Acquiring and Modeling
Abstract Commonsense Knowledge via Conceptualization

Yangqiu Song

Department of CSE, HKUST



Work done with Mutian He, Tianqing Fang, Weiqi Wang.

Outline

- Motivation
- Abstract ATOMIC

"Commonsense Knowledge"

- When we communicate,
  - we omit a lot of "common sense" knowledge, which we assume the hearer/reader possesses
  - we keep a lot of ambiguities, which we assume the hearer/reader knows how to resolve

- Attributes of objects
  - A lemon is sour.
- Condition/consequence of actions
  - To open a door, you must usually first turn the doorknob.
- Cause/effect between events and states
  - If you forget someone's birthday, they may be unhappy with you.

- Social:
  - If you forget your friend's birthday, he/she may be mad at you.
- Physical, temporal, spatial:
  - Apples fall instead of floating in the air.
- World entities:
  - Lions are bigger than cats.

Many Possible Applications

- How to plan a wedding, what to do and what to buy?
  - Understand the timeline
  - Understand the events
    - Precedents
    - Consequences
  - Understand the people
    - Social conventions

18 MONTH WEDDING PLANNER
Your monthly wedding to-do list from Rock My Wedding

YOUR WEDDING DATE:

18 MONTHS Here we go...
- [ ] Get the inspiration book 'Your Day Your Way' by Rock My Wedding.
- [ ] Decide how to collate your inspiration. Maybe a folder, Pinterest, Instagram.
- [ ] Work out a budget.
- [ ] Work out guest list and choose bridal party.

16 MONTHS to go...
- [ ] Look at venues and check availability.
- [ ] Book officiant/church etc and sort wedding licence.
- [ ] Start researching suppliers via Rock My Wedding recommended suppliers.
- [ ] Sort wedding insurance.

14 MONTHS to go...
- [ ] Arrange appointments at wedding dress boutiques.
- [ ] Book a wedding planner if you want one and budget allows.
- [ ] Book your photographer and videographer.

12 MONTHS to go...
- [ ] Consider underwear and try on dresses.
- [ ] Look at honeymoon options.

9 MONTHS to go...
- [ ] Book florist.
- [ ] Send Save The Dates.
- [ ] Look into Grooms attire.

6 MONTHS to go...
- [ ] Shop for Bridesmaid dresses.
- [ ] Confirm catering.
- [ ] Taste cakes and book.
- [ ] Book entertainment musical and other.
- [ ] Book stationery with a professional or plan your DIY stationery.
- [ ] Consider transport.
- [ ] Research and book any items you may need to hire.
- [ ] Decide on a gift list company and register.
- [ ] Book hair and make-up trials.

4 MONTHS to go...
- [ ] Buy your underwear if you didn't get it before trying dresses.
- [ ] Shop for shoes.
- [ ] Buy Grooms suit and leave time for alterations.
- [ ] Send invitations.
- [ ] Choose wedding rings.
- [ ] Have hair and make-up trials and book.

2 MONTHS to go...
- [ ] Organise dress fittings.
- [ ] Choose music.
- [ ] Finalise readings.
- [ ] Finalise order of service and the day.
- [ ] Have a pre wedding shoot with your photogtapher.
- [ ] Chase RSVP's.

3 WEEKS to go...
- [ ] Arrange your seating plan.
- [ ] Start making your table plan if you're making yourself.
- [ ] Write vows.
- [ ] Book beauty and spa treatments.
- [ ] Collect wedding rings.
- [ ] Call vendors to check all your bookings are still ok and everyone knows what is what and check balance due dates.

1 WEEK to go...
- [ ] Arrange your seating plan.
- [ ] Start making your table plan if you're making yourself.
- [ ] Write vows.
- [ ] Pack for honeymoon.
- [ ] Book beauty and spa treatments.
- [ ] Collect wedding outfits.

THE DAY BEFORE
- [ ] Drop off any decor items to the venue.
- [ ] Go through roles with everyone.
- [ ] Have a good meal and make sure you have arranged breakfast for the morning of.
- [ ] Get an early night and have an amazing day!

rockmywedding.co.uk | @rockmywedding

ROCK MY WEDDING

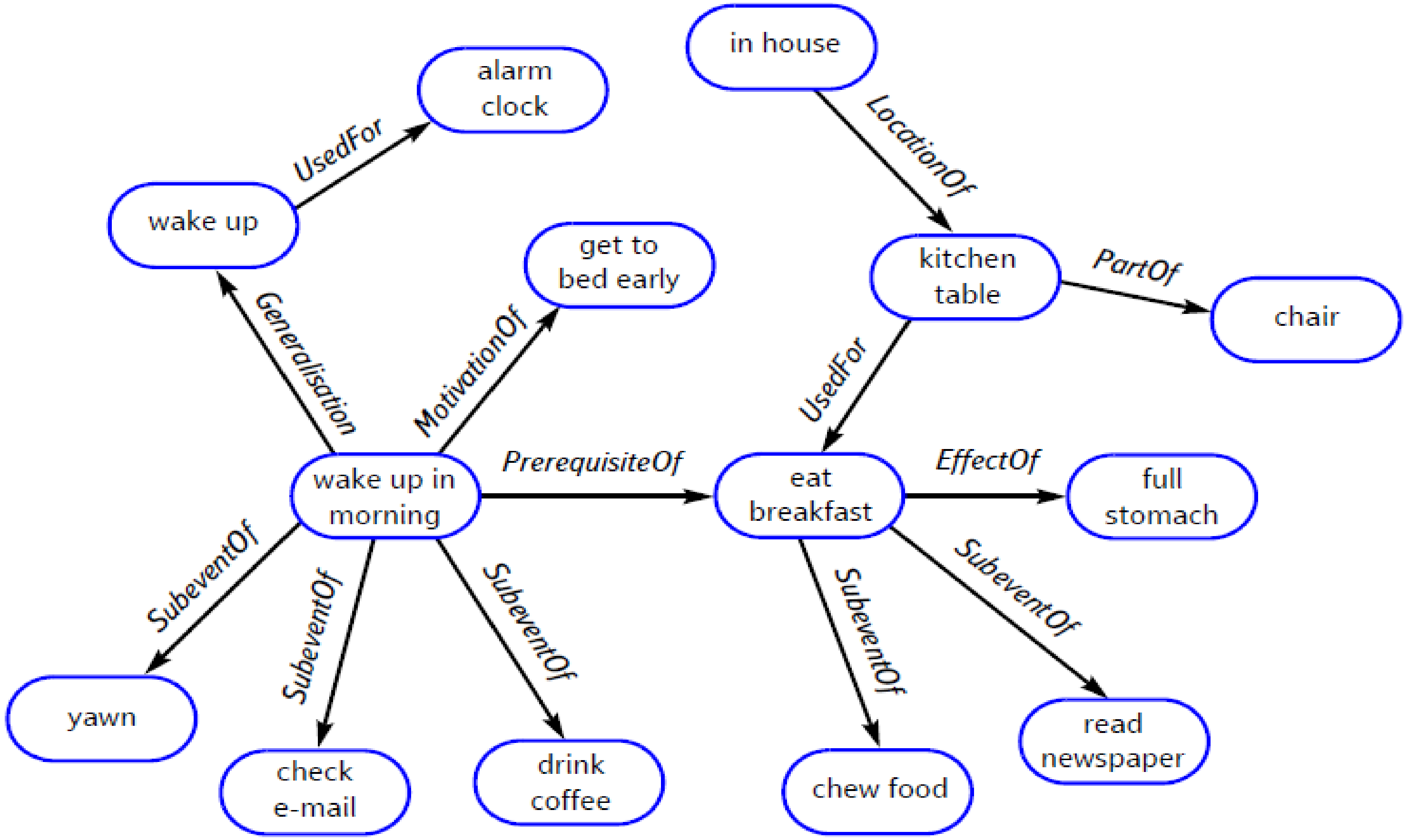
How to Define Commonsense Knowledge as Computer Scientists?

- According to 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.”

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

ConceptNet: An Approach Developed 18 Years Ago

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

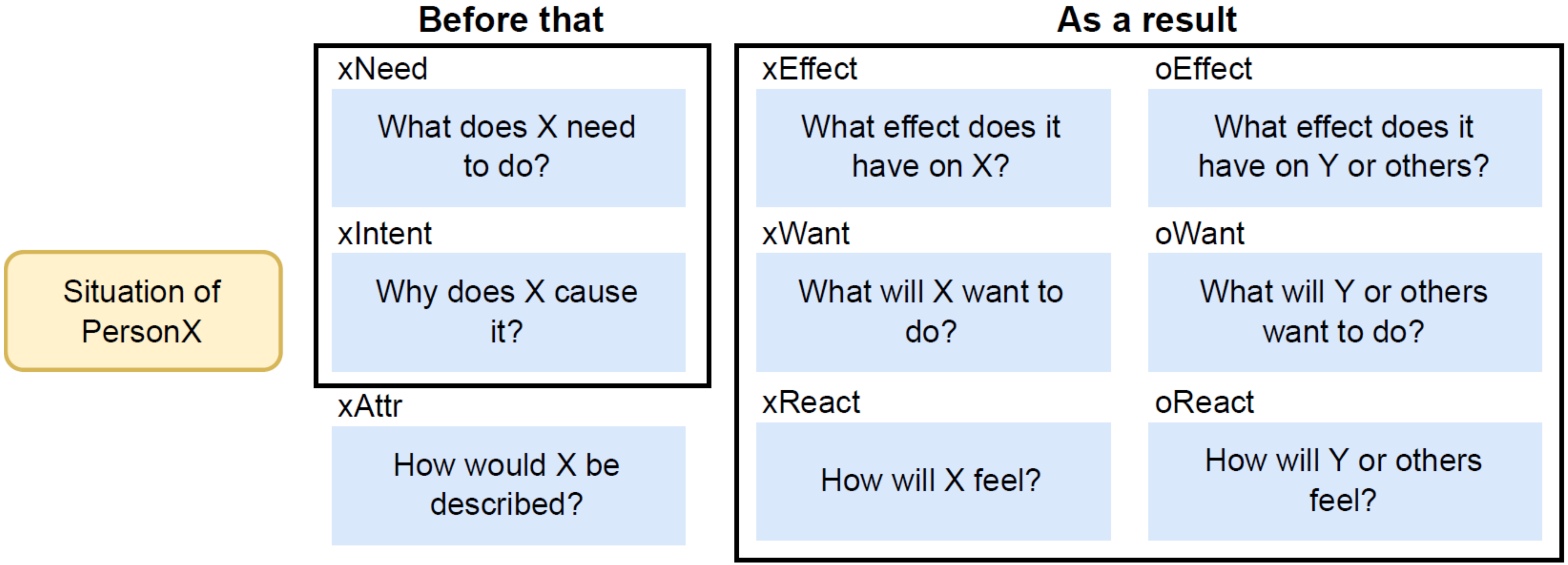


Essentially a crowdsourcing based approach + text mining

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

ATOMIC: Everyday If-then Commonsense Knowledge

- Crowdsourcing 9 Types of IF-THEN inferential knowledge
- All personal entity information has been removed to reduce ambiguity
- Mostly 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.

Usually in the Form of a Knowledge Graph

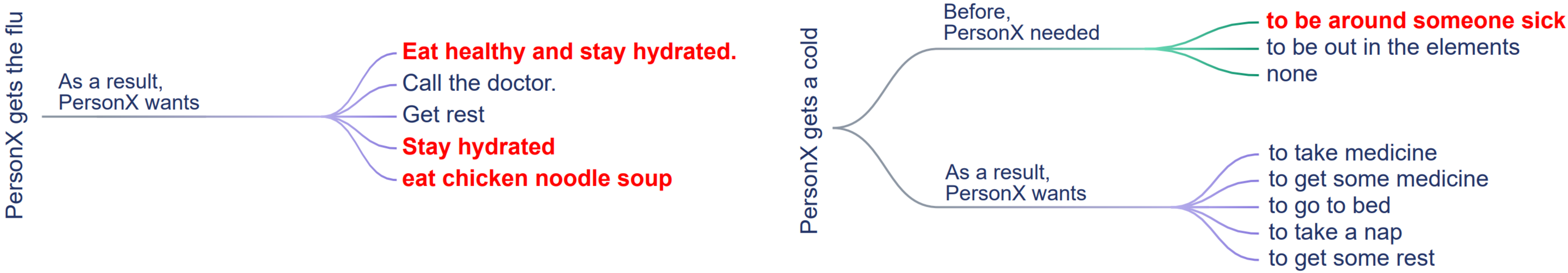
- Commonsense knowledge graph (CKG): represented in triples of texts
  - ⟨*h*: PersonX is hungry, *r*: xWant (then PersonX wants), *t*: to have lunch⟩

- An excerpt of the world with prototypical events and causes/consequences
  - Correspond to real-world situations

- But how could we know if it is generalizable?

Limited Coverage: Symbolic CKG

- CKGs can’t cover all the entities and situations
  - Not to say corresponding triples



Reasonable consequences for “PersonX gets the flu” are not covered by “PersonX gets a cold”

Source: https://mosaickg.apps.allenai.org/kg_atomic2020/

Conceptualization: A Missed Point

- Countless Entities and Situations in the real world
  - So many different things and situations we encounter new things everyday
- Humans understand the world through concepts
  - Summarize previous experiences into abstract mental representation
    - Entity concepts: animal (cat, dog, pet, tiger, …)
    - Situational concepts: a relaxing event (have a cup of coffee, take a break, …)
- CKGs are not enough (checked with Probase)

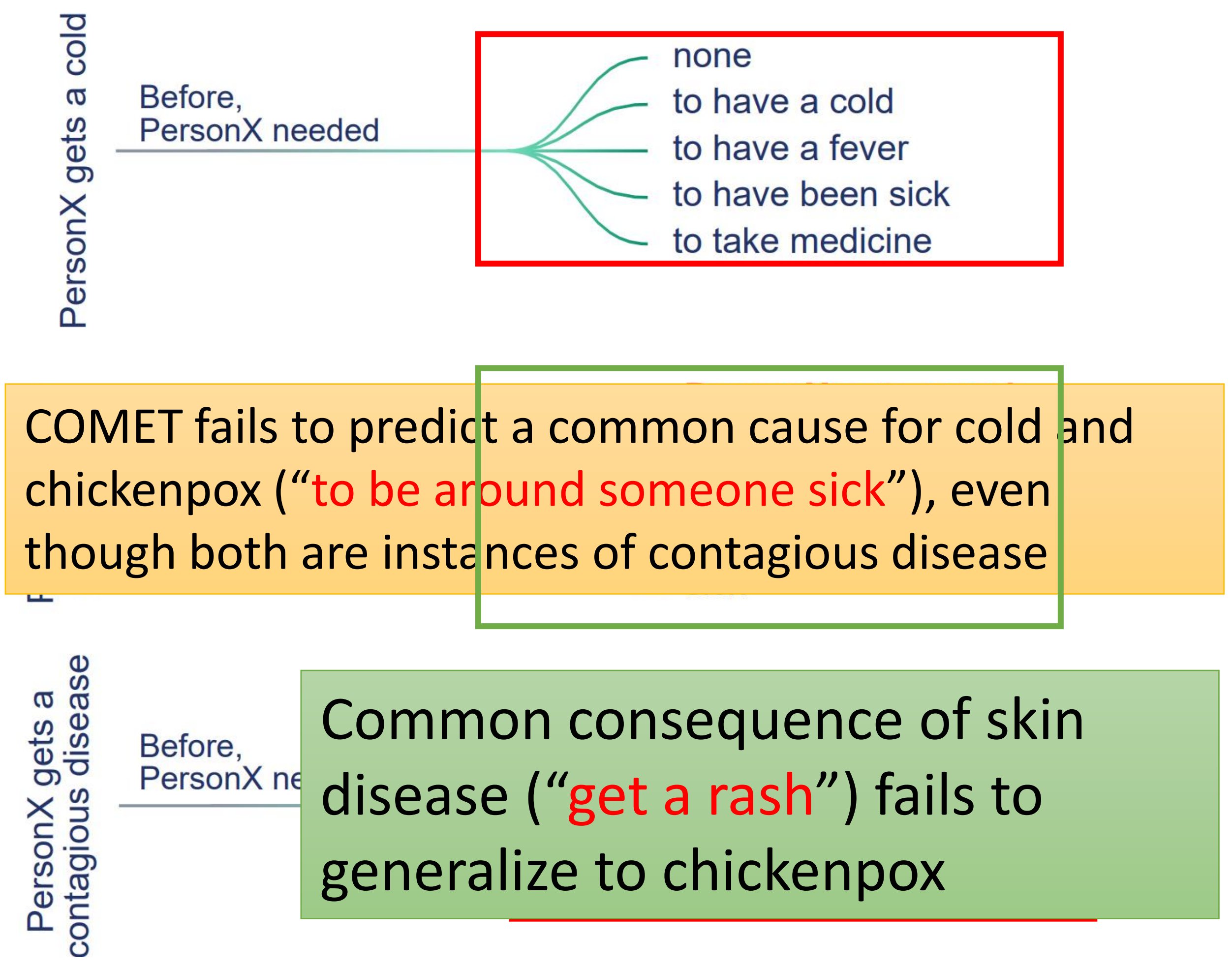
	ATOMIC [1]	ATOMIC-2020 [7]	DISCOS [4]
#Head	24.3K	44.0K	1,103.0K
#Triple	793.3K	1246.6K	3,235.9K
Average Degree	32.6	28.4	2.9
Concept Coverage	0.34%	0.76%	8.00%
Average Distinct Concept	0.093	0.114	0.048

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

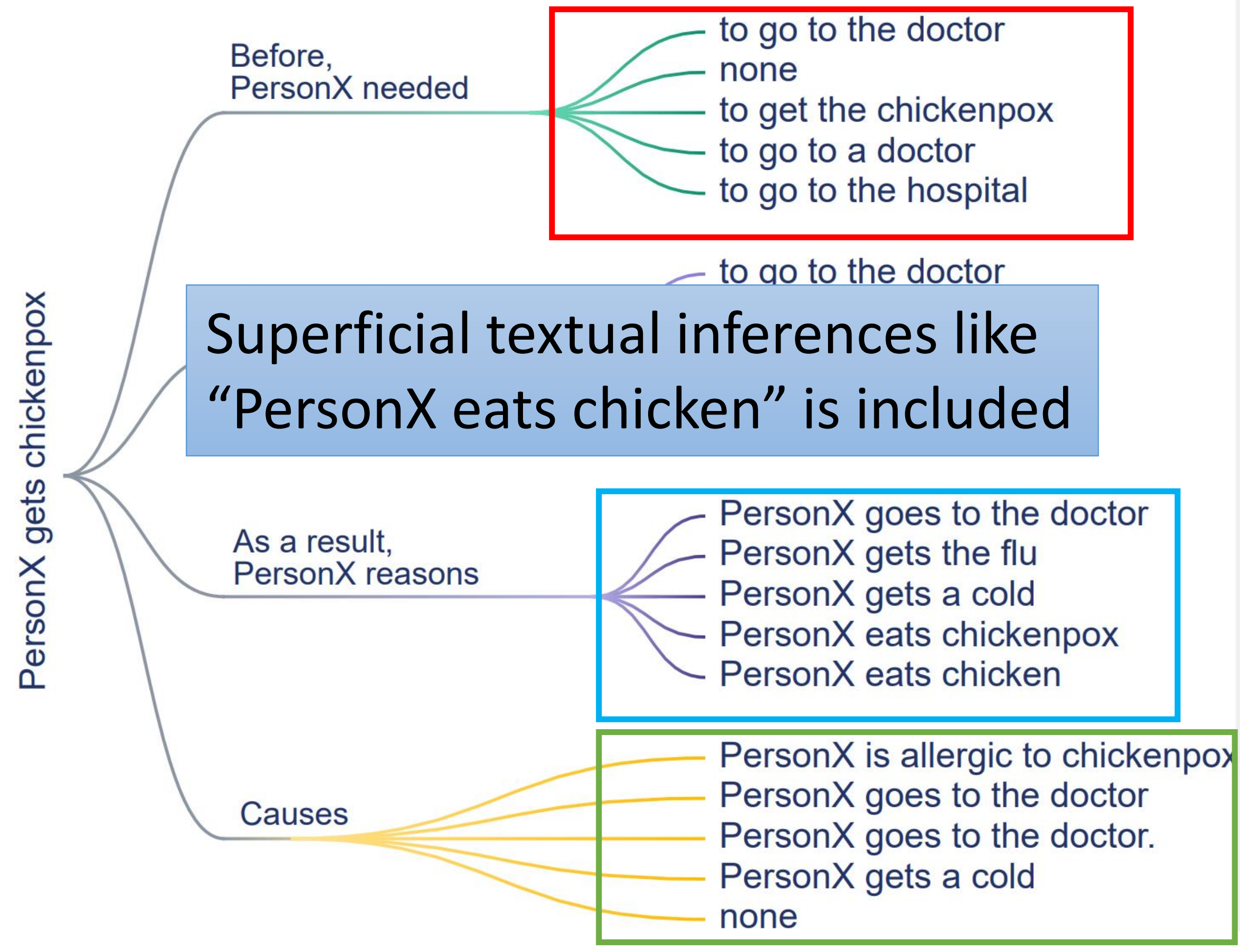
Limited Coverage: Neural Models

- Neural commonsense models to handle arbitrary texts?



COMET fails to predict a common cause for cold and chickenpox ("to be around someone sick"), even though both are instances of contagious disease

Common consequence of skin disease ("get a rash") fails to generalize to chickenpox



Superficial textual inferences like "PersonX eats chicken" is included

COMET: Commonsense Transformers for Automatic Knowledge Graph Construction Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, Yejin Choi. ACL, 2019.

Commonsense Reasoning in General

- Conceptual induction
  - Conceptualization and compositionality are keys to commonsense reasoning (generalization), but there is still lack of study

Induction

X: Item does not fit in container, REASON, item is big

Trophy is an item; Suitcase is a container

Conceptualization

Instantiation

Y: Trophy does not fit in suitcase, REASON, it is big

Deduction

Current deep learning usually just preforms induction by learning from examples, and then performs verification of a new claim, instead of decution.

CSKB/Training Data

- Computer not fit in parcel, REASON, Computer is big
- Rock not fit in carrier, REASON, rock is big
- …

- If we can explicitly instantiate new claims from abstractive claims, then we can have much more examples for learning models to learn.
- Conceptual induct is yet to be too difficult for language models: so many concepts

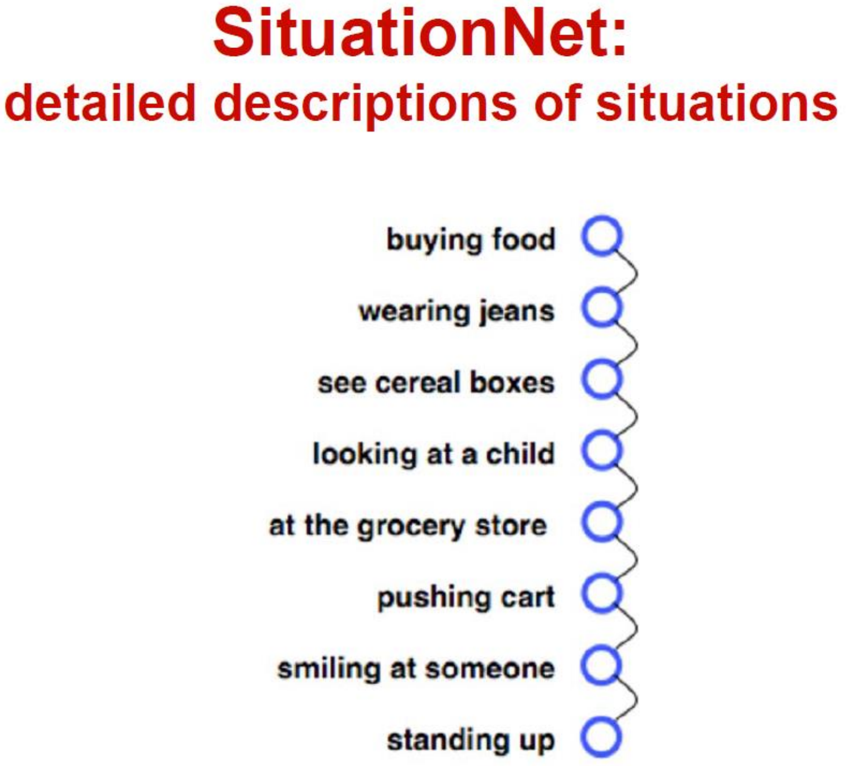
Conceptualization: Related Theories

- Vagueness
  - No strict hierarchy
  - Effort-taking on borderline cases
  - Diverse and context-dependent

- Abstraction of the world
  - K-line theory: not only conceptualize entities, but also events and mental states

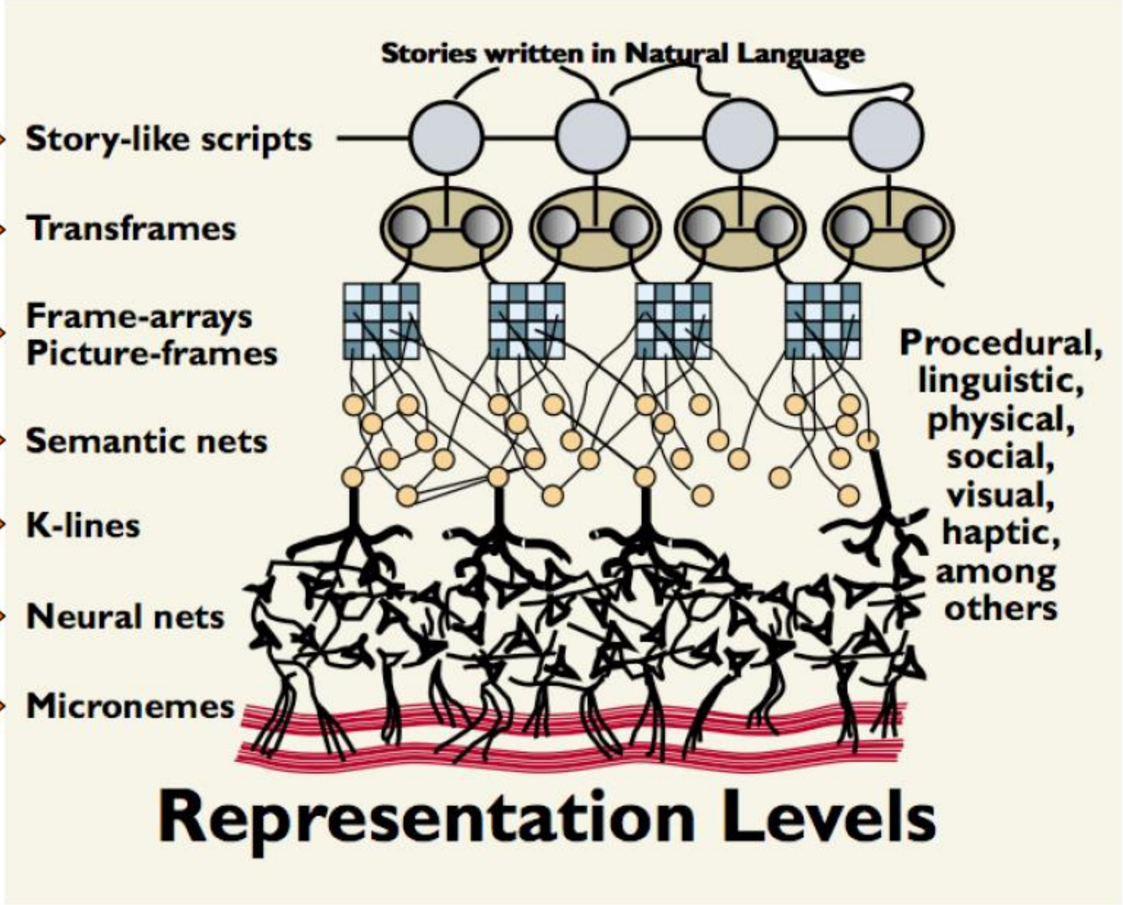
Murphy, Gregory. The big book of concepts. MIT press, 2004.
M. Minsky, “K-Lines: A theory of Memory," Cognitive Science 4 (1980). 117-133.

Representing Knowledge in Multiple Ways



Prototypical situations

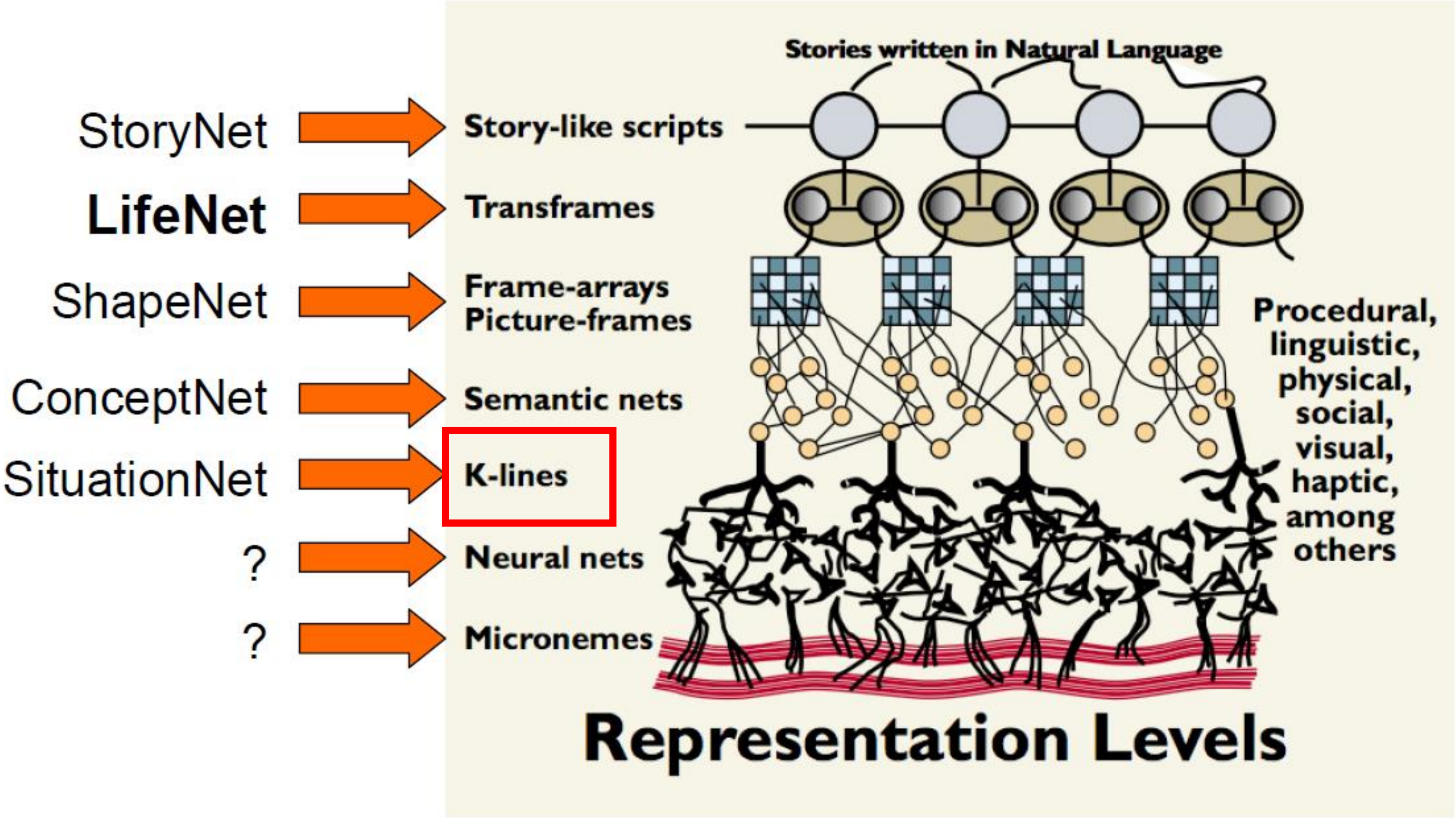
StoryNet
LifeNet
ShapeNet
ConceptNet
SituationNet
?
?



Push Singh, Barbara Barry, and Hugo Liu (2004). **Teaching machines about everyday life.** *BT Technology Journal*, 22(4):227-240.
Figure Credit: https://ocw.mit.edu/courses/media-arts-and-sciences/mas-961-ambient-intelligence-spring-2005/lecture-notes/week4_push_singh.pdf

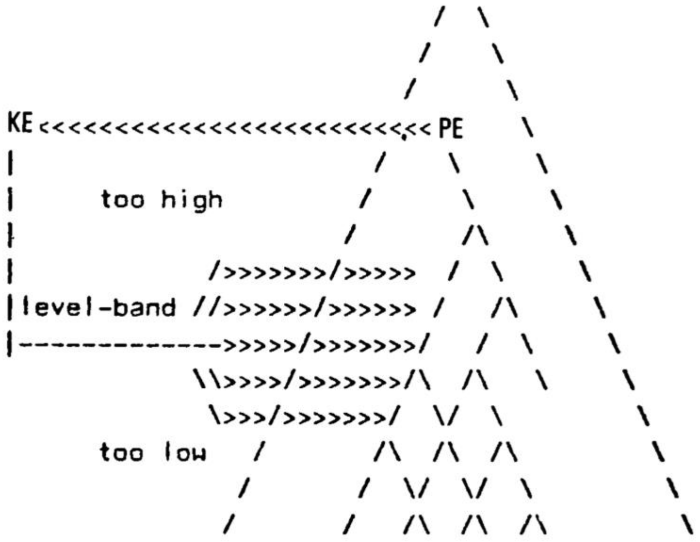
Representing Knowledge in Multiple Ways

- Encode memories in "abstract" form.
- Search all memory for the "nearest match."
- Remember "methods," not "answers."



M. Minsky, "K-Lines: A theory of Memory," Cognitive Science 4 (1980). 117-133.

The K-Line Theory



- Attach a K-node (a mental state, KE) to a "Pyramid" agent (PE) at a certain level
  - The pyramid is a tree structure that we use to conceptualize the world
  - The mapping has a lower-band limit and a higher-band limit, to compare the right common, non-conflicting properties
- Then the PE will help us to make abstraction, logical and procedural reasoning

- When comparing Tesla with Google, Toyota, some small company, we need the right level and right perspective of comparison
- E.g., mapping Tesla to a company, big company, IT company, AI company, high-tech company, automobile company

M. Minsky, "K-Lines: A theory of Memory," Cognitive Science 4 (1980). 117-133.

Outline

- Motivation
- Abstract ATOMIC: to develop a commonsense knowledge graph with more abstraction and conceptual induction capability

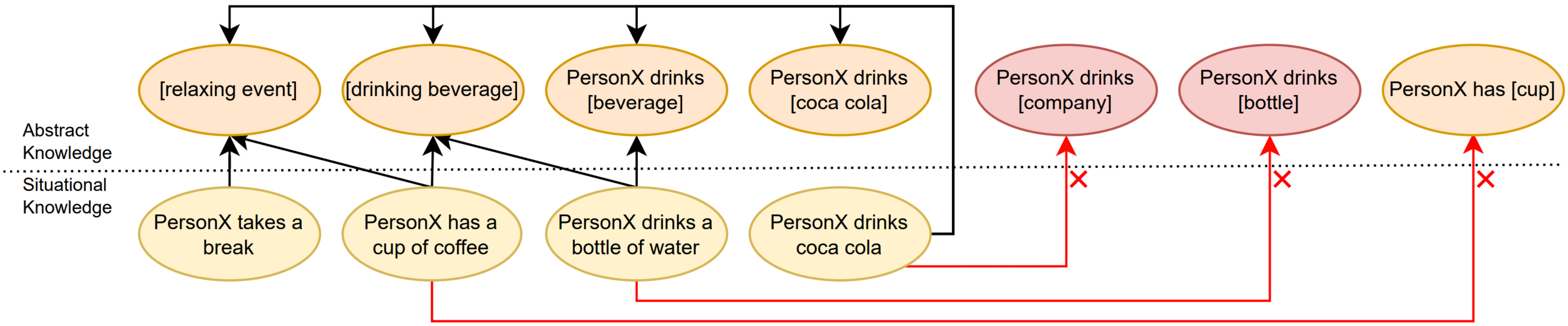
Abstract Events

- We call the original triple in ATOMIC the situational knowledge
  - ⟨*h*: PersonX is hungry, *r*: xWant (then PersonX wants), *t*: to have lunch⟩
  - Head is usually a description regarding everyday events on some unspecified PersonX
  - Tails are less complete as sentences and more difficult to parse for conceptualization

- Event conceptualization:
  - A **conceptualization** is an abstract simplified view of some selected part of the world
  - Different levels of abstractness
    - "PersonX drinks coca cola" as "[drinking coca cola]," "[drinking beverage]," "[event]"
  - Different perspectives
    - "Coca cola" as "[sugary beverage]," "[phosphate containing beverage]," "[iced drink]," not in a strict hierarchical taxonomy
      - PersonX drinks [iced drink], xReact, refreshed
      - PersonX drinks [sugary beverage], xEffect, gain weight

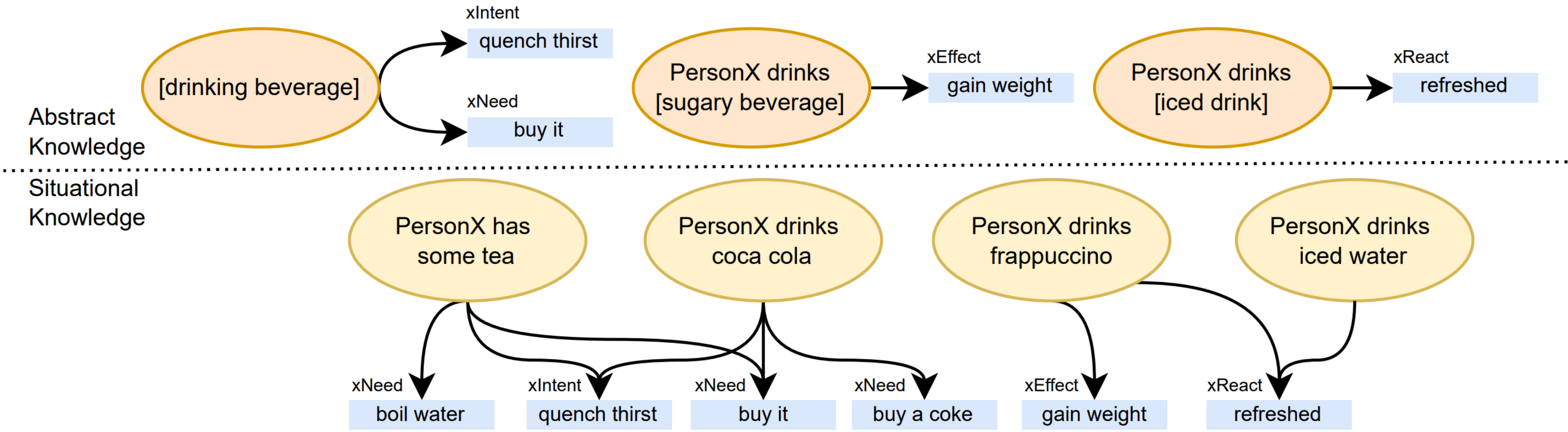
Abstract Events

- Formally defined as a textual template with a slot, filled by a concept, like “PersonX drinks [beverage],” “[drinking beverage],” “[relaxing event]”
- We construct a “bipartite graph” between situational knowledge and abstract knowledge



Abstract Triples

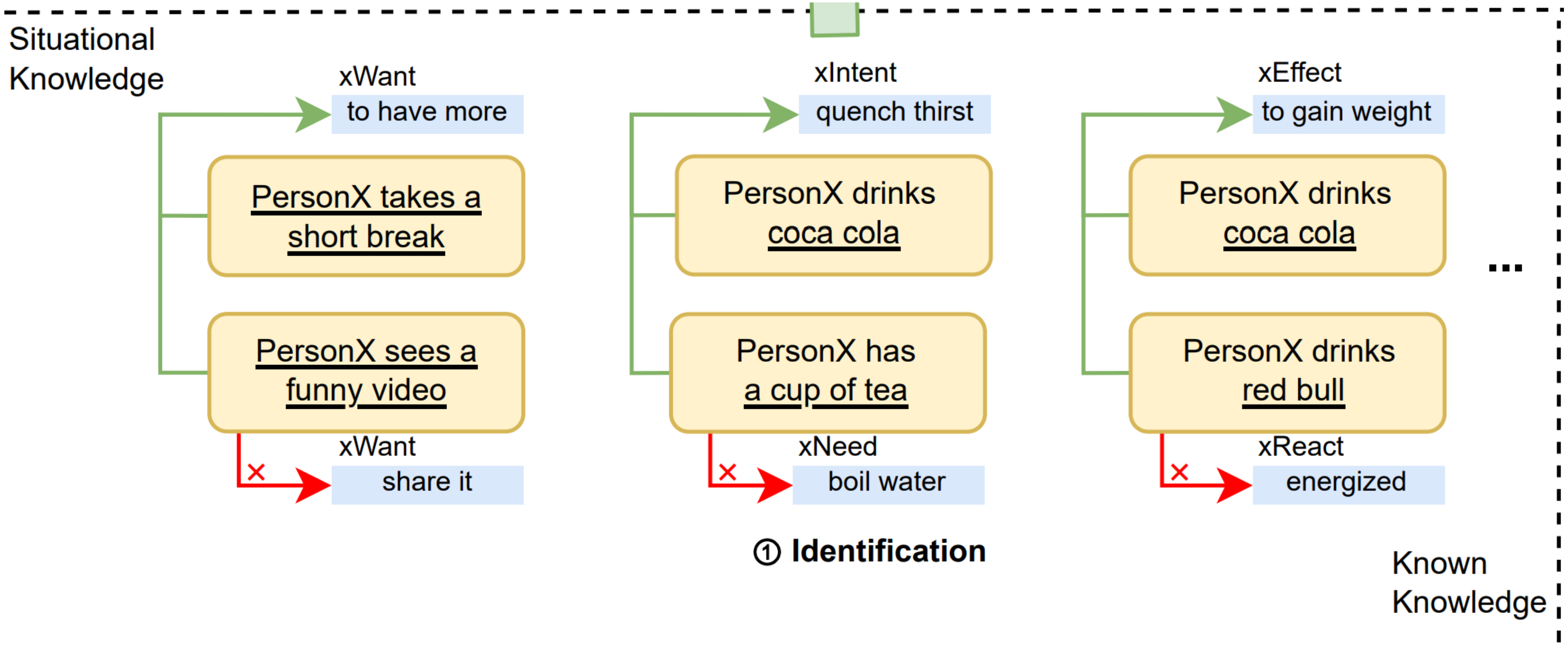
- Triples with an Abstract Event as the Head
  - E.g., link "PersonX drinks [sugary beverage]" to an effect of "gain weight"



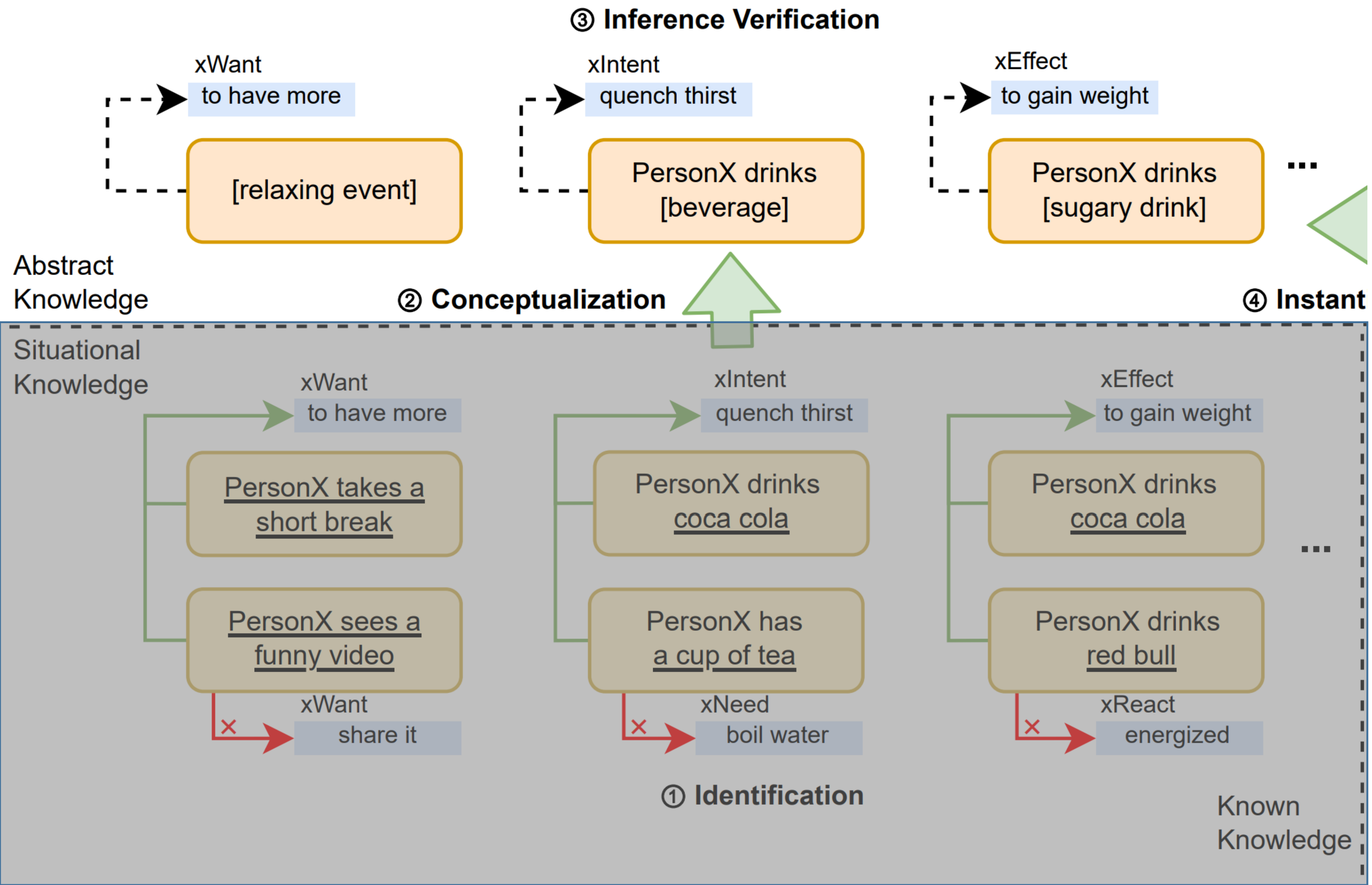
Validity

- Situational events and triples
  - Defeasibility of commonsense: each piece of commonsense knowledge is often termed as factoid that is only plausible, or typically true, according to human intuition without more formal reasoning
    - “a dog is smaller than a person” could be a false claim but is a commonsense
- Abstract events and triples
  - By whether it is typically valid among instantiations
    - PersonX has a cup of coffee → [relaxing event]
  - Even if itself is not ordinary CKG event or triple in natural language
    - PersonX drinks [phosphate containing beverage]
    - PersonX spends [time interval] reading
- Event conceptualizations
  - By whether it covers the meaning in the context
    - PersonX has a cup of coffee → PersonX has [cup]?

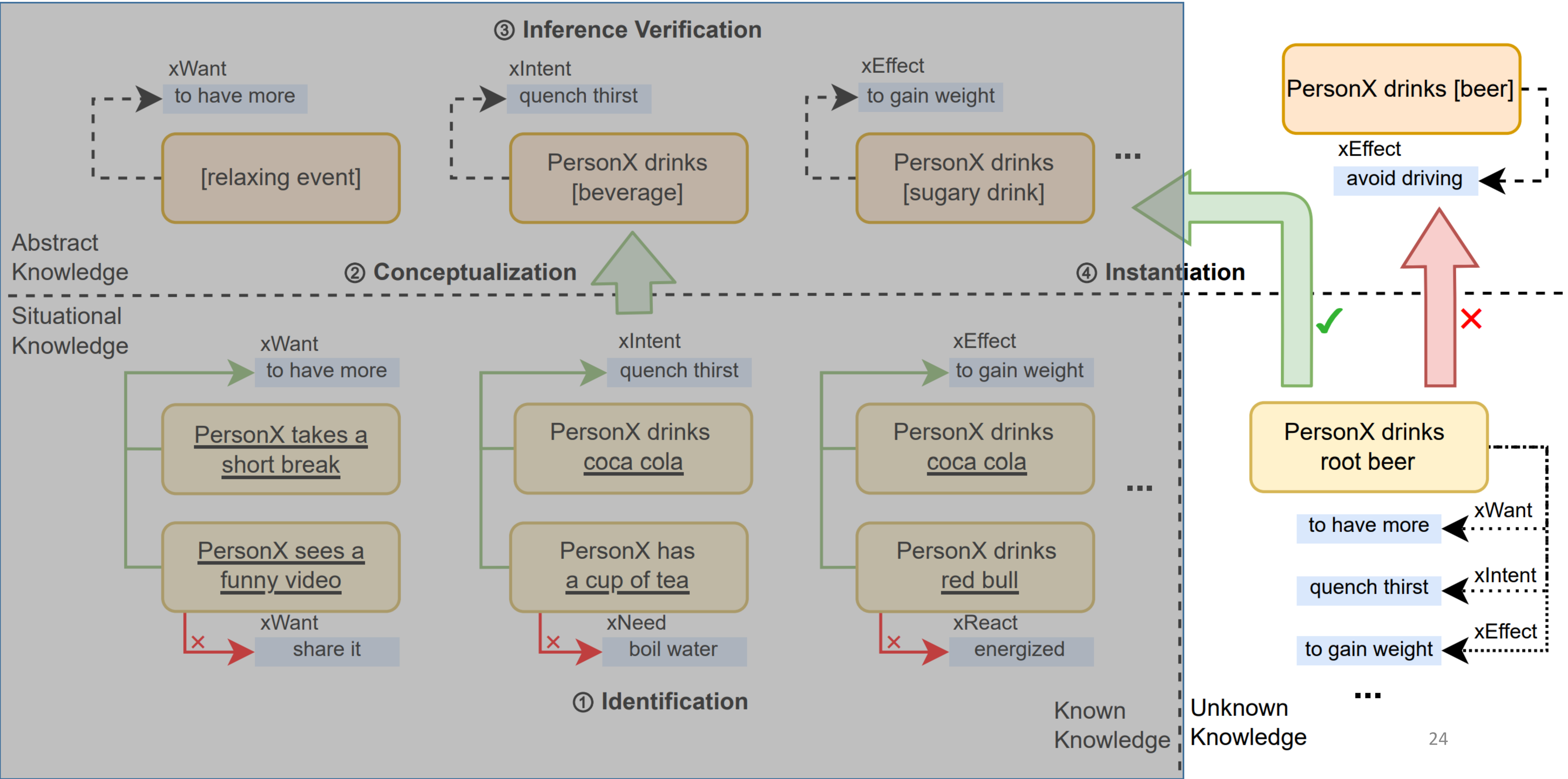
Overall Running Example



Overall Running Example

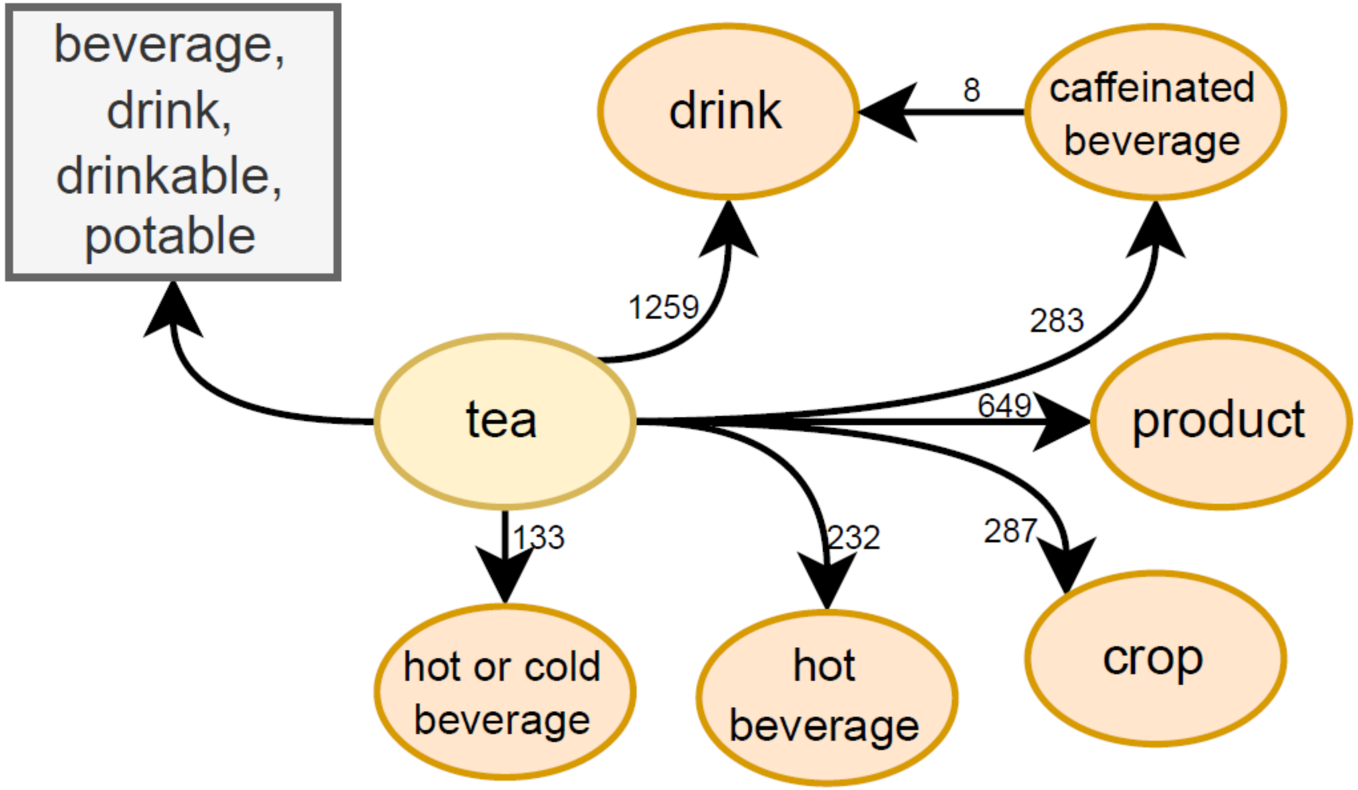


Overall Running Example



Concept Bank

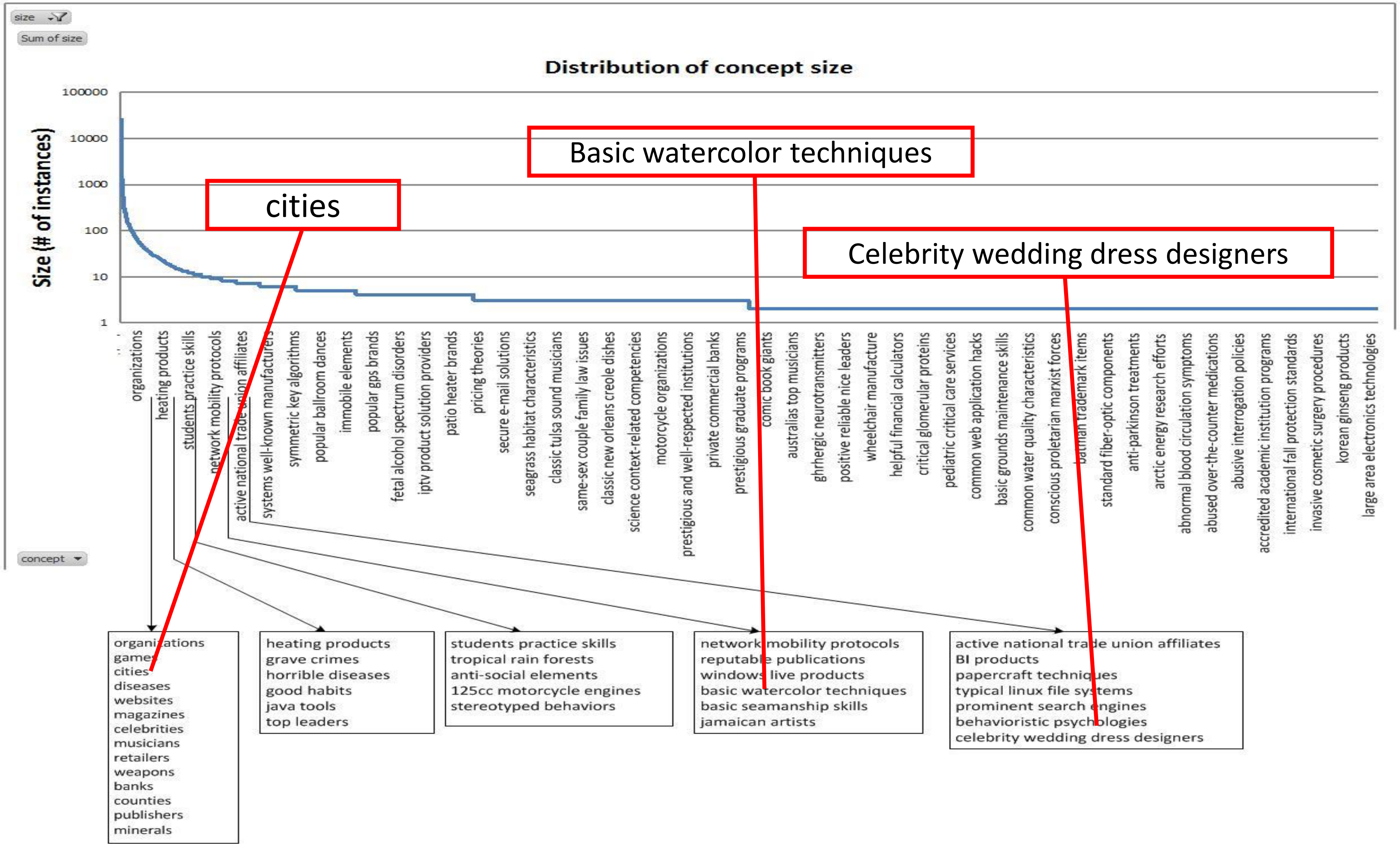
- Use both WordNet and Probase
  - To cover the flexibility of human conceptualization

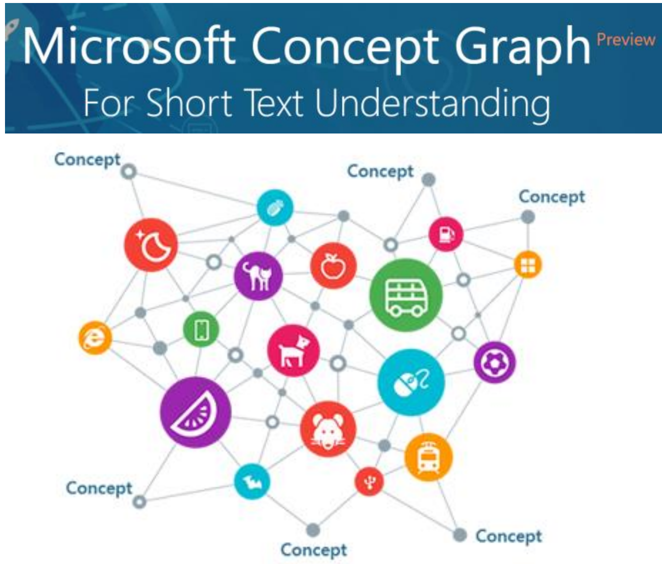


Data are available at https://concept.research.microsoft.com/

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

ProBase





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

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

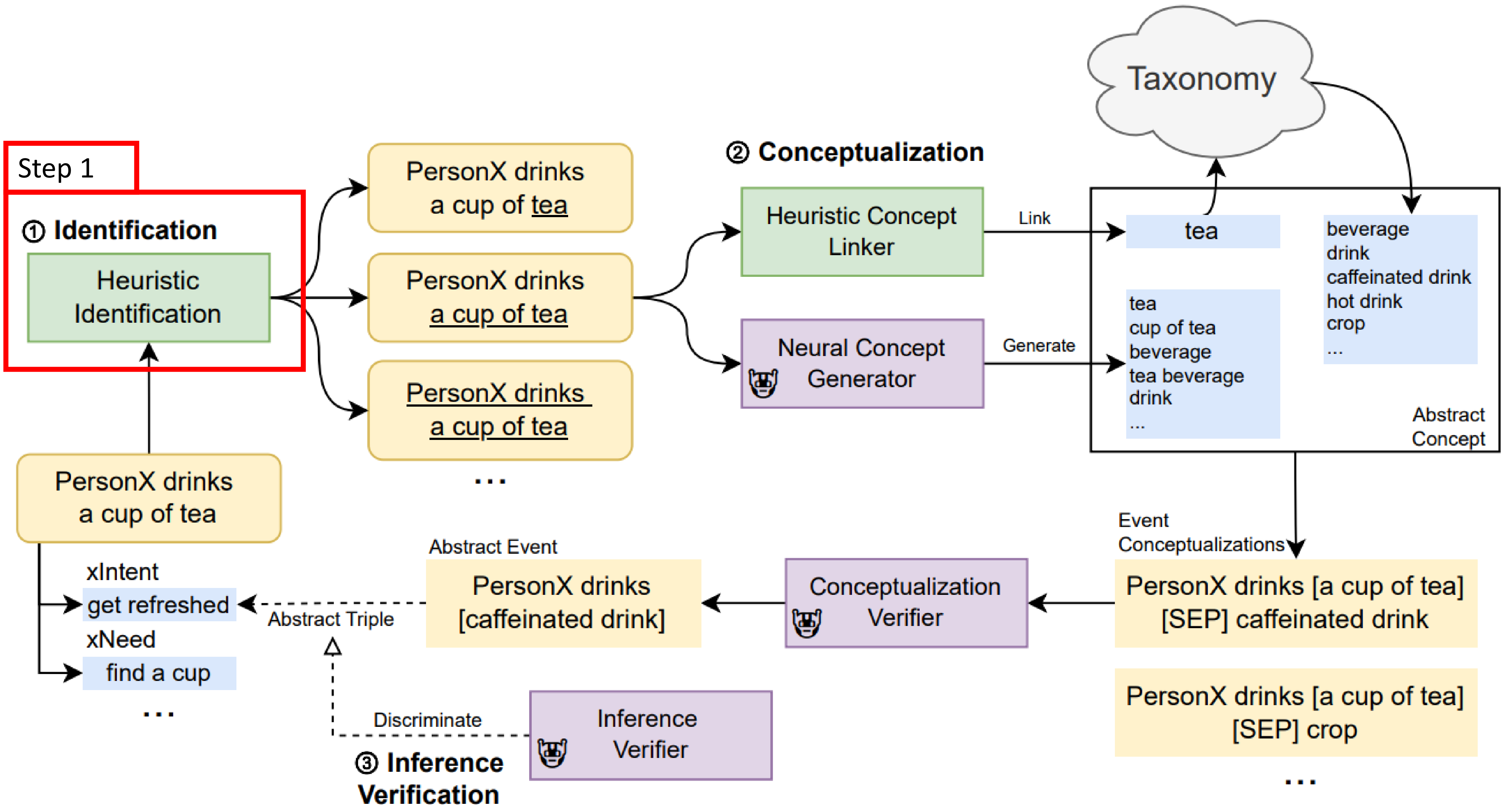
Wentao Wu, Hongsong Li, Haixun Wang, Kenny Qili Zhu: Probase: a probabilistic taxonomy for text understanding. SIGMOD Conference 2012: 481-492

Slide Credit: Haixun Wang

Outline

- Motivation

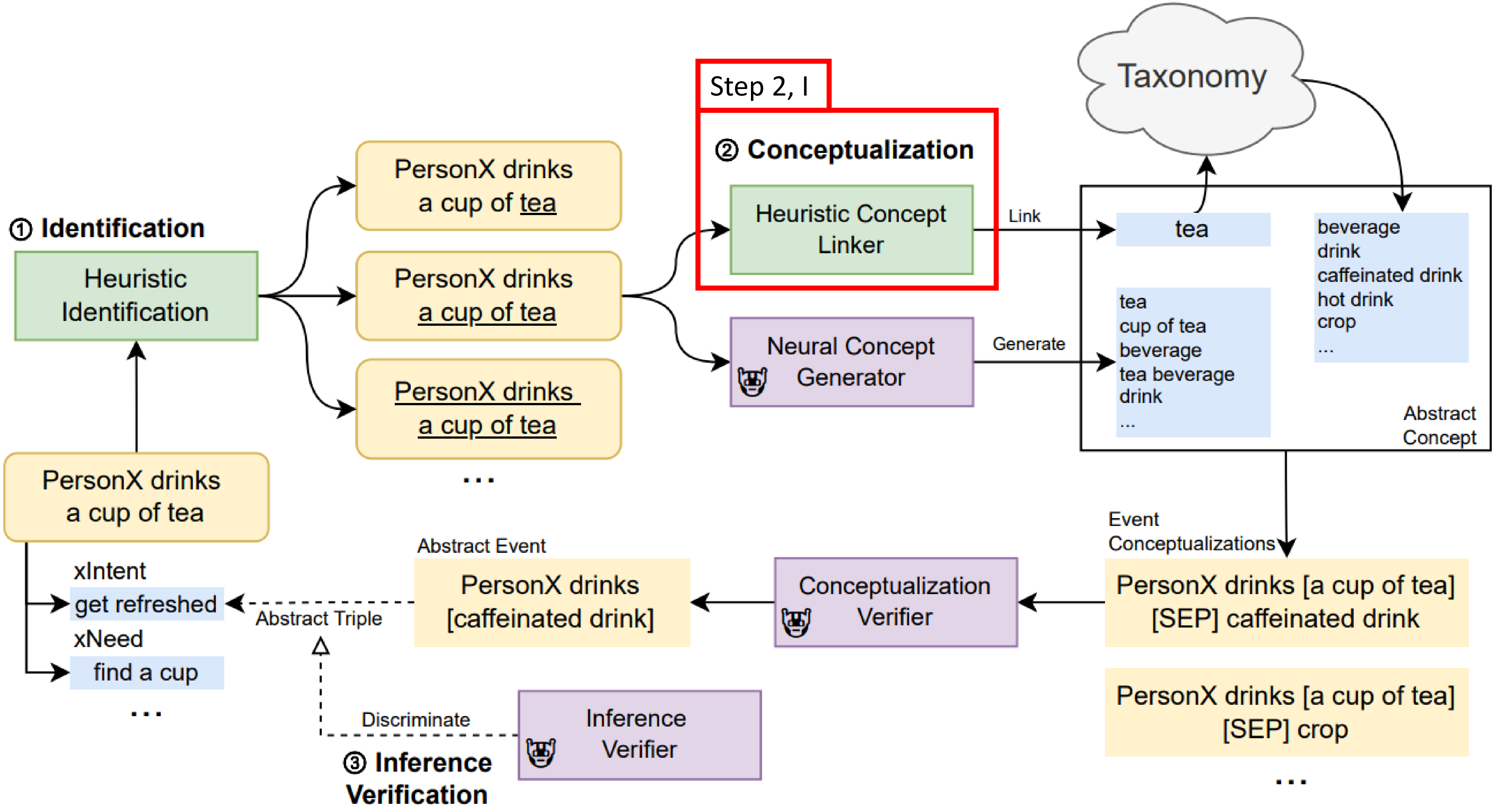
- Abstract ATOMIC
  - Construction steps


Step 1
① Identification
Heuristic Identification
PersonX drinks a cup of tea
PersonX drinks a cup of tea
PersonX drinks a cup of tea
PersonX drinks a cup of tea
...
② Conceptualization
Heuristic Concept Linker
Neural Concept Generator
Link
Generate
Taxonomy
tea
beverage
drink
caffeinated drink
hot drink
crop
...
tea
cup of tea
beverage
tea beverage
drink
...
Abstract Concept
Event Conceptualizations
PersonX drinks [a cup of tea] [SEP] caffeinated drink
PersonX drinks [a cup of tea] [SEP] crop
...
Conceptualization Verifier
Abstract Event
PersonX drinks [caffeinated drink]
Abstract Triple
xIntent
get refreshed
xNeed
find a cup
...
Discriminate
Inference Verifier
③ Inference Verification

Step 1: Identification

- Identify candidates
  - Word matching is not enough
    - “She gives her pet food” (She gives food to her pet)

- Work on dependency tree
  - Check each constituent/subtree
  - Nominal candidate or predicative candidate
  - Determined by linguistic tags by a set of rules (customized to ATOMIC data)
    - E.g., if a constituent is a noun according to POS, and a direct object of a verb, then it should be an entity


Step 2, I
② Conceptualization
Heuristic Concept Linker
Neural Concept Generator
Link
Generate
Taxonomy
tea
beverage
drink
caffeinated drink
hot drink
crop
...
tea
cup of tea
beverage
tea beverage
drink
...
Abstract Concept
① Identification
Heuristic Identification
PersonX drinks a cup of tea
PersonX drinks a cup of tea
PersonX drinks a cup of tea
PersonX drinks a cup of tea
...
xIntent
get refreshed
xNeed
find a cup
...
Abstract Triple
Abstract Event
PersonX drinks [caffeinated drink]
Conceptualization Verifier
Event Conceptualizations
PersonX drinks [a cup of tea] [SEP] caffeinated drink
PersonX drinks [a cup of tea] [SEP] crop
...
Discriminate
Inference Verifier
③ Inference Verification

Step 2: Conceptualization (I: Concept Linking)

- To link an entity or situation to the concept bank
  - Require sophisticated natural language understanding
- Possibly distinct from the text
  - “a cup of coffee” -> coffee (instead of the head word cup)
  - “Alice lives with her boyfriend” -> cohabitation (instead of simply living)

Step 2: Conceptualization (I: Heuristic Rules)

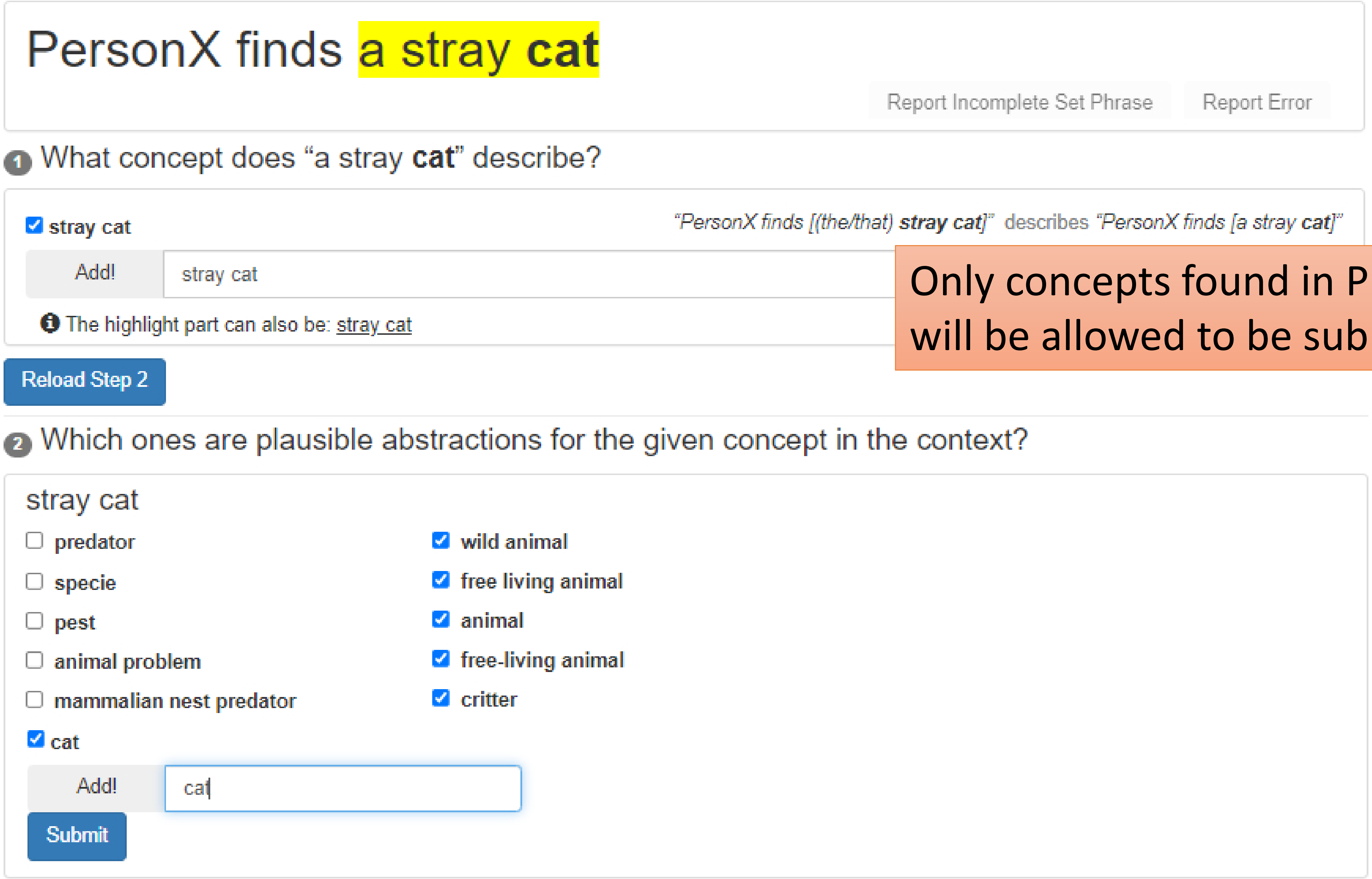
- Rules to derive possible concepts
  - As WordNet or Probase nodes
  - Use linguistic features for the constituent
  - Aided by WordNet relations, NOMBANK, and GlossBERT

- Impossible to be accurate
  - Goal: to provide good candidates

Type	Example	Concepts	Method
Word	PersonX finds some cats	cat	Directly use the headword, possibly lemmatized
Compound/ Phrase	PersonX sees many stray cats	stray cat	Collect compounds or phrases in nominal candidates by word matching in the constituent, subject to inflections
Predicate (Verb)	PersonX drinks coffee	drinking	Directly use the gerund of the verb
	PersonX says he enjoys himself	enjoyment	Check WordNet and NOMBANK for the noun form of the verb
Predicate (Adj.)	PersonX is happy	happiness	Same as above, for a copula with adjective complement
Conjunction	PersonX sees doctors and nurses	doctor, nurse	Use concepts from both conjuncts
Nominal Candidate with Classifiers	PersonX has a cup of tea	tea	Directly return results from the accompanied argument, if the head and the preposition form a NOMBANK transparent construction
	PersonX sees a group of people	group, people	If an argument is connected to another one by "of" but not a NOMBANK transparent construction, both the head and accompanied argument are used.
Verb Phrase	PersonX drinks coffee	drinking coffee	Combine the verb with its arguments
Phrasal Verb	PersonX gets up late	get up	Check WordNet for combination of the verb and one or two particle in the text
Light Verb	PersonX gives a speech	giving, speech	For light verbs like *give*, *take*, *have*, etc., the predicand can be the actual concept
	PersonX goes shopping	shopping	Directly use results from the predicand when it is a gerund
Raising-to-subject	PersonX seems to be happy	happiness	Directly use results from the predicand for cases like *seem*, *appear*, *used*, etc.
Control Verb	PersonX wants to leave	want, leave	Use both the head (superordinate verb) and results from its complement
Adj. + Infinitive	PersonX is likely to leave	leave	Directly use results from the infinitive for a copula with some adjectives like *likely*, *going*, *about*, *able*, etc. as complement

GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge. Luyao Huang, Chi Sun, Xipeng Qiu, Xuanjing Huang. EMNLP, 2019.

Annotation Phase 1



Only concepts found in Probase will be allowed to be submitted

Annotation Phase 2

PersonX finds a stray cat

Do the sentences below cover the meaning above?

PersonX finds stray cat	● Yes	○ No
PersonX finds animal	● Yes	○ No
PersonX finds animal problem	○ Yes	● No
PersonX finds free living animal	● Yes	○ No
PersonX finds critter	● Yes	○ No
PersonX finds free-living animal	● Yes	○ No
PersonX finds mammalian nest predator	○ Yes	● No
PersonX finds predator	○ Yes	● No
PersonX finds wild animal	● Yes	○ No
PersonX finds feline	● Yes	○ No
PersonX finds specie	○ Yes	● No
PersonX finds mammalian	○ Yes	● No

Feel free to add these determiners before the substitution to make it valid: the, a, PersonX's, some, the event of, the action of, at (the state of), with (the attribute of), ...

Submit

Annotation Phase 2

I heard that **PersonX gets hit by a car**

Do the sentences below cover the meaning above?

I heard about (the/PersonX's) accident	◉ Yes	○ No
I heard about (the/PersonX's) emergency	◉ Yes	○ No
I heard about (the/PersonX's) traffic incident	◉ Yes	○ No
I heard about (the/PersonX's) offense	○ Yes	◉ No
I heard about (the/PersonX's) serious violation	○ Yes	◉ No
I heard about (the/PersonX's) traffic offense	○ Yes	◉ No
I heard about (the/PersonX's) emergency situation	◉ Yes	○ No
I heard about (the/PersonX's) unexpected event	◉ Yes	○ No
I heard about (the/PersonX's) car accident	◉ Yes	○ No
I heard about (the/PersonX's) event	◉ Yes	○ No
I heard about (the/PersonX's) injury	◉ Yes	○ No
I heard about (the/PersonX's) traumatic event	◉ Yes	○ No

Feel free to add these determiners before the substitution to make it valid: the, a, PersonX's, some, the event of, the action of, at (the state of), with (the attribute of), ...

Submit

Annotation: Quality Control

- Rigorous worker enrollment
  - 95% acceptance on at least 1000 tasks
  - One or two qualification tests
- Detailed instructions
  - Over 30 examples
- In-progress monitoring
  - Disqualify underperformed workers, and discard their annotations

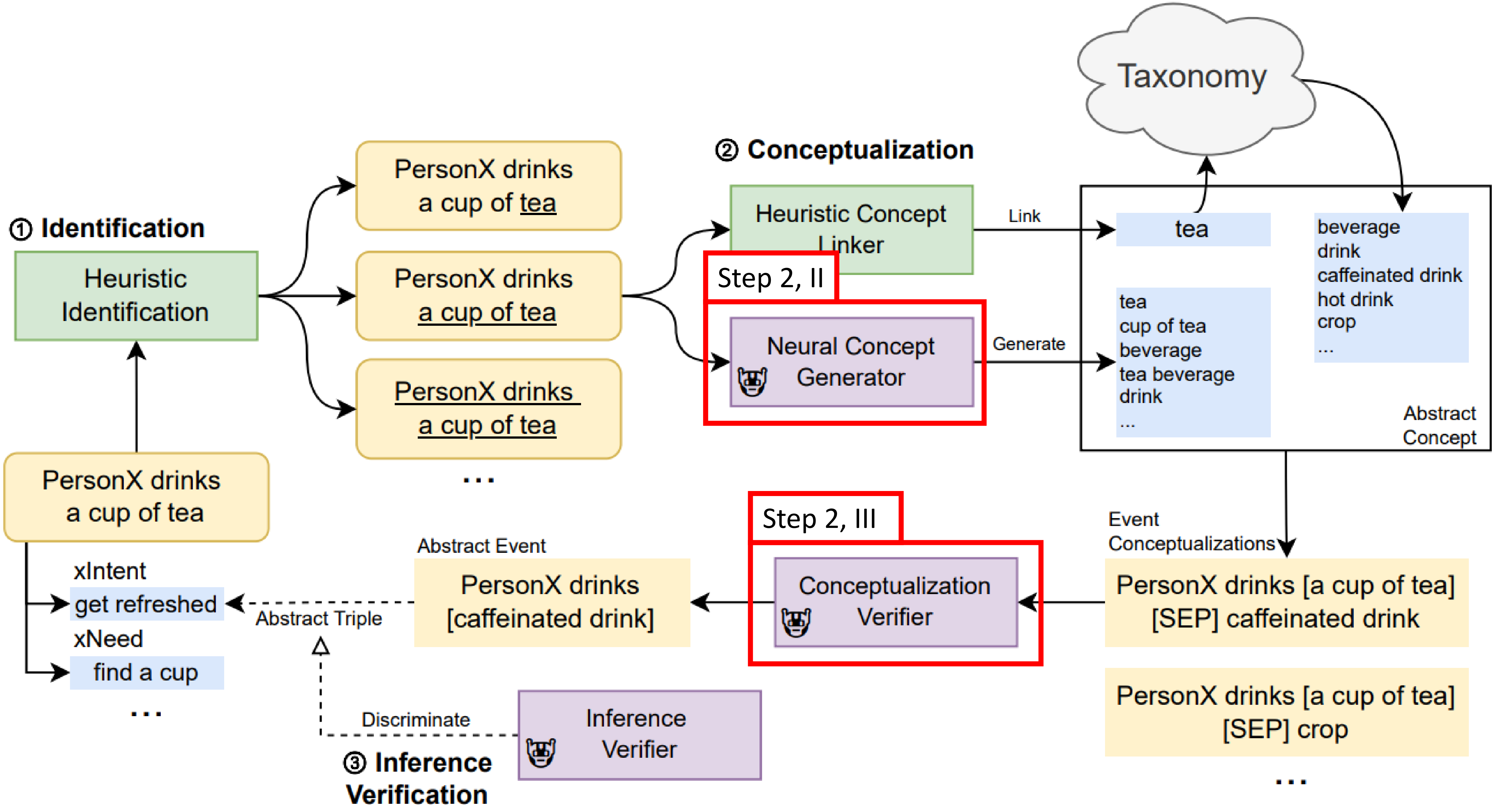
Annotation Results

- For 24.3K events in ATOMIC, after filtering general concepts, idioms, and duplicates, 15.9K events are valid candidates
- Then 10K events (with entities or situations to be conceptualized) are randomly selected, from 8,045 original ATOMIC events (around 1/3 of ATOMIC heads)
- After annotation, 7,019 ATOMIC events were used to form 18,964 different positive abstract events

	Event	Triple
Total Questions	131,004	90,207
Questions with Agreement	92,235	81,197
Positives	40,833	65,900
Positive Rate	44.27%	81.16%
Inter-annotator Agreement	71.4%	73.0%
Manual Inspection Agreement	87.0%	86.5%

Conceptualization: Rules and Models

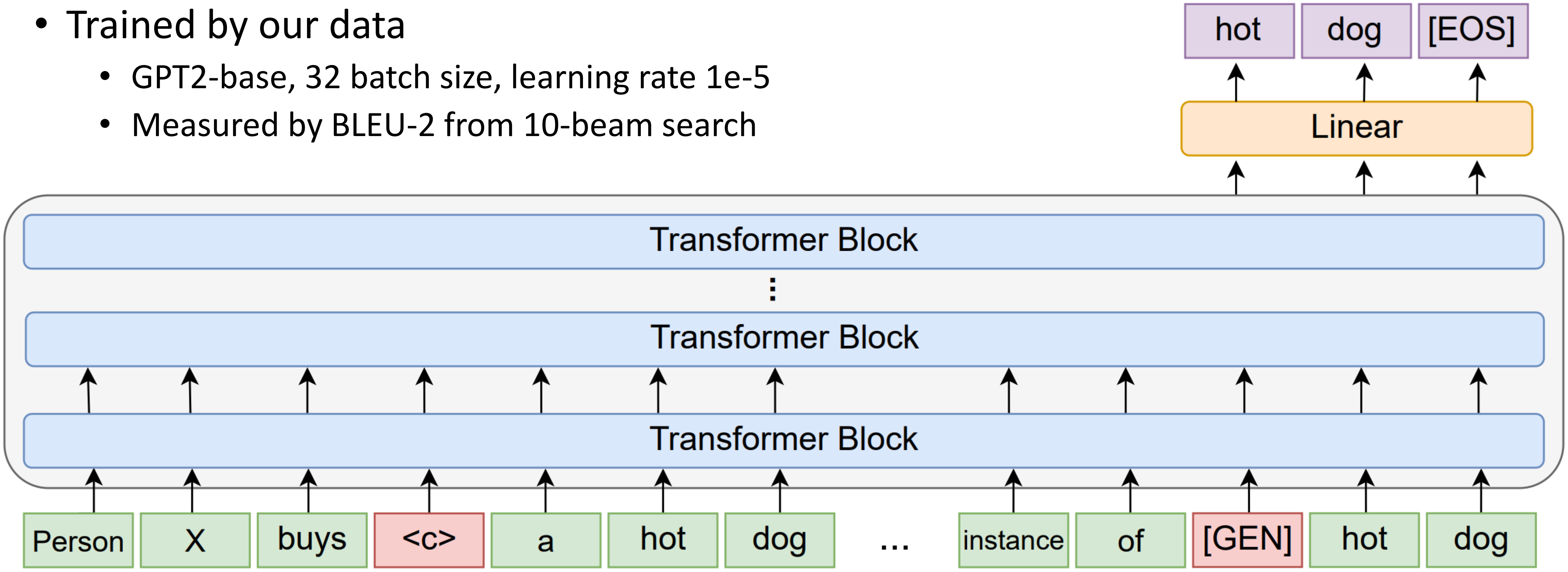
- Conceptualization needs both NLU and taxonomic knowledge
- Rules use taxonomic KG explicitly
  - But lack of contexts
- Neural models are contextualized
  - Doubtful diversity
  - …even the training data is aided by the taxonomic KG
- We combine both approaches and introduce a gatekeeper module



Taxonomy
② Conceptualization
① Identification
Heuristic Identification
PersonX drinks a cup of tea
PersonX drinks a cup of tea
PersonX drinks a cup of tea
...
Heuristic Concept Linker
Link
tea
beverage
drink
caffeinated drink
hot drink
crop
...
Step 2, II
Neural Concept Generator
Generate
tea
cup of tea
beverage
tea beverage
drink
...
Abstract Concept
PersonX drinks a cup of tea
xIntent
get refreshed
xNeed
find a cup
...
Abstract Event
PersonX drinks [caffeinated drink]
Abstract Triple
Step 2, III
Conceptualization Verifier
Event Conceptualizations
PersonX drinks [a cup of tea] [SEP] caffeinated drink
PersonX drinks [a cup of tea] [SEP] crop
...
Discriminate
Inference Verifier
③ Inference Verification

Step 2: Conceptualization (II: Neural Concept Generator)

- Use prompt to directly generate possible abstractions
  - With a neural autoregressive generator
    - PersonX buys <c> a hot dog </c>. <c> hot dog </c> is an instance of [GEN]
  - Trained by our data
    - GPT2-base, 32 batch size, learning rate 1e-5
    - Measured by BLEU-2 from 10-beam search



(a) Architecture of LM generator for Event Conceptualization based on pretrained GPT-2.

Neural Concept Generator: Results

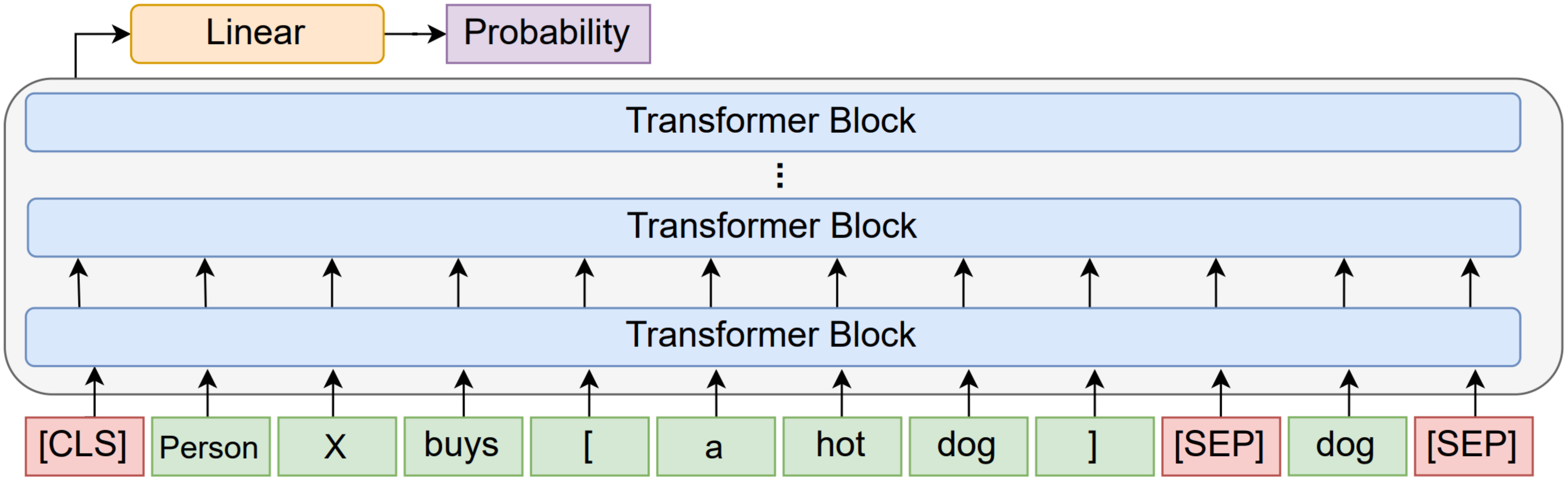
- Trained on the annotated data
  - Split by the original ATOMIC event partition
  - Use best models on its *dev* set
  - Alternative prompts and hyper-parameters show no improvements

BLEU-1	Dev Set	Test Set
Supervised Generator	65.1	68.0
GPT2 Zero-shot	25.0	20.4

BLEU-2	Dev Set	Test set
Supervised Generator	61.1	56.5
GPT2 Zero-shot	4.8	2.6

Step 2: Conceptualization (III: Verifier)

- Gatekeeping all event abstractions we found
  - With our annotated data
  - RoBERTa-base, 64 batch size, learning rate 2e-5
  - Measured by accuracy, threshold from *dev* set



(b) Architecture of LM disriminator for Event Conceptualization and Abstract Triple, based on pretrained RoBERTa.

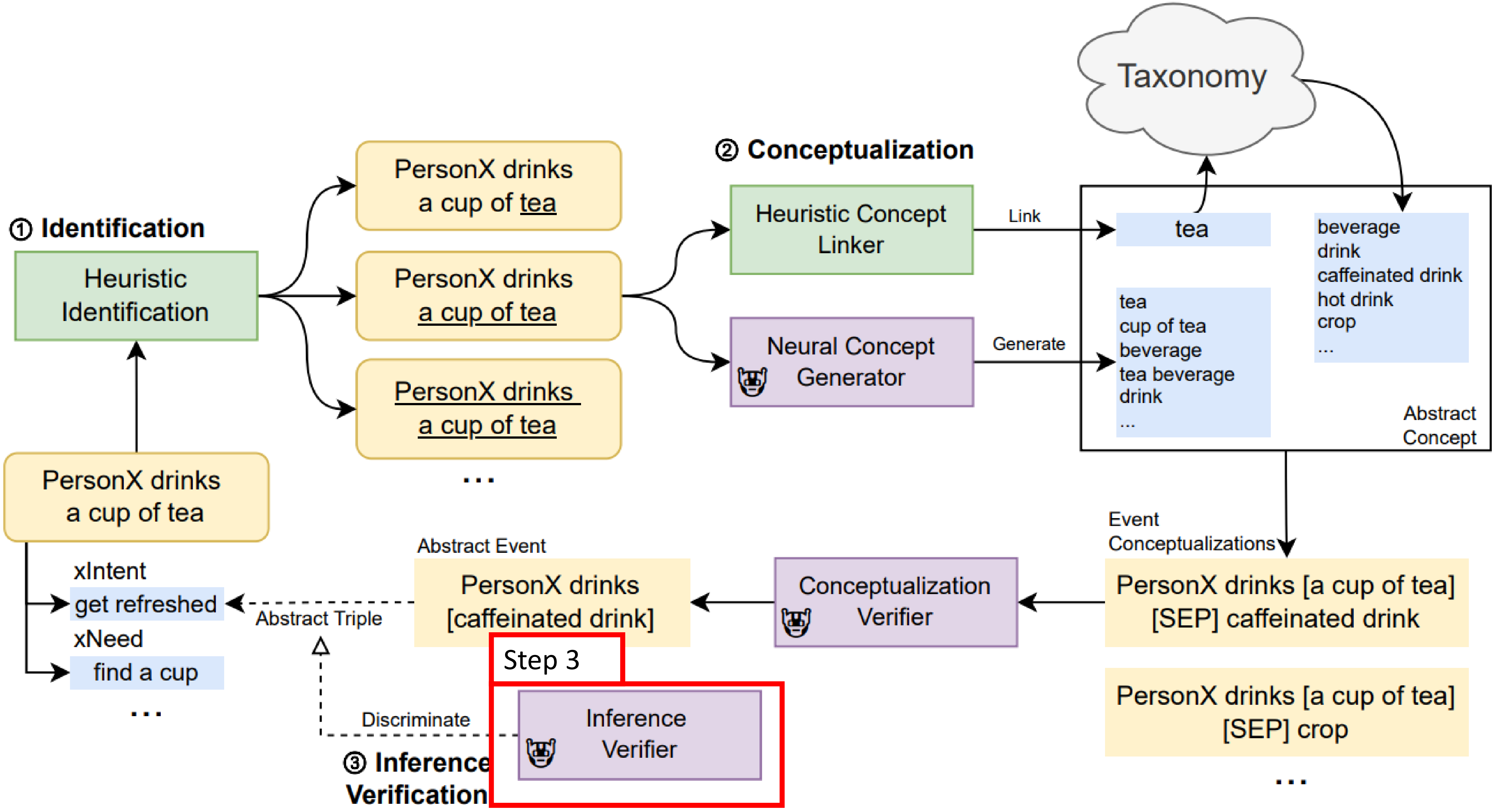
Verifier: Results

- Baselines
  - Negative Sampling
    - Using positive samples only, 5 negatives per positive
  - Generator for discrimination
    - Use losses as scores

Accuracy	Dev Set	Test Set
Supervised Learning	83.9	85.0
Negative Sampling	78.4	78.8
Generator – Supervised (GPT-2)	76.9	77.6
Generator - Zero-shot (GPT-2)	55.1	58.5

Examples

Event	Linked concepts	Conceptualization (Heuristic linking)	Conceptualization (Neural Generation)
PersonX buys [a new iphone]	iphone, new iphone	IOS device, smart phone, mobile devide, device, apple device, product, …	smart phone, phone, product, electronic device, device, mobile device
PersonX pays [PersonX's water bill]	bill, water bill	bill, expense, basic household expense, user charge, utility bill, …	bill, expense, …
[PersonX gets a cold]	contracting, cold	cold, common illness, illness, infection, minor ailment, respiratory infection, upper respiratory infection, viral infection, …	cold, condition, sickness, …
[PersonX gives birth to children]	giving birth, birth, gift	occasion, birth, life change, life event, happy event, …	creation, birth, event, life event, life cycle event, life changing event
[PersonX has a bad day at work]	experience	experience	difficulty, negative event, problem, unpleasant experience, …



Taxonomy
② Conceptualization
① Identification
Heuristic Identification
PersonX drinks a cup of tea
PersonX drinks a cup of tea
PersonX drinks a cup of tea
PersonX drinks a cup of tea
...
Heuristic Concept Linker
Neural Concept Generator
Link
Generate
tea
beverage
drink
caffeinated drink
hot drink
crop
...
tea
cup of tea
beverage
tea beverage
drink
...
Abstract Concept
Event Conceptualizations
PersonX drinks [a cup of tea] [SEP] caffeinated drink
PersonX drinks [a cup of tea] [SEP] crop
...
Conceptualization Verifier
Abstract Event
PersonX drinks [caffeinated drink]
Abstract Triple
xIntent
get refreshed
xNeed
find a cup
...
Step 3
Inference Verifier
Discriminate
③ Inference Verification

Inference

- Collect all tails from the instantiations in ATOMIC
- Verify if it applies to the abstract event
  - With a neural model on our annotated data

Annotation: Inference

Question 1

If **PersonX asks PersonY [marriage]**, ***after that/as result, the Effect on PersonX is:*** **personX puts a ring on personY's finger.**

How likely is this sentence going to happen? (Invalid if it does not make sense to you)

- ◯ Always / Usually
- ◯ Typical: Often / Probable
- ◯ Atypical: Farfetched / Never
- ◯ Invalid: Can't understand the meaning

Annotation Results

- Based on the positive abstract events, total 1,149,118 possible abstract triples are collected
- 2,756 abstract events are sampled for annotation

	Event	Triple
Total Questions	131,004	90,207
Questions with Agreement	92,235	81,197
Positives	40,833	65,900
Positive Rate	44.27%	81.16%
Inter-annotator Agreement	71.4%	73.0%
Manual Inspection Agreement	87.0%	86.5%

Inference Verification Modeling

- Similar to conceptualization verifier
- Adopt RoBERTa with following prompts

PersonX drinks coffee [EOS] [GEN] [xReact] refreshed [EOS]

Relation	Prompt
xNeed	Before that PersonX needs:
xIntent	PersonX's intention is:
xAttr	PersonX will be described as:
xEffect	Effects on PersonX will be:
xWant	After that PersonX wants:
xReact	After that PersonX feels:
oEffect	Effects on others will be:
oWant	After that others want:
oReact	After that others feel:

- Measured by AUC
- Other prompts and hyperparameters attempted as well

Inference Verification Modeling: Results

- Baselines
  - Negative Sampling
    - Positive examples from ATOMIC (same size as annotated data)
    - Positive examples from our annotated positives
    - Mixed
  - ATOMIC generator: COMET (GPT2-medium), 30.3 BLEU-2 on ATOMIC subset

AUC	Dev Set	Test Set
ATOMIC + Negative Sampling	0.62	0.65
Annotation	**0.72**	**0.74**
w/ Negative Sampling	0.67	0.67
Annotation + ATOMIC	**0.73**	**0.76**
w/ Negative Sampling	0.63	0.65
Generator based on ATOMIC	0.49	0.50

Another experiment on knowledge base population shows generative model is as effective as KB completion models.

Supervised Learning	KG-BERT (BERT-base) *110M*	62.5
	KG-BERT (BERT-large) *340M*	67.7
	KG-BERT (DeBERTa-base) *10M*	64.5
	KG-BERT (DeBERTa-large) *350M*	69.2
	KG-BERT (BART-base) *139M*	65.1
	KG-BERT (BART-large) *406M*	70.4
	KG-BERT (RoBERTa-base) *110M*	68.0
	KG-BERT (RoBERTa-large) *340M*	70.9
	COMET (GPT2-small) *117M*	69.6
	COMET (GPT2-medium) *345M*	69.7
	COMET (GPT2-large) *774M*	70.6
	COMET (GPT2-XL) *1558M*	70.7

Abstract ATOMIC

- Selection scores from neural models
  - Event conceptualization score > 0.8
  - Triple score > 0.9

- Heuristic concept linker produce much more diverse candidates but much less accurate

Numbers of selected data	0.7~0.8	0.8~0.9	≥0.9
Event Conceptualization (by Neural Concept Generator)	10.3K	17.7K	171.1K
Event Conceptualization (by Heuristic Concept Linker)	8.3K	11.5K	81.3K
Event Conceptualization (Total)	16.7K	26.2K	203.0K
Different Abstract Event	4.3K	7.0K	63.0K
Abstract Triple	542.2K	937.2K	2,947.9K

Human evaluation based on sampled data	0.7~0.8	0.8~0.9	≥0.9
Event Conceptualization (by Neural Concept Generator)	0.64	0.72	0.88
Event Conceptualization (by Heuristic Concept Linker)	0.67	0.74	0.90
Abstract Triple	0.41	0.55	0.71

ATOMIC's human score: 86.18%

Instantiation from Abstract Knowledge

Event	Instantiation	Relation	Positive Tails	Negative Tails
PersonX calls [health professional]	the doctor, the dentist	xWant	set an appointment, to ask the doctor a question, to tell the doctor their problems, …	to take their pet there, to ask a question
		xIntent	to schedule an appointment, to help pet, to be healthy, to feel better, …	to know about their pet, to be informed
		xNeed	dial the number, find the number, look up things online, to pick up the phone, …	to have a sick animal, to get doctor's phone number
[homecoming]	PersonX comes back, PersonX comes to PersonY’s house	oWant	to greet PersonX, to hug him, to help him relax, to eat out, to invite PersonX inside, …	to go eat, to have dinner, to talk to PersonX
		xIntent	see their family, to get home, to sleep, see their family, to attend some competition, …	have a break from learning, to attend the wedding
		xReact	cozy, happy, nostalgic, relaxed	drink, ready to eat, sleepy

Concept-aided Situational Commonsense Modeling

- A more abstract view may help the model to learn?
  - Augment ATOMIC with abstract knowledge
    - Especially with limited data
  - Use the ATOMIC subset that constitute the base of events in annotated triples
    - Mix with annotated or the corresponding automatically-built triples
    - Further finetune on ATOMIC

BLEU-2	GPT2-base	GPT2-medium
Baseline (COMET)	17.7	19.6
+Conceptualization (Human)	20.6	23.5
+Conceptualization (Auto)	19.3	21.0
+Conceptualization (Both)	19.0	22.9

Conclusion and Future Work

Thank you ☺

- A framework for machine conceptualization is formulated and implemented
  - A dataset for validity of conceptualization is annotated
  - Heuristic rules and neural models to generate and verify conceptualization are developed
  - A large scale abstract CKG is inferred
    - 70K abstract events and 2.9M abstract triples
- Future work
  - Better models
  - More downstream tasks
  - Integrating more data, e.g., ATOMIC-10X

https://github.com/HKUST-KnowComp/atomic-conceptualization