

CS291K – Advanced Data Mining

Instructor: Xifeng Yan
Computer Science
University of California at Santa Barbara

Semantic Parsing and Question Answering

Lecturer: Izzeddin Gur
Computer Science
University of California at Santa Barbara

Source of slides

- Percy Liang from Stanford
- MSR
- Images from Despicable Me



Question Answering

Who did Humphrey Bogart marry in 1928?



?



Web Search

Google who did humphrey bogart marry in 1928

Web Images Maps Shopping More ▾ Search tools

About 3,070,000 results (0.33 seconds)

Humphrey Bogart - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Humphrey_Bogart ▾
Humphrey Bogart by Karsh (Library and Archives Canada).jpg Bogart's birthday has been a subject of controversy; according to Warner Bros, On April 3, 1928, he married Mary Philips at her mother's apartment in Hartford, Connecticut.

Humphrey Bogart's fears that he was gay 'almost drove him to ...
www.dailymail.co.uk/.../Humphrey-Bogarts-fears-gay-drove-suicide.htm... ▾
Sep 27, 2010 - Now it has emerged that before marrying her in 1945, Bogie had bedded ... the night before their marriage in 1928 with a friend of Bogart's 'as a ...

Who Was Married to Humphrey Bogart - Ask.com
[www.ask.com/.../Arts and Humanities/Theater and Performing Arts](https://www.ask.com/.../Arts_and_Humanities/Theater_and_Performing_Arts) ▾
Helen Menken was married to Humphrey Bogart from 1926 to 1927 when they divorced. He then married Mary Phillips in 1928 but divorced her in 1937. In 1938 he married. ... Who did Humphrey Bogart marry?

HUMPHREY & BOGEY - Psykick Girl Productions
www.psykickgirl.com/lulu/bogey.html ▾
HUMPHREY BOGART SPENT THE LAST twenty-one years of his life that after Humphrey was married and divorced by Helen Menken he would marry Mary Except for a two-week revival of Saturday's Children in 1928, Humphrey did not ...

Who is Humphrey Bogart? - wiseGEEK
www.wisegEEK.com/who-is-humphrey-bogart.htm ▾



Semantic Parsing

Who did Humphrey Bogart marry in 1928?



semantic parsing

Type.Person \sqcap Marriage.(Spouse.HumphreyBogart \sqcap StartDate.1928)



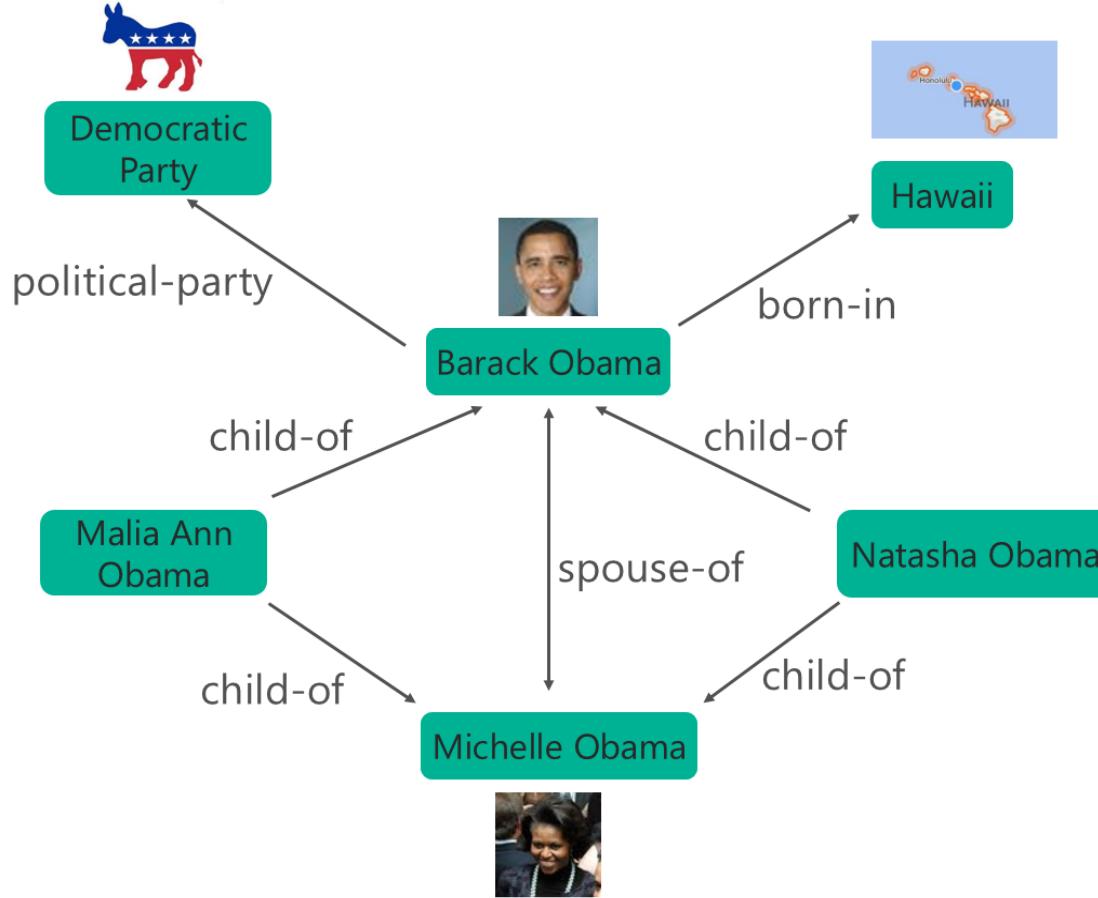
execute logical form

Mary Philips

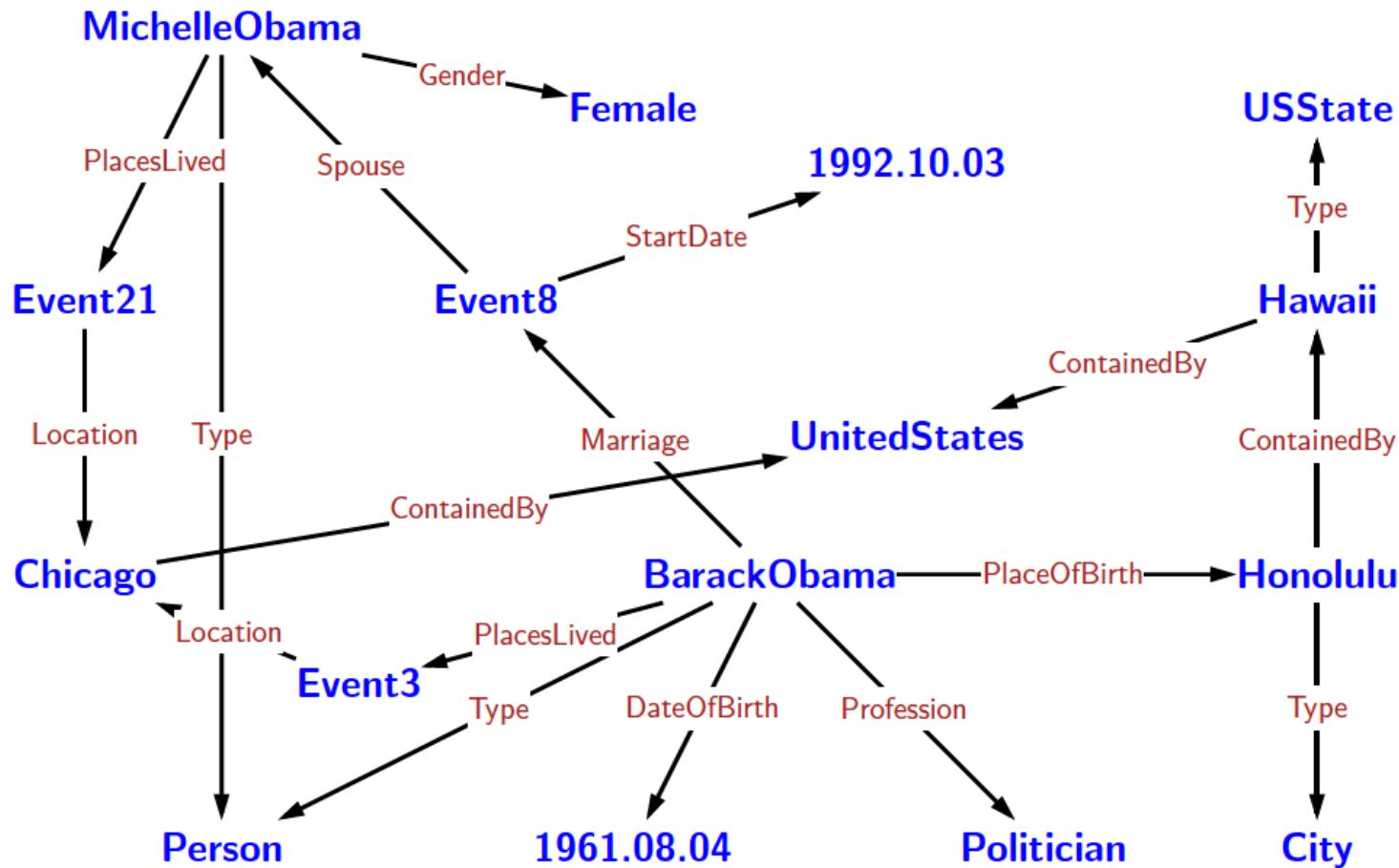
Motivation : Natural Language interface to large structured knowledge bases



Knowledge Bases



Freebase



Traditional Semantic Parsing

Supervision: manually annotated logical forms

What's California's capital?

Capital.California

How long is the Mississippi river?

RiverLength.Mississippi

...

...

Hard to obtain “ground truth” logical forms



Weakly Supervised Parsers

Supervision: question/answers pairs

What's California's capital?

Sacramento

How long is the Mississippi river?

3,734km

...

...

Easy to obtain
More natural



Logical Forms

Who lived in Seattle?

Lambda Calculus :

$$\lambda x. \exists e. \text{PlacesLived}(x, e) \wedge \text{Location}(e, \text{Seattle})$$

Dependency Based Compositional Semantics (Lambda-DCS) :

PlacesLived.Location.Seattle



Lambda DCS

Unary Base Case : Seattle

Binary Base Case : PlaceOfBirth

Join : PlaceOfBirth.Seattle

Intersection : Profession.Scientist \sqcap PlaceOfBirth.Seattle

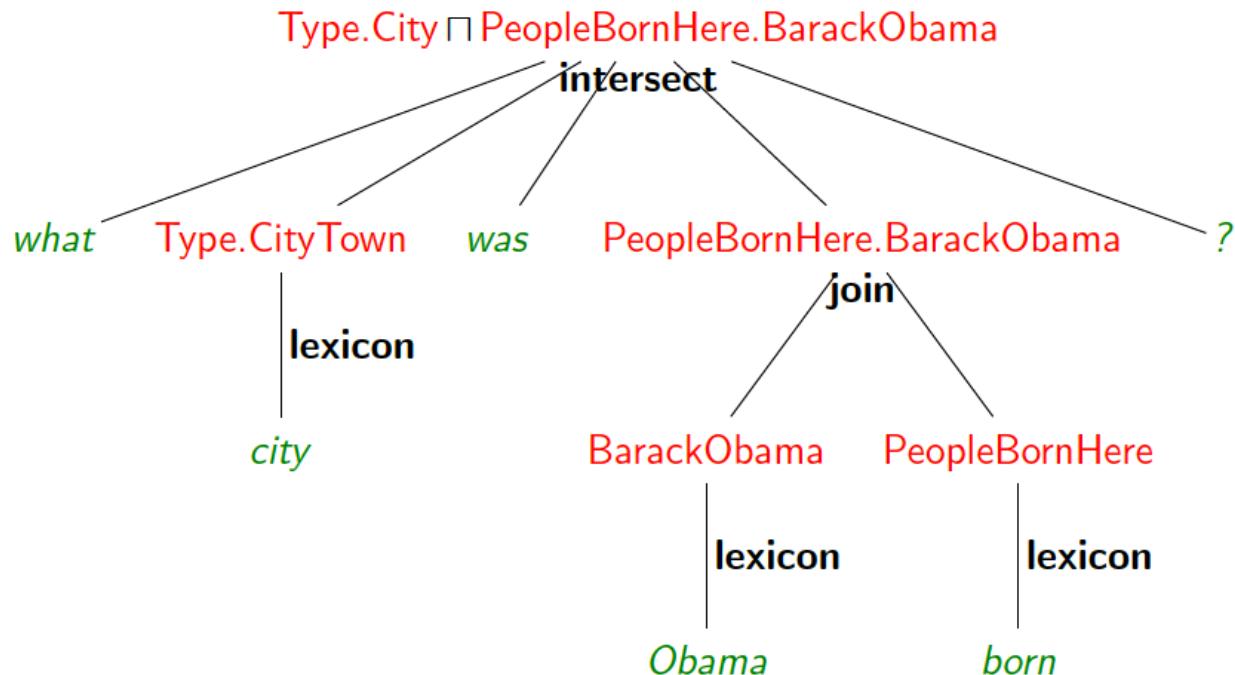
Union : Oregon \sqcup Washington \sqcup Type.CanadianProvince

Join : Type.USSState \sqcap \neg Border.California

Higher Order Functions : $count(Type.USSState)$



Traditional Semantic Parsing



Large-scale **Lexicon** construction is difficult



Incomplete Knowledge Base

St. Patrick's day happens on

March 17th

HolidayDate

St. Patrick's day occurs on

March 17th

StPatricksDay March 17th

Purim happens on

March this year

Purim Adar 14th

Purim occurs on

March

Spring break happens usually around March

Spring break usually occurs around March

happen \Rightarrow HolidayDate (1)

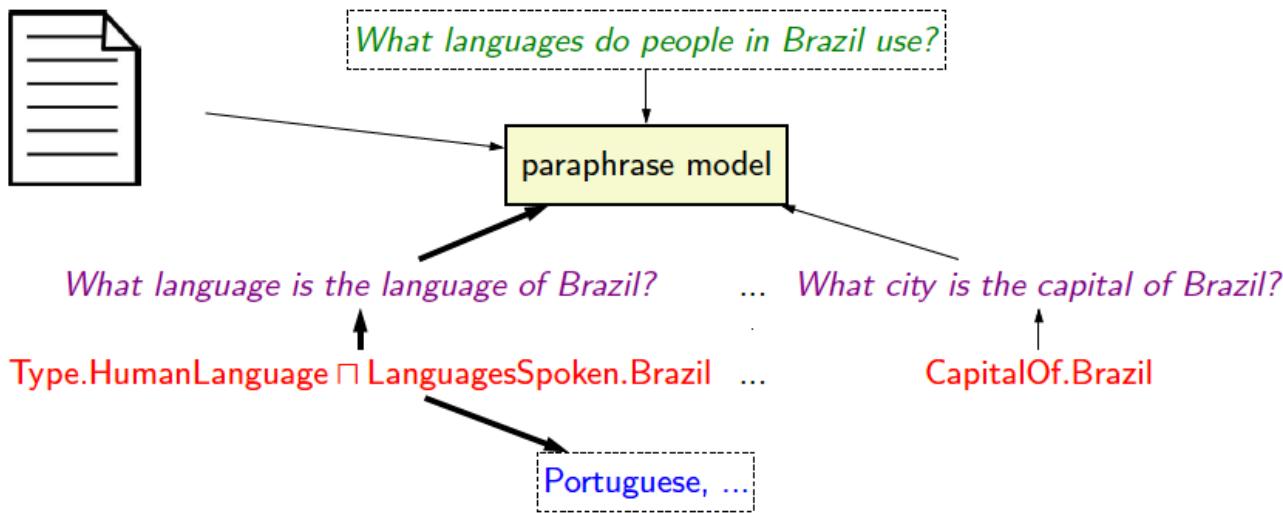
occur \Rightarrow HolidayDate (1)

occur \Leftrightarrow *happen* (3)

No corresponding ***KB Relation*** for every ***Natural Relation***



Semantic Parsing via Paraphrasing



Generate candidate logical forms
Generate template based paraphrases
Rank generated canonical utterances



Setup

Input:

- Knowledge-base \mathcal{K}
- Training set of question-answer pairs $\{(x_i, y_i)\}_1^n$

What are the main cities in California? SF, LA, ...

Output:

- Semantic parser that maps questions x to answers y through logical forms z

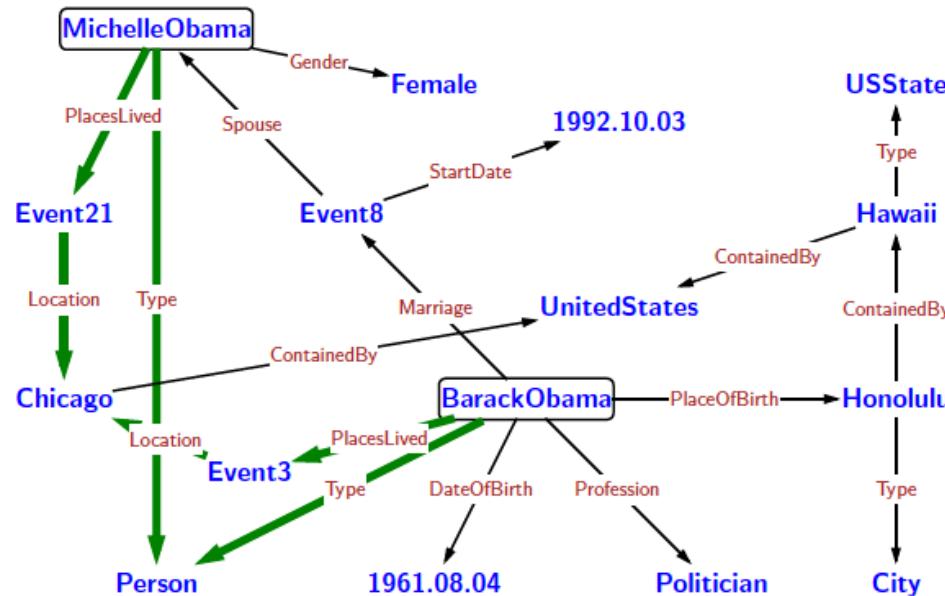
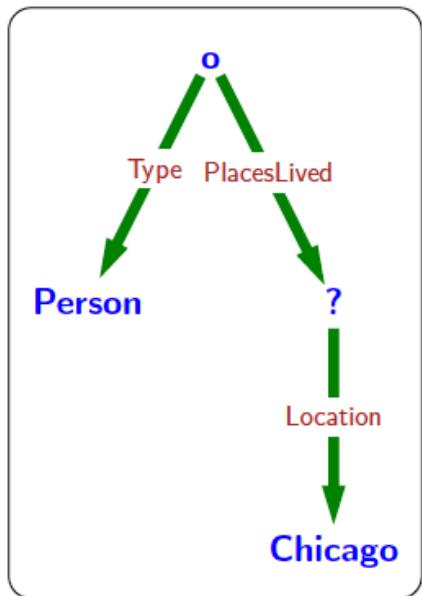
countries in Asia \Rightarrow Type.Country \sqcap ContainedBy.America

\Rightarrow China, Japan, India, ...



Logical Forms

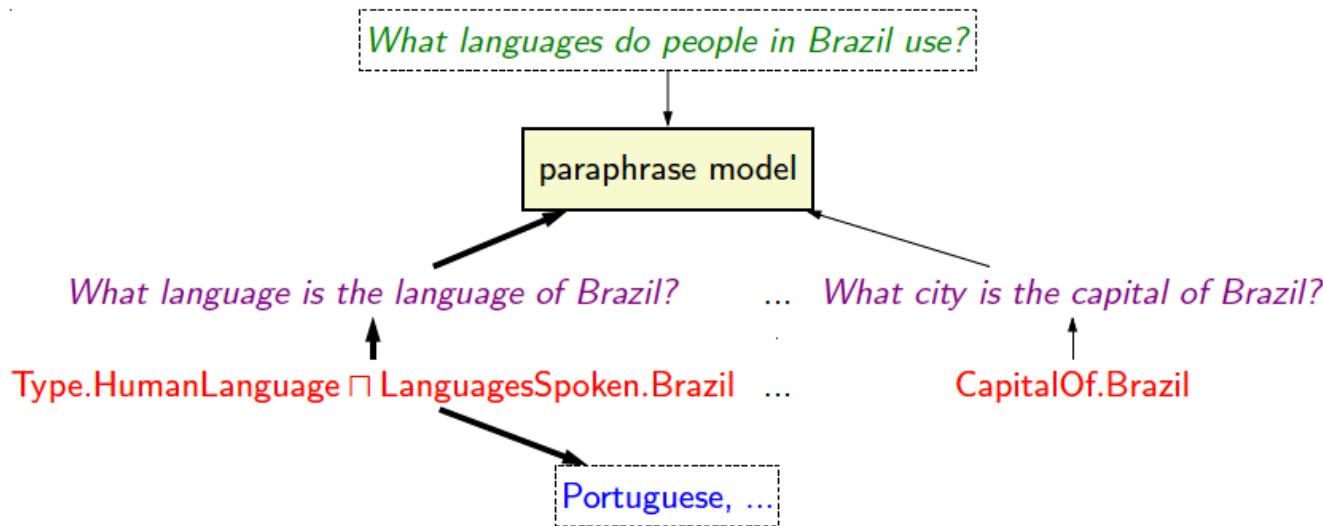
Type.Person \sqcap PlacesLived.Location.Chicago



Logical Form == Graph Template



Overview



x : input question

Z_x : candidate logical forms

C_z : generated canonical utterances

y : answer



Model

Model: distribution over logical forms and canonical utterances

$$p_{\theta}(c, z \mid x) = \frac{\exp(\phi(x, c, z)^{\top} \theta)}{\sum_{z' \in Z_x, c' \in C_z} \exp(\phi(x, z', c')^{\top} \theta)}$$

Decomposition to paraphrase model and logical form model:

$$\phi(x, c, z) = \phi_{\text{pr}}(x, c) + \phi_{\text{lf}}(z)$$

Need to estimate parameters θ_{pr} and θ_{lf}



Learning

Training data: $\{(x_i, y_i)\}_{i=1}^n$

Objective function:

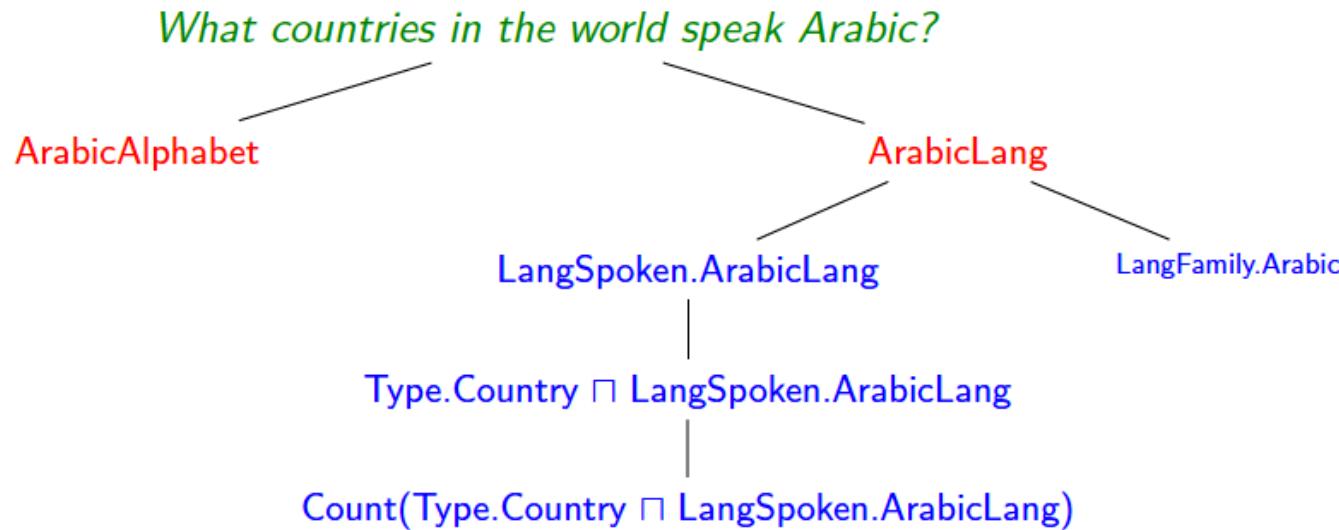
$$p_{\theta}(y \mid x) = \sum_{z \in Z_x : y=[z]_{\mathcal{K}}} \sum_{c \in C_z} p_{\theta}(c, z \mid x)$$

$$O(\theta) = \sum_{i=1}^n \log p_{\theta}(y_i \mid x_i) - \lambda \|\theta\|_1$$



Generating Candidate Logical Forms

Grow logical forms around entities:



Generating Candidate Logical Forms

Template	Example	Question
$p.e$	Directed.TopGun	<i>who directed Top Gun</i>
$p_1.p_2.e$	Employment.EmployerOf.SteveBalmer	<i>Where does Steve Balmer work?</i>
$p.(p_1.e_1 \sqcap p_2.e_2)$	Character.(Actor.BradPitt \sqcap Film.Troy)	<i>Who did Brad Pitt play in Troy?</i>
Type. $t \sqcap z$	Type.Composer \sqcap SpeakerOf.French	<i>What composers spoke French?</i>
count(z)	count(BoatDesigner.NatHerreshoff)	<i>How many ships were designed by Nat Herreshoff?</i>

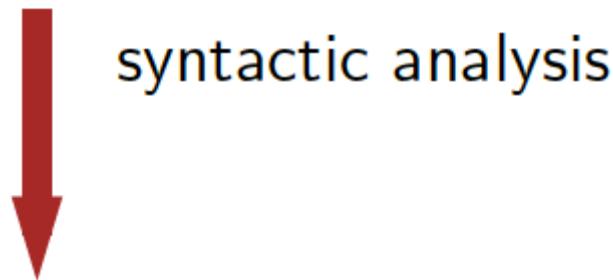
Detect entities => Use templates to generate logical forms



Canonical Utterance Generation

Type.Country \sqcap LangSpoken.ArabicLang

country spoken in languages spoken Arabic language



What country is Arabic language spoken in?

What country spoken the languages Arabic language?



Canonical Utterance Generation

	$d(p)$ Categ.	Rule	Example
$p.e$	NP	WH $d(t)$ has $d(e)$ as NP ?	<i>What election contest has George Bush as winner?</i>
	VP	WH $d(t)$ (AUX) VP $d(e)$?	<i>What radio station serves area New-York?</i>
	PP	WH $d(t)$ PP $d(e)$?	<i>What beer from region Argentina?</i>
	NP VP	WH $d(t)$ VP the NP $d(e)$?	<i>What mass transportation system served the area Berlin?</i>
$R(p).e$	NP	WH $d(t)$ is the NP of $d(e)$?	<i>What location is the place of birth of Elvis Presley?</i>
	VP	WH $d(t)$ AUX $d(e)$ VP ?	<i>What film is Brazil featured in?</i>
	PP	WH $d(t)$ $d(e)$ PP ?	<i>What destination Spanish steps near travel destination?</i>
	NP VP	WH NP is VP by $d(e)$?	<i>What structure is designed by Herod?</i>

Use syntactic category => Logical Form Template => Utterance



Paraphrasing

At this point our problem involves text only:

What countries in the world speak Arabic?

What country is Arabic language spoken in?

Simple paraphrase model utilizing **a lot of text**

- Association model - Paralex
- Vector space model - Wikipedia

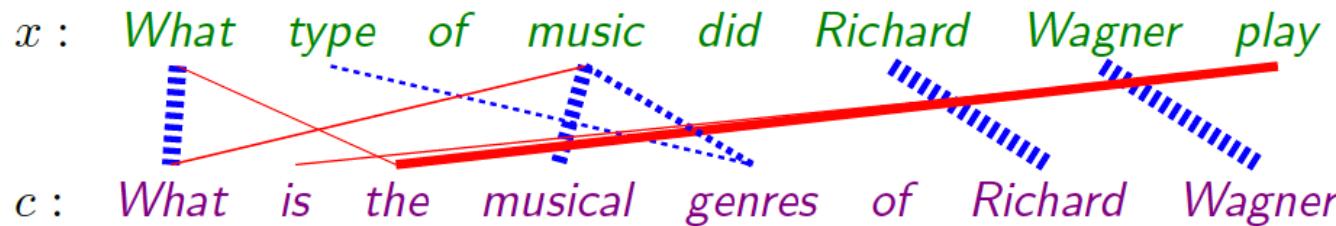
$$\phi_{\text{pr}}(x, c) = \phi_{\text{as}}(x, c) + \phi_{\text{vs}}(x, c)$$



Association Model

Association: pair of spans $(x_{ij}, c_{i'j'})$

type of music \Leftrightarrow musical genre



Generate all associations and extract features:

identical lemma	3
type of music \wedge musical genre	1
play \wedge the	1
WN derivation	1
delete IN	1
delete of	1
...	...

Use large parallel corpora : **Paralex**



Vector Space Model

Associations disadvantage: coverage

Train word vectors $v(w)$:

C : content words in utterance x

$$v(x) = \frac{1}{|C|} \sum_{x_i \in C} v(x_i)$$

Learn a matrix W to estimate “similarity” score

$$s(x, c) = v(x)^\top W v(c)$$

Options for W

- Identity: dot product
- Diagonal: dot product with scaling
- Full matrix: interactions between dimensions



Association vs Vector Space

x : What type of music did Richard Wagner play?

as : What is the musical genres of Richard Wagner?

vs : What composition has Richard Wagner as lyricist?

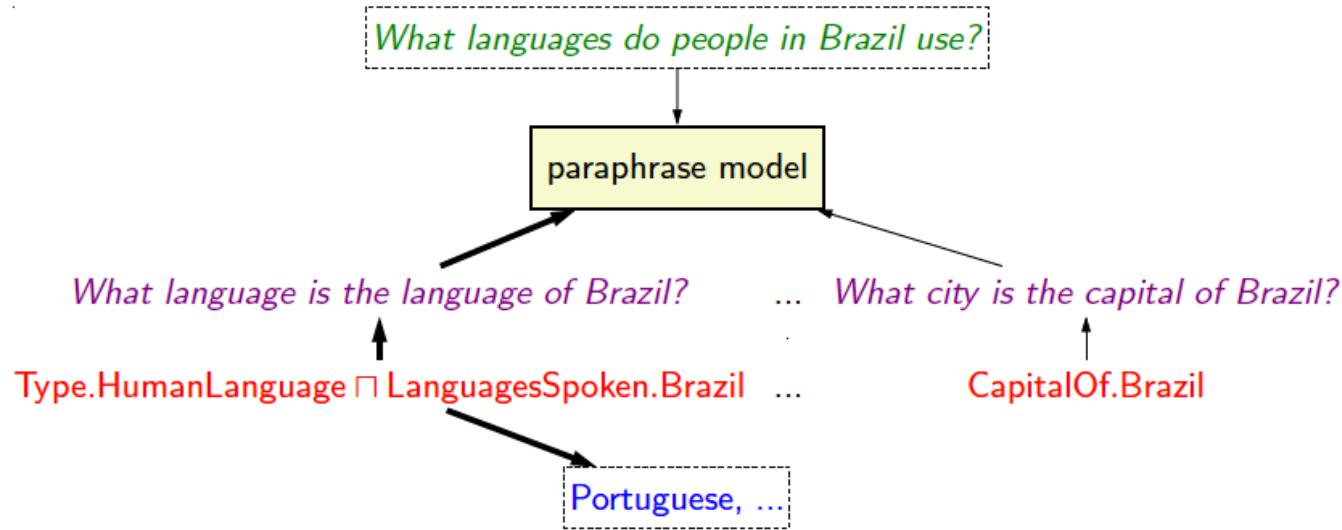
x : Where is made Kia car?

as : What place is founded by Kia motors?

vs : What city is Kia motors a headquarters of?



Recap



x : input question

Z_x : candidate logical forms

C_z : generated canonical utterances

y : answer



Dataset : WebQuestions

What character did Natalie Portman play in Star Wars? ⇒ Padmé Amidala

What kind of money to take to Bahamas? ⇒ Bahamian dollar

What currency do you use in Costa Rica? ⇒ Costa Rican colón

What did Obama study in school? ⇒ political science

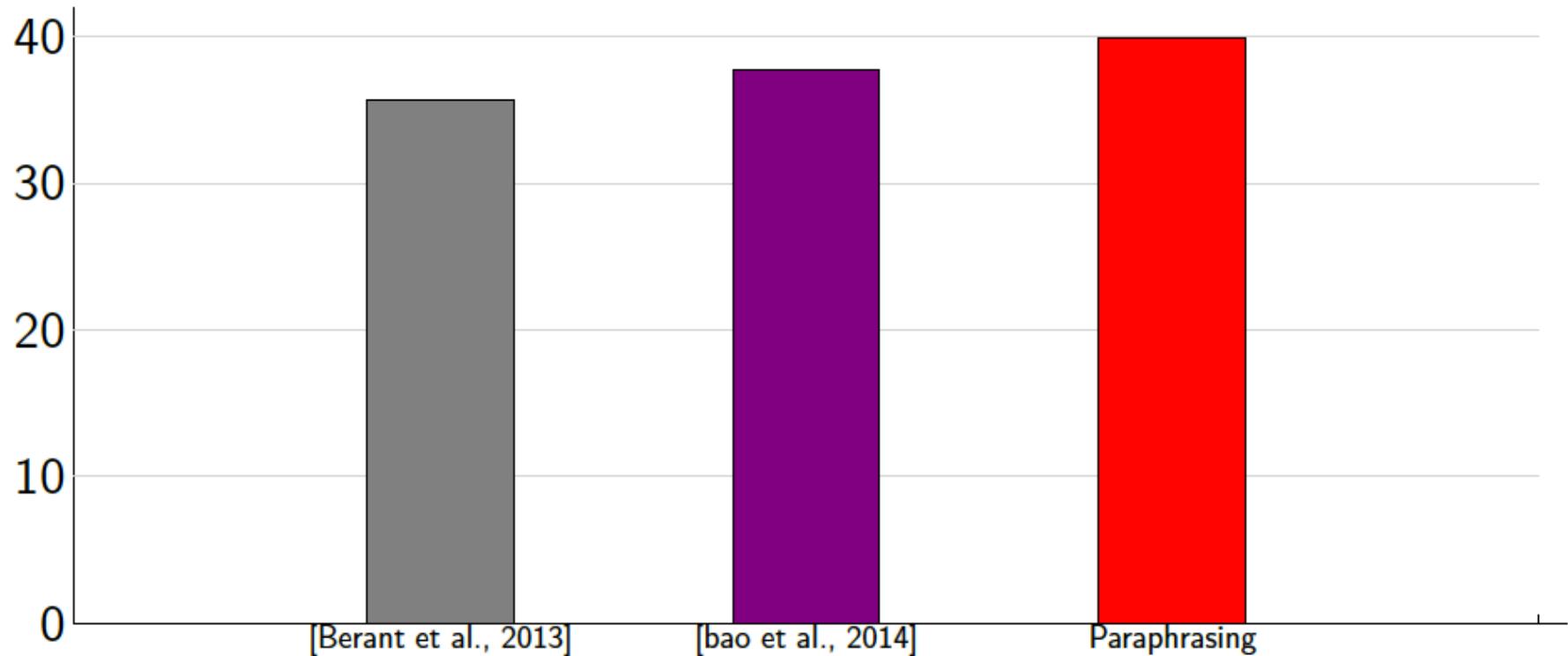
What do Michelle Obama do for a living? ⇒ writer, lawyer

What killed Sammy Davis Jr? ⇒ throat cancer

5,810 questions crawled from Google Suggest and answered using AMT



Results



Semantic Parsing via Staged Graph Query Generation

Query graph

Resembles subgraphs of the knowledge base

Can be *directly* mapped to a logical form

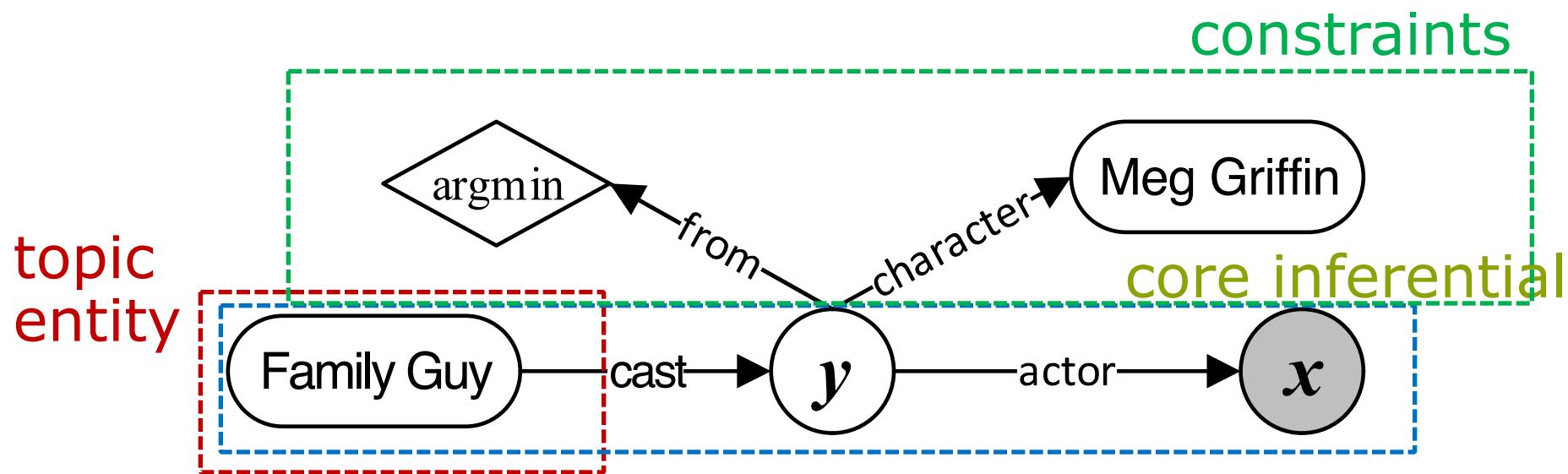
Semantic parsing

A search problem that *grows* the graph through *staged* state-actions



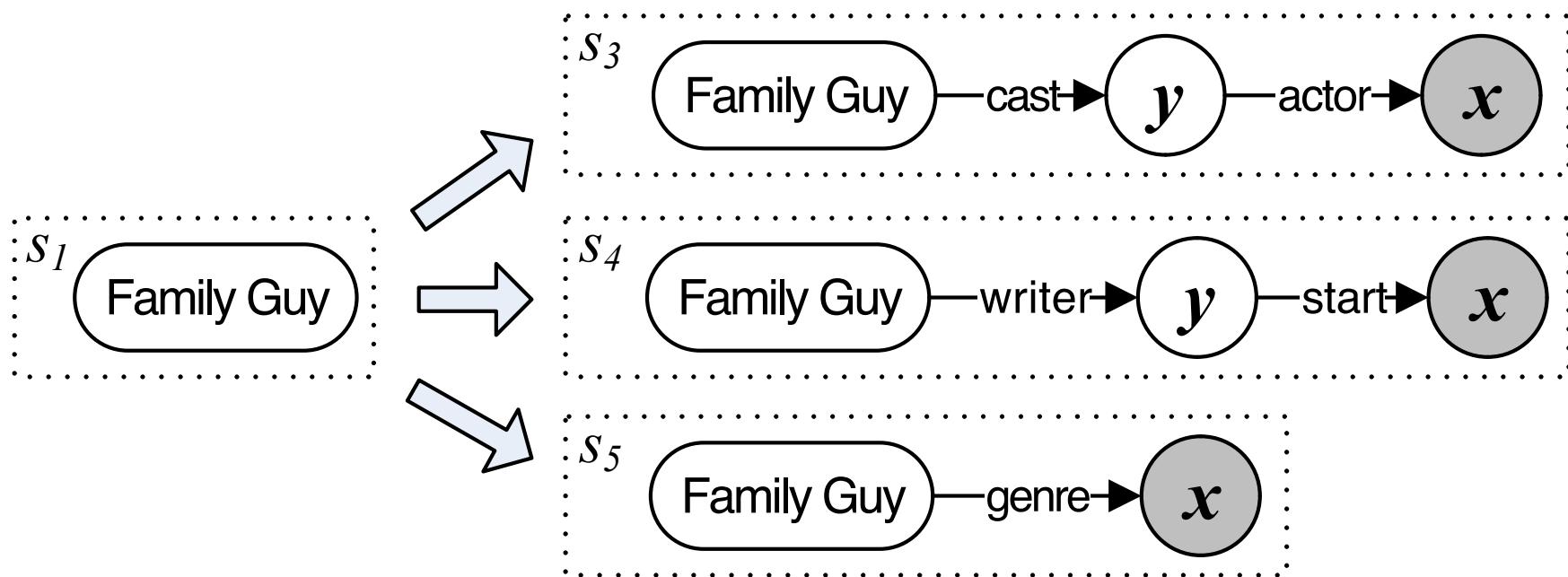
Query Graph

Who first voiced Meg on Family Guy?

$$\lambda x. \exists y. \text{cast}(\text{FamilyGuy}, y) \wedge \text{actor}(y, x) \wedge \text{character}(y, \text{MegGriffin})$$


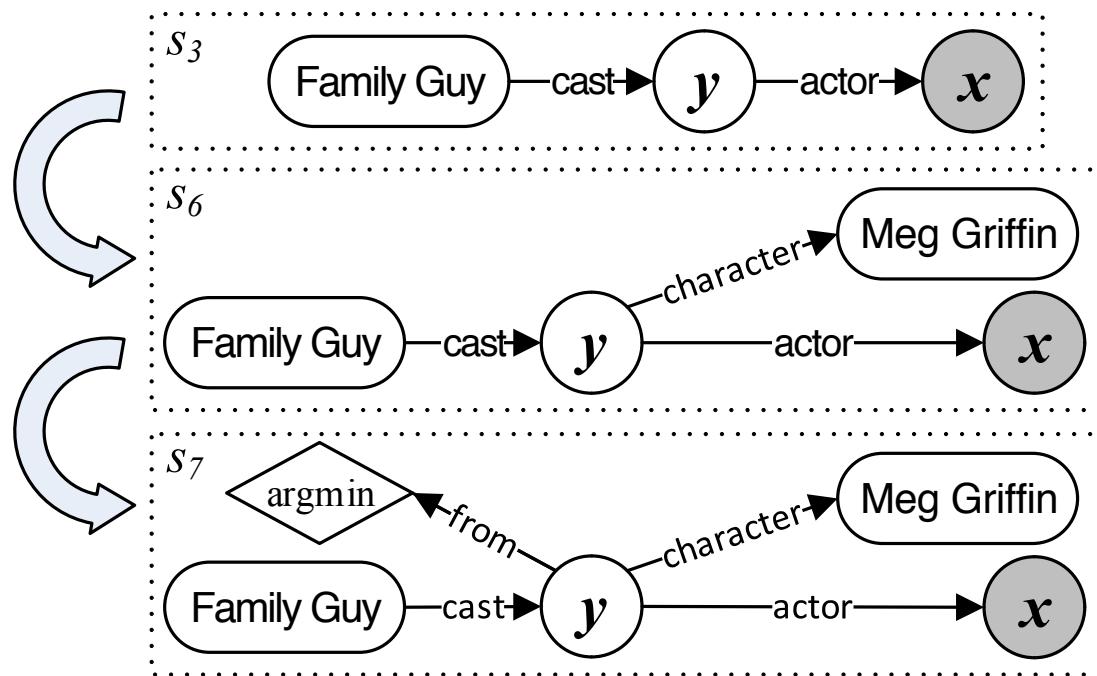
Generating Core Inferential Chain

Who first voiced Meg on **Family Guy**?



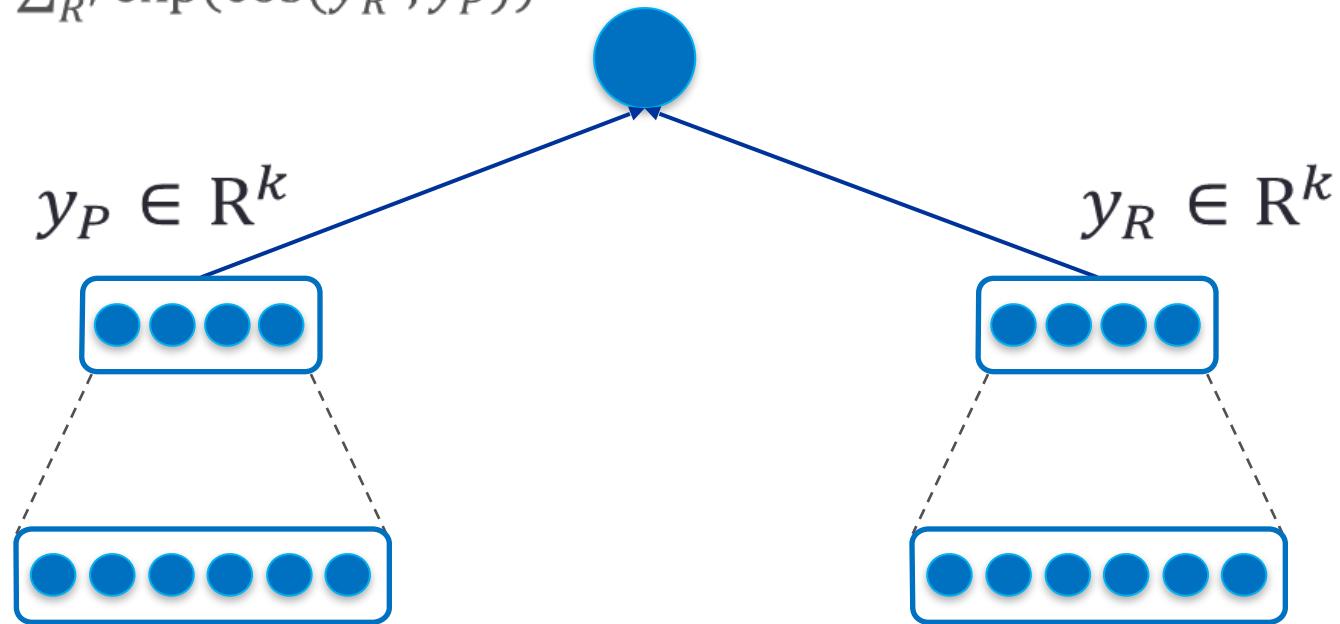
Constraint Augmentation

Who **first** voiced Meg on **Family Guy**?



Model

$$P(R|P) = \frac{\exp(\cos(y_R, y_P))}{\sum_{R'} \exp(\cos(y_{R'}, y_P))}$$



who voiced meg on $\langle e \rangle$

cast-actor



Deep Convolutional Neural Networks

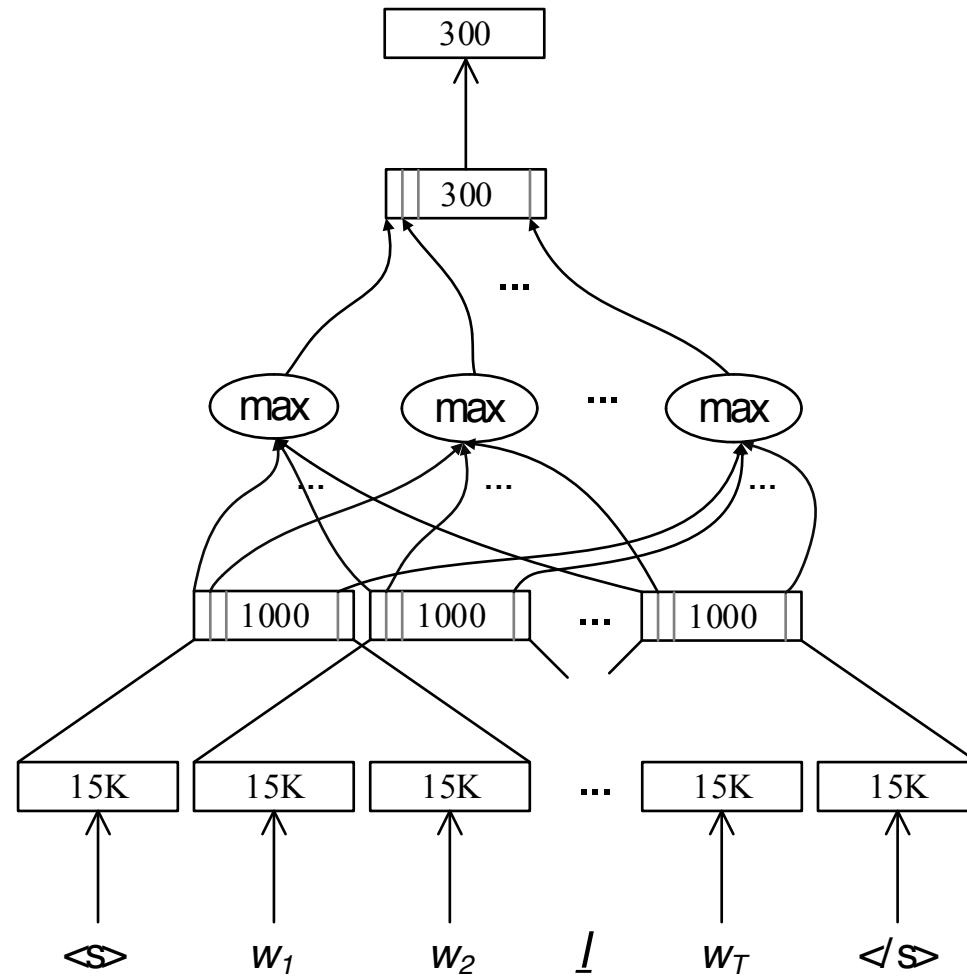
Softmax

MLP

Pooling

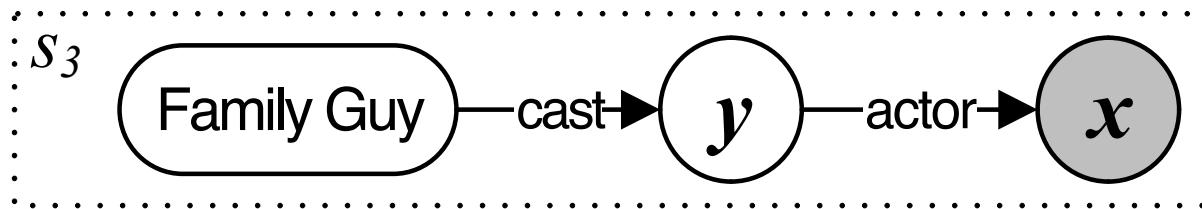
Convolution

Embedding



Augmentation Features

- Who first voiced Meg on Family Guy?


$$\lambda x. \exists y. \text{cast}(\text{FamilyGuy}, y) \wedge \text{actor}(y, x)$$

- One or more constraint nodes can be added to y or x
 - y : Additional property of this event (e.g., $\text{character}(y, \text{MegGriffin})$)
 - x : Additional property of the answer entity (e.g., gender)
- Only subset of constraint nodes are considered
 - e.g., entities detected in the question (more detail in Appendix)



Combining Features

- • Topic Entity
 - Entity linking scores
- Core Inferential Chain
 - Relation matching scores (NN models)
- Constraints: Keyword and entity matching
 - $\text{ConstraintEntityWord}(\text{"Meg Griffin"}, q) = 0.5$
 - $\text{ConstraintEntityInQuestion}(\text{"Meg Griffin"}, q) = 1$
- Overall
 - $\text{NumNodes}(s) = 5$
 - $\text{NumAnswers}(s) = 1$



Sample Inferential Chains

Pattern	Inferential Chain
<i>what was <e> known for</i>	<i>people.person.profession</i>
<i>what kind of government does <e> have</i>	<i>location.country.form_of_government</i>
<i>what year were the <e> established</i>	<i>sports.sports_team.founded</i>
<i>what city was <e> born in</i>	<i>people.person.place_of_birth</i>
<i>what did <e> die from</i>	<i>people.deceased_person.cause_of_death</i>
<i>who married <e></i>	<i>people.person.spouse_s</i> <i>people.marriage.spouse</i>



Results

Avg. F1 (Accuracy) on WebQuestions Test Set

