



Natural Language Processing

Info 159/259

Lecture 22: Question Answering (April 14, 2020)

David Bamman, UC Berkeley



us congress



[All](#)

[News](#)

[Images](#)

[Videos](#)

[Maps](#)

[More](#)

[Settings](#)

[Tools](#)

About 793,000,000 results (0.57 seconds)

Ad · www.ballotpedia.org/116th/Congress

116th U.S. Congress | Full List | ballotpedia.org

Find out who holds leadership positions in **Congress** and see a list of members. Find out who is part of the 116th **Congress** including membership and leadership. Candidate Facts. Encyclopedic information. Highlights: Professional Staff Available, Founded In 2007.

About Ballotpedia

Digital Encyclopedia
Covering U.S. Elections

Support Our Work

Grow The Encyclopedia
Of American Politics

www.congress.gov

Congress.gov | Library of Congress

U.S. Congress legislation, Congressional Record debates, Members of Congress , legislative process educational resources presented by the Library of ...

[Show all members](#) · [Roll Call Votes](#) · [Browse](#) · [Contact Us](#)

www.house.gov

House.gov: Home

Speaker Nancy Pelosi addresses the Members of the 116th Congress. Watch Live House Proceedings. The Office of the Clerk gives you a front-row seat. U.S. ...

[Find Your Representative](#) · [Representatives](#) · [The House Explained](#) · [Committees](#)

en.wikipedia.org/wiki/United_States_Congress

United States Congress - Wikipedia

The United States **Congress** is the bicameral legislature of the federal government of the United States, and consists of two chambers: the House of Representatives and the Senate. The **Congress** meets in the United States Capitol in Washington, D.C.

House of Representatives last election: [Nove...](#)

House of Representatives next election: [Nove...](#)

Houses: [Senate](#); [House of Representatives](#)

Speaker of the House: [Nancy Pelosi](#) (D); since ...



United States Congress



congress.gov

The United States Congress is the bicameral legislature of the federal government of the United States, and consists of two chambers: the House of Representatives and the Senate. The Congress meets in the United States Capitol in Washington, D.C.
[Wikipedia](#)

President of the Senate: [Mike Pence](#) (R); since [January 20, 2017](#)

President pro tempore of the Senate: [Chuck Grassley](#) (R); since January 3, 2019

Senate political groups: [Republican](#) (53); [Democratic](#) (45); Independent (2);

Senate last election: November 6, 2018

Senate next election: November 3, 2020

Number of members: Congress has 535 voting members: 435 representatives and 100 senators. [wikipedia.org](https://en.wikipedia.org)



who is the speaker of the house



All

News

Images

Books

Shopping

More

Settings

Tools

About 1,350,000,000 results (0.69 seconds)

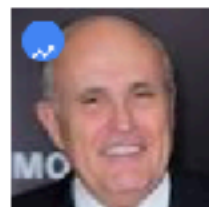
United States House of Representatives / Speaker of the House

Nancy Pelosi

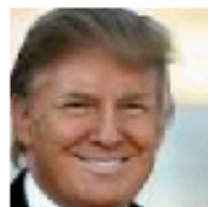


People also search for

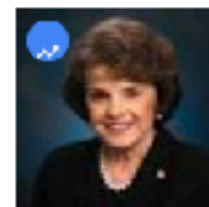
[View 15+ more](#)



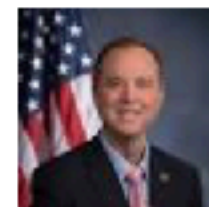
Rudy
Giuliani
Trending



Donald
Trump



Dianne
Feinstein
Trending



Adam
Schiff



Devin
Nunes
Trending



Chuck
Schumer



Alexan
Ocasio

[Claim this knowledge panel](#)

[Feedback](#)

Hey Siri who won the Alaska democratic primary?

Tap to Edit >

According to the Associated Press, Joe Biden has won the Alaska Democratic primary with 55.3% of the vote.



2020 ELECTIONS COVERAGE

Alaska Democratic Party Primary

2020 Presidential Primaries Results

CANDIDATE	VOTE%	DELEGATES
 Joe Biden	55.3%	11
Bernie Sanders	44.7%	4
Elizabeth Warren	—	—
Tulsi Gabbard	—	—

100% of precincts reporting

Updated at 07:27PM PDT
Election results from The Associated Press



Watson



Factoid questions

- Questions that can be answered with simple facts:
 - Who founded Virgin Airlines?
 - What is the average age of the onset of autism?
 - Where is Apple Computer based?
 - ~~How do I get to be a cook at the French Laundry?~~

Question Answering

- Knowledge-based question answering
- IR-based question answering

Knowledge-based QA

- Parse a question into a meaning representation, and execute that meaning representation against a database.

SQL	What record company did conductor Mikhail Snitko record for after 1996?
	SELECT RecordCompany WHERE (YearOfRecording > 1996) AND (Conductor = Mikhail Snitko)

Semantics

Lexical semantics is concerned with representing the meaning of words (and their relations)

Logical semantics is concerned with representing the meaning of sentences.

Meaning representation

- A representation of the meaning of a sentence needs to bridge **linguistic** aspects of the sentence with **non-linguistic** knowledge about the world.

Following directions



Linguistic	Non-linguistic
Verbs like <i>grab</i> , <i>release</i>	Actions a robot can execute
Nouns like <i>cup</i> , <i>block</i>	Specific entities in the world
transitive VP (V NP)	An action executed upon an object

Question answering

Barack Obama was the 44th President of the United States. Obama was born on August 4 in Honolulu, Hawaii. In late August 1961, Obama's mother moved with him to the University of Washington in Seattle for a year...

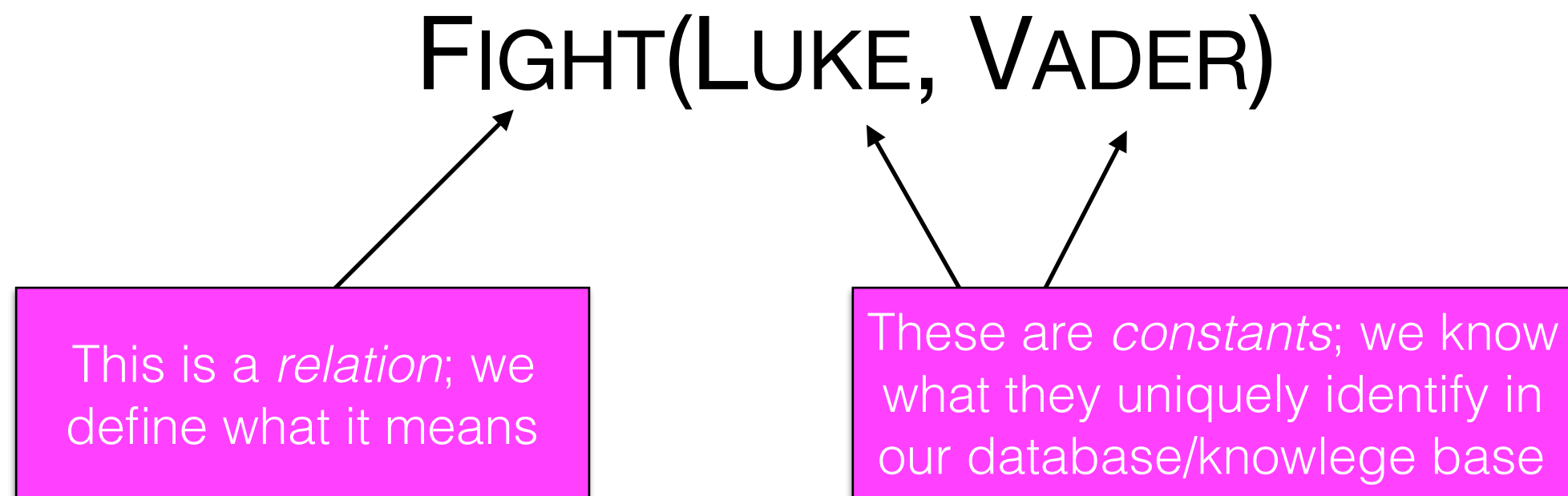
Is Barack Obama from the United States?

Non-linguistic
Hawaii is a place
Hawaii is a place within the United States
“born in X” entails “from x”

Meaning representation

Geo	which state has the most rivers running through it?
	(argmax \$0 (state:t \$0) (count \$1 (and (river:t \$1) (loc:t \$1 \$0))))
ATIS	all flights from dallas before 10am
	(lambda \$0 e (and (flight \$0) (from \$0 dallas:ci) (< (departure time \$0) 1000:ti)))
SQL	What record company did conductor Mikhail Snitko record for after 1996?
	SELECT Record Company WHERE (Year of Recording > 1996) AND (Conductor = Mikhail Snitko)
Django	if length of bits is lesser than integer 3 or second element of bits is not equal to string 'as'
	if len(bits) < 3 or bits[1] != 'as':

Representation of meaning



Representation of meaning

- Relations: likes(x,y) is scoped over two variables
- We can represent the partial representation of meaning with lambda expressions:

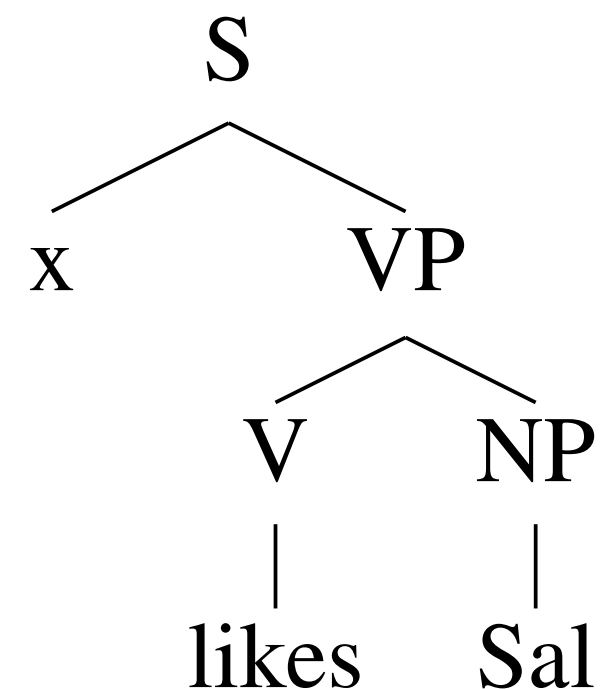
$\lambda x.\text{likes}(x,\text{Sal})$

Expect one other argument to complete the meaning of this relation

Representation of meaning

$\lambda x.\text{likes}(x, \text{Sal})$

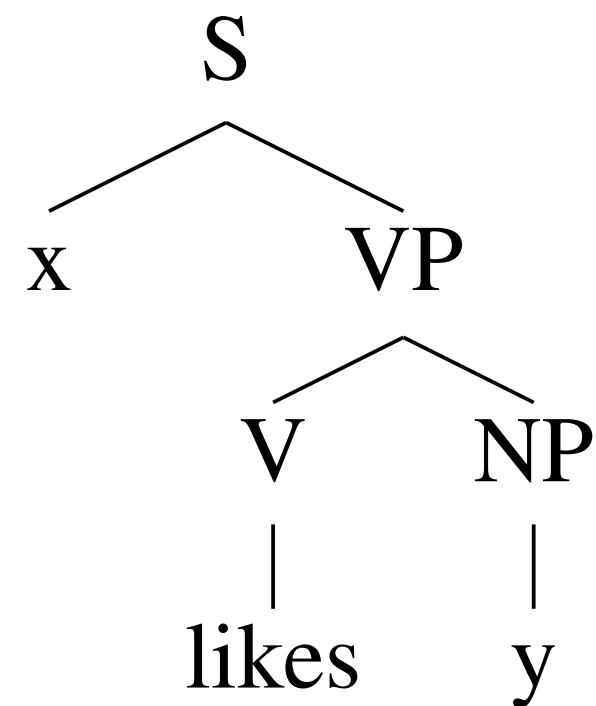
Lambda expressions
let us tie semantics
explicitly to phrases
(subtrees in syntax)



Representation of meaning

$\lambda y.\lambda x.\text{likes}(x,y)$

Lambda expressions
let us tie semantics
explicitly to phrases
(subtrees in syntax)



Learning from logical forms

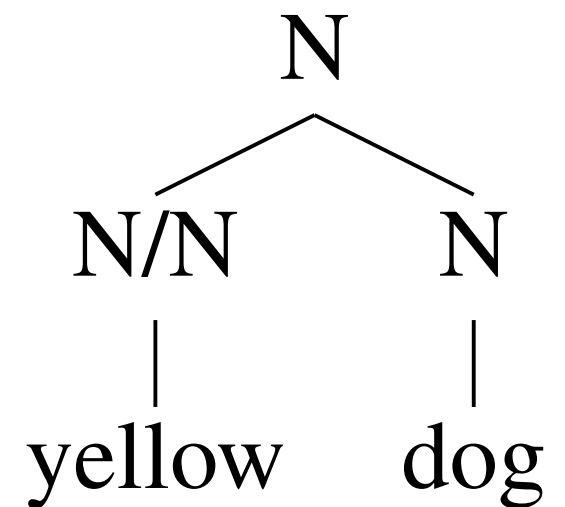
sentence	what states border texas
logical form	$\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{texas})$

CCG

- Combinatory categorial grammar is a phrase-structure grammar that ties syntactic representation to semantic structure.
- Two components:
 - **Combinators** that tell us how to compose syntactic types (S, VP, NP, etc.)
 - Mapping between syntactic operations and equivalent semantic operation
 - **Lexicon** mapping each word in a vocabulary to its syntactic/semantic information.

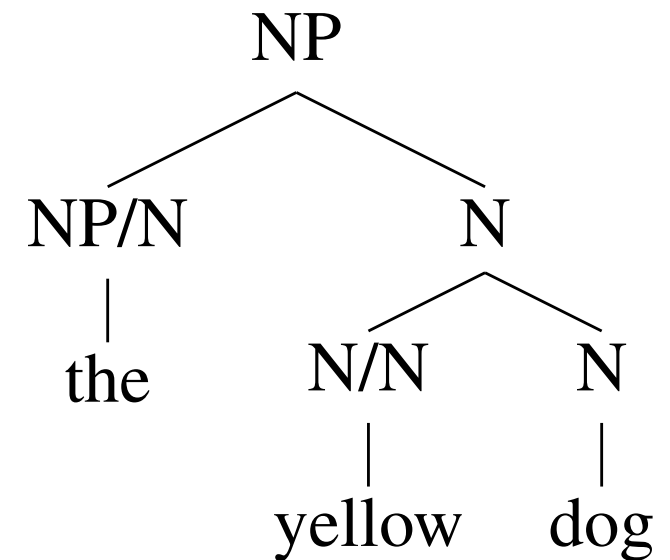
CCG Combinators

- Forward application combinator ($X/Y \rightarrow X$)
- $N/N \rightarrow N$



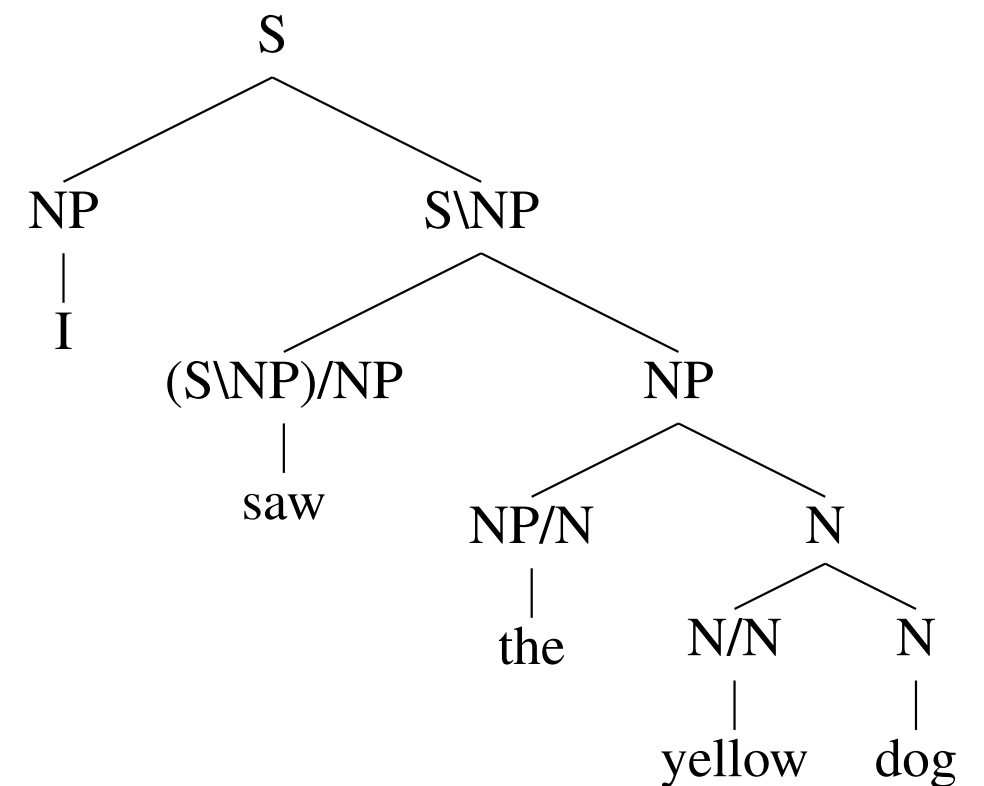
CCG Combinators

- Forward application combinator ($X/Y \rightarrow X$)
 - $N/N \rightarrow N$
 - $NP/N \rightarrow NP$



CCG Combinators

- Backward application combinator ($X \backslash Y \rightarrow X$)
- $S \backslash NP \rightarrow S$



CCG Lexicon

Utah	NP	utah
Idaho	NP	idaho
borders	$(S \backslash NP) / NP$	$\lambda x. \lambda y. \text{borders}(y, x)$
adjoins	$(S \backslash NP) / NP$	$\lambda x. \lambda y. \text{adjoins}(y, x)$
abuts	$(S \backslash NP) / NP$	$\lambda x. \lambda y. \text{abuts}(y, x)$

CCG Combinators

- Each combinator tells us what to do with the corresponding semantics
- Forward application:

$X/Y : f$	$Y : g$	$\rightarrow$	$X f(g)$	
$(S \backslash NP) / NP :$ $\lambda x . \lambda y . borders(y, x)$	$NP :$ idaho	$\rightarrow$	$S \backslash NP :$ $\lambda x . \lambda y . borders(y, x)(idaho)$	$S \backslash NP :$ $\lambda y . borders(y, idaho)$

Semantics

S

borders(utah,idaho)

S\NP

$\lambda y.$ borders(y,idaho)

NP

utah

Utah

(S\NP)/NP

$\lambda x.\lambda y.$ borders(y,x)

borders

NP

idaho

Idaho

Semantics

Utah	NP	utah
Idaho	NP	idaho
borders	$(S \backslash NP) / NP$	$\lambda x. \lambda y. \text{borders}(y, x)$
adjoins	$(S \backslash NP) / NP$	$\lambda x. \lambda y. \text{adjoins}(y, x)$
abuts	$(S \backslash NP) / NP$	$\lambda x. \lambda y. \text{abuts}(y, x)$

Semantics

Utah	NP	utah
Idaho	NP	idaho
borders	(S\NP)/NP	$\lambda x.\lambda y.borders(y,x)$
adjoins	(S\NP)/NP	$\lambda x.\lambda y.borders(y,x)$
abuts	(S\NP)/NP	$\lambda x.\lambda y.borders(y,x)$

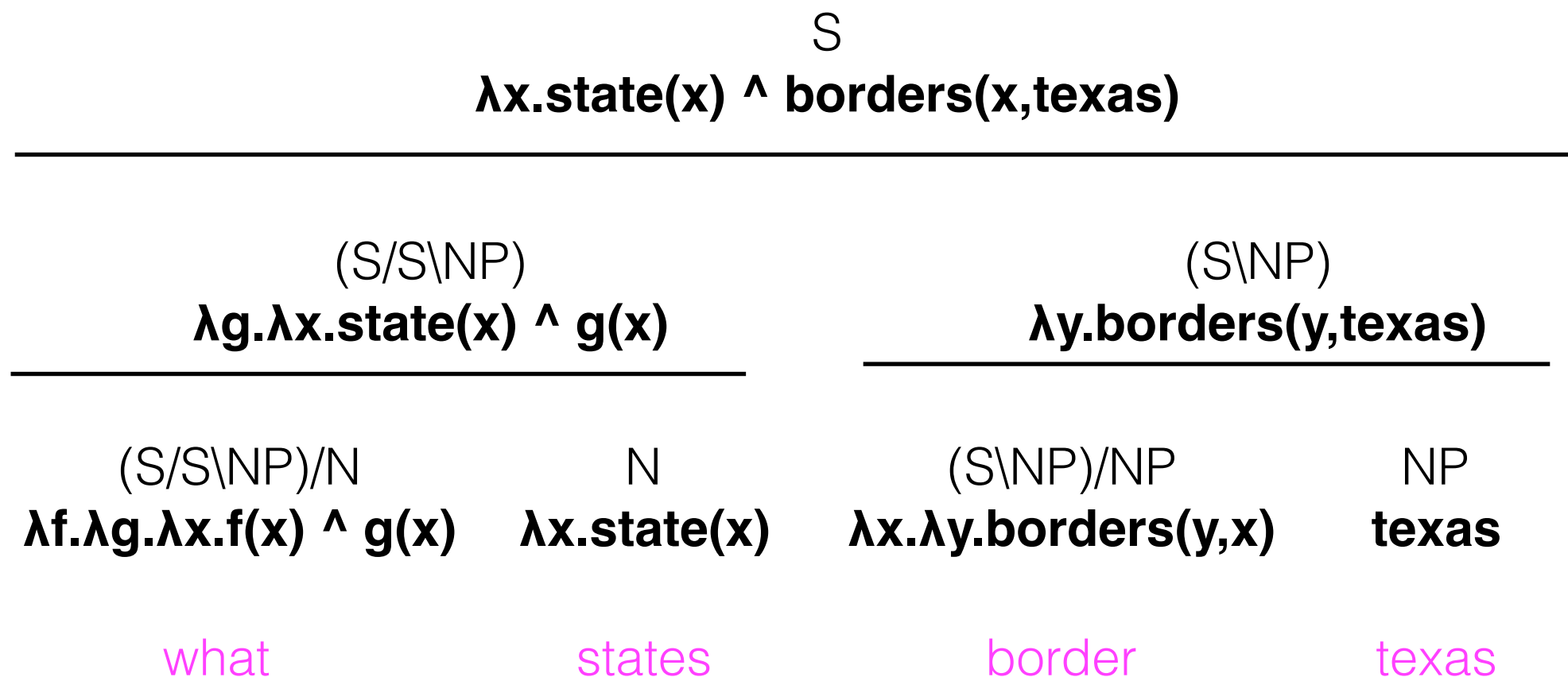
Semantics

- Semantic parsing with CCG is simply syntactic parsing, assuming mapping from syntactic primitives to logical forms.
- But this encounters two problems:
 - We don't have those manual mappings (task-specific).
 - We can't parse anything not in our lexicon.

Learning from logical forms

- We can train a semantic parser in a number of ways:
 - Full derivational trees (CCGBank)
 - Logical forms (Zettlemoyer and Collins 2005)
 - Denotations (Berant et al. 2013)

Learning from trees



Learning from logical forms

sentence	what states border texas
logical form	$\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{texas})$

Two core ideas:

- We'll learn the lexicon (including the lambda expressions)
- We'll learn CCG parser from that lexicon, and treat the true tree as a **latent variable**

Learning from logical forms

- Calculate the joint probability of a logical form L and derivation T for sentence S as:

$$P(L, T \mid S; \theta) = \frac{\exp(f(L, T, S)^\top \theta)}{\sum_{L, T} \exp(f(L, T, S)^\top \theta)}$$

sums over all valid trees/logical forms for the sentence

feature
Utah := NP : utah
Utah := NP : idaho
borders := (S\NP)/NP : $\lambda x. \lambda y. \text{borders}(y, x)$
borders := (S\NP)/NP : $\lambda x. \lambda y. \text{borders}(x, y)$

$f(L, T, S)$

Learning from denotations

sentence	what states border texas
logical form	$\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{texas})$
denotation	new_mexico, oklahoma, arkansas, louisiana

sentence	number of dramas starring tom cruise
logical form	$\text{count}(\lambda x.\text{genre}(x, \text{drama}) \wedge \exists y.\text{performance}(x, y) \wedge \text{actor}(y, \text{tom_cruise}))$
denotation	28

Learning from denotations

sentence	what states border texas
logical form	$\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{texas})$
denotation	new_mexico, oklahoma, arkansas, louisiana

sentence	number of dramas starring tom cruise
logical form	$\text{count}(\lambda x.\text{genre}(x, \text{drama}) \wedge \exists y.\text{performance}(x, y) \wedge \text{actor}(y, \text{tom_cruise}))$
denotation	28

Learning from denotations

- How could we use the principles of **learning from logical forms** to learn from denotations?
- The meaning of a sentence is the set of possible worlds consistent with that statement.

Utah borders Idaho	TRUE
number of dramas starring tom cruise	28

Learning from denotations

- Basic idea: maximize the probability of the tree T / logical form z that, when executed against a knowledge base $\mathcal{K}$, yield the correct denotation y

$$\sum_{i=1}^N \log \sum_{T: \llbracket T.z \rrbracket_{\mathcal{K}} = y_i} P(T \mid S_i, \theta)$$

objective function

IR-based QA

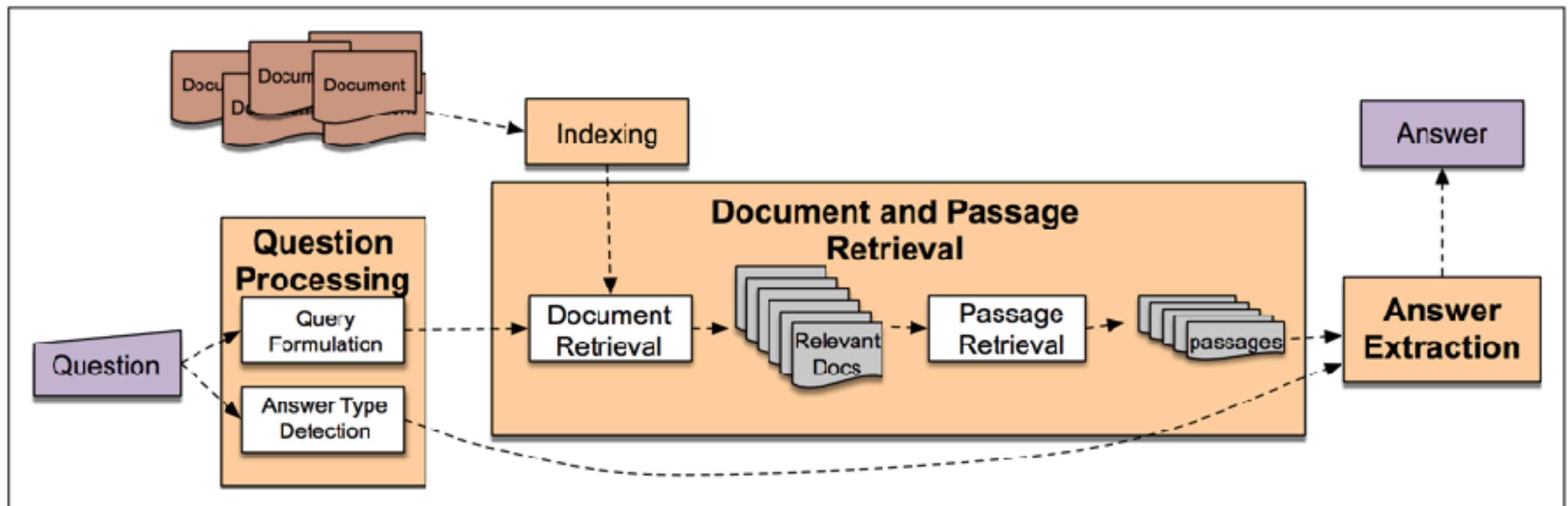


Figure 25.2 IR-based factoid question answering has three stages: question processing, passage retrieval, and answer processing.

Answer type detection

- If we know the type of answer expected for a question, we can exclude answers that aren't of that type (resulting in higher precision)

Who founded Virgin Airlines?	→	PERSON
What Canadian city has the largest population?	→	CITY

Answer type detection

ENTITY	
animal	What are the names of Odin's ravens?
body	What part of your body contains the corpus callosum?
color	What colors make up a rainbow?
creative	In what book can I find the story of Aladdin?
currency	What currency is used in China?
disease/medicine	What does Salk vaccine prevent?
event	What war involved the battle of Chapultepec?
food	What kind of nuts are used in marzipan?
instrument	What instrument does Max Roach play?
lang	What's the official language of Algeria?
letter	What letter appears on the cold-water tap in Spain?
other	What is the name of King Arthur's sword?
plant	What are some fragrant white climbing roses?
product	What is the fastest computer?
religion	What religion has the most members?
sport	What was the name of the ball game played by the Mayans?
substance	What fuel do airplanes use?
symbol	What is the chemical symbol for nitrogen?
technique	What is the best way to remove wallpaper?
term	How do you say "Grandma" in Irish?
vehicle	What was the name of Captain Bligh's ship?
word	What's the singular of dice?

Answer type detection

- Trained an answer type detection system with supervised learning:
 - Words/embeddings
 - POS for each word
 - Question headword
 - Which **city** in China has the largest number of foreign financial companies?
 - What is the state **flower** of California?

Document retrieval

- Rather than running answer extraction on all documents, let's just focus on the subset that are likely to be **relevant**.
- This is a problem of **information retrieval**.

IR

- Represent query **and** each document by tf-idf weighted vector, rank by cosine similarity

$$\cos(x, y) = \frac{\sum_{i=1}^F x_i y_i}{\sqrt{\sum_{i=1}^F x_i^2} \sqrt{\sum_{i=1}^F y_i^2}}$$

term	doc 1	doc 2	doc 3
alaska			4.5
primary			7.8
dog		4.8	3.4
friday			
nlp	6.4		
the	0.05	0.03	0.02
dem			8.1
tree	2.3	9.1	

Document retrieval

- Answer extraction typically operates on passages (not entire documents), so we can rank the passages as well using information about:
 - Number of entities of correct type
 - Number of question keywords
 - Longest exact sequence of keywords in passage
 - Rank of document
 - Ngram overlap

Answer extraction

- From a given passage, extract the answer.

Input passage

Talking Heads were an American rock band formed in 1975 in **New York City** and active until 1991.^[8] The band comprised David Byrne (lead vocals, guitar), Chris Frantz (drums), Tina Weymouth (bass), and Jerry Harrison (keyboards, guitar). Described by the critic Stephen Thomas Erlewine as "one of the most critically acclaimed bands of the '80s,"^[3] the group helped to pioneer new wave music by integrating elements of punk, art rock, funk, and world music with avant-garde sensibilities and an anxious, clean-cut image.^[3]

Question

Where did the Talking Heads originate?

Datasets

Dataset	Size	Source	Notes
SQUAD	150K	Wikipedia	
NewsQA	100K	CNN	
HotpotQA	113K	Wikipedia	Requires reasoning over multiple documents
NarrativeQA	46K	Books/movie scripts	Requires reasoning over long documents, answer not a span

HotpotQA

Paragraph A, Return to Olympus:

[1] *Return to Olympus is the only album by the alternative rock band Malfunkshun.* [2] *It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990.* [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

Paragraph B, Mother Love Bone:

[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] The band was active from 1987 to 1990. [6] *Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene.* [7] *Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success.* [8] The album was finally released a few months later.

Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?

A: Malfunkshun

Supporting facts: 1, 2, 4, 6, 7

NarrativeQA

Title: Ghostbusters II

Question: How is Oscar related to Dana?

Answer: her son

Summary snippet: ...Peter's former girlfriend Dana Barrett has had a son, Oscar...

Story snippet:

DANA (setting the wheel brakes on the buggy)

Thank you, Frank. I'll get the hang of this eventually.

She continues digging in her purse while Frank leans over the buggy and makes funny faces at the baby, OSCAR, a very cute nine-month old boy.

FRANK (to the baby)

Hiya, Oscar. What do you say, slugger?

FRANK (to Dana)

That's a good-looking kid you got there, Ms. Barrett.

SQUAD

- 100,000 question-answer pairs created by crowdworkers.
- Task: given an input passage from Wikipedia, create a question whose answer is **a span of text within it**.

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?

gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

graupel

Where do water droplets collide with ice crystals to form precipitation?

within a cloud

Figure 1: Question-answer pairs for a sample passage in the SQuAD dataset. Each of the answers is a segment of text from the passage.

SQUAD2.0

- 100K questions from SQUAD + 50K **unanswerable questions**.
- To perform well, models must know when to abstain from answering a question.

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**

Rajpurkar et al. (2018), "Know What You Don't Know: Unanswerable Questions for SQuAD"

Reasoning	Description	Example	Percentage
Lexical variation (synonymy)	Major correspondences between the question and the answer sentence are synonyms.	Q: What is the Rankine cycle sometimes called ? Sentence: The Rankine cycle is sometimes referred to as a <u>practical Carnot cycle</u> .	33.3%
Lexical variation (world knowledge)	Major correspondences between the question and the answer sentence require world knowledge to resolve.	Q: Which governing bodies have veto power? Sen.: <u>The European Parliament and the Council of the European Union</u> have powers of amendment and veto during the legislative process.	9.1%
Syntactic variation	After the question is paraphrased into declarative form, its syntactic dependency structure does not match that of the answer sentence even after local modifications.	Q: What Shakespeare scholar is currently on the faculty ? Sen.: Current faculty include the anthropologist Marshall Sahlins, ..., Shakespeare scholar <u>David Bevington</u> .	64.1%
Multiple sentence reasoning	There is anaphora, or higher-level fusion of multiple sentences is required.	Q: What collection does the V&A Theatre & Performance galleries hold? Sen.: The V&A Theatre & Performance galleries opened in March 2009. ... They hold the UK's biggest national collection of <u>material about live performance</u> .	13.6%
Ambiguous	We don't agree with the crowdworkers' answer, or the question does not have a unique answer.	Q: What is the main goal of criminal punishment? Sen.: Achieving crime control via <u>incapacitation and deterrence</u> is a major goal of criminal punishment.	6.1%

Evaluation

- Exact match: proportion of questions for which the predicted answer is an exact match for the correct answer.
- F1: Calculate F1 between predicted answer and gold answer (both bag of tokens).

SQUAD2.0

Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Mar 12, 2020	ALBERT + DAAF + Verifier (ensemble) PINGAN Omni-Sinitic	90.386	92.777
2 Jan 10, 2020	Retro-Reader on ALBERT (ensemble) Shanghai Jiao Tong University http://arxiv.org/abs/2001.09694	90.115	92.580
3 Nov 06, 2019	ALBERT + DAAF + Verifier (ensemble) PINGAN Omni-Sinitic	90.002	92.425
4 Sep 18, 2019	ALBERT (ensemble model) Google Research & TTIC https://arxiv.org/abs/1909.11942	89.731	92.215
4 Feb 25, 2020	Albert_Verifier_AA_Net (ensemble) QIANXIN	89.743	92.180

Input passage

Talking Heads were an American [rock](#) band formed in 1975 in [New York City](#) and active until 1991.^[8] The band comprised [David Byrne](#) (lead vocals, guitar), [Chris Frantz](#) (drums), [Tina Weymouth](#) (bass), and [Jerry Harrison](#) (keyboards, guitar). Described by the critic [Stephen Thomas Erlewine](#) as "one of the most critically acclaimed bands of the '80s,"^[3] the group helped to pioneer [new wave music](#) by integrating elements of [punk](#), [art rock](#), [funk](#), and [world music](#) with [avant-garde](#) sensibilities and an anxious, clean-cut image.^[3]

Question

Where did the Talking Heads originate?

- What is the [start token position](#) and [end token position](#) for the answer span in the input passage?

Featurized

- Answer type match: does the candidate answer contain a phrase with the answer type expected?
- Pattern match:

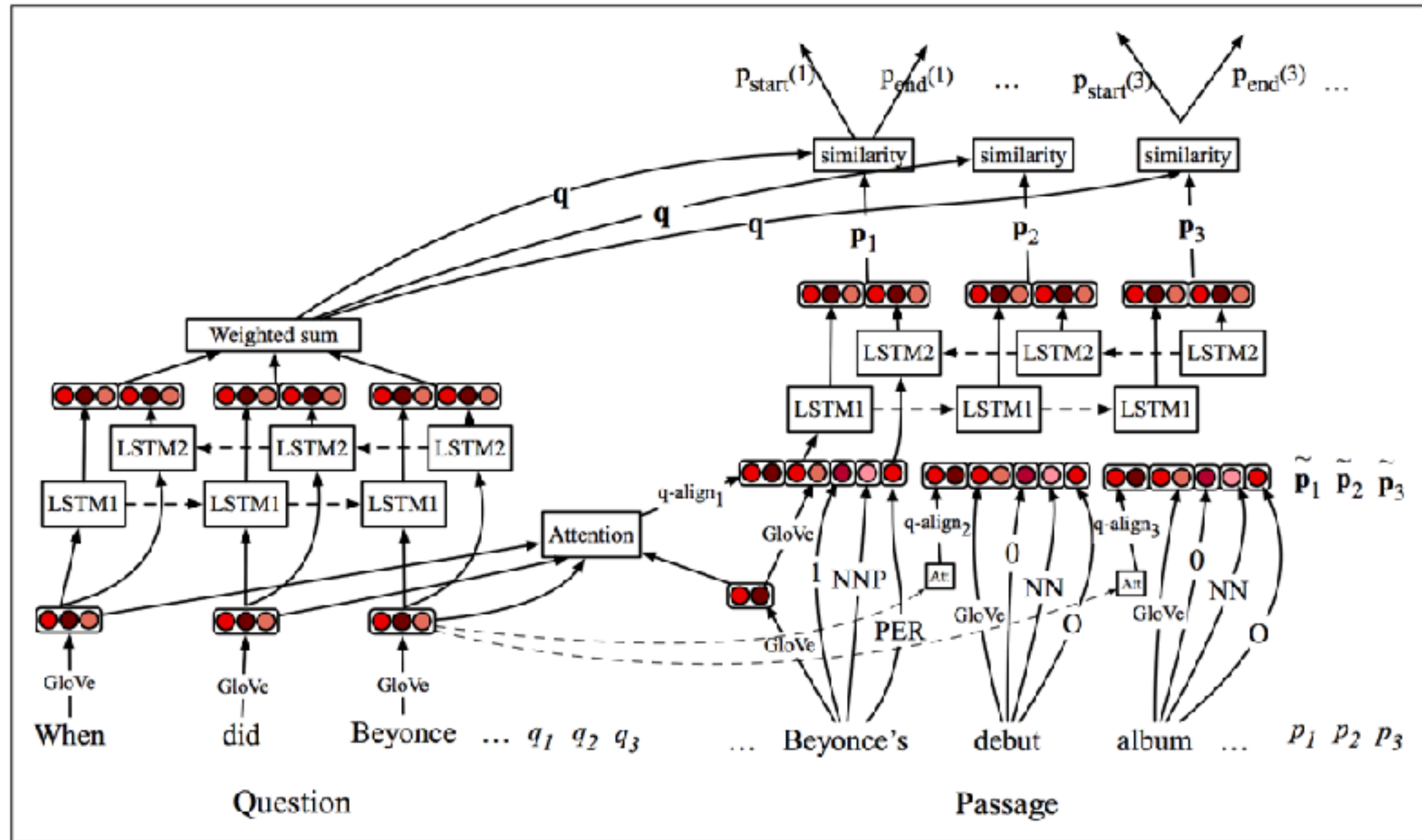
Pattern	Question	Answer
<AP> such as <QP>	What is autism?	“, <u>developmental disorders</u> such as autism”
<QP>, a <AP>	What is a caldera?	“the Long Valley caldera, a <u>volcanic crater</u> 19 miles long”

Figure 25.5 Some answer-extraction patterns using the answer phrase (AP) and question phrase (QP) for definition questions (Pasca, 2003).

Featurized

- Number of matched question keywords
- Keyword distance: distance between answer and query keywords
- Novelty factor: does the passage contain words that are not in the query?
- Length of longest sequence of question terms that occur in the candidate answer
- Apposition features: answer is in appositive construction to phrase containing several question words.
 - Question: What is a caldera?
 - Candidate: “the Long Valley caldera, a volcanic crater 19 miles long”

Neural



Input passage

Talking Heads were an American [rock](#) band formed in 1975 in [New York City](#) and active until 1991.^[8] The band comprised [David Byrne](#) (lead vocals, guitar), [Chris Frantz](#) (drums), [Tina Weymouth](#) (bass), and [Jerry Harrison](#) (keyboards, guitar). Described by the critic [Stephen Thomas Erlewine](#) as "one of the most critically acclaimed bands of the '80s,"^[3] the group helped to pioneer [new wave music](#) by integrating elements of [punk](#), [art rock](#), [funk](#), and [world music](#) with [avant-garde](#) sensibilities and an anxious, clean-cut image.^[3]

Question

Where did the Talking Heads originate?

- What is the [start token position](#) and [end token position](#) for the answer span in the input passage?

Input passage

Talking Heads were an American rock band formed in 1975 in New York City and active until 1991.

Question

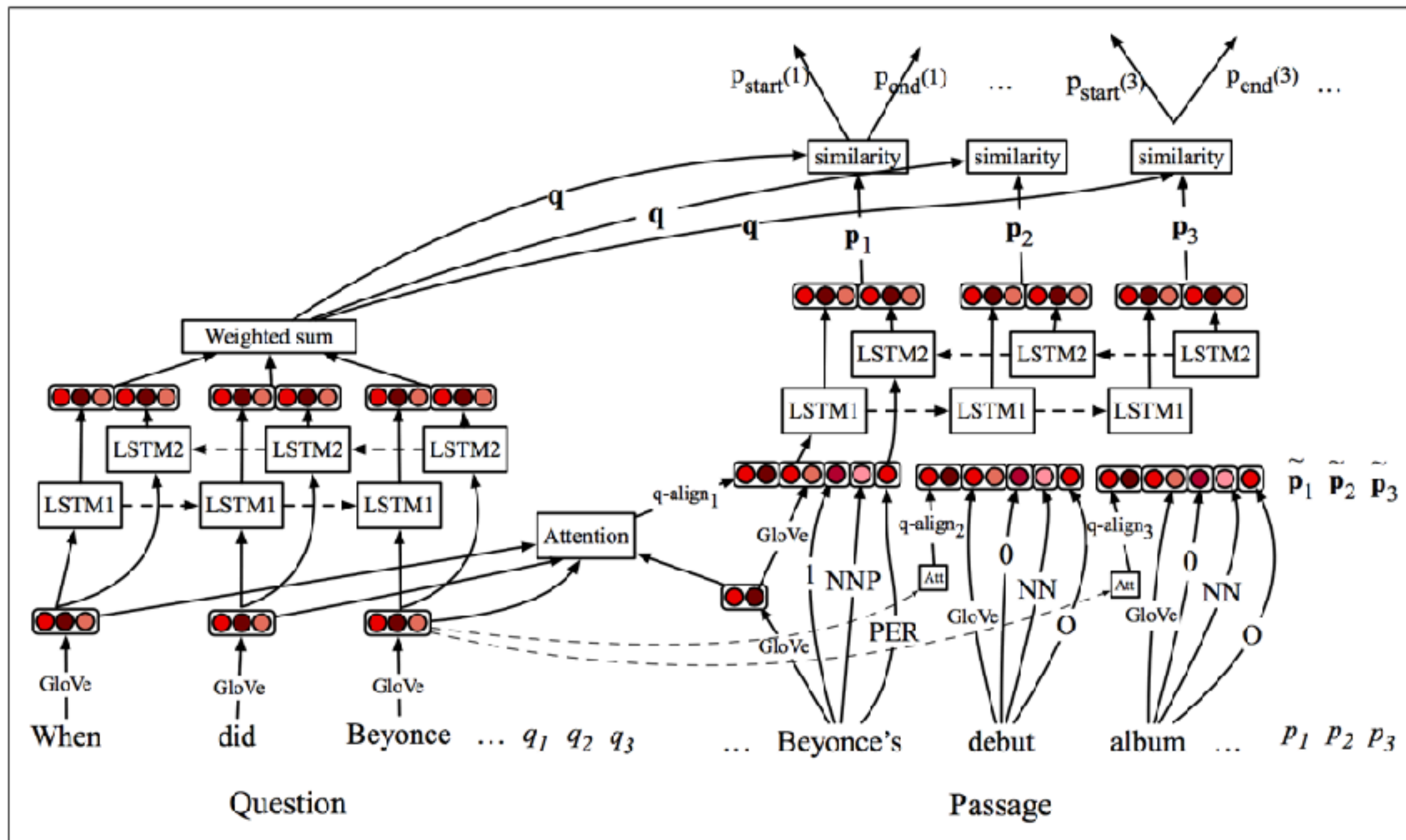
Where did the Talking Heads originate?

- What is the **start token position** and **end token position** for the answer span in the input passage?

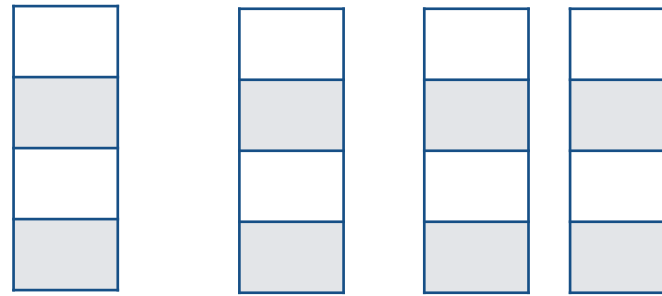
$p_{\text{start}}(i)$	0	0	0	0	0.10	0	0	0	0
$p_{\text{end}}(i)$	0	0	0	0	0.09	0	0	0	0
Text	Talking	Heads	were	an	American	rock	band	formed	in

$p_{\text{start}}(i)$	0	0	0.85	0	0.05	0	0	0	0
$p_{\text{end}}(i)$	0	0	0	0.16	0.75	0	0	0	0
Text	1975	in	New	York	City	and	active	until	1991

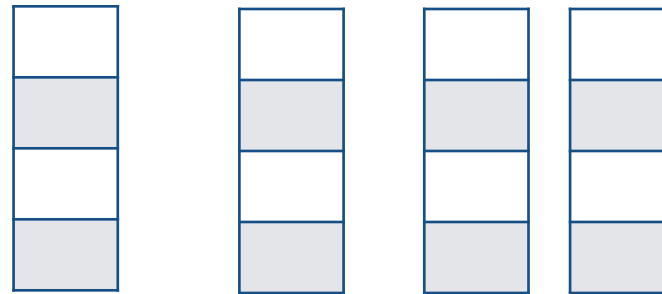
- Each token position i has two associated probabilities:
 - $p_{\text{start}}(i)$ — the probability of the answer starting at that token.
 - $p_{\text{end}}(i)$ — the probability of the answer ending at that token.



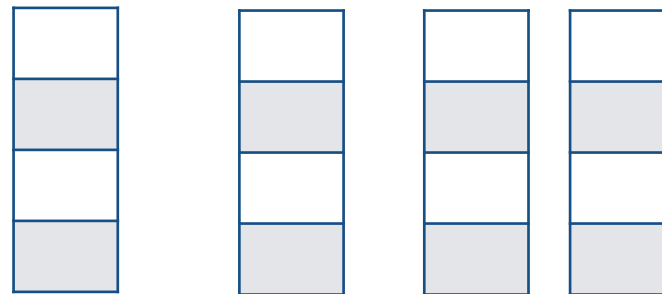
q-align



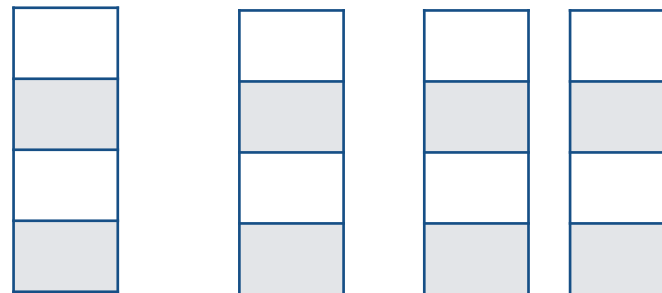
Token in question?



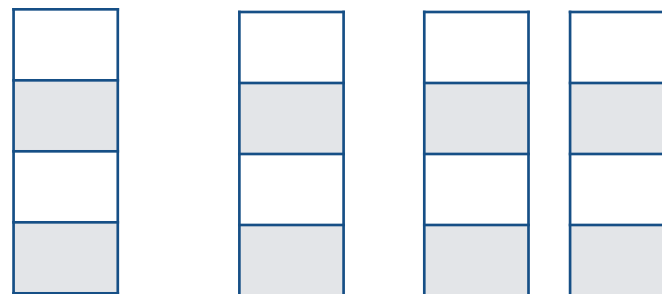
NER embeddings



POS embeddings

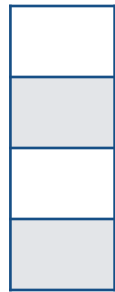


Word embeddings



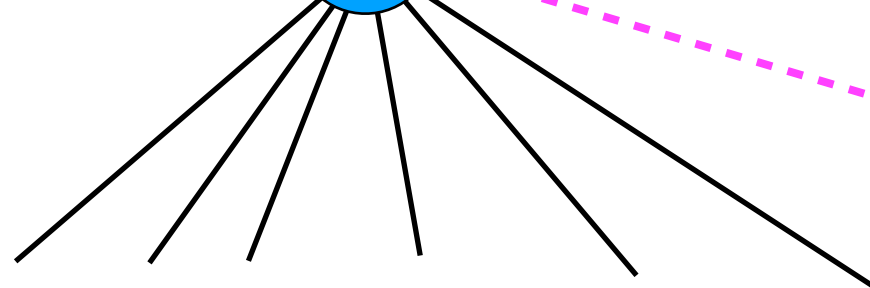
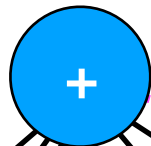
Talking Heads were an ...

Each passage token attends over all question tokens

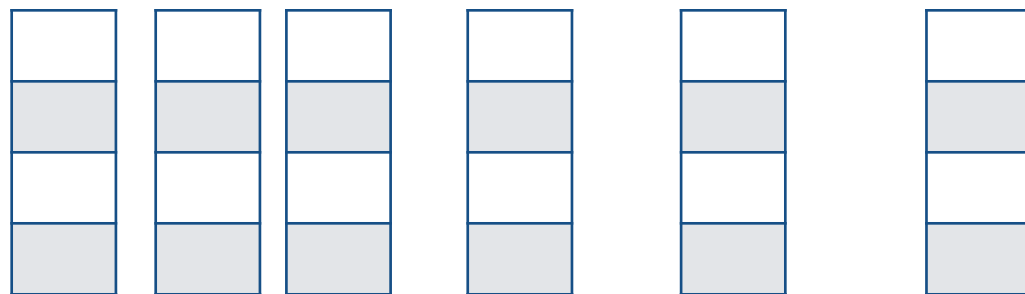


$$\mathbf{q-align}_i = \sum_j \alpha_{i,j} q_j$$

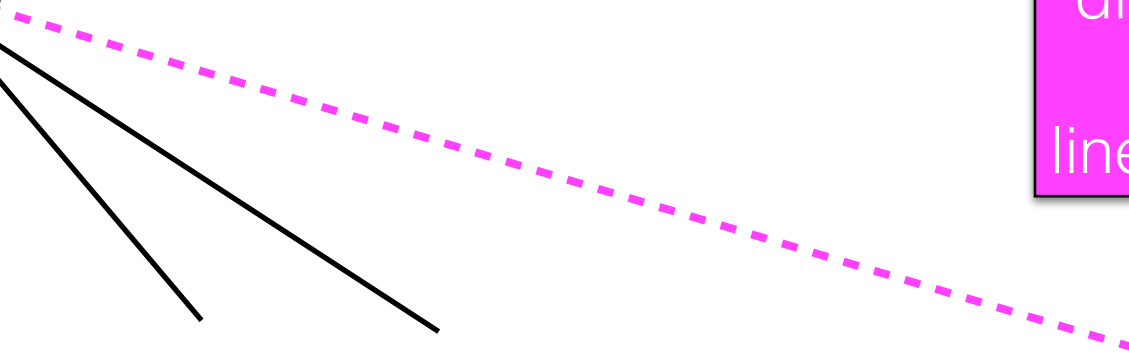
$$a_{i,j} = \frac{\exp \left(f(p_i)^\top f(q_j) \right)}{\sum_{j'} \exp \left(f(p_i)^\top f(q_{j'}) \right)}$$



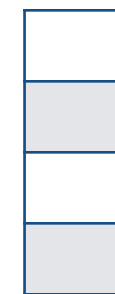
q₁ q₂ q₃ q₄ q₅ q₆



$f()$ here is some differentiable function over the embeddings (e.g., linear transformation, FFNN)



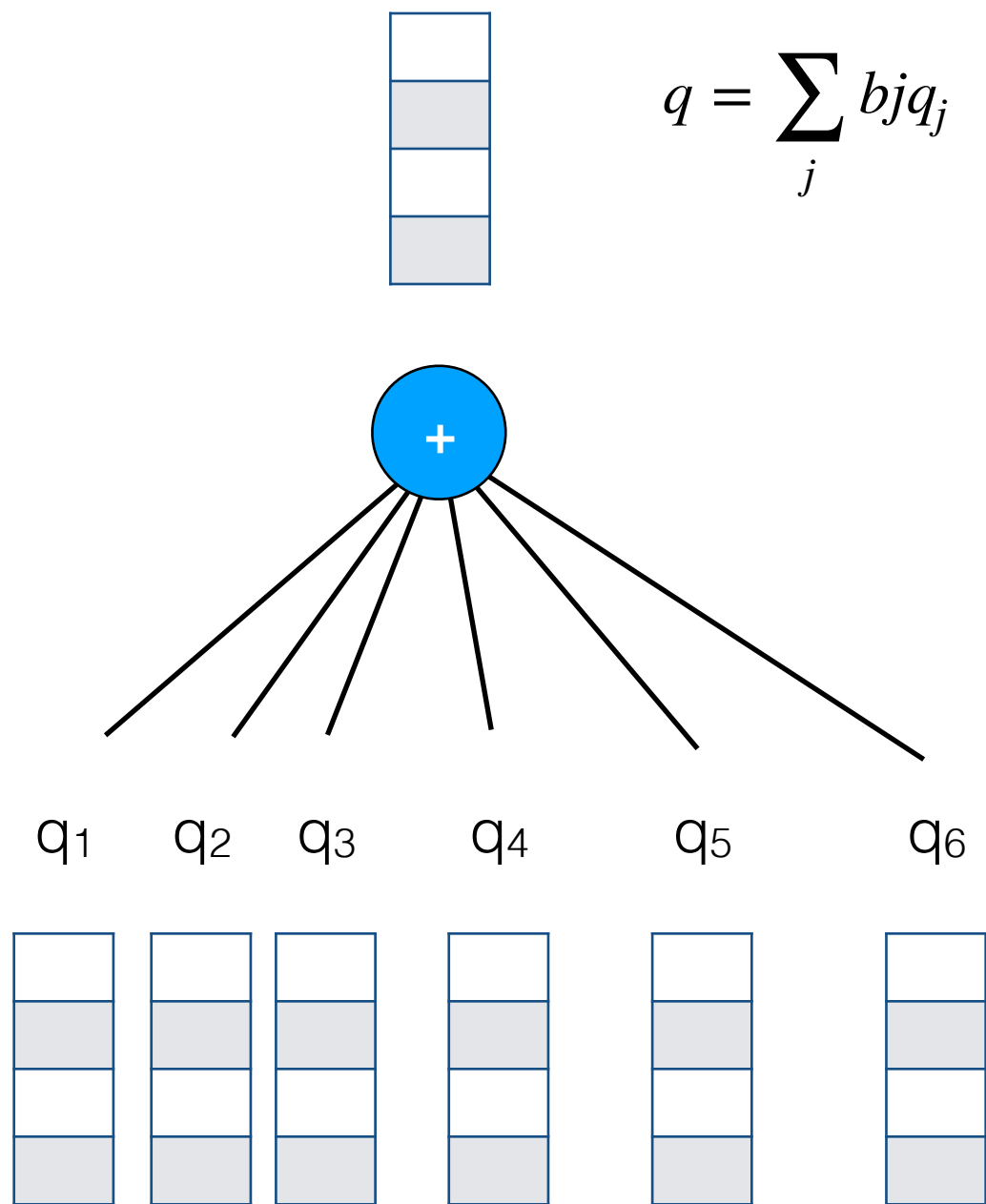
p₁



Where did the Talking Heads originate?

Talking Heads were an ...

Single representation (using attention) for the entire question

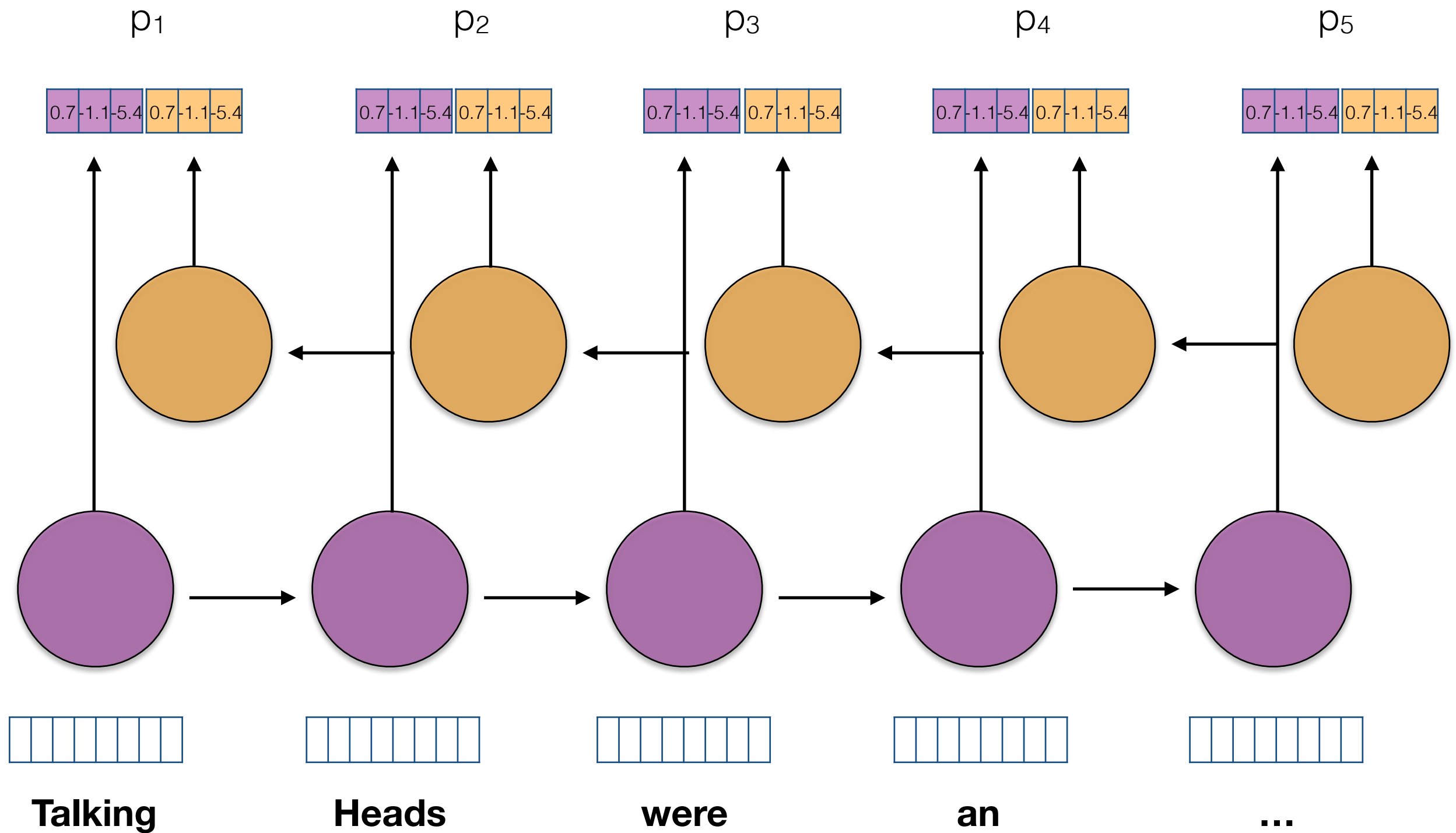


$$b_j = \frac{\exp(w^\top q_j)}{\sum_j \exp(w^\top q_j)}$$

Where did the Talking Heads originate?

$$p_{\text{start}}(i) \propto \exp(p_i W_s q)$$

$$p_{\text{end}}(i) \propto \exp(p_i W_e q)$$



SQUAD2.0

Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Mar 12, 2020	ALBERT + DAAF + Verifier (ensemble) PINGAN Omni-Sinitic	90.386	92.777
2 Jan 10, 2020	Retro-Reader on ALBERT (ensemble) Shanghai Jiao Tong University http://arxiv.org/abs/2001.09694	90.115	92.580
3 Nov 06, 2019	ALBERT + DAAF + Verifier (ensemble) PINGAN Omni-Sinitic	90.002	92.425
4 Sep 18, 2019	ALBERT (ensemble model) Google Research & TTIC https://arxiv.org/abs/1909.11942	89.731	92.215
4 Feb 25, 2020	Albert_Verifier_AA_Net (ensemble) QIANXIN	89.743	92.180

Question Answering

		Evaluated on					Avg Δ
		SQuAD	TriviaQA	NQ	QuAC	NewsQA	
Fine-tuned on	SQuAD	75.6	46.7	48.7	20.2	41.1	-17.2
	TriviaQA	49.8	58.7	42.1	20.4	10.5	-29.9
	NQ	53.5	46.3	73.5	21.6	24.7	-20.4
	QuAC	39.4	33.1	33.8	33.3	13.8	-36.9
	NewsQA	52.1	38.4	41.7	20.4	60.1	-22.2

Table 4: F1 scores of each fine-tuned model evaluated on each test set

Sen and Safari (2020), “What do Models Learn from Question Answering Datasets?”