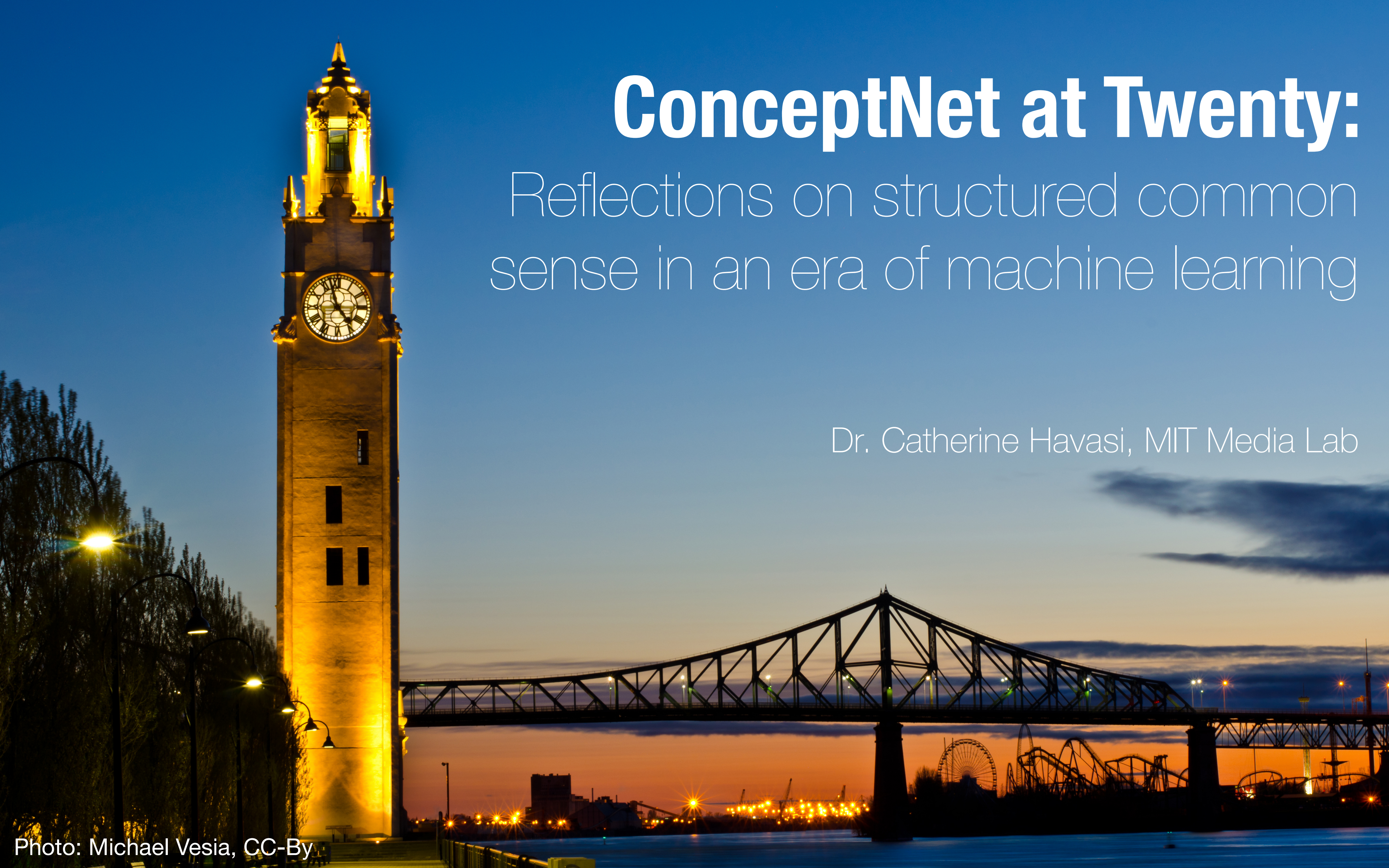
ConceptNet at Twenty:

Reflections on structured common sense in an era of machine learning

Dr. Catherine Havasi, MIT Media Lab



What's New · Check Email · YAHOO! · Personalize · Help

Yahoo! Mail
free email for life

Know when friends are online! Click to download Yahoo! Messenger

Yahoo! Auctions
coins, cards, stamps

Search · advanced search

Shopping - **Auctions** - Yellow Pages - People Search - Maps - Travel - Classifieds - Personals - Games - Chat - **Clubs**
Mail - Calendar - Messenger - **Companion** - My Yahoo! - News - Sports - Weather - TV - Stock Quotes - more...

Yahoo! Shopping - Thousands of stores. Millions of products.

Departments		Stores	Products
· Apparel	· Flowers	· Sports Authority	· Digital cameras
· Bath/Beauty	· Food/Drink	· Gap	· Pokemon
· Computers	· Music	· Eddie Bauer	· MP3 players
· Electronics	· Video/DVD	· Macy's	· DVD players

Arts & Humanities
Literature, Photography...

News & Media
Full Coverage, Newspapers, TV...

Business & Economy
Companies, Finance, Jobs...

Recreation & Sports
Sports, Travel, Autos, Outdoors...

Computers & Internet
Internet, WWW, Software, Games...

Reference
Libraries, Dictionaries, Quotations...

Education
College and University, K-12...

Regional
Countries, Regions, US States...

Entertainment
Cool Links, Movies, Humor, Music...

Science
Animals, Astronomy, Engineering...

In the News
- Mich. girl, 6, shot by classmate
- Bush, Gore win in primaries
- Israel releases Nazi Eichmann's memoirs

more...

Marketplace
- Free 56K Internet Access
- Y! Auctions - Peanuts, Pokemon, computers
- Yahoo! Bill Pay - free 3-month trial

more...

Inside Yahoo!
- Yahoo! GeoCities - build your free home page
- Play free Fantasy Soccer
- Yahoo! Clubs - create


Google
Search Stanford
10 results
clustering on
Search
Search The Web
10 results
clustering on
Search

MEANWHILE, BACK IN MASSACHUSETTS

Photo: Eric Baetscher

COLLECTING COMMON-SENSE KNOWLEDGE

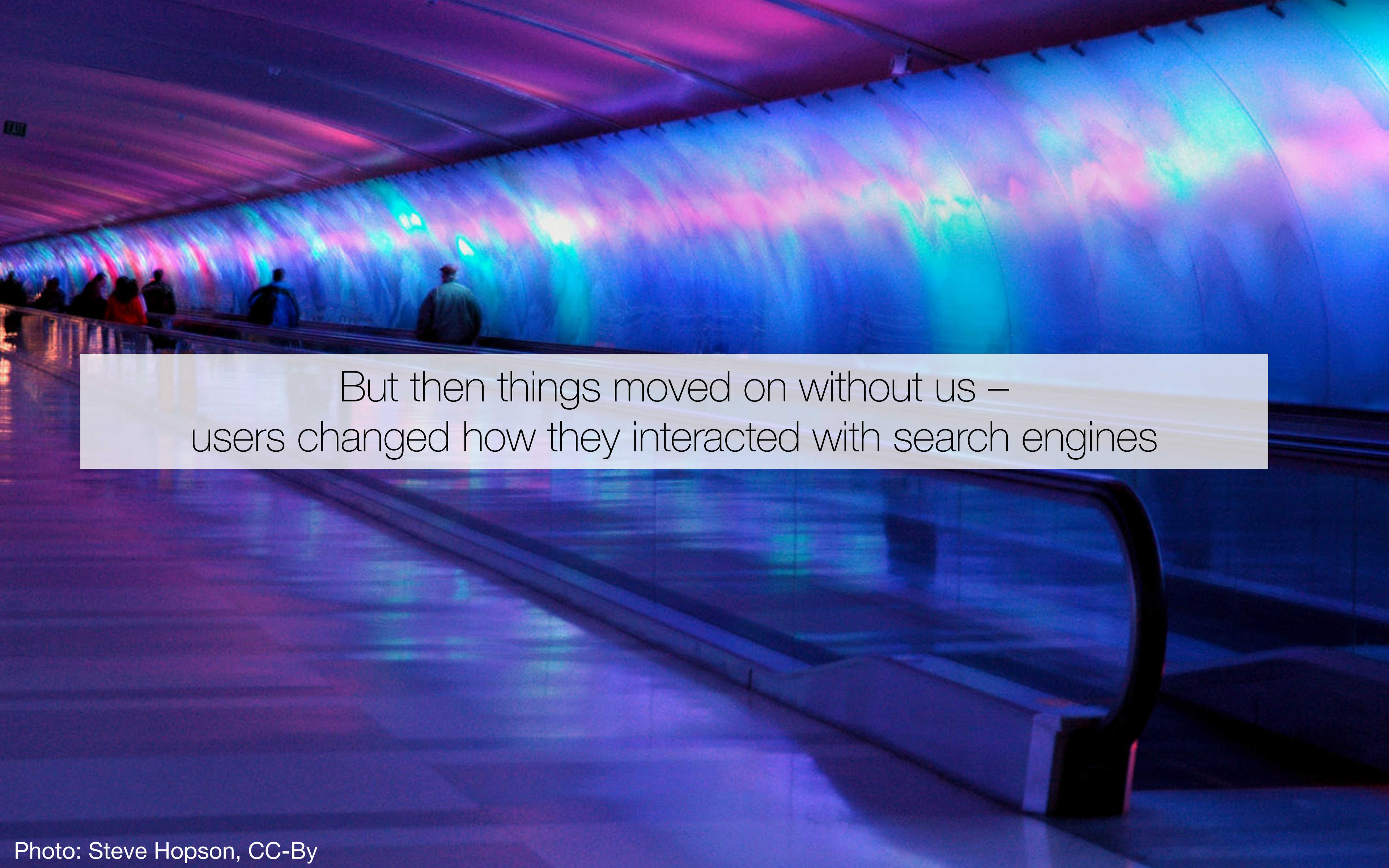
4	You are likely to find water in soup	by matrix29
4	Something you find at the supermarket is soup	by aww
3	You are likely to find chicken in soup.	by jbbdude
3	Something you find in a container is soup	by bedume
3	Something you find in a jar is soup	by smtango
3	soup is a type of hot liquid	by verbosity
2	You are likely to find a bean in a soup	by Fable
2	Soup is good food	by Visionsofkaos
2	You are likely to find a fungus in soup	by bedume
2	Soup is a light meal	by push
2	a bowl is for soup	by friscojohn
2	Billibi is a soup	by Ariadne
2	Something you find in a can is soup	by guru1

Open Mind wants to know...

Are these statements true?

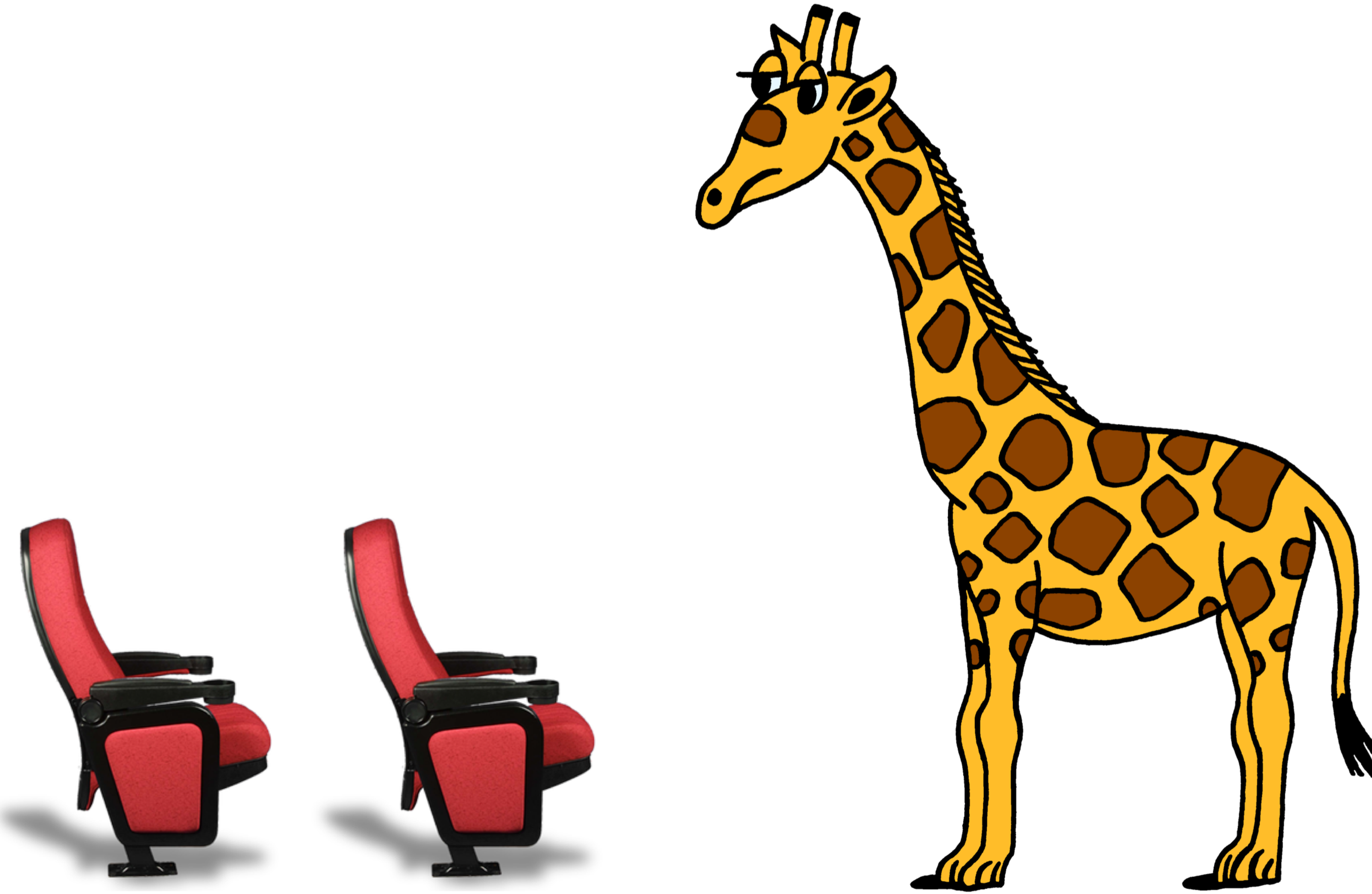
- You are likely to find **soup** in **a market**.
  Yes No Sort of
- You are likely to find **food** in **soup**.
  Yes No Sort of
- You are likely to find **soup** in **the cabinet**.
  Yes No Sort of
- **soup** can be **eaten**
  Yes No Sort of
- You are likely to find **soup** in **the freezer**.
  Yes No Sort of

Robyn Speer, Joshua Chin, and Catherine Havasi: "ConceptNet 5.5: An Open Multilingual Graph of General Knowledge." (2017)



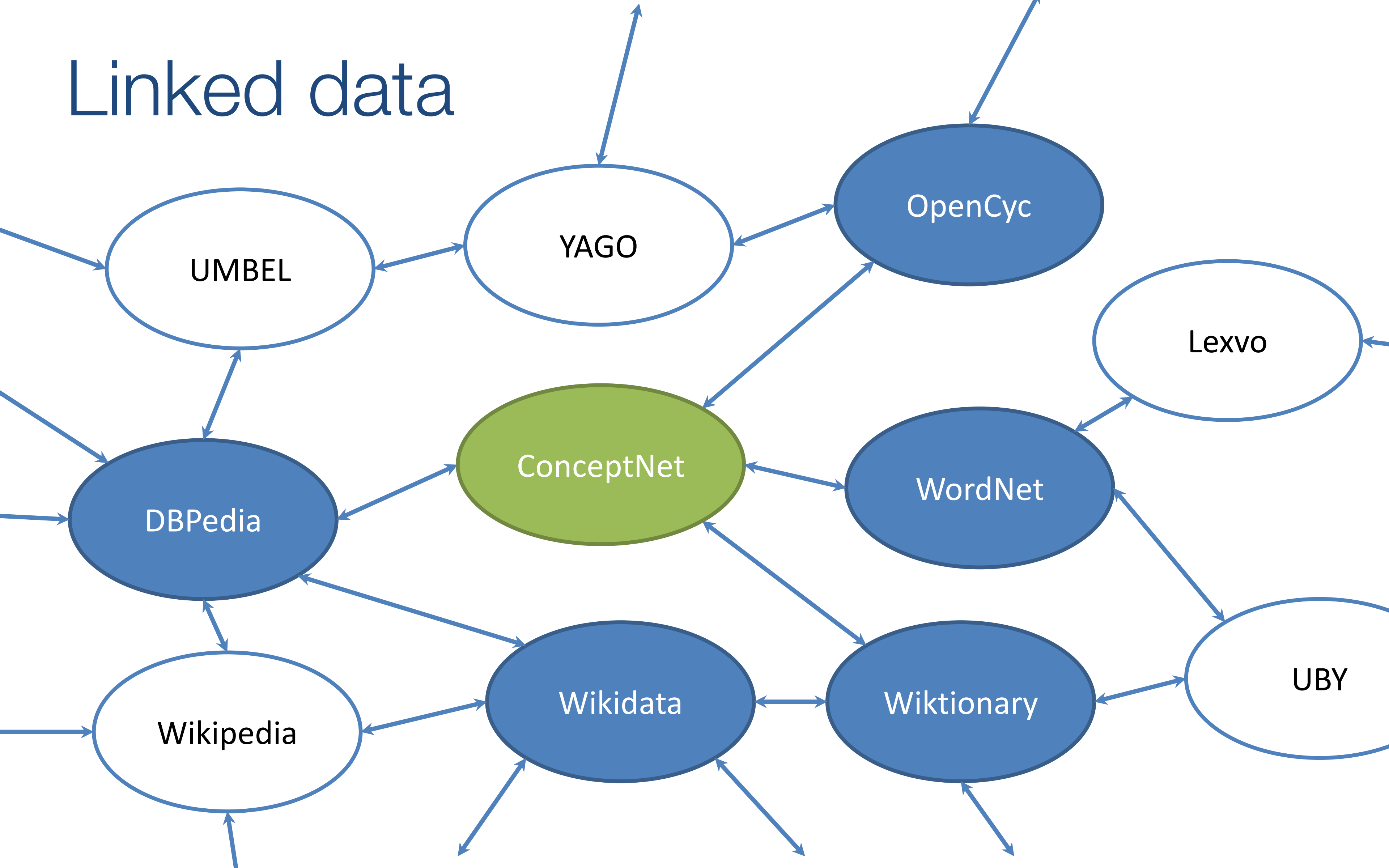
But then things moved on without us –
– users changed how they interacted with search engines

How do people do it?



We have built models of how people think about the world in 73 languages – called ConceptNet.

Linked data

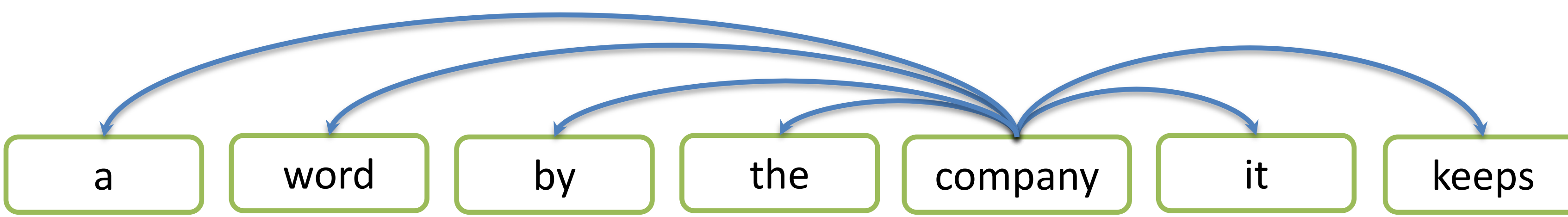

UMBEL
YAGO
OpenCyc
Lexvo
ConceptNet
WordNet
DBPedia
Wikidata
Wiktionary
UBY
Wikipedia

“I don’t have to actually experience crashing my car into a wall a few hundred times before I slowly start avoiding to do so.”

- Andrej Karpathy, Open AI

Distributional semantics

- “You shall know a word by the company it keeps.” – J. R. Firth
- *Skipgrams* are one way to train a neural net to understand words by their contexts



Source: *Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean*: Distributed representations of words and phrases and their compositionality. (2013)

Retrofitting

- A way of connecting knowledge graphs to traditionally created word embeddings

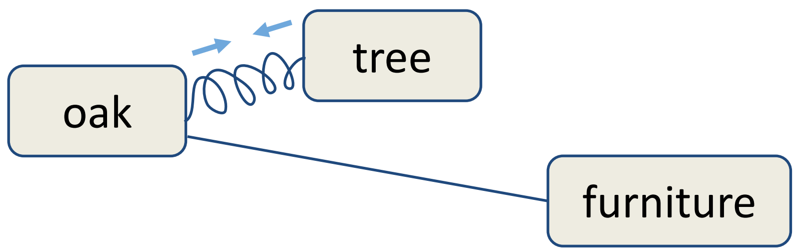
- Apply knowledge-based constraints *after* training distributional word vectors

- It works better than *during* training, for some reason

Source: *Manaal Faruqui, Jesse Dodge, Sujay K. Jauhar, Chris Dyer, Eduard Hovy, Noah A. Smith*: Retrofitting Word Vectors to Semantic Lexicons (2015)

Retrofitting

- Terms that are connected in the knowledge graph should have vectors that are closer together
- Many extensions now, such as “antonyms should move farther apart” and morphology



Source: *N. Mrkšić, D. Ó Séaghdha, B. Thomson, M. Gašić, L. Rojas-Barahona, P. Su, D. Vandyke, T. Wen, S. Young*: Counter-fitting Word Vectors to Linguistic Constraints (2016)

Retrofitting just works*

- Until 11/2018, the top-performing systems almost always use retrofitting
- At Semeval 2018, about half of the top results used external information in order to get a top result.
- Other areas of ML were seeing similar results from adding structure information

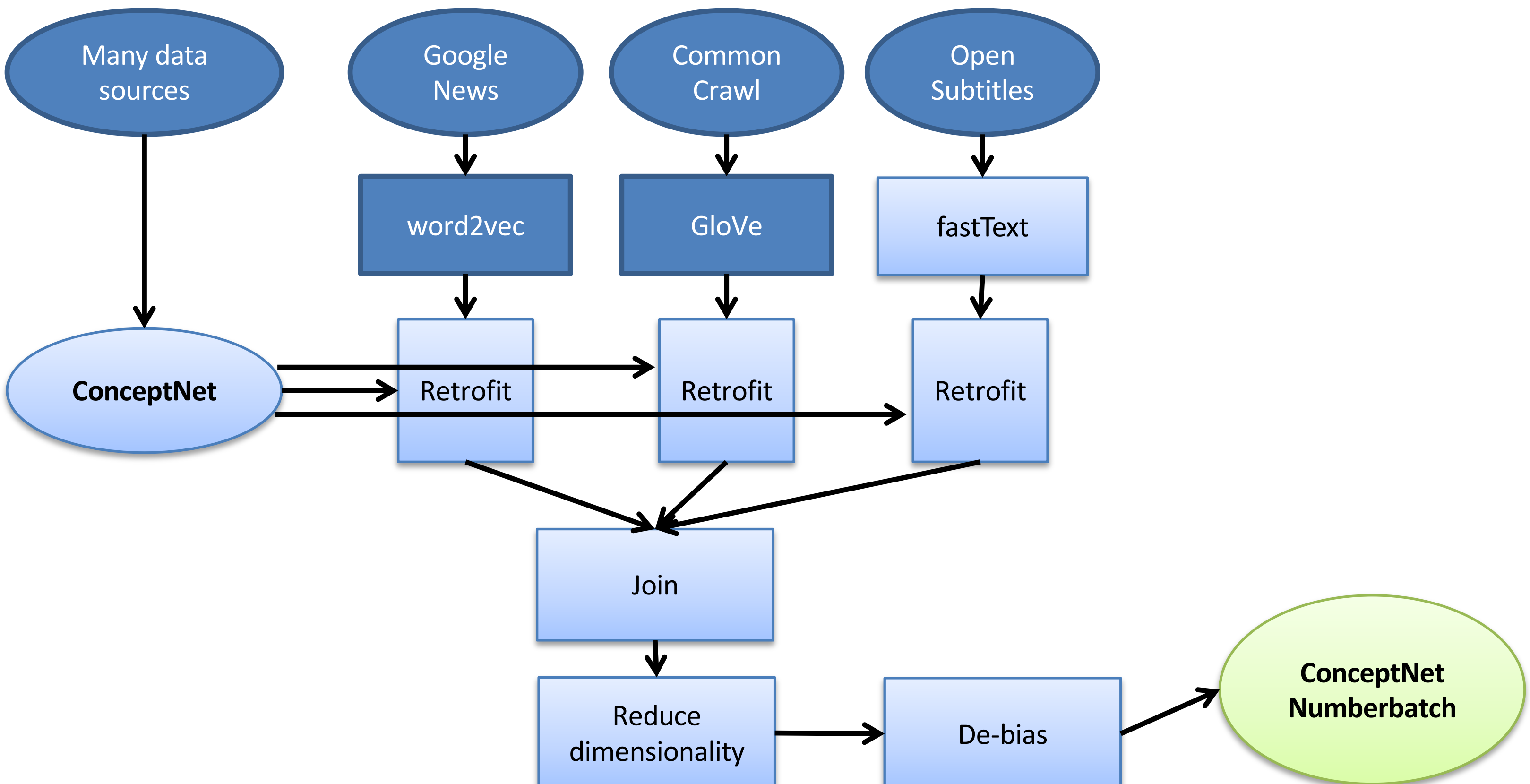
worked*

ConceptNet Numberbatch

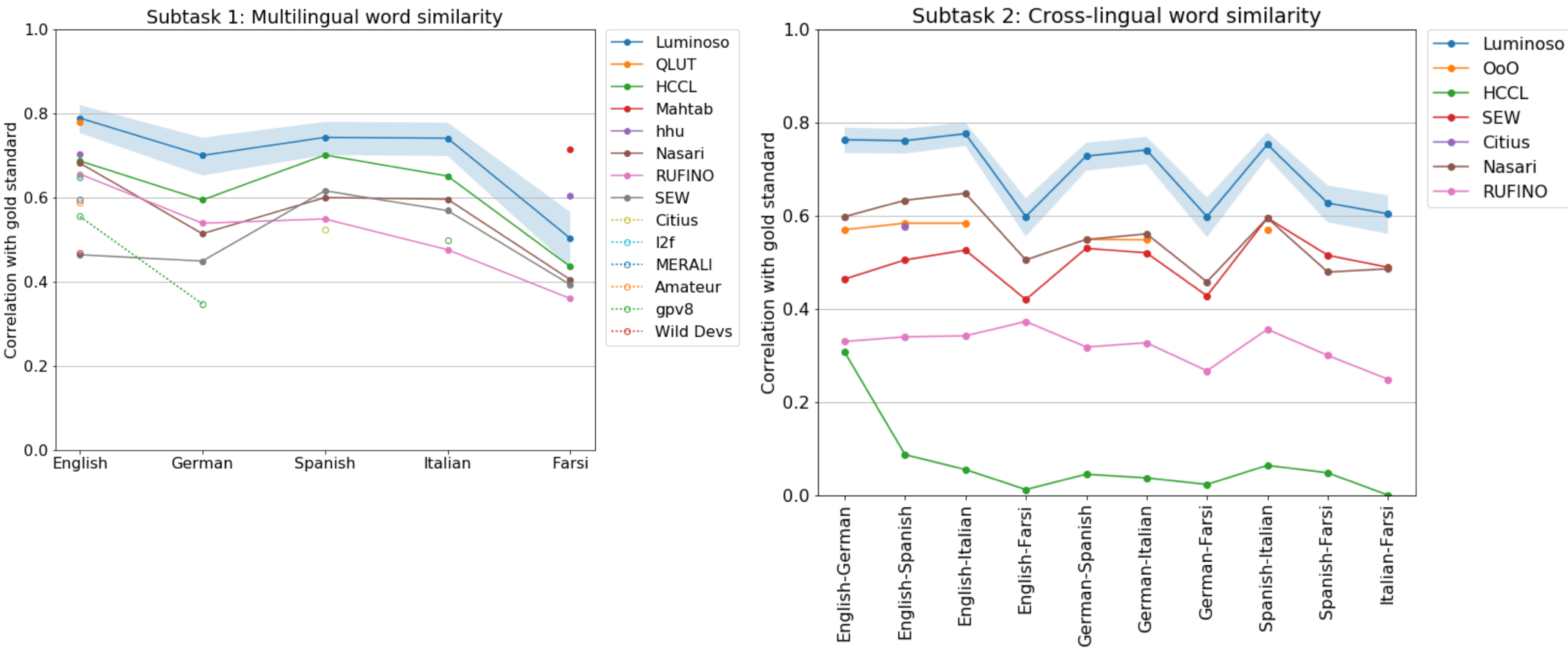
- Word vectors with common sense built in
- Hybrid of ConceptNet and distributional semantics
- Multilingual by design
- Open source, open data

Source: *Robyn Speer, Joshua Chin, and Catherine Havasi:* "ConceptNet 5.5: An Open Multilingual Graph of General Knowledge." (2017)

Building ConceptNet Numberbatch



Word relatedness (SemEval 2017)



Source: *Robyn Speer and Joanna Lowry-Duda*: Extending Word Embeddings with Multilingual Relational Knowledge (Semeval 2017)

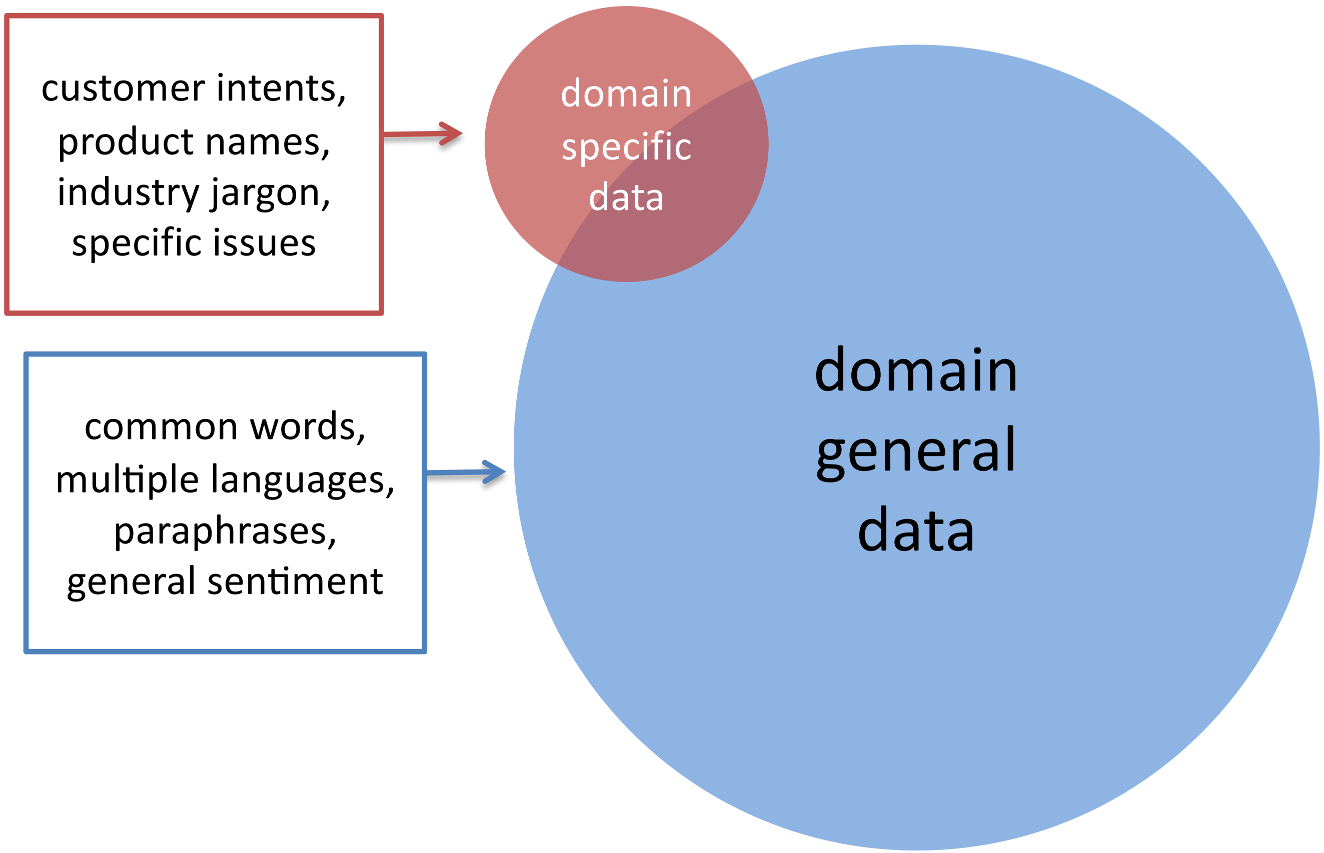
Distinguishing attributes using ConceptNet

Term 1	**Term 2**	**Attribute**	**Discriminative?**
lambs	cattle	wool	✔
__Lambs__ produce __wool__, while __cattle__ do not.			
shoulder	leg	arm	✔
A __shoulder__ is attached to an __arm__, while a __leg__ is not.			
train	subway	rails	✖
Both a __train__ and a __subway__ involve __rails__, so rails are not a discriminative attribute here.			
finger	soup	water	✖
__Soup__ may be related to __water__, but this is the wrong direction. In this task, a discriminative attribute must be related to the first term and not the second.			

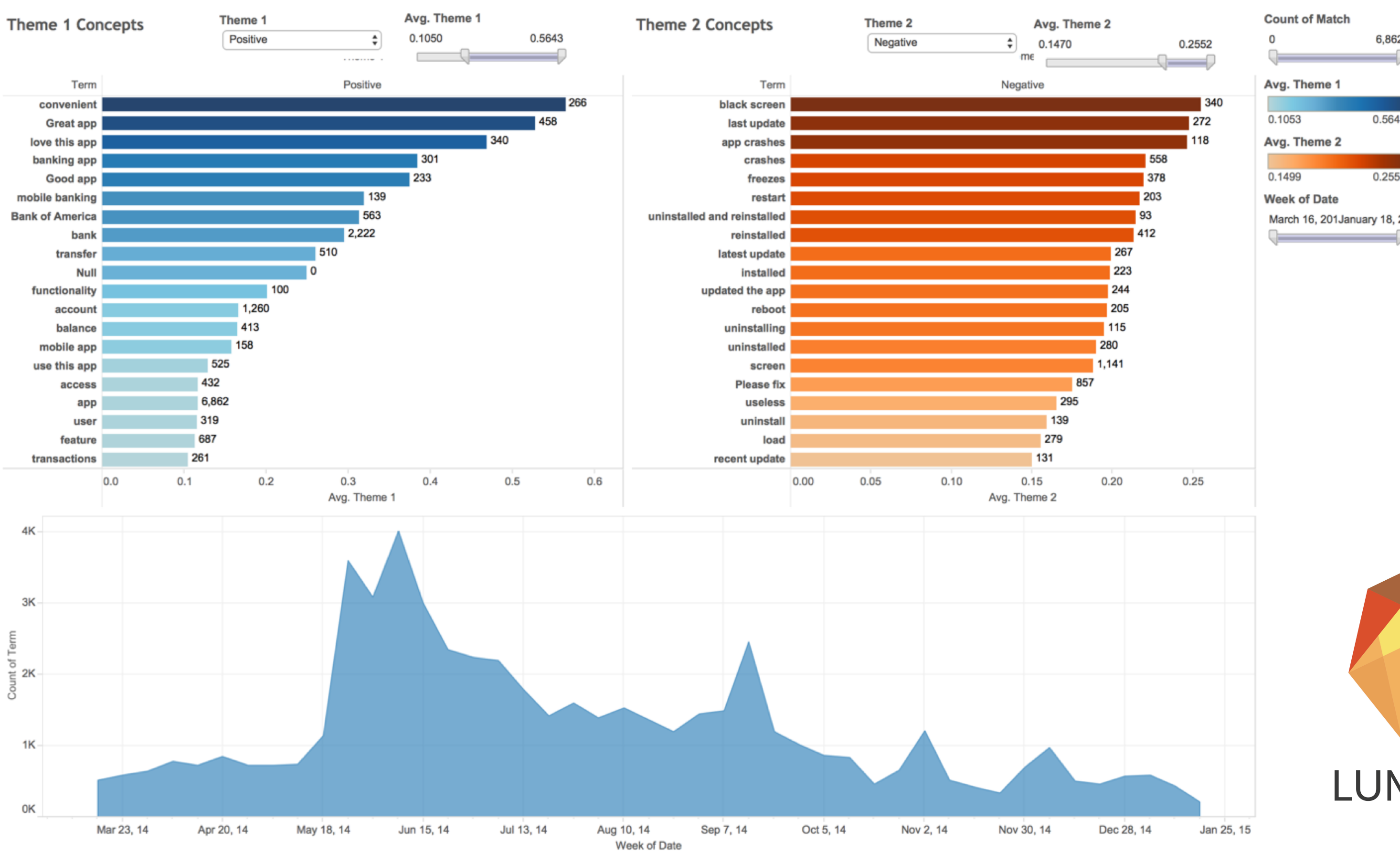
- A task at SemEval 2018
- We got **74%** accuracy (2nd place) by directly querying ConceptNet Numberbatch
- Additional features trained on the data didn't help on the test set
- All top systems used knowledge graphs

Source: *Robyn Speer and Joanna Lowry-Duda*: Distinguishing Attributes Using Text Corpora and Relational Knowledge (Semeval 2018)

What is domain Adaptation?



Spin-out: Luminoso (2010-)



LUMINOSO

What are other examples of transfer learning?

- Commercial transfer learning (Luminoso, etc)
- Pretraining
- Fine tuning and layer freezing for Elmo and Bert (and GPT-2)
- Fast.ai's ULMFiT (http://nlp.fast.ai/)

Computational Creativity and Common Sense

Story understanding evaluations

- The Story Cloze Test evaluates common sense
- Five-sentence stories, two possible endings, only one makes sense
  - Previous state of the art (GPT-1): 86.5%
  - Jiaao Chen et al., adding ConceptNet as an input: 87.6%

Story Cloze Test

Context	Right Ending	Wrong Ending
Karen was assigned a roommate her first year of college. Her roommate asked her to go to a nearby city for a concert. Karen agreed happily. The show was absolutely exhilarating.	Karen became good friends with her roommate.	Karen hated her roommate.
Jim got his first credit card in college. He didn't have a job so he bought everything on his card. After he graduated he amounted a $10,000 debt. Jim realized that he was foolish to spend so much money.	Jim decided to devise a plan for repayment.	Jim decided to open another credit card.
Gina misplaced her phone at her grandparents. It wasn't anywhere in the living room. She realized she was in the car before. She grabbed her dad's keys and ran outside.	She found her phone in the car.	She didn't want her phone anymore.

Source: *Jiaao Chen, Jianshu Chen, Zhou Yu*: Incorporating Structured Commonsense Knowledge in Story Completion (2018)

Story understanding evaluations

- SemEval-2018 task: answer simple multiple-choice questions about a passage

Text: It was a long day at work and I decided to stop at the gym before going home. I ran on the treadmill and lifted some weights. I decided I would also swim a few laps in the pool. Once I was done working out, I went in the locker room and stripped down and wrapped myself in a towel. I went into the sauna and turned on the heat. I let it get nice and steamy. I sat down and relaxed. I let my mind think about nothing but peaceful, happy thoughts. I stayed in there for only about ten minutes because it was so hot and steamy. When I got out, I turned the sauna off to save energy and took a cool shower. I got out of the shower and dried off. After that, I put on my extra set of clean clothes I brought with me, and got in my car and drove home.

Q1: Where did they sit inside the sauna?

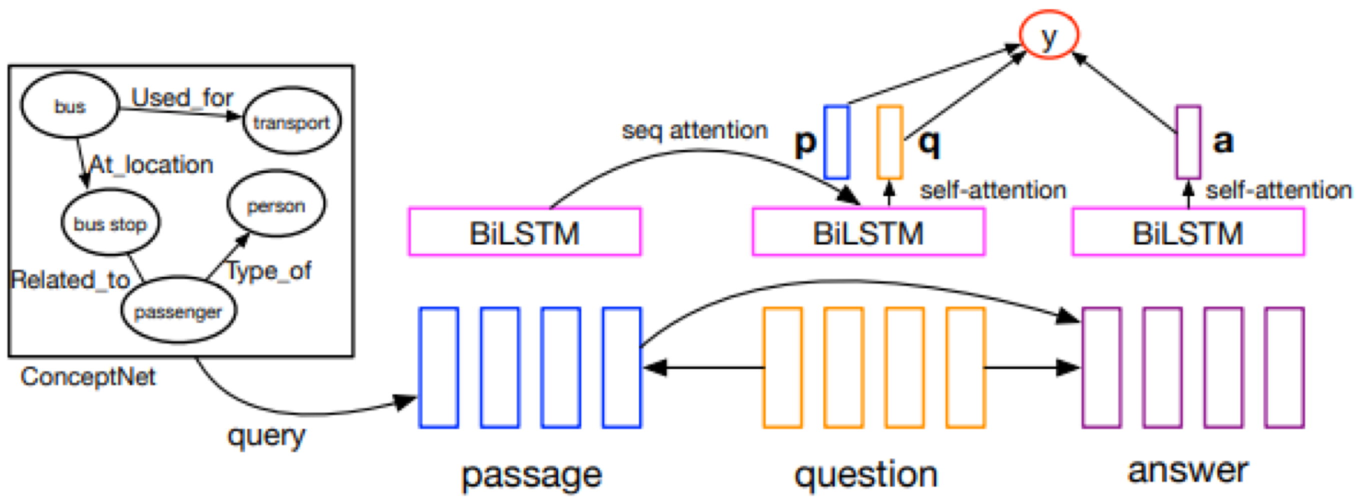
(a) on the floor
(b) on a bench

Q2: How long did they stay in the sauna?

(a) about ten minutes
(b) over thirty minutes

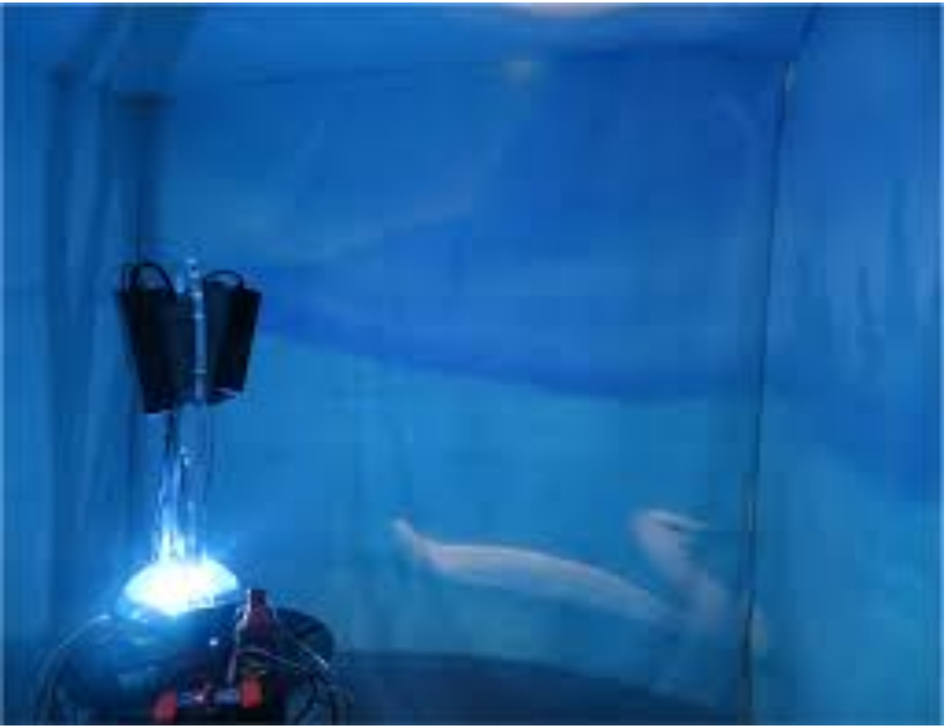
Story understanding using ConceptNet

- The winning system on the SemEval task (Yuanfudao Research) used ConceptNet to find unstated connections, on top of an LSTM + attention model



Source: *Liang Wang, Meng Sun, Wei Zhao, Kewei Shen, Jingming Liu*: Three-way Attention and Relational Knowledge for Commonsense Machine Comprehension (2018)

Narratarium



Source: *K. Hayden, D. Novy C. Havasi, M. Bove, S. Alfaro R. Speer*: Narratarium: An Immersive Storytelling Environment (2016)

```
>>> Where have you been?
well, as far as i know, they've been having a good time in there.
>>> And what are you doing?
well, i'm trying to set up a three phase energy beam to neutralise the neutron
>>> Isn't it already neutralized?
no, it won't be neutralised.
>>> Why not?
well, er, we don't know. there's no known cure.
>>> For a neutron?
yes, for a neutron star.
>>> Oh.
and we are not ready for it.
>>>
```

Intelligent Agents that focus on character-based interactions

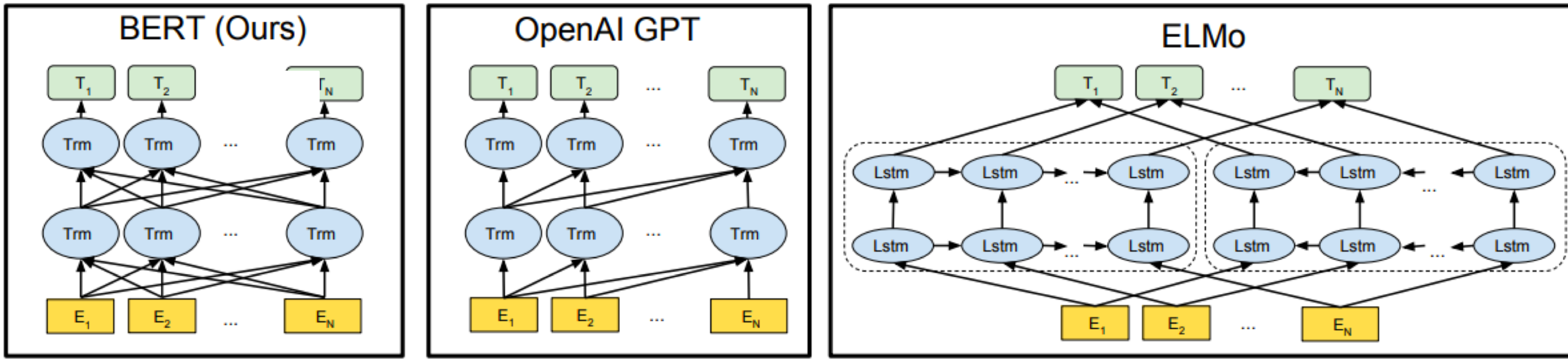
Jason Alonso, Catherine Havasi

Can we use this to bring experiences to life?

WITH: NINA LUTZ AND
JASON ALONSO ~ 2020

Can you tell me how to get to Sesame Street?

BERT: predicting masked words and next sentences



Input = [CLS] the man went to [MASK] store [SEP]

he bought a gallon [MASK] milk [SEP]

Label = IsNext

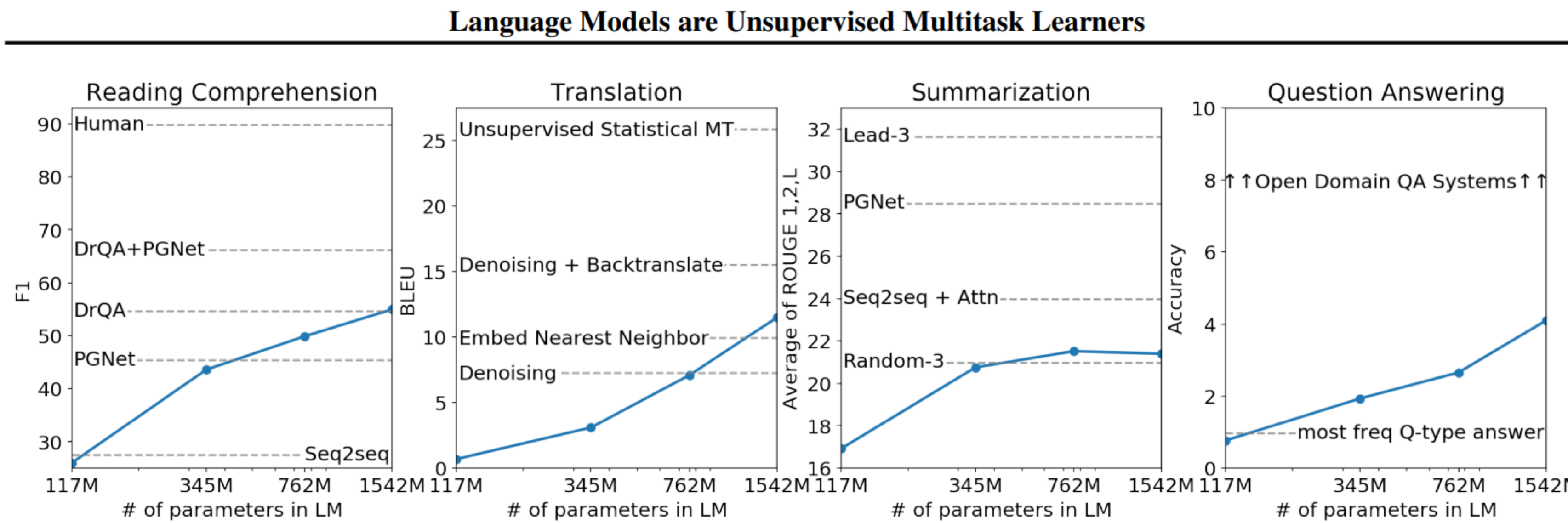
Input = [CLS] the man [MASK] to the store [SEP]

penguin [MASK] are flight ##less birds [SEP]

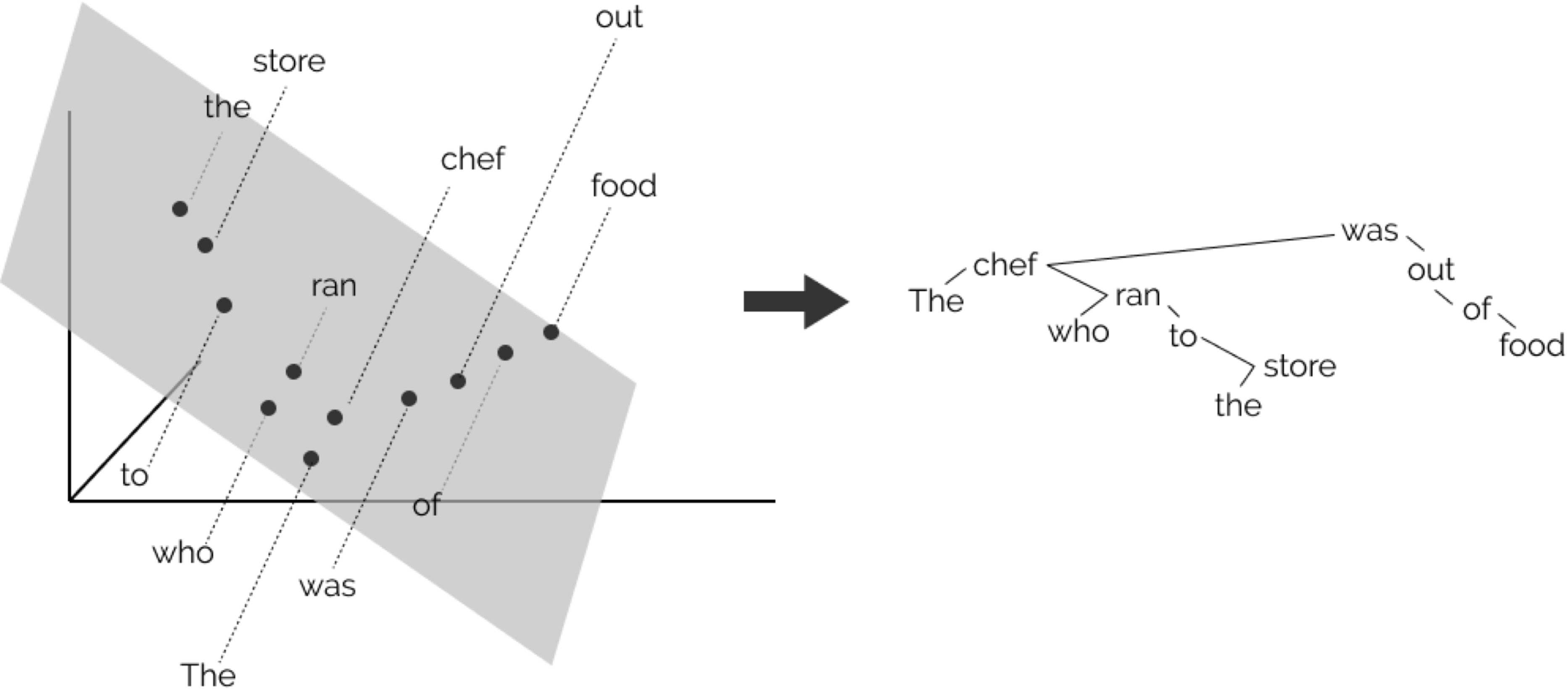
Label = NotNext

Source: *Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova*: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2019)

And are more generalizable than previous models…



Source: *Alec Radford, Jeffrey Wu, Rewon Child, David Luan,Dario Amodei, Ilya Sutskever* : Language Models are Unsupervised Multitask Learners (2019)

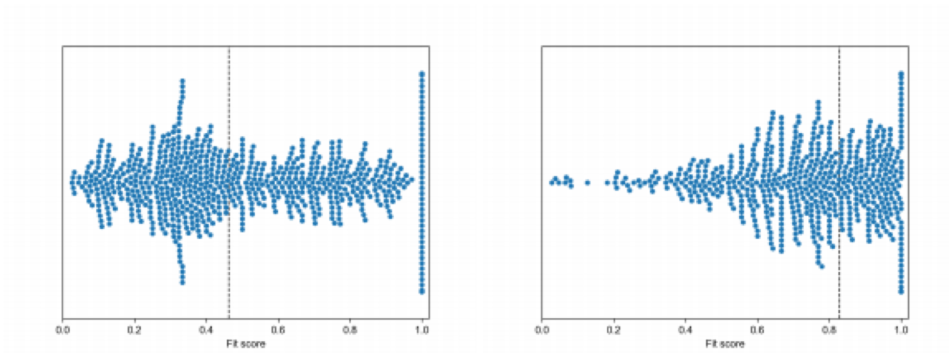


John Hewitt and Christopher D. Manning: A Structural Probe for Finding Syntax in Word Representations

Andy Coenen, Emily Reif, Ann Yuan, Been Kim, Adam Pearce, Fernanda Viégas, Martin Wattenberg.Visualizing and Measuring the Geometry of BERT

Was the knowledge inside all along?

Is BERT good at Common Sense?

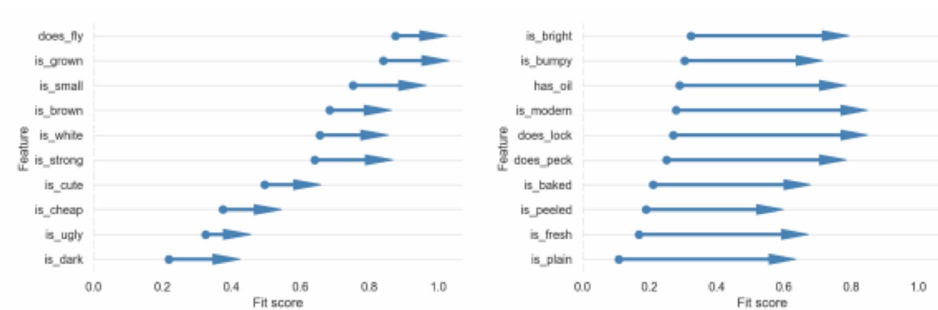


(a) GloVe (b) BERT

Figure 1: Swarm plots showing attribute fit scores for GloVe (left) and BERT (right). Each dot represents a single attribute, displayed along the x-axis according to the classifier's ability to fit that feature with the given embeddings. The y-axis is not significant, and instead, dots are displaced along the y-axis instead of overlapping to show quantity. The median fit score per embedding type is displayed with a dotted line.

Metric	Visual	Encyclopedic	Functional	Perceptual	Taxonomic	Overall
Median_{GloVe}	46.2	38.9	44.4	49.0	89.1	46.1
Median_{BERT}	83.3	76.2	78.3	80.0	100	82.7
Δ	**+37.1**	**+37.3**	**+33.9**	**+31.0**	**+10.9**	**+36.6**

Table 1: Comparison of median logistic classifier fit scores (out of 100 percent fit) across categories defined in CSLB.



(a) Small increase in fit score (< 0.15) (b) Large increase in fit score (> 0.3)

Figure 2: Differences between fit scores when using GloVe (start of arrow) or BERT (end of arrows) embeddings.

Source: *Jeff Da, Jungo Kasai*: Cracking the Contextual Commonsense Code: Understanding Commonsense Reasoning Aptitude of Deep Contextual Representations

Common sense understanding is not the default

- Question-answering AIs learn to give typical answers like “Yes” and “2”
- Story-understanding AIs learn that the wrong answers are written in a different style
- Reading-comprehension AIs learn to copy the nearest answer-shaped phrase

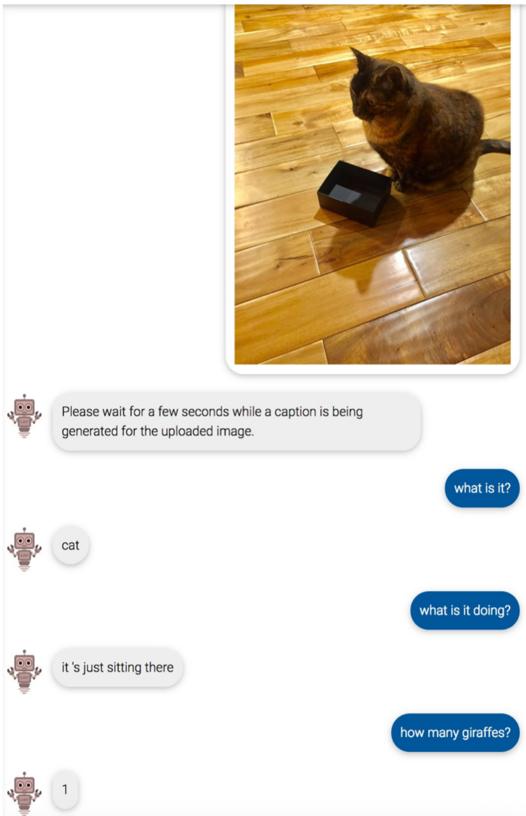
Article: Endangered Species Act
Paragraph: “ *. . . Other legislation followed, including the Migratory Bird Conservation Act of 1929, a 1937 treaty prohibiting the hunting of right and gray whales, and the Bald Eagle Protection Act of 1940. These later laws had a low cost to society—the species were relatively rare—and little opposition was raised.*”

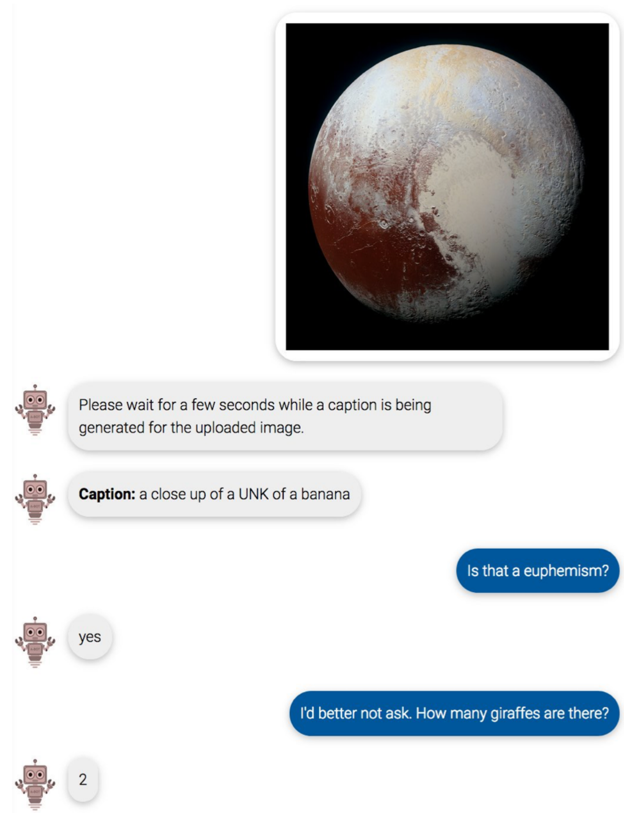
Question 1: “*Which laws faced significant opposition?*”
Plausible Answer: *later laws*

Question 2: “*What was the name of the 1937 treaty?*”
Plausible Answer: *Bald Eagle Protection Act*

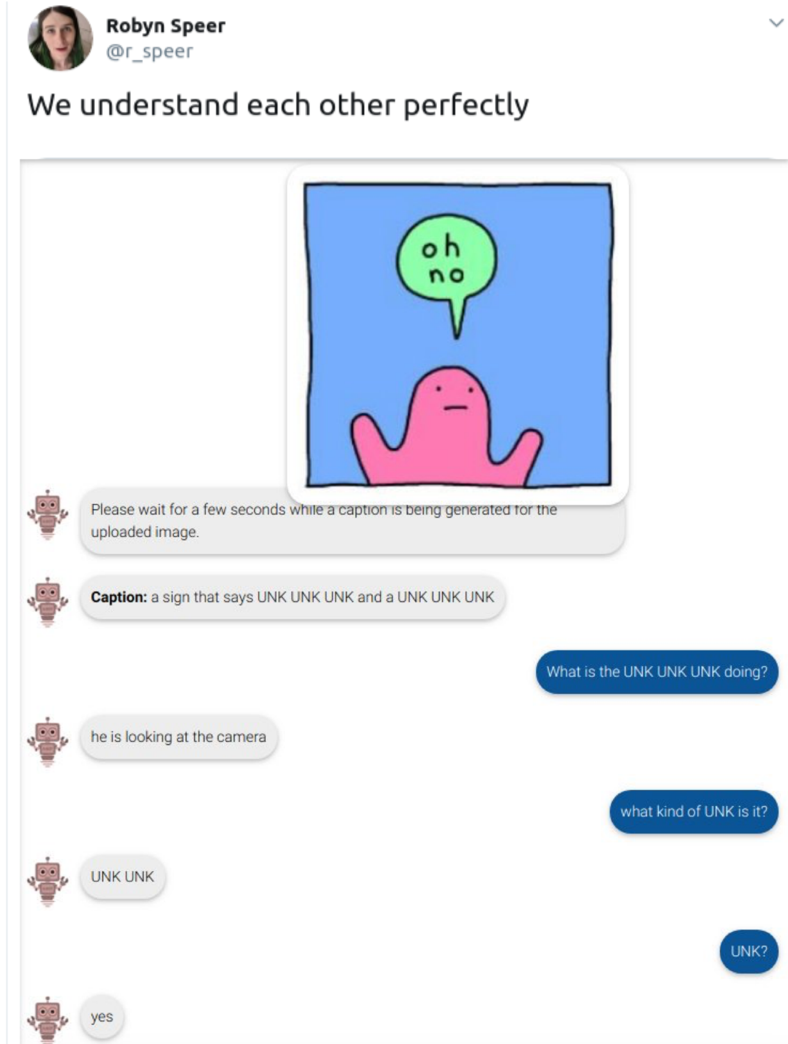
Source: *Pranav Rajpurkar, Robin Jia, Percy Liang*: Know What You Don’t Know: Unanswerable Questions for SQuAD (2018)

Common sense understanding is not the default



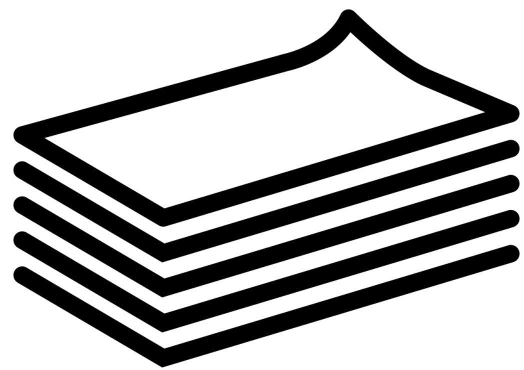


Source: Janelle Shane, on Twitter, using Visual Chatbot



Source: Robyn Speer, on Twitter, using Visual Chatbot

Training a transformer takes a lot of energy.

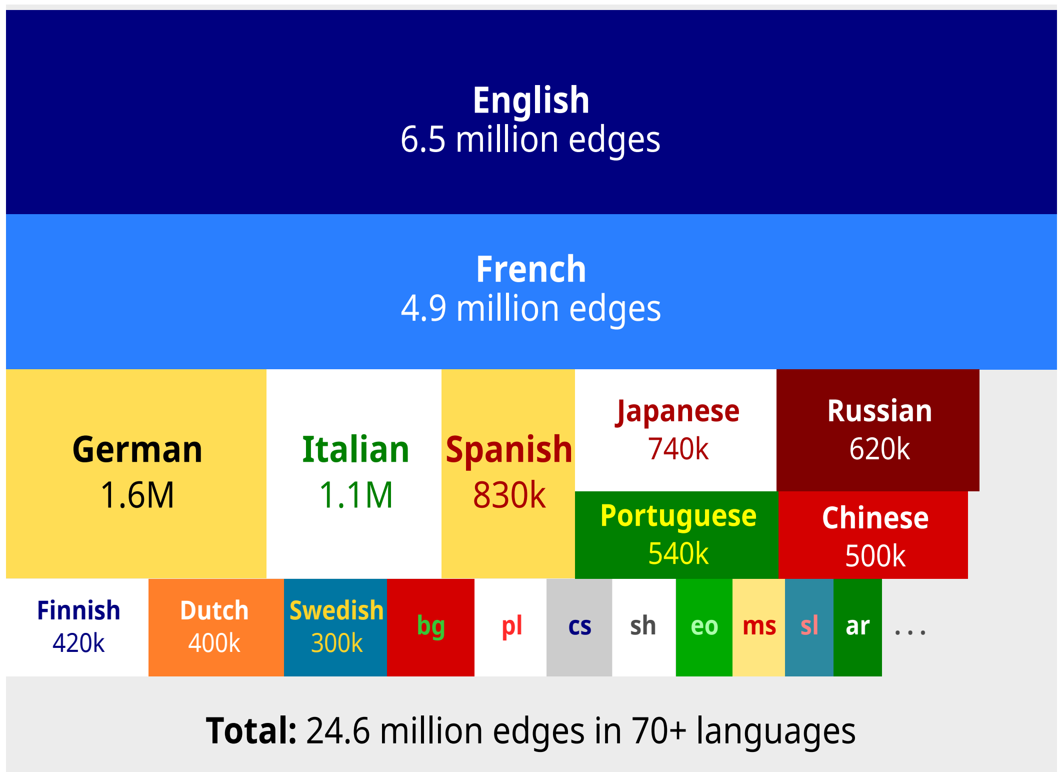


Often, there isn't enough data for a domain or language to train a new model.

One Size (Pre-trained) Model Does **Not** Fit All:

Language*
Domain
Cultural/Identity

Languages in ConceptNet



Evaluations in Portuguese

Table 15. Accuracy for answering cloze questions.

	Noun (1769)	Verb (1077)	Adj (809)	Adv (235)	Total (3890)
Baseline	34.43%	32.82%	25.28%	25.11%	31.52%
PAPEL	**44.19%**	**36.63%**	**33.47%**	22.13%	**38.53%**
DA	39.49%	32.87%	30.01%	24.36%	34.77%
Wikt.PT	39.85%	35.65%	31.15%	27.45%	36.13%
OpenWN-PT	38.72%	31.78%	25.28%	26.17%	33.25%
PULO	40.77%	31.43%	22.16%	23.19%	33.25%
TeP	41.72%	30.71%	31.49%	25.00%	35.53%
OpenThes.PT	35.01%	26.51%	26.21%	25.43%	30.24%
Port4Nooj	37.11%	26.86%	27.97%	**29.89%**	31.93%
WN.Br	24.82%	29.55%	24.44%	25.11%	26.07%
ConceptNet	37.00%	34.42%	32.55%	27.73%	34.79%
CARTÃO	46.78%	**36.86%**	**36.46%**	27.77%	**40.74%**
Redun3	40.54%	32.61%	28.83%	27.70%	35.13%
Redun2	45.00%	34.03%	30.44%	**28.09%**	37.90%
All	**49.90%**	33.05%	34.98%	26.81%	**40.72%**

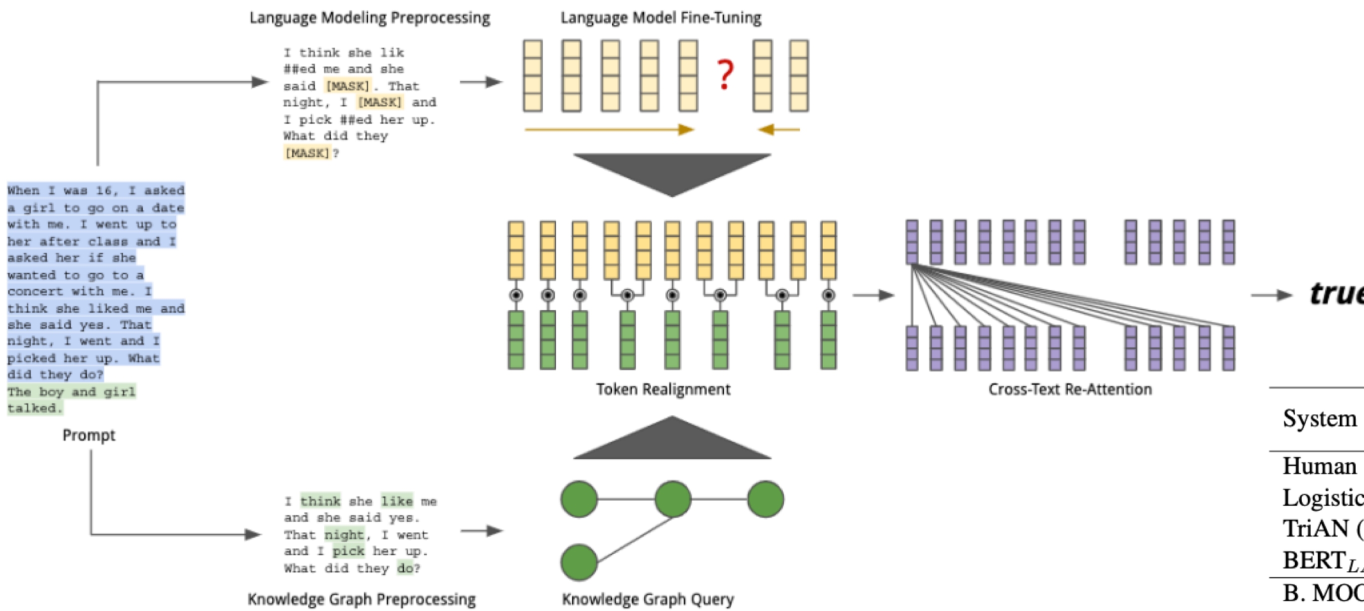
Table 13. Selection of results for the SimLex-999 test.

LKB	Relations	Algorithm	ρ
PAPEL	All	PR-Jac$_{800}$	0.49
DA	All	PR-Jac$_{400}$	0.38
Wikt.PT	All	PR-Jac$_{1600}$	0.42
OWN-PT	Syn + Hyp	Adj-Cos	0.44
PULO	Syn + Hyp	Adj-Cos	0.29
TeP	Syn + Hyp	Adj-Jac	0.36
OT.PT	Syn + Hyp	Adj-Cos	0.34
Port4Nooj	All	Adj-Jac	0.19
WN.Br	Syn + Hyper	Adj-Jac	0.04
ConceptNet	Syn + Hyp	Adj-Jac	0.43
CARTÃO	All	PR-CosV$_{1600}$	0.53
Redun3	Syn + Hyper	Adj-Jac	0.44
Redun2	Syn + Hyper	PR-Jac$_{50}$	0.49
All	Syn + Hyper	PR-CosV$_{50}$	0.57
All	Syn + Hyper	PR-CosV$_{100}$	0.59
All	Syn + Hyper	PR-CosV$_{200}$	**0.61**
All	Syn + Hyper	PR-CosV$_{400}$	**0.61**
All	Syn + Hyper	PR-CosV$_{800}$	**0.61**
All	Syn + Hyper	PR-CosV$_{1600}$	0.60
All	Syn + Hyper	PR-CosV$_{3200}$	0.60
All	Syn + Hyper	Adj-Cos	0.58
All	Syn + Hyper	Adj-Jac	0.57
All	All	PR-CosV$_{400}$	0.56

Source: *Hugo Gonçalo Oliveira*: A Survey on Portuguese Lexical Knowledge Bases: Contents, Comparison and Combination (2018)

Could you make transformers more accurate using data like we did with retrofitting?

Transformers + Knowledge

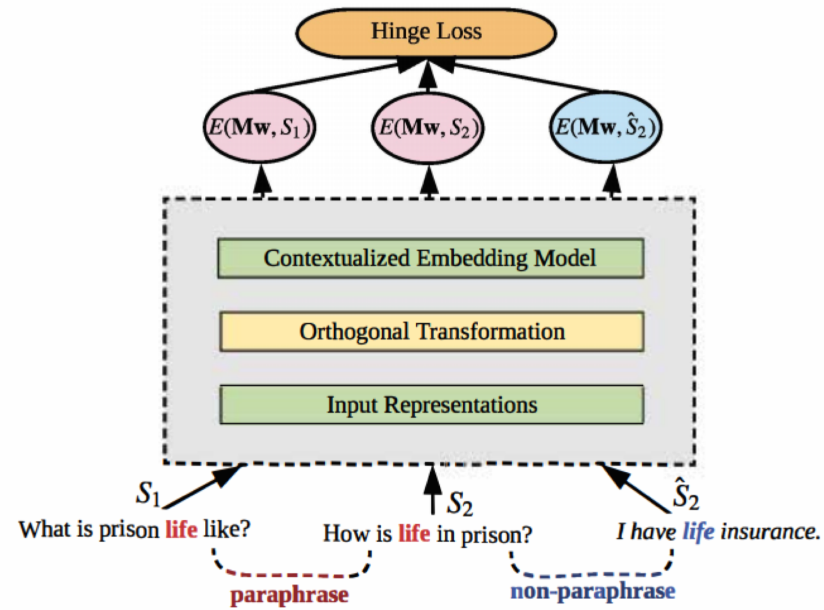


System	Accuracy Dev	Accuracy Test
Human	97.4	98.0
Logistic Baseline	-	60.8
TriAN (Wang, 2018)	76.1	-
BERT_{LARGE}	82.3	-
B. MOOD (with ConceptNet)	83.1	-
B. MOOD (with WebChild)	82.7	-
B. MOOD (with ATOMIC)	82.5	-
B. MOOD (w/o final attention)	82.4	-
B. MOOD (with all KB)	**83.3**	**80.7**

Source: *Jeff Da*: BIG MOOD: Relating Transformers to Explicit Commonsense Knowledge (COIN 2019)

Transformers + Knowledge

Source: *Weijia Shi, Muhao Chen, Pei Zhou, Kai-Wei Chang*: Retrofitting Contextualized Word Embeddings with Paraphrases (2019)



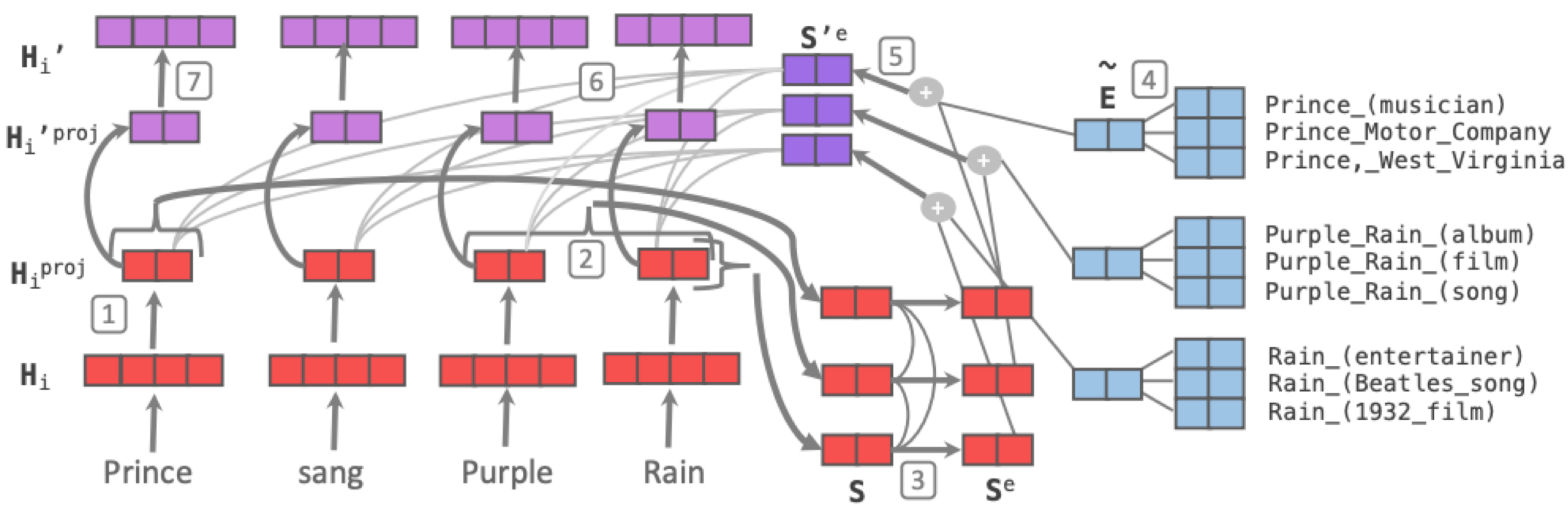
Transformers + ConceptNet

Source: *Z. Ye, Q. Chen, W. Wang, Z. Ling:* Align, Mask and Select: A Simple Method for Incorporating Commonsense Knowledge into Language Representation Models

(1) A triple from ConceptNet
(population, AtLocation, city)
(2) **Align** with the English Wikipedia dataset to obtain a sentence containing "population" and "city"
The largest **city** by **population** is Birmingham, which has long been the most industrialized city.
(3) **Mask** "city" with a special token "[QW]"
The largest **[QW]** by **population** is Birmingham, which has long been the most industrialized city?
4) **Select** distractors by searching (population, AtLocation, $*$) in ConceptNet
(population, AtLocation, Michigan) (population, AtLocation, Petrie dish) (population, AtLocation, area with people inhabiting) (population, AtLocation, country)
5) Generate a multi-choice question answering sample
question: The largest **[QW]** by **population** is Birmingham, which has long been the most industrialized city? **candidates**: *city*, Michigan, Petrie dish, area with people inhabiting, country

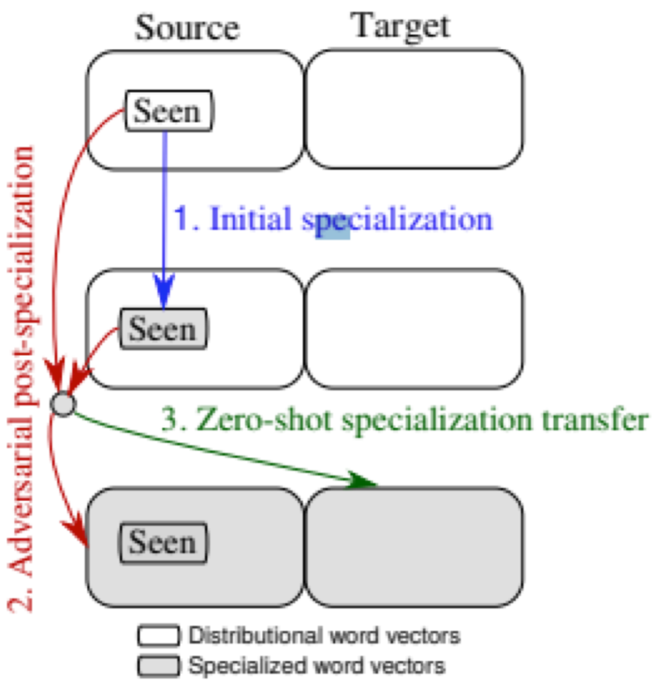
Table 2: The detailed procedure of constructing a multi-choice question answering sample with the proposed AMS method. The $*$ in the fourth step is a wildcard character. The correct answer for the question is underlined.

Transformers + Entity Graphs



Source: *M. Peters, M. Neumann, R. Logan, R. Schwartz, V. Joshi, S. Singh, N. Smith*: Knowledge Enhanced Contextual Word Representations

Can we teach a network to retrofit?

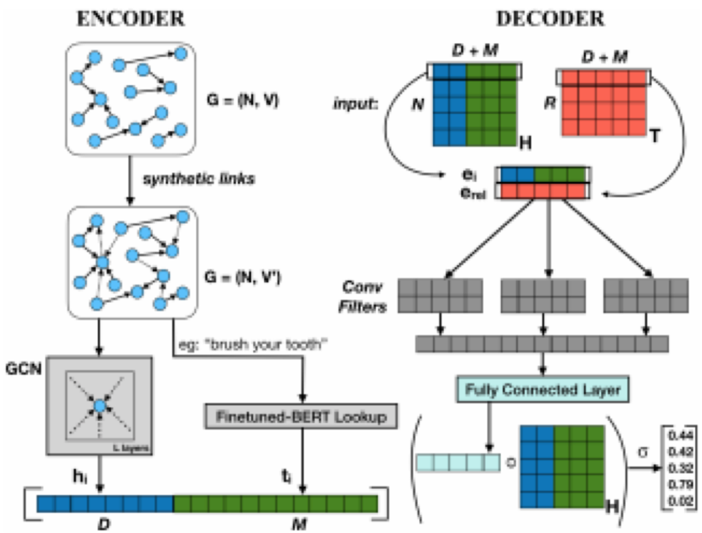


- Use a GAN to teach a network to add a knowledge graph to an embedding
- Essentially increasing the vocabulary of the kgraph
- Better downstream performance
- Potential for work in bi-lingual embedding space

Source: *Edoardo M. Ponti, Ivan Vulic, Goran Glavaš, Nikola Mrkšic, Anna Korhonen*
Adversarial Propagation and Zero-Shot Cross-Lingual Transfer of Word Vector Specialization (2019)

Can we use transformers to build knowledge graphs?

Can transformers give back to knowledge graphs?



- Big idea: Increase the density of a knowledge graph by adding additional links

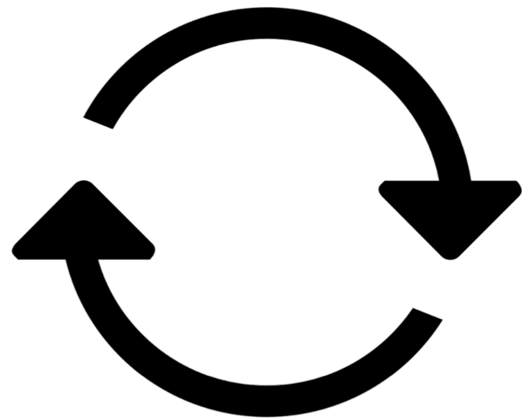
Source: C. Malaviya, C. Bhagavatula, A. Bosselut, Y. Choi: Exploiting Structural and Semantic Context for Commonsense Knowledge Base Completion (2019)

Inferring common sense with CoMET

- Uses ConceptNet as a training set instead of a knowledge resource
- Fine-tune a GPT language model to generate ConceptNet statements

Seed	Relation	Completion	Plausible
piece	`PartOf`	machine	✓
bread	`IsA`	food	✓
oldsmobile	`IsA`	car	✓
happiness	`IsA`	feel	✓
math	`IsA`	subject	✓
mango	`IsA`	fruit	✓
maine	`IsA`	state	✓
planet	`AtLocation`	space	✓
dust	`AtLocation`	fridge	
puzzle	`AtLocation`	your mind	🤔
college	`AtLocation`	town	✓
dental chair	`AtLocation`	dentist	✓
finger	`AtLocation`	your finger	
sing	`Causes`	you feel good	✓
doctor	`CapableOf`	save life	✓
post office	`CapableOf`	receive letter	✓
dove	`SymbolOf`	purity	✓
sun	`HasProperty`	big	✓
bird bone	`HasProperty`	fragile	✓
earth	`HasA`	many plant	✓
yard	`UsedFor`	play game	✓
get pay	`HasPrerequisite`	work	✓
print on printer	`HasPrerequisite`	get printer	✓
play game	`HasPrerequisite`	have game	✓
live	`HasLastSubevent`	die	✓
swim	`HasSubevent`	get wet	✓
sit down	`MotivatedByGoal`	you be tire	✓
all paper	`ReceivesAction`	recycle	✓
chair	`MadeOf`	wood	✓
earth	`DefinedAs`	planet	✓

Source: *A. Bosselut, H. Rashkin, M. Sap, C. Malaviya, A. Celikyilmaz, Y. Choi*: COMET: Commonsense Transformers for Automatic Knowledge Graph Construction (2019)




●●○○○ T-Mobile Wi-Fi
4:42 PM
80%
Go ahead,
I'm listening...

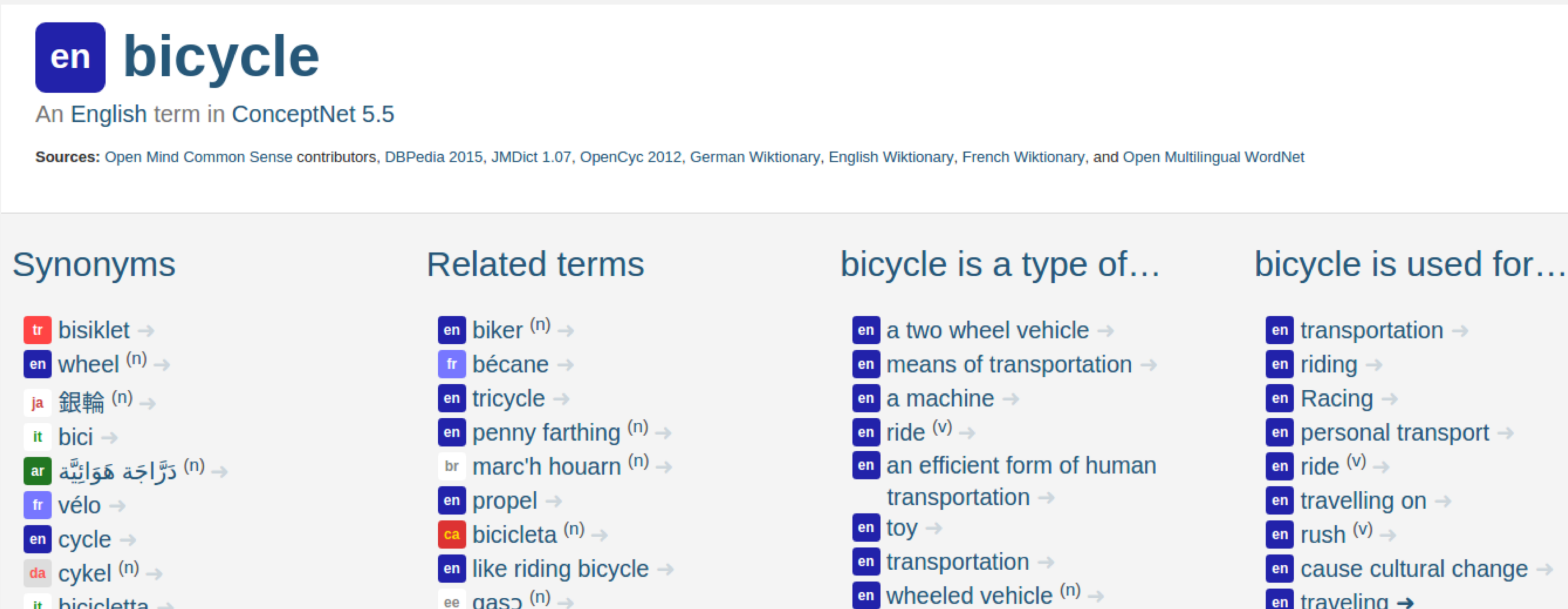
What about human-centric AI?

Knowledge graphs & explainablity?

Image Credit: Amazon


Let’s find a way to communicate with machines like we always dreamed of.

conceptnet.io – a browsable interface



api.conceptnet.io – a Linked Data API


```
{
  "@context": [
    "http://api.conceptnet.io/ld/conceptnet5.5/context.ld.json",
    "http://api.conceptnet.io/ld/conceptnet5.5/pagination.ld.json"
  ],
  "@id": "/c/en/bicycle",
  "edges": [
    {
      "@id": "/a/[/r/AtLocation/,/c/en/bicycle/,/c/en/garage/]",
      "dataset": "/d/conceptnet/4/en",
      "end": {
        "@id": "/c/en/garage",
        "label": "the garage",
        "language": "en",
        "term": "/c/en/garage"
      },
      "license": "cc:by/4.0",
      "rel": {
        "@id": "/r/AtLocation",
        "label": "AtLocation"
      },
```

We'd like to work with you…

- Building and expanding ConceptNet
- Cross-resource alignment
- Bias in machine learning and structural resources
- Explainablity in kgs
- Other ideas

Thank you to all ConceptNet Collaborators

- Co-director Robyn Speer
- Jason Alonso, Charlotte Chen, Pedro Colon-Hernandez, Joanna Lowry-Duda, Nina Lutz, Katherine Xiao
- So many others over the years…

blog.conceptnet.io


Thank you to all of my crowds over the years.
Photo: Moses Namkung, CC-By

ConceptNet
Numberbatch

Thank you!

Dr. Catherine Havasi
havasi@media.mit.edu
@catherinehavasi