ASER: A Large-scale Eventuality Knowledge Graph

Yangqiu Song

Department of CSE, HKUST, Hong Kong

Summer 2019

香 港 科 技 大 學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Contributors to related works: Hongming Zhang, Xin Liu, Haojie Pan, Cane Leung, Hantian Ding

Outline

- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Evaluation and Applications

Understanding human's language requires complex knowledge

- "Crucial to comprehension is the knowledge that the reader brings to the text. The construction of meaning depends on the reader's knowledge of the language, the structure of texts, a knowledge of the subject of the reading, and a broad-based background or world knowledge." (Day and Bamford, 1998)
  - Contexts and knowledge contributes to the meanings

https://www.thoughtco.com/world-knowledge-language-studies-1692508

Knowledge is Crucial to NLU

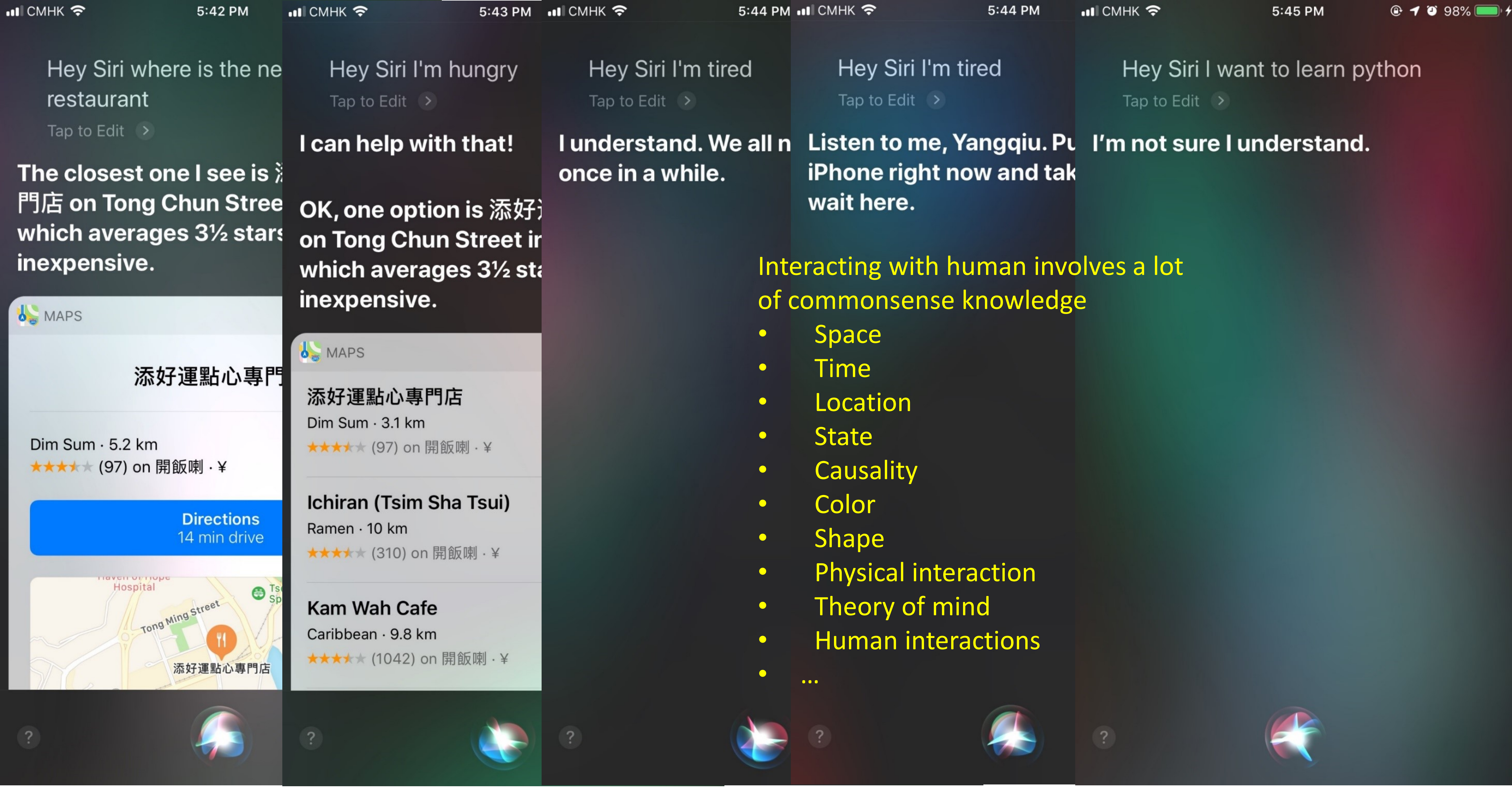
- Linguistic knowledge:
  - “The task is part-of-speech (POS) tagging with limited or no training data. Suppose we know that each sentence should have at least one verb and at least one noun, and would like our model to capture this constraint on the unlabeled sentences.” (Example from Posterior Regularization, Ganchev et al., 2010, JMLR)
- Contextual/background knowledge: conversational implicature

Ground Truth | a *pitcher* is leaning back about to throw a ball

Dialogue

A: Is the player wearing a ***uniform***?
B: Yes.
A: He is playing?
B: Yes.
A: Do he have ***baseball gear***?
B: Yes, a ***glove*** and a ***ball*** is in his hand.
A: They are in the ***baseball field***?
B: Yes.
A: He is wearing a ***cap***?
B: Yes.
A: How the weather looks?
B: It looks decent.
A: Do you see a ***baseball ball***?
B: Yes it 's in his hand.
A: The ***umpire*** is in the picture?
B: No.
A: The batter is in the picture?
B: No.
A: Do you see the fans?
B: No.

Example taking from VisDial (Das et al., 2017)



Interacting with human involves a lot of commonsense knowledge

- Space
- Time
- Location
- State
- Causality
- Color
- Shape
- Physical interaction
- Theory of mind
- Human interactions
- …

Judy Kegl, The boundary between word knowledge and world knowledge, TINLAP3, 1987
Ernie Davis, Building AIs with Common Sense, Princeton Chapter of the ACM, May 16, 2019

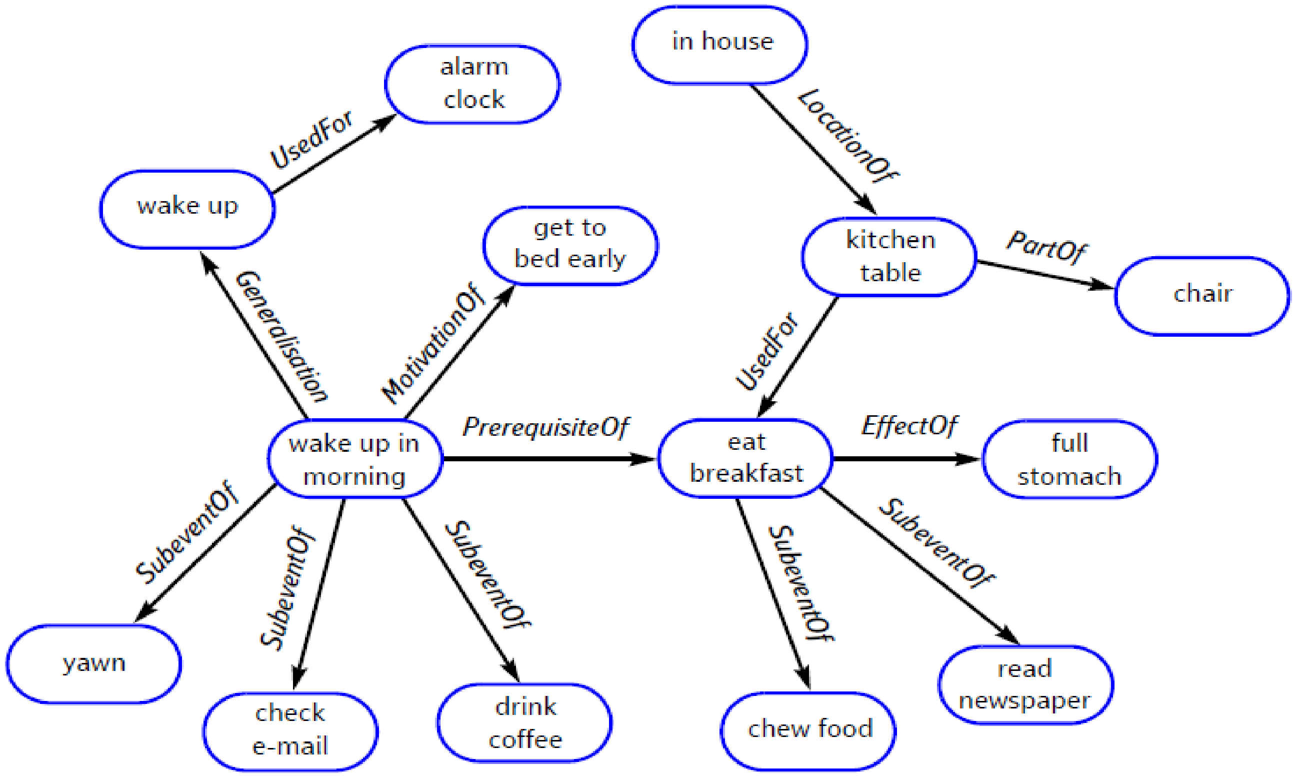
How to define commonsense knowledge? (Liu & Singh, 2004)

- "While to the average person the term 'commonsense' is regarded as synonymous with 'good judgement', "
- "the AI community it is used in a technical sense to refer to the millions of basic facts and understandings possessed by most people."
- "Such knowledge is typically omitted from social communications", e.g.,
  - If you forget someone's birthday, they may be unhappy with you.

H Liu and P Singh, ConceptNet - a practical commonsense reasoning tool-kit, BTTJ, 2004

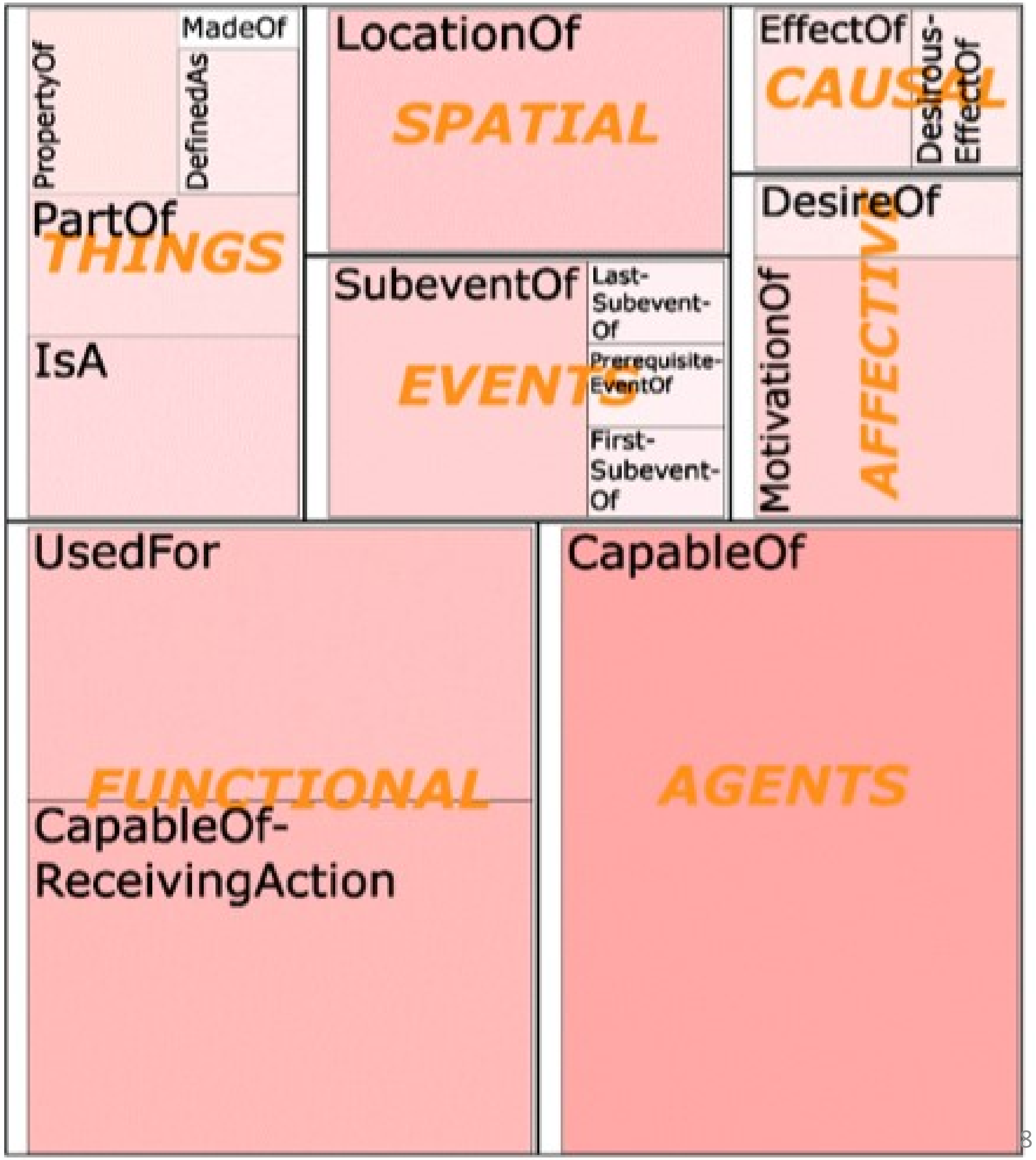
How to collect commonsense knowledge?

- ConceptNet5 (Speer and Havasi, 2012)
  - Core is from Open Mind Common Sense (OMCS) (Liu & Singh, 2004)



  - Essentially a crowdsourcing based approach + text mining

- Knowledge in ConceptNet
  - Things
  - Spatial
  - Location
  - Events
  - Causal
  - Affective
  - Functional
  - Agents



Comparison

	Database content	Resource	Capabilities	Scales
ConceptNet (2002-now)	Commonsense	OMCS (from the public) (automatic)	Contextual inference	1.6 million relations among 300,000 nodes (2004); now (2017) 21 million edges over 8 million nodes (1.5 million are English)
WordNet (1985)	Semantic Lexicon	Expert (manual)	Lexical categorization & word-similarity	200,000 word senses
Cyc (1984-now)	Commonsense	Expert (manual)	Formalized logical reasoning	1.6 million facts with 118,000 concepts (2004); now (2019) 20 million facts with 1.5 million concepts

Slides credit: Haixun Wang

The Scale

- “A founder of AI, Marvin Minsky, once estimated that ‘...commonsense is knowing maybe 30 or 60 million things about the world and having them represented so that when something happens, you can make analogies with others’.” (Liu & Singh, 2004)

H Liu and P Singh, ConceptNet - a practical commonsense reasoning tool-kit, BTTJ, 2004

What contribute to ConceptNet5.5 (21 million edges and over 8 million nodes)?

- Facts acquired from Open Mind Common Sense (OMCS) (Singh 2002) and sister projects in other languages (Anacleto et al. 2006)
- Information extracted from parsing Wiktionary, in multiple languages, with a custom parser ("Wikiparsec")
- "Games with a purpose" designed to collect common knowledge (von Ahn, Kedia, and Blum 2006) (Nakahara and Yamada 2011) (Kuo et al. 2009)
- Open Multilingual WordNet (Bond and Foster 2013), a linked-data representation ofWordNet (Miller et al. 1998) and its parallel projects in multiple languages
- JMDict (Breen 2004), a Japanese-multilingual dictionary
- OpenCyc, a hierarchy of hypernyms provided by Cyc (Lenat and Guha 1989), a system that represents commonsense knowledge in predicate logic
- A subset of DBPedia (Auer et al. 2007), a network of facts extracted from Wikipedia infoboxes

Most of them are entity-centric knowledge, there are only 74,989 nodes among 116,097 edges about events

Speer, Chin, and Havasi, ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. AAAI 2017.

Nowadays,

- Many large-scale knowledge graphs about entities and their attributes (property-of) and relations (thousands of different predicates) have been developed
  - Millions of entities and concepts
  - Billions of relationships

Google Knowledge Graph (2012)
570 million entities and 18 billion facts

However,

- Semantic meaning in our language can be described as ‘a finite set of mental primitives and a finite set of principles of mental combination (Jackendoff, 1990)’.
- The primitive units of semantic meanings include
  - Thing (or Object),
  - Activity,
  - State,
  - Event,
  - Place,
  - Path,
  - Property,
  - Amount,
  - etc.

How to collect more knowledge rather than entities and relations?

Jackendoff, R. (Ed.). (1990). Semantic Structures. Cambridge, Massachusetts: MIT Press.

Outline

- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Applications

Semantic Primitive Units

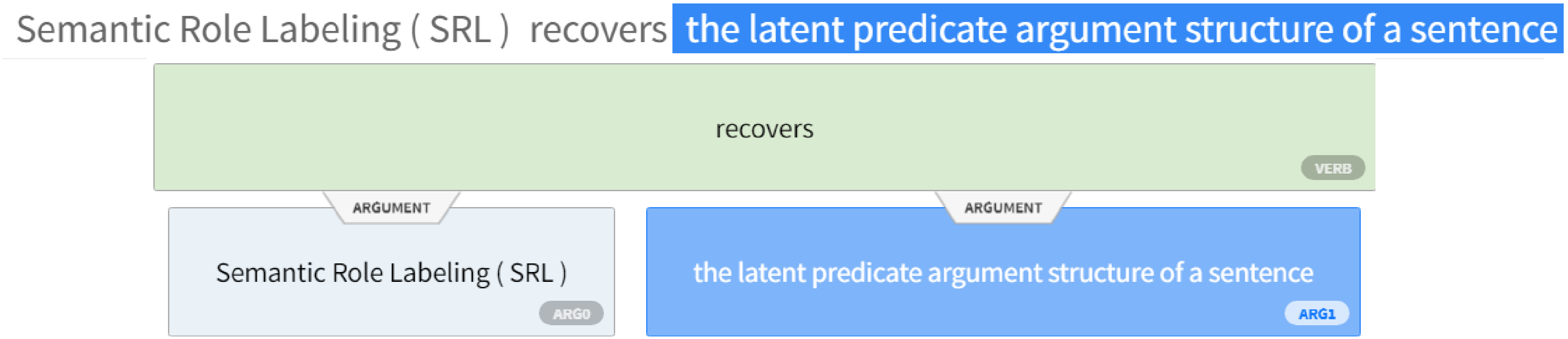
- Entities or concepts can be nouns or noun phrases
  - Concepts in Probase (2012):
    - Company,
    - IT company,
    - big company,
    - big IT company,
    - …
  - Hierarchy is partially based on head+modifier composition

- Let's think about verbs and verb phrases
  - How should we define semantic primitive unit for verbs?

Wentao Wu, Hongsong Li, Haixun Wang, Kenny Q Zhu. Probase: A probabilistic taxonomy for text understanding. SIGMOD, 2012. (now Microsoft concept graph https://concept.research.microsoft.com/)

Semantic Primitive Units Related to Verbs

- Too short?
  - v-o (verb + object): too general
    - have book
    - love reading
  - s-v-o (subject + verb + object): many overlaps with entity-centric KGs
    - Trump was born in 1946
    - Apple released iphone 8
- Too long?
  - Difficult to obtain frequent facts to reflect commonness of commonsense knowledge



https://demo.allennlp.org/semantic-role-labeling

Semantic Primitive Units

- Too fine grained, e.g., Event Extraction in ACE 2005?

After Sept. 11, 2001, Indonesia was quick to sign onto U.S. President George W. Bush's global war on terror.

Time-Starting

Attacker

Trigger

Attack

Event Type: Conflict.Attack

State-of-the-art overall F1: around 40% (Ji, & Huang, 2013)

In Baghdad, a cameraman died when an American tank fired on the Palestine Hotel.

Die.Place | Die.Victim | Trigger1 | Die.Instrument | Trigger2 | Attack.Target

Attack.Place | Attack.Target | Die | Attack.Instrument | Attack

Event Type: Life.Die

Event Type: Conflict.Attack

Qi Li, Heng Ji, Liang Huang: Joint Event Extraction via Structured Prediction with Global Features. ACL (1) 2013: 73-82

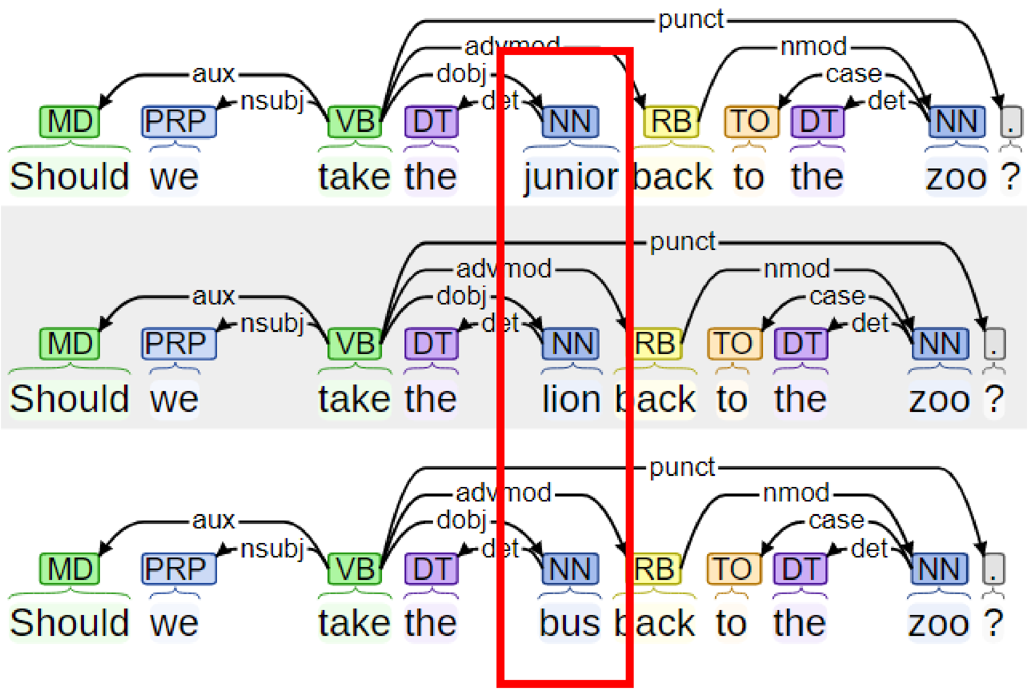
Commonsense Knowledge Construction

- The principle: a middle way of building primitive semantic units
  - Not too long
  - Not too short
  - Could be general
  - Better to be specific and semantically meaningful and complete

- Any linguistic foundation?

“Linguistic description – grammar = semantics”
The lower bound of a semantic theory (Katz and Fodor, 1963)

- Disambiguation needs both “the speaker's knowledge of his language and his knowledge about the world” (Katz and Fodor, 1963)
  - The **bill** is large.
    - Some document demanding a sum of money to discharge a debt exceeds in size most such documents
    - The beak of a certain bird exceeds in bulk those of most similar birds
  - Syntactically unambiguous
  - Compare semantic meanings by fixing grammar



Principle #1

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. Language, 39(2), 170–210.

Selectional Preference (SP)

Principle #2

- The need of language inference based on '**partial information**' (Wilks, 1975)
  - The soldiers fired at the women, and we saw several of them fall.
  - The needed partial information: hurt things tending to fall down
    - "not invariably true"
    - "tend to be of a very high degree of generality indeed"
- Selectional preference (Resnik, 1993)
  - A relaxation of selectional restrictions (Katz and Fodor, 1963) and as syntactic features (Chomsky, 1965)
  - Applied to isA hierarchy in WordNet and verb-object relations

Yorick Wilks. 1975. An intelligent analyzer and understander of English. Communications of the ACM, 18(5):264–274.
Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. Language, 39(2), 170–210.
Noam Chomsky. 1965. Aspects of the Theory of Syntax. MIT Press, Cambridge, MA.
Philip Resnik. 1993. Selection and information: A class-based approach to lexical relationships. Ph.D. thesis, University of Pennsylvania.

A Test of Commonsense Reasoning

- Proposed by Hector Levesque at U of Toronto
- An example taking from Winograd Schema Challenge

  - (A) The fish ate the worm. It was hungry.
  - (B) The fish ate the worm. It was tasty.

- On the surface, they simply require the resolution of anaphora
  - But Levesque argues that for Winograd Schemas, the task requires the use of knowledge and commonsense reasoning

http://commonsensereasoning.org/winograd.html
https://en.wikipedia.org/wiki/Winograd_Schema_Challenge

Why is it a challenge?

- Must also be carefully written not to betray their answers by selectional restrictions or statistical information about the words in the sentence

- Designed to be an improvement on the Turing test

The soldiers fired at the women, and we saw several of them fall.

woman fall

All Images News Videos

About 2,360,000,000 results (0.47 seconds)

soldier fall

All Images Videos News

About 244,000,000 results (0.65 seconds)

- (A) The fish ate the worm. It was hungry.
- (B) The fish ate the worm. It was tasty.

fish hungry

All Images Videos News

About 119,000,000 results (0.67 seconds)

worm hungry

All Images News Videos

About 9,490,000 results (0.47 seconds)

fish tasty

All Images Videos Maps

About 312,000,000 results (0.59 seconds)

worm tasty

All Images Videos News

About 17,600,000 results (0.60 seconds)

A Brief History of Datasets and Development

2011	2012	2014	Recent results
Levesque. AAAI Spring Symposium	The first large dataset. Rahman and Ng: EMNLP-CoNLL	Davis et al. "A Collection of Winograd Schemas"	

Human’s performance: 95% (Nangia and Bowman, 2018)

Author/year	System	Fine-tuned	Accuracy
Emami et al. (2018)	Knowledge Hunter	No	54.58%
Trieu H. Trinh and Quoc V. Le (2018)	Language models (single)	No	54.58%
	Language models (Ensemble)	No	63.74%
Alec Radford et al. (2019)	GPT-2	No details	70.70%
Ruan et al. (2019)	BERT-large + dependency	Rahman and Ng 2012 dataset	71.10%
Kocijan et al. (2019)	BERT-large	No	60.10%
	GPT	No	55.30%
		Wiki + Rahman and Ng 2012 dataset	72.20%

SP-10K: A Large-scale Evaluation Set

- Traditional evaluation
  - Small sets of one-hop direct dependency relations
    - McRae et al., 1998: 821 pairs of nsubj and dobj relations
    - Keller and Lapata, 2003: 540 pairs of dobj, noun-noun, and amod relations
    - Padó et al., 2006: 207 pairs of nsubj, dobj, and amod relations
    - Wang et al, 2018: 3062 (subject, verb, dobject) triplets
  - Pseudo-disambiguation (Ritter et al., 2010; de Cruys, 2014): corpus driven, no human annotation

- Ours:
  - 10K pairs of five relations, including two 2-hop relations

Hongming Zhang, Hantian Ding, and Yangqiu Song. SP-10K: A Large-Scale Evaluation Set for Selectional Preference Acquisition. ACL, 2019.

Examples in SP-10K

dobj	Plausibility
(eat, meal)	10.00
(close, door)	8.50
(touch, food)	5.50
(hate, investment)	4.00
(eat, mail)	0.00

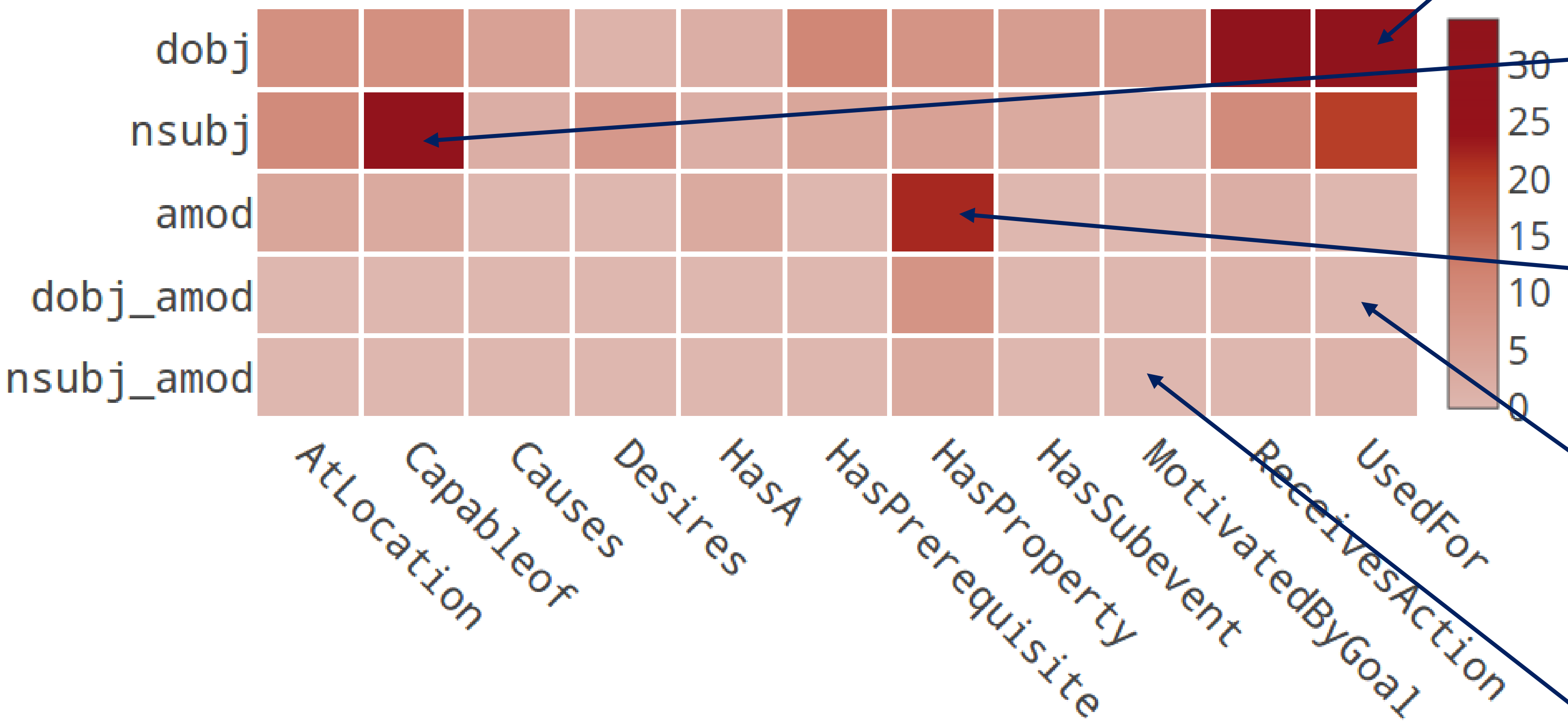
nsubj	Plausibility
(singer, sing)	10.00
(law, permit)	7.78
(women, pray)	5.83
(victim, contain)	2.22
(textbook, eat)	0.00

amod	Plausibility
(fresh, air)	9.77
(new, method)	8.89
(medium, number)	4.09
(immediate, food)	2.05
(secret, wind)	0.75

dobj_amod	Plausibility
(lift, heavy *object*)	9.17
(design, new *object*)	8.00
(attack, small *object*)	5.23
(inform, weird *object*)	3.64
(earn, rubber *object*)	0.63

nsubj_amod	Plausibility
(evil *subject*, attack)	9.00
(recent *subject*, demonstrate)	6.00
(random *subject*, bear)	4.00
(happy *subject*, steal)	2.25
(sunny *subject*, make)	0.56

Correlations with OMCS



(sing, song) (dobj, 9.25)
(song, UsedFor, sing)

(phone, ring) (nsubj, 8.75)
(phone, CapableOf, ring)

(cold, water) (amod, 8.86)
(water, HasProperty, cold)

(create, new) (dobj_amod, 8.25)
(create idea, UsedFor, invent new things)

(hungry, eat) (nsubj_amod, 10.00)
(eat, MotivatedByGoal, are hungry)

Performance on Winograd Schema

- 72 out of 273 questions satisfying nsubj_amod and dobj_amod relations
  - Jim yelled at Kevin because he was so upset.
  - We compare the scores
    - (yell, upset *object*) following nsubj_amod
    - (upset *object* , yell) following dobj_amod
- Results

Model	Correct	Wrong	NA	Accuracy (predicted)	Accuracy (overall)
Stanford	33	35	4	48.5%	48.6%
End2end (Lee et al., 2018)	36	36	0	50.0%	50.0%
PP* (Resnik, 1997)	36	19	17	65.5%	61.8%
SP-10K	13	0	56	100%	59.0%

dobj_amod	Plausibility
(lift, heavy *object*)	9.17
(design, new *object*)	8.00
(attack, small *object*)	5.23
(inform, weird *object*)	3.64
(earn, rubber *object*)	0.63

nsubj_amod	Plausibility
(evil *subject*, attack)	9.00
(recent *subject*, demonstrate)	6.00
(random *subject*, bear)	4.00
(happy *subject*, steal)	2.25
(sunny *subject*, make)	0.56

*PP: posterior probability for SP acquisition using Wikipedia data

Outline

- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Applications

Higher-order Selectional Preference

- The need of language inference based on ‘**partial information**’ (Wilks, 1975)
  - The soldiers fired at the women, and we saw several of them fall.
  - The needed partial information: hurt things tending to fall down
  - Many ways to represent it, e.g.,

(hurt, X) connection (X, fall)

- How to scale up the knowledge acquisition and inference?

ATOMIC

- Crowdsourcing 9 Types of IF-THEN relations

- All entity information has been removed to reduce ambiguity

Event	Type of relations	Inference examples	Inference dim.
"PersonX pays PersonY a compliment"	If-Event-Then-Mental-State	PersonX wanted to be nice PersonX will feel good PersonY will feel flattered	xIntent xReact oReact
	If-Event-Then-Event	PersonX will want to chat with PersonY PersonY will smile PersonY will compliment PersonX back	xWant oEffect oWant
	If-Event-Then-Persona	PersonX is flattering PersonX is caring	xAttr xAttr
"PersonX makes PersonY's coffee"	If-Event-Then-Mental-State	PersonX wanted to be helpful PersonY will be appreciative PersonY will be grateful	xIntent oReact oReact
	If-Event-Then-Event	PersonX needs to put the coffee in the filter PersonX gets thanked PersonX adds cream and sugar	xNeed xEffect xWant
	If-Event-Then-Persona	PersonX is helpful PersonX is deferential	xAttr xAttr
"PersonX calls the police"	If-Event-Then-Mental-State	PersonX wants to report a crime Others feel worried	xIntent oReact
	If-Event-Then-Event	PersonX needs to dial 911 PersonX wants to explain everything to the police PersonX starts to panic Others want to dispatch some officers	xNeed xWant xEffect oWant
	If-Event-Then-Persona	PersonX is lawful PersonX is responsible	xAttr xAttr

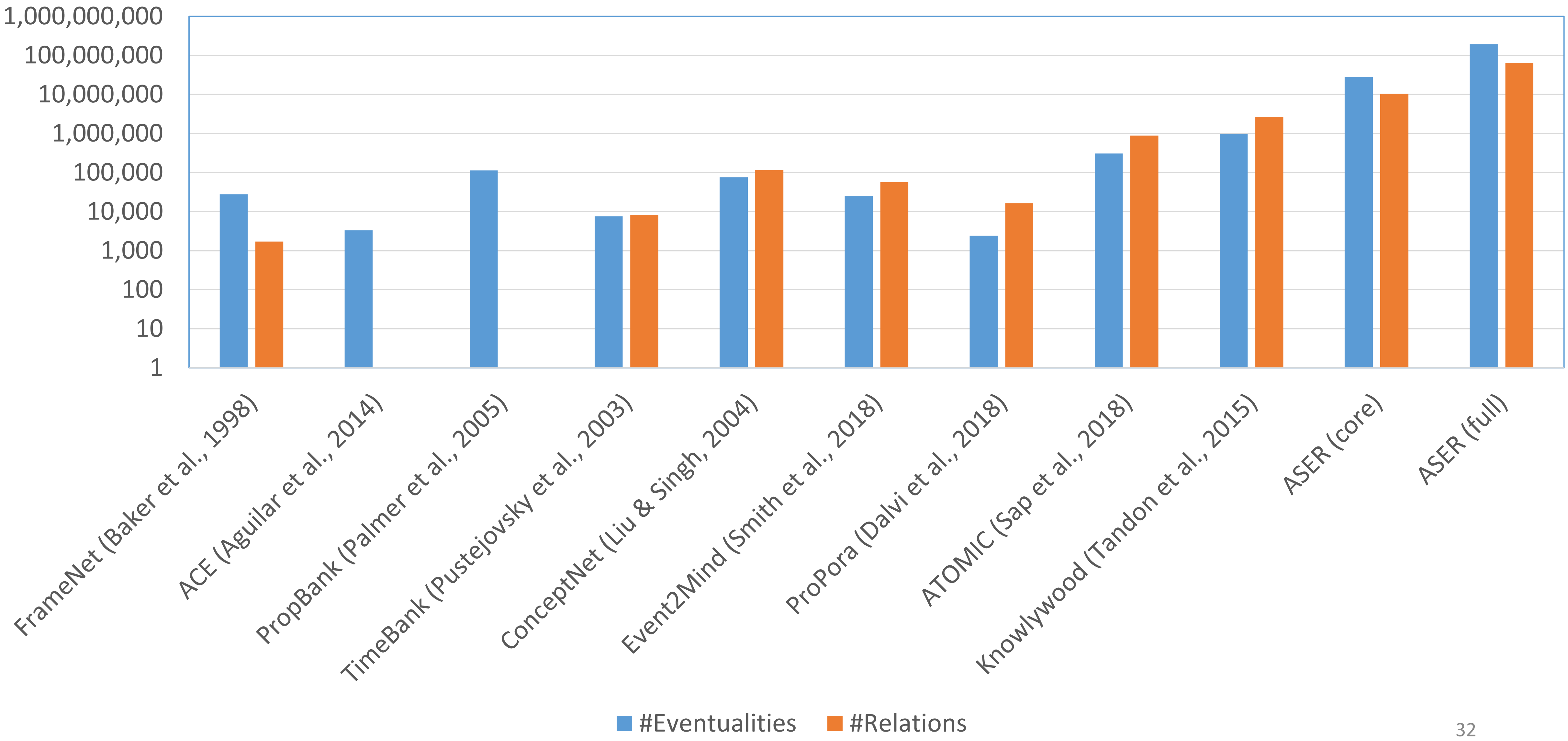
Maarten Sap, Ronan LeBras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, Yejin Choi: ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning. AAAI, 2019.

KnowlyWood

- Perform information extraction from free text
  - Mostly movie scripts and novel books
- Four relations: previous, next, parent, similarity
- No subject information
  - Only verb+object

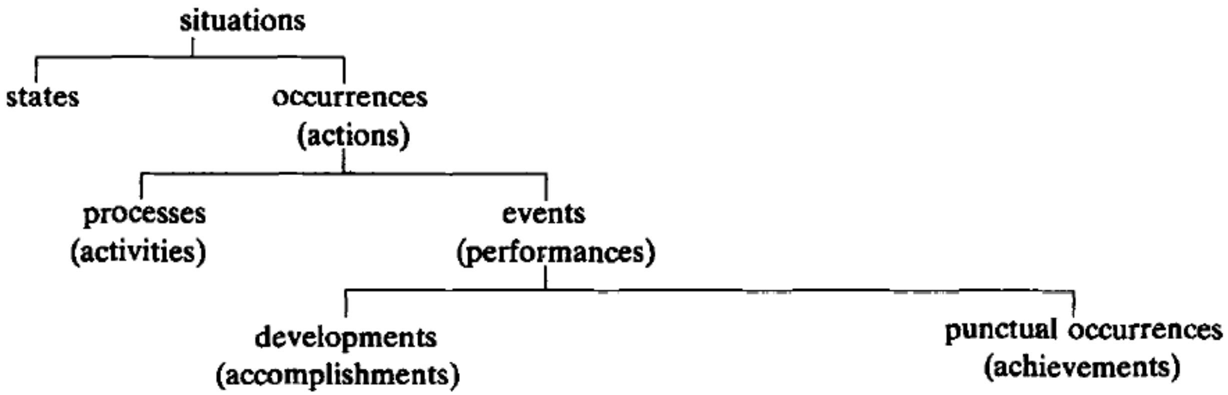
Niket Tandon, Gerard de Melo, Abir De, Gerhard Weikum: Knowlywood: Mining Activity Knowledge From Hollywood Narratives. CIKM 2015: 223-232

Scales of Verb Related Knowledge Graphs

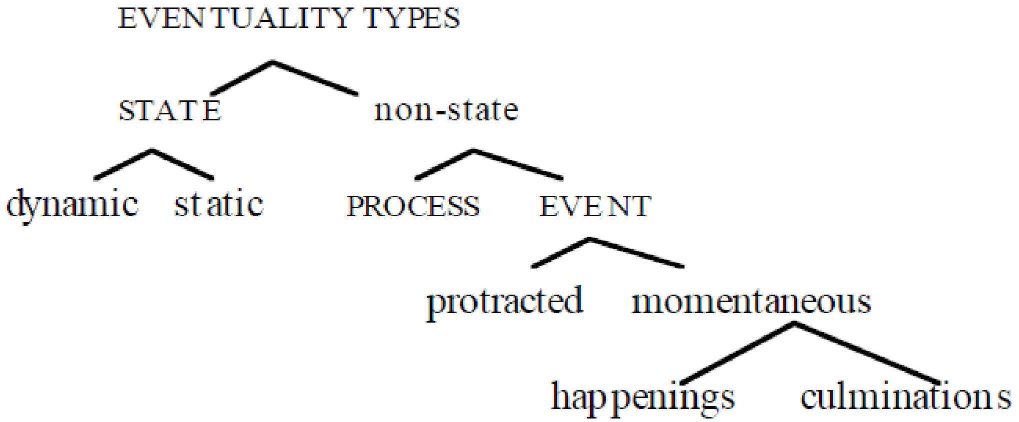


ASER
Activities, States, Events, and their Relations

Mourelatos' taxonomy (1978)



Bach's taxonomy (1986)



- **State**: The air smells of jasmine.
- **Process**: It's snowing.
- **Development**: The sun went down.
- **Punctual occurrence**: The cable snapped. He blinked. The pebble hit the water.

- **Static states**: be in New York, love (one's cat);
- **Dynamic states**: sit, stand, drunk, present, sick;
- **Processes**: walk, push a cart, sleep;
- **Protracted events**: build (a cabin), eat a sandwich, polish a shoe, walk to Boston;
- **Culminations**: take off; arrive, leave, depart;
- **Happenings**: blink, flash, knock, kick, hit, pat, wink;

Alexander P. D. Mourelatos. Events, processes, and states. Linguistics and Philosophy, 2, 415-434. 1978.
Emmon Bach. The algebra of events. Linguistics and philosophy, 9 (1), 5-16. 1986.

Many other and subtle definitions

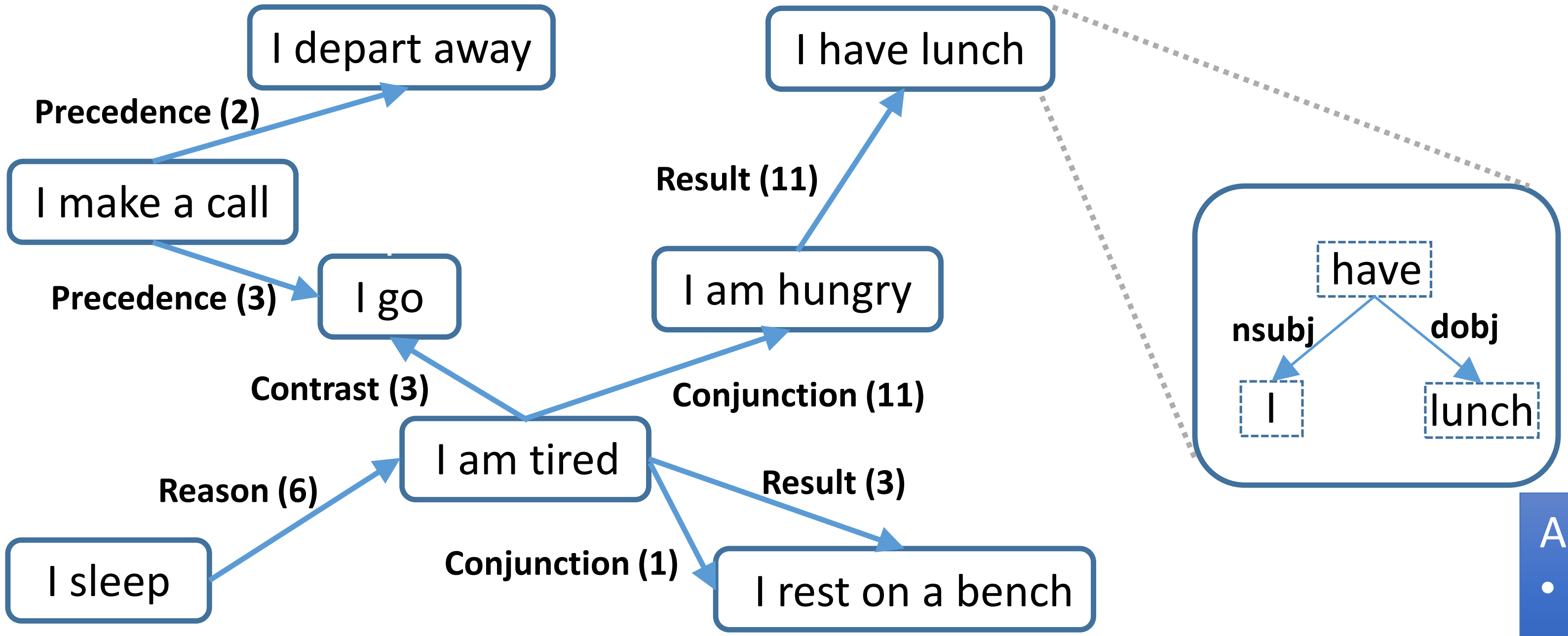
- According to Alexander Mourelatos (1978),
  - "Event, can be sharply differentiated by ... the contrast between perfective and imperfective aspect in verbs corresponds to the count/mass distinction in the domain of nouns."
    - "Cardinal count" adverbials versus frequency
    - adverbials occurrence versus associated occasion
  - "Mary capsized the boat" is an event predication because (a) it is equivalent to "There was at least one capsizing of the boat by Mary," or (b) because it admits cardinal count adverbials, e.g., "at least once," "twice," "three times."

- Learning based classification
  - English: state vs. non-state; ~93.9% accuracy; culminated/nonculminated ~74.0% (Siegel and McKeown, 2000)
  - Chinese: state, activity, change; ~73.6% accuracy (Liu et al., 2018)

Siegel, Eric V., and Kathleen R. McKeown. "Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights." *Computational Linguistics* 26.4 (2000): 595-628.
Hongchao Liu, Chu-ren Huang, Renkui Hou, and Hongzheng Li, Prediction of Mandarin Verbs' Event Types Based on Lingustic Features Vectors and Word Embedding Vectors. Journal of Chinese Information Processing. 2018.

Our Approach

- Use verb-centric patterns from dependency parsing
  - Principle #1: For comparing semantics by fixing syntax (Katz and Fodor, 1963)
- Maintain a set of key tags and a set of auxiliary tags
  - Principle #2: For obtaining frequent 'partial information' (Wilks, 1975)



A hybrid graph of
- Each eventuality is a hyper-edge of words
- Heterogeneous edges among eventualities

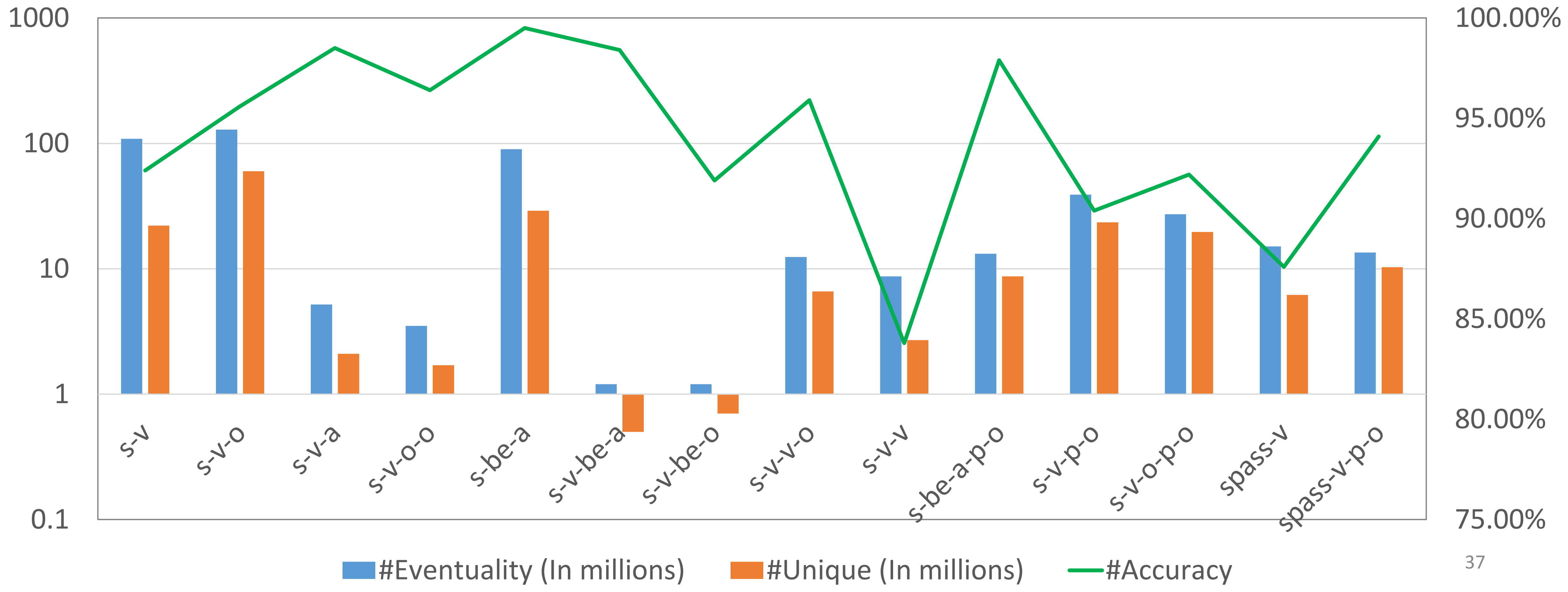
Eventualities

- Using patterns to collect partial information
- Six relations are also kept but treated as auxiliary edges
  - advmod,
  - amod,
  - nummod,
  - aux,
  - compound,
  - neg

Pattern	Code	Example
n1-nsubj-v1	s-v	`The dog barks'
n1-nsubj-v1-dobj-n2	s-v-o	`I love you'
n1-nsubj-v1-xcomp-a	s-v-a	`He felt ill'
n1-nsubj-(v1-iobj-n2)-dobj-n3	s-v-o-o	`You give me the book'
n1-nsubj-a1-cop-be	s-be-a	`The dog is cute'
n1-nsubj-v1-xcomp-a1-cop-be	s-v-be-a	`I want to be slim'
n1-nsubj-v1-xcomp-n2-cop-be	s-v-be-o	`I want to be a hero'
n1-nsubj-v1-xcomp-v2-dobj-n2	s-v-v-o	`I want to eat the apple'
n1-nsubj-v1-xcomp-v2	s-v-v	`I want to go'
(n1-nsubj-a1-cop-be)-nmod-n2-case-p1	s-be-a-p-o	`It' cheap for the quality'
n1-nsubj-v1-nmod-n2-case-p1	s-v-p-o	`He walks into the room'
(n1-nsubj-v1-dobj-n2)-nmod-n3-case-p1	s-v-o-p-o	`He plays football with me'
n1-nsubjpass-v1	spass-v	`The bill is paid'
n1-nsubjpass-v1-nmod-n2-case-p1	spass-v-p-o	`The bill is paid by me'

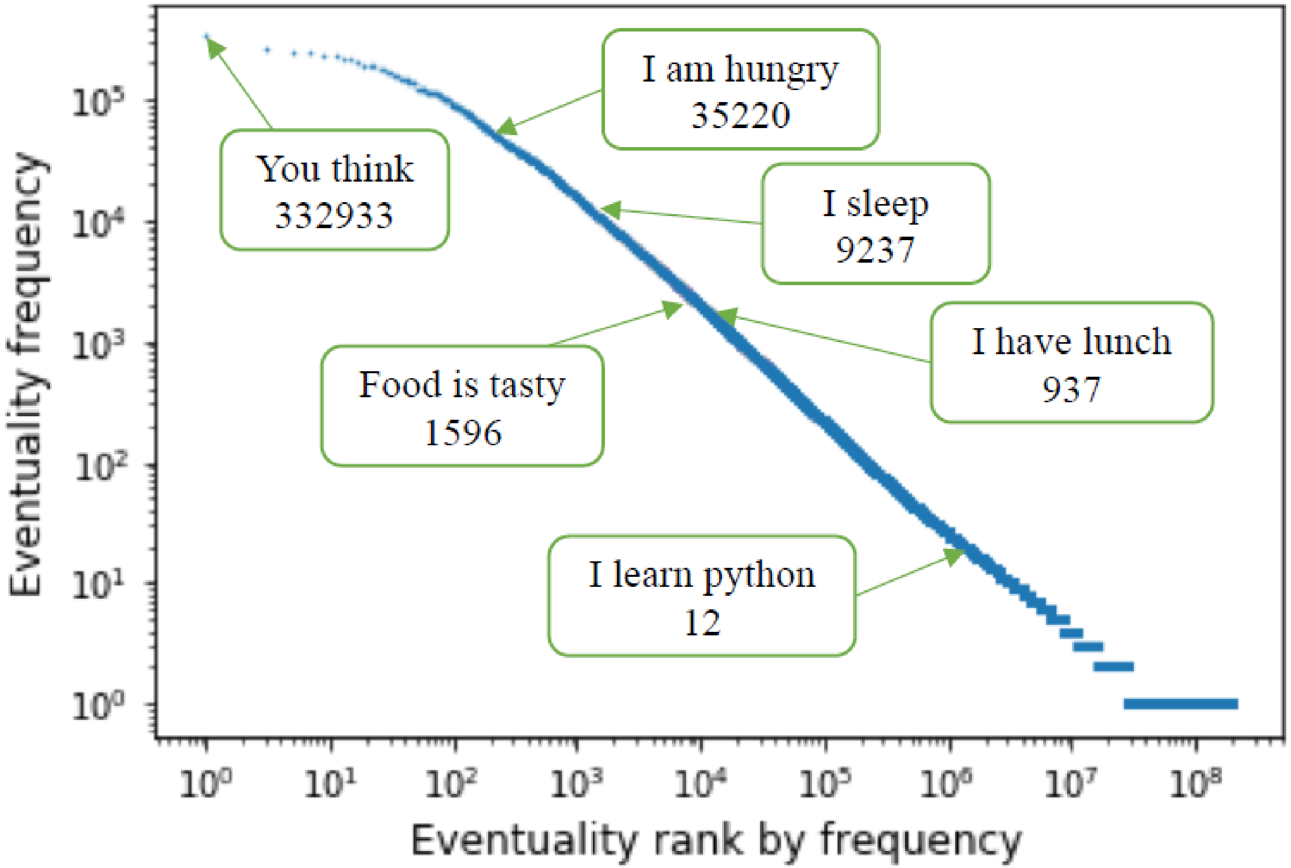
Extraction Results

- Extract examples from 11-billion tokens from Yelp, NYT, Wiki, Reddit, Subtitles, E-books
- Evaluate about 200 examples in each pattern using Amazon Turk



Distribution

- Frequency characterizes selectional preference, e.g.,
- `The dog is chasing the cat, it barks loudly‘
  - ‘dog barks’ appears 12,247
  - ‘cat barks’ never appears



Eventuality Relations: Pattern Matching + Bootstrapping

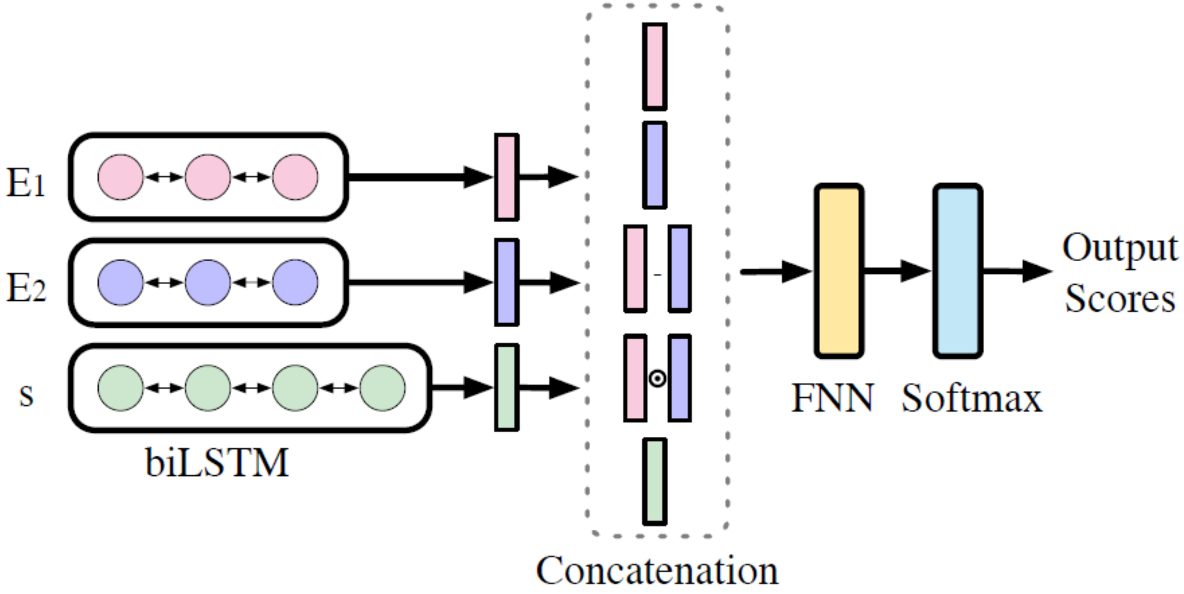
- Seeds from Penn Discourse Treebank (PDTB) (Prasad et al., 2007)
- 14 relations taking from CoNLL shared task
    - More frequent relations
- Less ambiguous connectives
    - 'so that' 31 times only in 'Result' relations
- Some are ambiguous
    - 'while': Conjunction 39 times, Contrast 111 times, Expectation 79 times, and Concession 85 times

Relation Type	Seed Patterns
Precedence	E1 **before** E2; E1 , **then** E2; E1 **till** E2; E1 **until** E2
Succession	E1 **after** E2; E1 **once** E2
Synchronous	E1, **meanwhile** E2; E1 **meantime** E2; E1, **at the same time** E2
Reason	E1, **because** E2
Result	E1, **so** E2; E1, **thus** E2; E1, **therefore** E2; E1, **so that** E2
Condition	E1, **if** E2; E1, **as long as** E2
Contrast	E1, **but** E2; E1, **however** E2; E1**, by contrast** E2; E1, **in contrast** E2; E1 , **on the other hand**, E2; E1, **on the contrary**, E2
Concession	E1, **although** E2
Conjunction	E1 **and** E2; E1, **also** E2
Instantiation	E1, **for example** E2; E1, **for instance** E2
Restatement	E1, **in other words** E2
Alternative	E1 **or** E2; E1, **unless** E2; E1, **as an alternative** E2; E1, **otherwise** E2
ChosenAlternative	E1, E2 **instead**
Exception	E1, **except** E2

Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., & Webber, B. L. (2007). The penn discourse treebank 2.0 annotation manual.
Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Rashmi Prasad, Christopher Bryant, Attapol T. Rutherford. The CoNLL-2015 Shared Task on Shallow Discourse Parsing.

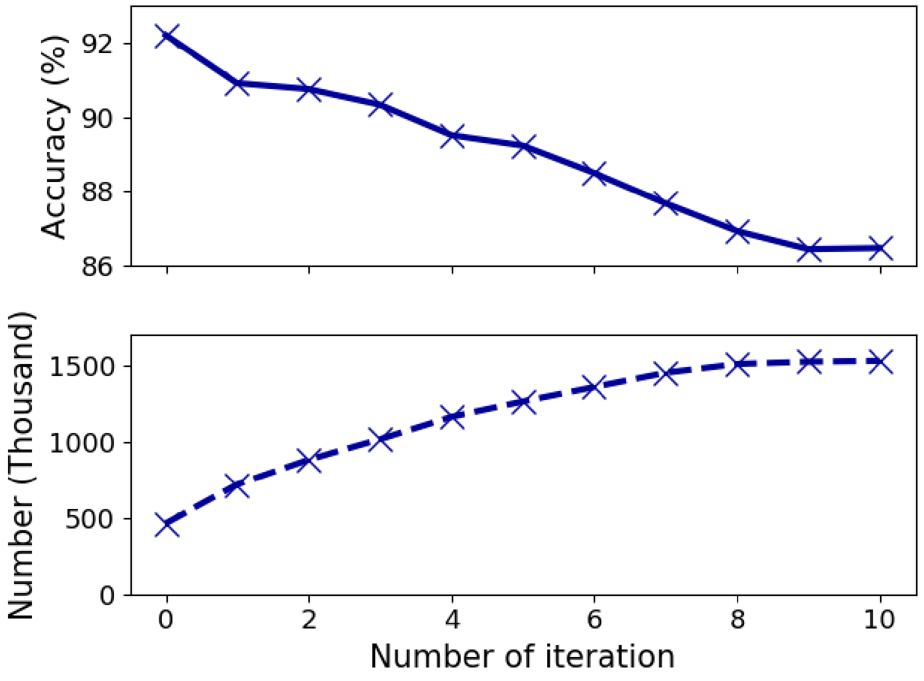
Eventuality Relations: Pattern matching + Bootstrapping

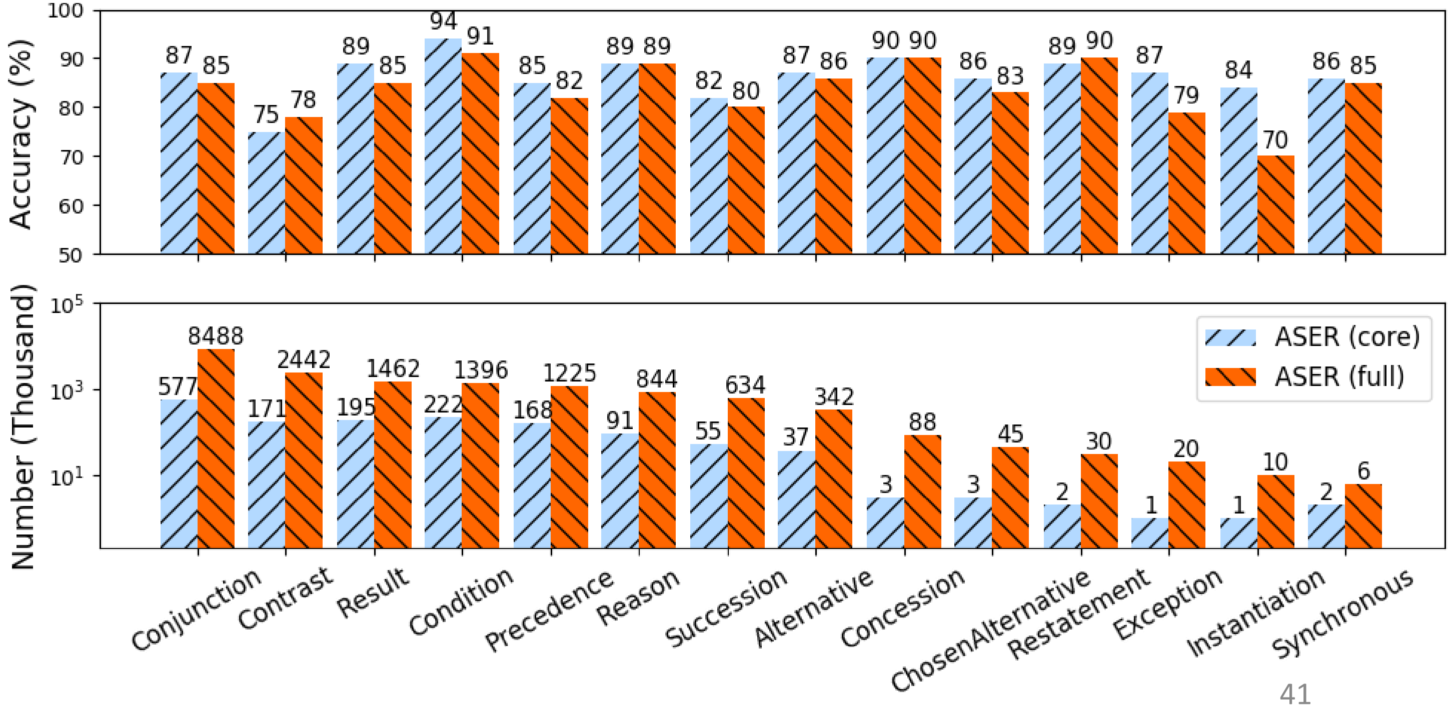
- Bootstrapping: incrementally self-supervised learning
- For each instance x = (E1;E2; sentence)
  - Use three bidirectional LSTMs
- Reduce the confident rate by iterations to reduce error propagation



Extraction Results

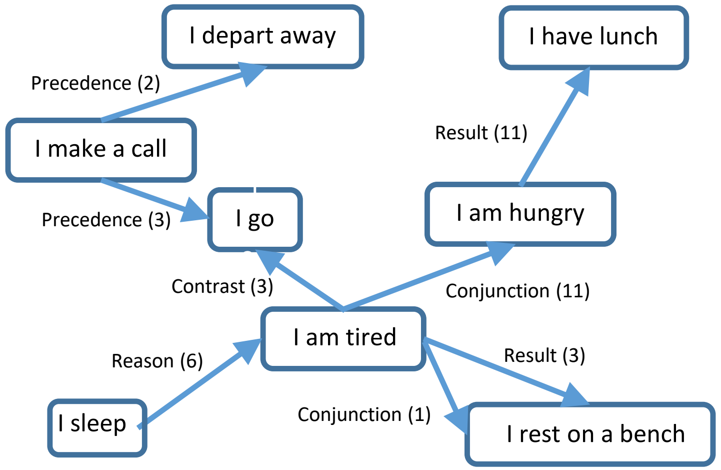
- Left: number of relations and overall accuracy
- Right: accuracy of each relations for the last iteration
- Each point is annotated with 200 examples by Amazon Turk





An Example of Inference over ASER

- We can support both eventuality-based and relation based inference
- We also do higher-order relation inference



	Eventuality Retrieval	Relation Retrieval
One-hop	P('I have lunch' | 'I am hungry', *Result*) = 1	P(*Result* | 'I am hungry', 'I have lunch') = 1
	P('I go' | 'I make a call', *Precedence*) = 0.6	P(*Result* | 'I am tired', 'I rest on a bench') = 0.75
	P('I depart away' | 'I make a call', *Precedence*) = 0.4	P(*Conjunction* | 'I am tired', 'I rest on a bench') = 0.25
Two-hop	P(' I rest one a bench' | 'I sleep', *Reason, Result*) = 1	P(*Reason*, *Conjunction* | 'I sleep', 'I am hungry') = 1
	P('I am hungry' | 'I sleep', *Reason, Conjunction*) = 0.91	P(*Reason*, *Result* | 'I sleep', 'I rest on a bench') = 0.75
	P('I rest one a bench' | 'I sleep', *Reason, Conjunction*) = 0.09	P(*Reason*, *Conjunction* | 'I sleep', 'I rest on a bench') = 0.25

Outline

- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Applications

Inference for Winograd Schema Challenge

Question

97. The fish ate *the worm*. **It** was hungry.

98. *The fish* ate the worm. **It** was tasty.

Extracted Eventualities

The fish: ('X ate Y', 'X was hungry')

the worm: ('X ate Y', 'Y was hungry')

The fish: ('X ate Y', 'X was tasty')

the worm: ('X ate Y', 'Y was tasty')

ASER Knowledge

ASER('X ate Y', 'X was hungry') = 18

ASER('X ate Y', 'Y was hungry') = 1

ASER('X ate Y', 'X was tasty') = 0

ASER('X ate Y', 'Y was tasty') = 7

Prediction

The fish

the worm

Results

- We selected a subset of 165 questions
  - The sentence does not have a subordinate clause
  - The targeting pronoun is covered by a pattern we used

Methods	Correct	Wrong	NA	Predicted	Overall
Random Guess	83	82	0	50.30%	50.30%
Deterministic (Raghunathan et al., 2010)	75	71	19	51.40%	51.20%
Statistical (Clark & Manning, 2015)	75	78	12	49.00%	49.10%
Deep-RL (Clark & Manning, 2016)	80	76	9	51.30%	51.20%
End2end (Lee et al., 2018)	79	84	2	48.50%	48.50%
Knowledge Hunting (Emami et al., 2018)	94	71	0	56.90%	56.90%
LM (single) (Trinh & Le, 2018)	90	75	0	54.50%	54.50%
SP (human) (Zhang et al., 2019)	15	0	150	100%	54.50%
SP (PP) (Zhang et al., 2019)	50	26	89	65.80%	57.30%
ASER	63	27	75	70.00%	60.90%

Dialogue Generation

- DailyDialog dataset (Li et al., 2017)

Post	I should eat some food .
Response	Yeah, you must be hungry. Do you like to eat some beaf?
ConceptNet	\`eat food', MotivatedByGoal, \`you are hungry' \`eat food', HasPrerequisite, \`open your mouth'
KnowlyWood	(eat,food), next, (keep, eating) (eat,food), next, (enjoy, taste) (eat,food), next, (stick, wasp) ...
ASER	i eat food [s-v-o], Conjunction, beef is good [s-be-a] i eat food [s-v-o], Condition, i am hungry [s-be-a] i eat food [s-v-o], Concession, i take picture [s-v-o] ...

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, Shuzi Niu: DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset. IJCNLP(1) 2017: 986-995

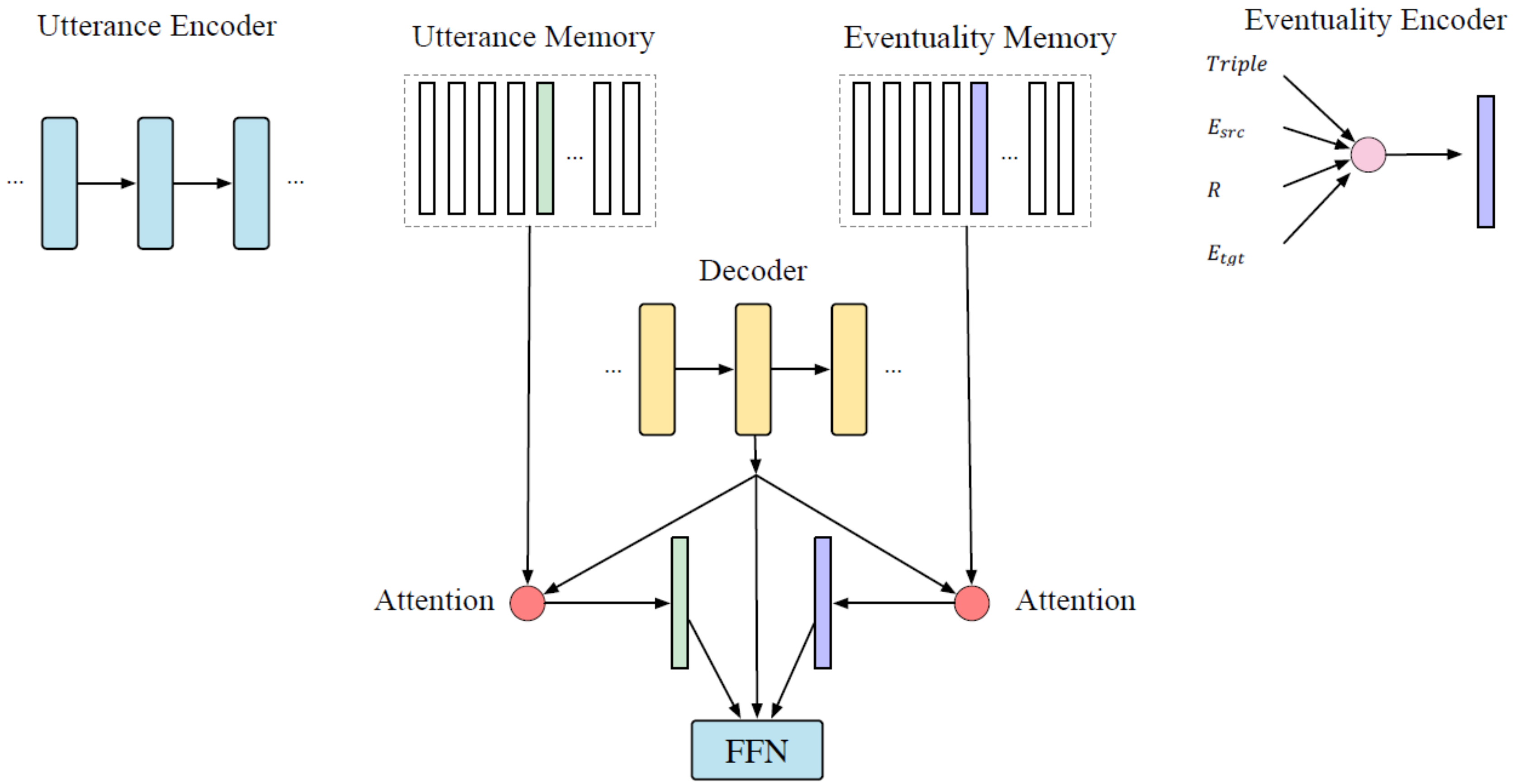
Coverage

- We select all pairs that at least one KG can cover
  - 30,145 of 49,188 conversation pairs are selected

KG	# Covered Pairs	Coverage Rate	# Unique matched events
ConceptNet	7,246	24.04%	1,195
Knowlywood	17,183	57.00%	30,036
ASER	20,494	67.98%	9,511

Dialogue Generation Model

- A typical seq2seq model with memories



Results

- Metric: BLEU score of generated responses

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Base	30.16	5.75	2.28	0.98
+ConceptNet	30.89	6.14	2.60	1.21
+KnowlyWood	30.72	6.26	2.68	1.29
+ASER	**32.10**	**7.14**	**3.54**	**2.07**

Conclusions and Future Work

- We extended the concept of selectional preference for commonsense knowledge extraction

- A very preliminary work with many potential extensions
  - More patterns to cover
  - More links in the KG
  - More types of relations
  - More applications

Thank you ☺

- Code and data
  - https://github.com/HKUST-KnowComp/ASER
- Project Homepage
  - https://hkust-knowcomp.github.io/ASER/