Knowledge Bases

Lara J. Martin (she/they)

https://laramartin.net/neurosymbolic-text-gen/

Learning Objectives

- Find out about existing popular commonsense knowledge bases
- Distinguish when you would use one knowledge base over another
- Connect knowledge graphs to planning

Logistics

Homework 3 & Project Milestone 2 are due today!

- Also the submission for the extra credit if your team is doing that

I have office hours today, but will be virtually joining a meeting at 4pm

- (I'd prefer it if you came earlier than that)

Dr. Manas Gaur will guest lecture on Thursday 10/29

Review: Definition of Common Sense

The basic level of **practical knowledge** and **reasoning**

concerning **everyday situations and events**

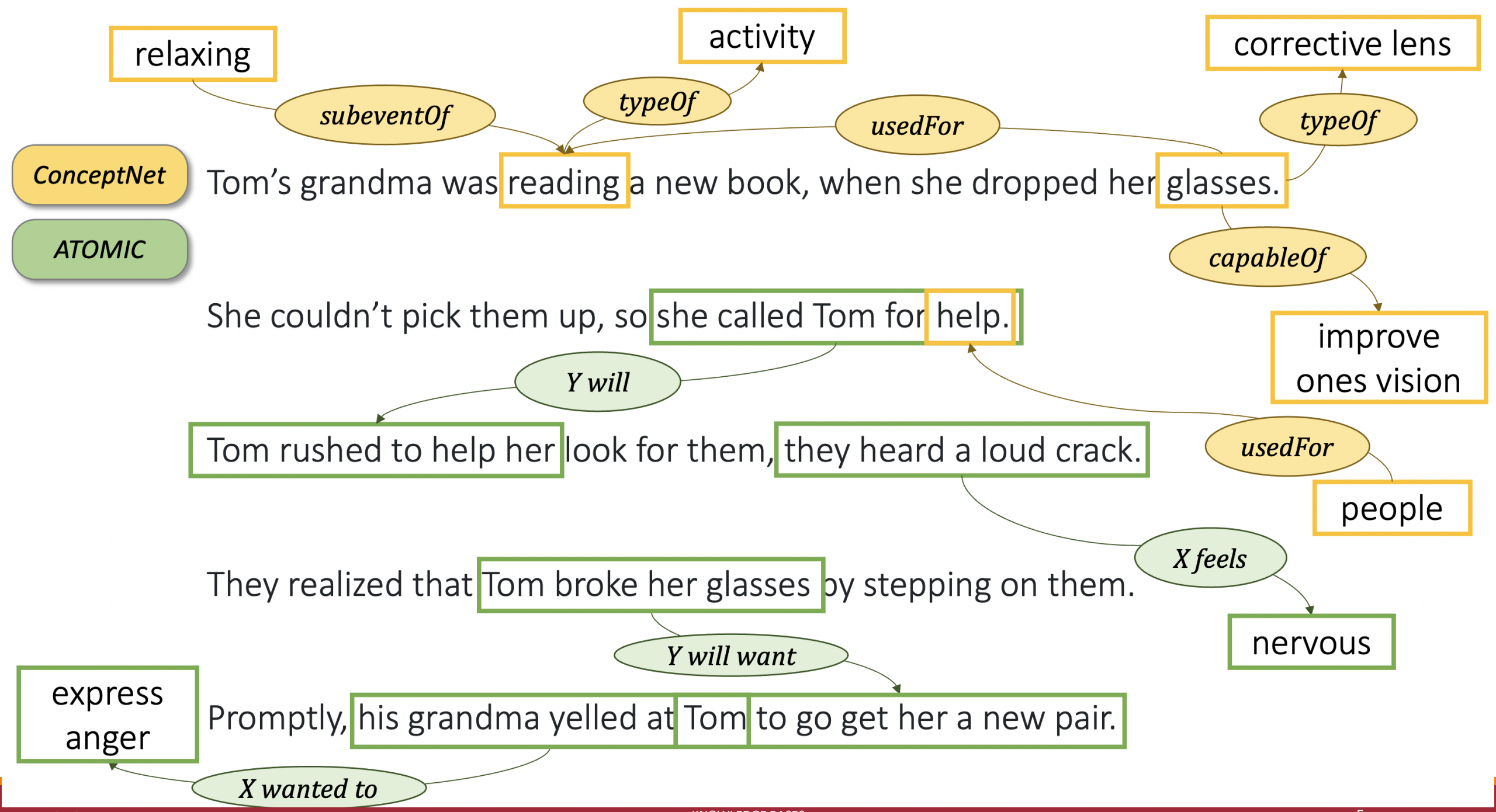
that are **commonly shared** among most people.

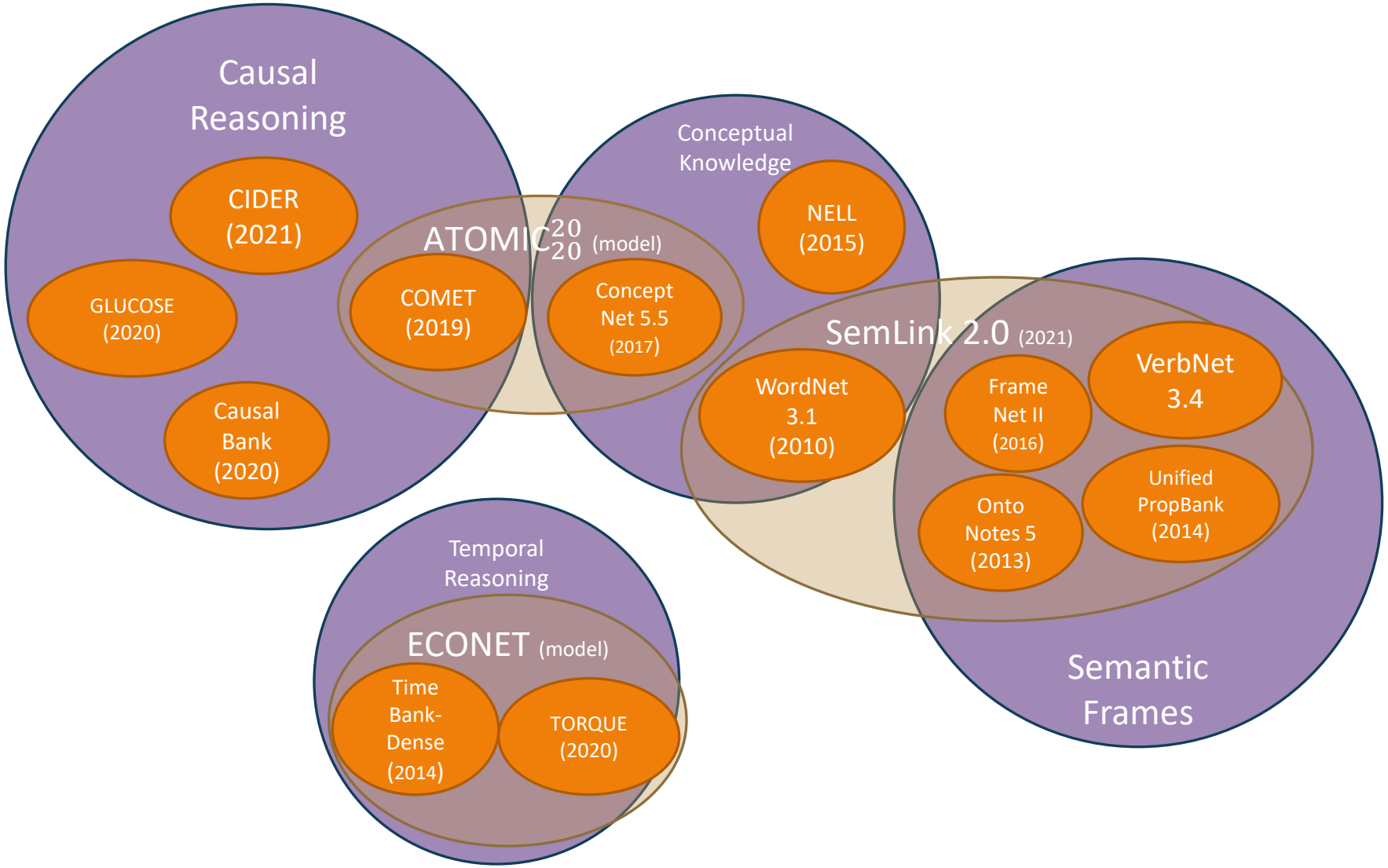
It's OK to keep the closet door open

It's not OK to keep the refrigerator door open because the food might go bad

Essential for humans to live and interact with each other in a reasonable and safe way

Essential for AI to understand human needs and actions better


relaxing
activity
corrective lens
subeventOf
typeOf
usedFor
typeOf
ConceptNet
Tom's grandma was reading a new book, when she dropped her glasses.
ATOMIC
capableOf
She couldn't pick them up, so she called Tom for help.
improve ones vision
Y will
Tom rushed to help her look for them, they heard a loud crack.
usedFor
people
X feels
They realized that Tom broke her glasses by stepping on them.
nervous
Y will want
express anger
Promptly, his grandma yelled at Tom to go get her a new pair.
X wanted to


Causal Reasoning
CIDER (2021)
GLUCOSE (2020)
Causal Bank (2020)
ATOMIC2020 (model)
COMET (2019)
Conceptual Knowledge
Concept Net 5.5 (2017)
NELL (2015)
SemLink 2.0 (2021)
WordNet 3.1 (2010)
Frame Net II (2016)
VerbNet 3.4
Onto Notes 5 (2013)
Unified PropBank (2014)
Semantic Frames
Temporal Reasoning
ECONET (model)
Time Bank-Dense (2014)
TORQUE (2020)

How do you create a commonsense resource?

Desirable properties for a commonsense resource

COVERAGE

Large scale

Diverse knowledge types

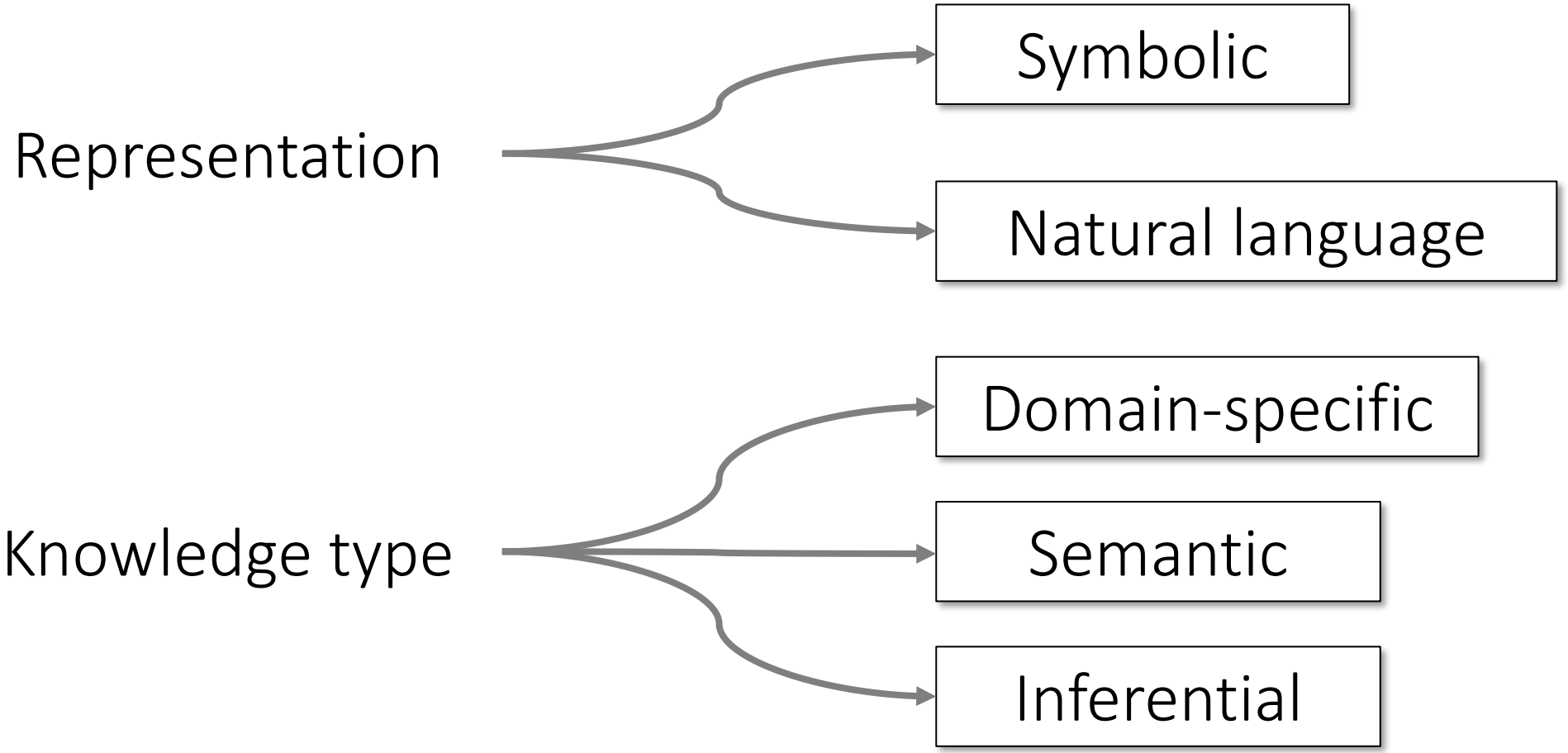
USEFUL

High quality knowledge

Usable in downstream tasks

Multiple resources tackle different knowledge types

Creating a commonsense resource



What is a semantic frame?

"people understand the meaning of words largely by virtue of the frames which they evoke"

- Understanding words in context
- Based on recurring experiences

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L Petruck, Christopher R. Johnson, Collin F. Baker, & Jan Scheffczyk. *FrameNet II: Extended Theory and Practice* (Revised November 1, 2016.)

Fillmore, Charles J. (1982). "Frame semantics". In The Linguistic Society of Korea, eds. *Linguistics in the Morning Calm*. Seoul: Hanshin. 111-37.

CONCEPTNET:

semantic knowledge in *natural language* form

http://conceptnet.io/

Related terms

- en book →
- en books →
- en book →

Effects of reading

- en learning →
- en ideas →
- en a headache →

reading is a type of...

- en an activity →
- en a good way to learn →
- en one way of learning →
- en one way to learn →

reading is a subevent of...

- en you learn →
- en turning a page →
- en learning →

en reading

An English term in ConceptNet 5.8

Subevents of reading

- en relaxing →
- en study →
- en studying for a subject →

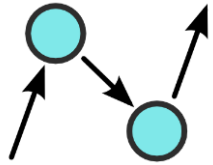
Things used for reading

- en article →
- en a library →
- en literature →
- en a paper page →

Types of reading

- en browse (n, communication) →
- en bumf (n, communication) →
- en clock time (n, time) →
- en miles per hour (n, time) →

What is ConceptNet?



General commonsense knowledge

21 million edges and over 8 million nodes (as of 2017)

- Over 85 languages
- In English: over 1.5 million nodes

Knowledge covered:

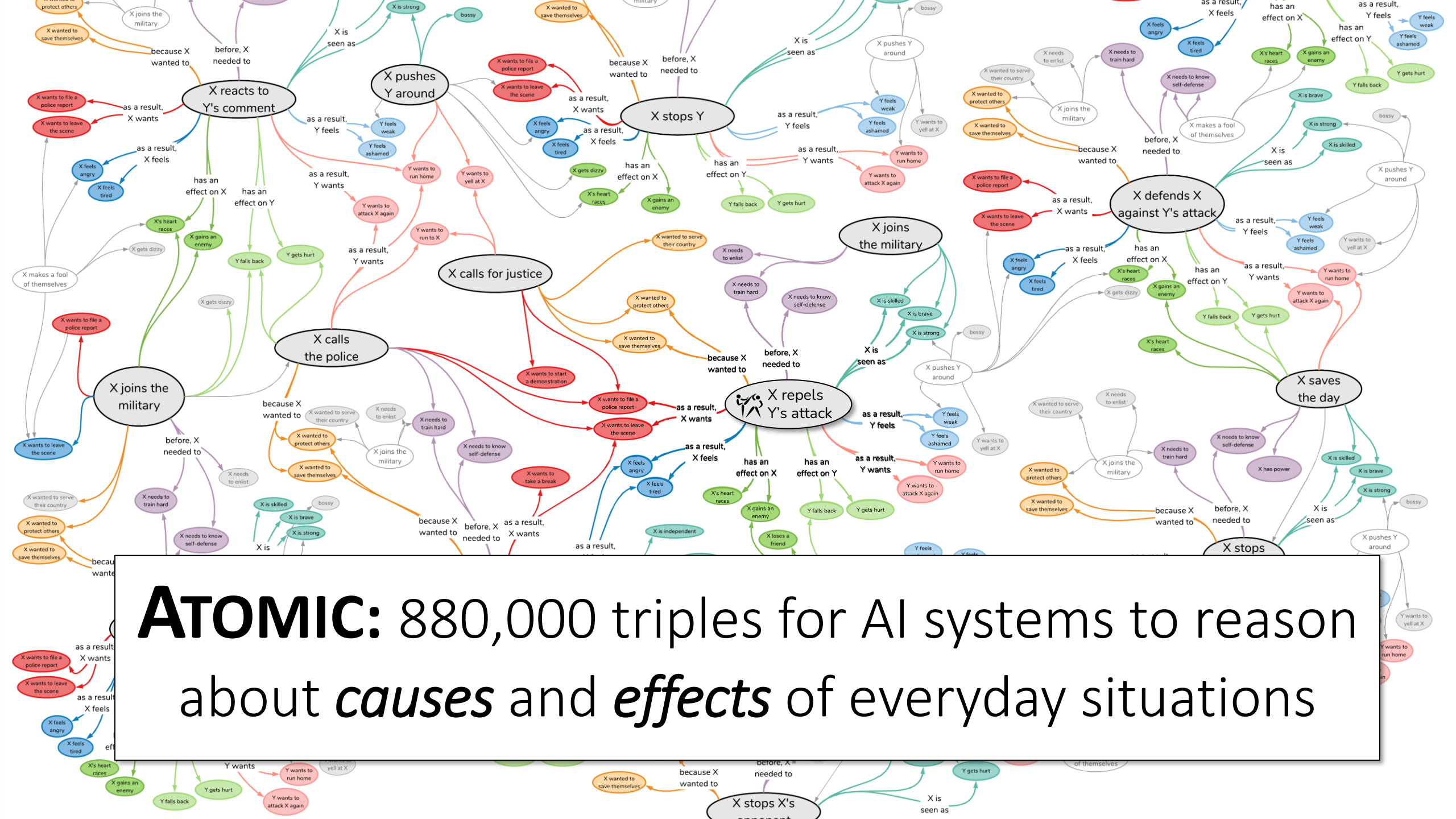
- Open Mind Commonsense assertions
- Wikipedia/Wiktionary semantic knowledge
- WordNet, Cyc ontological knowledge

http://conceptnet.io/

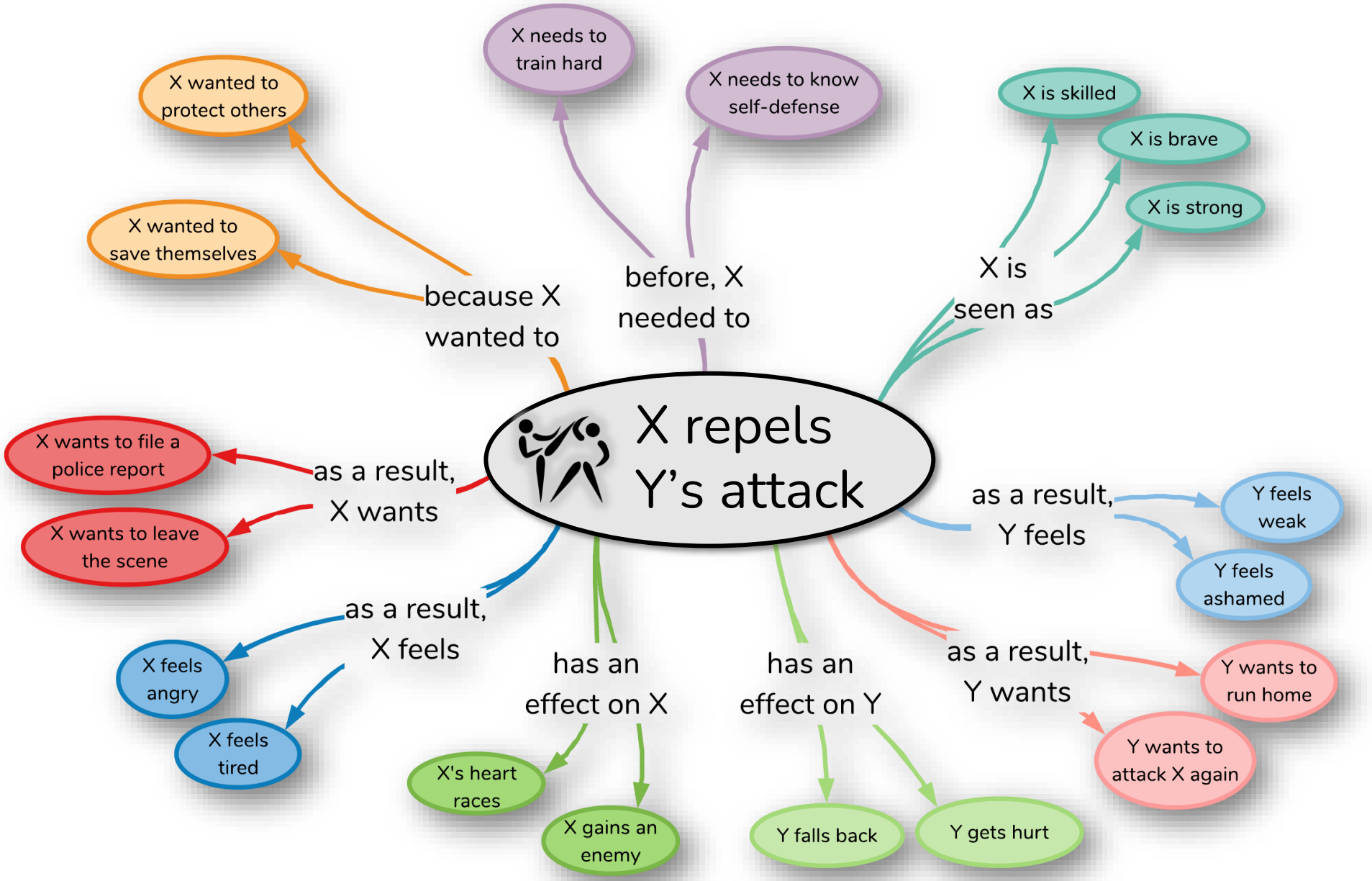
ATOMIC:

inferential knowledge in *natural language* form

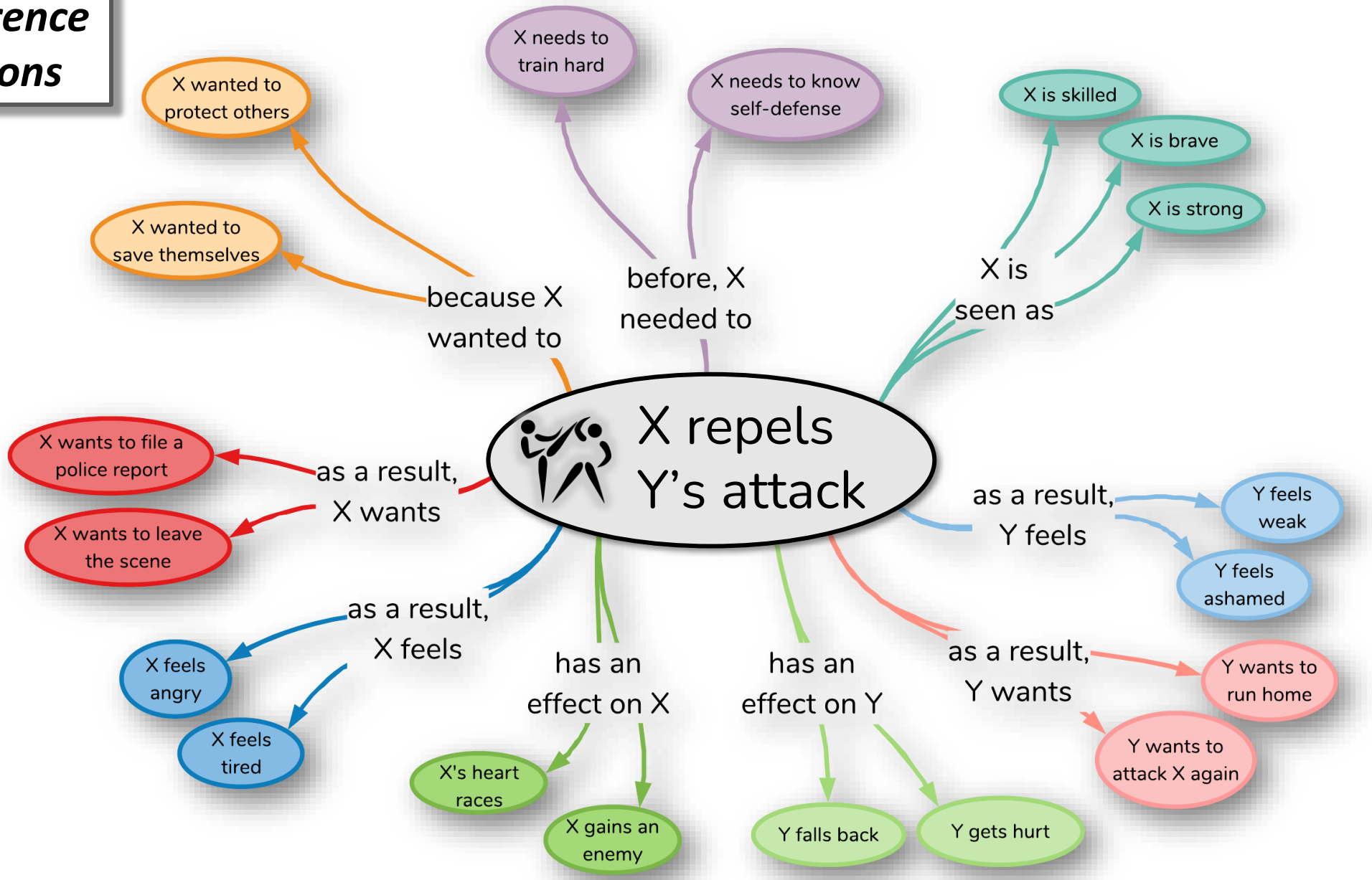
https://github.com/allenai/comet-atomic-2020

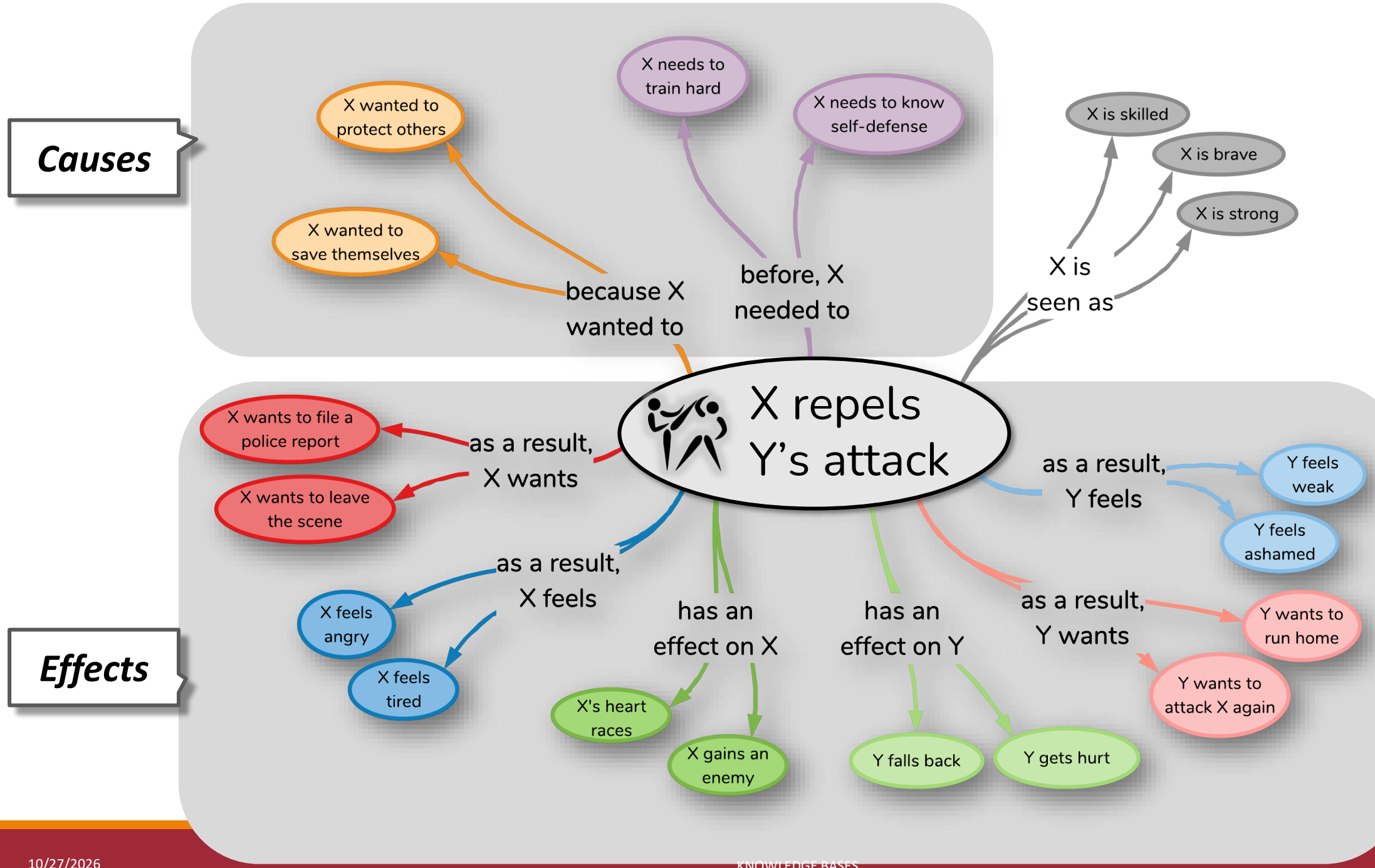


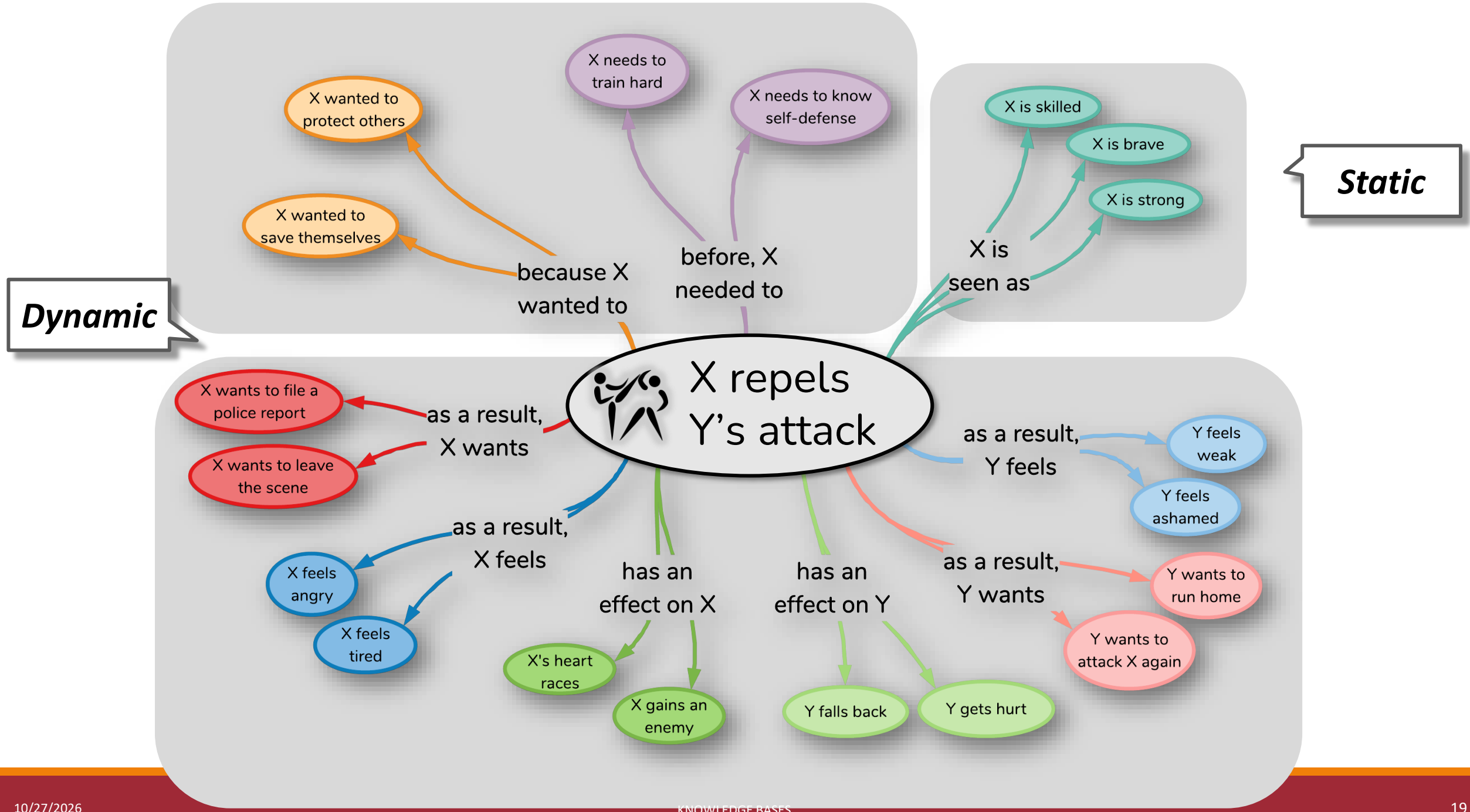
ATOMIC: 880,000 triples for AI systems to reason about ***causes*** and ***effects*** of everyday situations

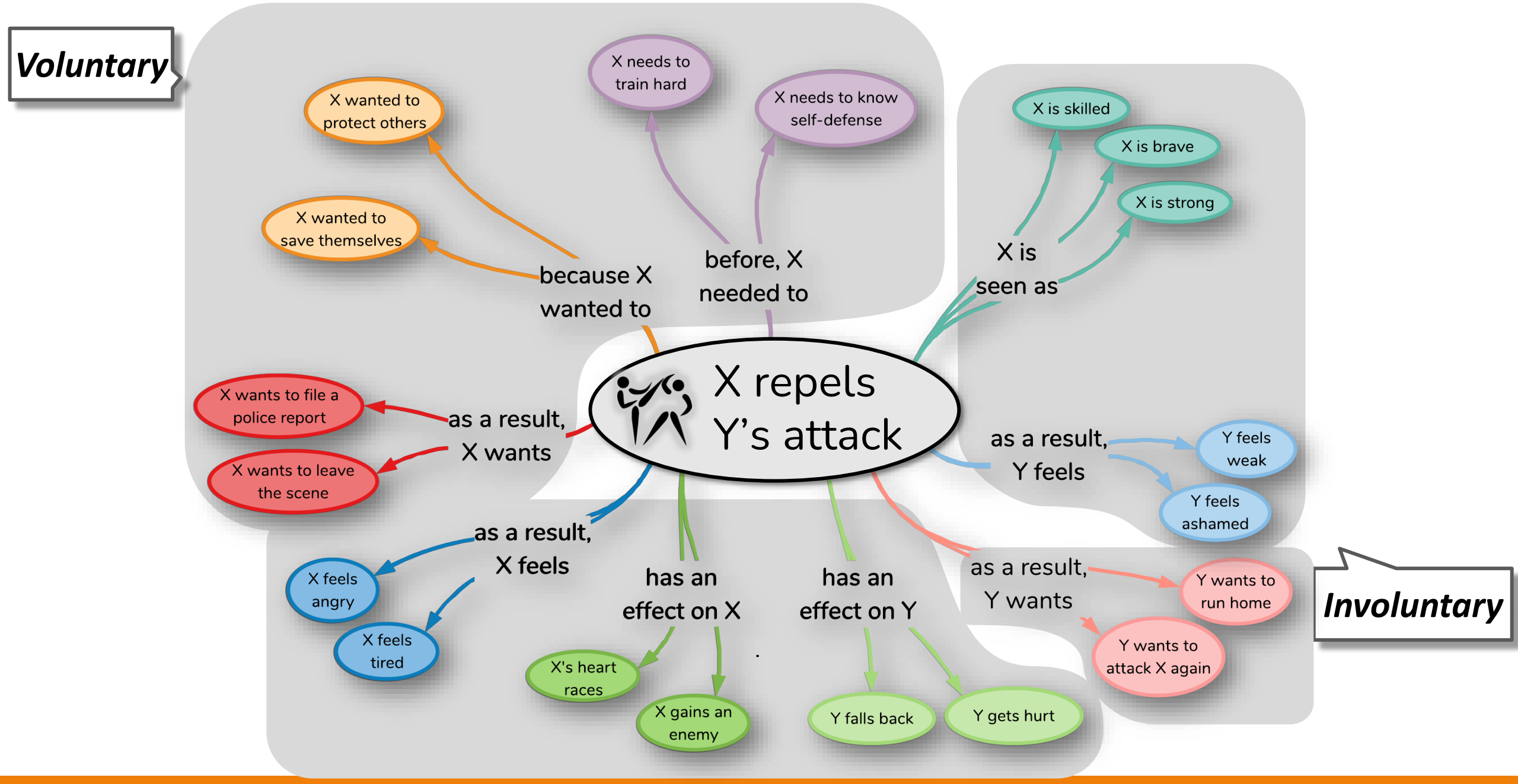

X repels Y's attack
because X wanted to
X wanted to protect others
X wanted to save themselves
before, X needed to
X needs to train hard
X needs to know self-defense
X is seen as
X is skilled
X is brave
X is strong
as a result, X wants
X wants to file a police report
X wants to leave the scene
as a result, X feels
X feels angry
X feels tired
has an effect on X
X's heart races
X gains an enemy
has an effect on Y
Y falls back
Y gets hurt
as a result, Y wants
Y wants to run home
Y wants to attack X again
as a result, Y feels
Y feels weak
Y feels ashamed

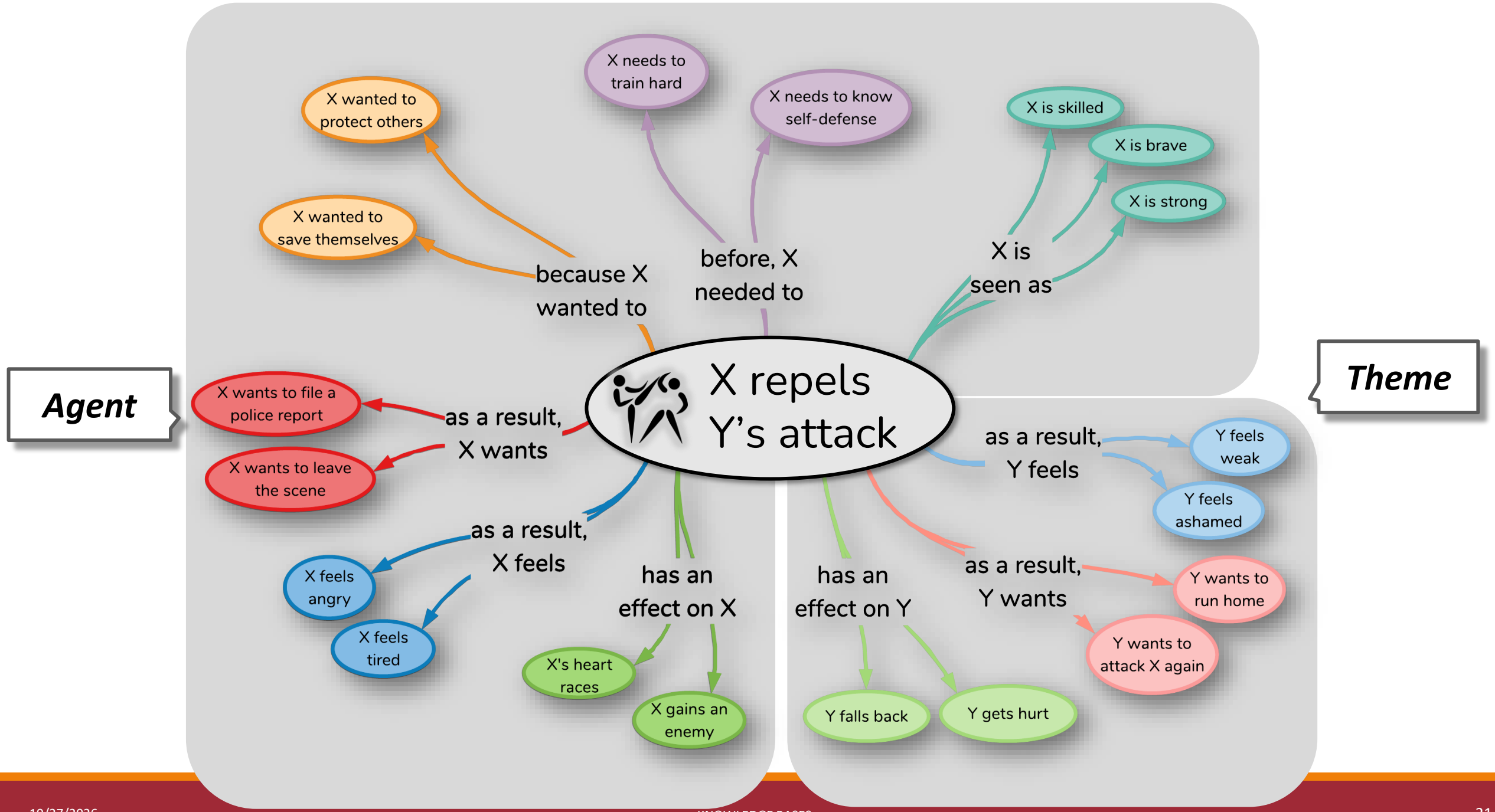
nine inference dimensions

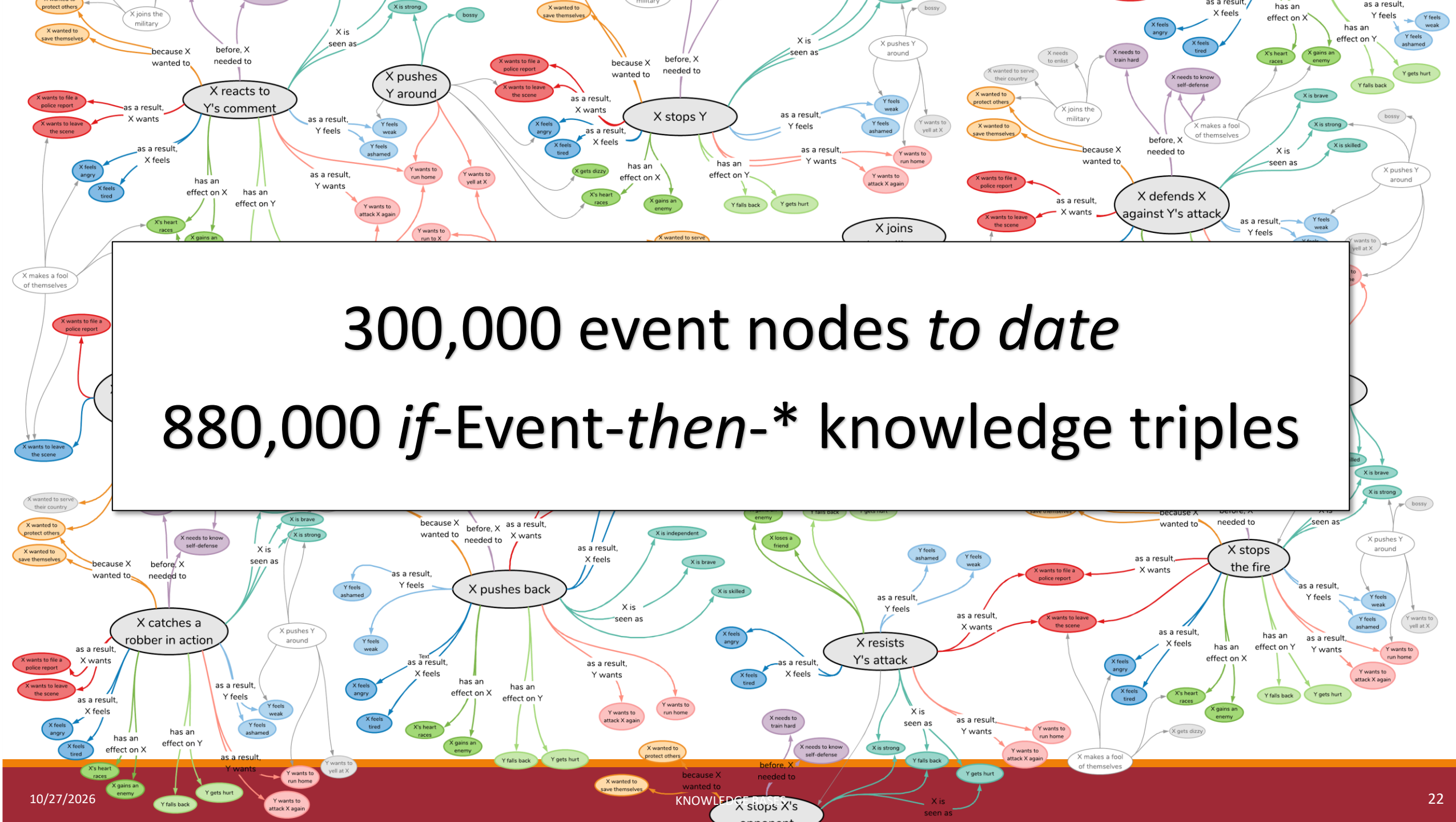



Causes
X wanted to protect others
X needs to train hard
X needs to know self-defense
X wanted to save themselves
because X wanted to
before, X needed to
X is skilled
X is brave
X is strong
X is seen as
X repels Y's attack
X wants to file a police report
X wants to leave the scene
as a result, X wants
as a result, X feels
X feels angry
X feels tired
has an effect on X
X's heart races
X gains an enemy
has an effect on Y
Y falls back
Y gets hurt
as a result, Y feels
Y feels weak
Y feels ashamed
as a result, Y wants
Y wants to run home
Y wants to attack X again
Effects


Dynamic
Static
X repels Y's attack
because X wanted to
X wanted to protect others
X wanted to save themselves
before, X needed to
X needs to train hard
X needs to know self-defense
X is seen as
X is skilled
X is brave
X is strong
as a result, X wants
X wants to file a police report
X wants to leave the scene
as a result, X feels
X feels angry
X feels tired
has an effect on X
X's heart races
X gains an enemy
has an effect on Y
Y falls back
Y gets hurt
as a result, Y wants
Y wants to run home
Y wants to attack X again
as a result, Y feels
Y feels weak
Y feels ashamed


Voluntary
X needs to train hard
X wanted to protect others
X needs to know self-defense
X wanted to save themselves
because X wanted to
before, X needed to
X is skilled
X is brave
X is strong
X is seen as
X repels Y's attack
X wants to file a police report
as a result, X wants
X wants to leave the scene
as a result, Y feels
Y feels weak
Y feels ashamed
as a result, X feels
X feels angry
X feels tired
has an effect on X
has an effect on Y
as a result, Y wants
Y wants to run home
Y wants to attack X again
X's heart races
X gains an enemy
Y falls back
Y gets hurt
Involuntary


Agent
Theme
X repels Y's attack
because X wanted to
X wanted to protect others
X wanted to save themselves
before, X needed to
X needs to train hard
X needs to know self-defense
X is seen as
X is skilled
X is brave
X is strong
as a result, X wants
X wants to file a police report
X wants to leave the scene
as a result, X feels
X feels angry
X feels tired
has an effect on X
X's heart races
X gains an enemy
as a result, Y feels
Y feels weak
Y feels ashamed
has an effect on Y
Y falls back
Y gets hurt
as a result, Y wants
Y wants to run home
Y wants to attack X again


300,000 event nodes *to date*
880,000 *if*-Event-*then*-* knowledge triples

ATOMIC: knowledge of *cause* and *effect*

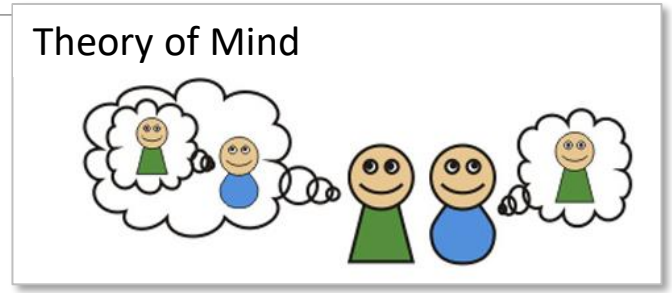
Humans have **theory of mind**, allowing us to

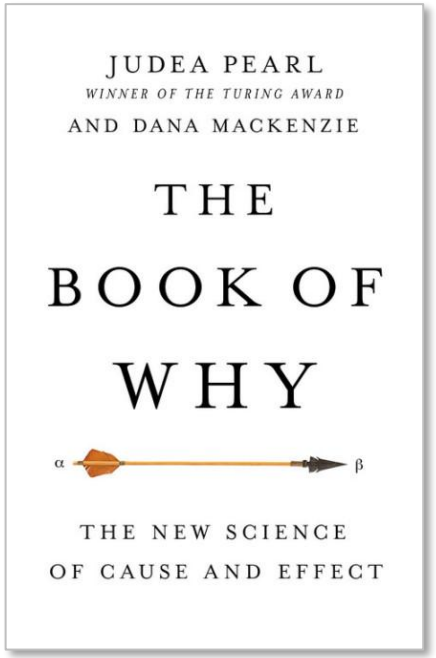
- make inferences about **people's mental states**
- understand **likely events** that precede and follow (Moore, 2013)

AI systems struggle with ***inferential*** reasoning

- only find **complex correlational patterns** in data
- **limited to the domain** they are trained on

(Pearl; Davis and Marcus 2015; Lake et al. 2017; Marcus 2018)





Ways of categorizing existing knowledge bases

ATOMIC
(Sap et al., 2019)

NELL
(Mitchell et al., 2015)

ConceptNet 5.5
(Speer et al., 2017)

OpenCyc 4.0
(Lenat, 2012)

Ways of categorizing existing knowledge bases

Represented in **symbolic logic**
(e.g., LISP-style logic)

Represented in **natural language**
(how humans *talk* and *think*)

NELL
(Mitchell et al., 2015)

OpenCyc 4.0
(Lenat, 2012)

ConceptNet 5.5
(Speer et al., 2017)

ATOMIC
(Sap et al., 2019)

```
(#$implies
  (#$and
    (#$isa ?OBJ ?SUBSET)
    (#$genls ?SUBSET ?SUPERSET))
  (#$isa ?OBJ ?SUPERSET))
```

Ways of categorizing existing knowledge bases

Represented in **symbolic logic**
(e.g., LISP-style logic)

Represented in **natural language**
(how humans *talk* and *think*)

NELL
(Mitchell et al., 2015)

OpenCyc 4.0
(Lenat, 2012)

ConceptNet 5.5
(Speer et al., 2017)

When would you want to use a taxonomic knowledge base?

How about an inferential?

Knowledge of "***what***"
(taxonomic: `A isA B`)

Knowledge of "***why***" and "***how***"
(inferential: *causes* and *effects*)

ATOMIC
(Sap et al., 2019)

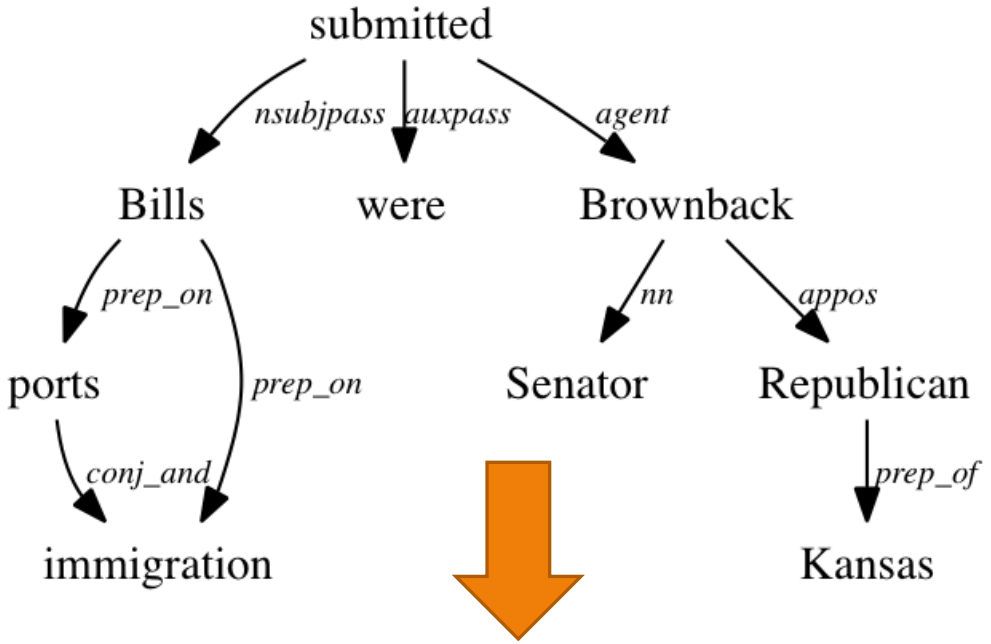
Q: How do you gather commonsense knowledge at scale?

A: It depends on the type of knowledge

Extracting commonsense from text

Based on information extraction (IE) methods

1. Read and parse text
2. Create candidate rules
3. Filter rules based on quality metric




```
isA(senator,Brownback)
location(Kansas,Brownback)
~~isA(senator,Kansas~~)
...
```

Advantage:
can extract knowledge automatically

Example system:
Never Ending Language Learner (NELL; Carlson et al., 2010)

Some commonsense cannot be extracted

Text is subject to **reporting bias**
(Gordon & Van Durme, 2013)

Noteworthy events

- Murdering 4x more common than exhaling

Commonsense is not often written

- Grice's maxim of quantity



found when extracting commonsense knowledge on four large corpora using Knext (Gordon & Van Durme, 2013)

When communicating, people try to be as informative as they possibly can, and give as much information as is needed, and no more.

Eliciting commonsense from humans

EXPERTS CREATE KNOWLEDGE BASE

NON-EXPERTS WRITE KNOWLEDGE IN NATURAL LANGUAGE PHRASES

What is an “expert” in this case?

OpenCyc 4.0 (Lenat, 2012)

WordNet (Miller et al., 1990)

ATOMIC (Sap et al., 2019)

ConceptNet 5.5 (Speer et al., 2017)

Eliciting commonsense from humans

EXPERTS CREATE KNOWLEDGE BASE

Advantages:

- Fewer errors
- Capture nuances → Potentially more useable?
- Structured
- More specific

Drawbacks:

- Less data
- More expensive (time/money) → harder to update
- Not “common” sense/ not aligned to real world

NON-EXPERTS WRITE KNOWLEDGE IN NATURAL LANGUAGE PHRASES

Natural language

- Accessible to non-experts
- Different phrasings allow for more diverse knowledge

Fast and scalable collection

- Crowdsourcing
- Games with a purpose

Knowledge bases and mitigating biases

PersonX clutches a gun → because X wanted to

- to be safe
- to protect himself
- to protect themselves
- to defend themselves
- to defend himself

ATOMIC (Sap et al., 2019)

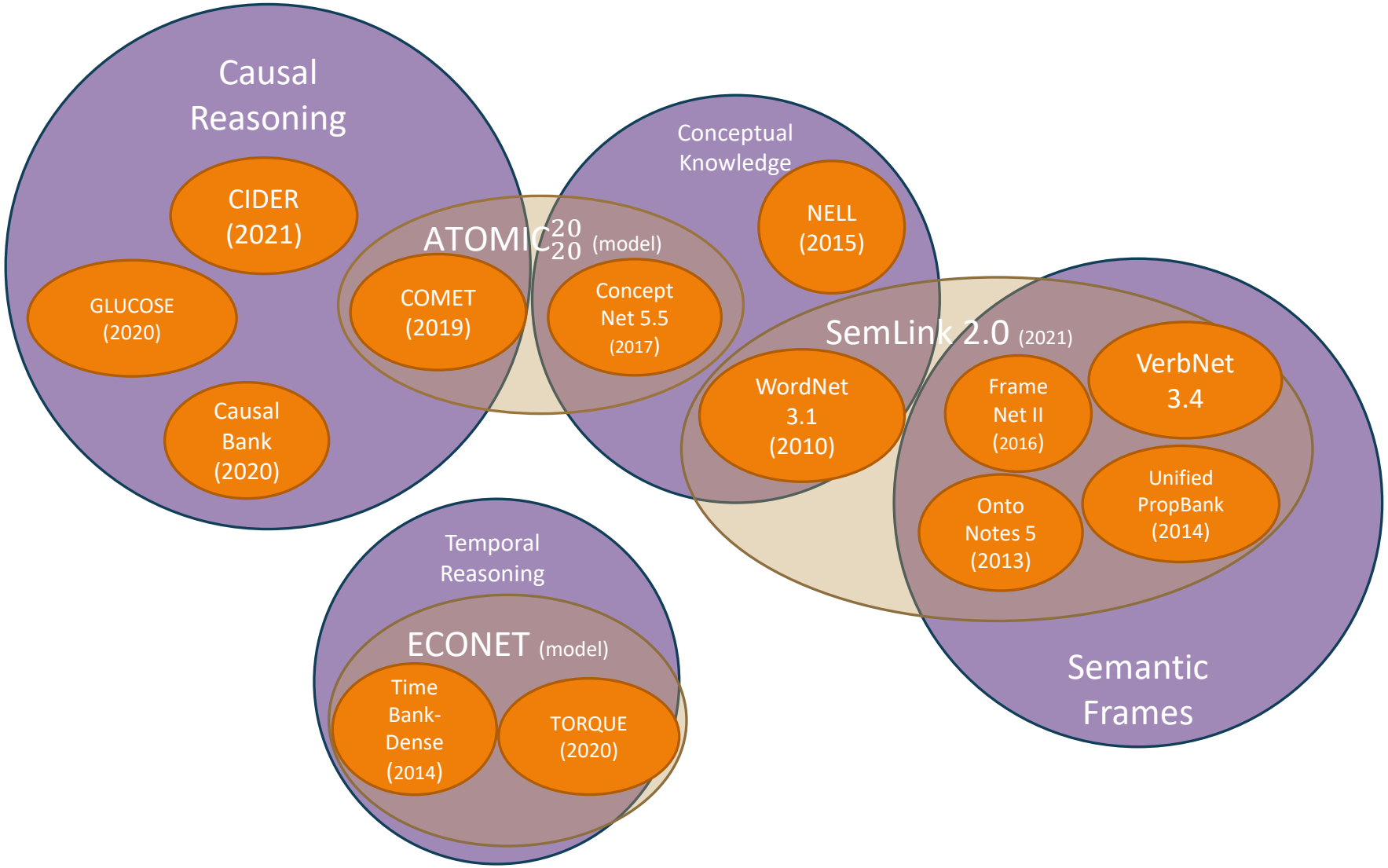
Jaquain clutches a gun → because X wanted to

- to kill someone
- none
- to protect himself
- to be safe
- to protect themselves

Karen clutches a gun → because X wanted to

- to be safe
- to protect himself
- to shoot
- to get the gun
- none

COMET (Bosselut et al., 2019): ATOMIC + OpenAI GPT

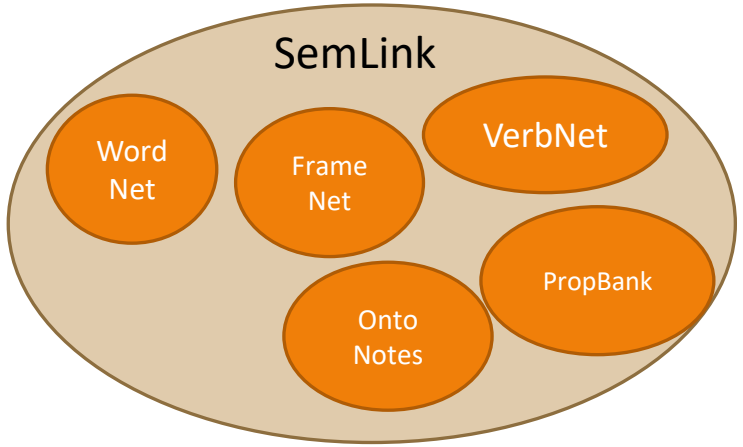

Causal Reasoning
CIDER (2021)
GLUCOSE (2020)
Causal Bank (2020)
ATOMIC2020 (model)
COMET (2019)
Conceptual Knowledge
Concept Net 5.5 (2017)
NELL (2015)
SemLink 2.0 (2021)
WordNet 3.1 (2010)
Frame Net II (2016)
VerbNet 3.4
Onto Notes 5 (2013)
Unified PropBank (2014)
Semantic Frames
Temporal Reasoning
ECONET (model)
Time Bank-Dense (2014)
TORQUE (2020)

SemLink/Unified Verb Index 2.0

https://github.com/cu-clear/semlink

Combines 5 systems:

VerbNet, PropBank, FrameNet, WordNet and OntoNotes

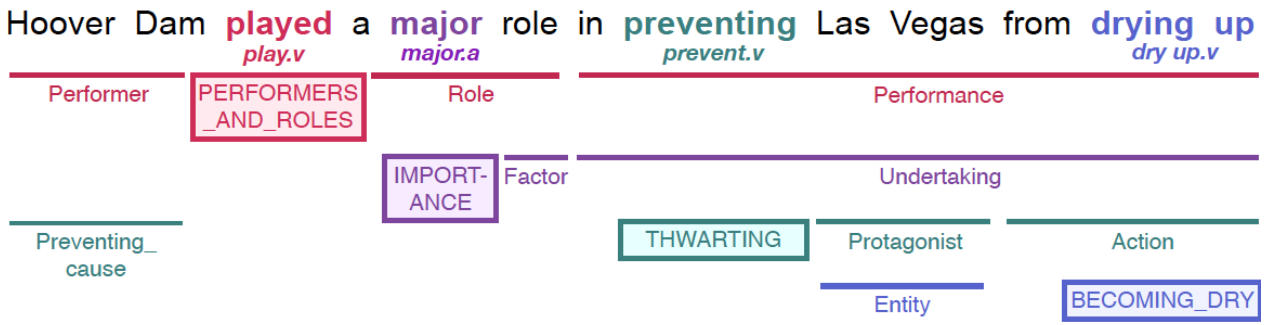


Use: above link

Kevin Stowe, Jenette Preciado, Kathryn Conger, Susan Windisch Brown, Ghazaleh Kazeminejad, James Gung, and Martha Palmer. 2021. SemLink 2.0: Chasing Lexical Resources. In *Proceedings of the 14th International Conference on Computational Semantics (IWCS)*, pages 222–227, Groningen, The Netherlands (online). Association for Computational Linguistics.

FrameNet II

https://framenet.icsi.berkeley.edu/



Data Source: British National Corpus, US newswire, American National Corpus; annotated

Languages: English, global initiative: https://www.globalframenet.org/

Use: Open-SESAME; Raw data needs to be requested

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L Petruck, Christopher R. Johnson, Collin F. Baker, & Jan Scheffczyk. *FrameNet II: Extended Theory and Practice* (Revised November 1, 2016.)

Picture from Open-SESAME (Swabha Swayamdipta, Sam Thomson, Chris Dyer, & Noah A. Smith. "Frame-Semantic Parsing with Softmax-Margin Segmental RNNs and a Syntactic Scaffold" on arXiv.

VerbNet v3.4

https://verbs.colorado.edu/verbnet/

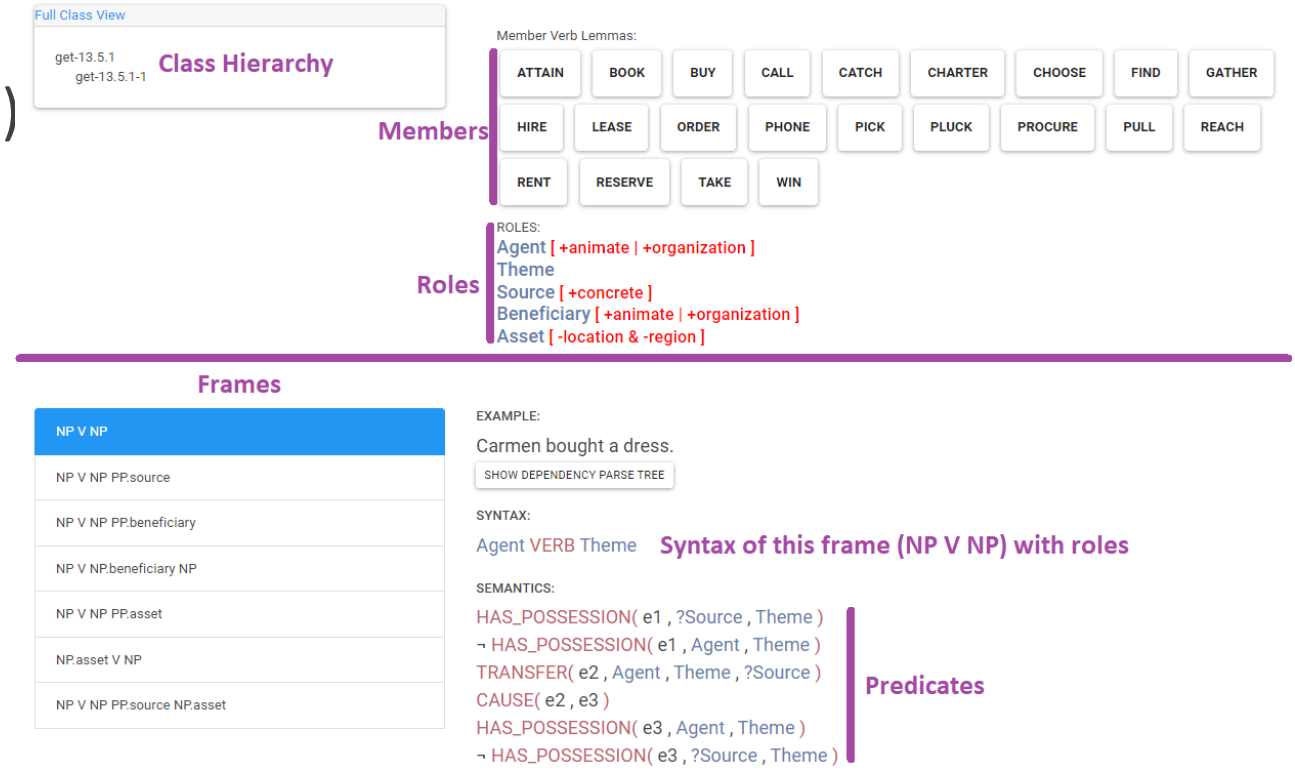
Verb classes based on Beth Levin (1993)

Data Source: hand-crafted

Languages: English

Use: raw data

Demo: https://uvi.colorado.edu/uvi_search



K. Kipper Schuler, "VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon," University of Pennsylvania, 2005.

Levin, B. (1993) "English Verb Classes and Alternations: A Preliminary Investigation", University of Chicago Press, Chicago, IL.

Unified* PropBank

http://propbank.github.io/

Proposition→ true/false statement

Data Source: hand-crafted; added to PennTreebank

Languages: English, Hindi, Chinese, Arabic, Finnish, Portuguese, Basque, Turkish (Plus a way to map English to different languages

Use: raw data

*semantic propositions regardless of part of speech (e.g. create & creation)

Event relation: Offer

25. **Predicate**: *offer*-verb
Roleset id: offer.01 transaction
Roles: Arg0: entity offering
Arg1: commodity
Arg2: price
Arg3: benefactive or entity offered to
Example: *He offered to buy the house.*

26. **Predicate**: *offer*-noun
Roleset id: offer.01 transaction
Roles: Arg0: entity offering
Arg1: commodity
Arg2: price
Arg3: benefactive or entity offered to
Example: *His offer to buy the house...*
He made an offer to buy the house.

27. **UNIFIED ROLESET**
Predicate aliases: *offer*-verb, *offer*-noun
Roleset id: offer.01 transaction
Roles: Arg0: entity offering
Arg1: commodity
Arg2: price
Arg3: benefactive or entity offered to
Example: *He offered to buy the house.*
His offer to buy the house..
He made an offer to buy the house.

```
(o / offer-01
  :ARG0 (h2 / he)
  :ARG1 (b2 / buy-01
          :ARG0 h2
          :ARG1 (h3 / house)))
```

Martha Palmer, Dan Gildea, Paul Kingsbury, The Proposition Bank: A Corpus Annotated with Semantic Roles *Computational Linguistics Journal*, *31:1*, 2005.

Claire Bonial, Julia Bonn, Kathryn Conger, Jena Hwang and Martha Palmer (2014) PropBank: Semantics of New Predicate Types. *The 9th edition of the Language Resources and Evaluation Conference*. Reykjavik, Iceland.

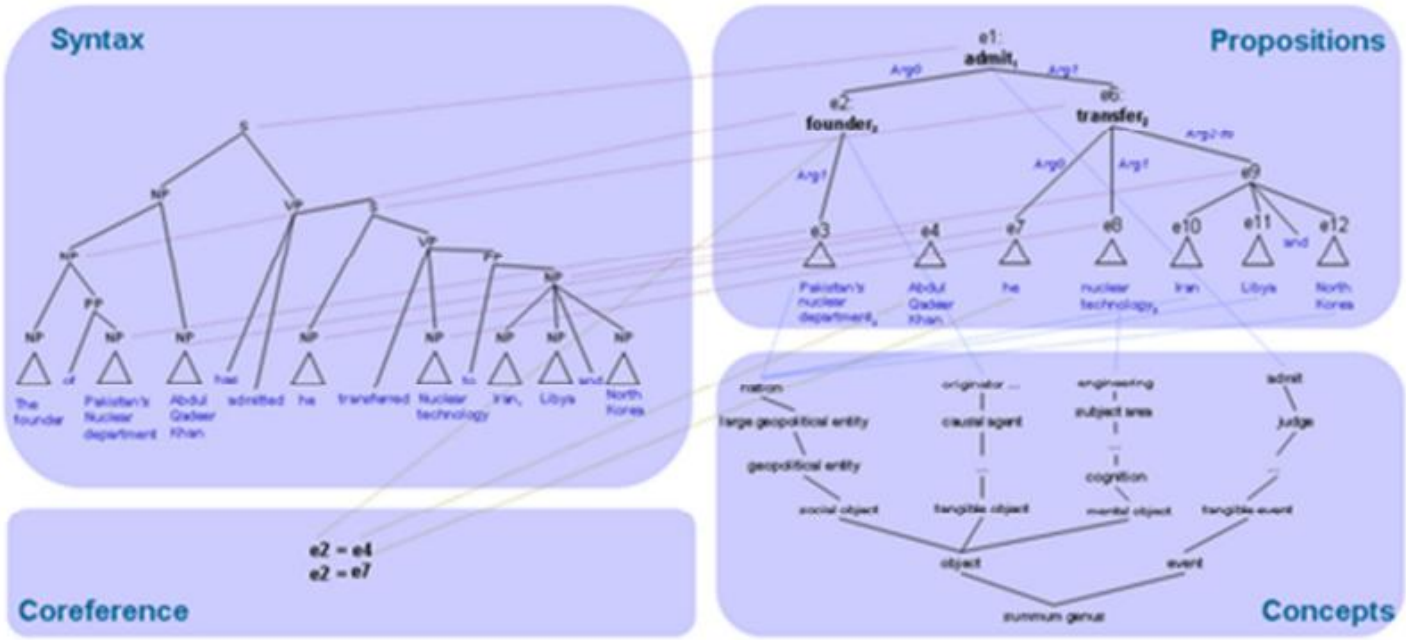
OntoNotes 5.0

https://catalog.ldc.upenn.edu/LDC2013T19

Data Source: news, telephone conversations, blogs, talk shows, etc.

Languages: English, Chinese, Arabic

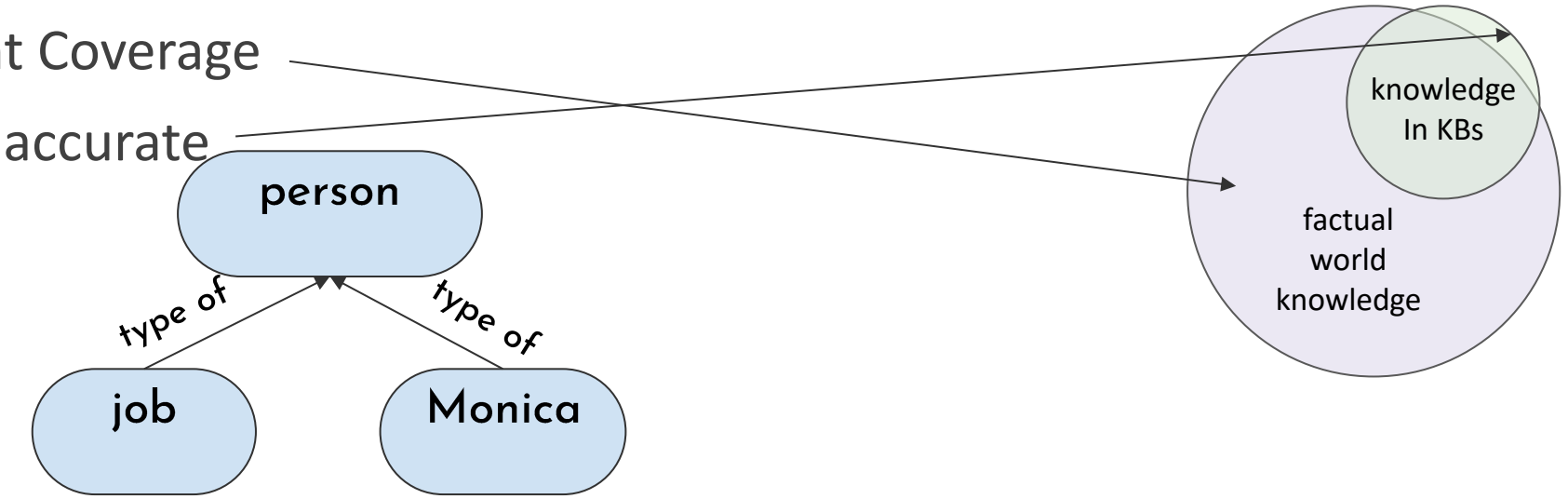
Use: raw data (same link)



S. S. Pradhan, E. Hovy, M. Marcus, M. Palmer, L. Ramshaw and R. Weischedel, "OntoNotes: A Unified Relational Semantic Representation," International Conference on Semantic Computing (ICSC 2007), 2007, pp. 517-526, doi: 10.1109/ICSC.2007.83.

Limitations

- Insufficient Coverage
- Not 100% accurate



- Easy to incorporate simple resources with stationary facts (ConceptNet) but they are limited in expressiveness:



Think-Pair-Share

How might knowledge bases be comparable to planning?

