# Commonsense Knowledge Base Population

Yangqiu Song

Department of CSE, HKUST, Hong Kong

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

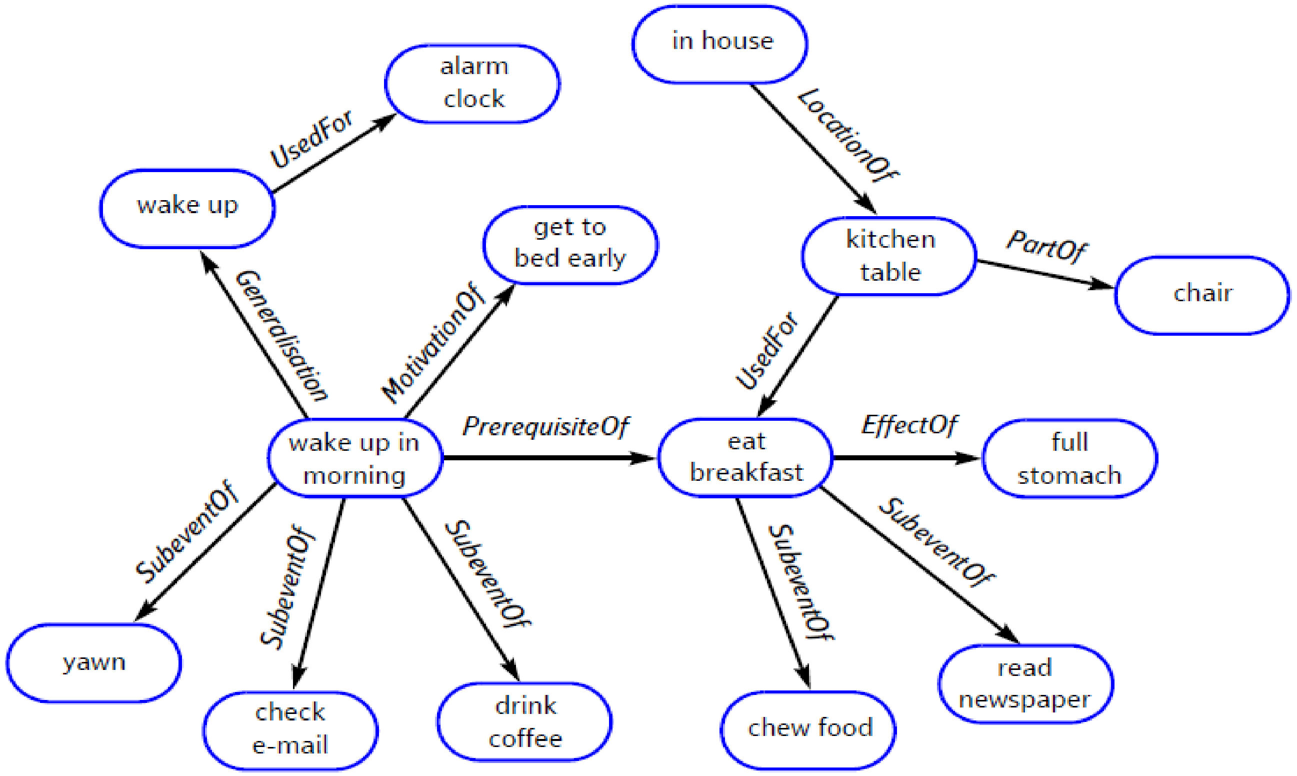
# How to Define Commonsense Knowledge? (Liu & Singh, 2004)

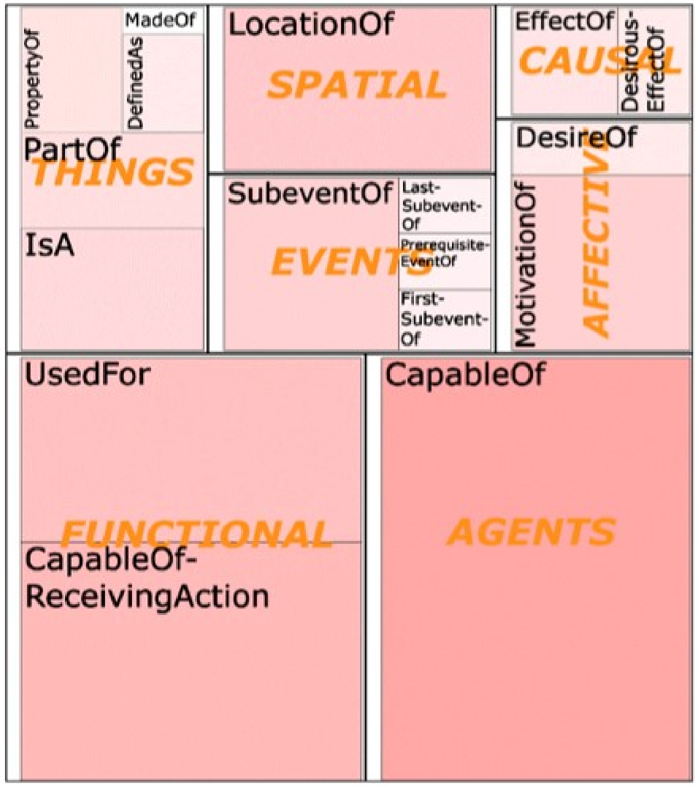
- “While to the average person the term ‘commonsense’ is regarded as synonymous with ‘good judgement’, ”
- “in the AI community it is used in a technical sense to refer to the millions of basic facts and understandings possessed by most people.”
- “Such knowledge is typically omitted from social communications”, 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

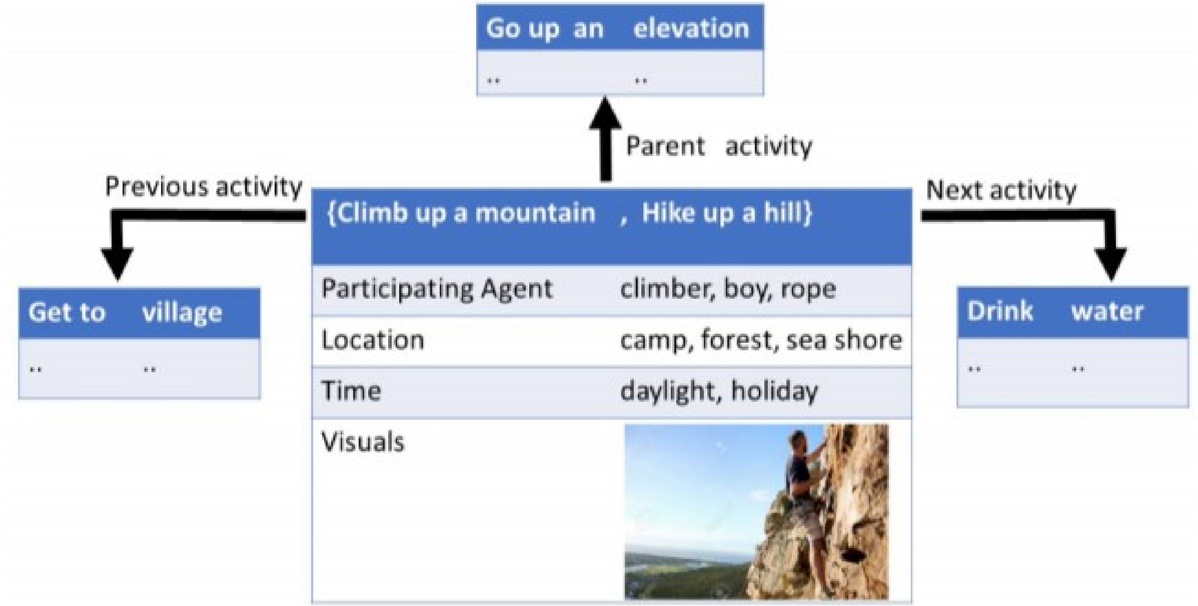
# ATOMIC

- Crowdsourcing 9 Types of IF-THEN relations
- Arbitrary texts: Human annotation
- All personal entity information has been removed to reduce ambiguity


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

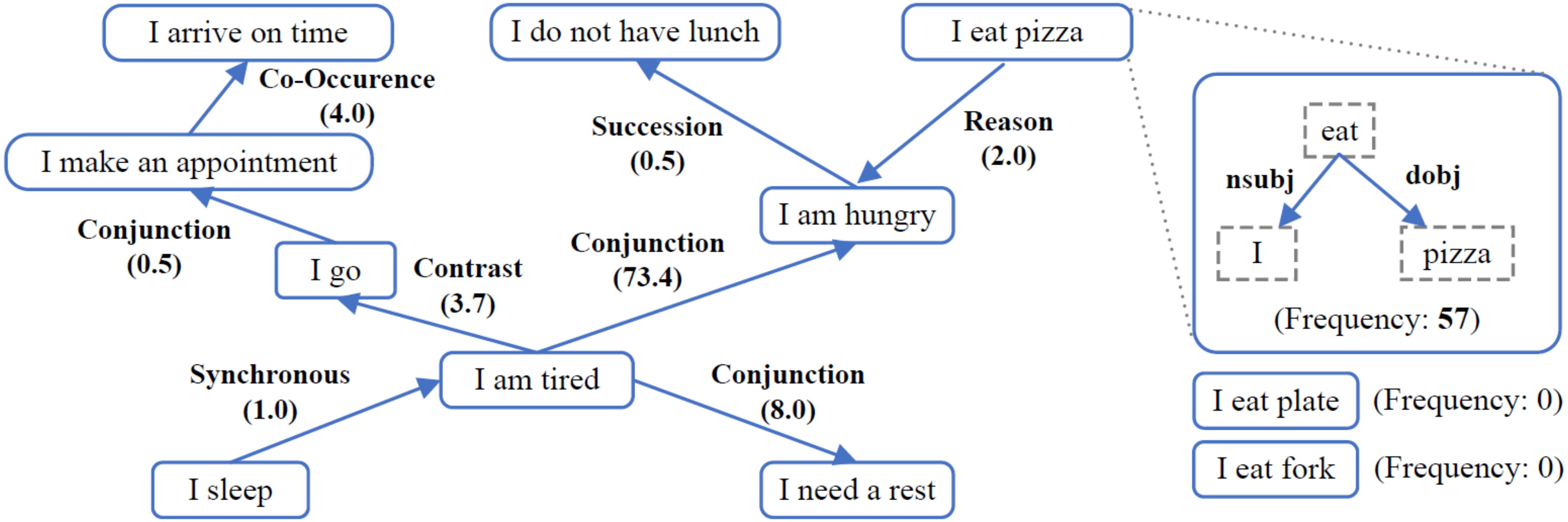
- Perform information extraction from free text
  - Mostly movie scripts and novel books
- Four relations: previous, next, parent, similarity
- Only verb+object




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

# ASER: **A**ctivities, **S**tates, **E**vents, and their **R**elations

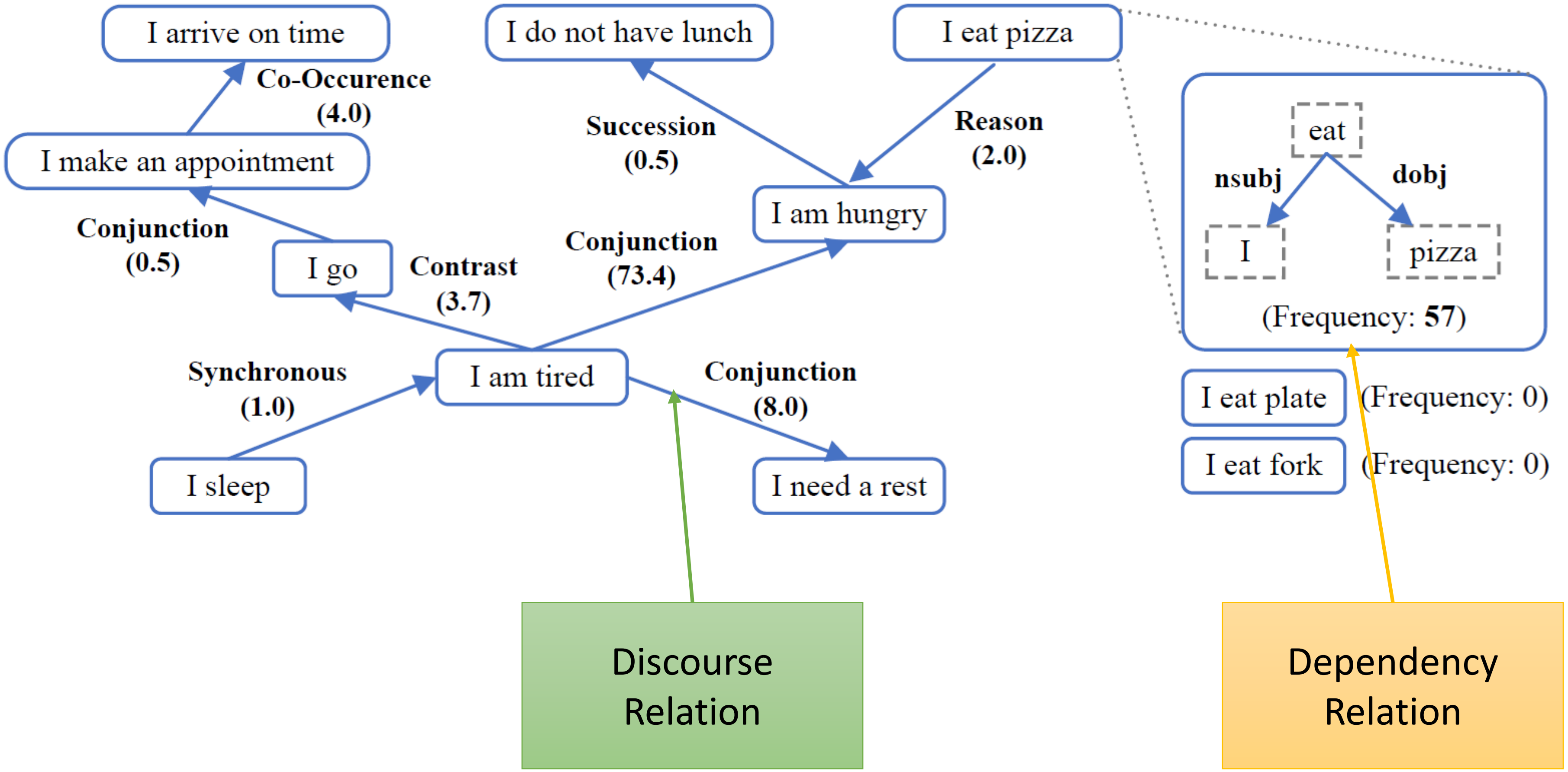
- Use verb-centric patterns from dependency parsing
  - Principle #1: to compare semantics by fixing syntax (Katz and Fodor, 1963)
- Maintain a set of key tags and a set of auxiliary tags
  - Principle #2: to obtain frequent 'partial information' (Wilks, 1975)



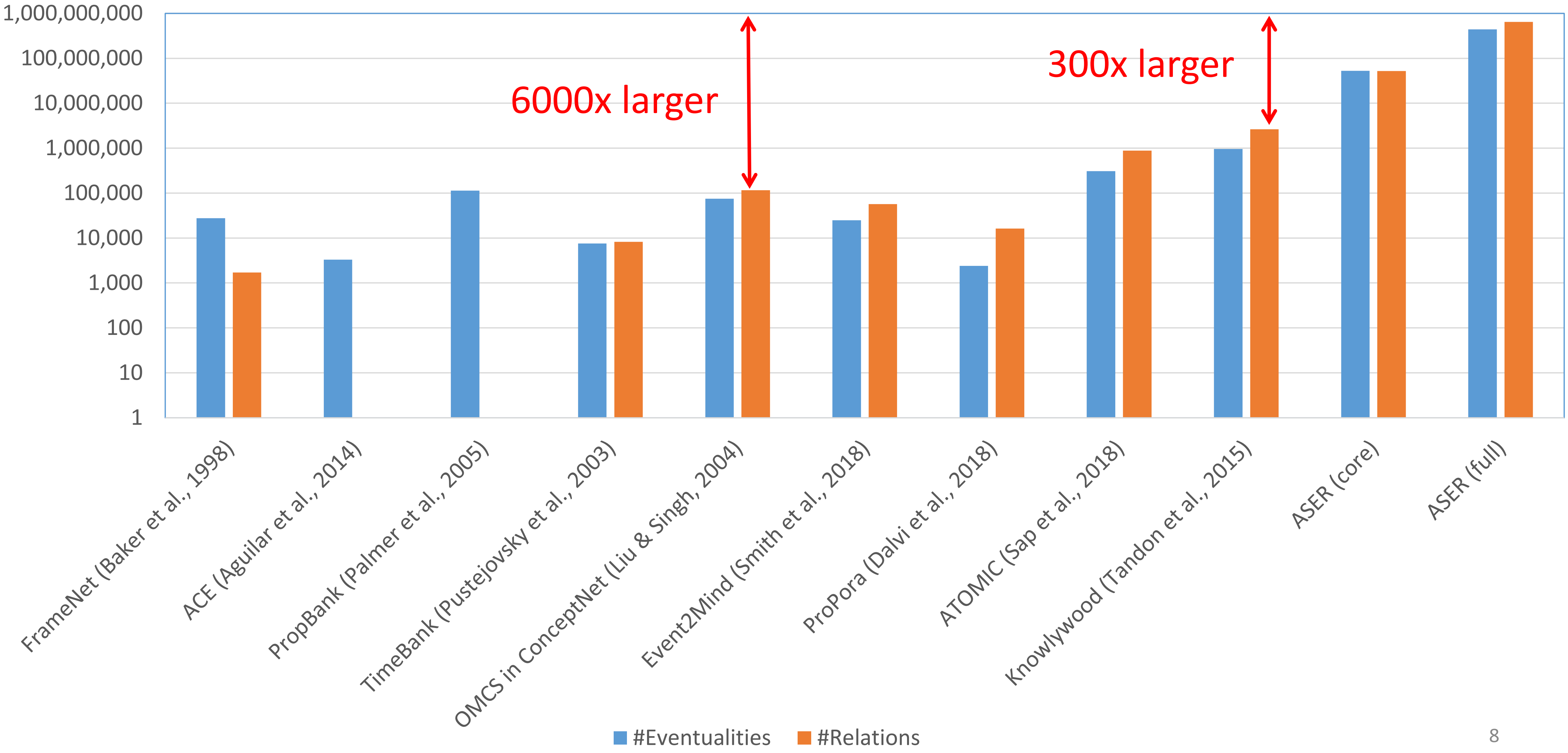

A hybrid graph of
- Each eventuality is a hyper-edge of words
- Heterogeneous edges among eventualities

Hongming Zhang*, Xin Liu*, Haojie Pan*, Yangqiu Song, and Cane Wing-Ki Leung. ASER: A Large-scale Eventuality Knowledge Graph. The Web Conference (WWW). 2020.

# ASER is Essentially a Knowledge Graph based on Linguistics



# Scales of Verb Related Knowledge Graphs



# How and Where to Automatically Collect Commonsense Knowledge?

- “Such knowledge is typically omitted from social communications” (Liu & Singh, 2004)
- Do we really omit or just not realized we have mentioned them in our daily communication?

Is it possible to transfer from linguistic knowledge to existing definition of commonsense knowledge?

# Commonsense Knowledge Base Population

- Human annotated knowledge bases are usually more accurate
- Traditional knowledge base population includes
  - Entity linking
  - Relation extraction
  - Etc.
- We define a new task of commonsense knowledge population
- How?
  - In fact, different commonsense knowledge bases have different properties

# Outline

- ConceptNet Population
  - Selectional preference

- ATOMIC Population
  - Latent variables (events and states) of commonsense

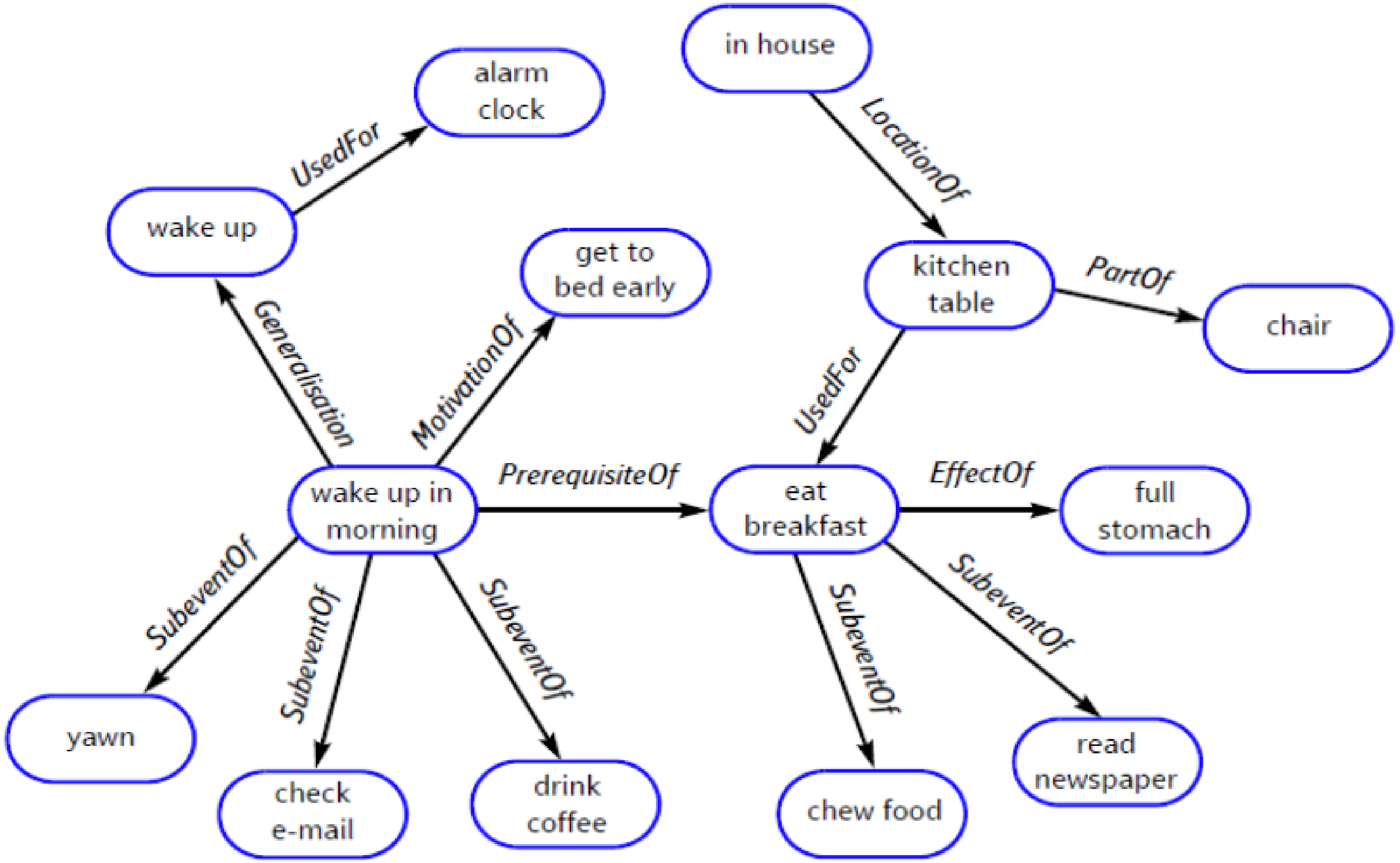


Slides credit for this part: Hongming Zhang

# ConceptNet (Speer & Havasi, 2012)

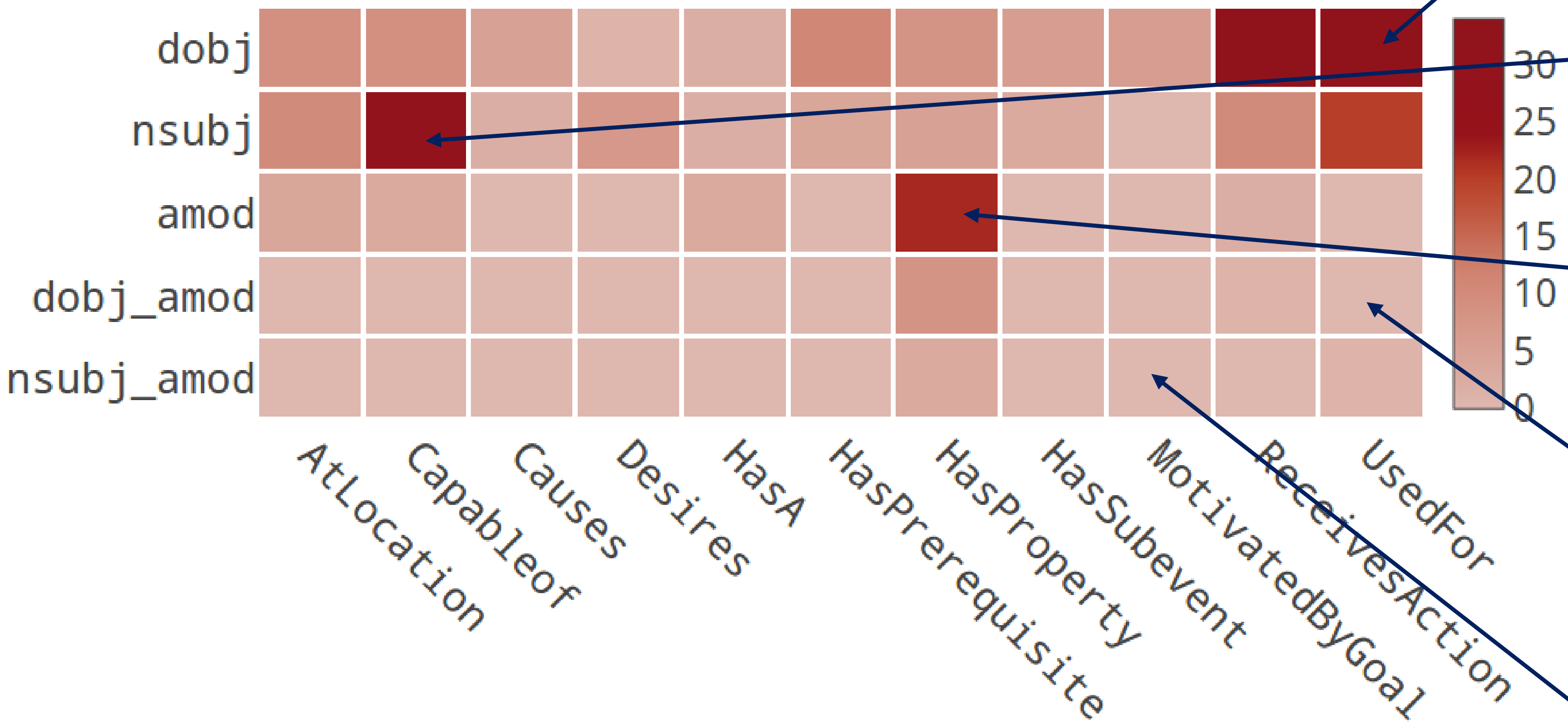
Core is OMCS (Liu & Singh 2004)

- Commonsense knowledge base
  - Commonsense knowledge about noun-phrases, or entities.




Speer and Havasi. "Representing General Relational Knowledge in ConceptNet 5." *LREC*. 2012.

# Revisit the Correlations of Selectional Preference and OMCS




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

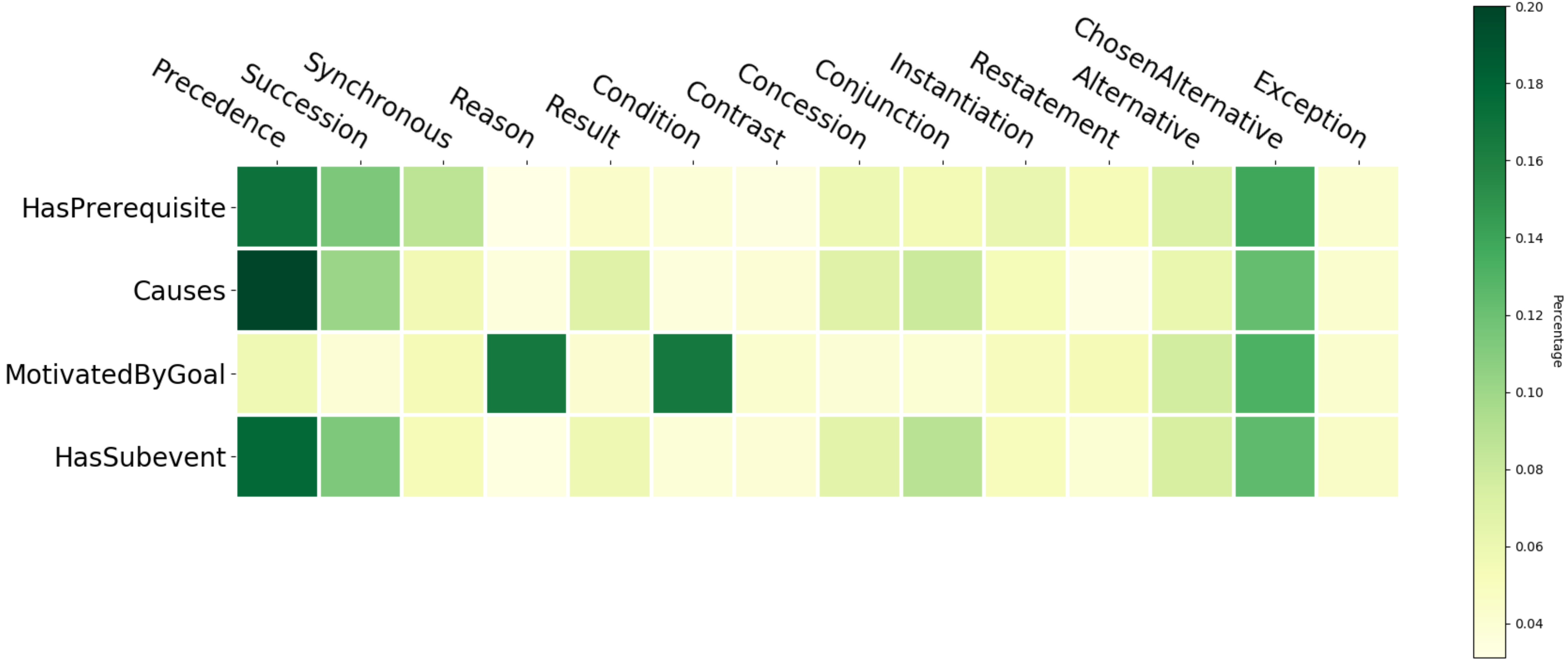
(phone, ring) (nsubj, 8.75)
(phone, CapableOf, ring)

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

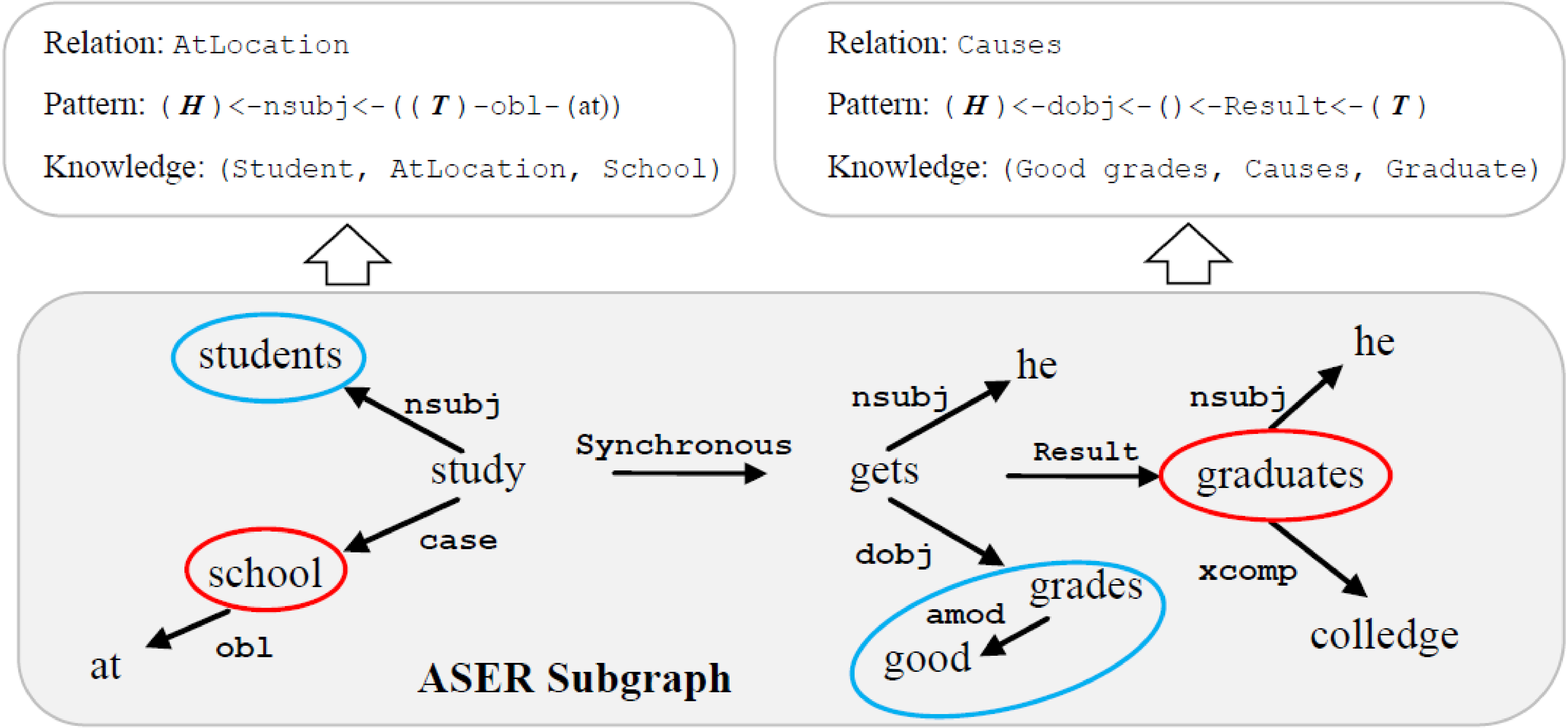
(create, new) (dobj_amod, 8.25)
(create idea, UsedFor, invent new things)

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

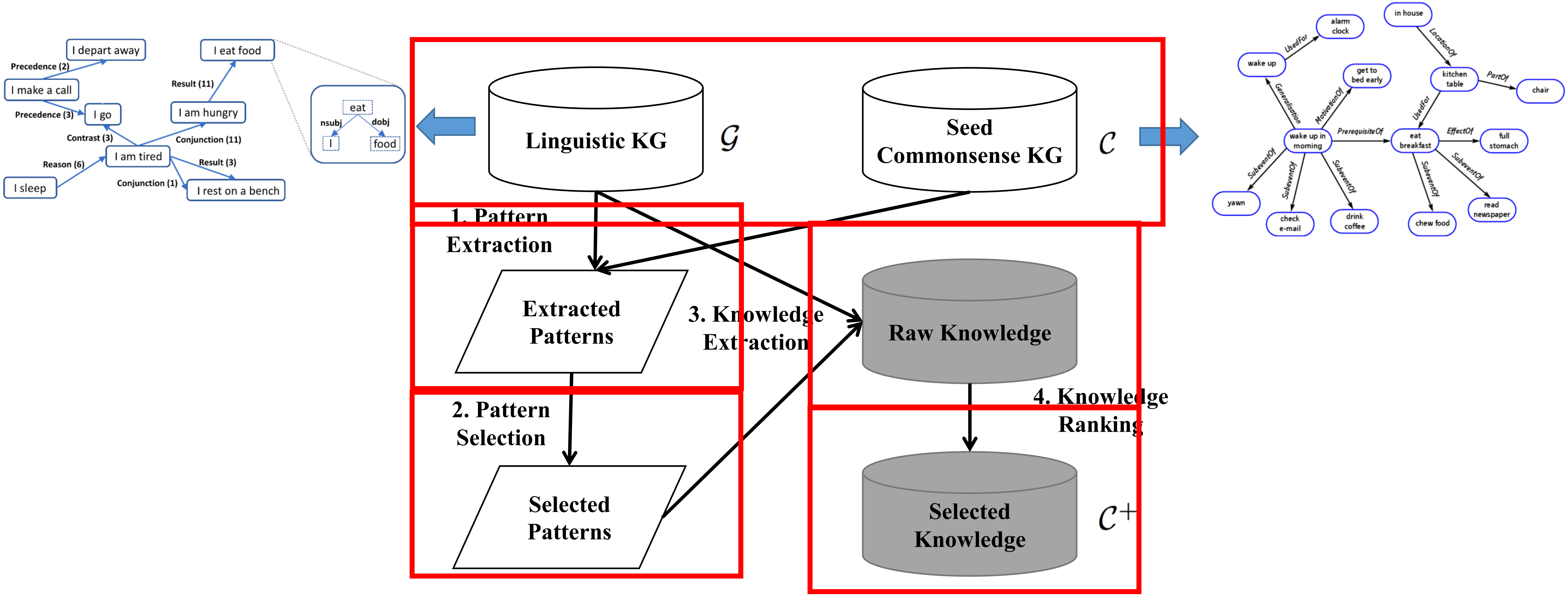
# Revisit the Correlations of ASER and OMCS



# TransOMCS


Relation: AtLocation
Pattern: ( H )<-nsubj<-(( T )-obl-(at))
Knowledge: (Student, AtLocation, School)
Relation: Causes
Pattern: ( H )<-dobj<-()<-Result<-( T )
Knowledge: (Good grades, Causes, Graduate)
students
nsubj
study
case
school
obl
at
Synchronous
ASER Subgraph
he
nsubj
gets
dobj
grades
amod
good
Result
he
nsubj
graduates
xcomp
colledge

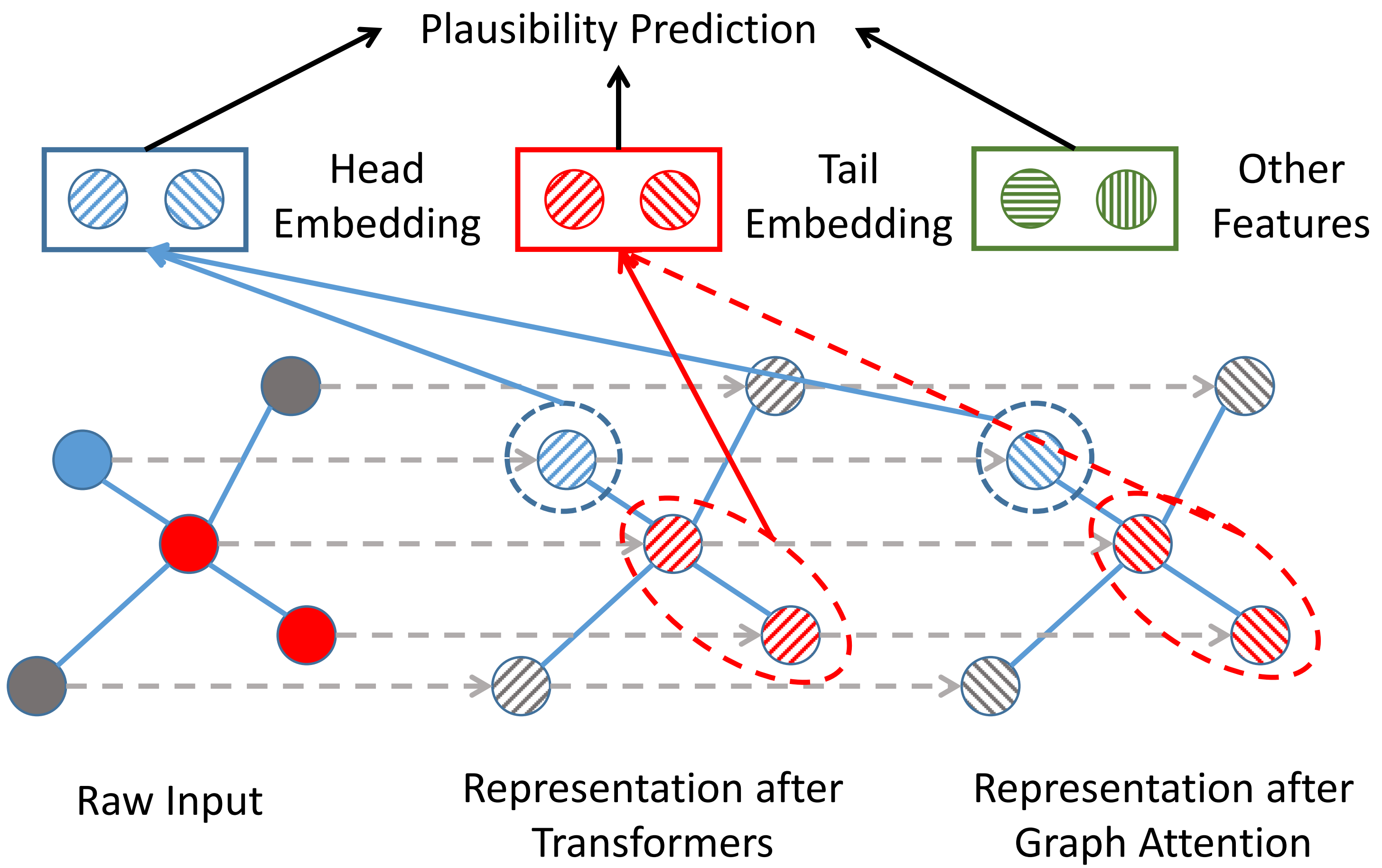
# TransOMCS: ASER to OMCS




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

# Knowledge Ranking

- Assigning confidence score to each piece of extracted commonsense
  - Leverage the semantics of the original sentences
  - Leverage the frequency information



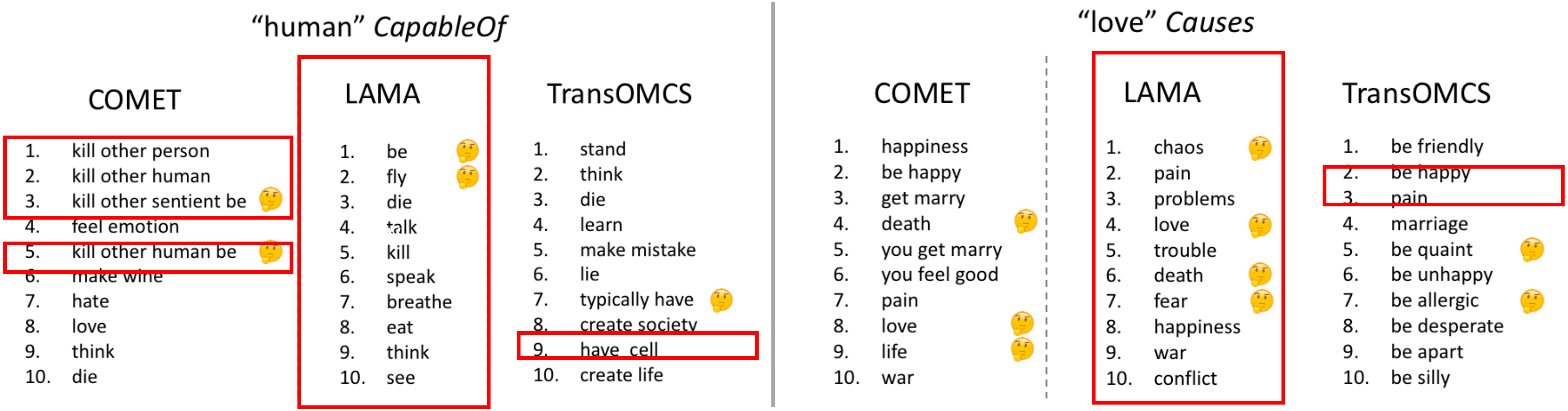
# Transferring ASER to ConceptNet

| Model | # Vocab | # Tuple | $\text{Novel}_t$ | $\text{Novel}_c$ | $\text{ACC}_n$ | $\text{ACC}_o$ |
|---|---|---|---|---|---|---|
| $\text{COMET}_{Original}$ (Greedy decoding) | 715 | 1,200 | 33.96% | 5.27% | 58% | 90% |
| $\text{COMET}_{Original}$ (Beam search - 10 beams) | 2,232 | 12,000 | 64.95% | 27.15% | 35% | 44% |
| $\text{COMET}_{Extended}$ (Greedy decoding) | 3,912 | 24,000 | **99.98%** | 55.56% | 34% | 47% |
| $\text{COMET}_{Extended}$ (Beam search - 10 beams) | 8,108 | 240,000 | **99.98%** | 78.59% | 23% | 27% |
| $\text{LAMA}_{Original}$ (Top 1) | 328 | 1,200 | - | - | - | 49% |
| $\text{LAMA}_{Original}$ (Top 10) | 1,649 | 12,000 | - | - | - | 20% |
| $\text{LAMA}_{Extended}$ (Top 1) | 1,443 | 24,000 | - | - | - | 29% |
| $\text{LAMA}_{Extended}$ (Top 10) | 5,465 | 240,000 | - | - | - | 10% |
| $\text{TransOMCS}_{Original}$ (no ranking) | 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%** |

Transferability from linguistic knowledge to commonsense knowledge

SP over eventualities can effectively represent interesting commonsense knowledge

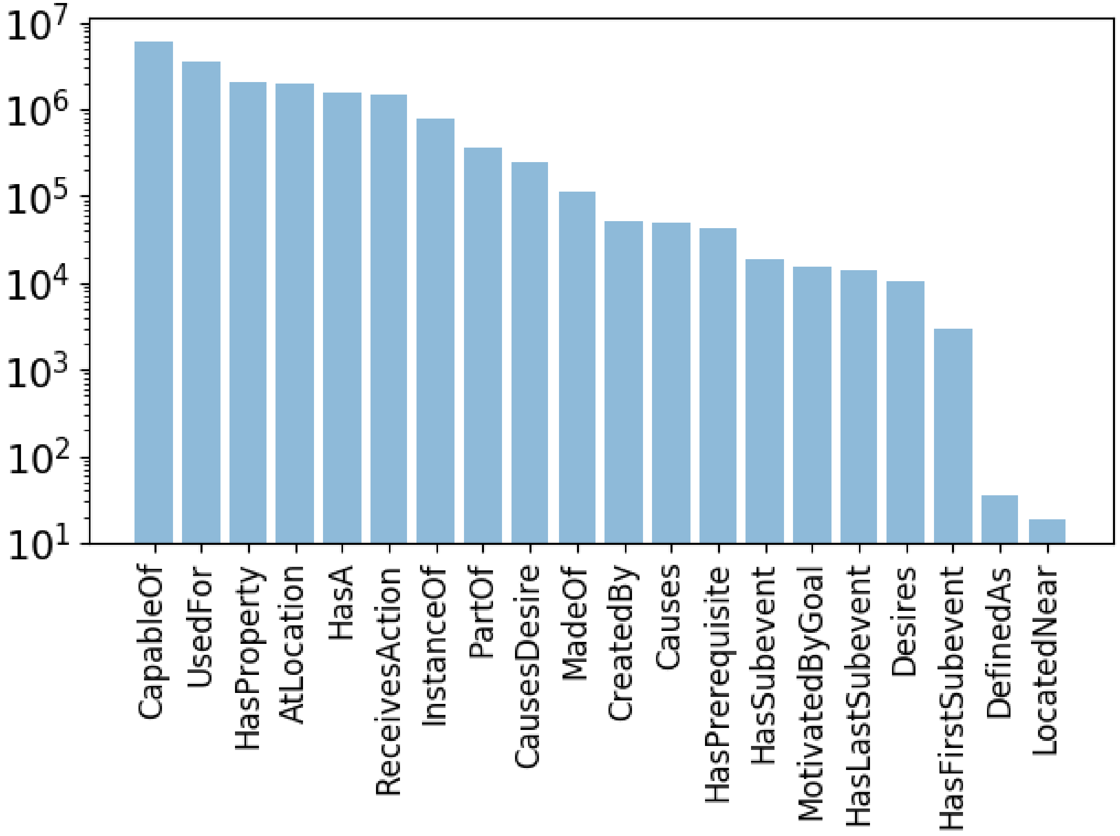
# Case Study



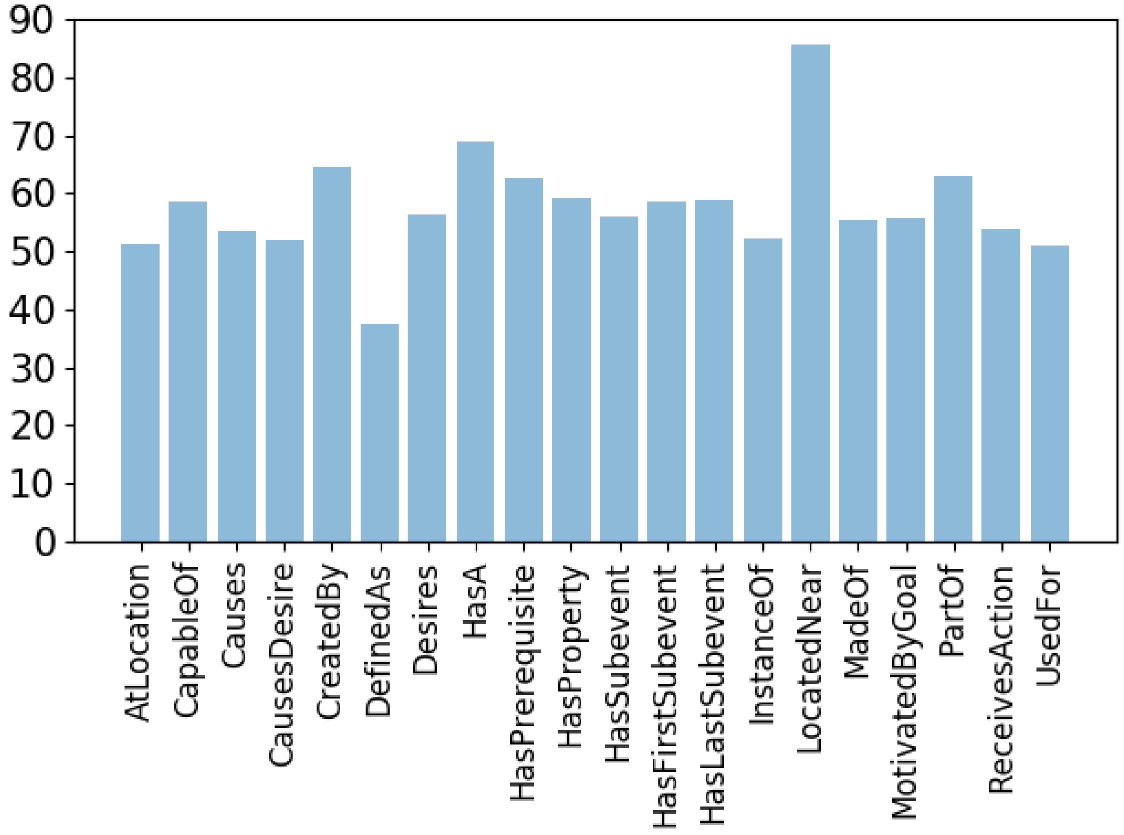

(a) Internal Setting (b) External Setting


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 2019.

# Distribution of Relations and Accuracy




Distribution of Relations




Accuracy

# Outline

- ConceptNet Population
    - Selectional preference

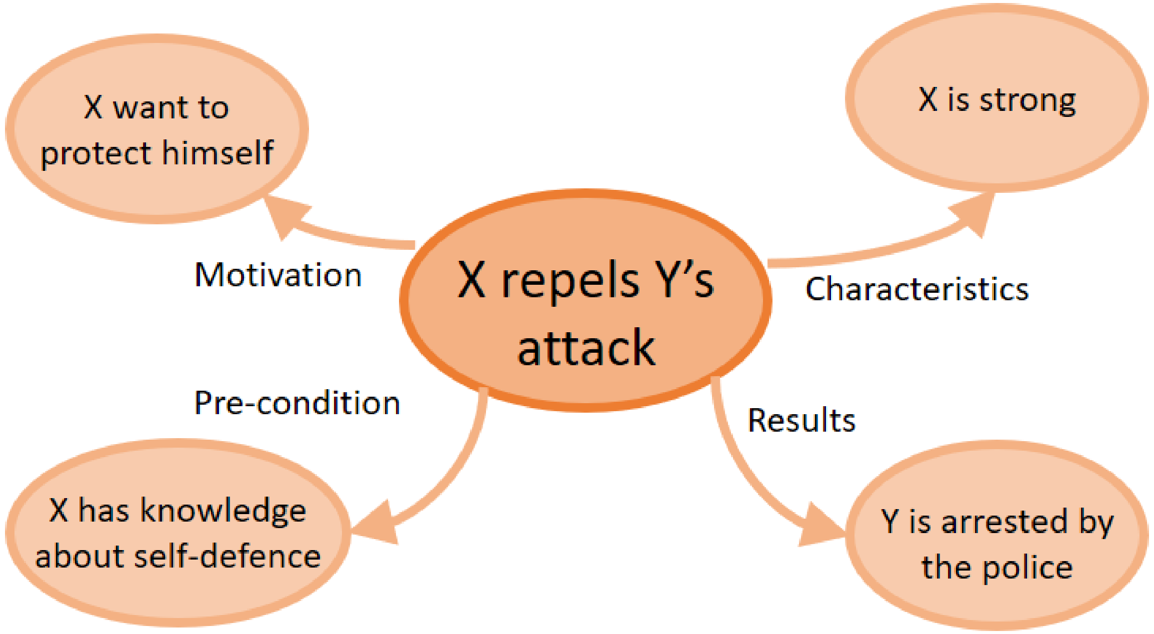
- ATOMIC Population
    - Latent variables (events and states) of commonsense



Slides credit for this part: Tianqing Fang

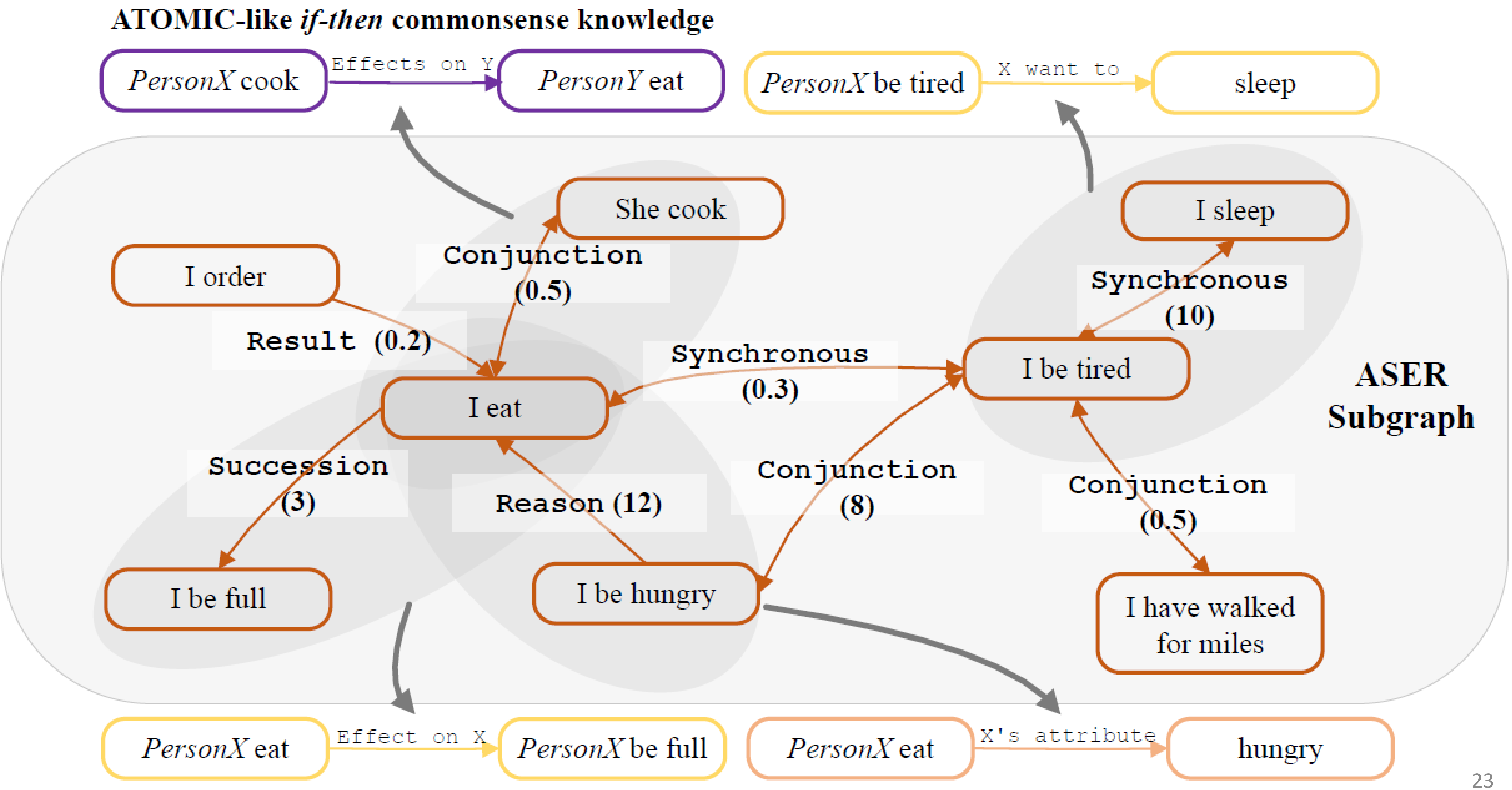
# ATOMIC (Sap, Maarten, et al. 2019)

- Everyday if-then (social) commonsense knowledge
- These are day-to-day knowledge that help us understand each other.
  - If a person *X did* something, human beings are able to inference:
    - Motivation: Why person X did this.
    - Pre-conditions: What enables X to do this.
    - Characteristics: What are attributes of X.
    - Result: What will affect X/others




Sap, Maarten, et al. "ATOMIC: An atlas of machine commonsense for if-then reasoning.", *AAAI* 2019.

# Transform ASER to ATOMIC



# Coverage and Implicit Edges

- Most event related commonsense relations are implicit on ASER
  - ConceptNet (Event-related relations), ATOMIC, ATOMIC 2020, and GLUCOSE

| | $\text{ASER}_{norm}$ Coverage | | | | Avg. Degree in $\text{ASER}_{norm}$ | | | | Avg. Degree in $\mathcal{C}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | In-Degree | | Out-Degree | | In-Degree | | Out-Degree | |
| | head(%) | tail(%) | edge(%) | #hops | head | tail | head | tail | head | tail | head | tail |
| ATOMIC | 79.76 | 77.11 | 59.32 | 2.57 | 90.9 | 61.3 | 91.2 | 61.6 | 4.2 | 3.4 | 34.6 | 1.5 |
| $\text{ATOMIC}_{20}^{20}$ | 80.39 | 47.33 | 36.73 | 2.65 | 96.9 | 66.9 | 97.3 | 67.3 | 4.3 | 2.9 | 34.6 | 1.5 |
| ConceptNet | 77.72 | 54.79 | 43.51 | 2.37 | 210.7 | 88.9 | 211.6 | 88.9 | 15.1 | 8.0 | 26.2 | 4.1 |
| GLUCOSE | 91.48 | 91.85 | 81.01 | 2.37 | 224.9 | 246.4 | 226.6 | 248.0 | 7.2 | 7.7 | 6.7 | 5.5 |

Table 3: The overall matching statistics for the four CSKBs. The *edge* column indicates the proportion of edges where their heads and tails can be connected by paths in ASER. Average (in and out)-degree on $\text{ASER}_{norm}$ and $\mathcal{C}$ for nodes from the CSKBs is also presented. The statistics in $\mathcal{C}$ is different from (Malaviya et al., 2020) as we check the degree on the aligned CSKB $\mathcal{C}$ instead of each individual CSKB.

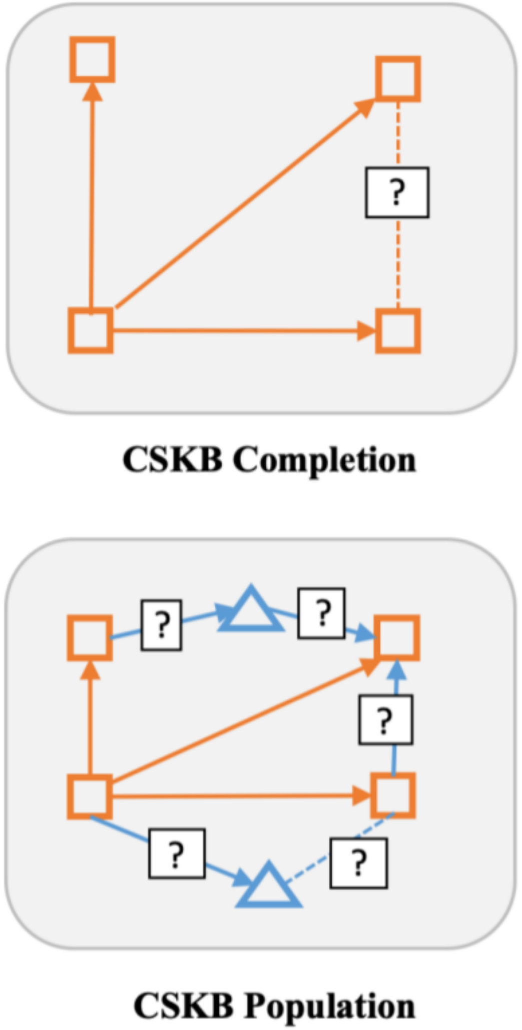
Maarten Sap, et al. ATOMIC: An atlas of machine commonsense for if-then reasoning. AAAI 2019.
Jena D Hwang, et al. (Comet-) Atomic 2020: On Symbolic and Neural Commonsense Knowledge Graphs. AAAI 2021.
Nasrin Mostafazadeh, et al. Glucose: Generalized and contextualized story explanations. NAACL 2020.

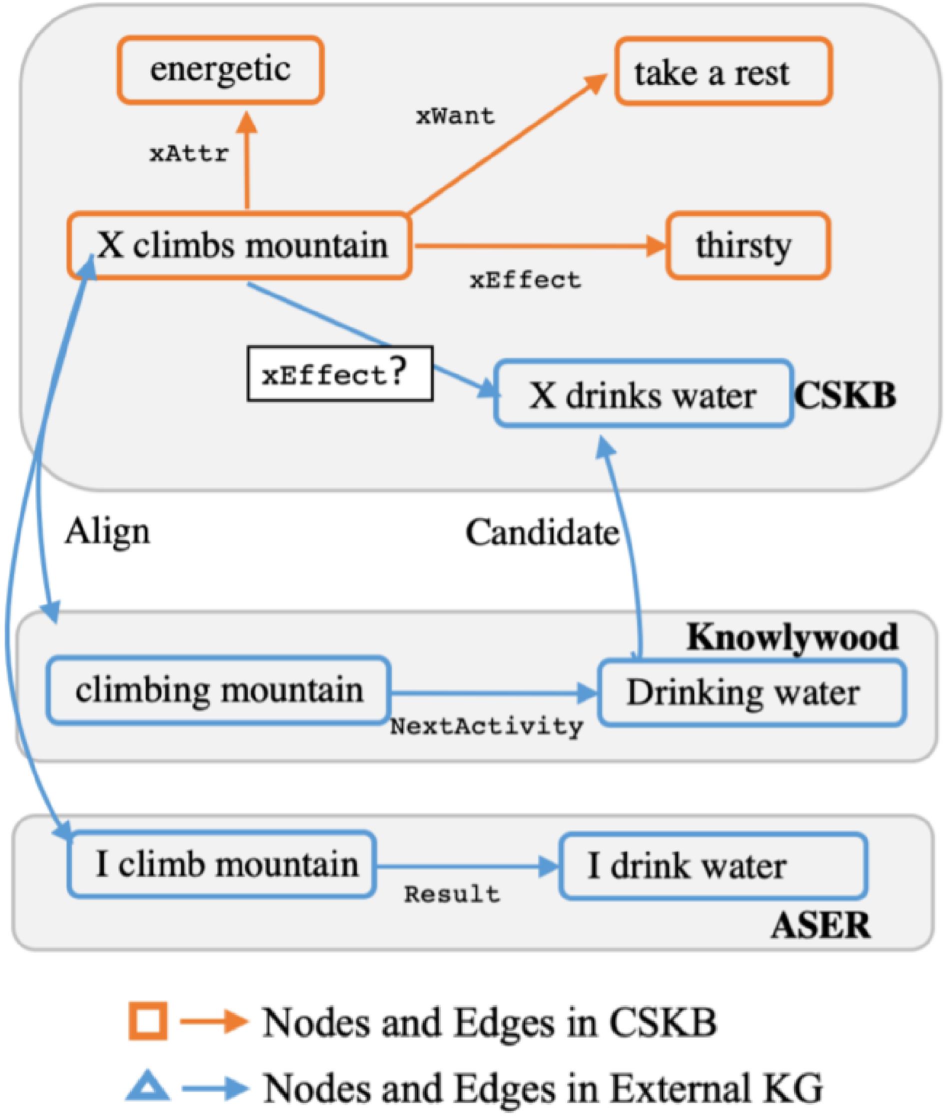
# Commonsense Acquisition and Graph based Learning

- Current ways of commonsense acquisition from graphs formalize the problem as a link prediction task:
  - Commonsense Knowledge Base (CKB) Completion (Li, 2016)
  - Add synthetic edges to CSKB for KBC (Malaviya, 2020)
  - Inductive Learning (Wang, 2020)
- In event-centric commonsense, nodes can be arbitrary text instead of structured nodes!
  - The graph of ATOMIC is almost a bipartite graph.

Li, Xiang, et al. "Commonsense knowledge base completion." ACL. 2016.
Malaviya, Chaitanya, et al. “Exploiting structural and semantic context for commonsense knowledge base completion.” IJCAI 2020.
Wang, Bin, et al. "Inductive Learning on Commonsense Knowledge Graph Completion." arXiv preprint (2020).

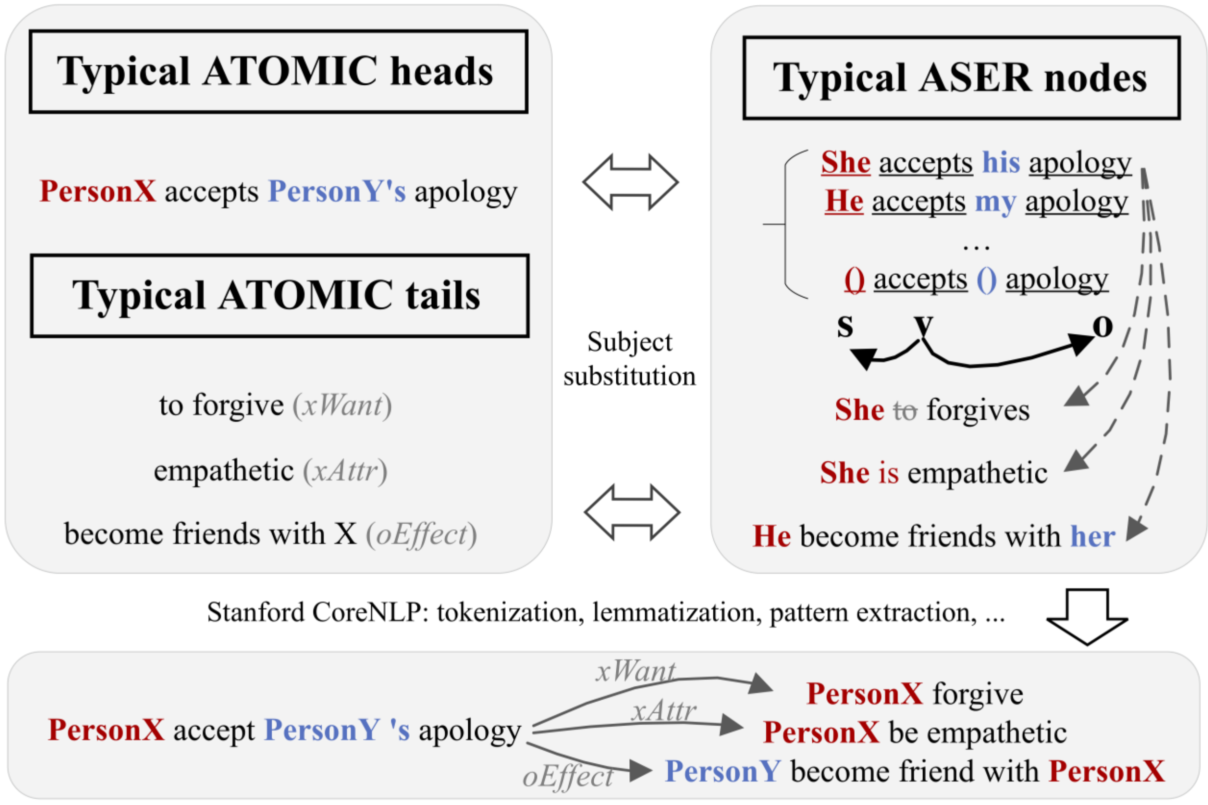
# CKGC (Completion) vs. CKGP (Population)

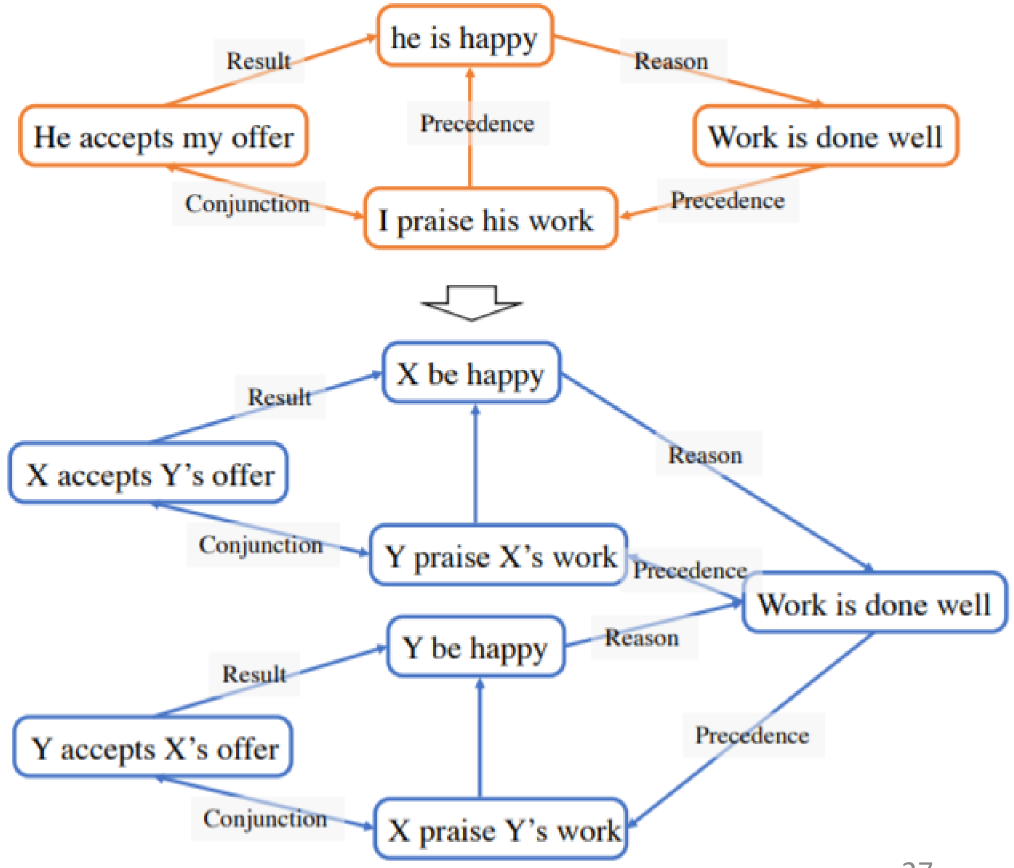


# Node Alignment with ASER

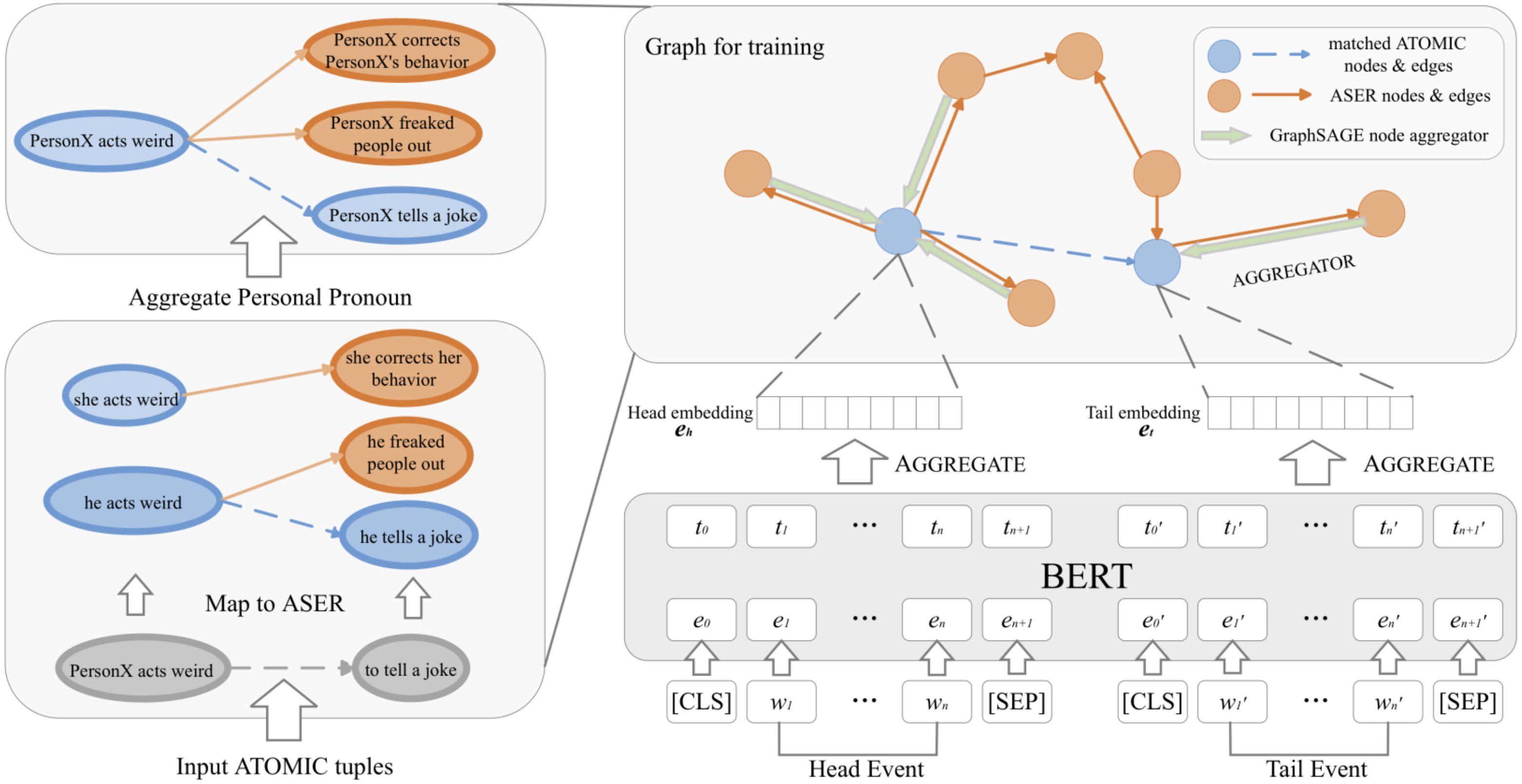
- ASER and other CSKB take different forms of representing personal entities
- Develop simple rules for aligning the two resources.



# DISCOS (DIScourse to COmmonSense): BertSAGE [WWW 2021]

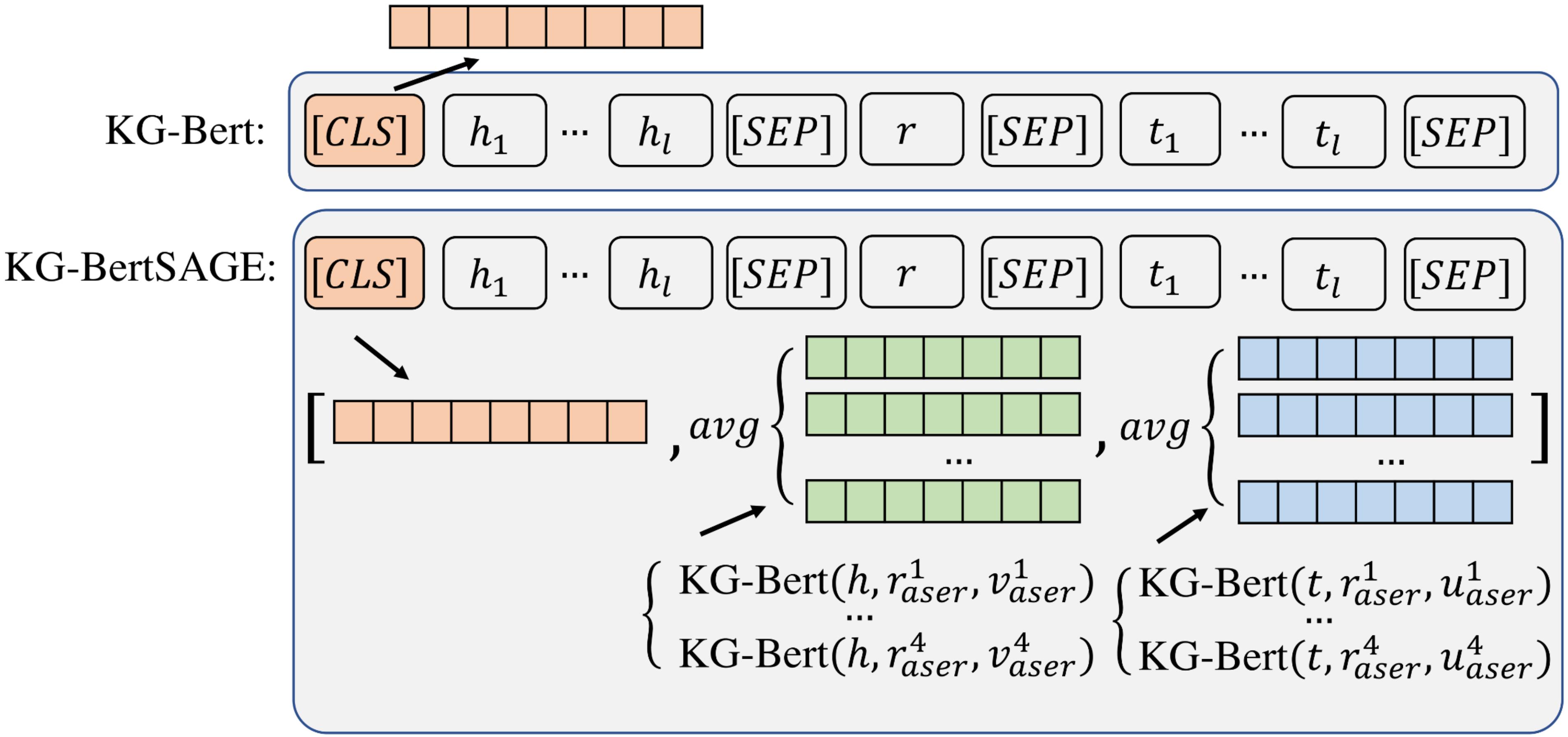
- Use BERT to encode the eventuality sentences
- Use GraphSAGE (Hamilton 2017) to aggregate the neighboring information in ASER





Hamilton, William L., Rex Ying, and Jure Leskovec. "Inductive representation learning on large graphs." NeurIPS. 2017.
Tianqing Fang, Hongming Zhang, Weiqi Wang, Yangqiu Song, and Bin He. DISCOS: Bridging the Gap between Discourse Knowledge and Commonsense Knowledge. WWW, 2021.

# Another Model: KG-BertSAGE [EMNLP 2021]




Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. KG-Bert: Bert for knowledge graph completion. arXiv preprint arXiv:1909.03193.
Tianqing Fang, Weiqi Wang, Sehyun Choi, Shibo Hao, Hongming Zhang, Yangqiu Song, and Bin He. Benchmarking Commonsense Knowledge Base Population with an Effective Evaluation Dataset. EMNLP. 2021.

# Training and Testing Data

- Training: four commonsense knowledge bases
  - ConceptNet (event-related relations)
  - ATOMIC
  - ATOMIC 2020
  - GLUCOSE
- Graph Data: normalized nodes/edges in ASER
- Testing: ~30K annotated data

| | Dev | Test | Train |
|---|---|---|---|
| # Triples | 6,217 | 25,514 | 1,100,362 |
| % Plausible | 51.05% | 51.74% | - |
| % Novel Nodes | 67.40% | 70.01% | - |

| Relation | ATOMIC$\binom{20}{20}$ | ConceptNet | GLUCOSE |
|---|---|---|---|
| oEffect | 21,497 | 0 | 7,595 |
| xEffect | 61,021 | 0 | 30,596 |
| gEffect | 0 | 0 | 8,577 |
| oWant | 35,477 | 0 | 1,766 |
| xWant | 83,776 | 0 | 11,439 |
| gWant | 0 | 0 | 5,138 |
| oReact | 21,110 | 0 | 3,077 |
| xReact | 50,535 | 0 | 13,203 |
| gReact | 0 | 0 | 2,683 |
| xAttr | 89,337 | 0 | 7,664 |
| xNeed | 61,487 | 0 | 0 |
| xIntent | 29,034 | 0 | 8,292 |
| isBefore | 18,798 | 0 | 0 |
| isAfter | 18,600 | 0 | 0 |
| HinderedBy | 87,580 | 0 | 0 |
| xReason | 189 | 0 | 0 |
| Causes | 0 | 42 | 26,746 |
| HasSubEvent | 0 | 9,934 | 0 |
| Total | 578,252 | 10,165 | 126,776 |

| Relation | number of edges |
|---|---|
| Precedence | 4,957,481 |
| Succession | 1,783,154 |
| Synchronous | 8,317,572 |
| Reason | 5,888,968 |
| Result | 5,562,565 |
| Condition | 8,109,020 |
| Contrast | 23,208,195 |
| Concession | 1,189,167 |
| Alternative | 1,508,729 |
| Conjunction | 37,802,734 |
| Restatement | 159,667 |
| Instantiation | 33,840 |
| ChosenAlternative | 91,286 |
| Exception | 51,502 |
| Co_Occurrence | 124,330,714 |
| Total | 222,994,594 |

# Main Population Results

- We use AUC as the evaluation metric. The break-down scores for all models are presented below.

| Relation | xWnt | oWnt | gWnt | xEfct | oEfct | gEfct | xRct | oRct | gRct | xAttr | xInt | xNeed | Cause | xRsn | isBfr | isAft | Hndr. | HasSubE. | all |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BERT | 57.7 | 64.9 | 66.3 | 59.1 | 66.2 | 60.0 | 50.6 | **68.7** | 72.3 | 56.2 | 63.9 | 56.4 | 48.3 | 34.5 | 59.2 | 58.0 | 66.1 | 73.0 | 59.4 |
| BERTSAGE | 54.7 | 58.9 | 58.0 | 58.0 | 70.0 | 54.7 | 52.8 | 62.4 | **76.6** | 55.0 | 61.0 | 57.1 | 46.2 | 45.5 | 66.7 | 64.9 | 69.6 | **80.4** | 60.0 |
| KG-BERT | 63.2 | **69.8** | **69.0** | 68.0 | 70.6 | 61.0 | 57.0 | 64.0 | 73.8 | **59.5** | **64.9** | 64.6 | 47.4 | **90.9** | 78.0 | **77.5** | 75.9 | 68.5 | 66.1 |
| KG-BERTSAGE | **66.0** | 68.9 | 68.6 | **68.2** | **70.8** | **62.3** | **60.5** | 64.6 | 74.1 | 59.1 | 63.0 | **65.4** | **50.0** | 76.4 | **78.2** | 77.4 | **77.5** | 67.0 | **67.2** |
| Human | 86.2 | 86.8 | 83.3 | 85.2 | 83.9 | 79.8 | 81.1 | 82.6 | 76.5 | 82.6 | 85.6 | 87.4 | 80.1 | 73.7 | 89.8 | 89.9 | 85.3 | 85.7 | 84.4 |

# Error Analysis and Discussion

- Break-down AUC scores for different types of testing edges.

| Model | *Original Test Set* | *CSKB head + ASER tail* | *ASER edges* |
|---|---|---|---|
| BERT | 65.0 | 47.9 | 44.6 |
| BERTSAGE | 67.2 | 49.4 | 46.2 |
| KG-BERT | 77.8 | 55.2 | 50.3 |
| KG-BERTSAGE | **78.2** | **57.5** | **52.3** |

Models achieve relatively poorer performance on novel edges from ASER

Evaluation Set:

A ***diverse*** evaluation set, consisting of edges from different domains

(1) Edges from the original test sets (according to the original split in CSKBs)

(2) Edges from ASER, where heads are in CSKBs, while tails are out of CSKBs

(3) Edges from ASER where neither head nor tail is from CSKBs

# Case Study

| Head | Relation | Tail | Label | Source |
|---|---|---|---|---|
| *PersonX* give *PersonY* ride | `xNeed` | *PersonX* need to wear proper clothes | Plau. | Triples in CSKBs (*Original Test Set*) |
| *PersonX* be wait for taxi | `isAfter` | *PersonX* hail a taxi | Plau. | |
| *PersonX* be diagnose with something | `Causes` | *PersonX* be sad | Plau. | |
| *PersonX* feel something | `xEffect` | *PersonX* figure | Implau. | Randomly sampled examples |
| *PersonX* be patient with ignorance | `HinderedBy` | *PersonY* have the right vocabulary | Implau. | |
| *PersonY* grasp *PersonY* meaning | `HasSubEvent` | *PersonY* open it mechanically | Implau. | |
| *PersonX* spill coffee | `oEffect` | *PersonY* have to server | Plau. | *CSKB head + ASER tail* |
| *PersonX* care for *PersonY* | `xNeed` | *PersonX* want to stay together | Plau. | |
| *PersonX* be save money | `HasSubEvent` | PeopleX can not afford something | Plau. | |
| *PersonX* decide to order a pizza | `xReact` | *PersonX* have just move | Implau. | |
| it be almost christmas | `gReact` | *PersonX* be panic | Implau. | |
| arm be break | `isBefore` | *PersonY* ask | Implau. | |
| *PersonX* go early in morning | `xEffect` | *PersonX* do not have to deal with crowd | Plau. | *ASER edges* |
| *PersonX* have take time to think it over *PersonX* | `xReact` | *PersonX* be glad | Plau. | |
| *PersonX* have a good work-life balance | `xIntent` | *PersonX* be happy | Plau. | |
| *PersonX* weight it by value | `oWant` | *PersonY* bet | Implau. | |
| *PersonX* be hang out on reddit | `oReact` | *PersonY* can not imagine | Implau. | |
| *PersonX* can get *PersonY* out shell | `xIntent` | *PersonX* just start poach *PersonY* | Implau. | |

# Conclusions

- We can acquire novel and high-quality commonsense knowledge from linguistic knowledge graph, i.e., ASER
- TransOMCS and DISCOS can also help
  - Mining patterns for commonsense knowledge
  - Completing (_, r, t)
  - Completing (_, r, _)
- Current model make uses of only one-hop neighbors of ASER
  - More ideas need to be tried for reasoning on graphs

Code and data

https://github.com/HKUST-KnowComp/ASER
https://github.com/HKUST-KnowComp/TransOMCS
https://github.com/HKUST-KnowComp/DISCOS-commonsense
https://github.com/HKUST-KnowComp/CSKB-Population

Project Homepage

https://hkust-knowcomp.github.io/ASER/

Thank you ☺