ASER: Building a Commonsense Knowledge Graph by Higher-order Selectional Preference

Yangqiu Song

Department of CSE, HKUST, Hong Kong

2020.07

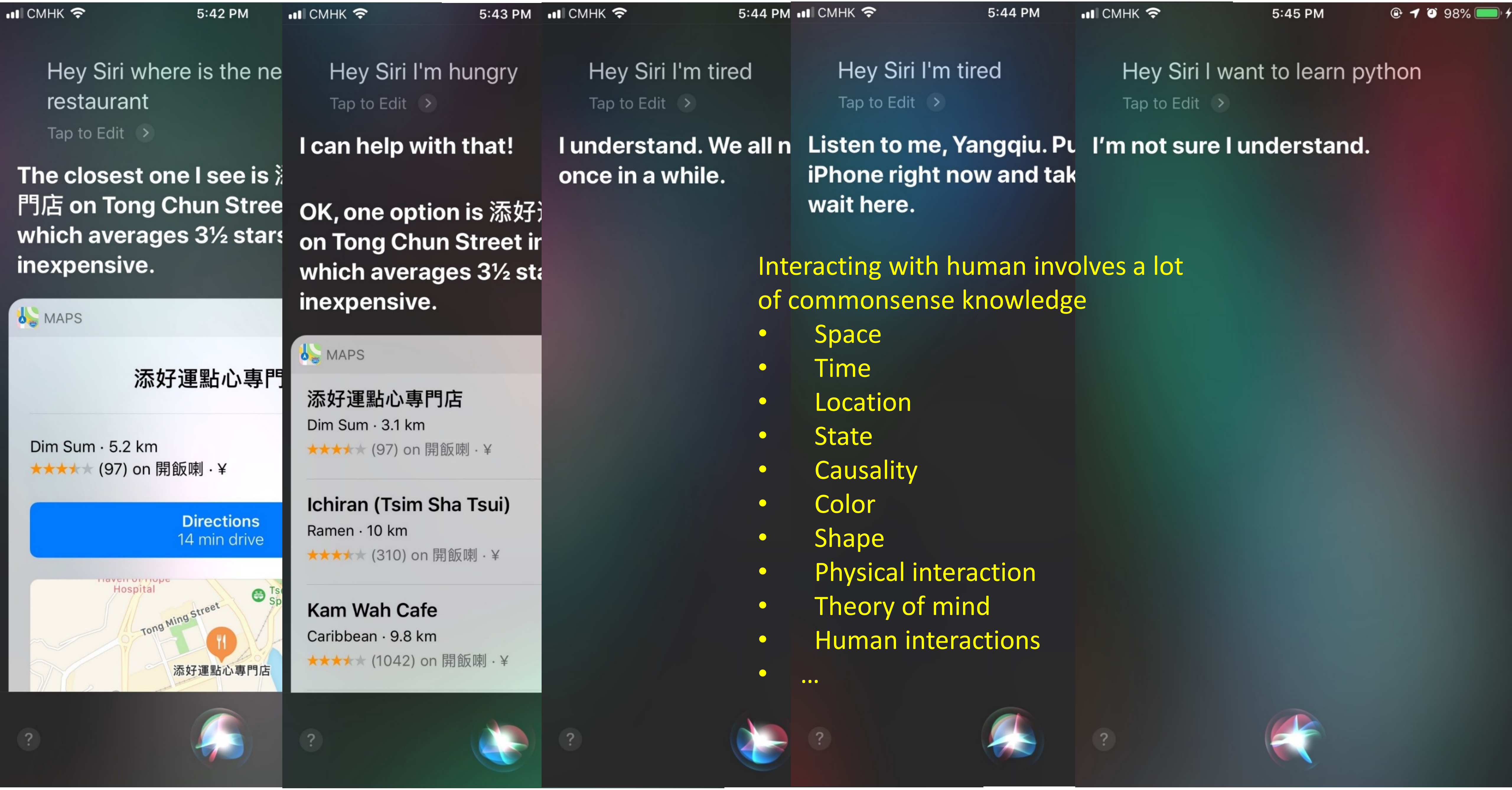
香 港 科 技 大 學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Contributors and Acknowledgements

- PhD/MPhil Students
  - Hongming Zhang, Xin Liu, Haojie Pan, Changlong Yu
- Undergraduate Students
  - Hantian Ding, Jiefu Ou, Xinran Zhao
- Other Collaborators
  - Cane Wing-Ki Leung, Daniel Khashabi, Dan Roth, Wilfred Ng, Lifeng Shang

Outline

- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Application on the Winograd Schema Challenge
- Extensions



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

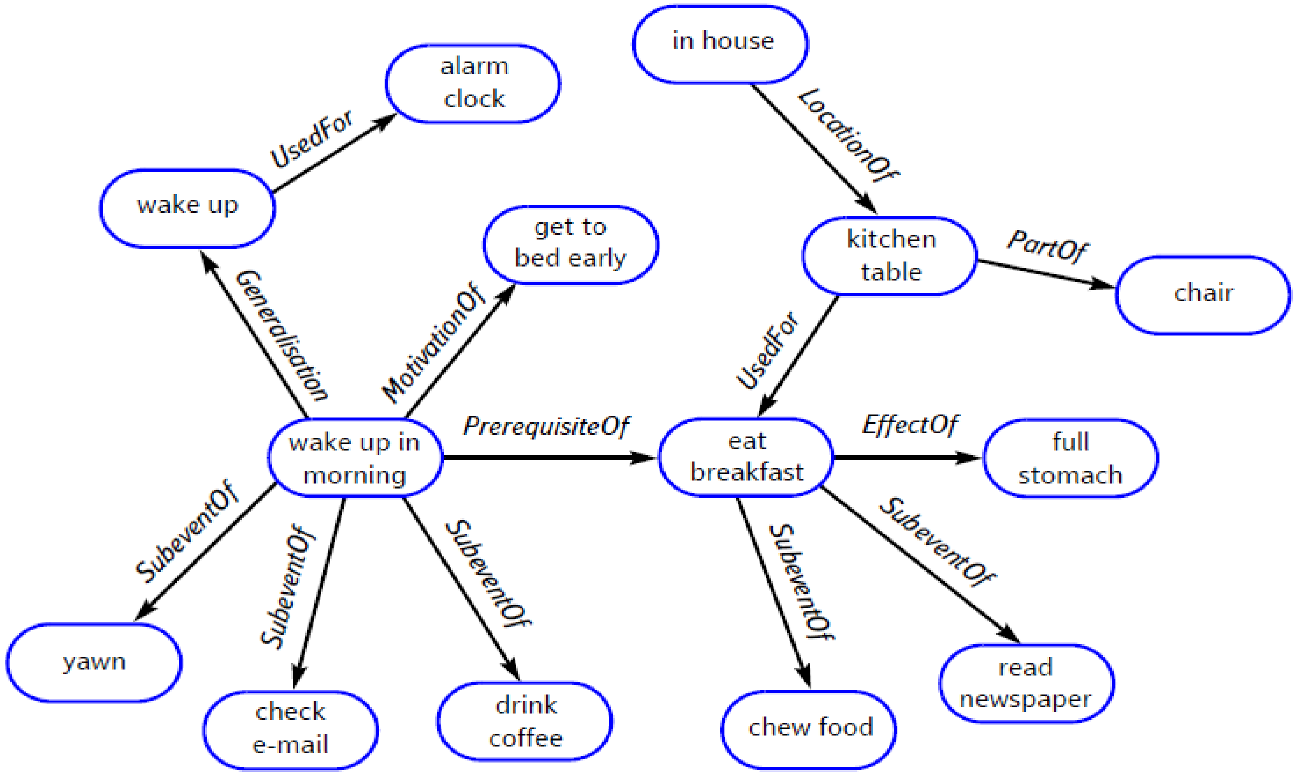
Commonsense Knowledge is the Key

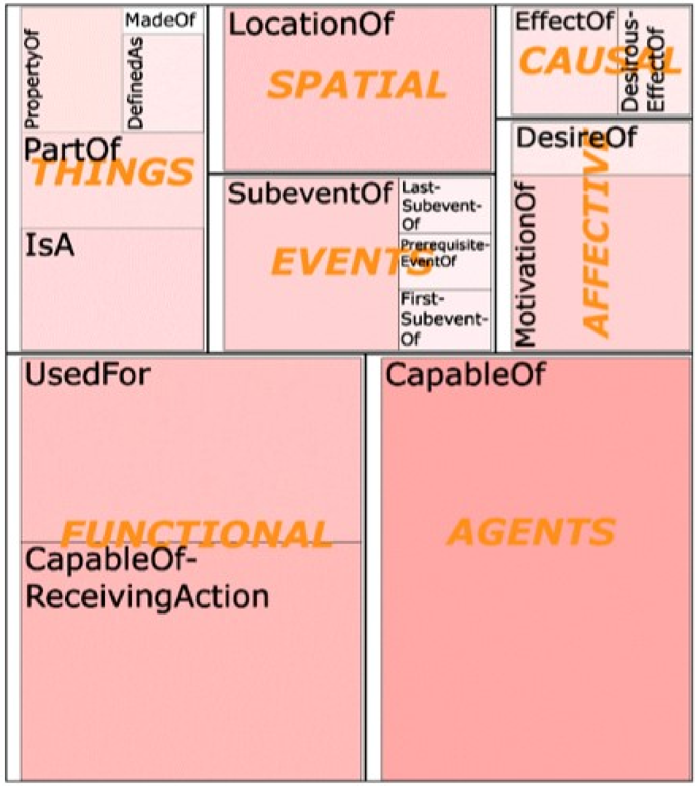
- How to define commonsense knowledge? (Liu & Singh, 2004)
  - "While to the average person the term 'commonsense' is regarded as synonymous with 'good judgement', "
  - "in the AI community it is used in a technical sense to refer to the millions of basic facts and understandings possessed by most people."
  - "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

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)

- ConceptNet
  - 2004: 1.6 million relations among 300,000 nodes
  - 2017: 21 million edges over 8 million nodes
    - 1.5 million nodes are English

Slides credit: Haixun Wang

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
116,097 edges among
74,989 nodes about
events

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

Most Existing KBs are Entity-centric

- 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, Entity, Concept, Instance, etc.),
  - Property,
  - Place,
  - Path,
  - Amount,
  - Activity,
  - State,
  - Event,
  - etc.

Eventuality

How to collect more knowledge about eventualities rather than entities and relations?

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

Outline

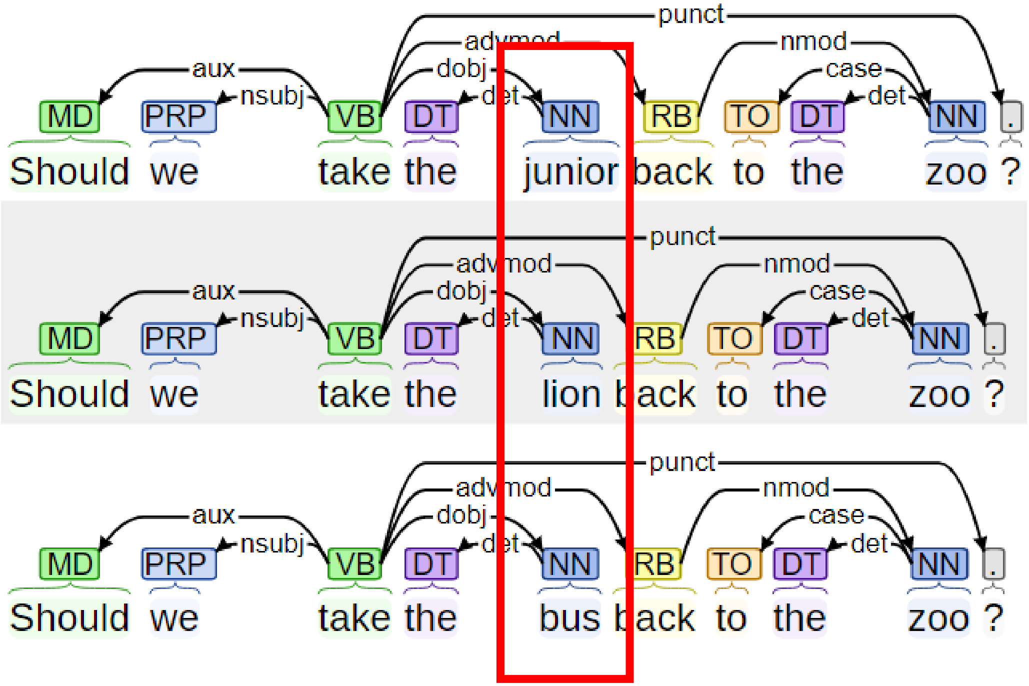
- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Application on the Winograd Schema Challenge
- Extensions

“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)

- Compare semantic meanings
  grammar
  - Syntactically unambiguous

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"

(hurt, X) connection (X, fall)

- 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

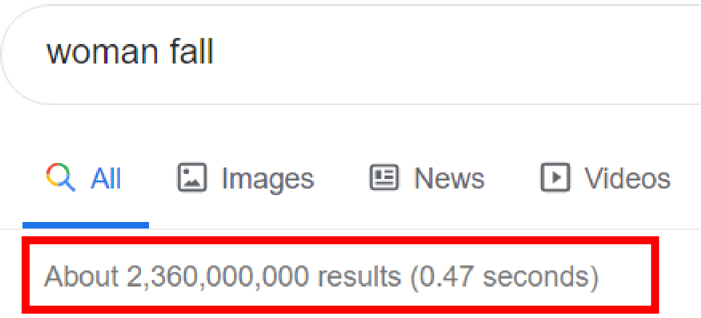
http://commonsensereasoning.org/winograd.html
https://en.wikipedia.org/wiki/Winograd_Schema_Challenge

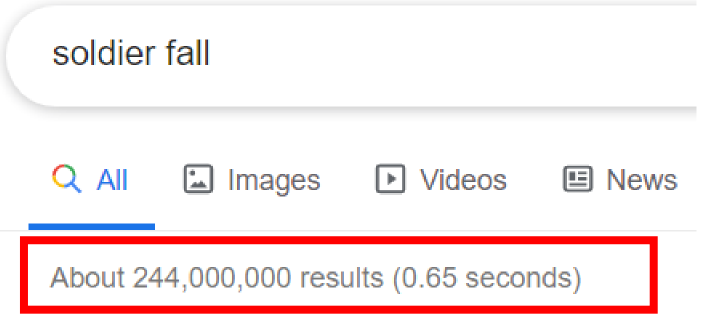
The soldiers fired at the women, and we saw several of them fall.

Why is it a challenge?

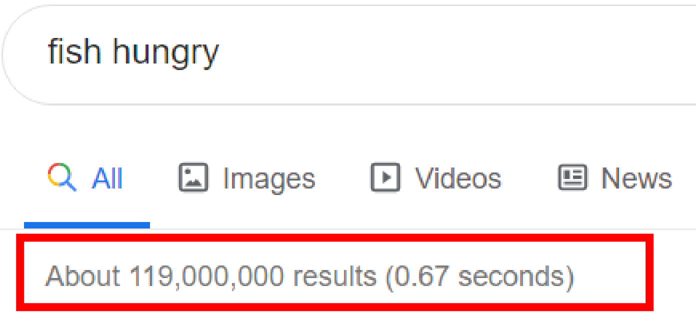
- Must also be carefully written not to betray their answers by selectional restrictions or statistical information about the words in the sentence

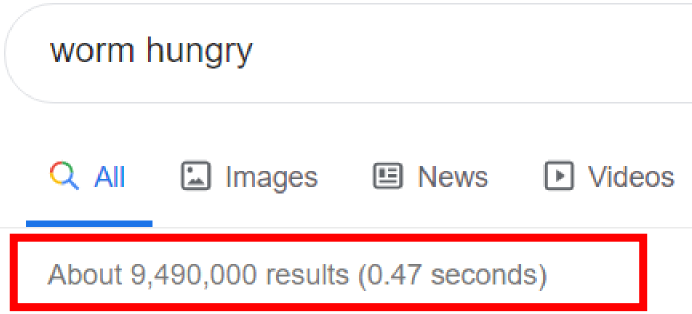
- Designed to be an improvement on the Turing test

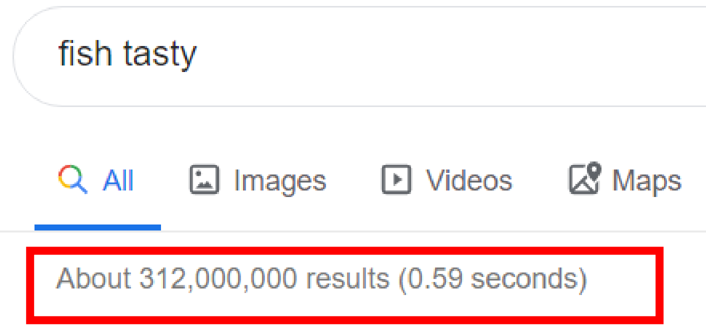




- (A) The fish ate the worm. It was hungry.
- (B) The fish ate the worm. It was tasty.







SP-10K: A Large-scale Evaluation Set of Selectional Preference

- 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

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

Outline

- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Application on the Winograd Schema Challenge
- Extensions

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)

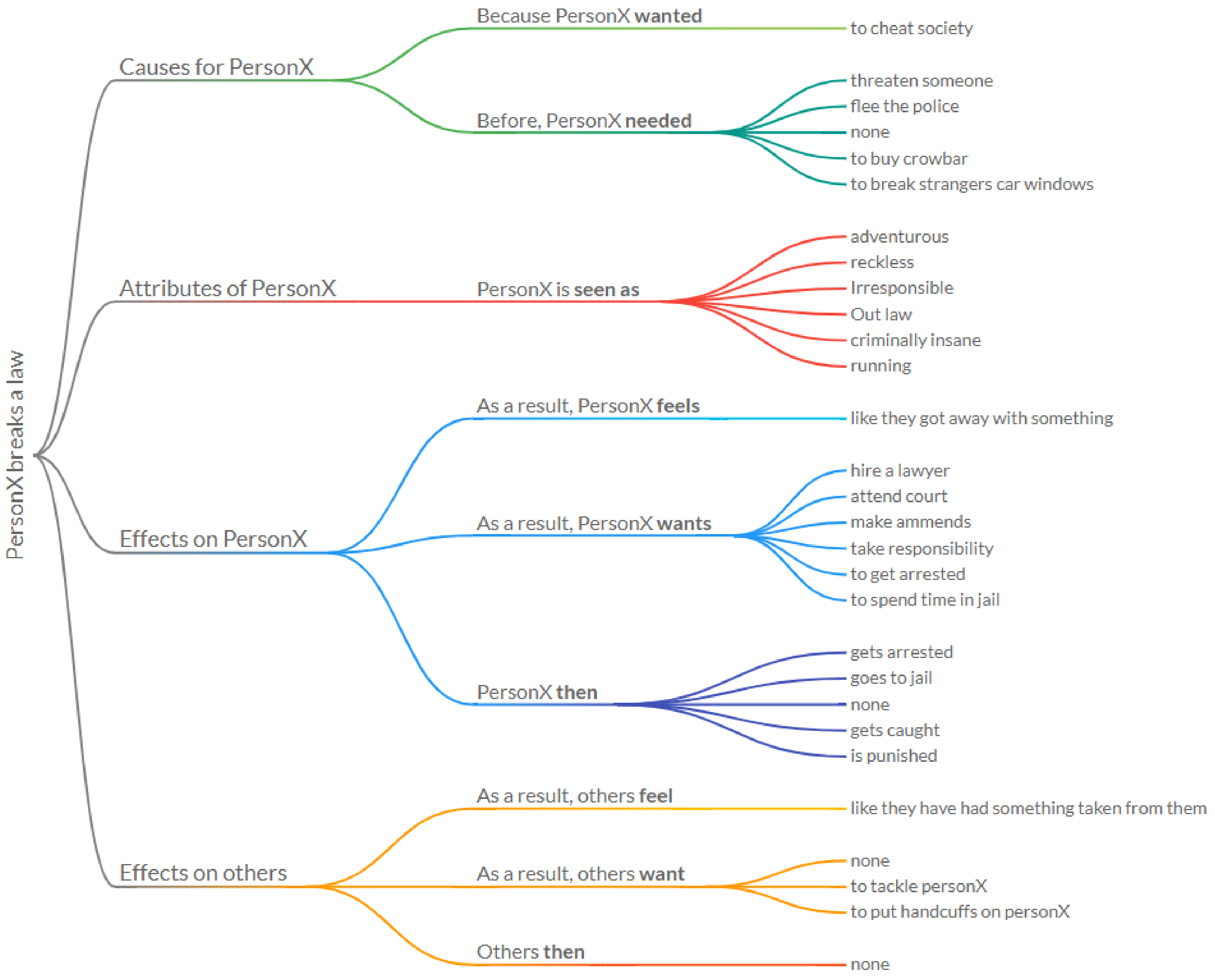
- How to scale up the knowledge acquisition and inference?

ATOMIC

- Crowdsourcing 9 Types of IF-THEN relations

- All personal entity information has been removed to reduce ambiguity

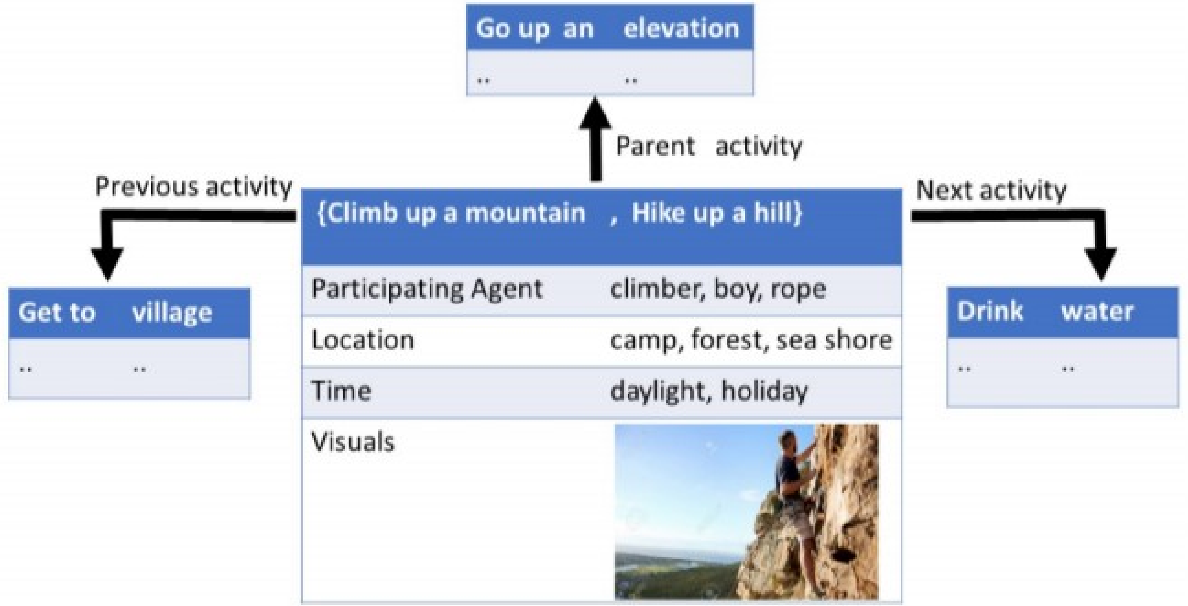
- Arbitrary texts



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
- Only verb+object

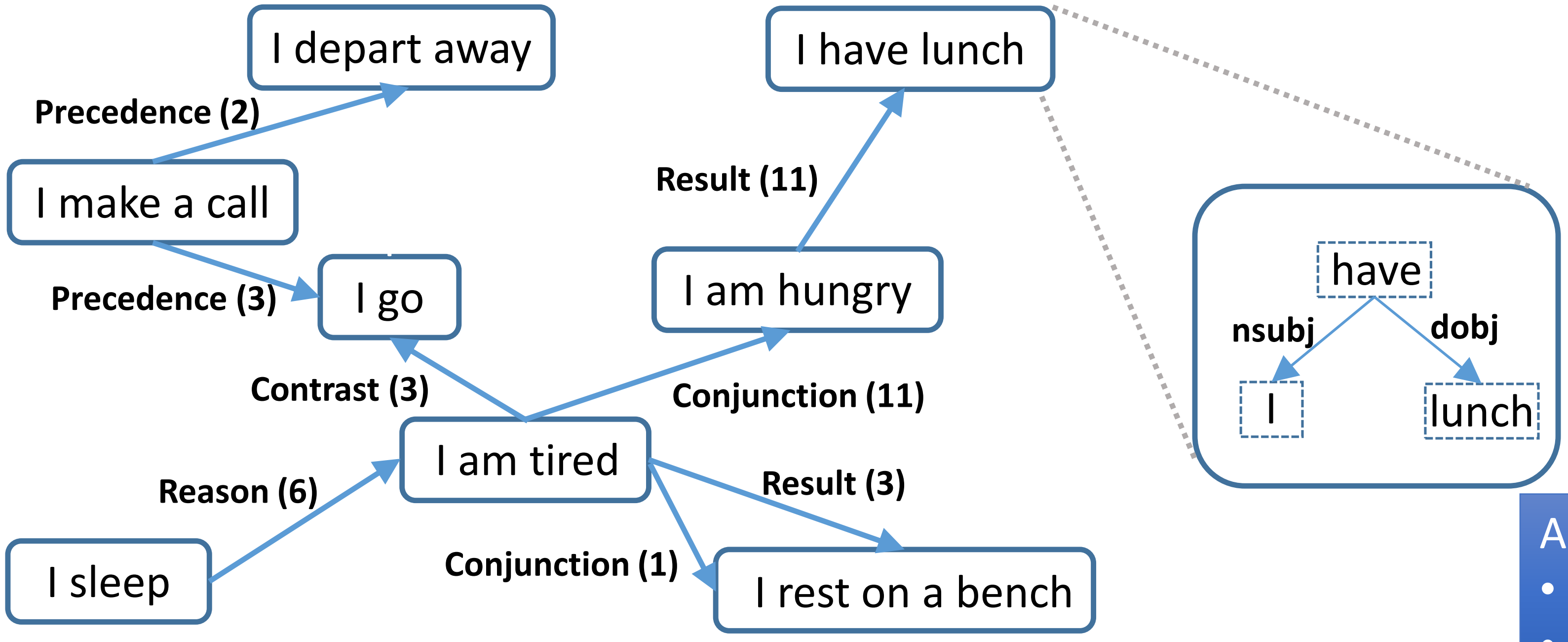


Niket Tandon, Gerard de Melo, Abir De, Gerhard Weikum: Knowlywood: Mining Activity Knowledge From Hollywood Narratives. CIKM 2015: 223-232

A New Knowledge Graph: ASER

Activities, **S**tates, **E**vents, and their **R**elations

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



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

Eventualities

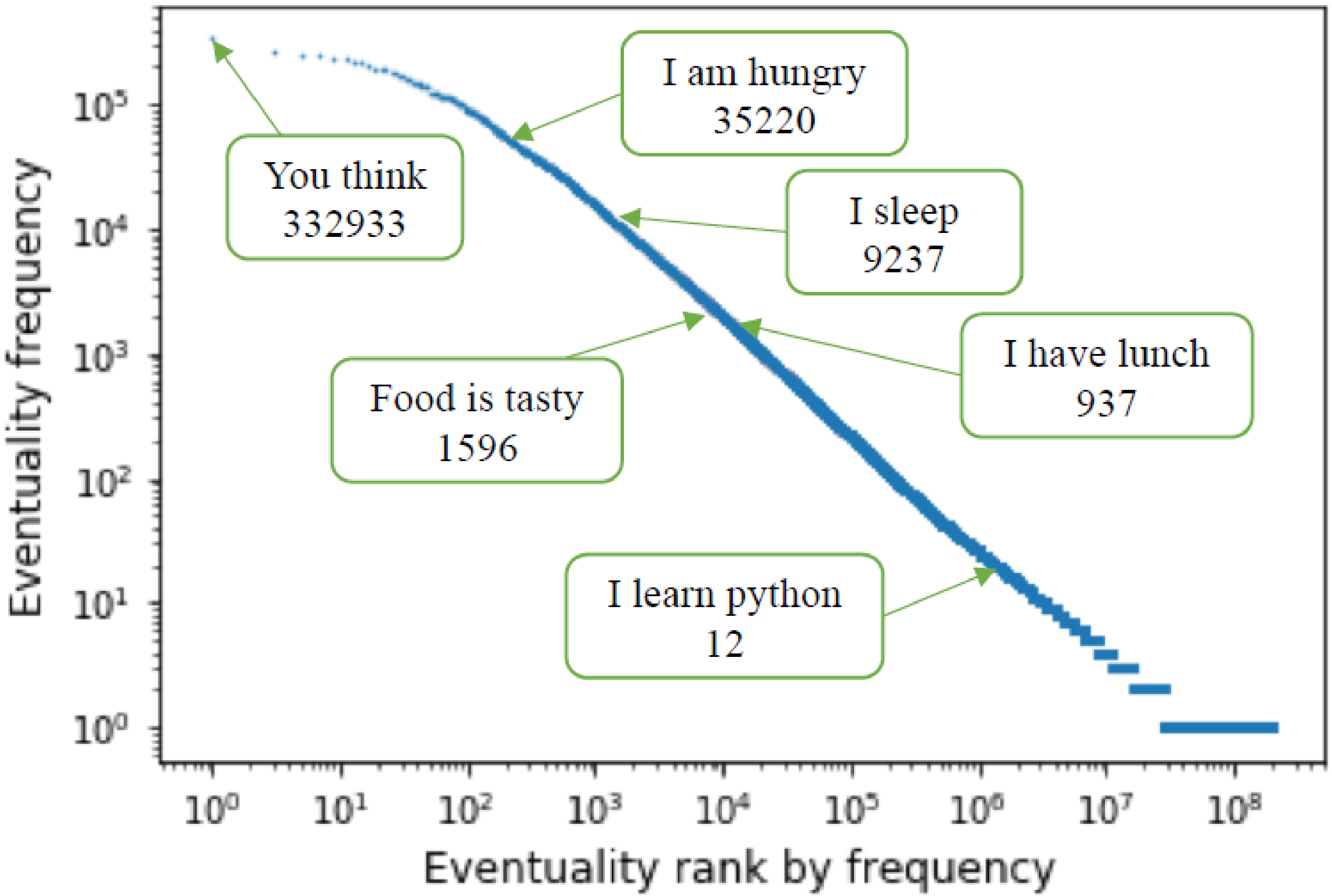
- Using patterns to collect partial information

- 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'

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

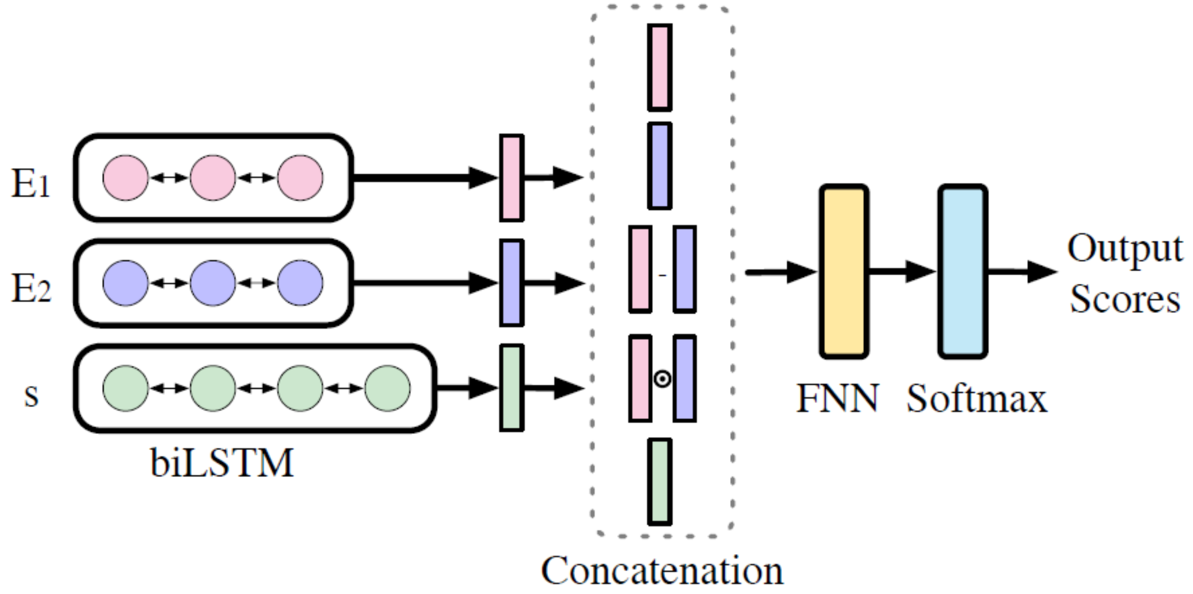
- Seeds from Penn Discourse Treebank (PDTB) (Prasad et al., 2007)
- 14 relations taking from CoNLL shared task
  - More frequent relations
- Less ambiguous connectives
  - ‘so that’ 31 times only in ‘Result’ relations
- Some are ambiguous
  - ‘while’: Conjunction 39 times, Contrast 111 times, Expectation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

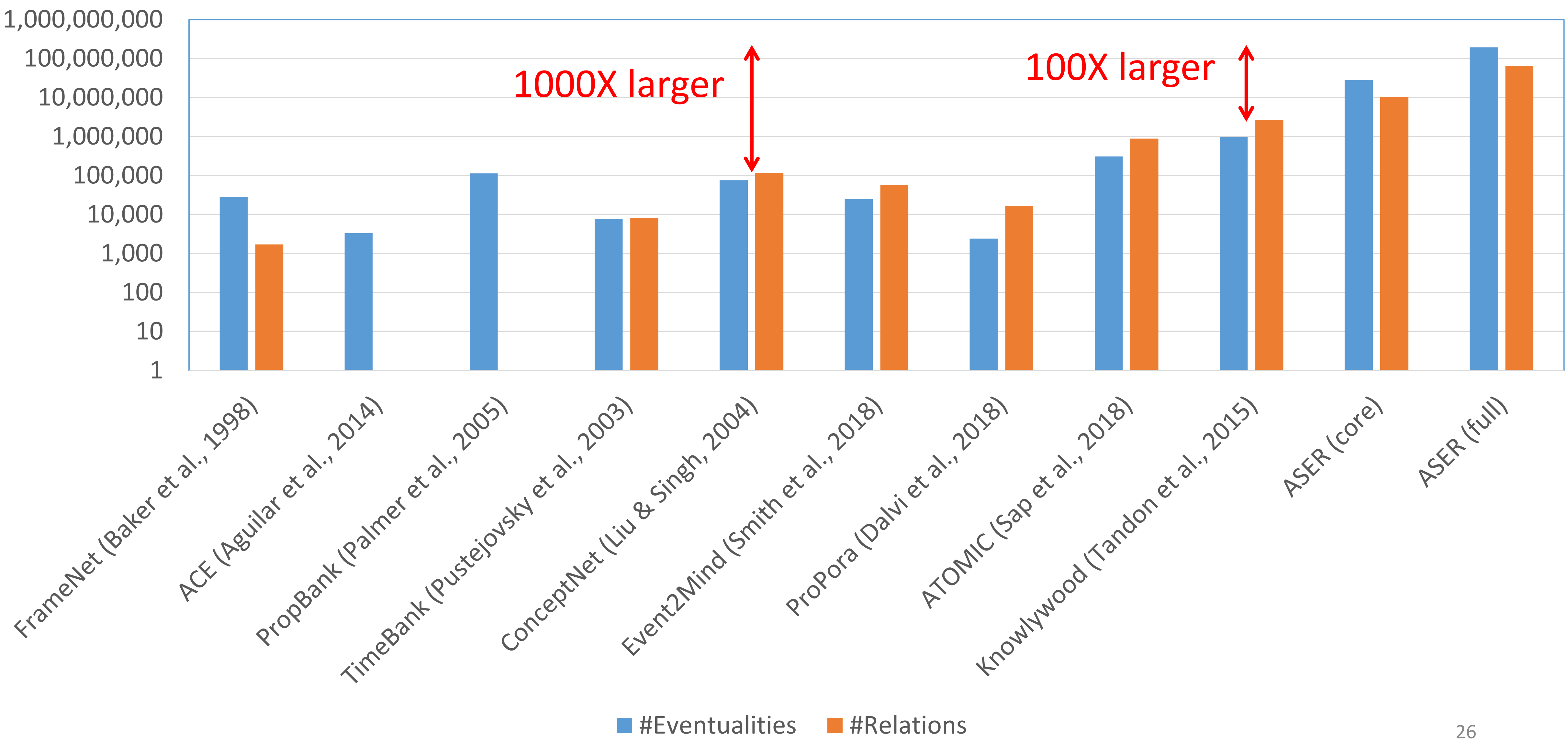
Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., & Webber, B. L. (2007). The penn discourse treebank 2.0 annotation manual.
Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Rashmi Prasad, Christopher Bryant, Attapol T. Rutherford. The CoNLL-2015 Shared Task on Shallow Discourse Parsing.

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



Scales of Verb Related Knowledge Graphs



Multi-hop Reasoning based on Selectional Preference

- One-hop
  - frequency(`sing’-nsubj-`singer’-) > frequency(`sing’-nsubj-`house’)
  - frequency(`eat’-dobj-`food’) > frequency(`eat’-dobj-`rock’)
- Two-hop
  - frequency(`eat’-nsubj-X-amod-`hungry’) > frequency(`eat’-dobj-Y-amod-`hungry’)
- Multi-hop
  - frequency(`X eat dinner’->Causes->`X be full’) > frequency(`X eat dinner’->Causes->`X be hungry’)

Outline

- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Application on the Winograd Schema Challenge
- Extensions

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 on Cases Consistent with Our Patterns

- 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 Accuracy	Overall Accuracy
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%

Hongming Zhang, Hantian Ding, and Yangqiu Song. SP-10K: A Large-Scale Evaluation Set for Selectional Preference Acquisition. ACL, 2019.
Hongming Zhang*, Xin Liu*, Haojie Pan*, Yangqiu Song, and Cane Wing-Ki Leung. ASER: A Large-scale Eventuality Knowledge Graph. WWW. 2020.

Overall Results based on Fine-tuning

Methods	Supervision	Overall Accuracy
Random Guess	NA	50.2%
Knowledge Hunting (Emami et al., 2018)	NA	57.3%
LM (single) (Trinh & Le, 2018)	NA	54.5%
LM (Ensembel) (Trinh & Le, 2018)	NA	61.5%
SP (human) (Zhang et al., 2019)	NA	52.7%
SP (PP) (Zhang et al., 2019)	NA	54.4%
GPT-2 (Radford et al., 2019)	NA	70.7%
BERT (Kocijan et al., 2019)	NA	61.9%
BERT+WSCR (Kocijan et al., 2019)	WSCR	71.4%
ASER (inference)	NA	56.6%
BERT+ASER	WSCR	64.5%
BERT+WSCR+ASER	WSCR+ASER	72.5%

WSCR: Rahman and Ng’s dataset (2012)

ASER: Automatically constructed patterns as training examples

Outline

- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Application on the Winograd Schema Challenge
- Extensions
  - ASER 2.0
  - ASER-EEG
  - TransOMCS

Partial Information Aggregation

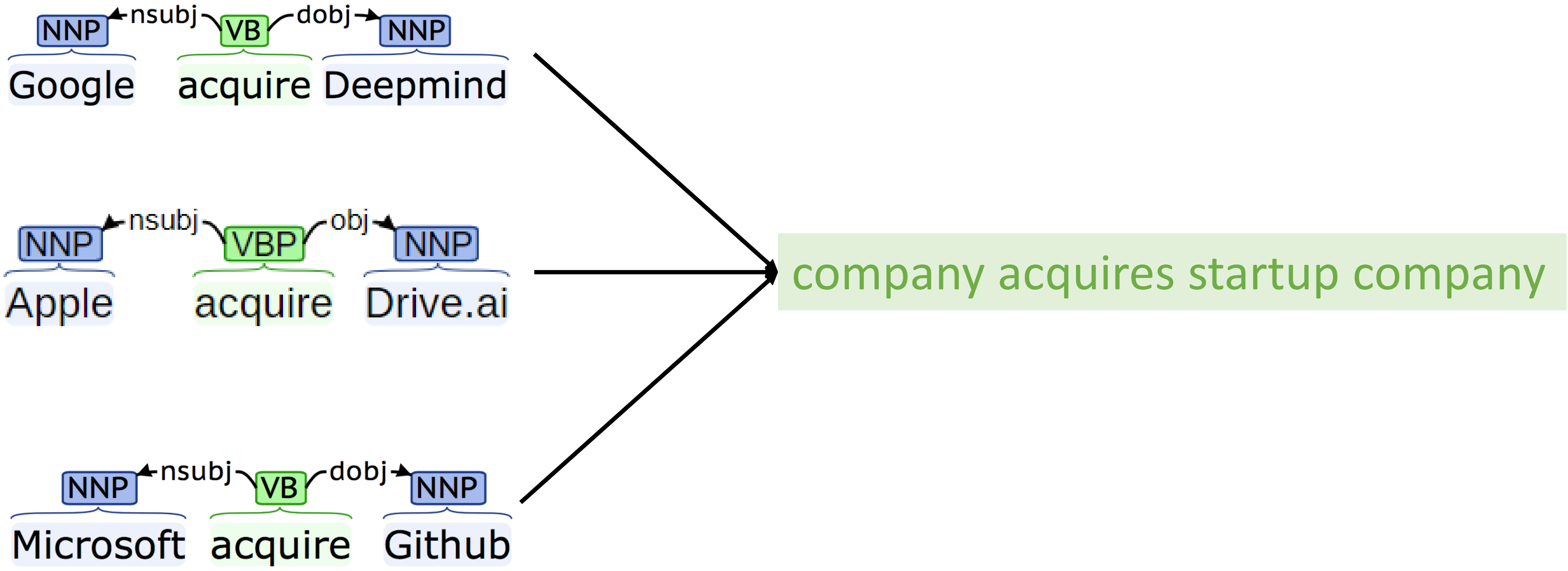
- “hurt things tending to fall down”

(hurt, X) connection (X, fall)

- “stocks price may increase when company X acquire a start-up”

(company, acquire, start-up) result-in (stock, increase)

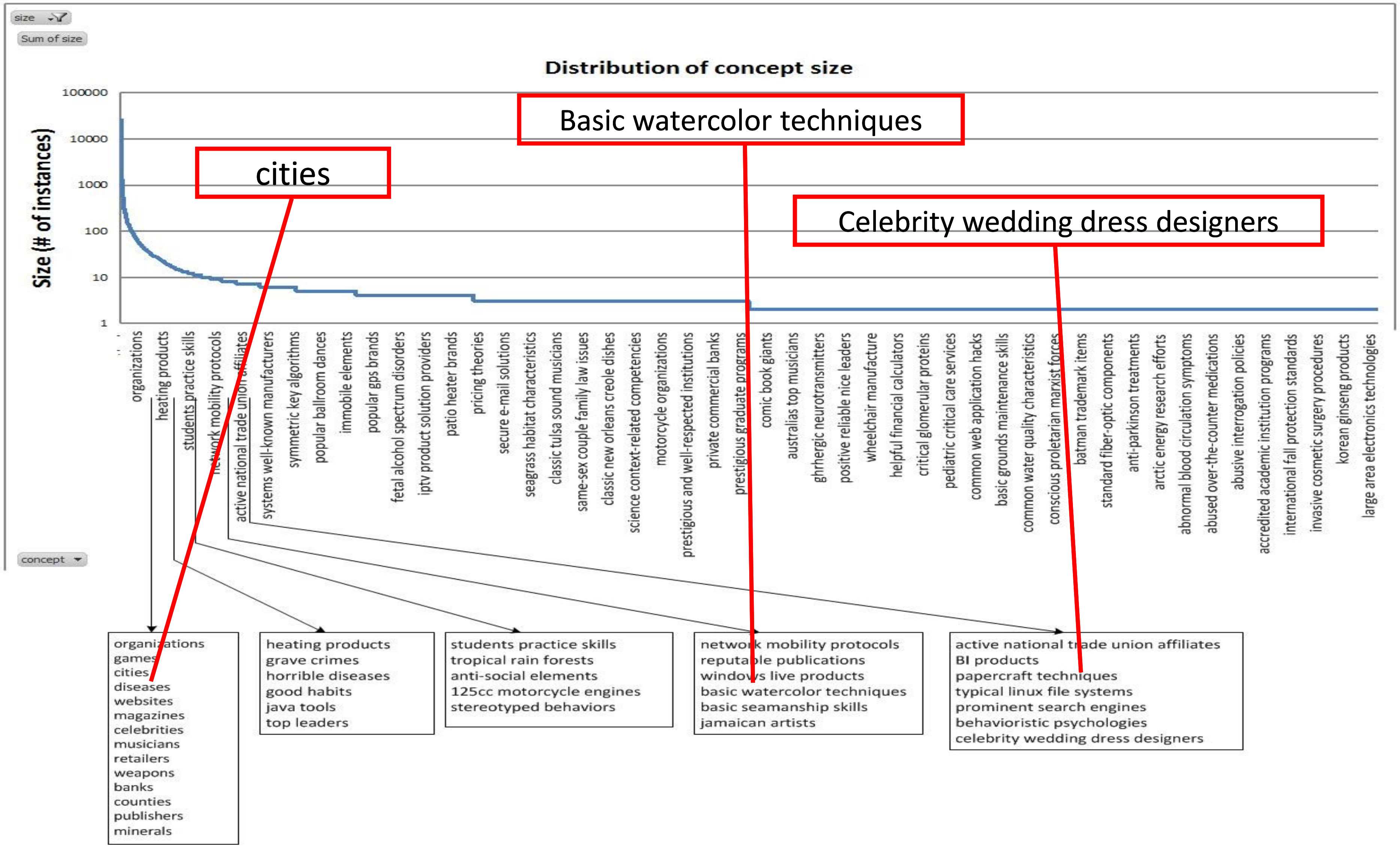
Conceptualization: The Goal

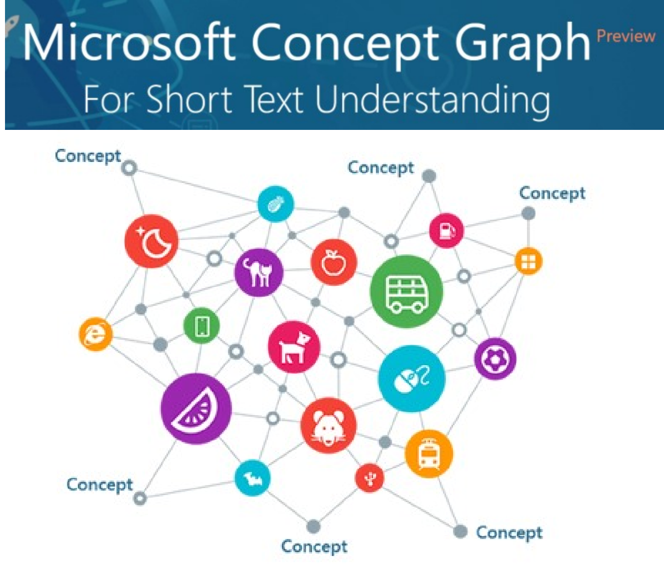


Normalization

			Probability
He, she, I, Bob, …	→	__PERSON__	1.0
1996, 2020, 1949, …	→	__YEAR__	1.0
23, 20, 333, ….	→	__DIGIT__	1.0
www.google.com, …	→	__URL__	1.0

Conceptualization with ProBase





Concepts are the glue that holds our mental world together.

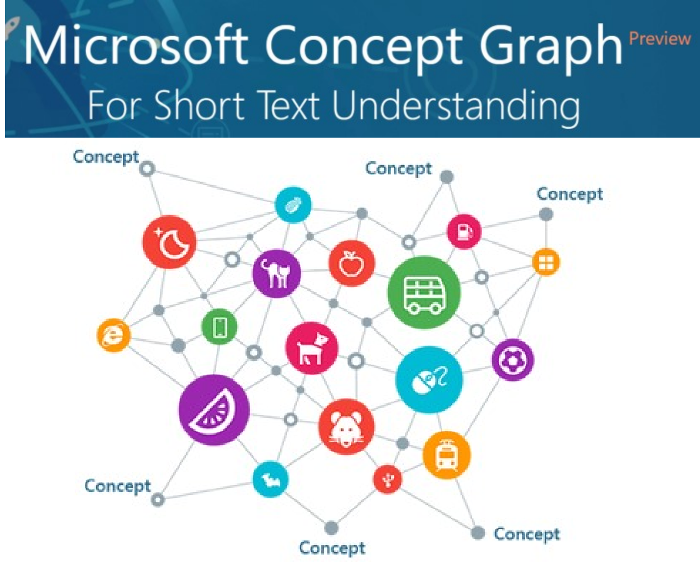
Gregory L. Murphy
NYU

Probase is a *large, universal, probabilistic* knowledge base with **an extremely large concept space**

Data are available at https://concept.research.microsoft.com/

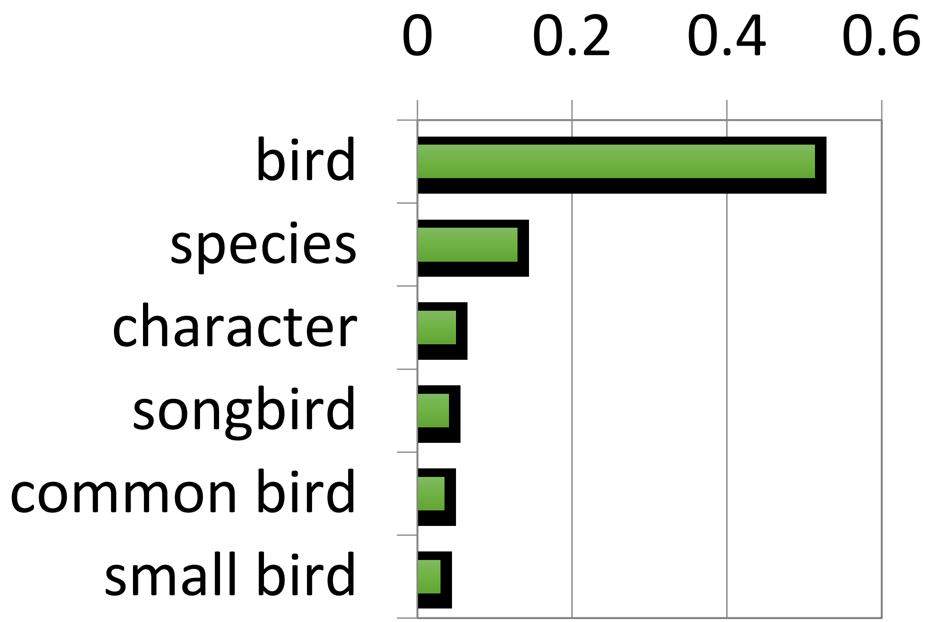
Wentao Wu, Hongsong Li, Haixun Wang, Kenny Qili Zhu: Probase: a probabilistic taxonomy for text understanding. SIGMOD Conference 2012: 481-492

Conceptualization with ProBase

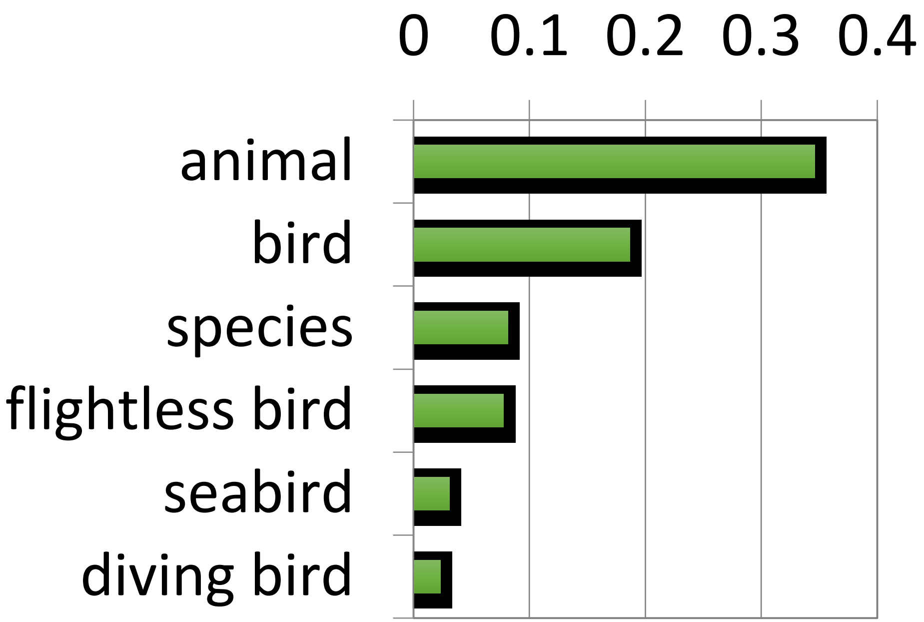


Typicality $$P(concept \mid instance) = \frac{\#(concept, instance)}{\#(instance)}$$

- Robin



- Penguin



Data are available at https://concept.research.microsoft.com/

Wentao Wu, Hongsong Li, Haixun Wang, Kenny Qili Zhu: Probase: a probabilistic taxonomy for text understanding. SIGMOD Conference 2012: 481-492
Yangqiu Song, Haixun Wang, Zhongyuan Wang, Hongsong Li, Weizhu Chen: Short Text Conceptualization Using a Probabilistic Knowledgebase. IJCAI 2011: 2330-2336

A Running Example

Obama

(politician, 0.0855)
(democrat, 0.0560)
(liberal, 0.0560)

dog

(animal, 0.2811)
(pet, 0.1377)
(domestic animal, 0.0525)

(Obama, have, dog)

(obama have animal, 0.2811)
(obama have pet, 0.1377)
(politician have dog, 0.0855)
(democrat have dog, 0.05604)
...
(politician have animal, 0.0240)
(democrat have animal, 0.01575)
...

$$\prod_{i=1}^{N} P(C_{i,k} | E_i)$$

$$P(politician \mid Obama) \times P(animal \mid dog) = 0.0855 \times 0.2811 = 0.0240$$

A Running Example

Obama

(politician, 0.0855)
(democrat, 0.0560)
(liberal, 0.0560)

dog

(animal, 0.2811)
(pet, 0.1377)
(domestic animal, 0.0525)

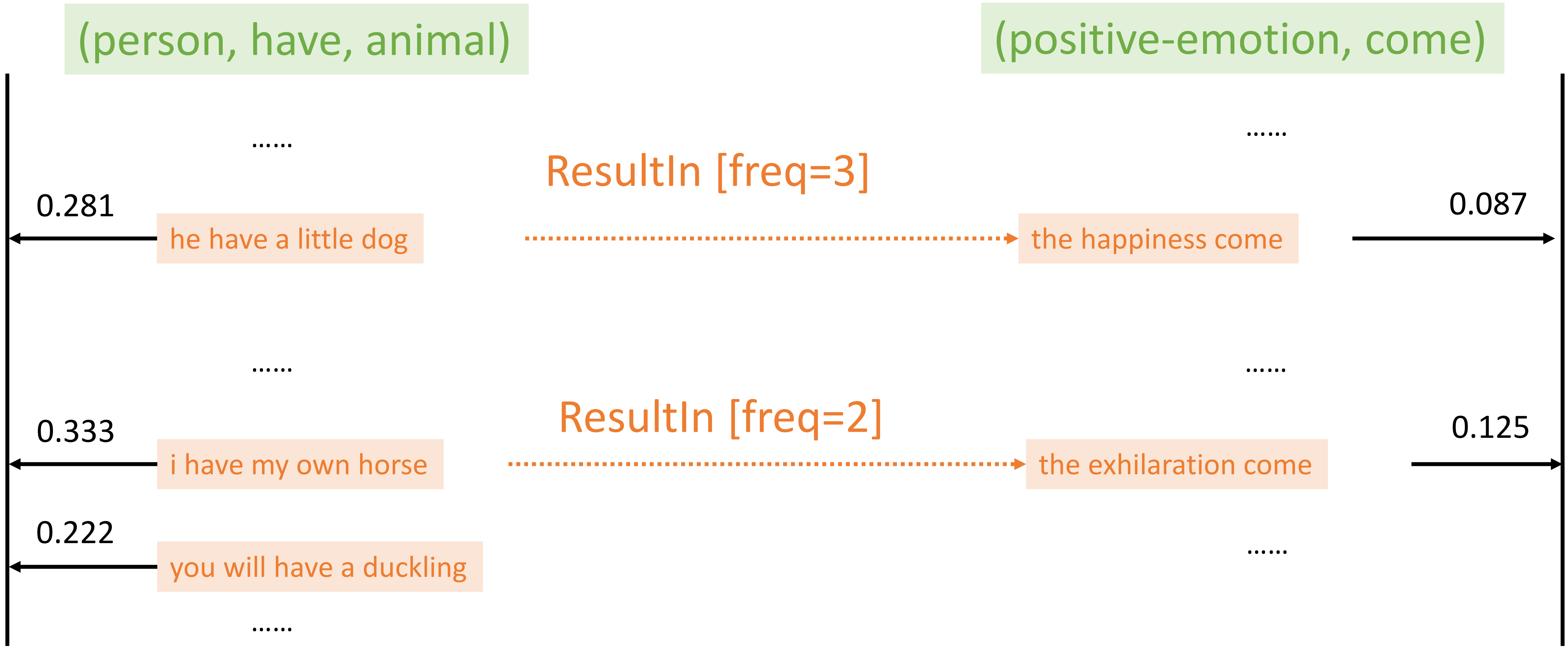
(Obama, have, dog)

(obama have animal, 0.2811)
(obama have pet, 0.1377)
(politician have dog, 0.0855)
(democrat have dog, 0.05604)
...
(politician have animal, 0.0240)
(democrat have animal, 0.01575)
...

Number of ASER-concepts:

$$C_N^1 \times K + C_N^2 \times K^2 + \cdots + C_N^N K^N$$

K is Top K probase-concept for each entity, N is #entity in an eventuality



$$P(\text{ResultIn} \mid \text{(person, have, animal)}, \text{(positive-emotion, come)}) = 0.281 \times 3 \times 0.087 + 0.333 \times 2 \times 0.125 = 0.157$$

ASER 2.0

- Rule based extraction (14 Eventuality Patterns, Improved Version)

Data	#Eventualities	#Unique Eventualities	#Relations	#Unique Relations
Core	349,296,240	34,212,258	65,997,575	15,339,027
Full	587,290,657	272,206,675	265,681,802	205,758,398

- Discourse Parser (18 Eventuality Patterns + Wang and Lan 2015)

Data	#Eventualities	#Unique Eventualities	#Relations	#Unique Relations
Core	477,383,662	42,964,177	120,995,415	25,880,127
Full	799,191,666	364,772,181	463,640,100	368,635,332

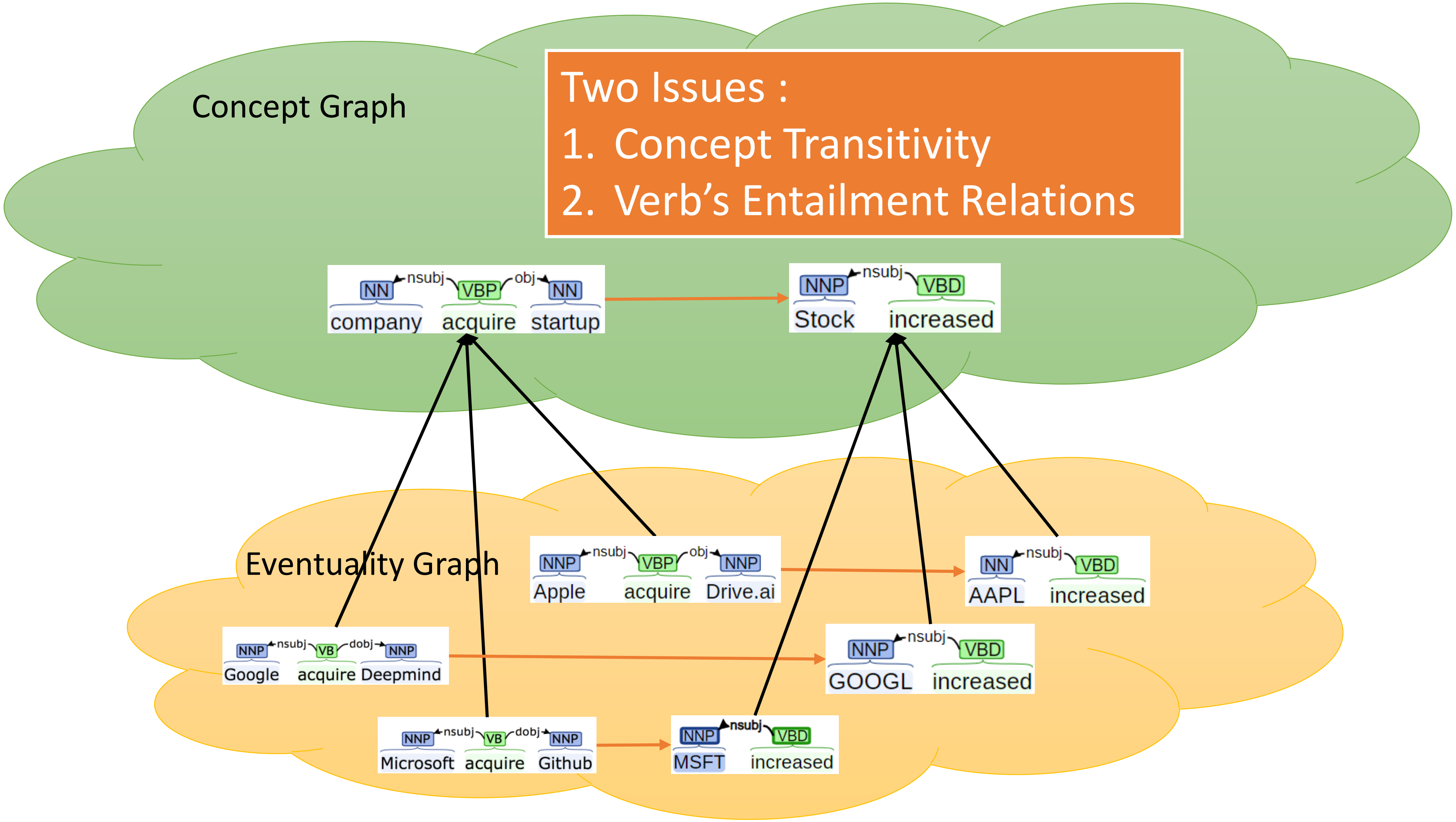
- Conceptualization Core:
  - Concepts: 65,837,819 (1.5 times larger)
  - Concept Relations: 289,735,387 (11 times larger)

Jianxiang Wang and Man Lan. A Refined End-to-End Discourse Parser. CONLL Shared Task 2015.

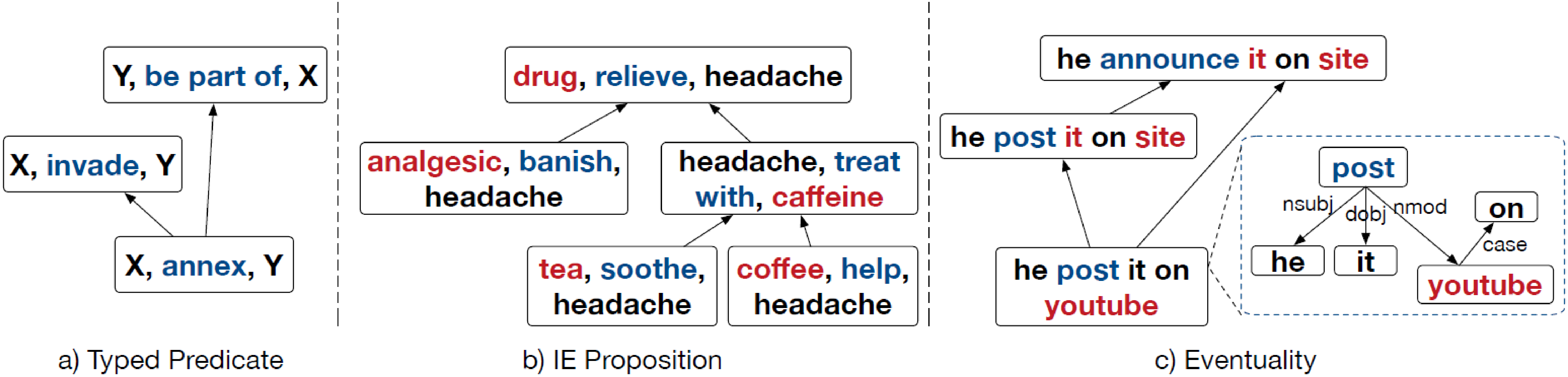
Outline

- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Application on the Winograd Schema Challenge
- Extensions
  - ASER 2.0
  - ASER-EEG
  - TransOMCS

Incorporating More Relations

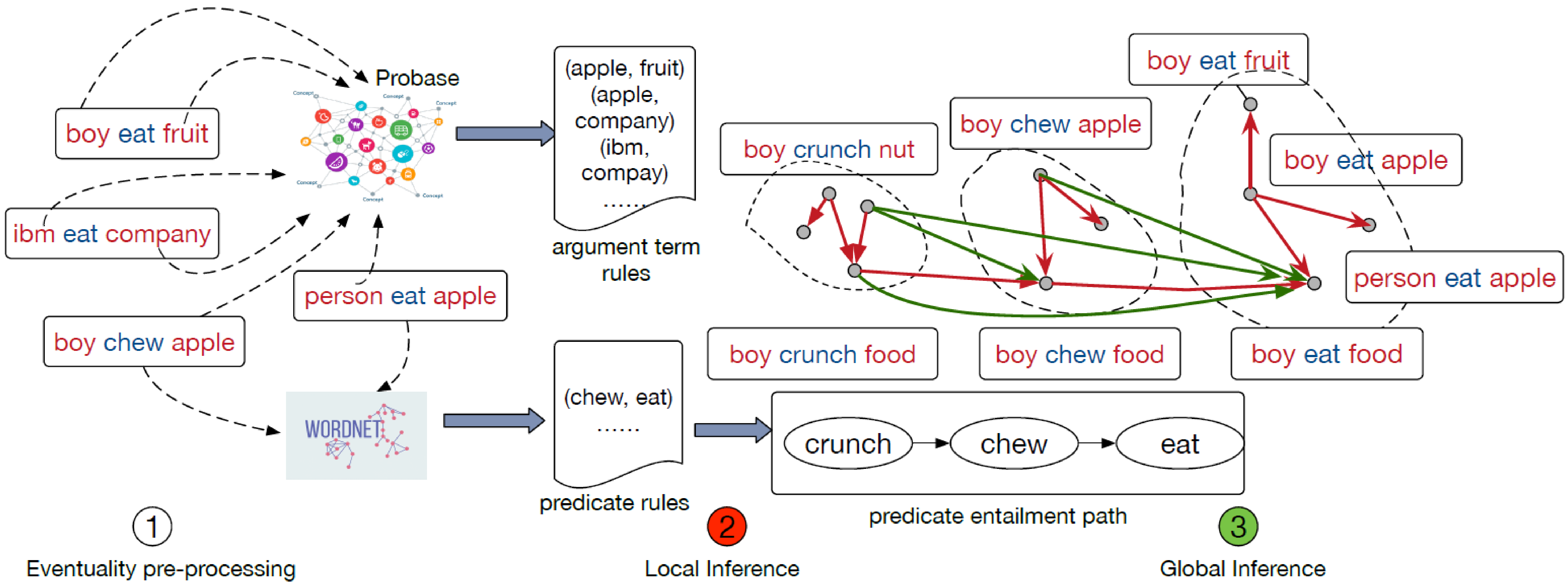


Entailment Graph Construction



Node Type	Reference	#Graphs	#Nodes	#Edges	Domain
Typed Predicate	Berant et al., ACL, 2011 Hosseini et al. TACL, 2018	2,303 363	10,672 101K	263,756 66M	Place/disease News
Open IE Proposition	Levy et al., CoNLL, 2014	30	5,714	1.5M	Healthcare
Eventuality	Ours	473	10M	103M	Commonsense

Three-step Construction



Changlong Yu, Hongming Zhang, Yangqiu Song, Wilfred Ng, Lifeng Shang . Enriching Large-Scale Eventuality Knowledge Graph with Entailment Relations. AKBC. 2020.

Results

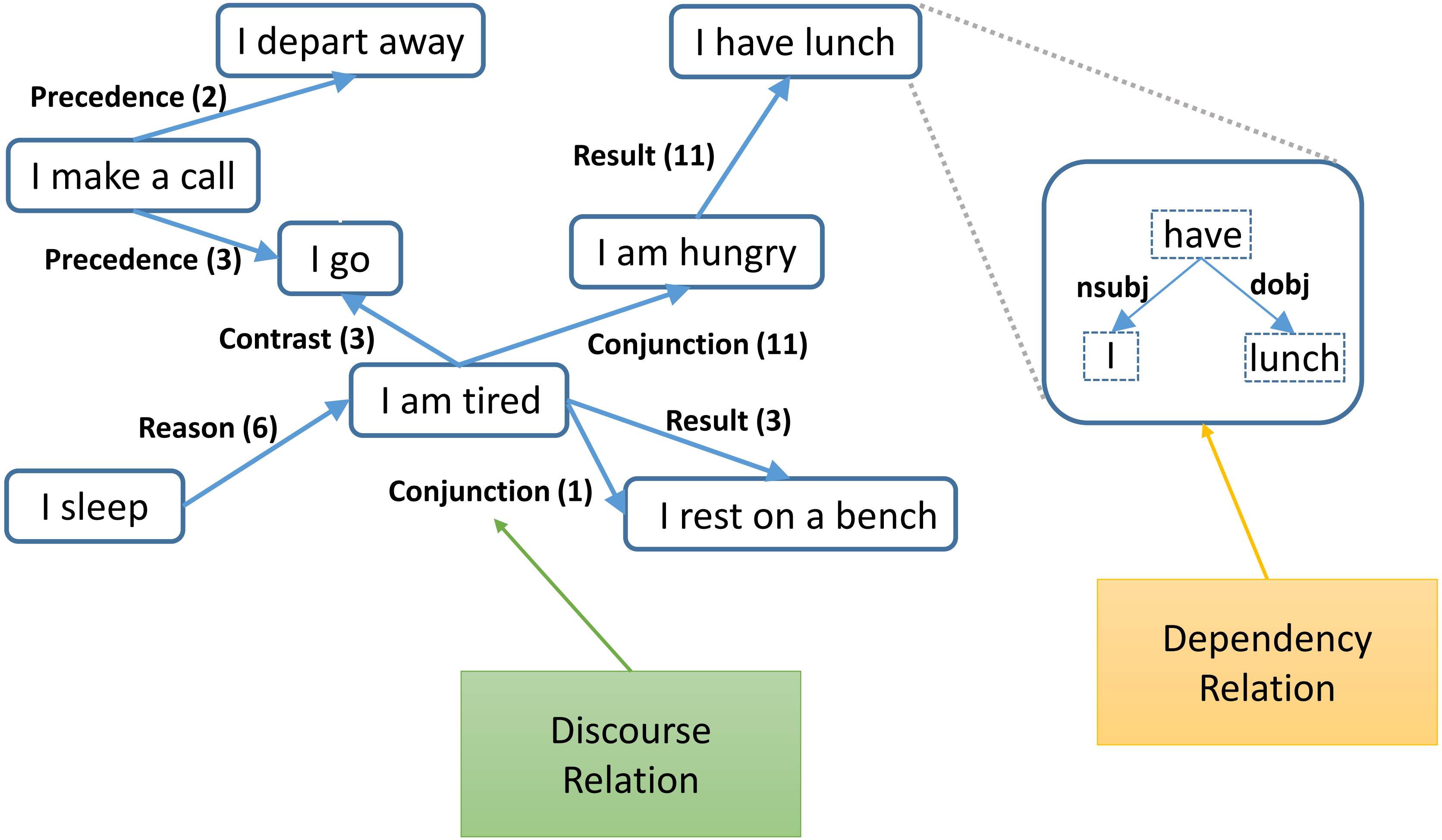
- We can generate 10 times of edges

	# Eventuality	# ER (global)	# ER (local)	Acc (local)	Acc (all)
s-v ⊨ s-v	3.3M	32.7M	10.7M	89.1%	85.7%
s-v-o ⊨ s-v-o	5.3M	45.2M	14.8M	90.1%	89.3%
s-v-p-o ⊨ s-v-p-o	1.9M	12.6M	5.3M	88.3%	87.4%
s-v-o-p-o ⊨ s-v-o	0.5M	0.8M	0.8M	91.4%	90.0%
s-v-p-o ⊨ s-v-o	1.1M	2.7M	0.9M	88.5%	87.2%
s-v-o ⊨ s-v-p-o	0.9M	5.4M	2.2M	87.8%	86.7%
s-v-o-p-o ⊨ s-v-o-p-o	2.4M	3.2M	2.1M	89.4%	88.4%
s-v-a ⊨ s-be-a	0.2M	0.1M	0.1M	97.9%	97.9%
s-be-a-p-o ⊨ s-be-a	0.8M	0.4M	0.4M	96.0%	95.8%
s-be-a-p-o ⊨ s-be-a-p-o	0.1M	0.1M	0.1M	95.1%	94.7%
Overall	10.0M	103.2M	37.4M	91.4%	90.3%

Outline

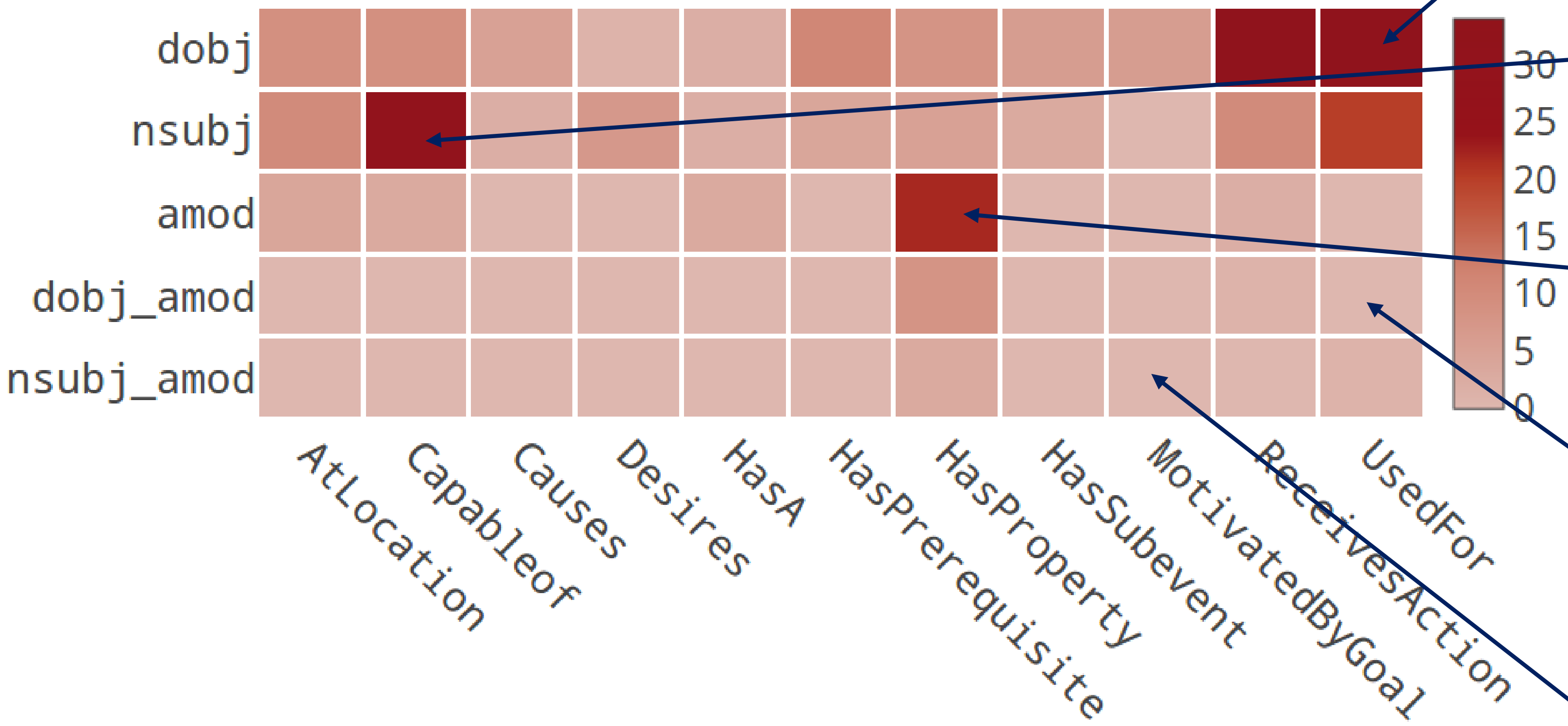
- Motivation: NLP and commonsense knowledge
- Consideration: selectional preference
- New proposal: large-scale and higher-order selectional preference
- Application on the Winograd Schema Challenge
- Extensions
    - ASER 2.0
    - ASER-EEG
    - TransOMCS

ASER is Essentially a Knowledge Graph based on Linguistics



How is it transferrable from linguistic knowledge to existing definition of commonsense knowledge?

Revisit the Correlations of SP and OMCS



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

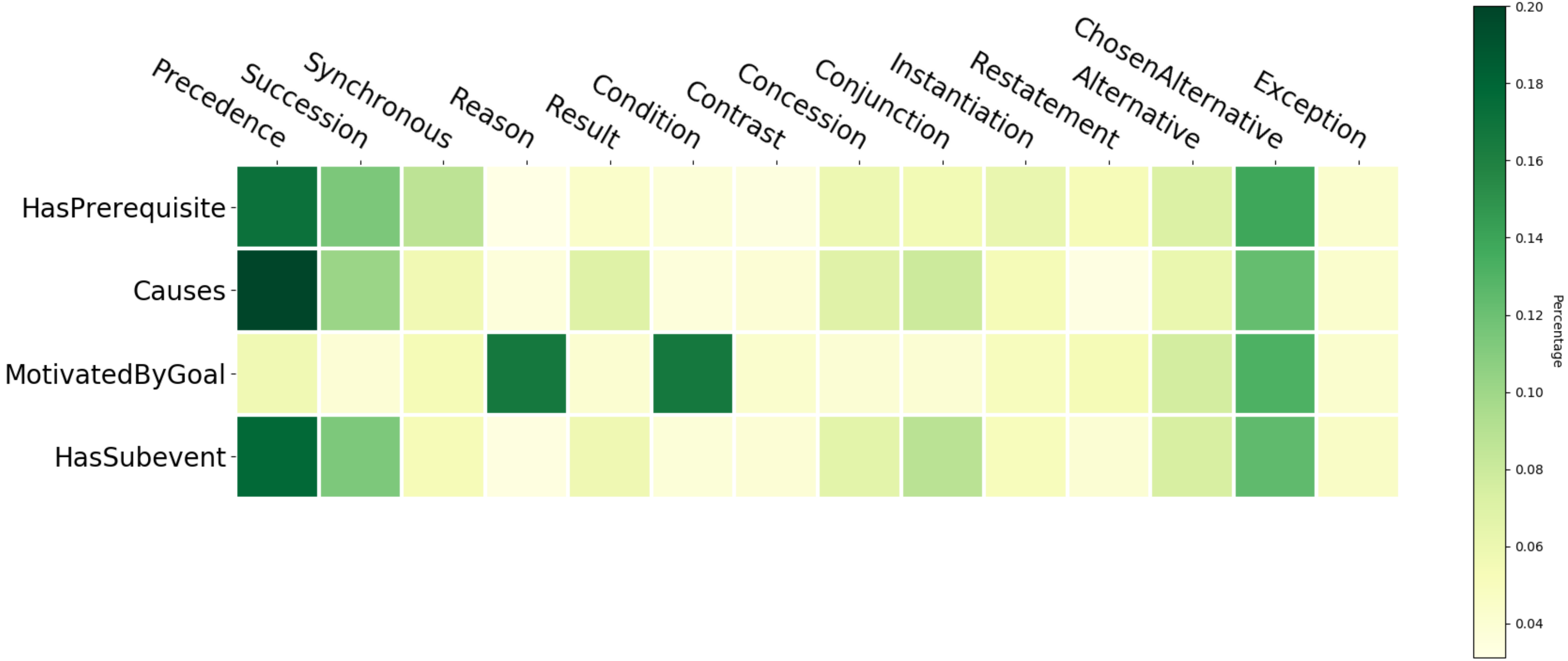
(phone, ring) (nsubj, 8.75)
(phone, CapableOf, ring)

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

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

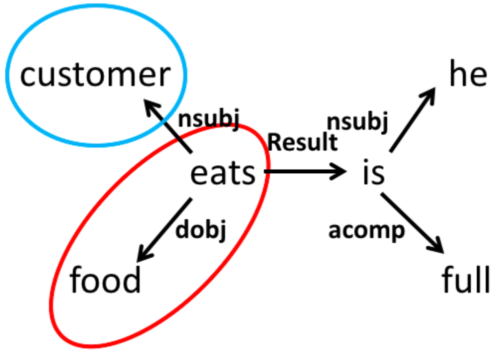
(hungry, eat) (nsubj_amod, 10.00)
(eat, MotivatedByGoal, are hungry)

Revisit the Correlations of ASER and OMCS



Can we Discover more OMCS Knowledge from ASER?

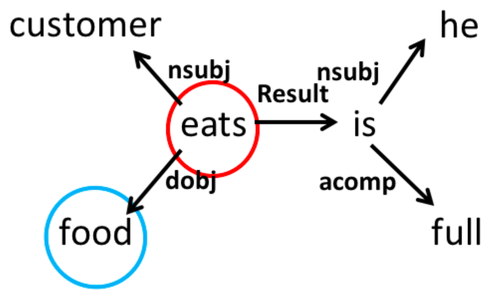
Step 1: Pattern mining by heuristic scoring
Step 2: Learning to rank from 1,000 annotated tuples in each relation



Relation: CapableOf

Pattern: ()<-nsubj<-(-dobj-)

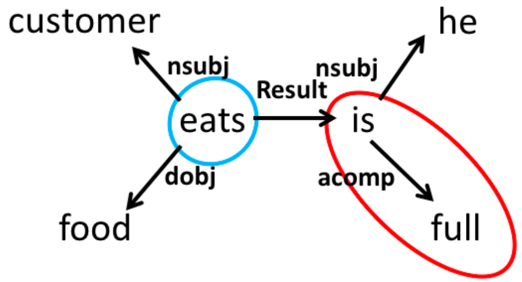
Knowledge: (‘customer’-*CapableOf*-‘eat food’)



Relation: ReceivesAction

Pattern: ()<-dobj<-()

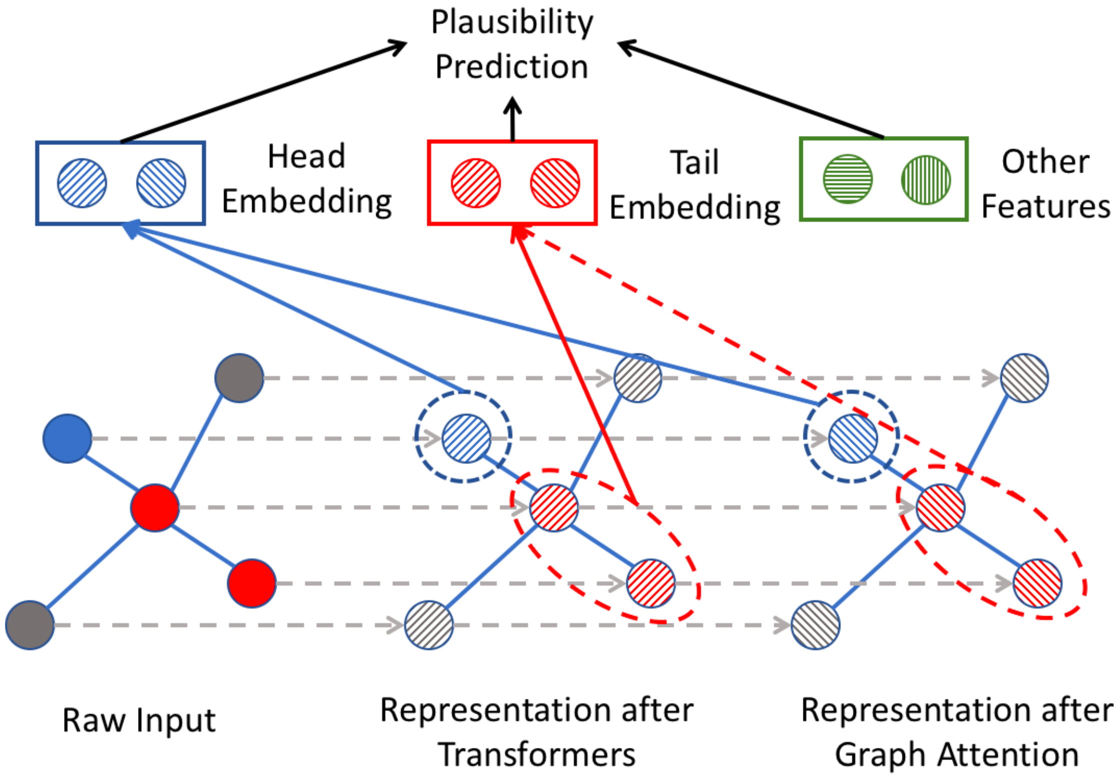
Knowledge: (‘food’-*ReceivesAction*-‘eat’)



Relation: Causes

Pattern: ()->Result->(-acomp-)

Knowledge: (‘eat’-*Causes*-‘be full’)

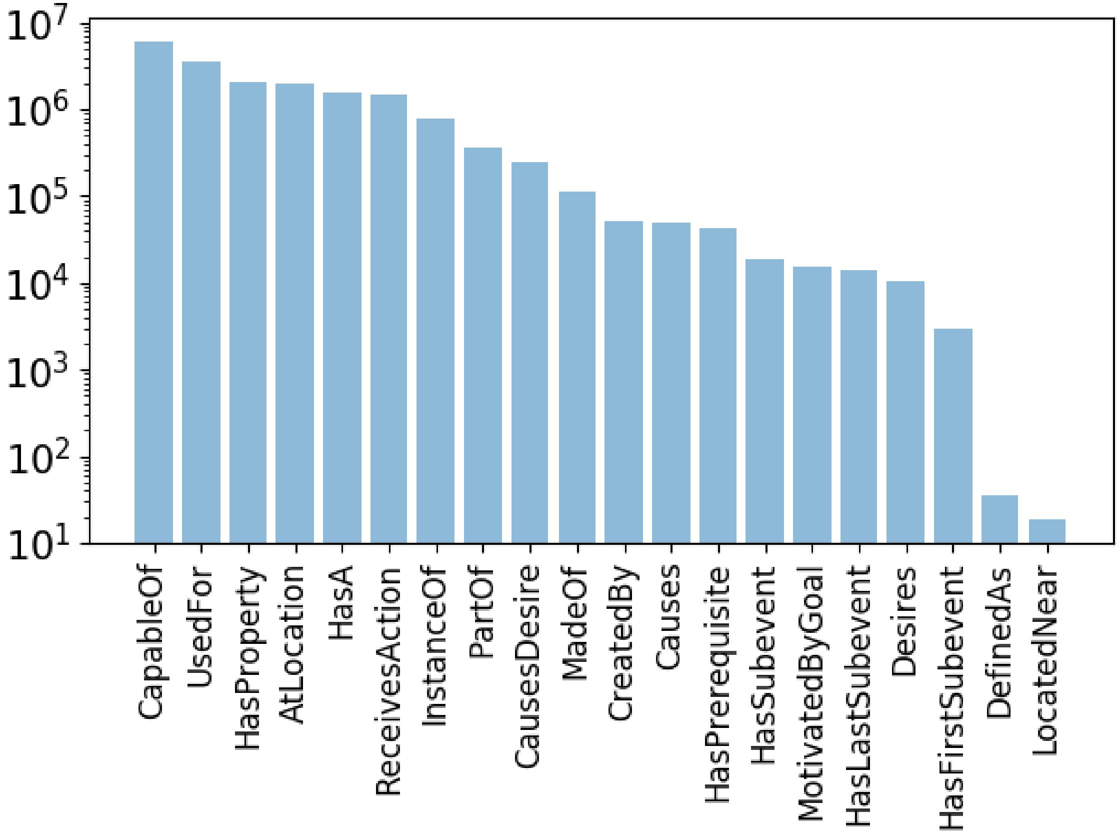


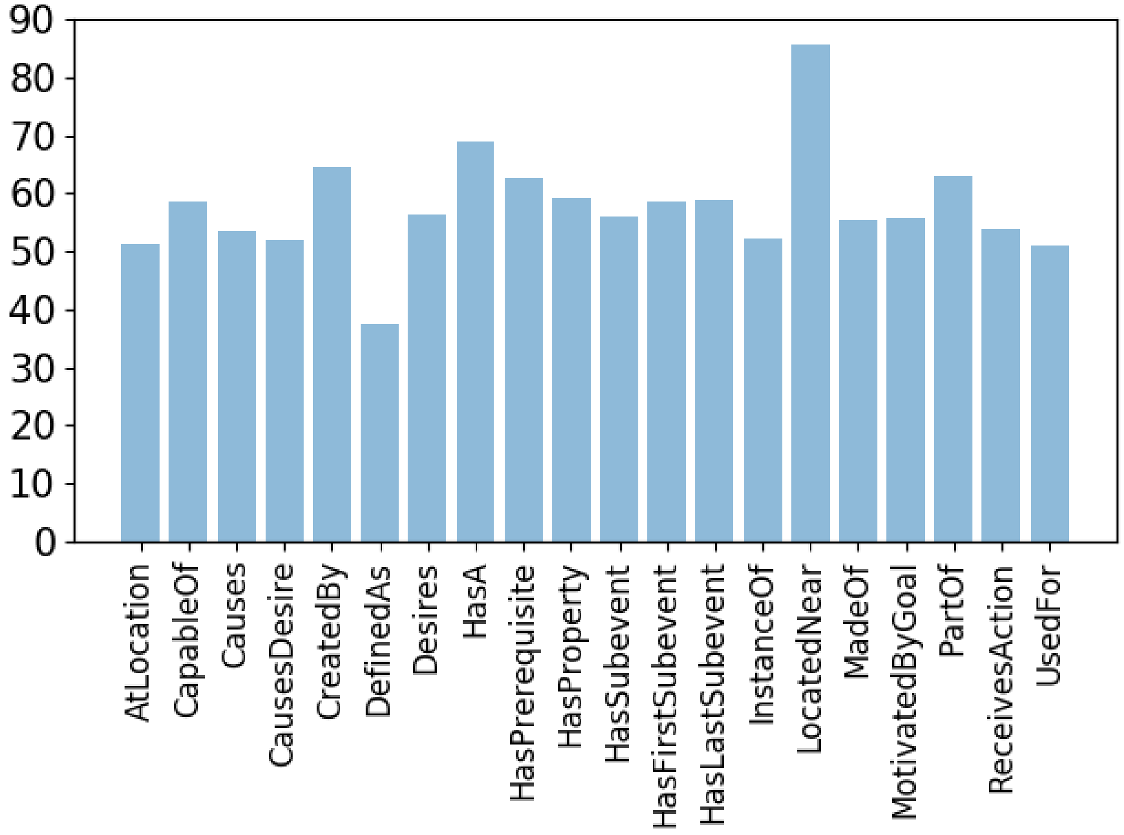
Hongming Zhang, Daniel Khashabi, Yangqiu Song, and Dan Roth. TransOMCS: From Linguistic Graphs to Commonsense Knowledge. IJCAI. 2020.

Model	# Vocab	# Tuple	Novel (Tuple)	Novel (Concept)	ACC (Novel)	ACC (Overall)
COMET 1.2K ConceptNet Test Set (Greedy)	715	1,200	33.96%	5.27%	58%	90%
COMET 1.2K ConceptNet Test Set (10 Beams)	2,232	12,000	64.95%	27.15%	35%	44%
COMET 24K ASER Sampled Graphs (Greedy)	3,912	24,000	99.98%	55.56%	34%	47%
COMET 24K ASER Sampled Graphs (10 Beams)	8,108	240,000	99.98%	78.59%	23%	27%
LAMA 1.2K ConceptNet Test Set (Top 1)	328	1,200	-	-	-	49%
LAMA 1.2K ConceptNet Test Set (Top 10)	1,649	12,000	-	-	-	20%
LAMA 1.2K ASER Sampled Graphs (Top 1)	1,443	24,000	-	-	-	29%
LAMA 1.2K ASER Sampled Graphs (Top 10)	5,464	240,000	-	-	-	10%
TransOMCS overlapped with 1.2K ConceptNet	33,238	533,449	99.53%	89.20%	72%	74%
TransOMCS (Top 1%)	37,517	184,816	95.71%	75.65%	86%	87%
TransOMCS (Top 10%)	56,411	1,848,160	99.55%	92.17%	69%	74%
TransOMCS (Top 30%)	68,438	5,544,482	99.83%	95.22%	67%	69%
TransOMCS (Top 50%)	83,823	9,240,803	99.89%	96.32%	60%	62%
TransOMCS (no ranking)	100,659	18,481,607	99.94%	98.30%	54%	56%
OMCS in ConceptNet 5.0	36,954	207,427	-	-	-	92%

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. COMET: commonsense transformers for automatic knowledge graph construction. ACL, 2019.
Fabio Petroni, Tim Rocktaschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander H. Miller. Language models as knowledge bases? EMNLP-IJCNLP, 2019.

Distribution of Relations and Accuracy





Case Studies

"human" *CapableOf*

COMET

1. kill other person
2. kill other human
3. kill other sentient be 🤔
4. feel emotion
5. kill other human be 🤔
6. make wine
7. hate
8. love
9. think
10. die

LAMA

1. be 🤔
2. fly 🤔
3. die
4. talk
5. kill
6. speak
7. breathe
8. eat
9. think
10. see

TransOMCS

1. stand
2. think
3. die
4. learn
5. make mistake
6. lie
7. typically have 🤔
8. create society
9. have cell
10. create life

(a) Original Setting

"love" *Causes*

COMET

1. happiness
2. be happy
3. get marry
4. death 🤔
5. you get marry
6. you feel good
7. pain
8. love 🤔
9. life 🤔
10. war

LAMA

1. chaos 🤔
2. pain
3. problems
4. love 🤔
5. trouble
6. death 🤔
7. fear 🤔
8. happiness
9. war
10. conflict

TransOMCS

1. be friendly
2. be happy
3. pain
4. marriage
5. be quaint 🤔
6. be unhappy
7. be allergic 🤔
8. be desperate
9. be apart
10. be silly

(b) Extended Setting

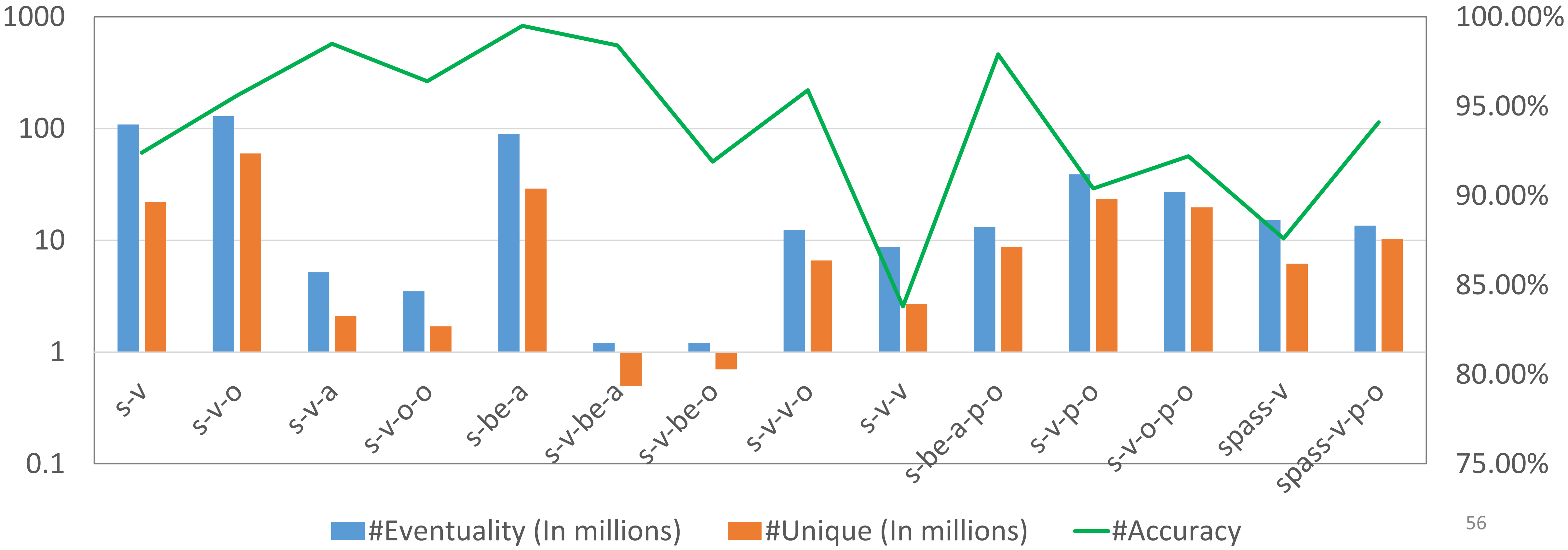
Conclusions

- We extended the concept of selectional preference for commonsense knowledge acquisition
- Many potential extensions
  - More links with knowledge base completion and population
  - Many downstream tasks
- Project Homepage
  - https://hkust-knowcomp.github.io/ASER/

Thank you ☺

Eventuality 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



Relation 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