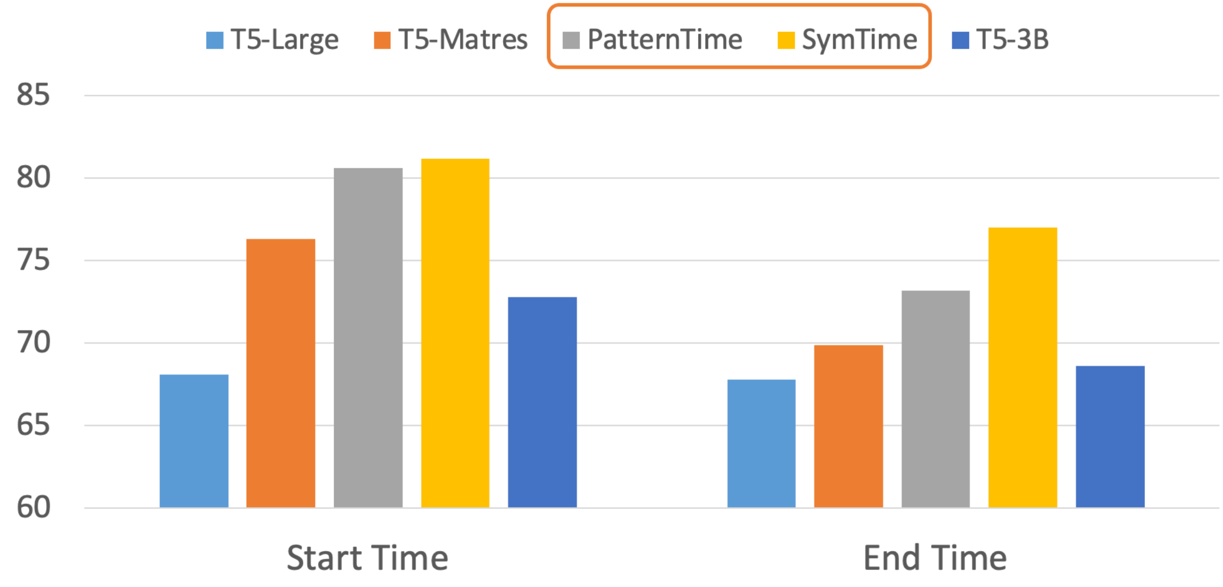
cross-sentence

[I went to the park, I wrote a review]: **before**, weeks

Zhou et al. Temporal Common Sense Acquisition with Minimal Supervision. ACL 2020

Weak Supervision – Linguistic Patterns

- Zhou et al. (2021): Temporal relation extraction from patterns
  - Automatically extracts weak supervision instances from unannotated texts
- Evaluation done on TRACIE (from the same paper)
  - Evaluates temporal relation of both start and end time



SymTime: explicitly computes end time with the start time and duration estimations from PatternTime (detail omitted)

Zhou et al. Temporal Common Sense Acquisition with Minimal Supervision. ACL 2020

Weak Supervision

- Mark joined Amazon a month ago.
  - What is the entity type?
- Weak Supervision:
  - From knowledge bases
    - Amazon.com, Inc is an American multinational technology company.
  - From weak but richer label representations
    - Word-embedding(company) is close to Word-embedding(Amazon)
  - From pre-trained LMs
    - Amazon is a [MASK] <- [MASK] = company.
  - From linguistic patterns
    - PER *join* company

Similarity: None of them directly reveals the type in the given context, but they all hint/suggest it.

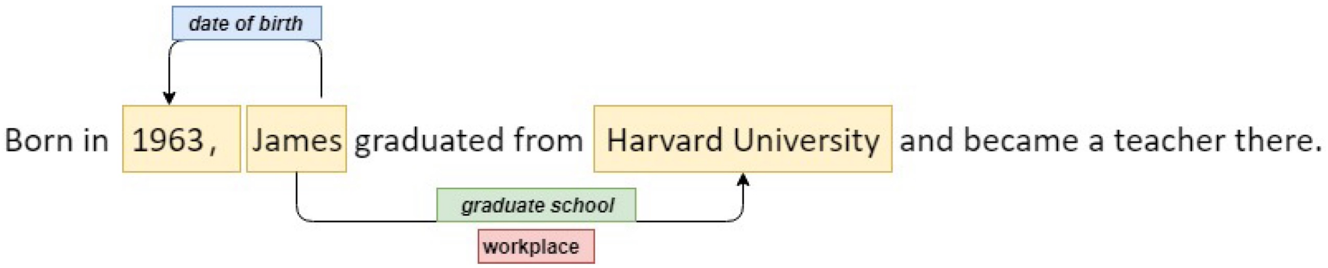
Alternative Supervision Sources

- **Weak supervision**: from task-related distant signals
  - Easy to acquire
  - Task-specific
  - May be noisy
- **Indirect supervision**: from other tasks
  - Human annotations from other popular tasks such as NLI and QA
  - Non-task-specific
  - Needs clever ways to be applied

Indirect Supervision for IE

- Intuition: Information extraction tasks can benefit from <u>other tasks</u>'
  - Formulation
  - Supervision
- Comparing to weak supervision:
  - clean human annotations
  - good baselines

Direct IE:
(James, Harvard),
relation, graduate_from



Indirect IE from QA:
Q: Where did James graduate from?
A: Harvard University

LMs use natural languages better..

we may train on Squad first…

Indirect Supervision for IE

- Part 1: Indirect Supervision from task formulation
  - Transform an IE task to another task
  - Not use additional supervision (even though indirect) to make fair comparisons
  - Question answering
    - Relation extraction
    - Named entity recognition

Indirect Supervision from QA

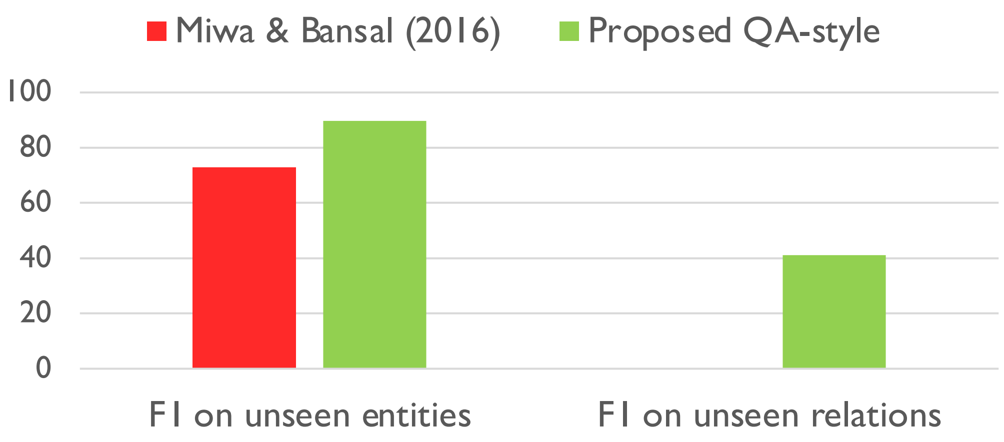
› Levy et al. (2017): Relation extraction formulated as QA

Relation	Question	Sentence & Answers
educated_at	What is **Albert Einstein**'s alma mater?	**Albert Einstein** was awarded a PhD by the **University of Zürich**, with his dissertation titled...
occupation	What did **Steve Jobs** do for a living?	**Steve Jobs** was an American **businessman**, **inventor**, and **industrial designer**.
spouse	Who is **Angela Merkel** married to?	**Angela Merkel**'s second and current husband is quantum chemist and professor **Joachim Sauer**, who has largely...

Schema Querification (crowdsourced)

» Why would it work?

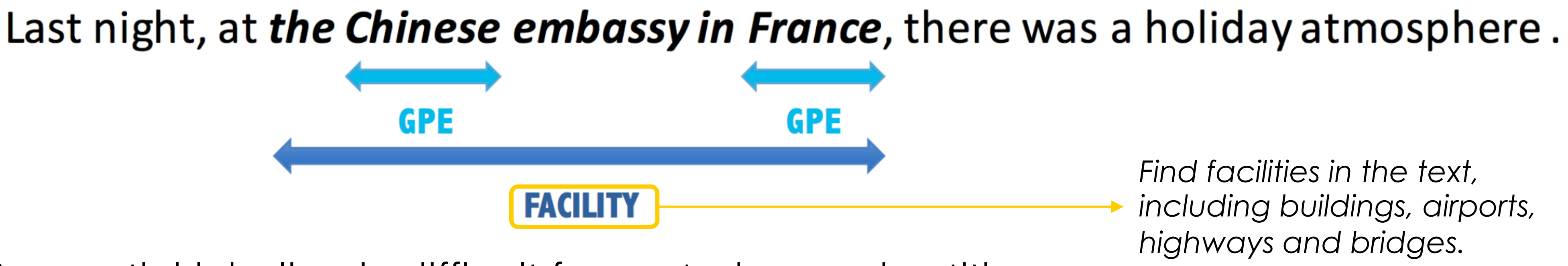
» Question provides "indirect" information on relation labels



Levy et al. Zero-Shot Relation Extraction via Reading Comprehension. CoNLL 2017

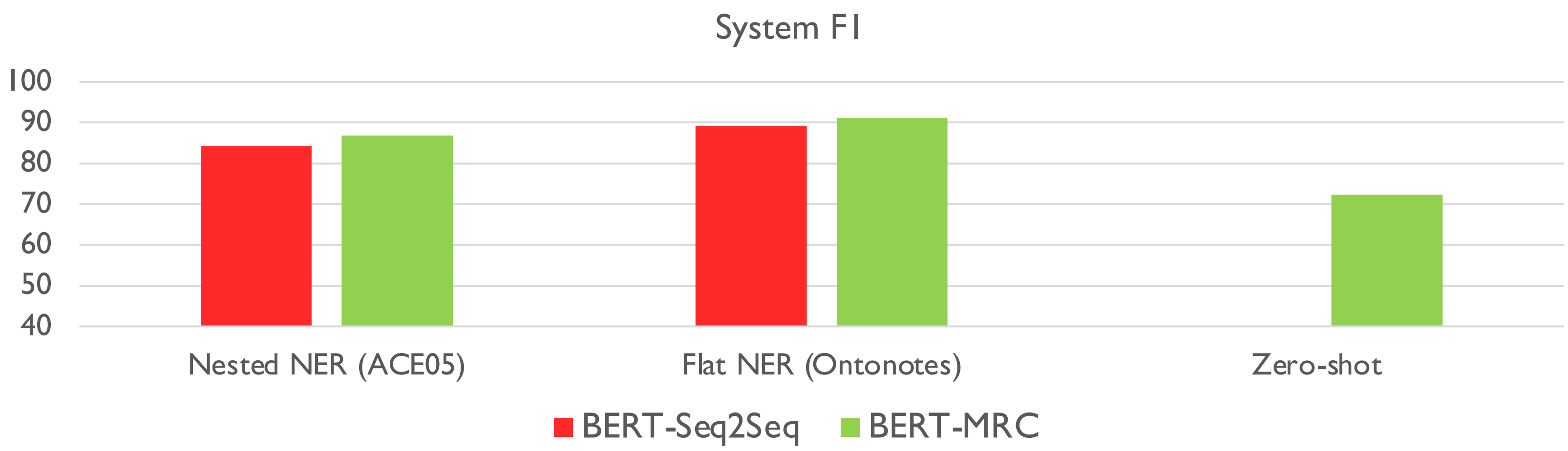
Indirect Supervision from QA

› Li et al. (2020): NER formulated as QA



» Sequential labeling is difficult for nested named entities

» Formulate as span-selection QA task will help



Li et al. A Unified MRC Framework for Named Entity Recognition. ACL 2020

Indirect Supervision from QA

- Li et al. (2020): NER as QA, where does the improvement come from?
  - Questions serve as "label definitions", provides additional indirect supervision

Entity	**Natural Language Question**
Location	Find locations in the text, including non-geographical locations, mountain ranges and bodies of water.
Facility	Find facilities in the text, including buildings, airports, highways and bridges.
Organization	Find organizations in the text, including companies, agencies and institutions.

Li et al. A Unified MRC Framework for Named Entity Recognition. ACL 2020

Indirect Supervision from QA

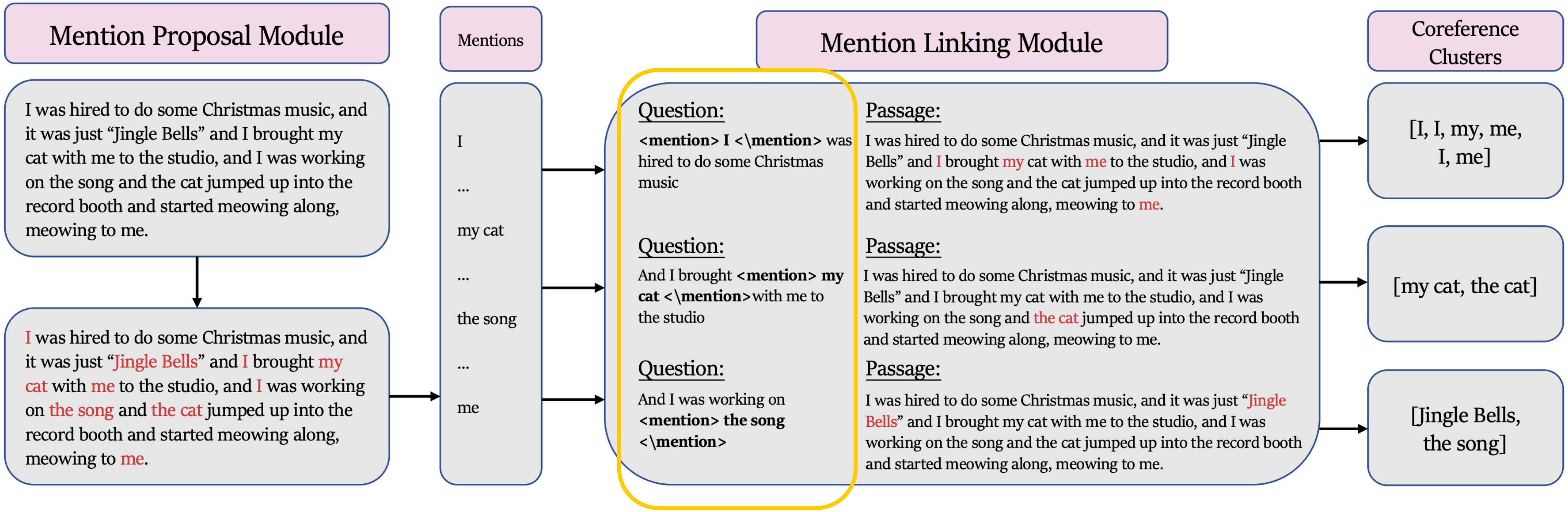
- Li et al. (2020): NER as QA, where does the improvement come from?
  - Pre-trained language models understand "natural language" better than "labels"

English OntoNotes 5.0	
Model	**F1**
BERT-Tagger	89.16
Position index of labels	88.29 (-0.87)
Keywords	89.74 (+0.58)
Wikipedia	89.66 (+0.59)
Rule-based template filling	89.30 (+0.14)
Synonyms	89.92 (+0.76)
Keywords+Synonyms	90.23 (+1.07)
Annotation guideline notes	91.11 (+1.95)

Li et al. A Unified MRC Framework for Named Entity Recognition. ACL 2020

Indirect Supervision from QA

› Wu et al. (2020): Coreference as QA

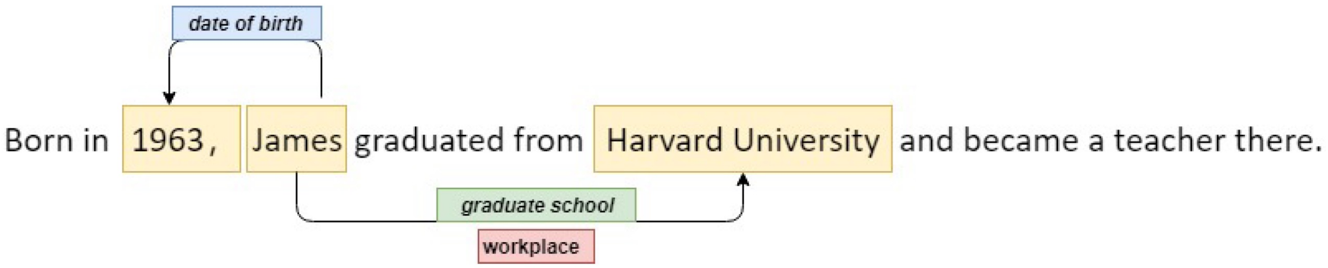


Use the sentence that each mention is in as the "question", all other spans belonging to the same cluster as "answers"

Wu et al. CorefQA: Coreference Resolution as Query-based Span Prediction. ACL 2020

Indirect Supervision for IE

- Part 2: Indirect Supervision from task formulation + supervision
  - Transform an IE task to another task, which has a representation that's easier for models
  - Use additional supervision from the original task format (e.g., QA, NLI)

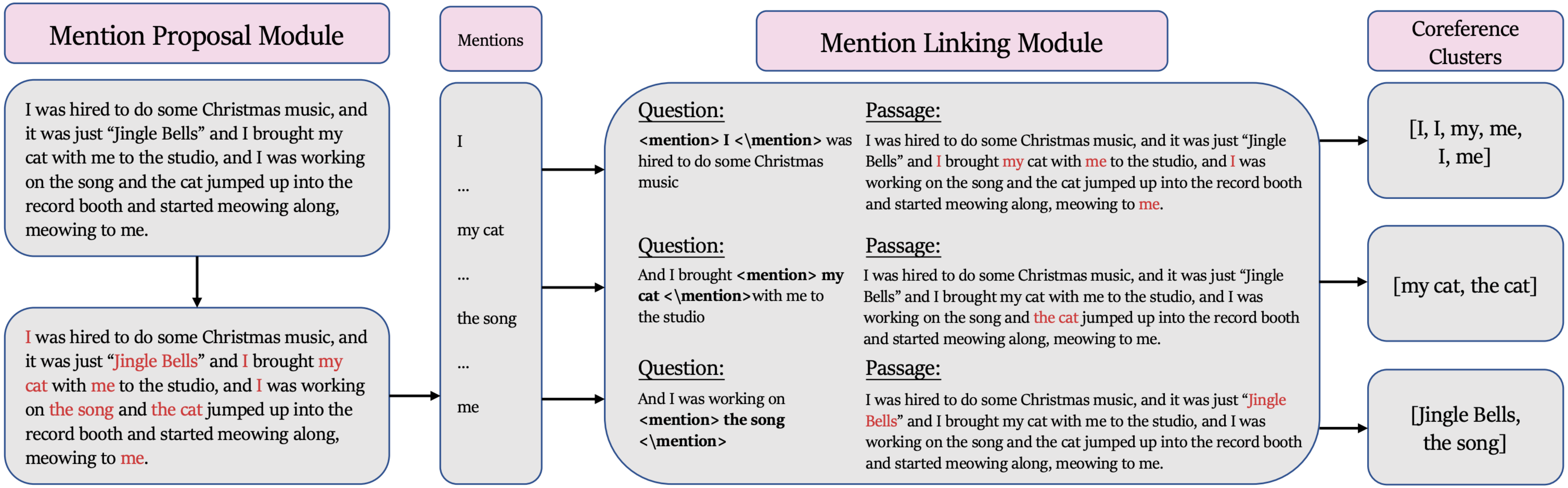


Indirect IE from QA:
Q: Where did James graduate from?
A: Harvard University

we may train on Squad first…

Indirect Supervision from QA

› Wu et al. (2020): Coreference as QA

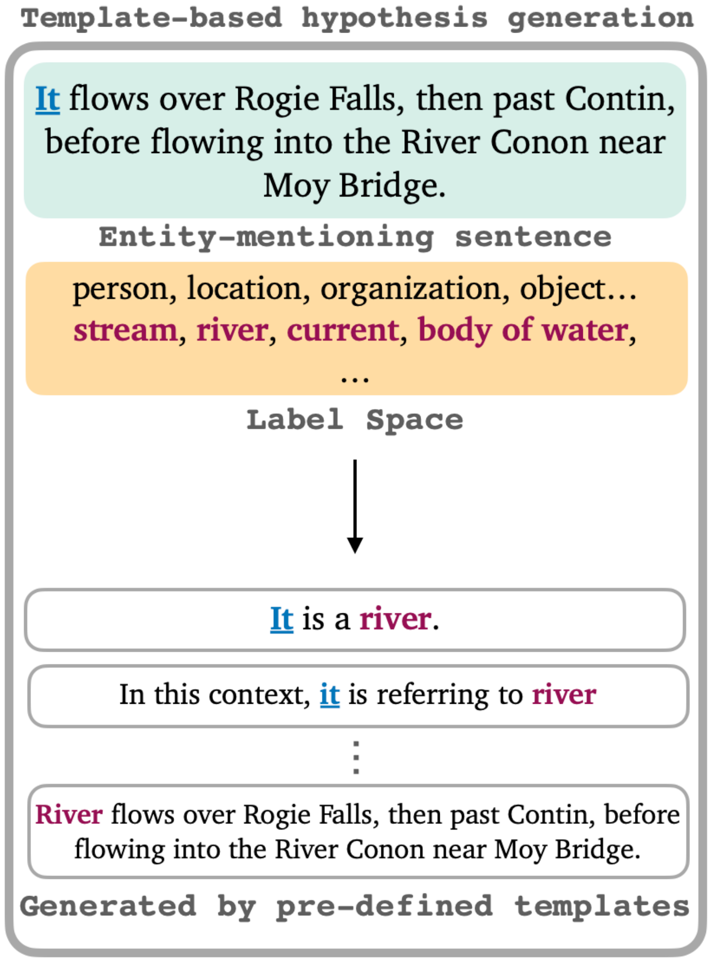


› Pre-train on Quoref + SQuAD improves ~1%, while the overall system improves 3.5%

Wu et al. CorefQA: Coreference Resolution as Query-based Span Prediction. ACL 2020

Indirect Supervision from NLI

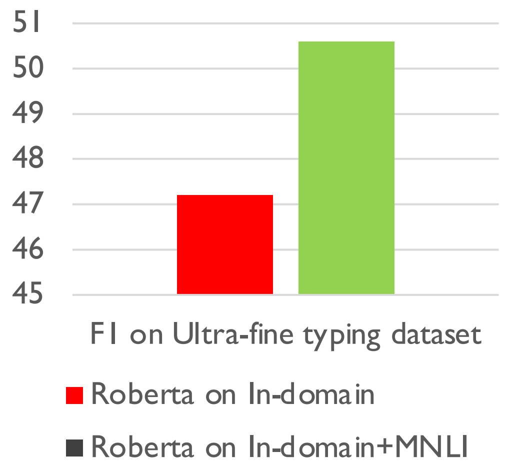
› Li et al. (2022): Entity typing formulated as textual entailment (NLI)

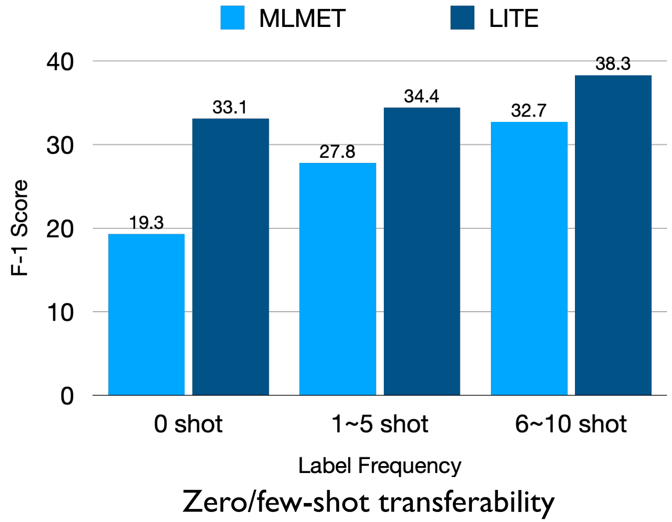


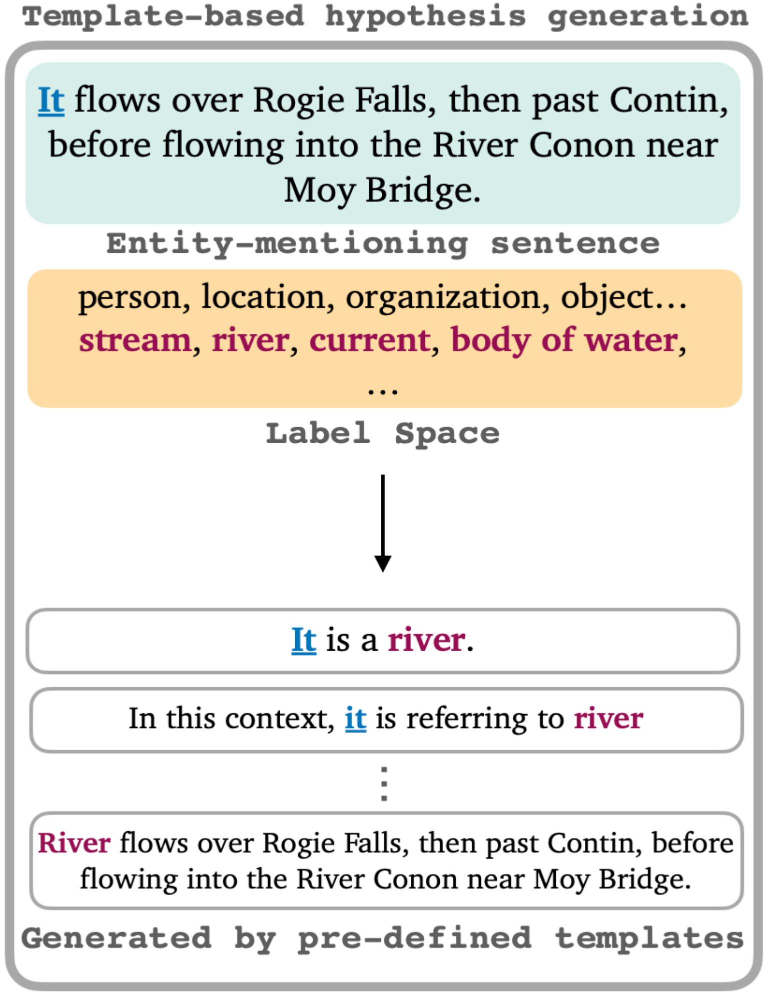
Li et al. Ultra-fine Entity Typing with Indirect Supervision from Natural Language Inference. TACL 2022

Indirect Supervision from NLI

- Li et al. (2021): Entity typing formulated as textual entailment (NLI)
  - “[Entity] is [Label]”
  - “In this context, [Entity] is referring to [Label]”
  - Replace [Entity] with [Label] in original context
- Advantages
  - “Natural language” representation
  - Existing entailment dataset transfers well
  - Open label space







Li et al. Ultra-fine Entity Typing with Indirect Supervision from Natural Language Inference. TACL 2022

Indirect Supervision from NLI

- Lyu et al. (2021), Sainz (2022)
  - Event trigger / argument extraction via QA and NLI
  - Better zero/few-shot performances

Event type: TRANSFER-OWNERSHIP

China has *purchased* *two nuclear submarines* from *Russia* *last month*.

Buyer-Arg Trigger Artifact-Arg Seller-Arg Time-Arg

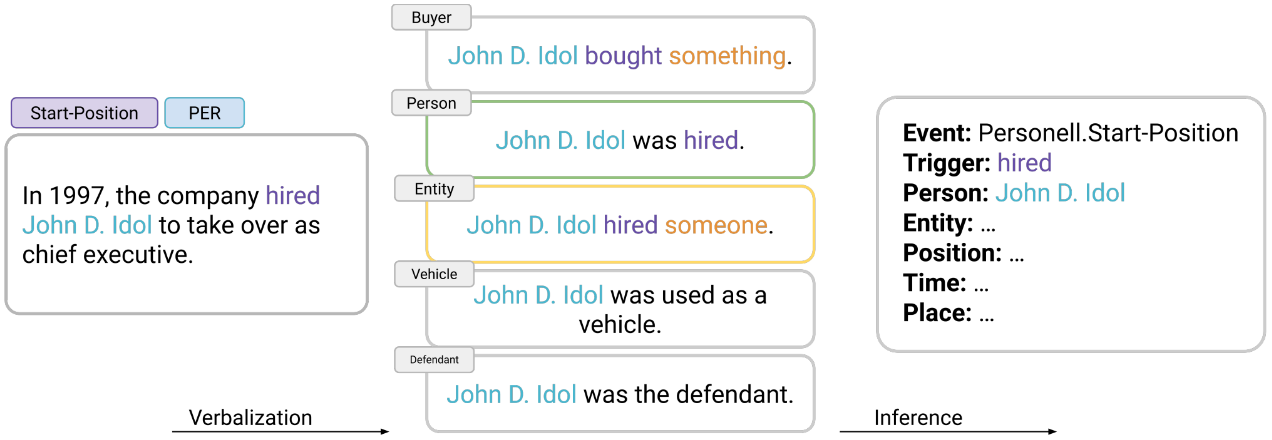
Q$_1$: Who bought something? A$_1$: *China*

Q$_2$: Who sold something? A$_2$: *Russia*

Q$_3$: What is bought? A$_3$: *Two nuclear submarines*

Q$_4$: Where is the purchase? A$_4$: No Answer

Trigger: this text is about a purchase (NLI)

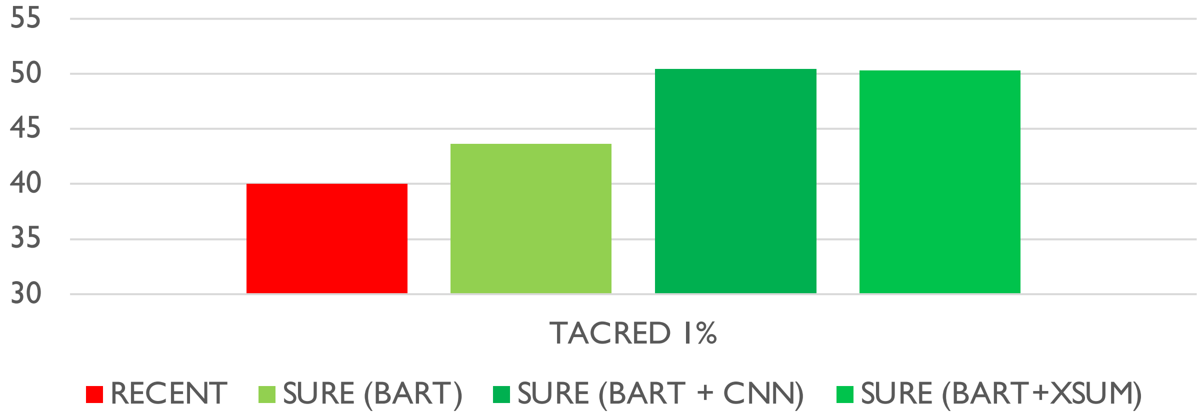


Lyu et al. Zero-shot Event Extraction via Transfer Learning: Challenges and Insights. IJCNLP 2021
Sainz et al. Textual Entailment for Event Argument Extraction: Zero- and Few-Shot with Multi-Source Learning. NAACL 2022

Indirect Supervision with Summarization

- Summarization requires advanced reading comprehension
  - Can also be tailored to specific tasks
- Lu et al. (2022): Relation extraction as summarization
  - Format input subject/object types as natural language
  - Ask a summarization model to generate verbalized relations

Pre-trained on CNN/Dailymail (Hermann et al., 2015) and XSum (Narayan et al., 2021)



Example

Input:

Subject: Mandelbrot **Type:** person

Object: Poland **Type:** country

Sentence:

Mandelbrot was born in Poland but as a child moved to France.

Input Sequence Construction:

The subject is Mandelbrot. The object is Poland. The type of Mandelbrot is person. The type of Poland is country. Mandelbrot was born in Poland but as a child moved to France.

Relation Verbalization:

r_1: Mandelbrot is a Poland

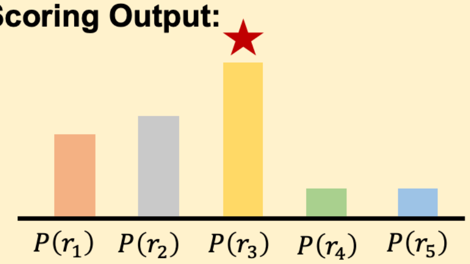
r_2: Mandelbrot was born in the city Poland

r_3: Mandelbrot was born in the country Poland

r_4: Mandelbrot was founded by Poland

r_5: Mandelbrot has no known relations to Poland

Trie Scoring Output:



(Mandelbrot, country of birth, Poland)

Lu et al. Summarization as Indirect Supervision for Relation Extraction. Arxiv 2022

Conclusion and Future Directions

- Direct supervision is not the answer to all problems
- Weak supervision
  - Knowledge bases and dictionaries
  - Label definitions
  - Pre-trained language models
  - Linguistic patterns
- Indirect supervision
  - From other task formulation
  - From other task formulation and supervision
- Future directions:
  - Quantify task-task relations
  - Unified framework