CS769 Advanced NLP

Information Extraction and Knowledge-based QA

Junjie Hu

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Slides adapted from Graham, Zhengbao

https://junjiehu.github.io/cs769-spring22/

Goal for Today

- Types of Knowledge Bases (KB)
- Information Extraction (IE) for Constructing KB
  1. IE w/ Pre-defined Relations
  2. OpenIE w/o Pre-defined Relations
- Using KB to Inform Neural Nets
- Probing Knowledge in LMs

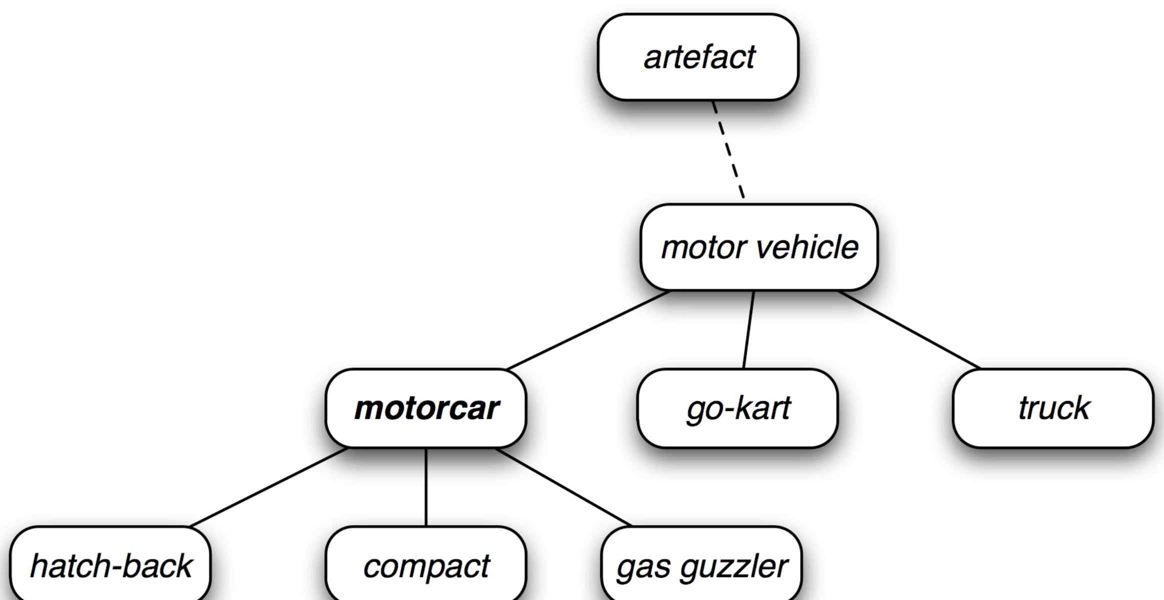
Knowledge Bases

- Structured databases of knowledge usually containing
  - Entities (nodes in a graph)
  - Relations (edges between nodes)
- How can we **learn to create/expand knowledge bases** with neural networks?
- How can we **learn from the information in knowledge bases** to improve neural representations?
- How can we use structured knowledge to answer questions (see also semantic parsing class)

Types of Knowledge Bases

WordNet (Miller 1995)

- WordNet is a large database of words including parts of speech, semantic relations



- Nouns: is-a relation (hatch-back/car), part-of (wheel/car), type/instance distinction
- Verb relations: ordered by specificity (communicate -> talk -> whisper)
- Adjective relations: antonymy (wet/dry)

Image Credit: NLTK

Cyc (Lenant 1995)

- A manually curated database attempting to encode all common sense knowledge, 30 years in the making

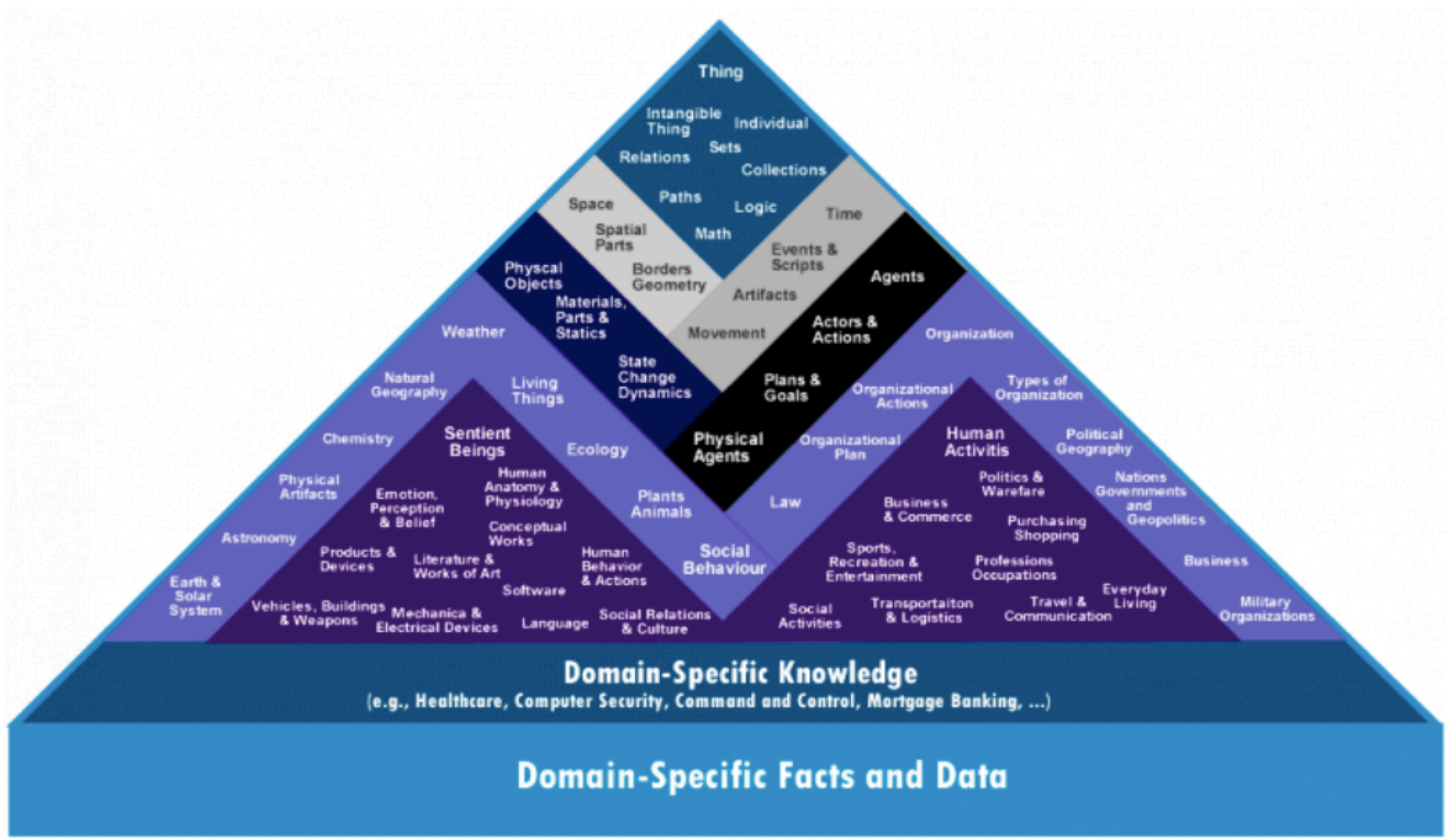
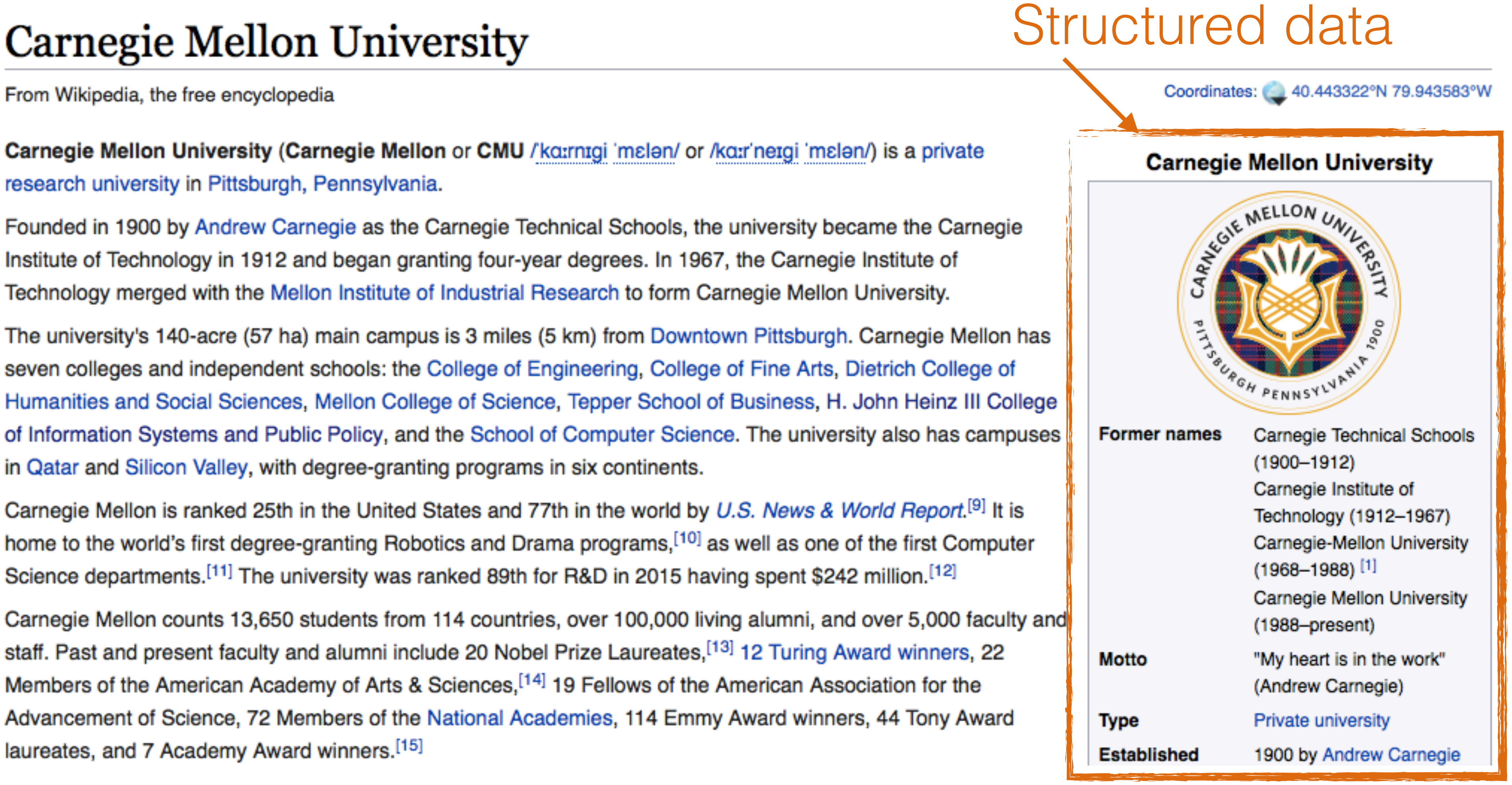


Image Credit: NLTK

DBPedia (Auer et al. 2007)

- Extraction of structured data from Wikipedia

Structured data



Carnegie Mellon University

From Wikipedia, the free encyclopedia

Coordinates: 40.443322°N 79.943583°W

Carnegie Mellon University (**Carnegie Mellon** or **CMU** /ˈkɑːrnɪɡi ˈmɛlən/ or /kɑːrˈneɪɡi ˈmɛlən/) is a private research university in Pittsburgh, Pennsylvania.

Founded in 1900 by Andrew Carnegie as the Carnegie Technical Schools, the university became the Carnegie Institute of Technology in 1912 and began granting four-year degrees. In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University.

The university's 140-acre (57 ha) main campus is 3 miles (5 km) from Downtown Pittsburgh. Carnegie Mellon has seven colleges and independent schools: the College of Engineering, College of Fine Arts, Dietrich College of Humanities and Social Sciences, Mellon College of Science, Tepper School of Business, H. John Heinz III College of Information Systems and Public Policy, and the School of Computer Science. The university also has campuses in Qatar and Silicon Valley, with degree-granting programs in six continents.

Carnegie Mellon is ranked 25th in the United States and 77th in the world by *U.S. News & World Report*.[9] It is home to the world's first degree-granting Robotics and Drama programs,[10] as well as one of the first Computer Science departments.[11] The university was ranked 89th for R&D in 2015 having spent $242 million.[12]

Carnegie Mellon counts 13,650 students from 114 countries, over 100,000 living alumni, and over 5,000 faculty and staff. Past and present faculty and alumni include 20 Nobel Prize Laureates,[13] 12 Turing Award winners, 22 Members of the American Academy of Arts & Sciences,[14] 19 Fellows of the American Association for the Advancement of Science, 72 Members of the National Academies, 114 Emmy Award winners, 44 Tony Award laureates, and 7 Academy Award winners.[15]

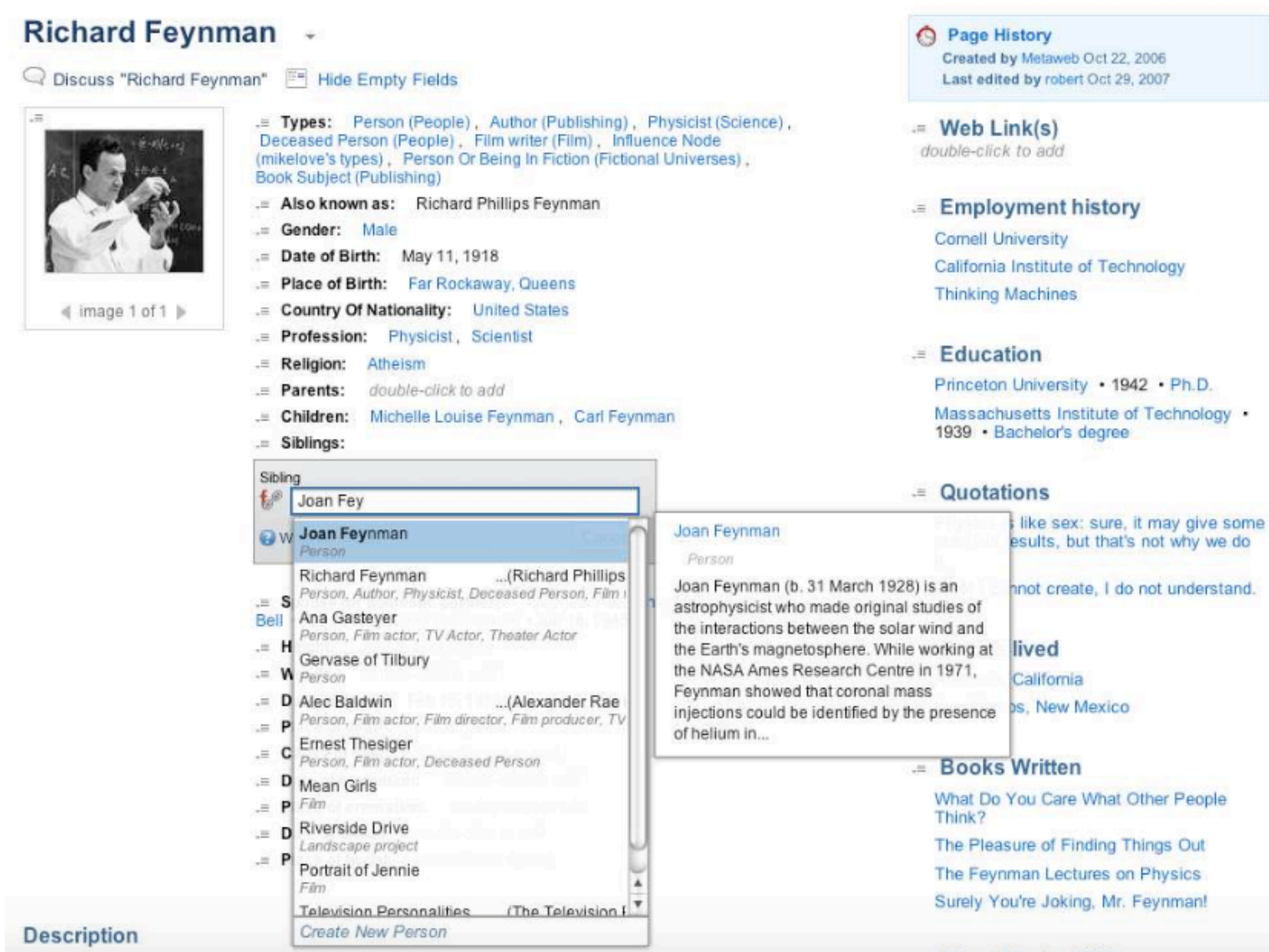
Carnegie Mellon University

Former names	Carnegie Technical Schools (1900–1912) Carnegie Institute of Technology (1912–1967) Carnegie-Mellon University (1968–1988) [1] Carnegie Mellon University (1988–present)
Motto	"My heart is in the work" (Andrew Carnegie)
Type	Private university
Established	1900 by Andrew Carnegie

- owl:Thing
- dul:Agent
- dul:SocialPerson
- wikidata:Q24229398
- wikidata:Q3918
- wikidata:Q43229
- dbo:Agent
- dbo:EducationalInstitution
- dbo:Organisation
- dbo:University
- geo:SpatialThing
- schema:CollegeOrUniversity
- schema:EducationalOrganization
- schema:Organization
- umbel-rc:Business
- umbel-rc:EducationalOrganization
- umbel-rc:Organization
- umbel-rc:University

WikiData (Bollacker et al. 2008)

- *Curated* database of entities, linked, and extremely large scale, multilingual



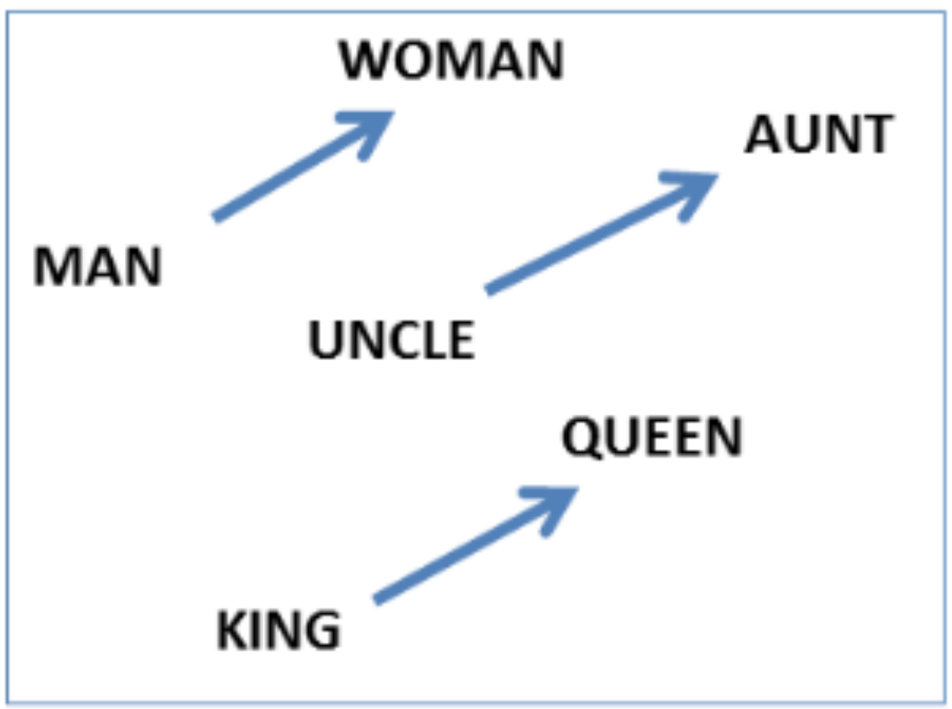
Learning Representations for Knowledge Bases

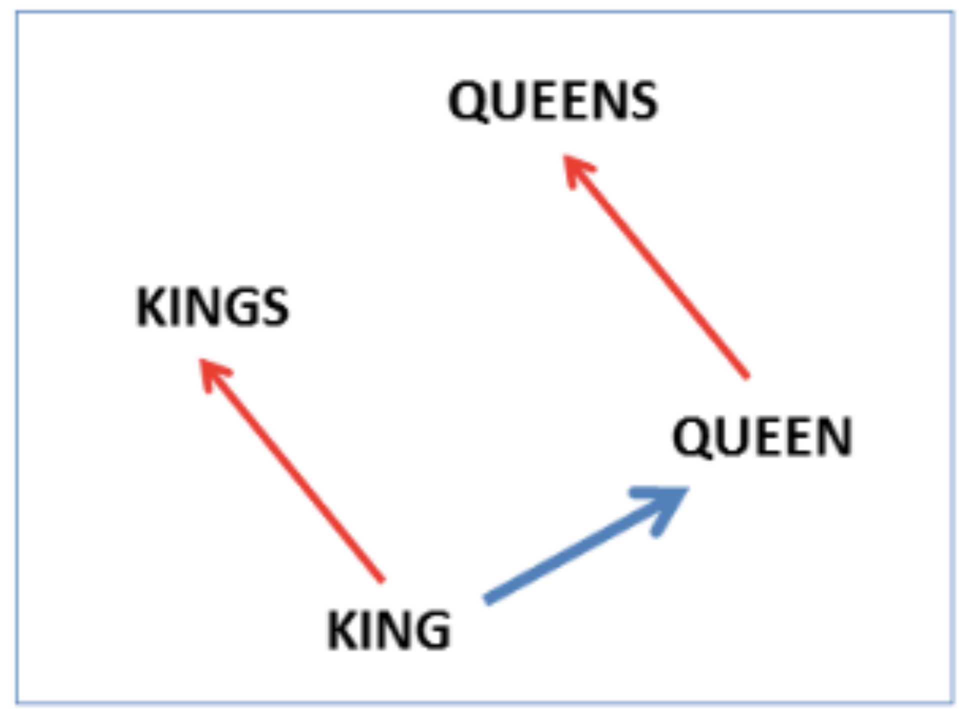
Knowledge Base Incompleteness

- Even w/ extremely large scale, knowledge bases are by nature incomplete
- e.g. in FreeBase 71% of humans were missing "date of birth" (West et al. 2014)
- Can we perform "relation extraction" to extract information for knowledge bases?

Remember: Consistency in Embeddings

e.g. king-man+woman = queen (Mikolov et al. 2013)



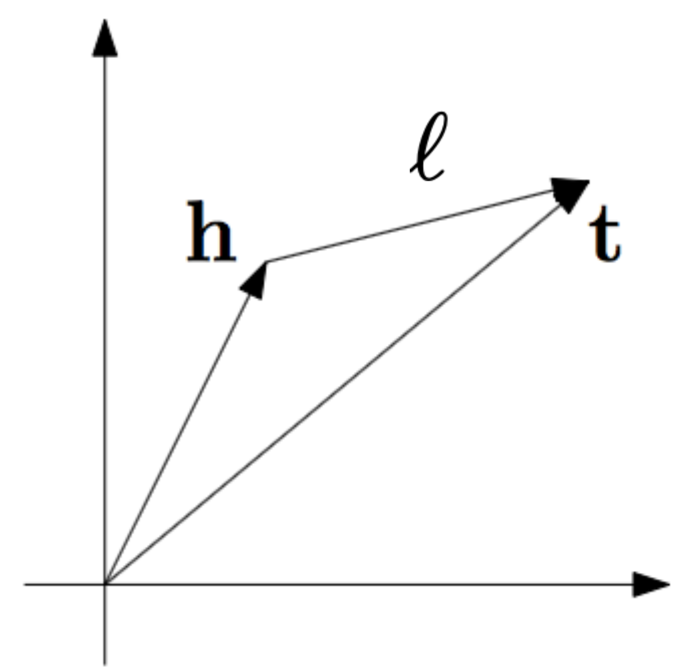


Learning Knowledge Graph Embeddings (Bordes et al. 2013)

- Motivation: express triples as additive transformation

- Method: minimize the distance of existin triples with a margin-based loss that

$$\sum_{(h,\ell,t)\in S} \sum_{(h',\ell,t')\in S'_{(h,\ell,t)}} \left[\gamma + d(\boldsymbol{h}+\boldsymbol{\ell},\boldsymbol{t}) - d(\boldsymbol{h'}+\boldsymbol{\ell},\boldsymbol{t'})\right]_+$$



(a) TransE

- Note: one vector for each relation, additive modification only, intentionally simpler than NTN

Relation Extraction w/ Neural Tensor Networks (Socher et al. 2013)

- A first attempt at predicting relations: a multi-layer perceptron that predicts whether a relation exists

$$u_R^T f\left(W_{R,1}e_1 + W_{R,2}e_2\right)$$

- Neural Tensor Network: Adds bi-linear feature extractors, equivalent to projections in space

$$g(e_1, R, e_2) = u_R^T f\left(e_1^T W_R^{[1:k]} e_2 + V_R \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} + b_R\right)$$

- Powerful model, but perhaps overparameterized!

Socher et al. 2013. Reasoning With Neural Tensor Networks for Knowledge Base Completion

Information Extraction w/ Pre-defined Relations

Pre-defined Relations

- Define a set of relations (a.k.a. schema) that we could extract for pairs of entities from text.

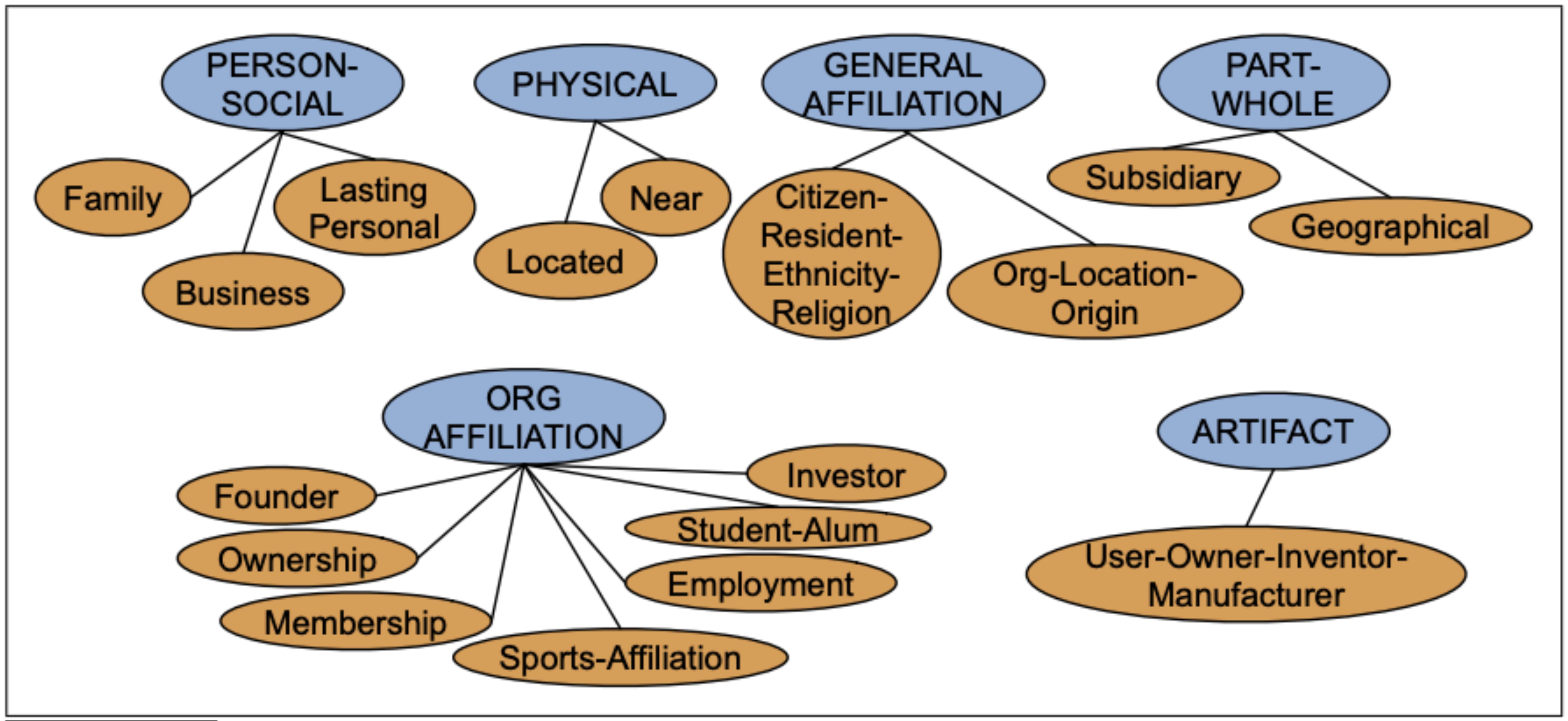


Figure 17.1 The 17 relations used in the ACE relation extraction task.

Relations	**Types**	**Examples**
Physical-Located	PER-GPE	**He** was in **Tennessee**
Part-Whole-Subsidiary	ORG-ORG	**XYZ**, the parent company of **ABC**
Person-Social-Family	PER-PER	**Yoko**'s husband **John**
Org-AFF-Founder	PER-ORG	**Steve Jobs**, co-founder of **Apple**...

Figure 17.2 Semantic relations with examples and the named entity types they involve.

Supervised Relation Extraction Baseline

- **Training**:
  - Labeled dataset: a KB triple t=<e1, r, e2> on a sentence s
  - Supervised training of models (e.g., logistic regression, NNs)
- **Test**:
  - Find any pairs of entities in a sentence
  - Apply the relation classifier on all entity pairs

function FINDRELATIONS(*words*) **returns** *relations*

relations ← *nil*
entities ← FINDENTITIES(*words*)
forall entity pairs $\langle e1, e2 \rangle$ **in** *entities* **do**
    if RELATED?($e1, e2$)
        relations ← *relations*+CLASSIFYRELATION($e1, e2$)

Distant Supervision for Relation Extraction (Mintz et al. 2009)

- **Motivation**: Supervised baseline is still limited to the labeled data size.
- Given an entity-relation-entity triple, extract all text that matches this and use it to train

[Steven Spielberg]'s film [Saving Private Ryan] is loosely based on the brothers' story.

Allison co-produced the Academy Award-winning [Saving Private Ryan], directed by [Steven Spielberg]...

- Extract hand-crafted features from this large corpus of (noisily) labeled text to train a system (e.g., multi-class logistic regression)

Relation Classification w/ CNNs (Zeng et al. 2014)

- Extract features w/o syntax using CNN
  - Lexical features of the words themselves
  - Features of the whole span extracted using convolution

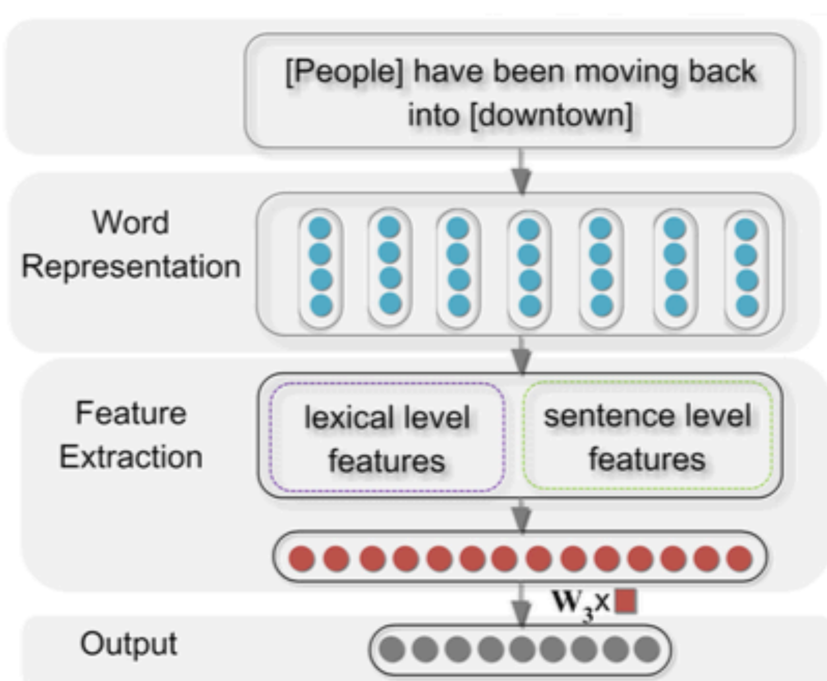


Figure 1: Architecture of the neural network used for relation classification.

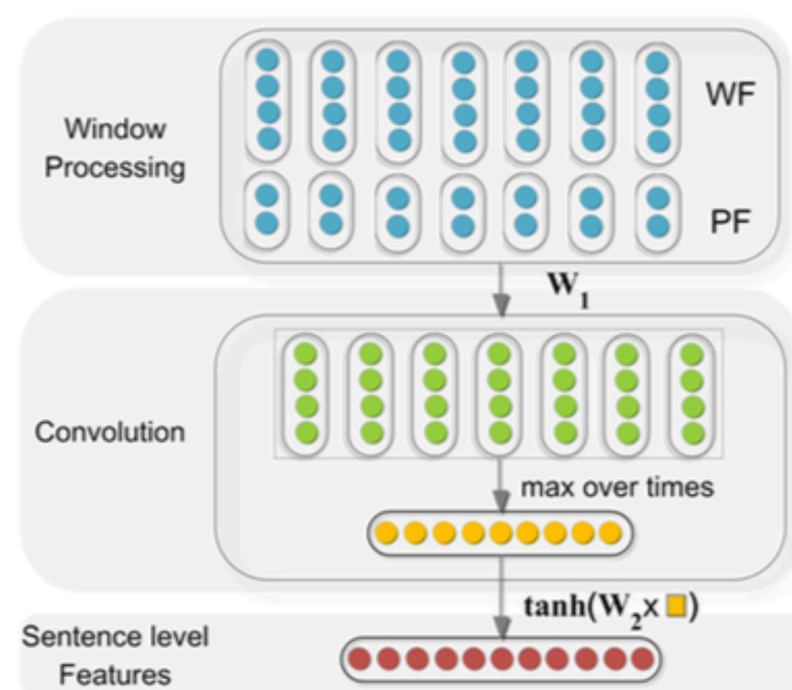


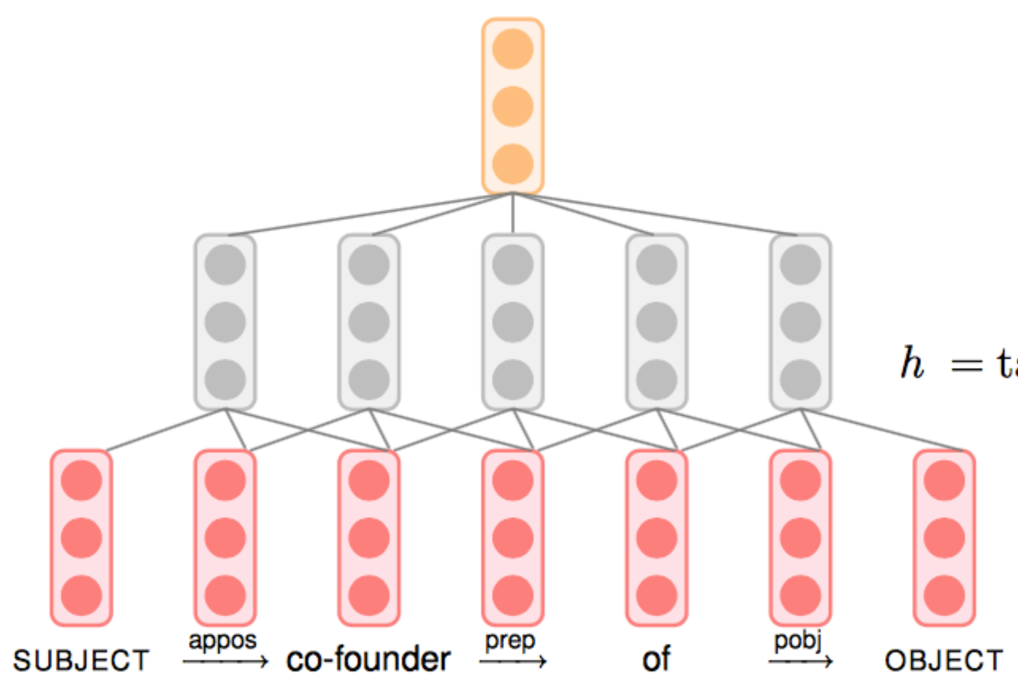
Figure 2: The framework used for extracting sentence level features.

Jointly Modeling KB Relations and Text (Toutanova et al. 2015)

- To model textual links between words w/ neural net: aggregate over multiple instances of links in dependency tree

Textual Pattern	**Count**
SUBJECT $\xrightarrow{\text{appos}}$ founder $\xrightarrow{\text{prep}}$ of $\xrightarrow{\text{pobj}}$ OBJECT	12
SUBJECT $\xleftarrow{\text{nsubj}}$ co-founded $\xrightarrow{\text{dobj}}$ OBJECT	3
SUBJECT $\xrightarrow{\text{appos}}$ co-founder $\xrightarrow{\text{prep}}$ of $\xrightarrow{\text{pobj}}$ OBJECT	3
SUBJECT $\xrightarrow{\text{conj}}$ co-founder $\xrightarrow{\text{prep}}$ of $\xrightarrow{\text{pobj}}$ OBJECT	3

- Model relations w/ CNN



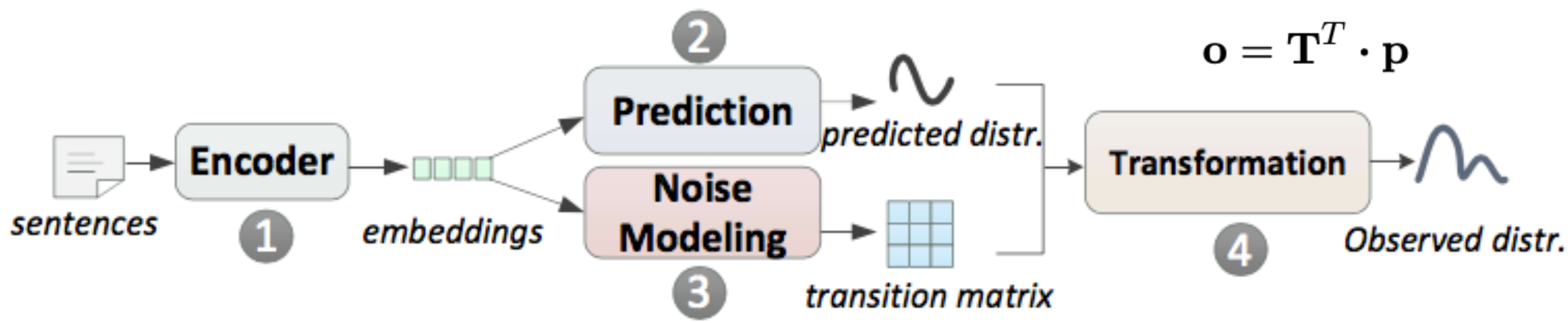
$$r = \max\{h\}$$

$$h = \tanh(W^{-1}v_{-1} + W^{0}v + W^{1}v_{+1} + b)$$

$$v = Ve$$

Modeling Distant Supervision Noise in Neural Models (Luo et al. 2017)

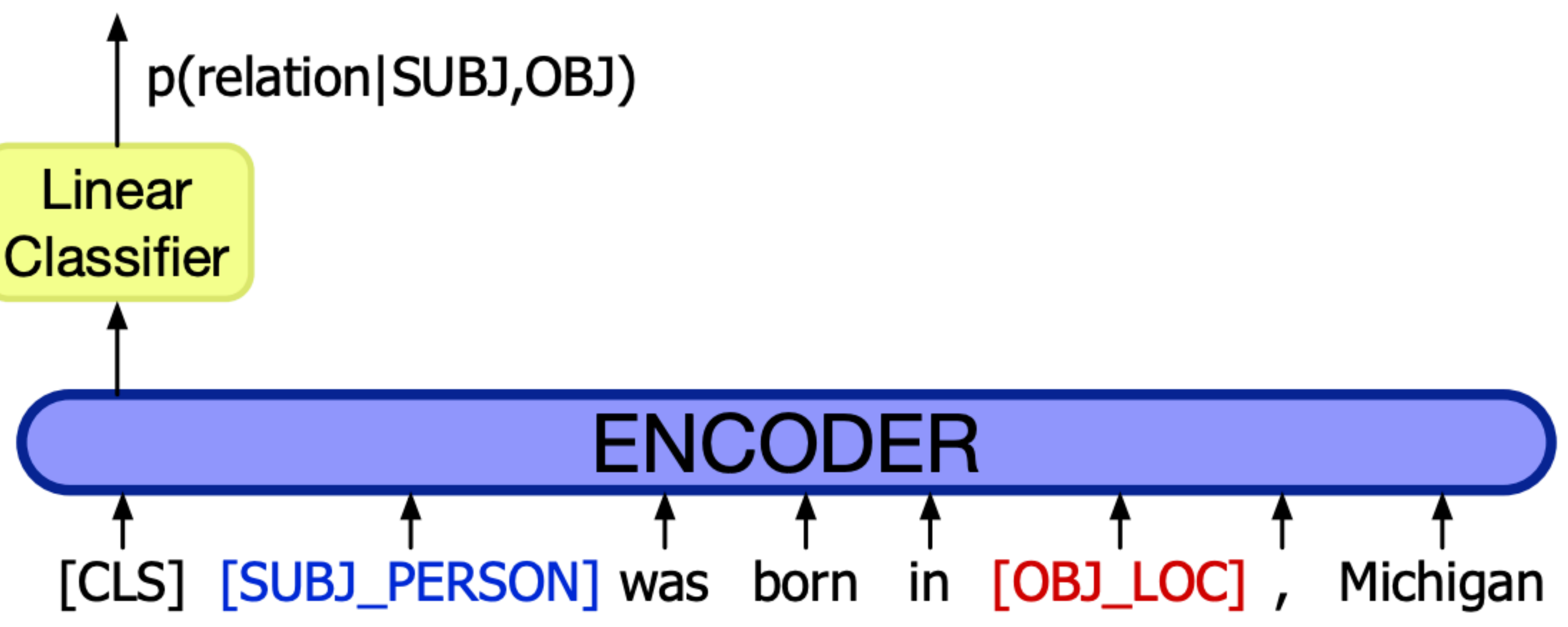
- Idea: there is noise in distant supervision labels, so we want to model it



- By controlling the "transition matrix", we can adjust to the amount of noise expected in the data
  - Trace normalization to try to make matrix close to identity
  - Start training w/ no transition matrix on data expected to be clean, then phase in on full data

Relation Extraction w/ Pre-trained LMs

- Relation extraction as a linear layer on top of an encoder (e.g., BERT), with the subject and object entities replaced in the input by their NER tags (Zhang et al. 2017, Joshi et al. 2020).



Schema-Free Extraction

Open Information Extraction
(Banko et al 2007)

- Basic idea: **the text is the relation. No pre-defined set of relation types!**

- e.g. "United has a hub in Chicago, which is the headquarters of United Continental Holdings"

    - {United; has a hub in; Chicago}

    - {Chicago; is the headquarters of; United Continental Holdings}

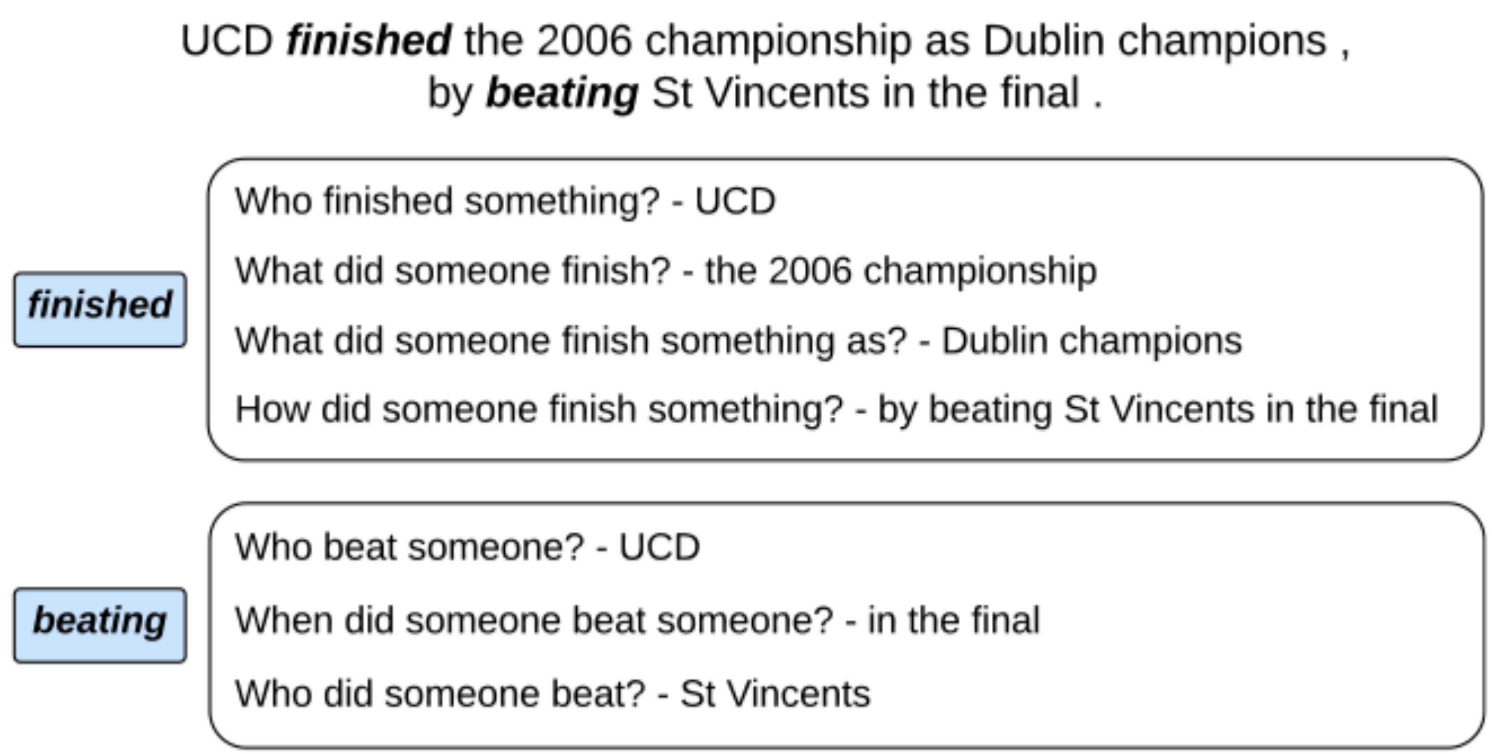
- Can extract any variety of relation strings, but does not abstract these relation strings to a relation type

Rule-based Open IE

- e.g. TextRunner (Banko et al. 2007), ReVerb (Fader et al. 2011)
- Use parser to extract according to rules
  - e.g. relation must contain a **predicate**, subject object must be **noun phrases**, etc.
- Train a fast model to extract over large amounts of data
- Aggregate multiple pieces of evidence (heuristically) to find common, and therefore potentially reliable, extractions

Neural Models for Open IE

- Unfortunately, heuristics are still not perfect
- Possible to create relatively large datasets by asking simple questions to annotators (He et al. 2015):



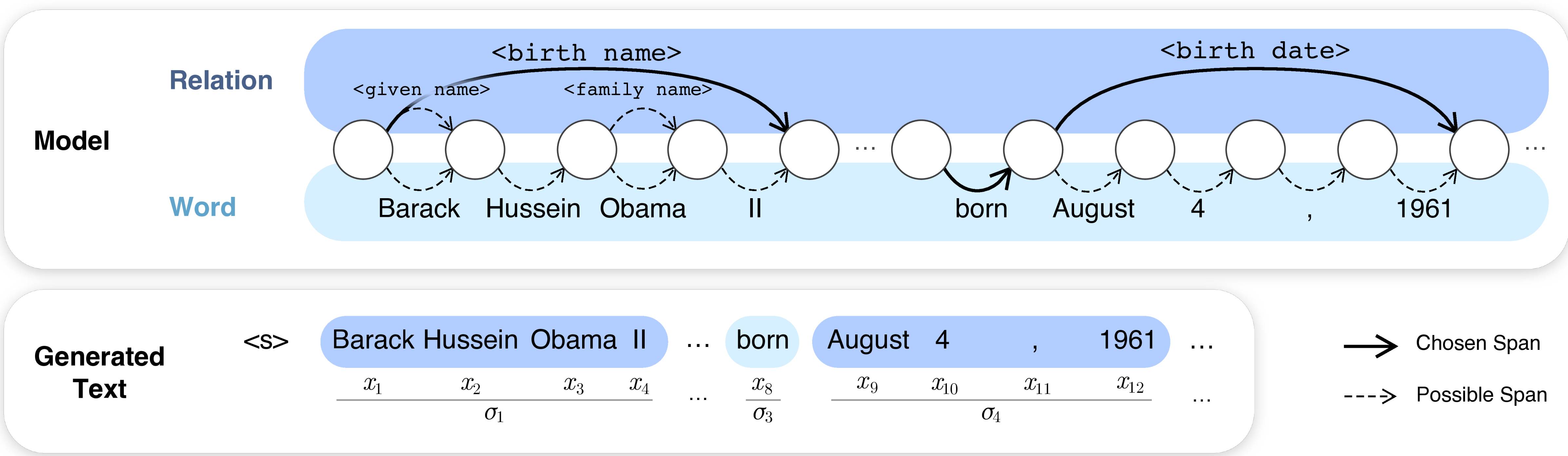
- Can be converted into OpenIE extractions

He et al. 2015. Question-Answer Driven Semantic Role Labeling: Using Natural Language to Annotate Natural Language

Using Knowledge Bases to Inform Neural Models

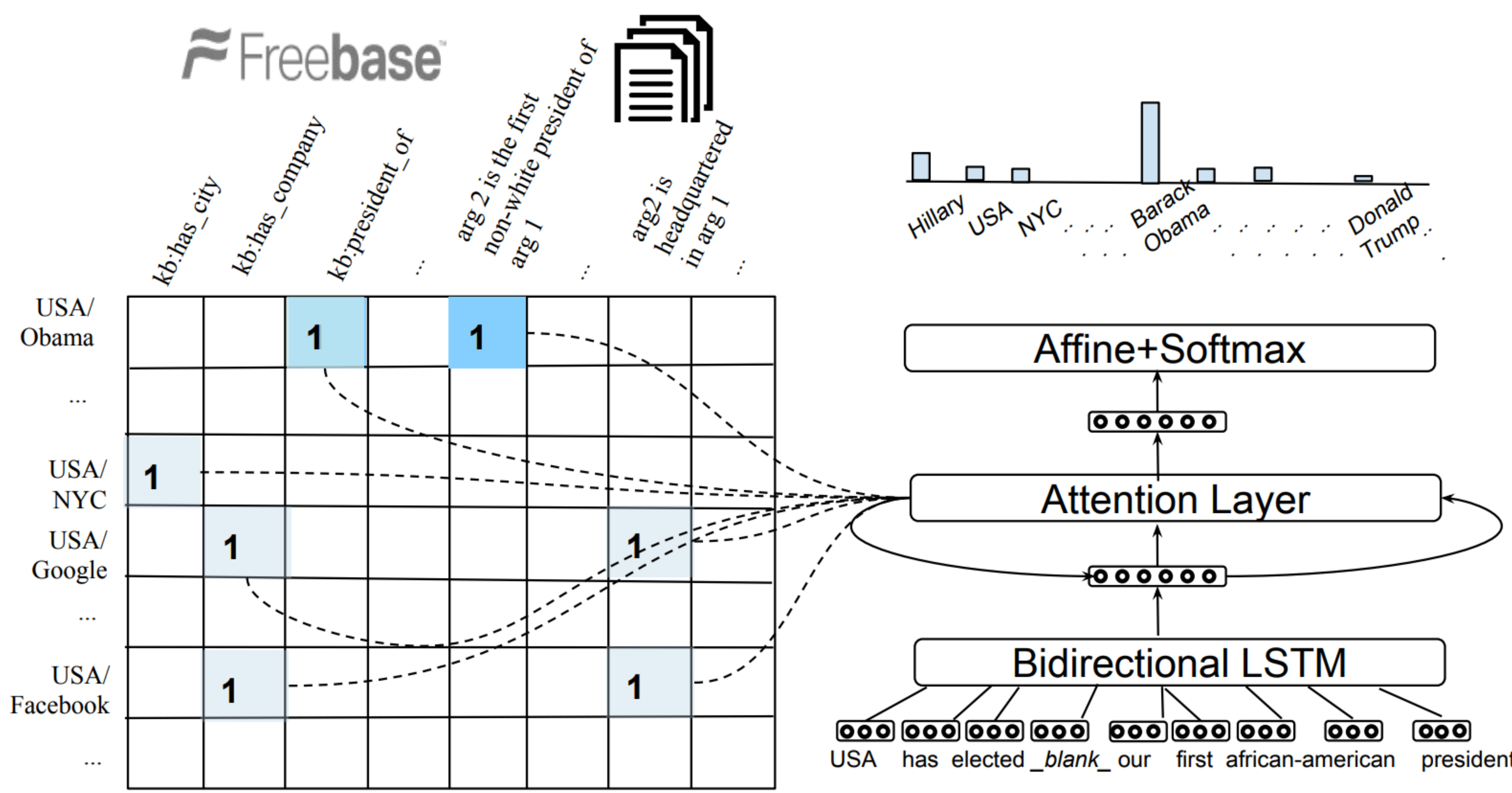
Injecting Knowledge into Language Models (Hayashi et al. 2020)

- Provide LMs with topical knowledge in the form of copiable graphs
    - Each (Wiki) text is given relevant KB taken from Wikidata
- Examine all possible decoding "paths" and maximize the marginal probability



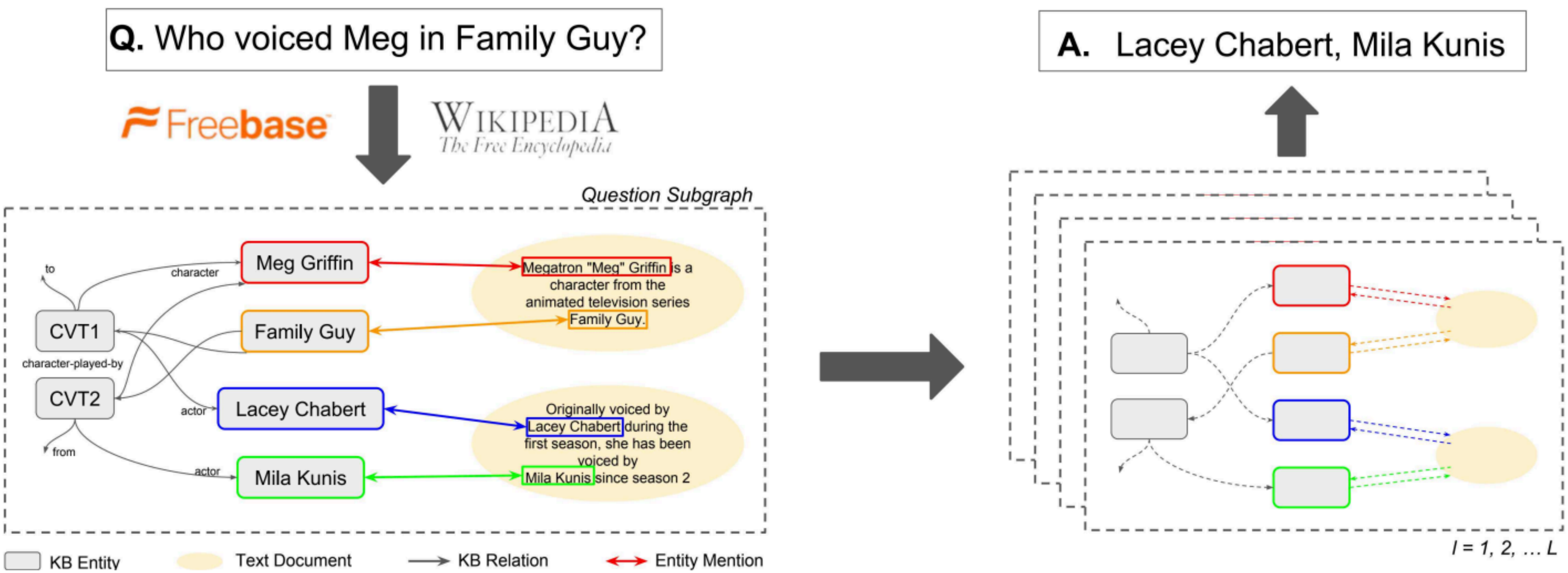
QA on KB w/ Universal Schema and Memory Network (Das et al. 2017)

- Represent each KB entity as a row in a memory matrix
- Use attention to retrieve relevant entities for QA



QA w/ KB and Text

- **Entity linking**: Link named entities extracted from text to their corresponding KB entities
- Use Graph NNs to encode the graph where each node in a graph can either be a KB entity or the entity with textual context.



Sun et al. 2018 Open Domain Question Answering Using Early Fusion of Knowledge Bases and Text

Probing Knowledge in LMs

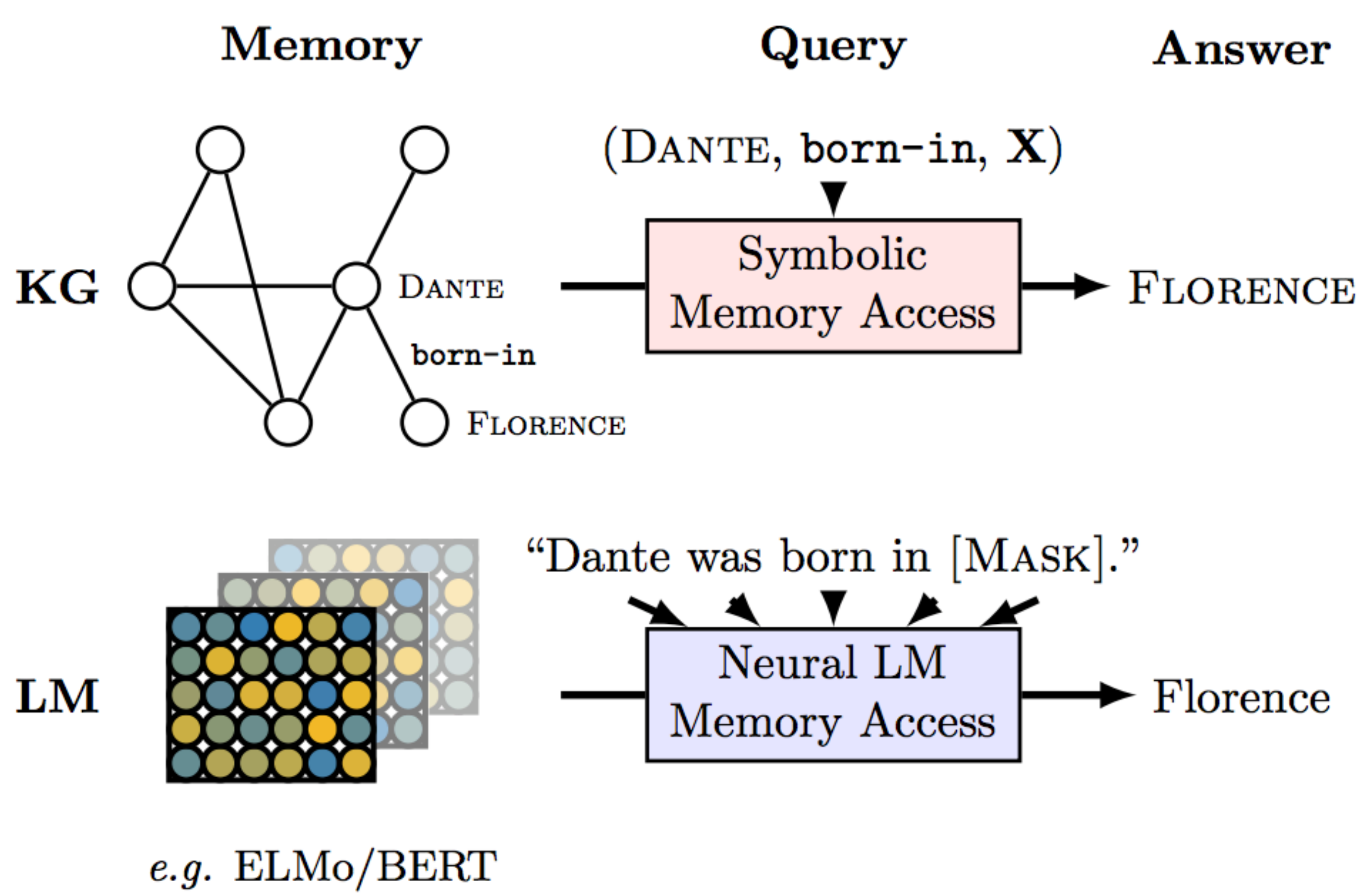
Probing Knowledge in LMs

- Traditional QA/MRC models usually refer to external resources to answer questions, e.g., Wikipedia articles or KGs.
- Do LMs pre-trained on a large text corpus already capture those knowledge?

LMs as KBs?

(Petroni et al. 2019)

- Structured queries (e.g., SQL) to query KBs.
- Natural language prompts to query LMs.



LMs as KBs?
(Petroni et al. 2019)

- LAMA benchmark
  - Manual prompts for 41 relations: “[X] was founded in [Y].”
  - Fill in subjects and have LMs (e.g., BERT) predict objects: “Bloomberg L.P. was founded in [MASK].”
  - Accuracy: ELMo 7.1%, Transformer-XL 18.3%, BERT-base 31.1%

Mask 1 Predictions:

5.2% **Chicago**

4.1% **London**

2.8% **Toronto**

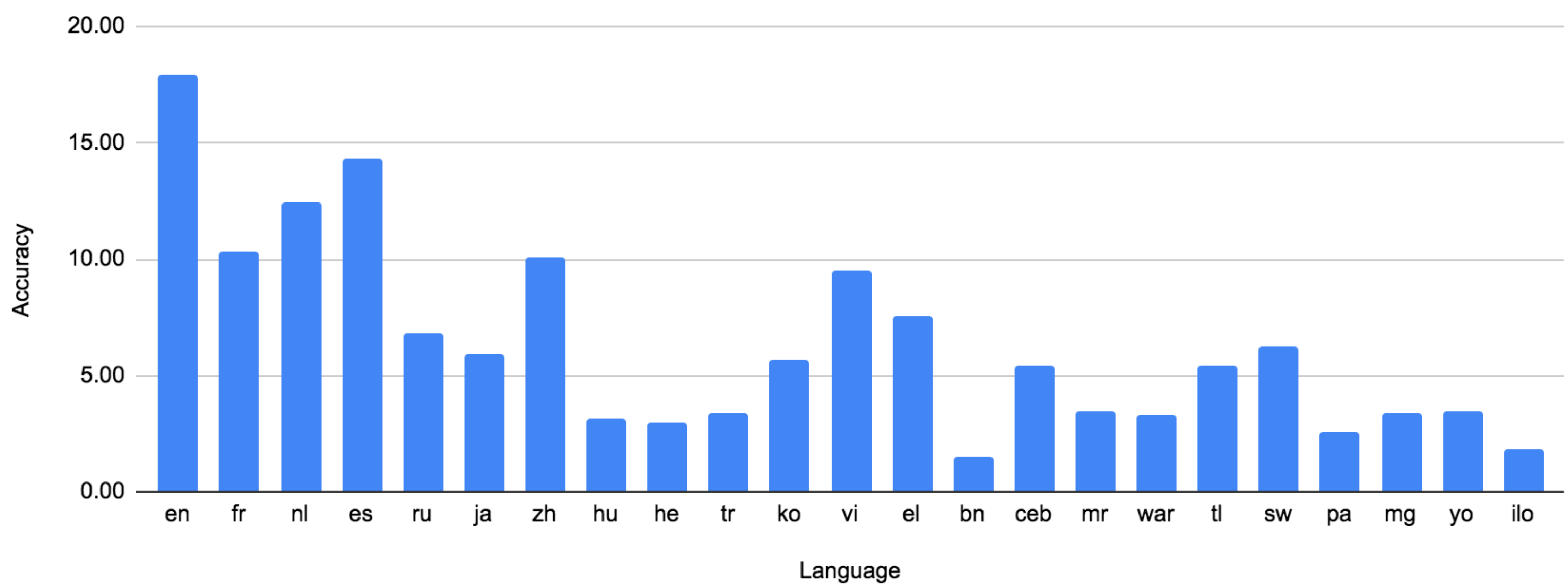
2.3% **c**

1.6% **India**

https://demo.allennlp.org/masked-lm/s/bloomberg-lp-was-founded-mask/I5Q1P2T5Z0

X-FACTR: Multilingual Factual Knowledge Probing (Jiang et al. 2020)

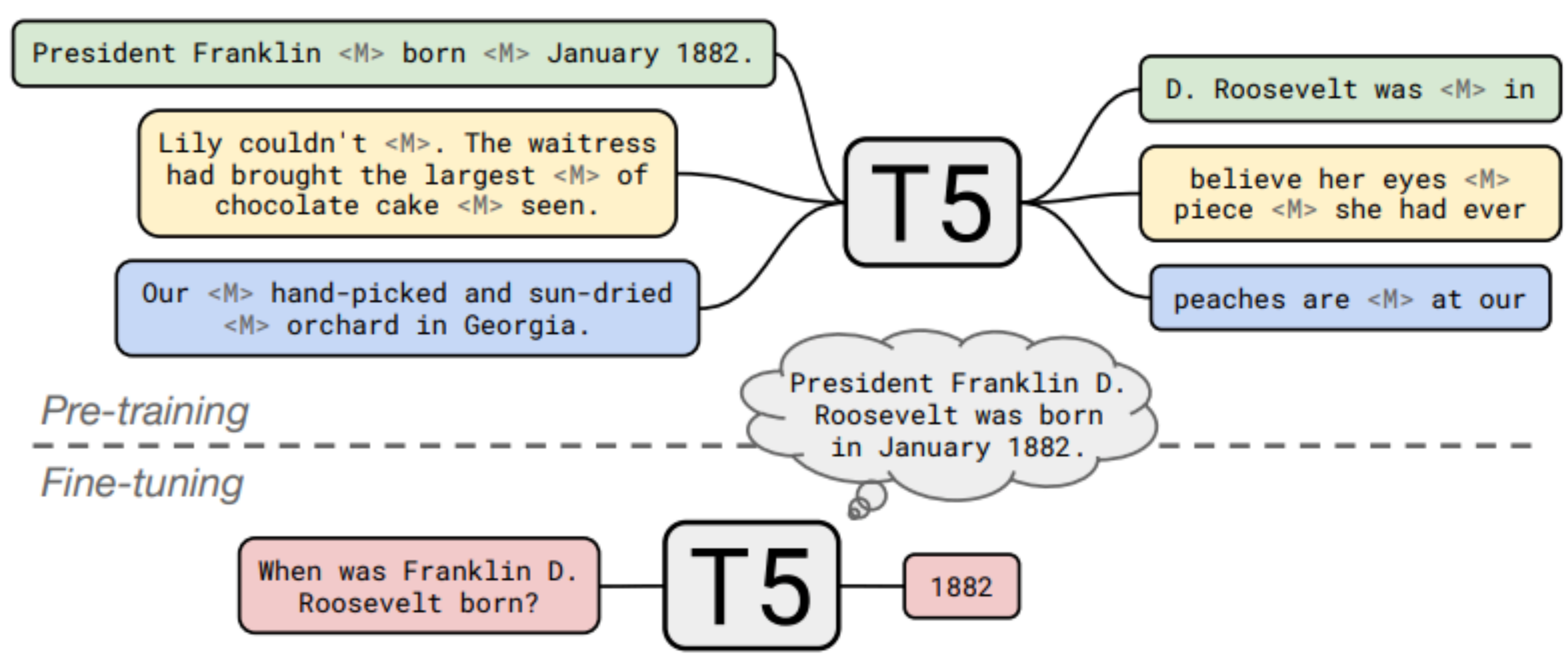
- Overall, factual knowledge in LMs is still limited, especially for low-resource languages.



Max performance of M-BERT, XLM, XLM-R

Close-book T5: Directly Fine-tune with QA Pairs (Roberts et al. 2020)

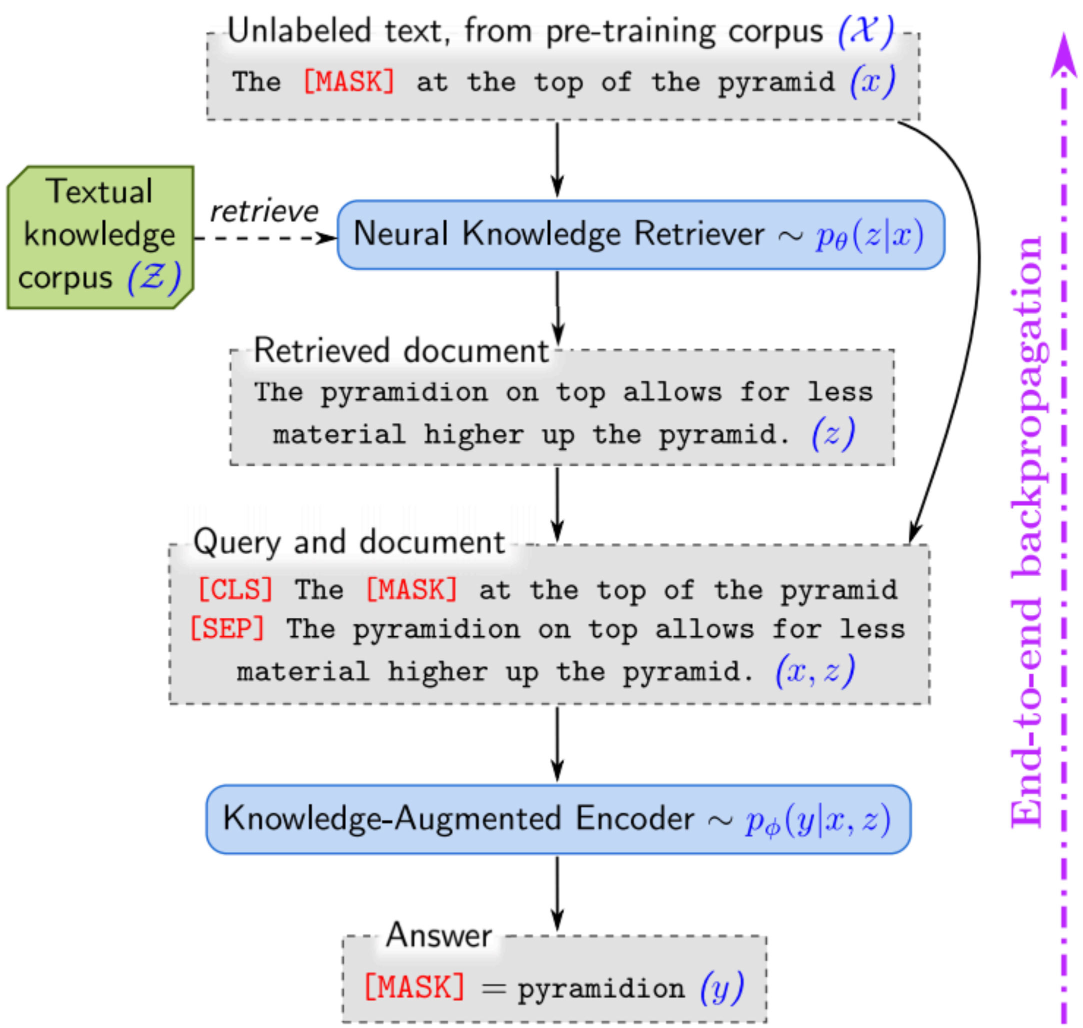
- Generate answers given questions without additional context.
- Performs even better than QA models with retrieved context.



Nonparametric Models Outperform Parametric Models

- For knowledge-intensive tasks like QA, nonparametric models (w/ retrieved context) outperform parametric models (w/o context) by a large margin.
- For example, REALM (Guu et al. 2020), RAG (Lewis et al. 2020) on the NaturalQuestion datasets.

Close-book T5	34.5
REALM	40.4
RAG	44.5



Questions?