Computing with Natural Language

Percy Liang

ACL Workshop on Semantic Parsing - June 15, 2014

Stanford University

[PaleoDeepDive (Shanan Peters, Chris Ré)]

Paleobiology

Paleobiology

The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period[1]. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations[2–5], but the mechanisms responsible for those perturbations remain uncertain[1,2]. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–480 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity[6]. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this palaeogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the 'Cambrian explosion' of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally tilted sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup[6]. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana[7,8], Baltica[8], Avalonia[9] and Siberia[10], making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first skeletonized crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution[1,6].

Here we use stratigraphic and lithologic data for 21,521 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence[11,12] that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences[11,12], but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth's surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth's crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreface system, often called the 'wave-base razor'[13], as well as adjacent transitional backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors[7] was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regolith[14,15], and

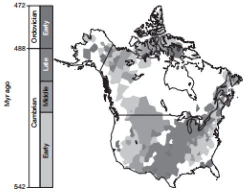


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

[1]Department of Geoscience, University of Wisconsin, Madison, Wisconsin 53706, USA. [2]Geology Department, Pomona College, Claremont, California 91711, USA.

Paleobiology

The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period[1]. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations[2–5], but the mechanisms responsible for those perturbations remain uncertain[1,2]. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–480 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity[6]. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this palaeogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the 'Cambrian explosion' of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally tilted sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup[6]. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana[7,8], Baltica[8], Avalonia[9] and Siberia[10], making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first skeletonized crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution[1,6].

Here we use stratigraphic and lithologic data for 21,521 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence[11,12] that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences[11,12], but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth's surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth's crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreface system, often called the 'wave-base razor'[13], as well as adjacent transitional backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors[7] was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regolith[14,15], and

Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

[1]Department of Geoscience, University of Wisconsin, Madison, Wisconsin 53706, USA. [2]Geology Department, Pomona College, Claremont, California 91711, USA.

paleobiodb.org

Paleobiology

The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period[1]. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations[2–5], but the mechanisms responsible for those perturbations remain uncertain[1,2]. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–480 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity[6]. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this palaeogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the 'Cambrian explosion' of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally tilted sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup[6]. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana[7,8], Baltica[9], Avalonia[9] and Siberia[10], making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first skeletonized crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution[1,6].

Here we use stratigraphic and lithologic data for 21,521 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence[11,12] that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences[11,12], but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth's surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth's crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreface system, often called the 'wave-base razor'[13], as well as adjacent transitional backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors[7] was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regolith[14,15], and

Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

[1]Department of Geoscience, University of Wisconsin, Madison, Wisconsin 53706, USA. [2]Geology Department, Pomona College, Claremont, California 91711, USA.

paleobiodb.org

Where was the last American Mastadon found?

Paleobiology

The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period[1]. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations[2–5], but the mechanisms responsible for those perturbations remain uncertain[1,2]. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–480 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity[6]. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this palaeogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the 'Cambrian explosion' of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally tilted sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup[6]. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana[7,8], Baltica[8], Avalonia[9] and Siberia[10], making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first skeletonized crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution[1,6].

Here we use stratigraphic and lithologic data for 21,521 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence[11,12] that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences[11,12], but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth's surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth's crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreface system, often called the 'wave-base razor'[13], as well as adjacent transitional backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors[7] was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regolith[14,15], and

Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

[1]Department of Geoscience, University of Wisconsin, Madison, Wisconsin 53706, USA. [2]Geology Department, Pomona College, Claremont, California 91711, USA.

Where was the last American Mastadon found?

How long do species tend to exist before going extinct?

Paleobiology

The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period[1]. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations[2–5], but the mechanisms responsible for those perturbations remain uncertain[1,2]. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–480 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity[6]. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this palaeogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the 'Cambrian explosion' of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally tilted sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup[6]. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana[7,8], Baltica[8], Avalonia[9] and Siberia[10], making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first skeletonized crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution[1,6].

Here we use stratigraphic and lithologic data for 21,521 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence[11,12] that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences[11,12], but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth's surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth's crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreface system, often called the 'wave-base razor'[13], as well as adjacent transitional backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors[7] was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regolith[14,15], and

Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

[1]Department of Geoscience, University of Wisconsin, Madison, Wisconsin 53706, USA. [2]Geology Department, Pomona College, Claremont, California 91711, USA.

paleobiodb.org

Where was the last American Mastadon found?

How long do species tend to exist before going extinct?

Goal: help scientists answer macro-questions

Challenge: requires computation / aggregation

Question answering via semantic parsing

Where was the last American Mastadon found?

Question answering via semantic parsing

Where was the last American Mastadon found?

↓ semantic parsing

$\text{LocationOf.argmax}(\text{Type.Occurrence} \sqcap \text{Genus.Mammut}, \text{Period})$

Question answering via semantic parsing

Where was the last American Mastadon found?

↓ semantic parsing

LocationOf.argmax(Type.Occurrence $\sqcap$ Genus.Mammut, Period)

↓ execute

`New Mexico`

Question answering via semantic parsing

Where was the last American Mastadon found?

↓ semantic parsing

↓ execute

`New Mexico`

Email assistant via semantic parsing

Send a reminder to all authors who haven't sent an abstract.

Email assistant via semantic parsing

Send a reminder to all authors who haven't sent an abstract.

↓ semantic parsing

$$\forall x \in (\text{Author} \sqcap \neg \text{Sent.Subject.Abstract}) : \text{Remind}(x)$$

Email assistant via semantic parsing

Send a reminder to all authors who haven't sent an abstract.

↓ semantic parsing

$$\forall x \in (\text{Author} \sqcap \neg\text{Sent.Subject.Abstract}) : \text{Remind}(x)$$

↓ execute

`[5 emails sent]`

Email assistant via semantic parsing

Send a reminder to all authors who haven't sent an abstract.

semantic parsing

execute

`[5 emails sent]`

Semantic parsing

[utterance: user input]

↓ semantic parsing

[program]

↓ execute

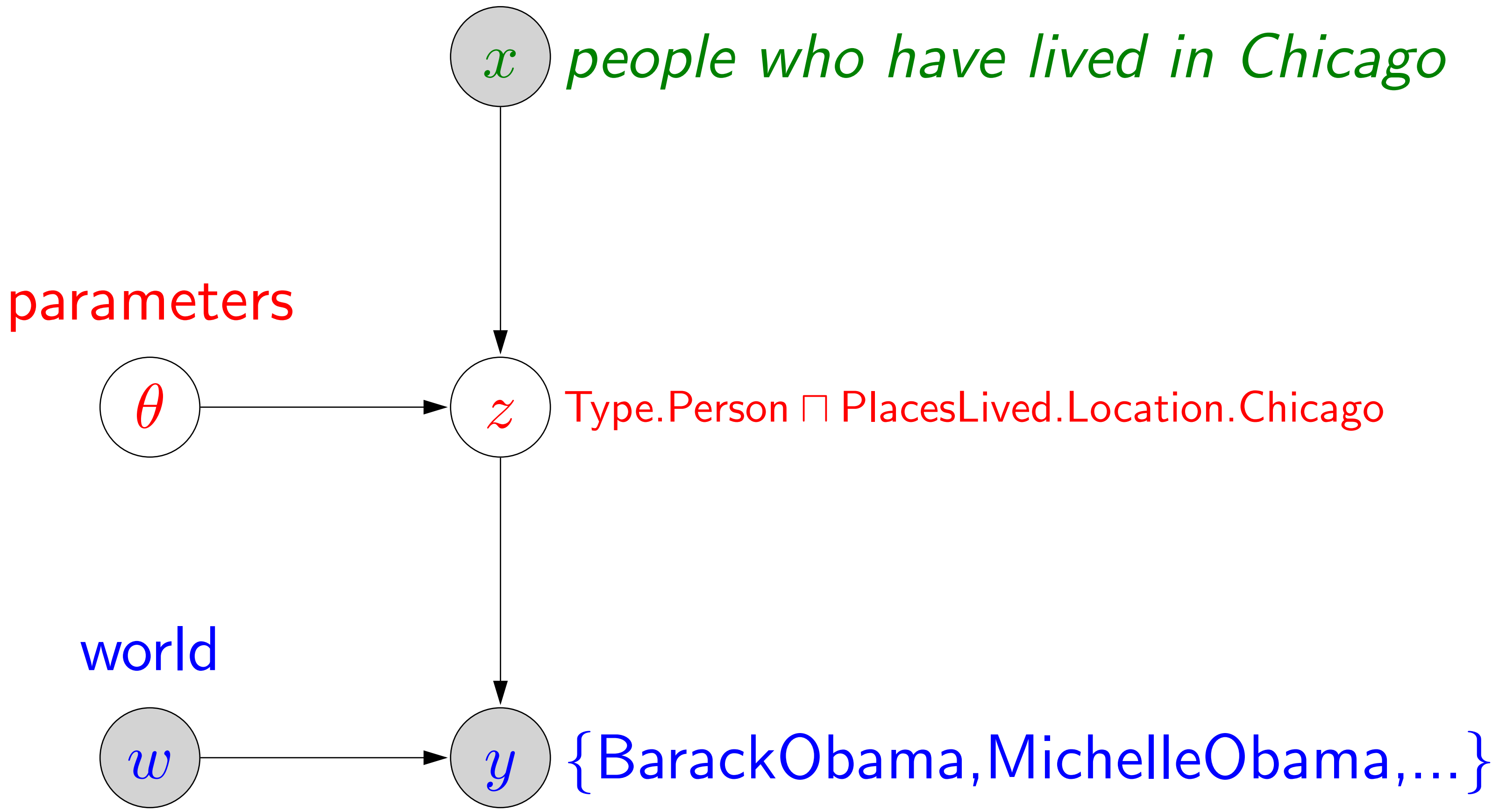
`[behavior: user output]`

Programs affect the world

Outline

- **Semantic parsing in 5 minutes**
- A closer look at the elements
  - Knowledge base incompleteness
  - Lexical coverage
  - Search over logical forms
  - Learning via bootstrapping
  - Leveraging denotations ("grounding")
  - Datasets
- Final remarks

Framework

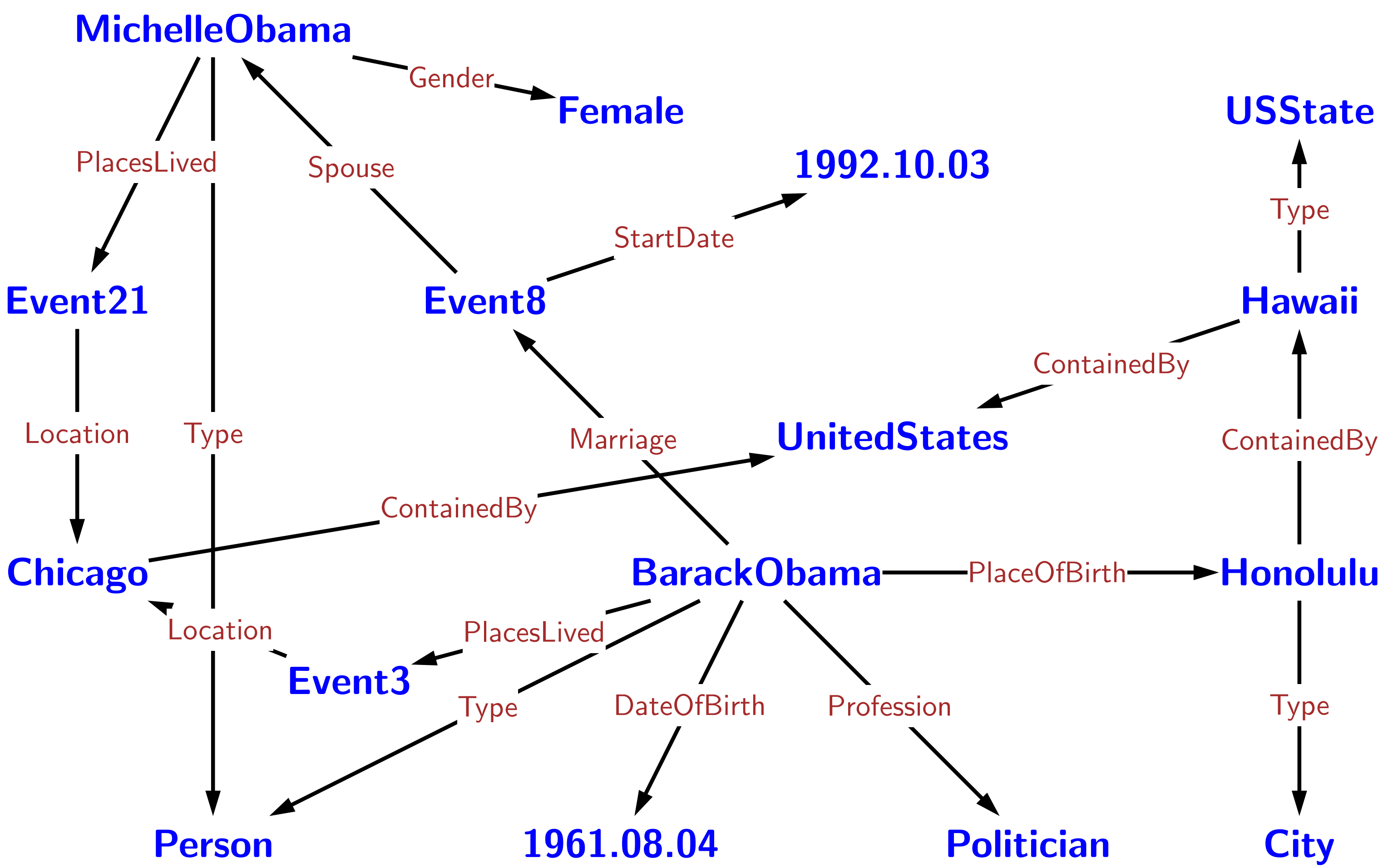


x
people who have lived in Chicago
parameters
θ
z
Type.Person ⊓ PlacesLived.Location.Chicago
world
w
y
{BarackObama,MichelleObama,...}

[Bollacker, 2008; Google, 2013]

World: Freebase

100M entities (nodes) 1B assertions (edges)



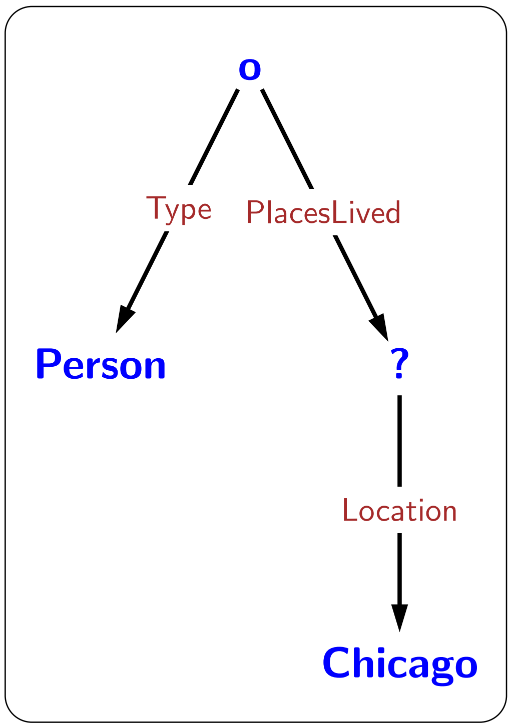
[Liang, 2013]

Logical forms

$\text{Type.Person} \sqcap \text{PlacesLived.Location.Chicago}$

Logical forms

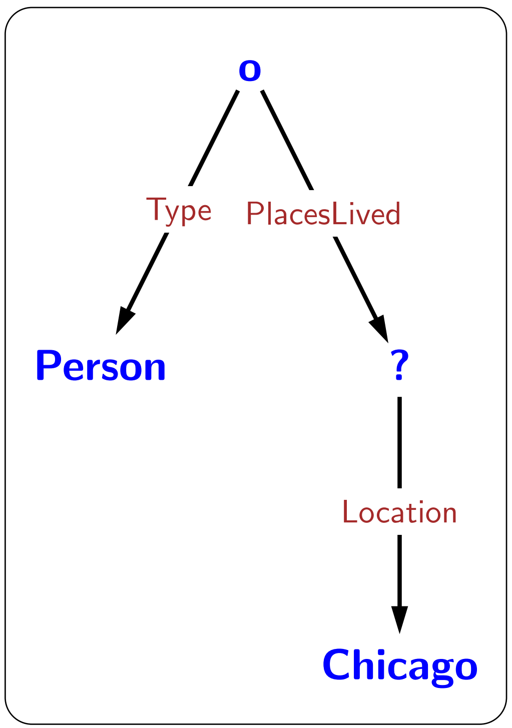
$\text{Type.Person} \sqcap \text{PlacesLived.Location.Chicago}$

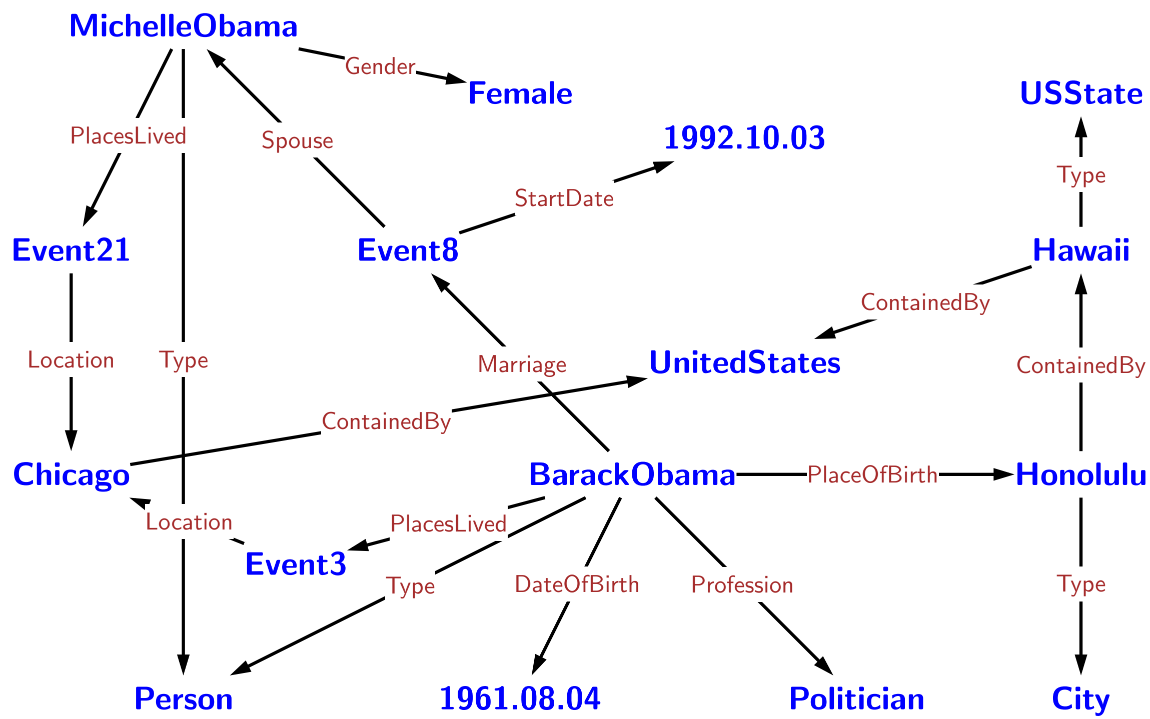


[Liang, 2013]

Logical forms

$\text{Type.Person} \sqcap \text{PlacesLived.Location.Chicago}$

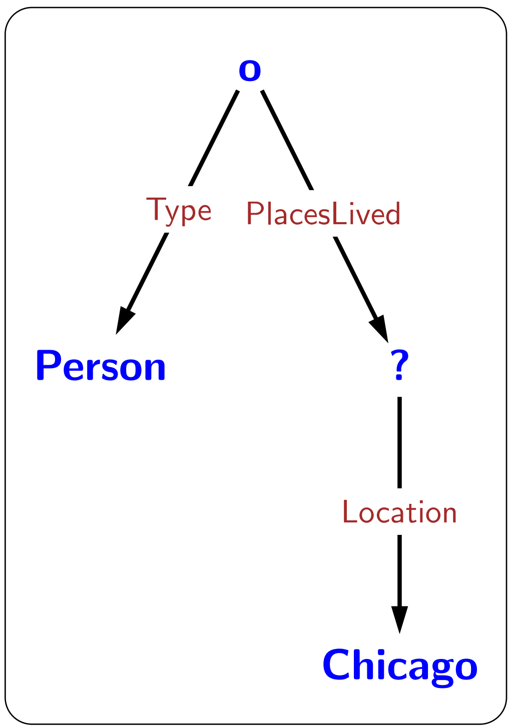


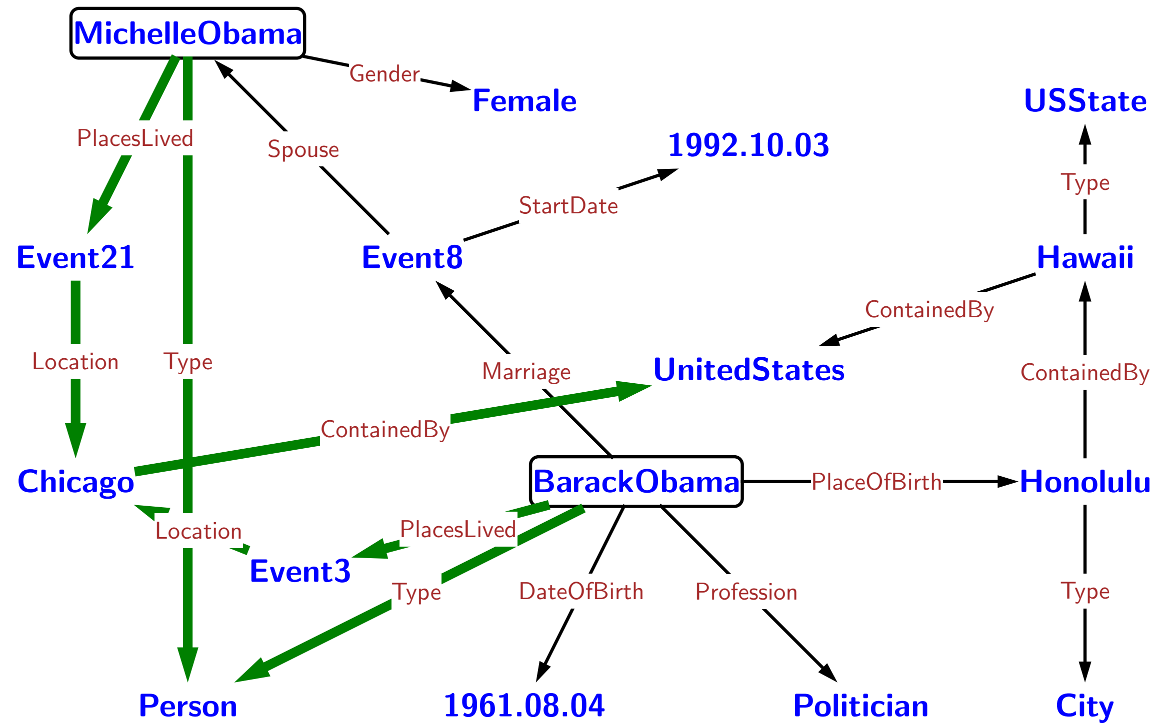


[Liang, 2013]

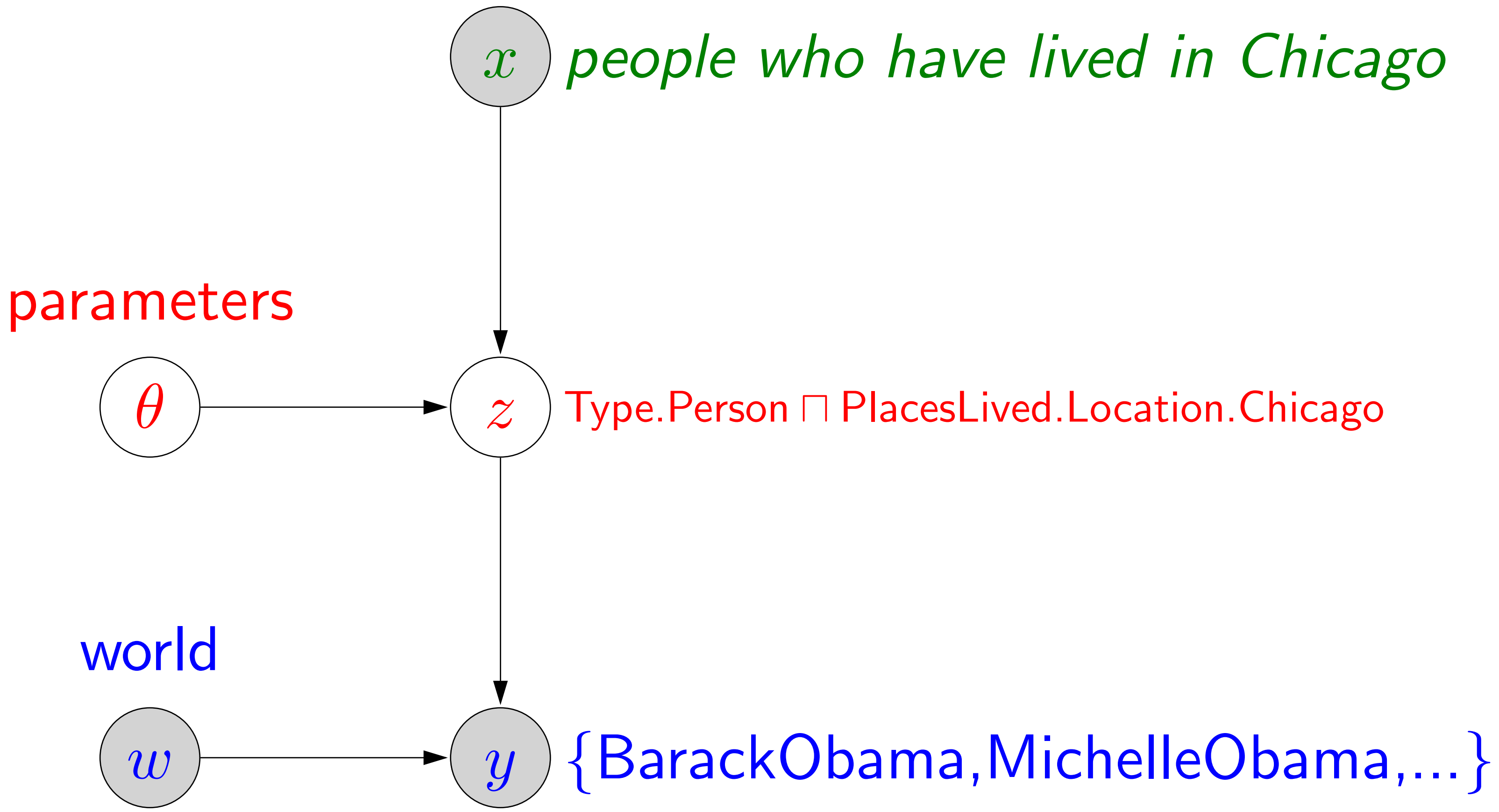
Logical forms

Type.Person ⊓ PlacesLived.Location.Chicago





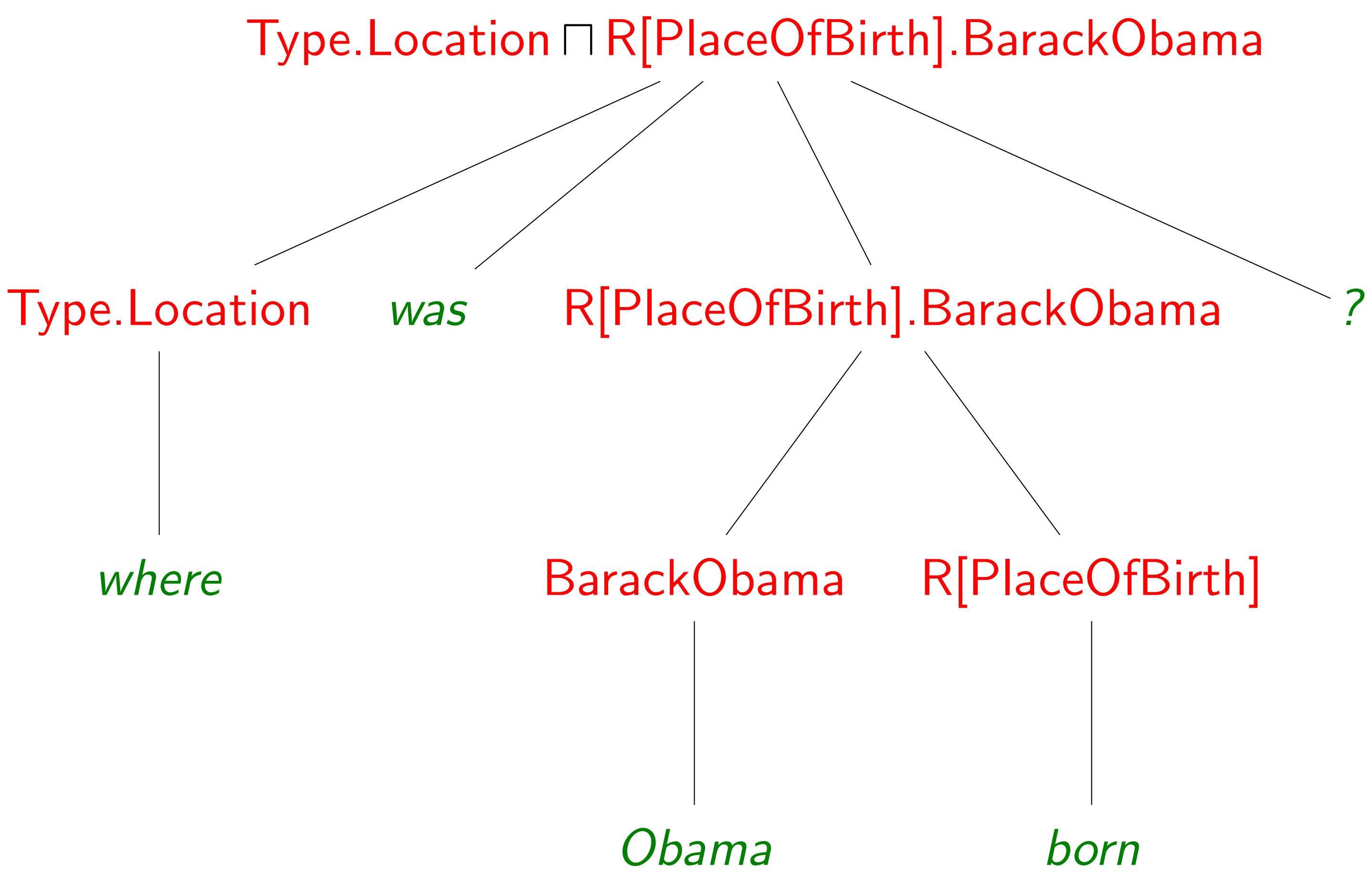
Framework



x
people who have lived in Chicago
parameters
θ
z
Type.Person ⊓ PlacesLived.Location.Chicago
world
w
y
{BarackObama,MichelleObama,...}

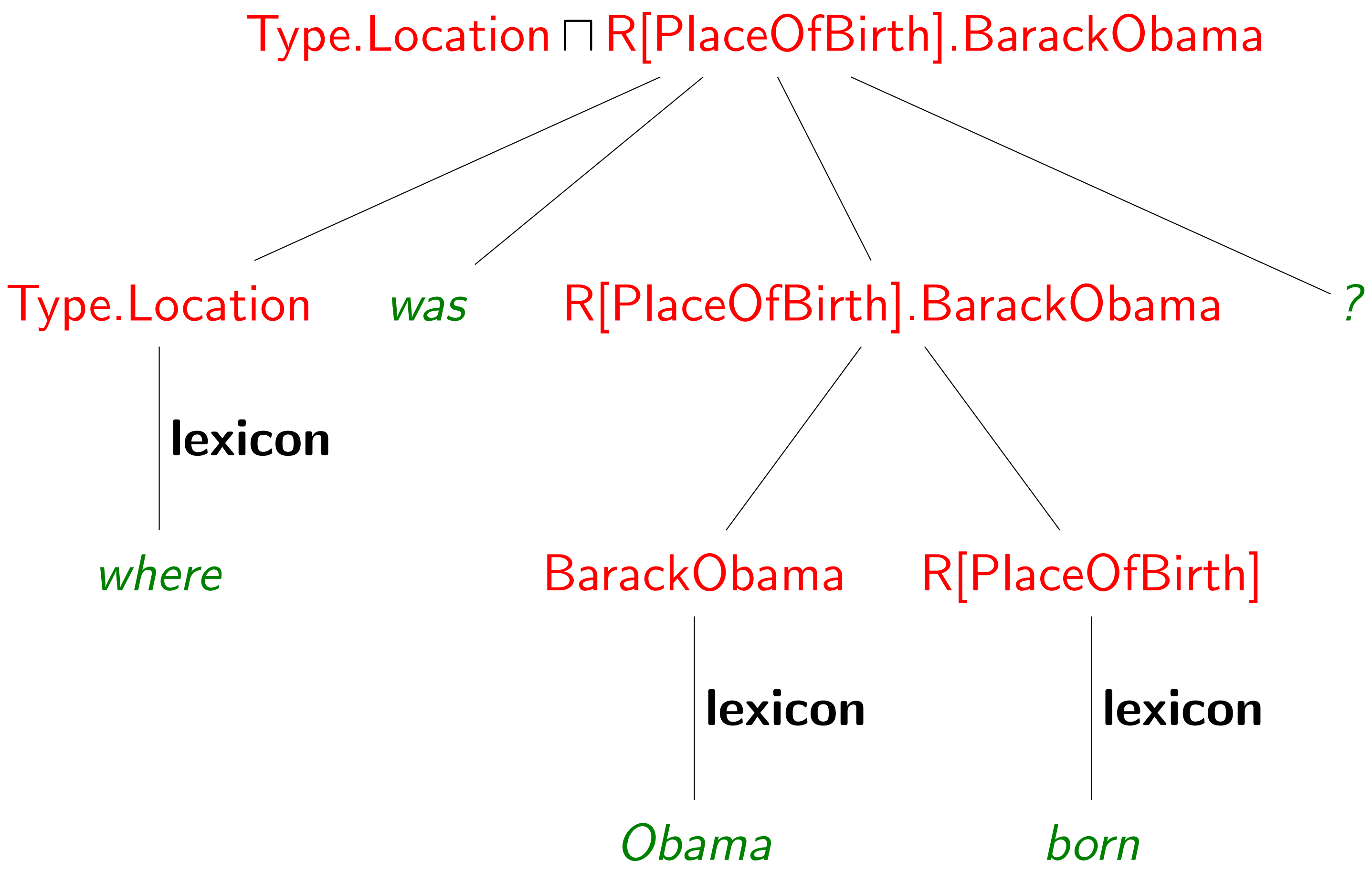
Derivations

Derivation: construction of logical form given utterance



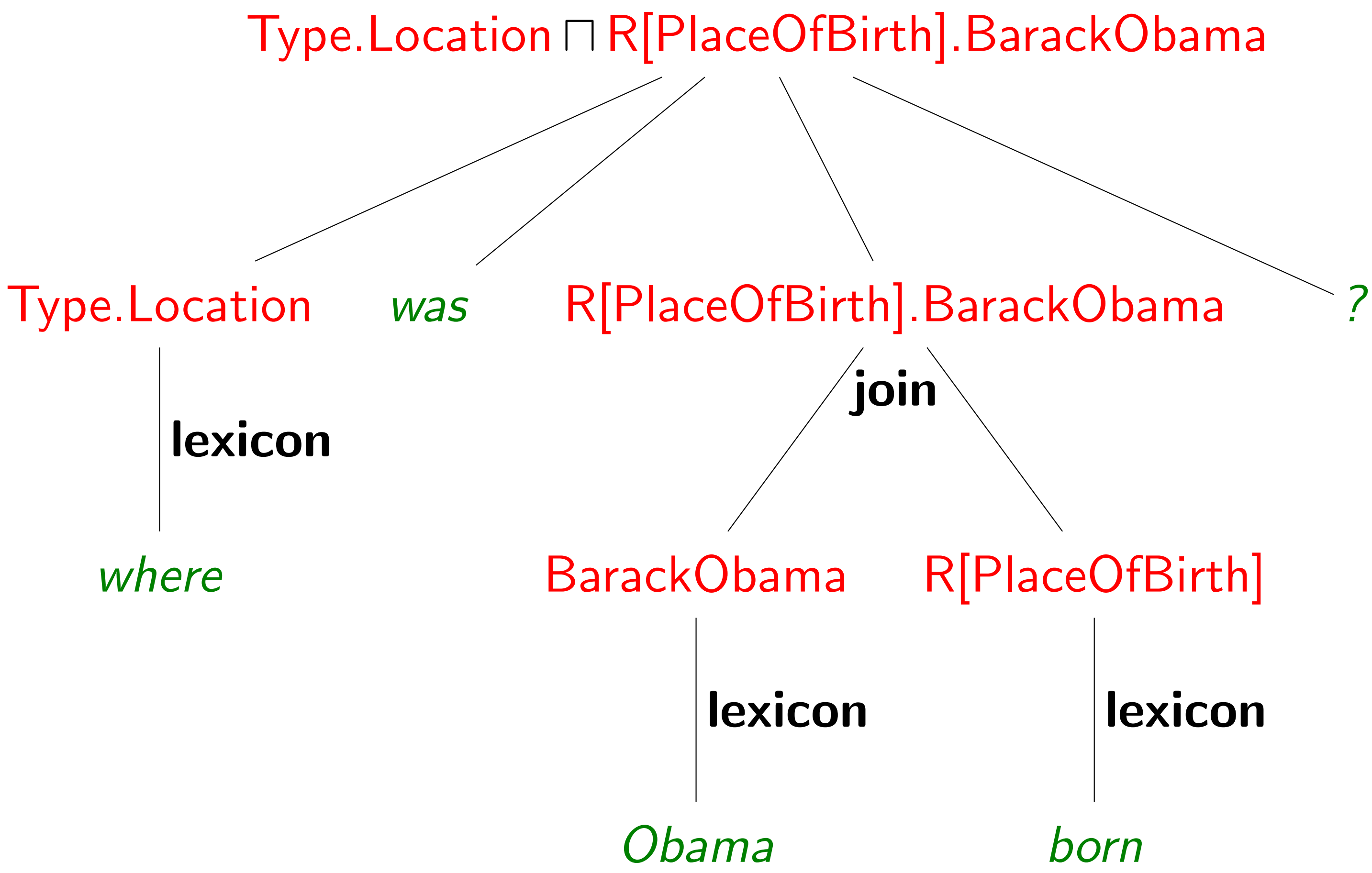
Derivations

Derivation: construction of logical form given utterance



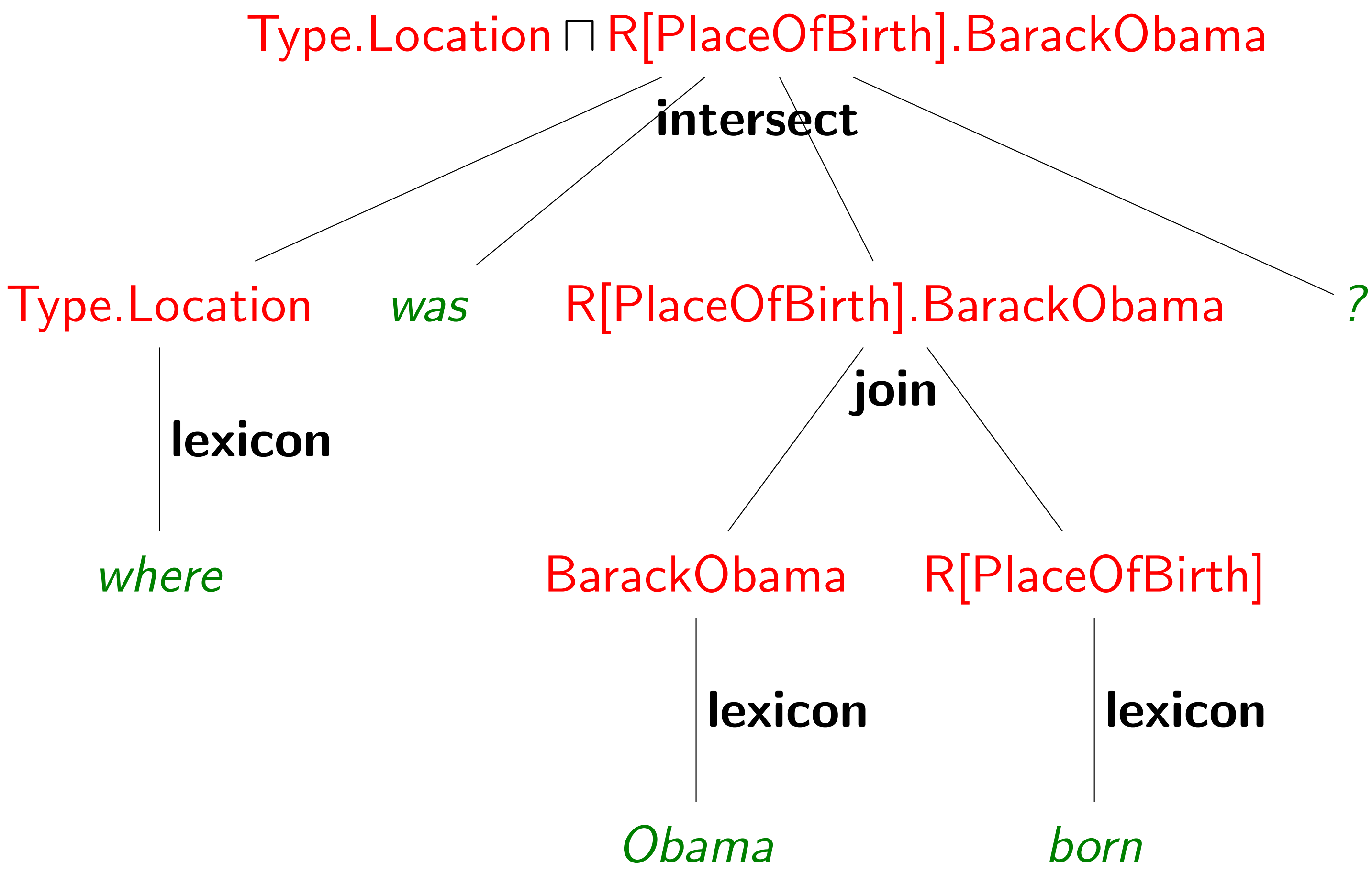
Derivations

Derivation: construction of logical form given utterance

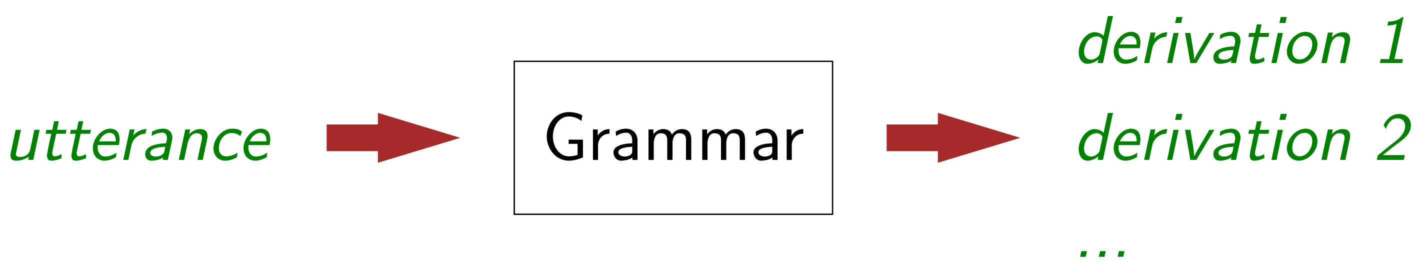


Derivations

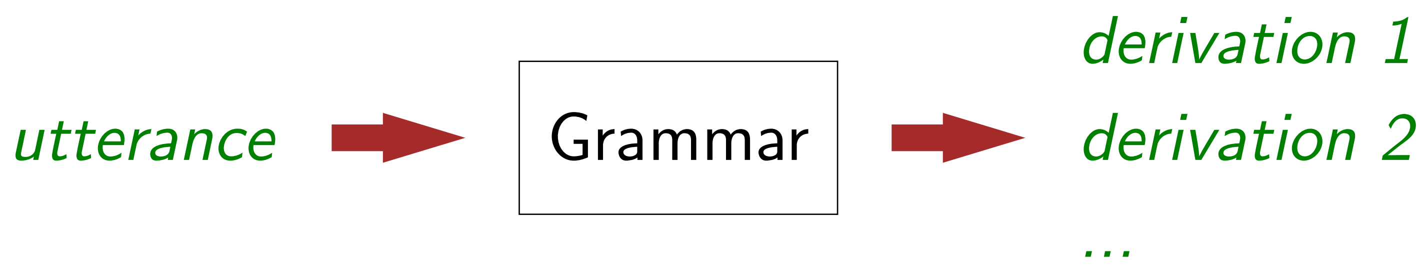
Derivation: construction of logical form given utterance



Grammar



Grammar



A Really Dumb Grammar

(lexicon) *Obama* $\Rightarrow$ Unary : BarackObama

(lexicon) *born* $\Rightarrow$ Binary : PlaceOfBirth

...

(join) Unary : u Binary : b $\Rightarrow$ Unary : $b.u$

(intersect) Unary : u Unary : v $\Rightarrow$ Unary : $u \sqcap v$

Many possible derivations!

Where was Obama born?

Many possible derivations!

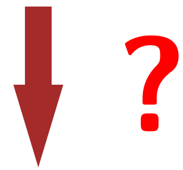
Where was Obama born?

?

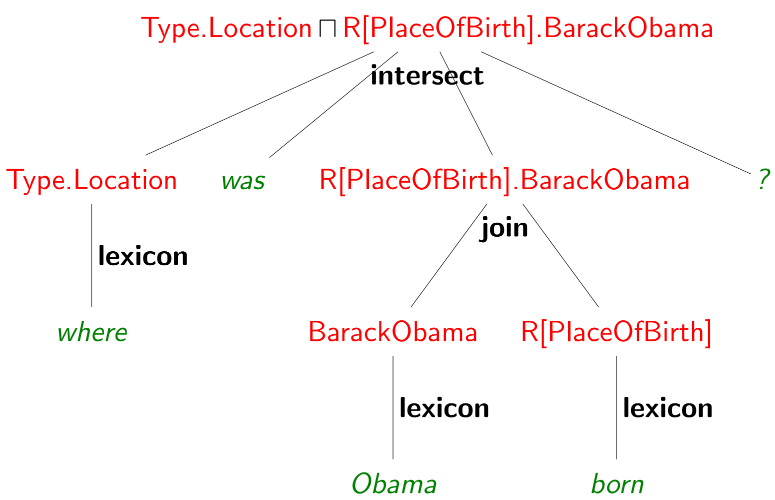
set of candidate derivations $\mathcal{D}(x)$

Many possible derivations!

Where was Obama born?

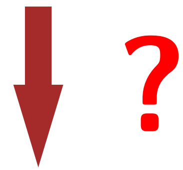


set of candidate derivations $\mathcal{D}(x)$

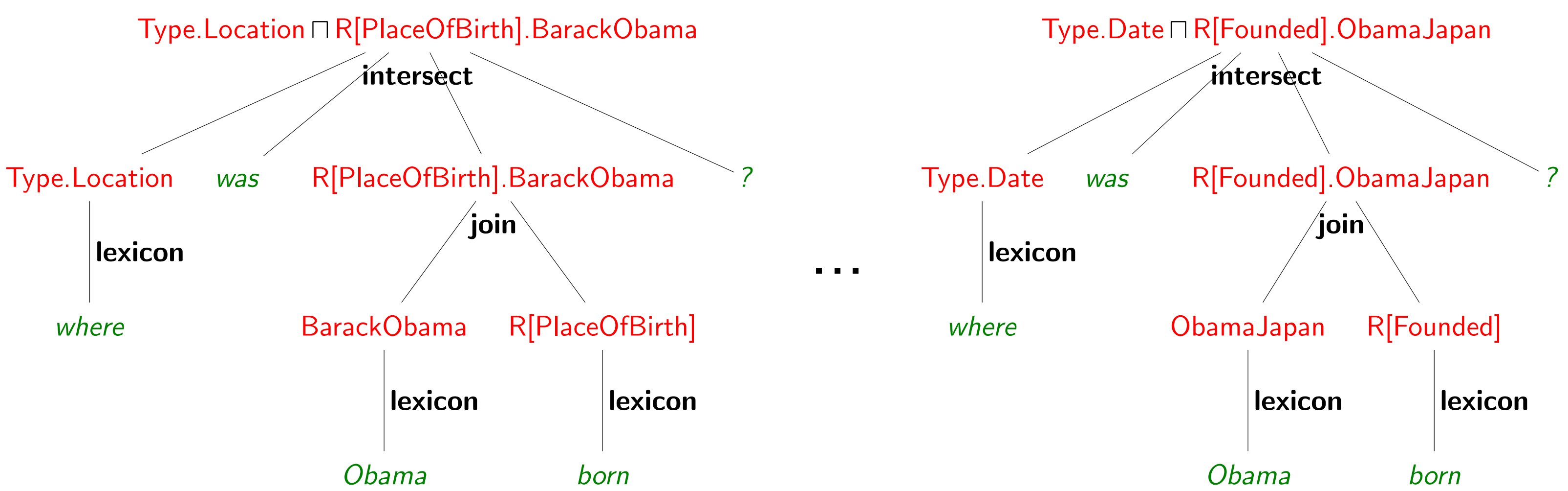


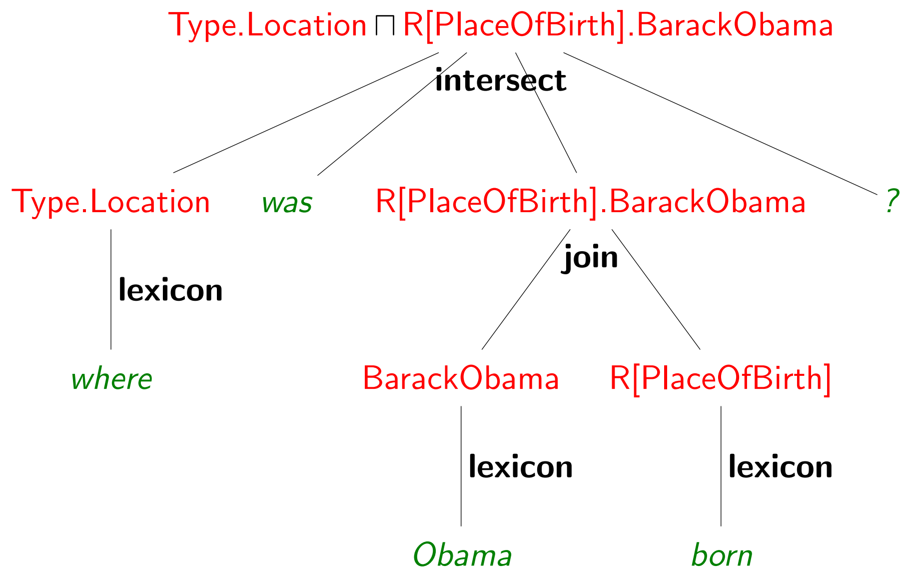
Many possible derivations!

Where was Obama born?



set of candidate derivations $\mathcal{D}(x)$





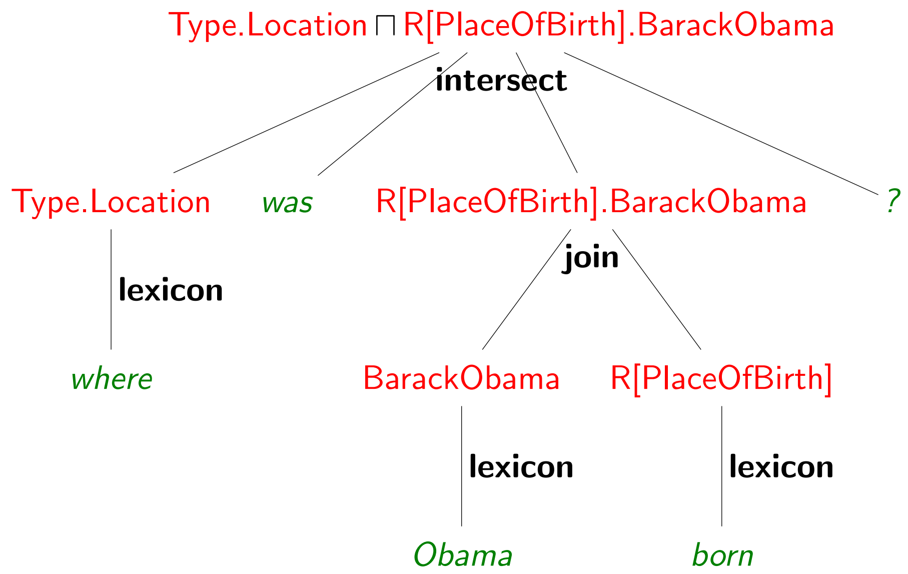
x: utterance

d: derivation

Feature vector $\phi(x, d) \in \mathbb{R}^f$:

x: utterance

d: derivation



Feature vector $\phi(x, d) \in \mathbb{R}^f$:

apply join	1
apply intersect	1
apply lexicon	3
skipped VBD-AUX	1
skipped NN	0
born maps to PlaceOfBirth	1
born maps to PlacesLived.Location	0
alignmentScore	1.52
denotation-size=1	1
...	...

Scoring derivations

Feature vector: $\phi(x, d) = [1.3, 2, 0, 1, 0, 0, \ldots] \in \mathbb{R}^F$

Parameter vector: $\theta = [1.2, -2.7, 3.4, \ldots] \in \mathbb{R}^F$

Scoring function:

$$\mathsf{Score}_\theta(x, d) = \phi(x, d) \cdot \theta$$

Log-linear model

Candidate derivations: $\mathcal{D}(x)$

Log-linear model

Candidate derivations: $\mathcal{D}(x)$

Model: distribution over derivations d given utterance x

$$p(d \mid x, \theta) = \frac{\exp(\mathsf{Score}_\theta(x,d))}{\sum_{d' \in \mathcal{D}(x)} \exp(\mathsf{Score}_\theta(x,d'))}$$

Learning

Training data:

What's Bulgaria's capital?

`Sofia`

When was Walmart started?

`1962`

What movies has Tom Cruise been in?

`TopGun,VanillaSky,...`

...

Learning

Training data:

What's Bulgaria's capital?

`Sofia`

When was Walmart started?

`1962`

What movies has Tom Cruise been in?

`TopGun,VanillaSky,...`

...

Objective: Maximum likelihood

$$\arg\max_\theta \sum_{i=1}^{n} \log p_\theta(y^{(i)} \mid x^{(i)})$$

Learning

Training data:

What's Bulgaria's capital?
`Sofia`
When was Walmart started?
`1962`
What movies has Tom Cruise been in?
`TopGun,VanillaSky,...`
...

Objective: Maximum likelihood

$$\arg\max_\theta \sum_{i=1}^n \log p_\theta(y^{(i)} \mid x^{(i)})$$

Algorithm:

AdaGrad (SGD with per-feature step size)

Training intuition

Where did Mozart tupress?

Vienna

Training intuition

Where did Mozart tupress?

`PlaceOfBirth.Mozart`

`PlaceOfDeath.Mozart`

`PlaceOfMarriage.Mozart`

Vienna

Training intuition

Where did Mozart tupress?

`PlaceOfBirth.Mozart` $\Rightarrow$ Salzburg

`PlaceOfDeath.Mozart` $\Rightarrow$ Vienna

`PlaceOfMarriage.Mozart` $\Rightarrow$ Vienna

Vienna

Training intuition

Where did Mozart tupress?

~~`PlaceOfBirth.Mozart`~~	~~$\Rightarrow$~~	~~Salzburg~~
`PlaceOfDeath.Mozart`	$\Rightarrow$	Vienna
`PlaceOfMarriage.Mozart`	$\Rightarrow$	Vienna

Vienna

Training intuition

Where did Mozart tupress?

~~`PlaceOfBirth.Mozart` ⇒ Salzburg~~

`PlaceOfDeath.Mozart` ⇒ Vienna

`PlaceOfMarriage.Mozart` ⇒ Vienna

Vienna

Where did William Hogarth tuppress?

Training intuition

Where did Mozart tupress?

~~`PlaceOfBirth.Mozart` ⇒ Salzburg~~

`PlaceOfDeath.Mozart` ⇒ Vienna

`PlaceOfMarriage.Mozart` ⇒ Vienna

Vienna

Where did William Hogarth tuppress?

`PlaceOfBirth.WilliamHogarth`

`PlaceOfDeath.WilliamHogarth`

`PlaceOfMarriage.WilliamHogarth`

London

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth.Mozart~~	~~⇒~~	~~Salzburg~~
PlaceOfDeath.Mozart	⇒	Vienna
PlaceOfMarriage.Mozart	⇒	Vienna

Vienna

Where did William Hogarth tuppress?

PlaceOfBirth.WilliamHogarth	⇒	London
PlaceOfDeath.WilliamHogarth	⇒	London
PlaceOfMarriage.WilliamHogarth	⇒	Paddington

London

Training intuition

Where did Mozart tupress?

~~`PlaceOfBirth.Mozart`~~	~~⇒~~	~~Salzburg~~
`PlaceOfDeath.Mozart`	⇒	Vienna
`PlaceOfMarriage.Mozart`	⇒	Vienna

Vienna

Where did William Hogarth tuppress?

`PlaceOfBirth.WilliamHogarth`	⇒	London
`PlaceOfDeath.WilliamHogarth`	⇒	London
~~`PlaceOfMarriage.WilliamHogarth`~~	~~⇒~~	~~Paddington~~

London

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth.Mozart ⇒ Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tuppress?

PlaceOfBirth.WilliamHogarth ⇒ London

PlaceOfDeath.WilliamHogarth ⇒ London

~~PlaceOfMarriage.WilliamHogarth ⇒ Paddington~~

London

Outline

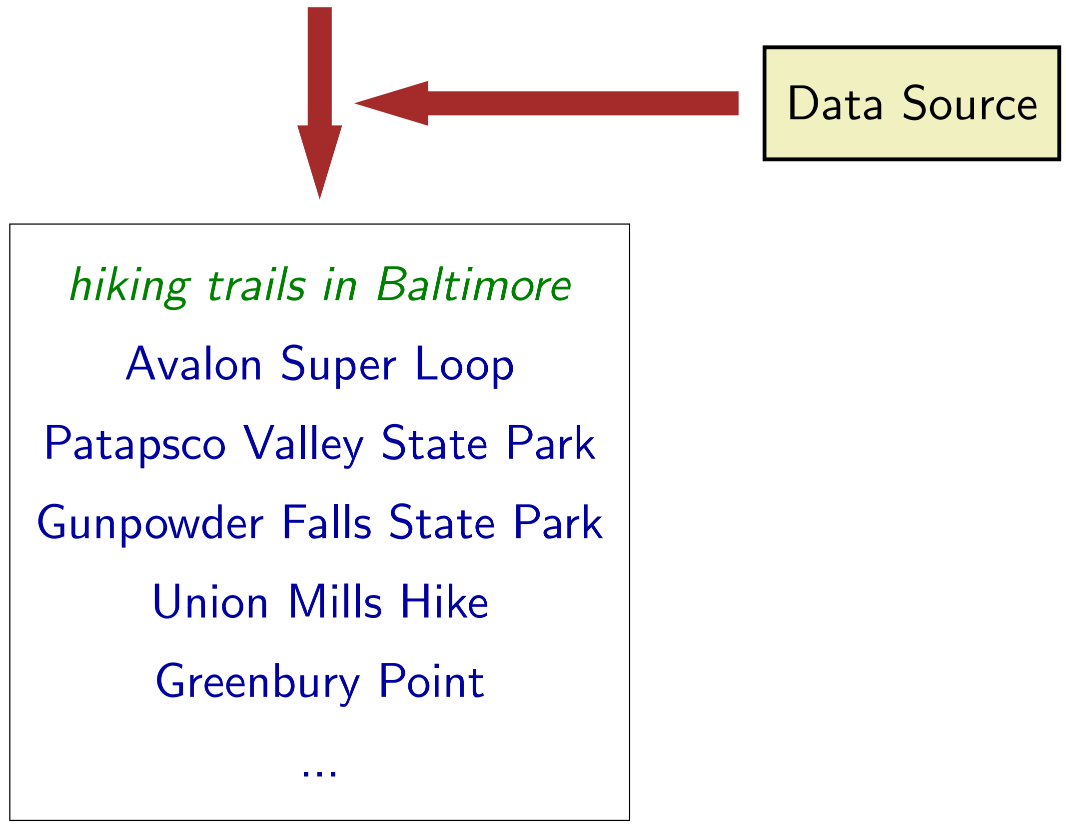
- Semantic parsing in 5 minutes
- **A closer look at the elements**
  - Knowledge base incompleteness
  - Lexical coverage
  - Search over logical forms
  - Learning via bootstrapping
  - Leveraging denotations ("grounding")
  - Datasets
- Final remarks

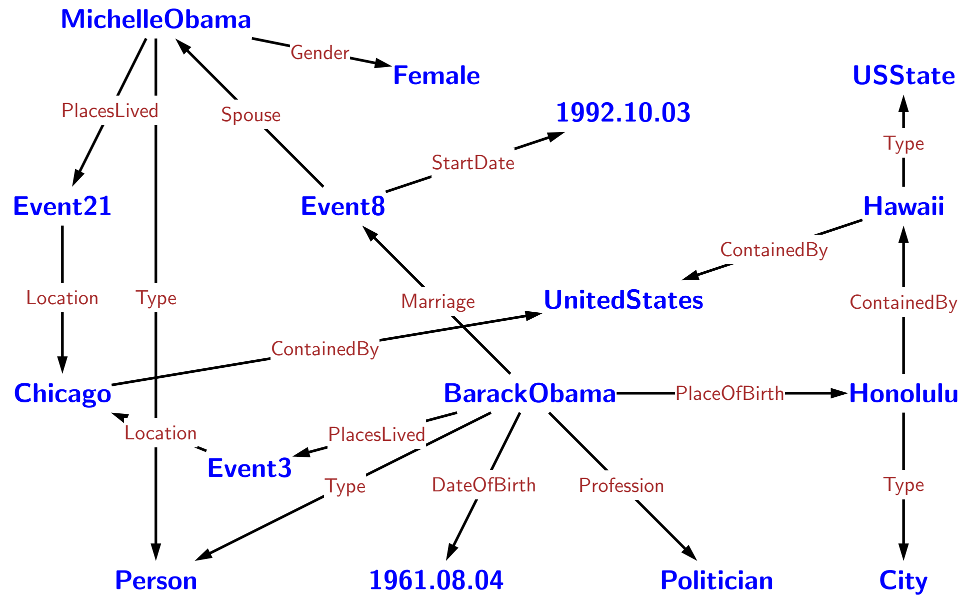
Outline

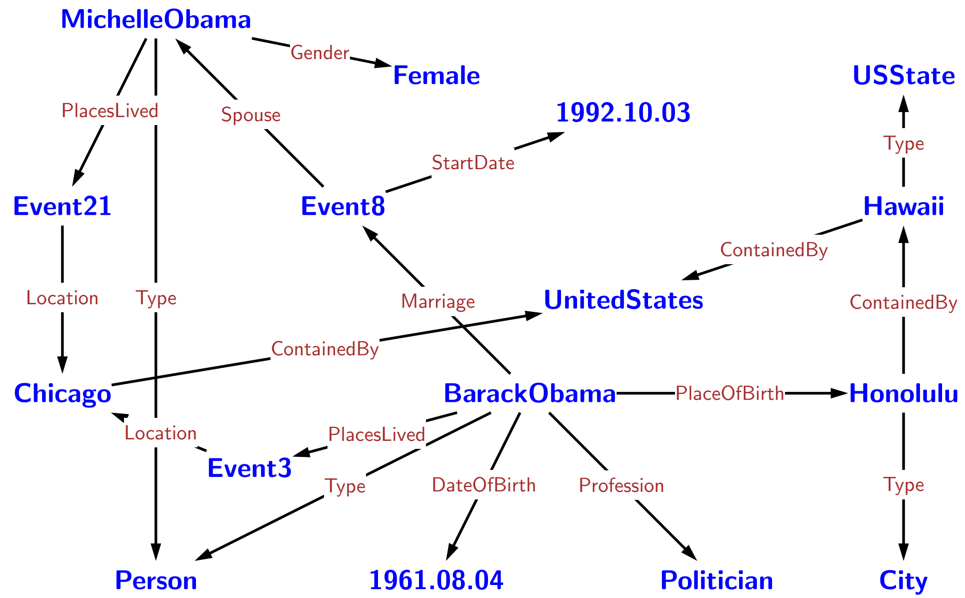
- Semantic parsing in 5 minutes
- A closer look at the elements
  - **Knowledge base incompleteness**
  - Lexical coverage
  - Search over logical forms
  - Learning via bootstrapping
  - Leveraging denotations ("grounding")
  - Datasets
- Final remarks

Challenge: incomplete knowledge base

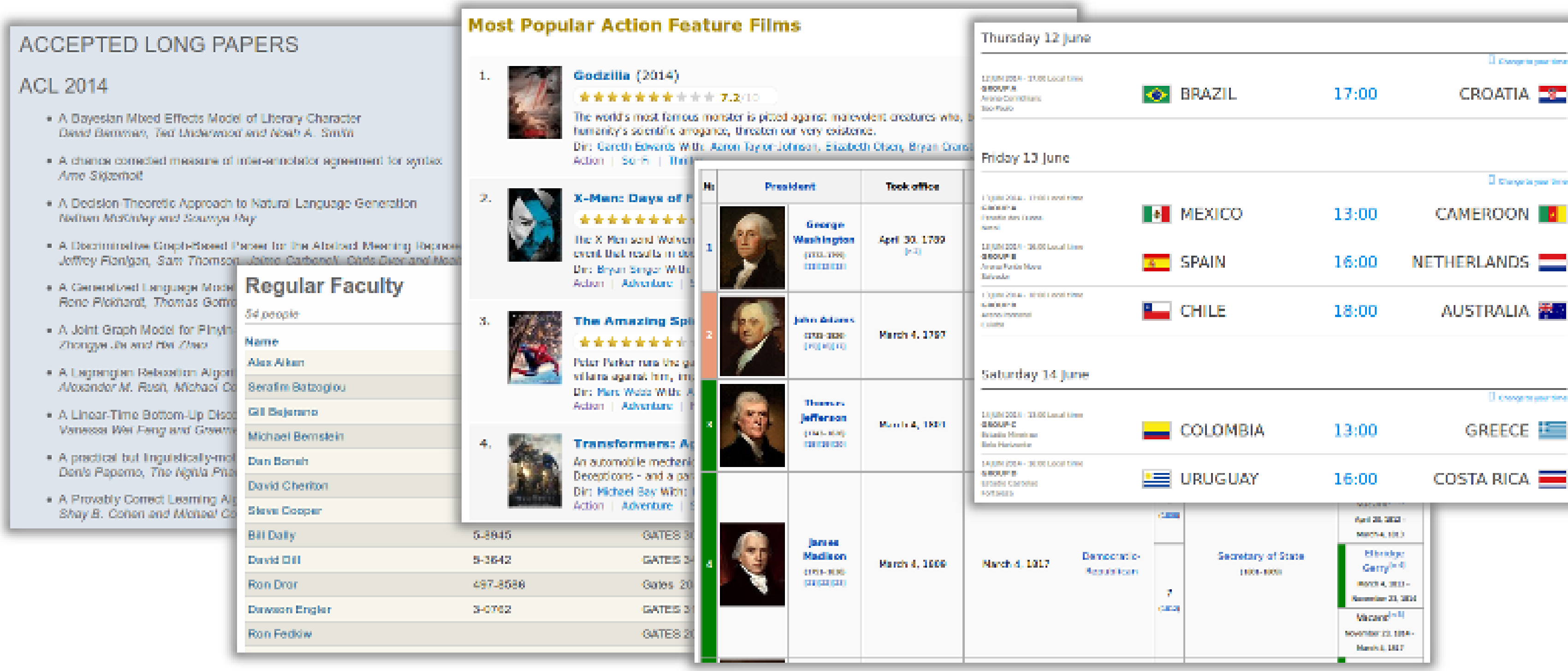
What are the longest **hiking trails in Baltimore***?*




MichelleObama
Gender
Female
USState
PlacesLived
Spouse
1992.10.03
Type
StartDate
Event21
Event8
Hawaii
ContainedBy
Location
Type
Marriage
UnitedStates
ContainedBy
ContainedBy
Chicago
BarackObama
PlaceOfBirth
Honolulu
Location
PlacesLived
Event3
Type
DateOfBirth
Profession
Type
Person
1961.08.04
Politician
City



Fewer than 10% general web questions can be answered via Freebase


ACCEPTED LONG PAPERS
ACL 2014
Regular Faculty
Name
Alex Aiken
Serafim Batzoglou
Gill Bejerano
Michael Bernstein
Dan Boneh
David Cheriton
Steve Cooper
Bill Dally
David Dill
Ron Dror
Dawson Engler
Ron Fedkiw
Most Popular Action Feature Films
Godzilla (2014)
President
Took office
George Washington
John Adams
Thomas Jefferson
James Madison
Thursday 12 June
BRAZIL 17:00 CROATIA
Friday 13 June
MEXICO 13:00 CAMEROON
SPAIN 16:00 NETHERLANDS
CHILE 18:00 AUSTRALIA
Saturday 14 June
COLOMBIA 13:00 GREECE
URUGUAY 16:00 COSTA RICA

[Pasupat & Liang, 2014]

Semantic parsing on the web

Input:

- query x

hiking trails near Baltimore

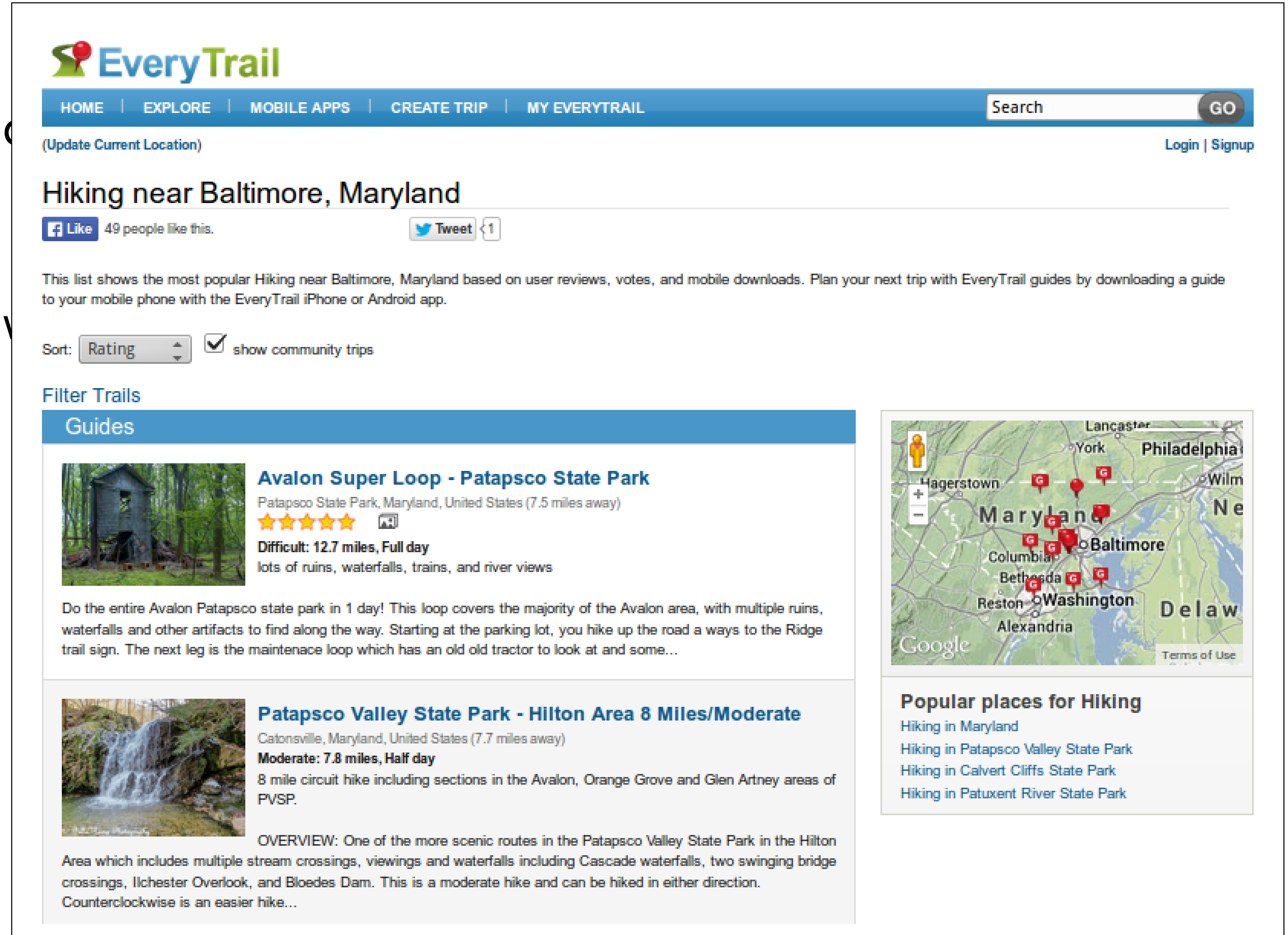
- web page w

[Pasupat & Liang, 2014]

Semantic parsing on the web

Input:

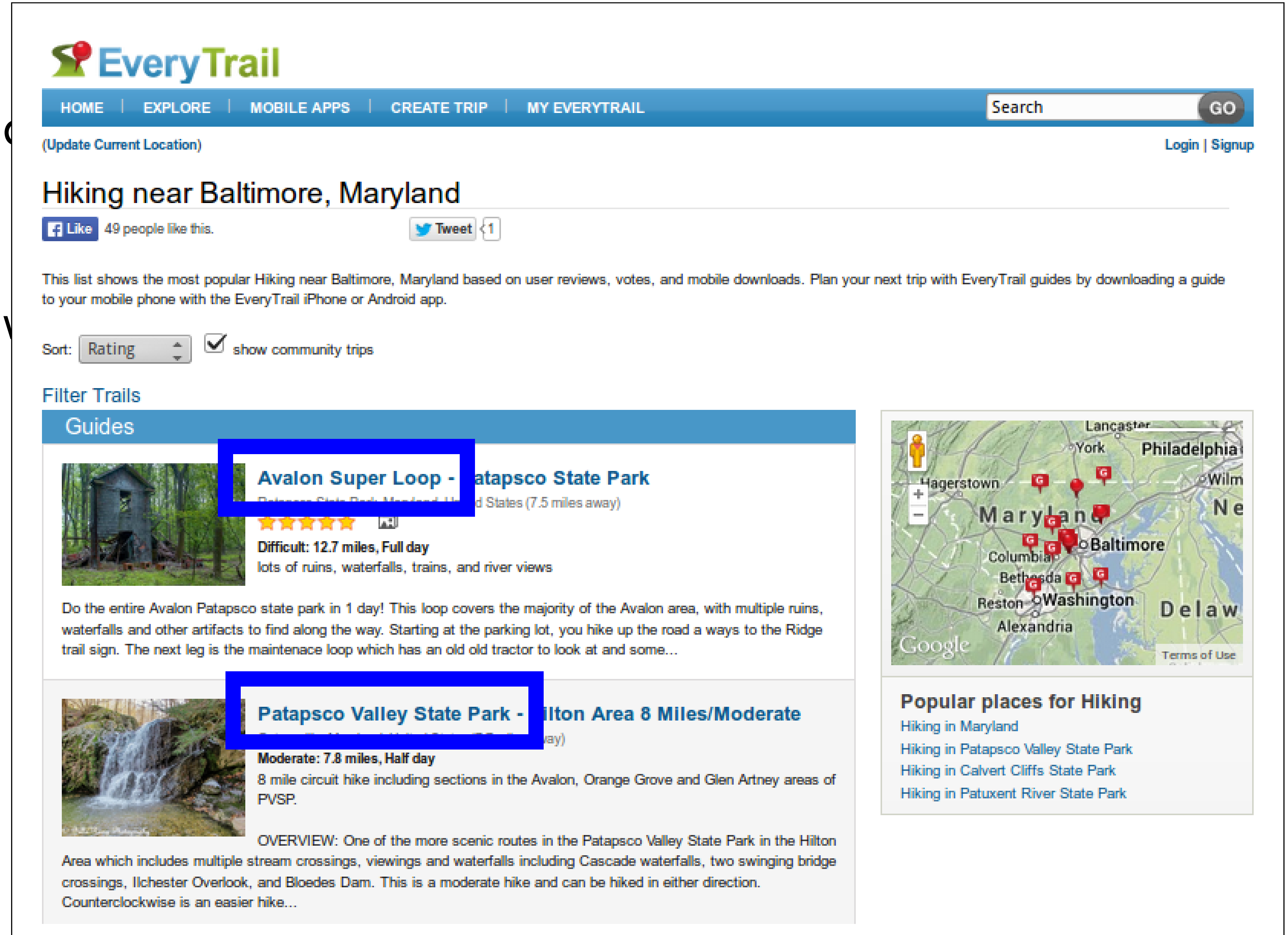
- C
- V



Semantic parsing on the web

Input:

- (
- V



[Pasupat & Liang, 2014]

Semantic parsing on the web

Input:

- query x

hiking trails near Baltimore

- web page w

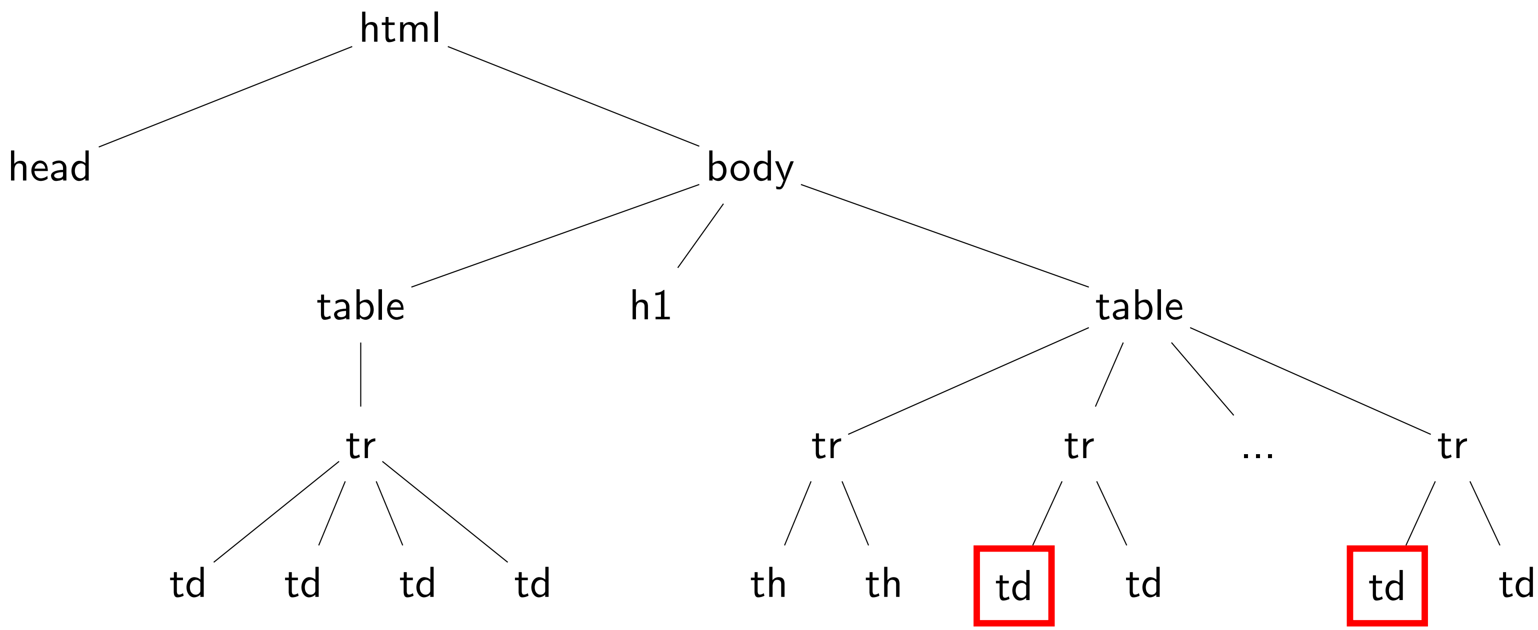
Output:

- list of entities y

[Avalon Super Loop, Patapsco Valley State Park, ...]

[Sahuguet and Azavant, 1999; Liu et al., 2000; Crescenzi et al., 2001]

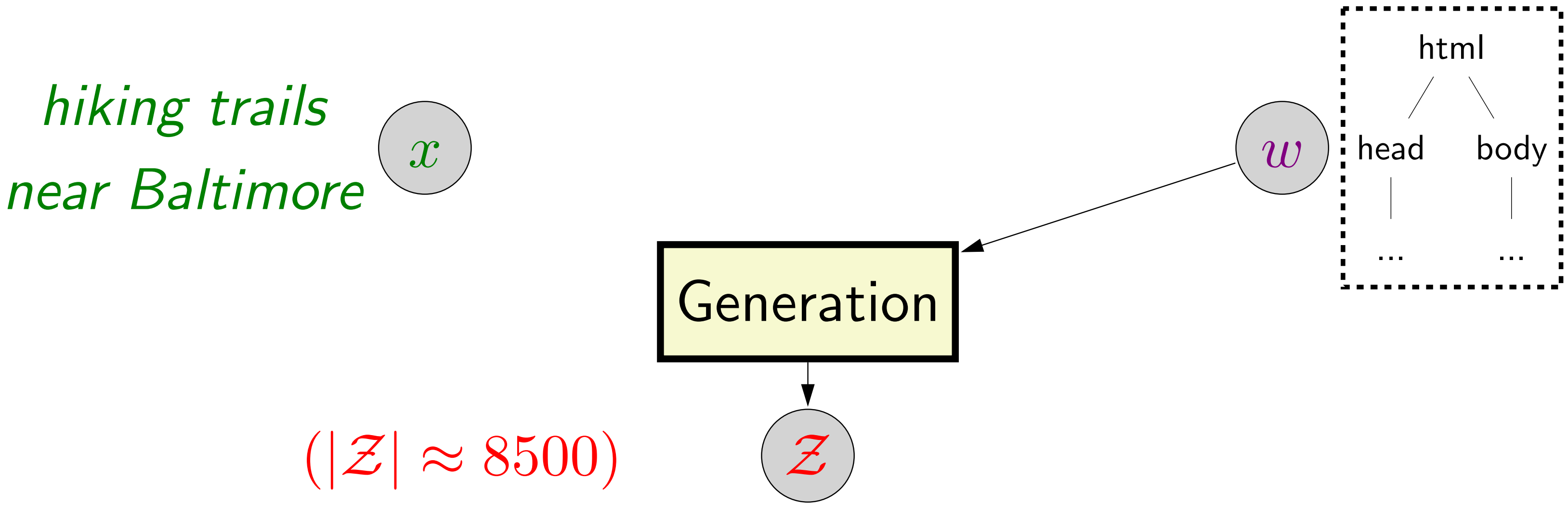
Logical forms: XPath expressions



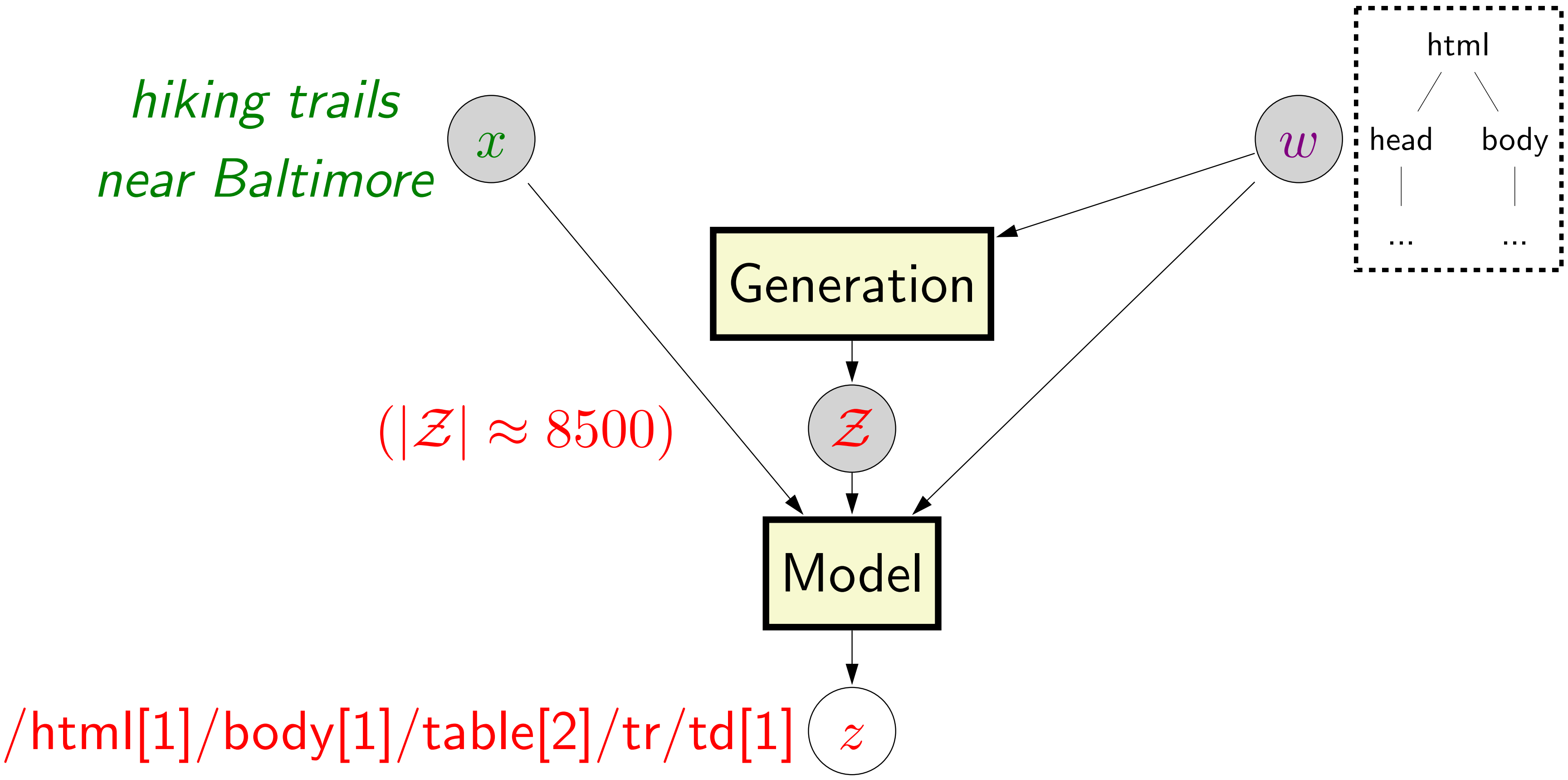
$z =$ `/html[1]/body[1]/table[2]/tr/td[1]`

Framework

Framework

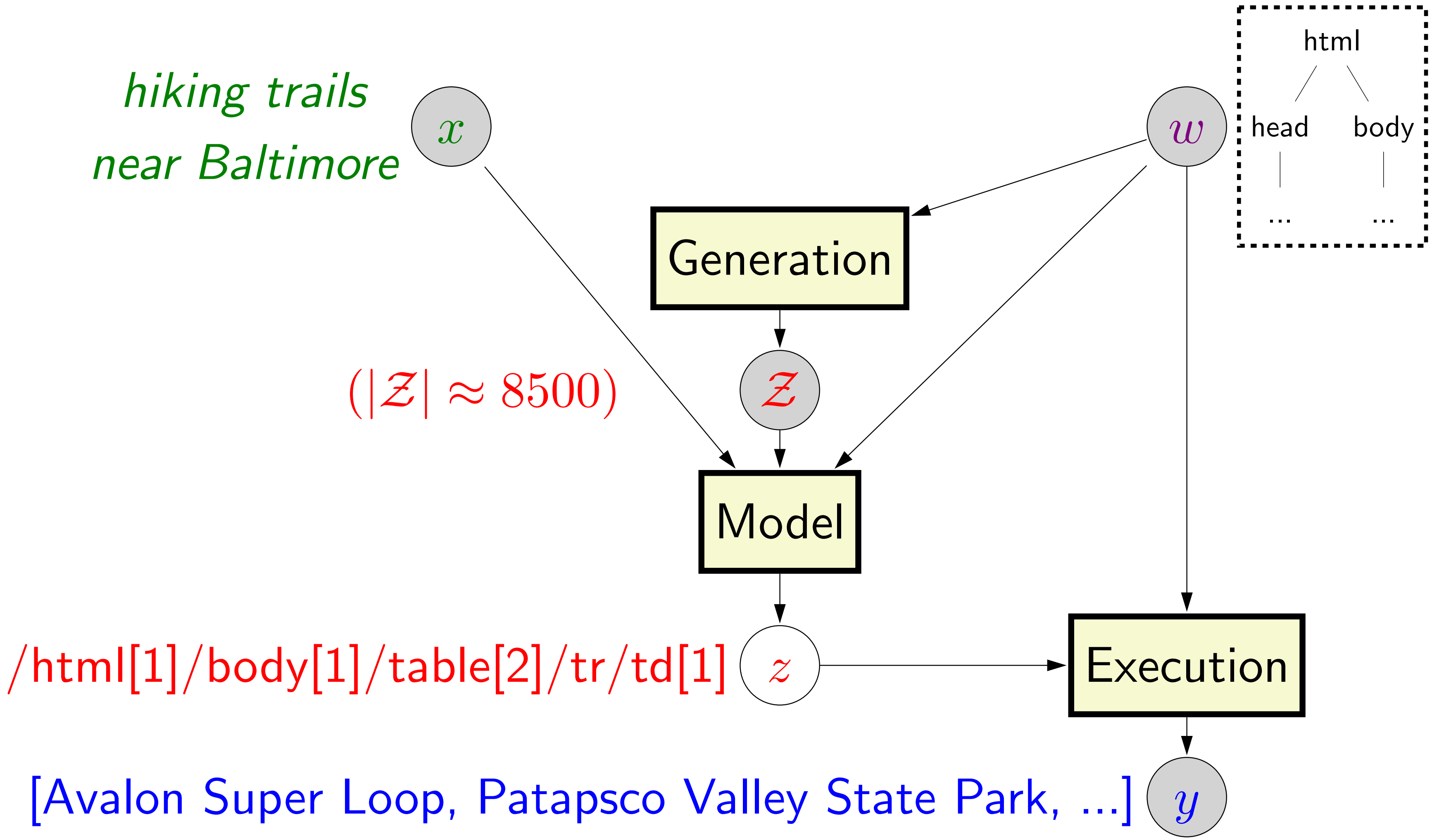

hiking trails
near Baltimore
x
w
html
head
body
...
...
Generation
$\mathcal{Z}$
$(|\mathcal{Z}| \approx 8500)$

Framework



hiking trails
near Baltimore
x
w
html
head
body
...
...
Generation
(|Z| ≈ 8500)
Z
Model
/html[1]/body[1]/table[2]/tr/td[1]
z

Framework



hiking trails
near Baltimore
x
w
html
head
body
...
...
Generation
(|Z| ≈ 8500)
Z
Model
/html[1]/body[1]/table[2]/tr/td[1]
z
Execution
[Avalon Super Loop, Patapsco Valley State Park, ...]
y

Outline

- Semantic parsing in 5 minutes
- A closer look at the elements
  - Knowledge base incompleteness
  - **Lexical coverage**
  - Search over logical forms
  - Learning via bootstrapping
  - Leveraging denotations ("grounding")
  - Datasets
- Final remarks

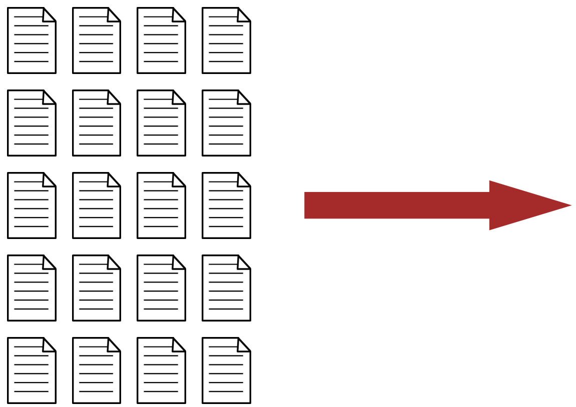
Challenge: lexical coverage

born $\Rightarrow$ Type.City, PeopleBornHere, Profession.Lawyer, ...

?

Solution: alignment

Open information extraction on ClueWeb09:



(*Barack Obama*, *was born in*, *Honolulu*)

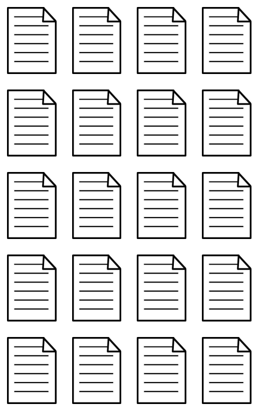
(*Albert Einstein*, *was born in*, *Ulm*)

(*Barack Obama*, *lived in*, *Chicago*)

... 15M triples ...

Solution: alignment

Open information extraction on ClueWeb09:



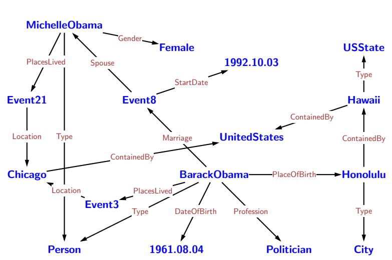
(*Barack Obama*, *was born in*, *Honolulu*)

(*Albert Einstein*, *was born in*, *Ulm*)

(*Barack Obama*, *lived in*, *Chicago*)

... 15M triples ...

Freebase:



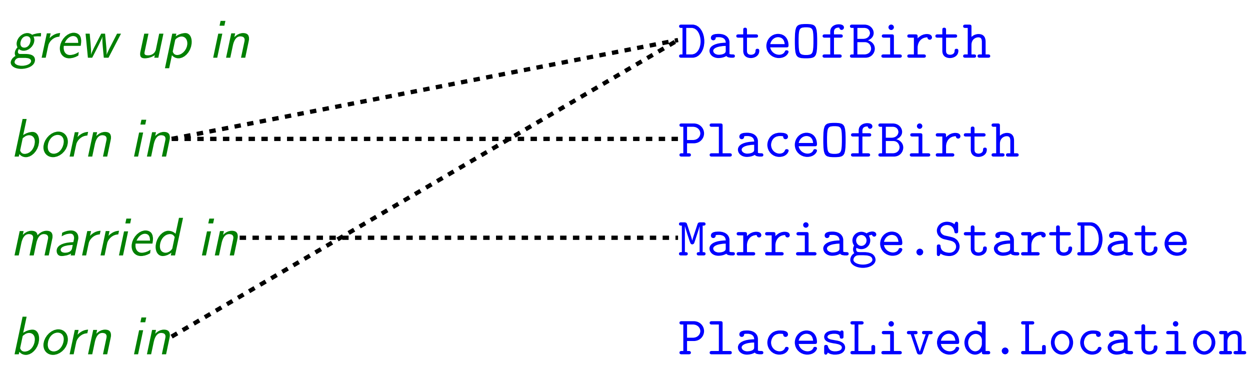
(`BarackObama`, `PlaceOfBirth`, `Honolulu`)

(`Albert Einstein`, `PlaceOfBirth`, `Ulm`)

(`BarackObama`, `PlacesLived.Location`, `Chicago`)

... 400M triples ...

Match text and Freebase predicates



Similar schema matching / alignment ideas [Cai & Yates, 2013, Fader et. al, 2013, Yao & van Durme, 2014; etc.]

Challenge: variability in language

What is the currency in the US?

Challenge: variability in language

What is the currency in the US?

What money do they use in the states?

How do you pay in America?

What's the currency of the US?

What money is accepted in the United States?

What money to take to the US?

...

A solution: paraphrasing

How many people live in Seattle?

paraphrase

What is the population of Seattle?

PopulationOf(Seattle)

`850,000`

Convert to a text-only problem

Challenge: "sub-lexical compositionality"

grandmother

$$\lambda x.\text{Gender.Female} \sqcap \text{Parent.Parent}.x$$

mayor

$$\lambda x.\text{GovtPositionsHeld.}(\text{Title.Mayor} \sqcap \text{OfficeOfJurisdiction}.x)$$

Challenge: "sub-lexical compositionality"

grandmother

$$\lambda x.\text{Gender.Female} \sqcap \text{Parent.Parent}.x$$

mayor

$$\lambda x.\text{GovtPositionsHeld.}(\text{Title.Mayor} \sqcap \text{OfficeOfJurisdiction}.x)$$

presidents who have served two **non-consecutive** *terms*

[requires higher-order quantification]

presidents who were **previously** *vice-presidents*

[anaphora]

every other *president*

[weird quantification anaphora]

Outline

- Semantic parsing in 5 minutes
- A closer look at the elements
  - Knowledge base incompleteness
  - Lexical coverage
  - **Search over logical forms**
  - Learning via bootstrapping
  - Leveraging denotations ("grounding")
  - Datasets
- Final remarks

Many possible derivations!

Where was Obama born?

A Really Dumb Grammar

(lexicon) *Obama* $\Rightarrow$ Unary : BarackObama

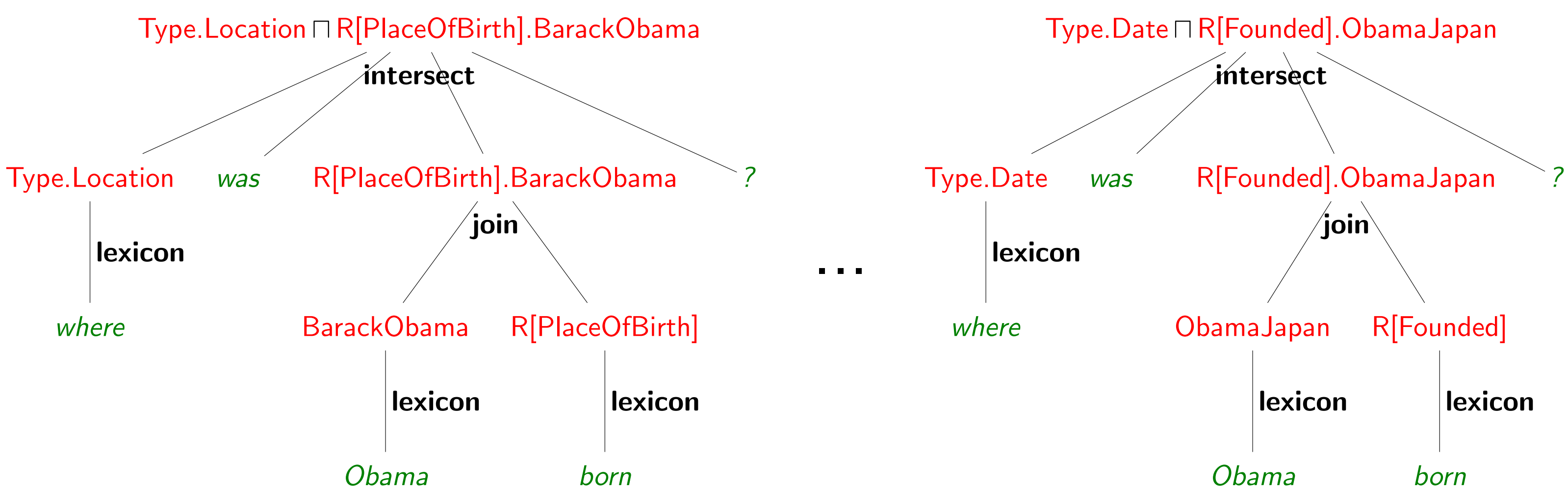
(lexicon) *born* $\Rightarrow$ Binary : PlaceOfBirth

...

(join) Unary : u Binary : b $\Rightarrow$ Unary : $b.u$

(intersect) Unary : u Unary : v $\Rightarrow$ Unary : $u \sqcap v$

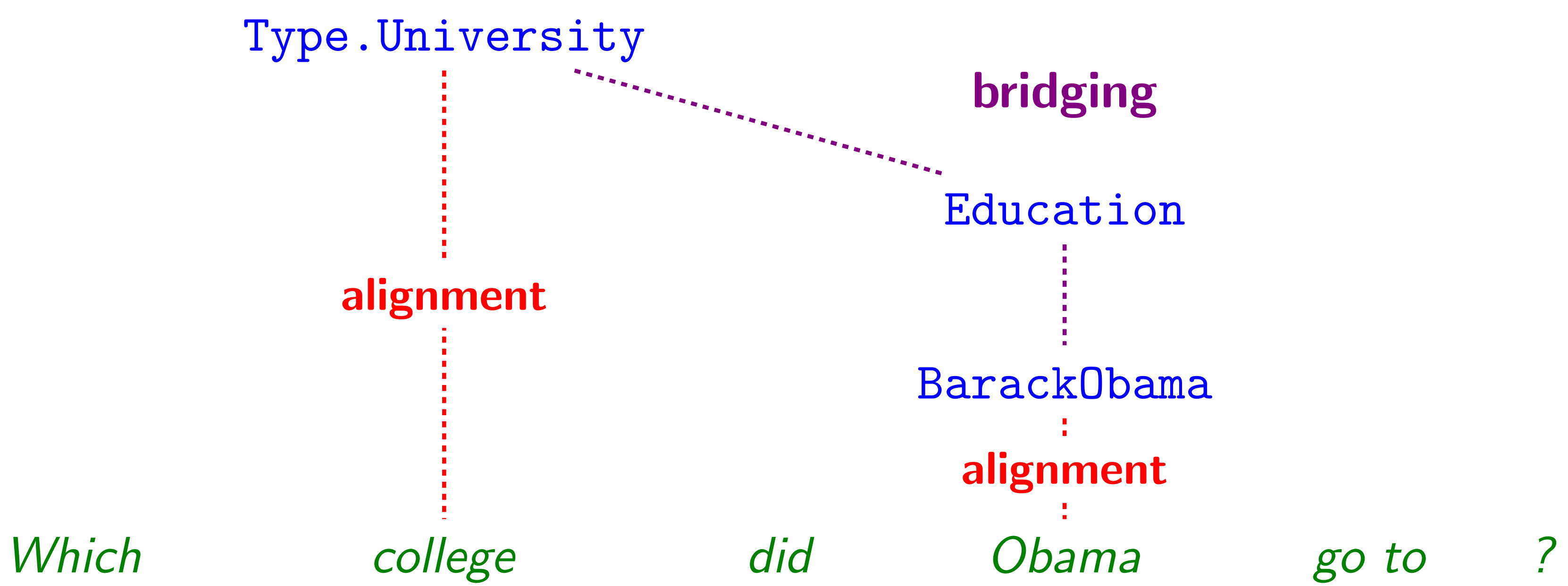
set of candidate derivations $\mathcal{D}(x)$



Bridging

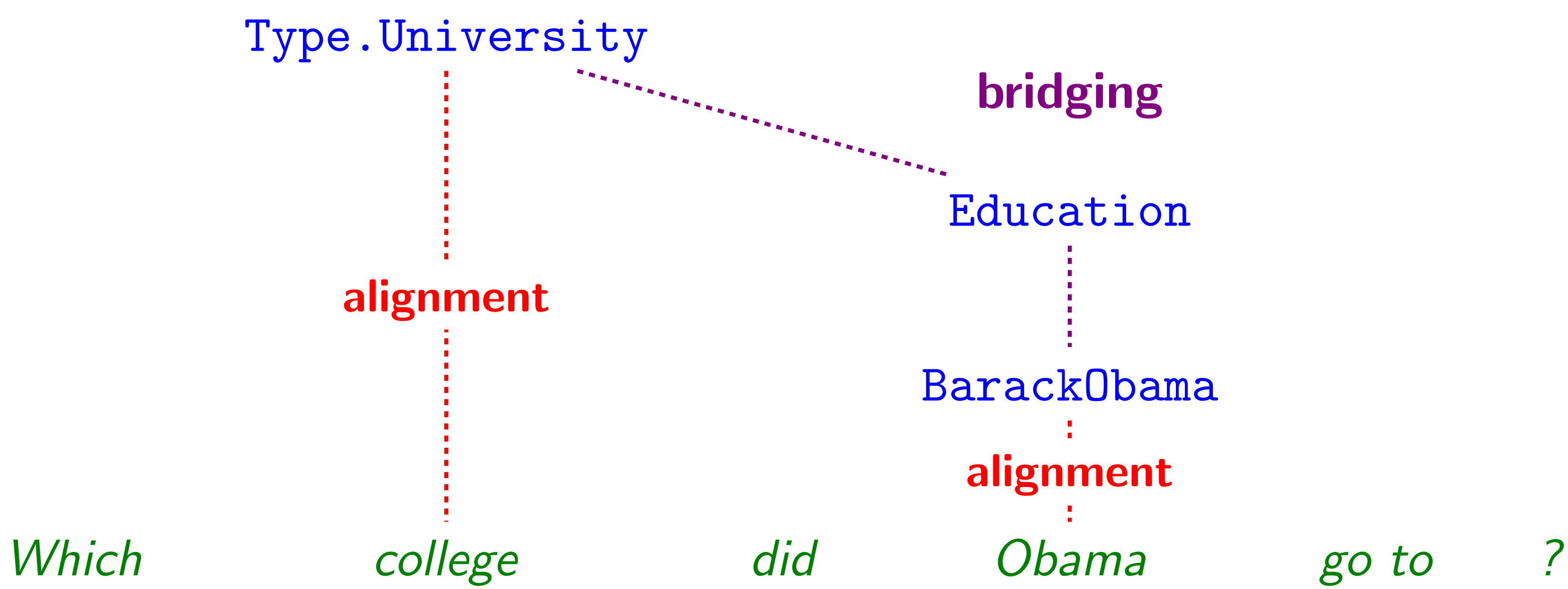

Type.University
alignment
BarackObama
alignment
Which
college
did
Obama
go to
?

Bridging



Bridging: use neighboring predicates / type constraints

Bridging



Bridging: use neighboring predicates / type constraints

Start building from parts with more certainty

[Berant & Liang, 2014]

Bridging to nowhere

Search logical forms based on "prior":

What countries in the world speak Arabic?

Bridging to nowhere

Search logical forms based on "prior":

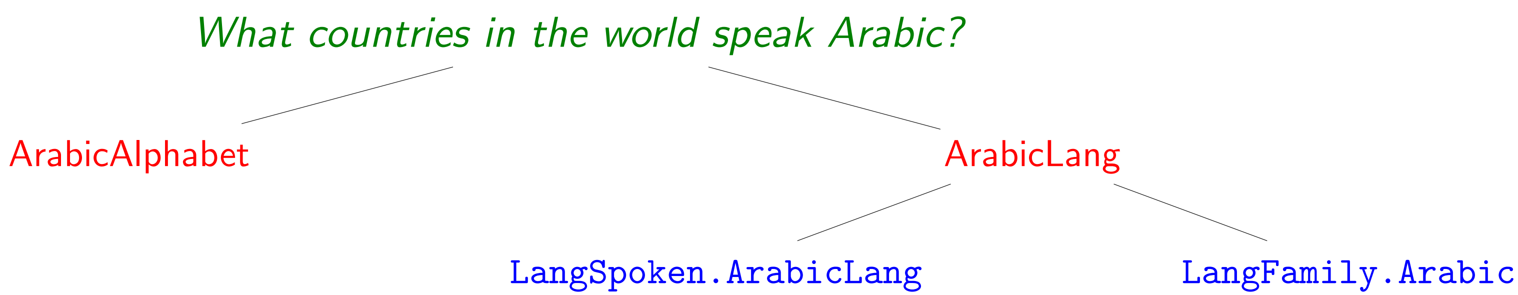
What countries in the world speak Arabic?

ArabicAlphabet

ArabicLang

Bridging to nowhere

Search logical forms based on "prior":



[Berant & Liang, 2014]

Bridging to nowhere

Search logical forms based on "prior":

What countries in the world speak Arabic?

ArabicAlphabet

ArabicLang

`LangSpoken.ArabicLang`

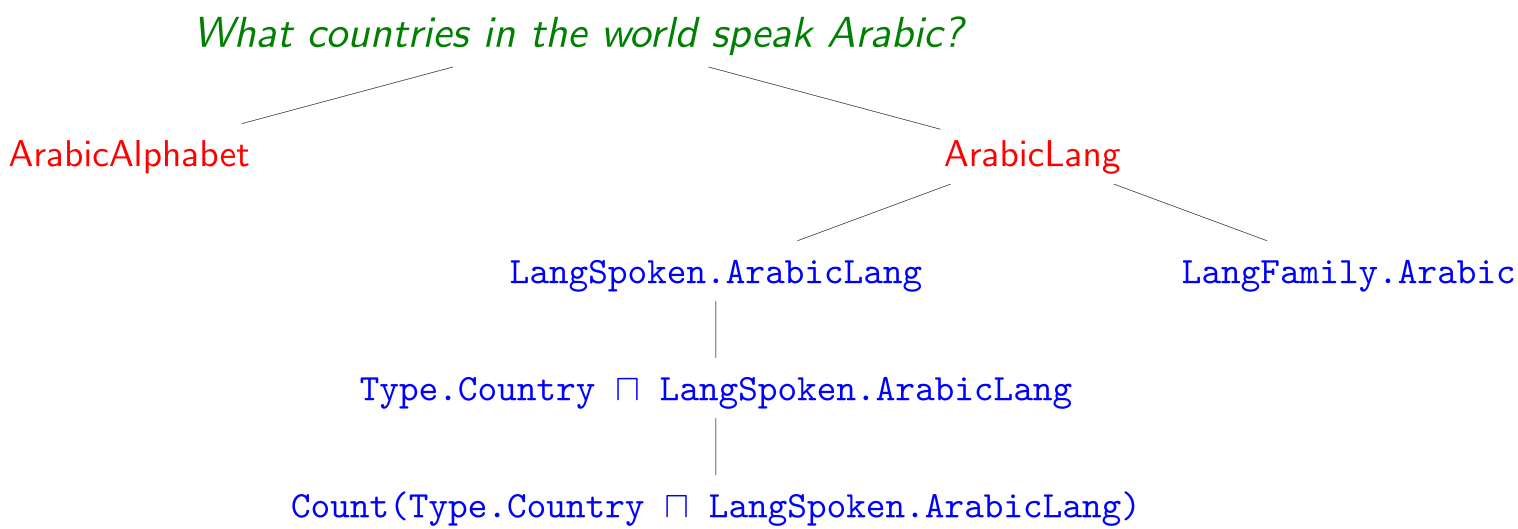
`LangFamily.Arabic`

`Type.Country` ⊓ `LangSpoken.ArabicLang`

`Count(Type.Country` ⊓ `LangSpoken.ArabicLang)`

Bridging to nowhere

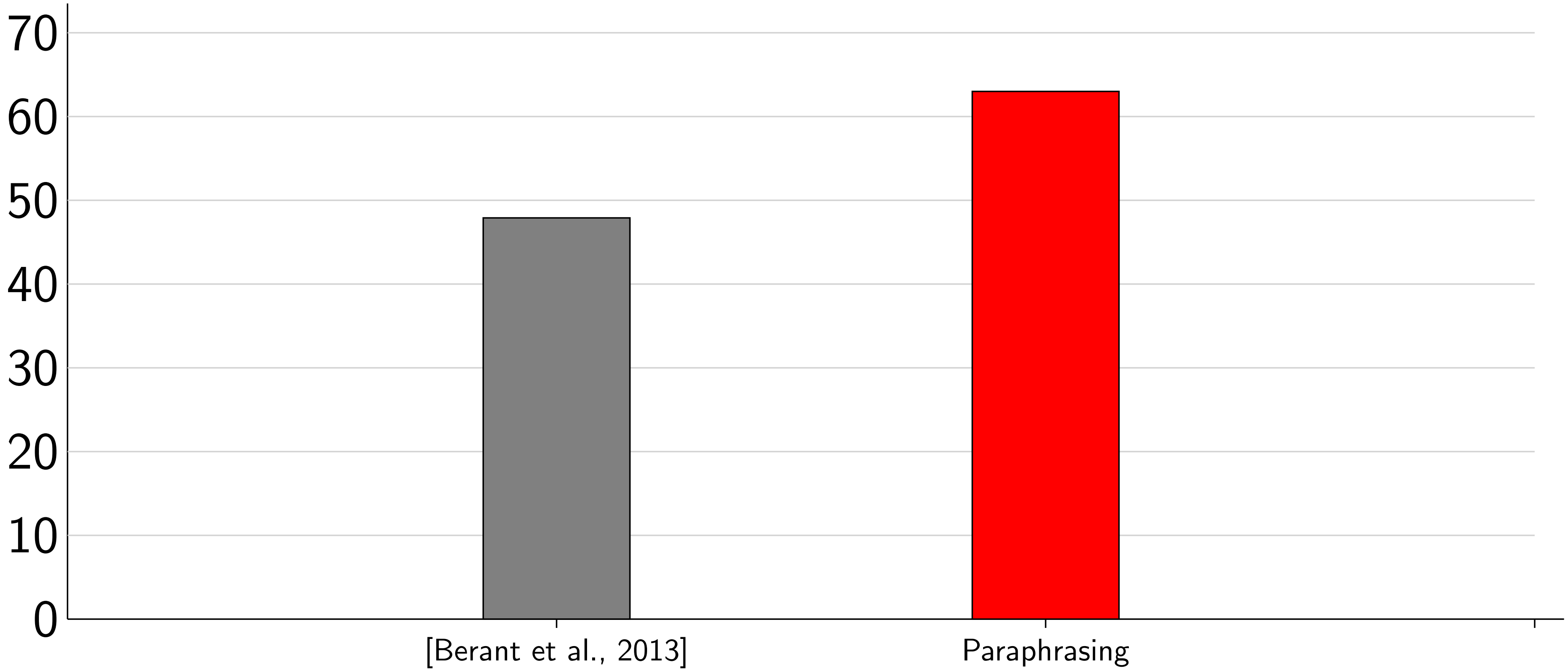
Search logical forms based on "prior":



Start building from parts with more certainty

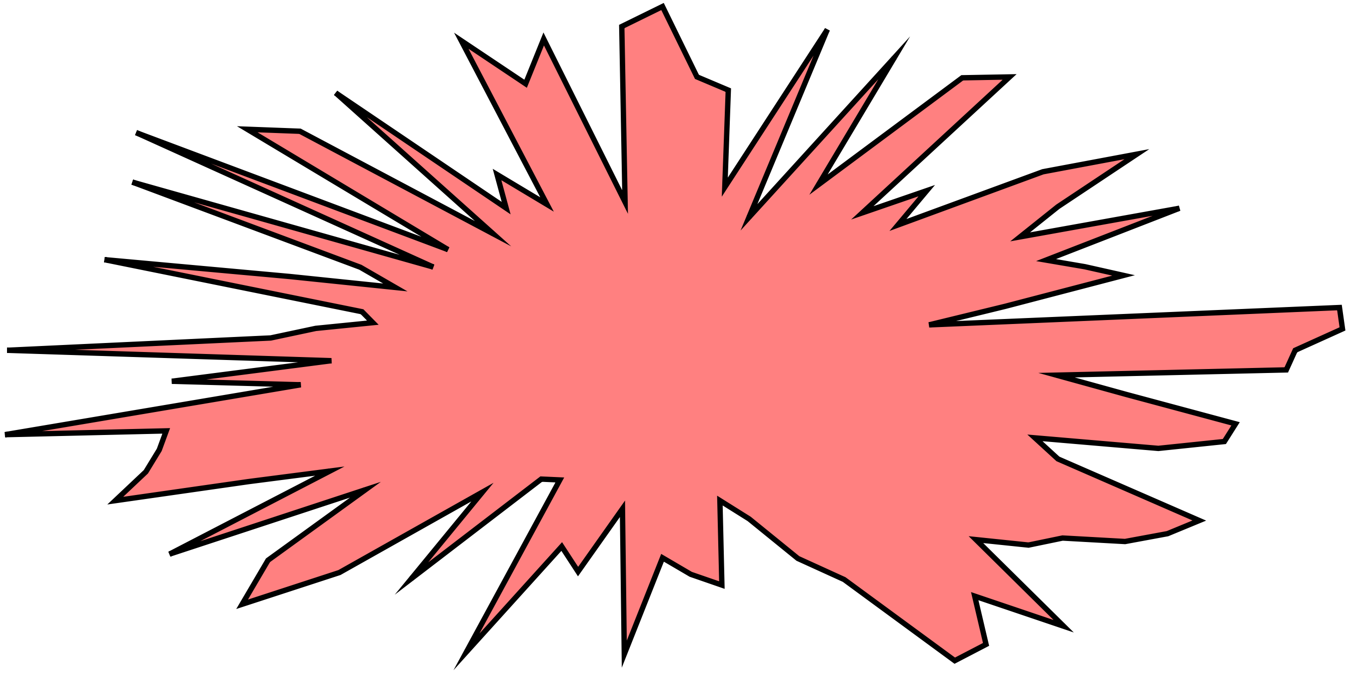
Oracle on WebQuestions

For what fraction of utterances was a candidate logical form correct?



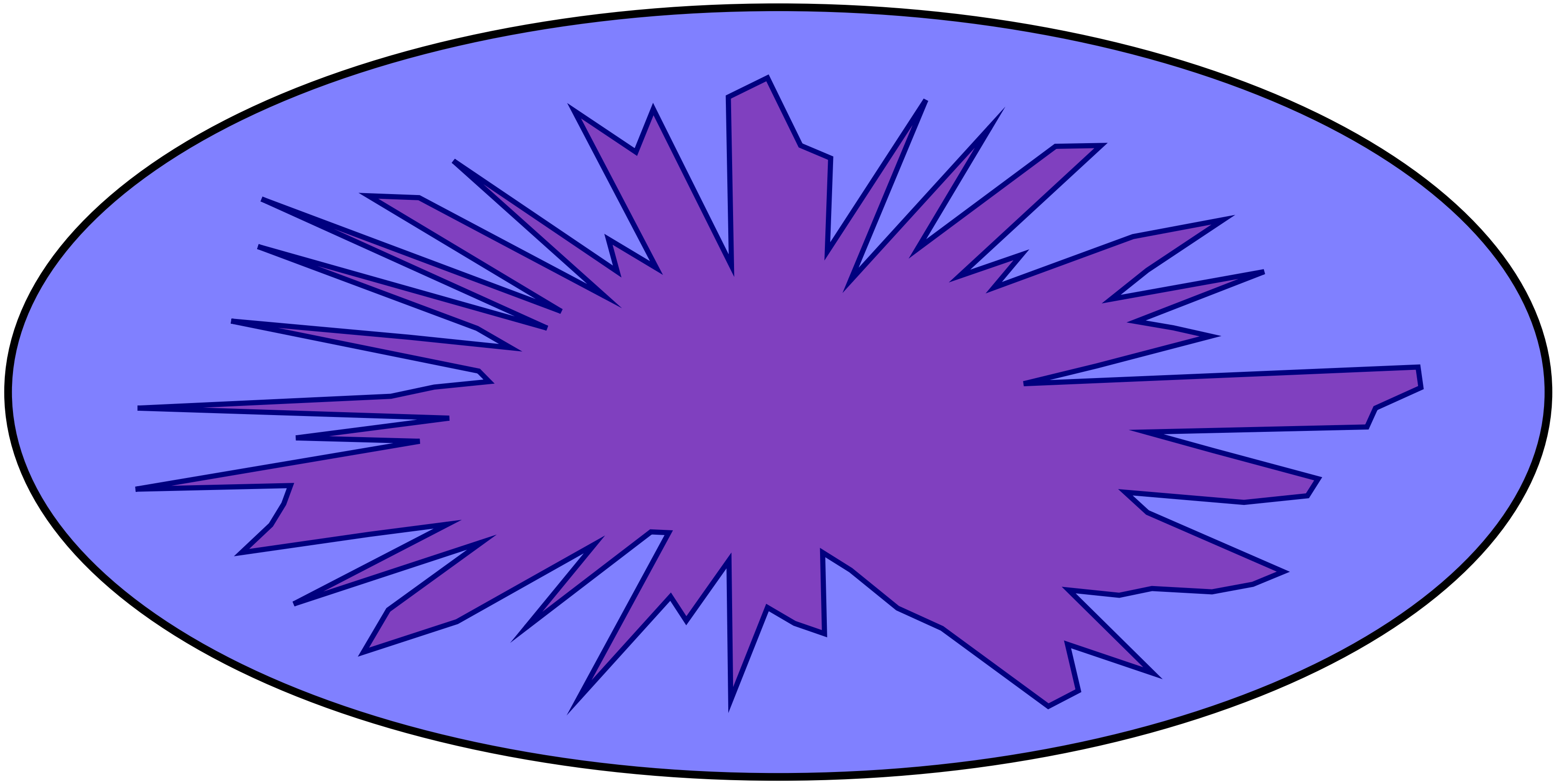
Overapproximation via simple grammars

- Modeling correct derivations requires complex rules



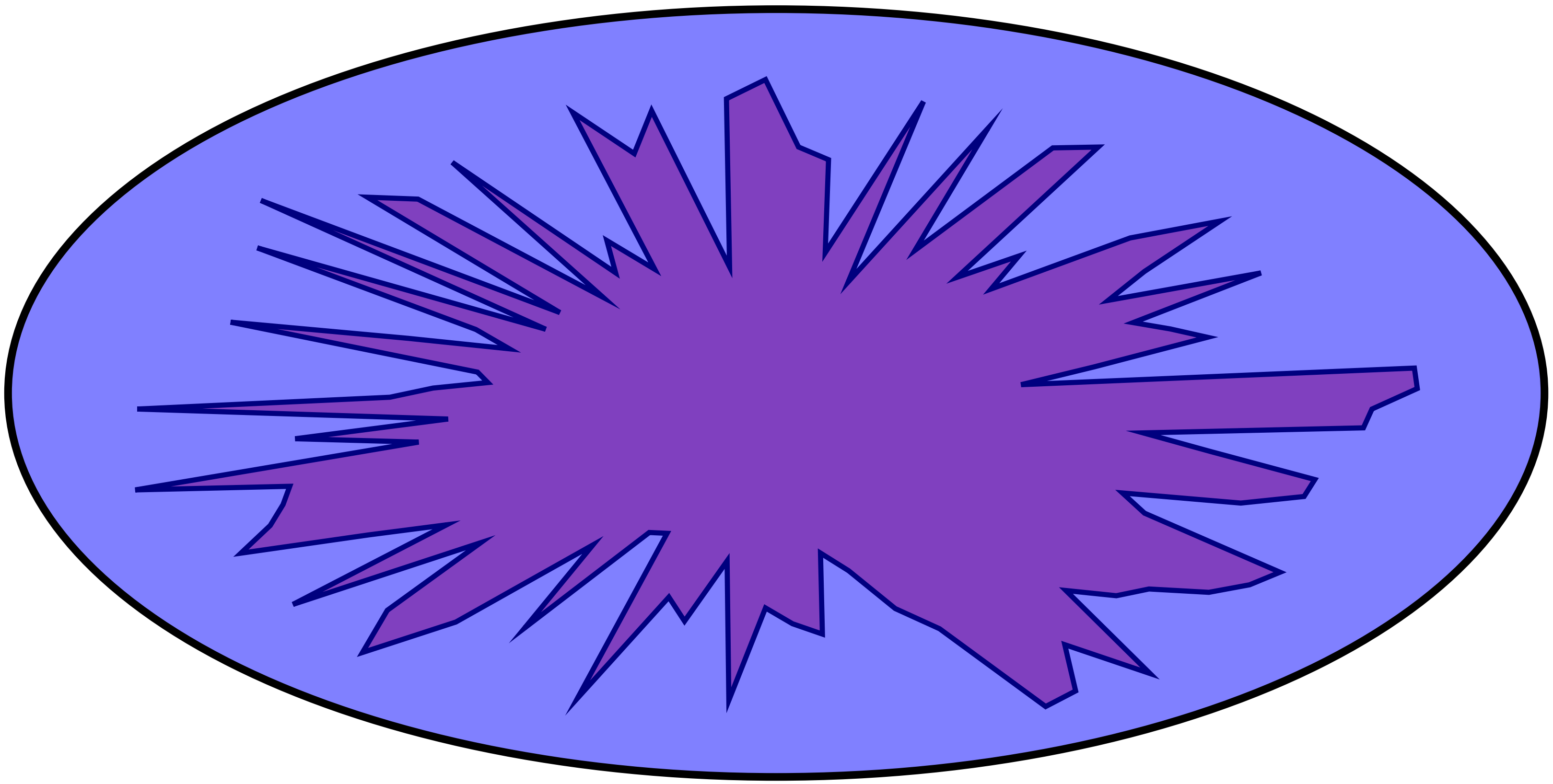
Overapproximation via simple grammars

- Modeling correct derivations requires complex rules
- Simple rules generate overapproximation of good derivations



Overapproximation via simple grammars

- Modeling correct derivations requires complex rules
- Simple rules generate overapproximation of good derivations



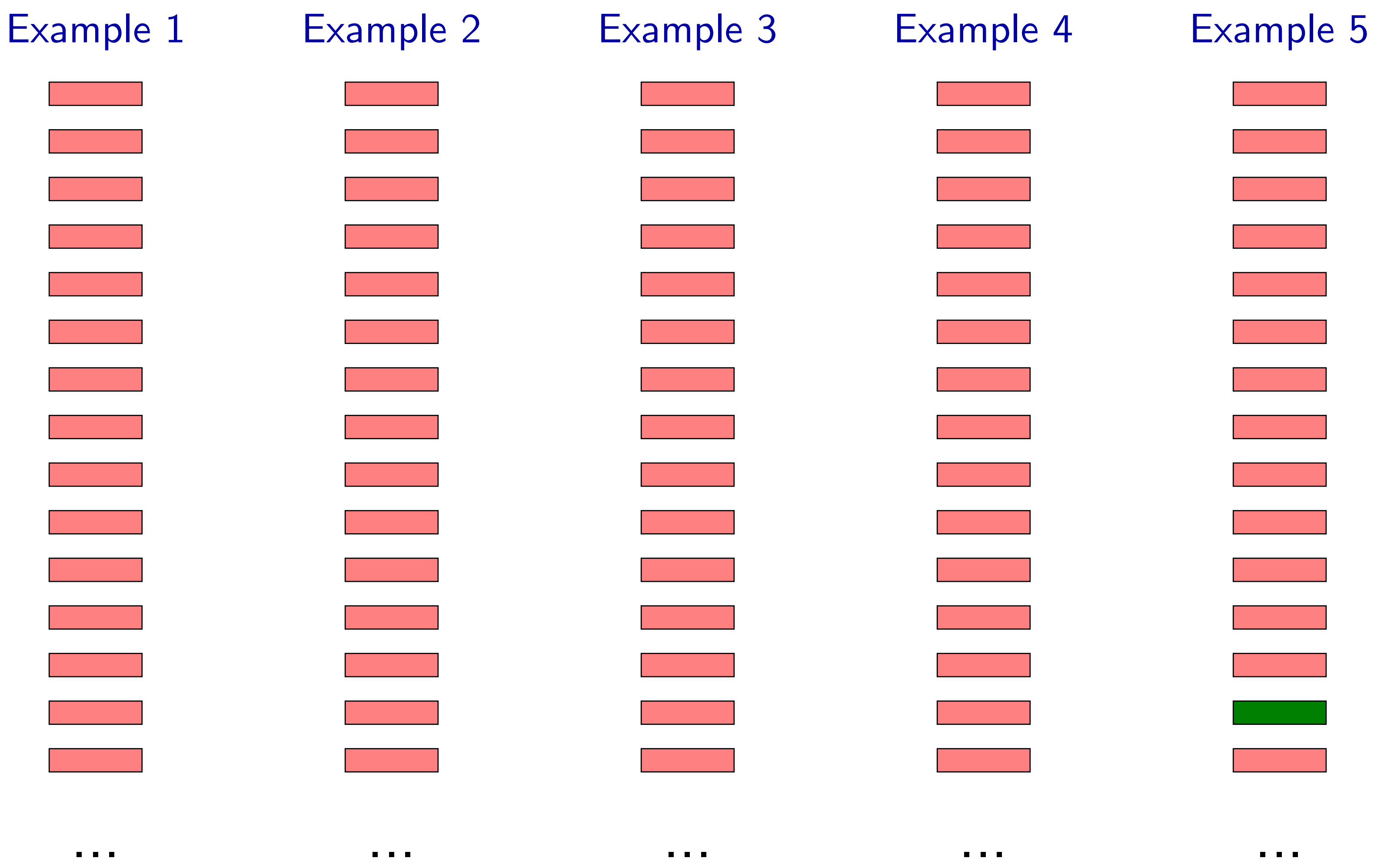
- Hard grammar rules ⇒ soft/overlapping features

Outline

- Semantic parsing in 5 minutes
- A closer look at the elements
  - Knowledge base incompleteness
  - Lexical coverage
  - Search over logical forms
  - **Learning via bootstrapping**
  - Leveraging denotations ("grounding")
  - Datasets
- Final remarks

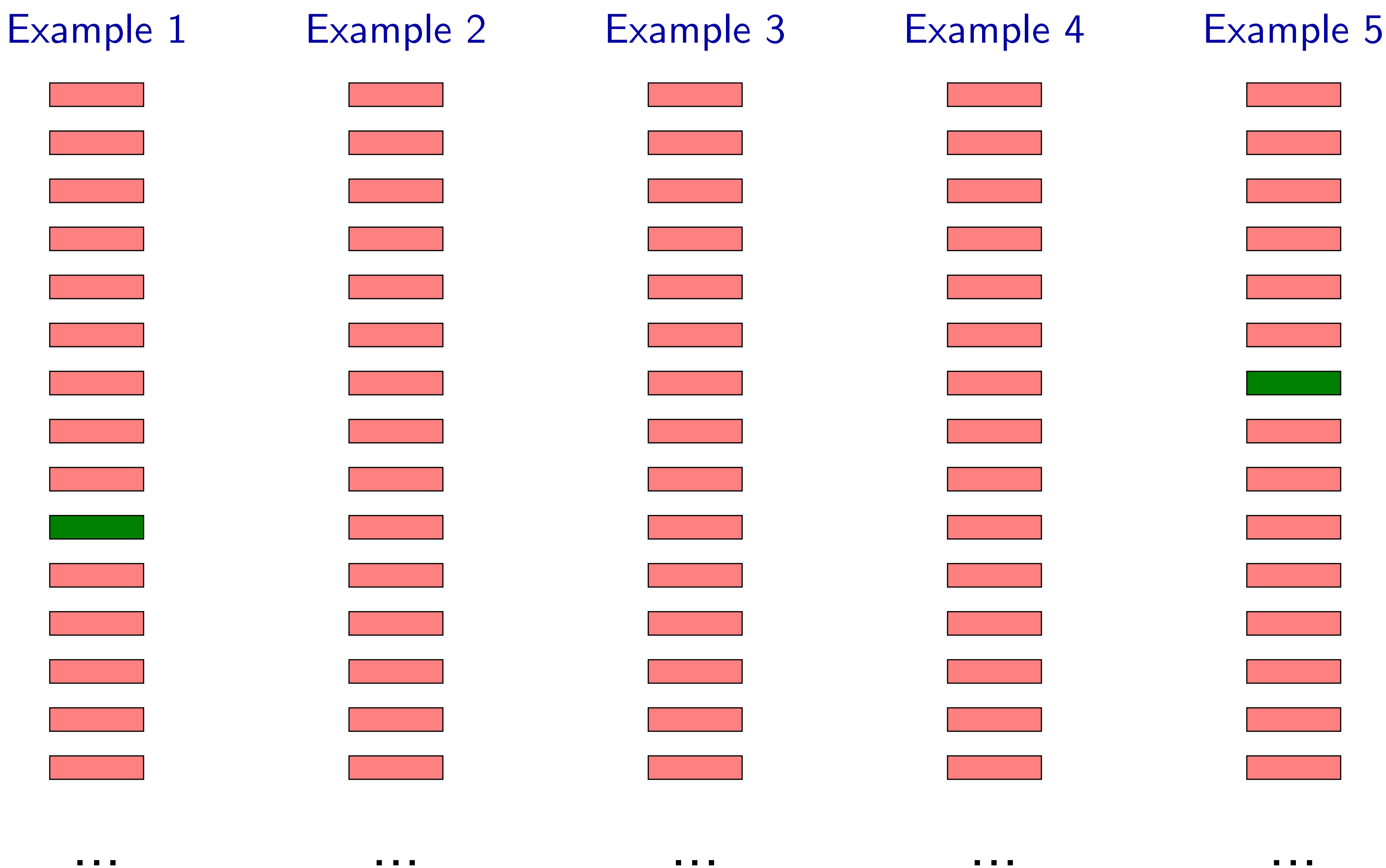
Bootstrapping from easy examples

Iteration 1



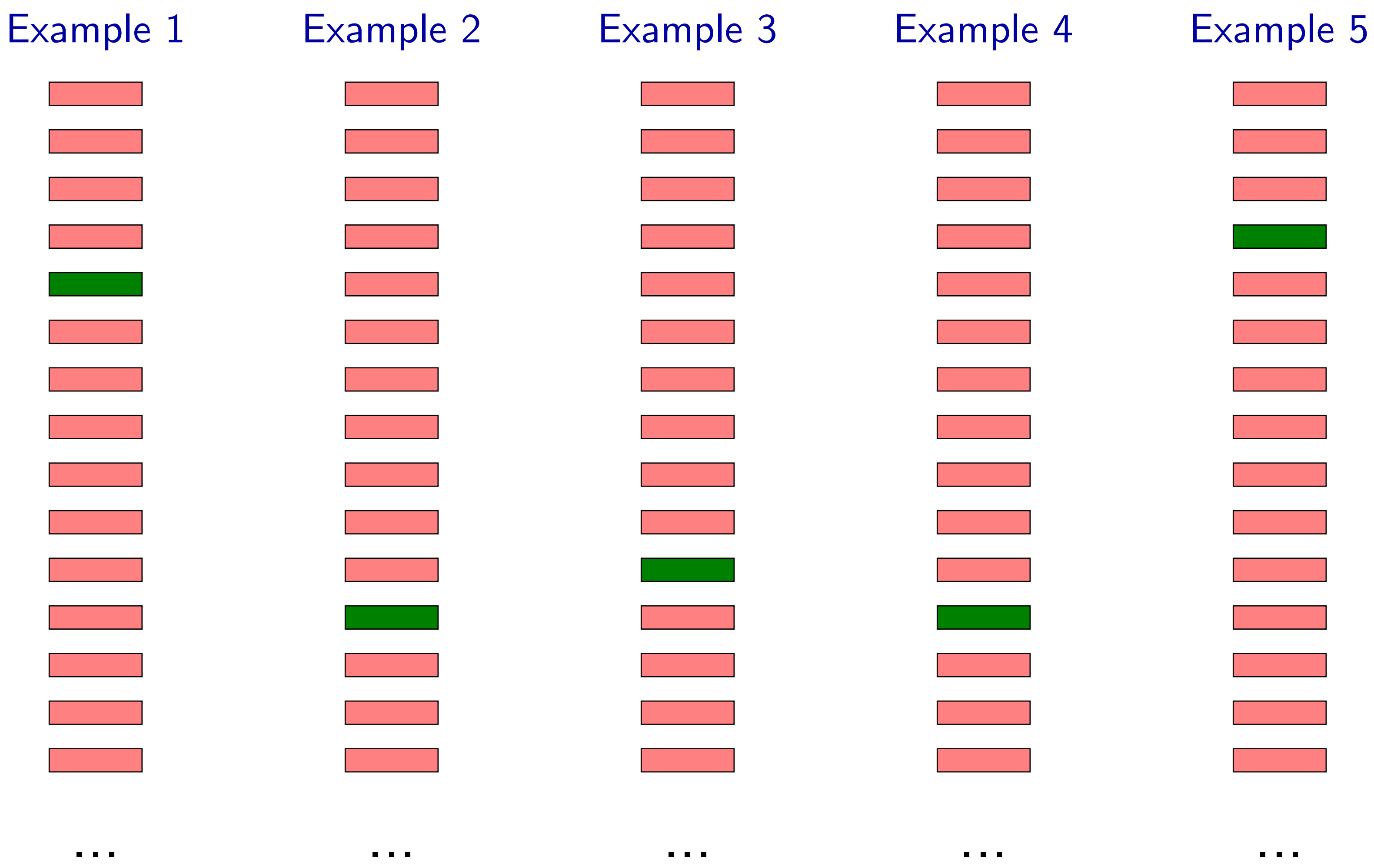
Bootstrapping from easy examples

Iteration 2



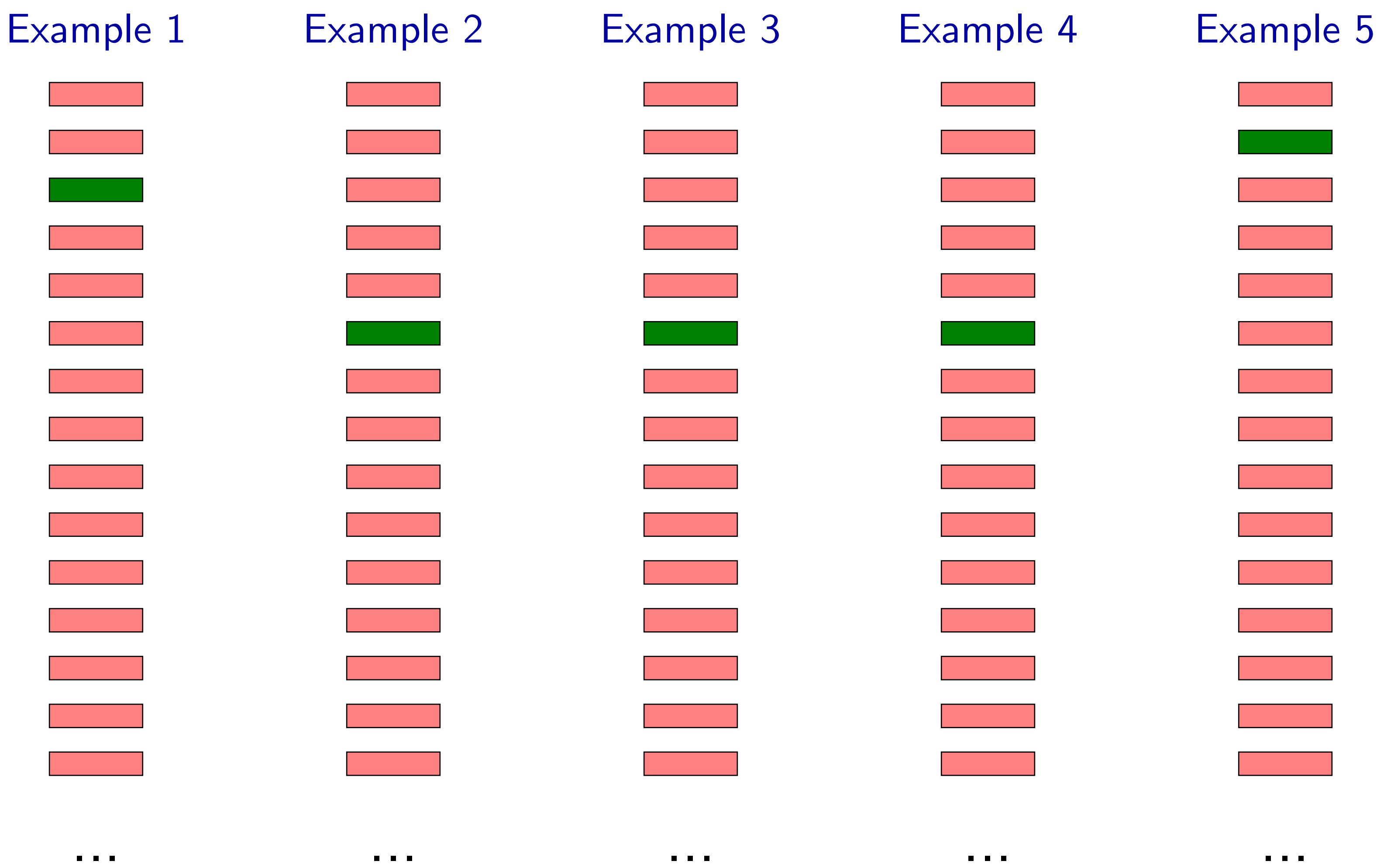
Bootstrapping from easy examples

Iteration 3



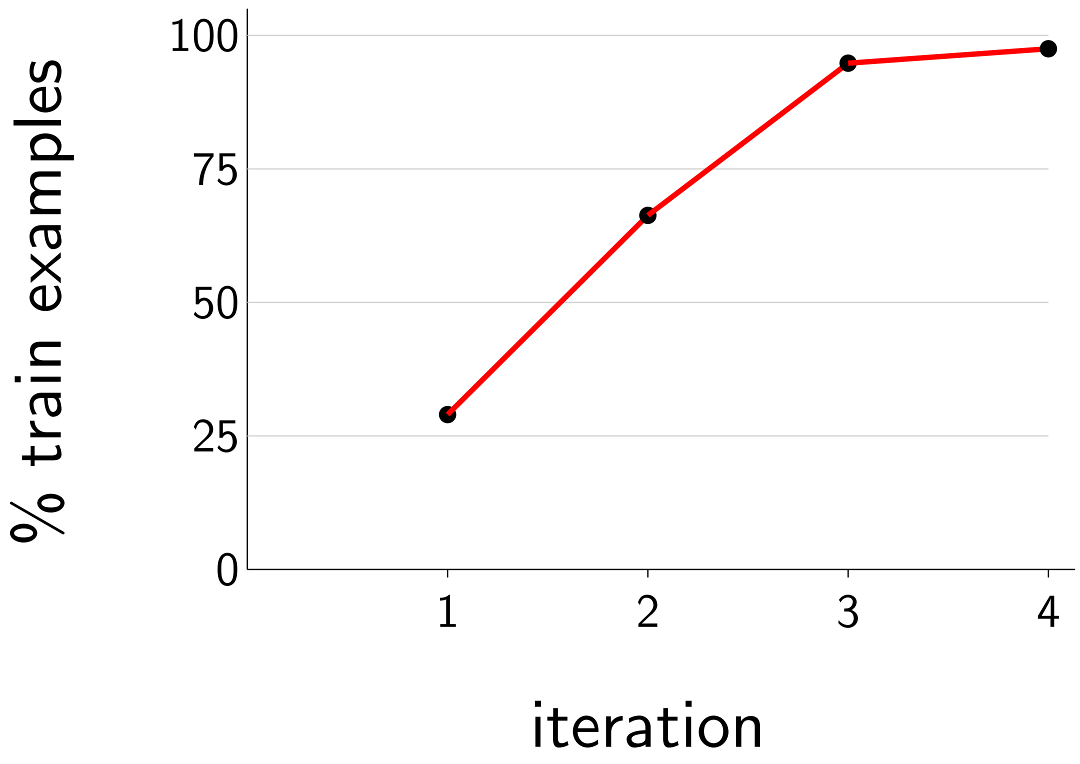
Bootstrapping from easy examples

Iteration 4



Bootstrapping from easy examples

On GeoQuery [Liang et al., 2011]:

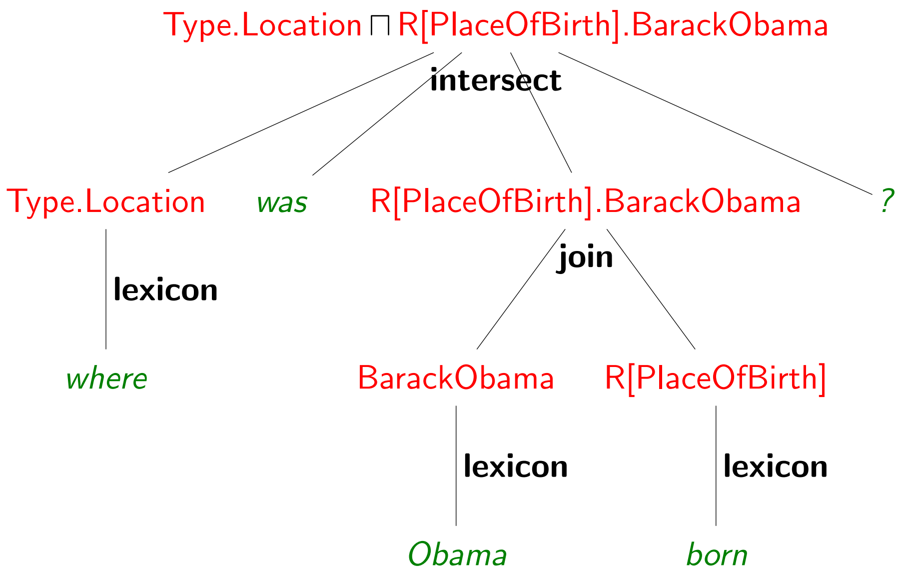


Outline

- Semantic parsing in 5 minutes
- A closer look at the elements
  - Knowledge base incompleteness
  - Lexical coverage
  - Search over logical forms
  - Learning via bootstrapping
  - **Leveraging denotations ("grounding")**
  - Datasets
- Final remarks

x: utterance

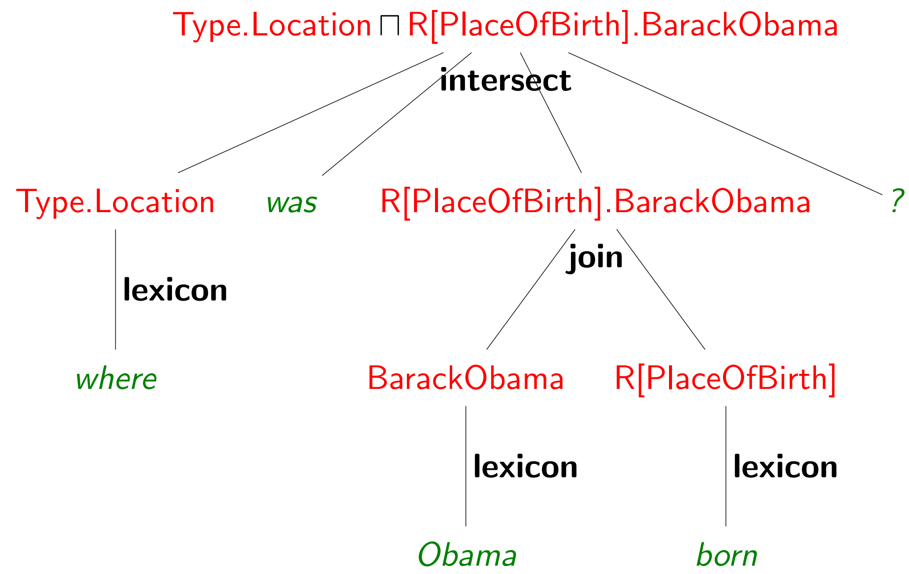
d: derivation



Feature vector $\phi(x, d) \in \mathbb{R}^f$:

x: utterance

d: derivation



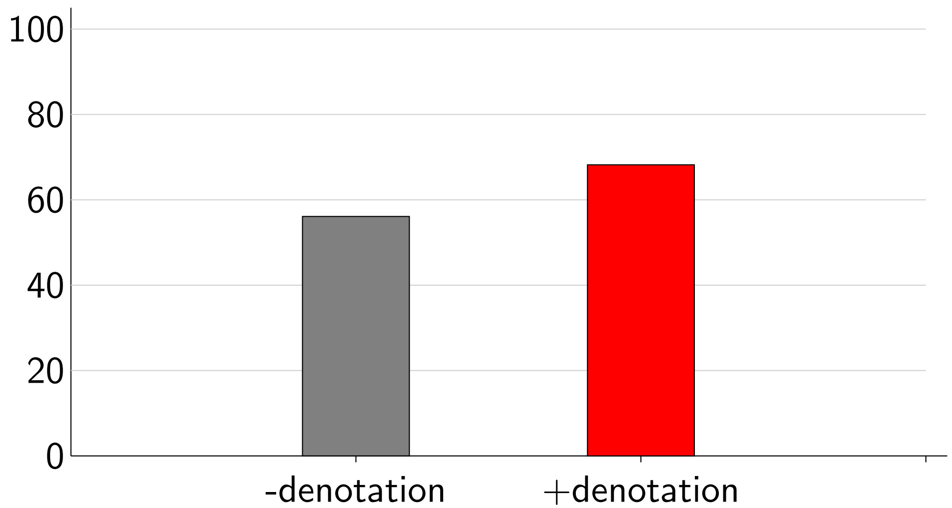
Feature vector $\phi(x, d) \in \mathbb{R}^f$:

apply join	1
apply intersect	1
apply lexicon	3
skipped VBD-AUX	1
skipped NN	0
born maps to PlaceOfBirth	1
born maps to PlacesLived.Location	0
alignmentScore	1.52
denotation-size=1	1
...	...

Denotation features for entity extraction

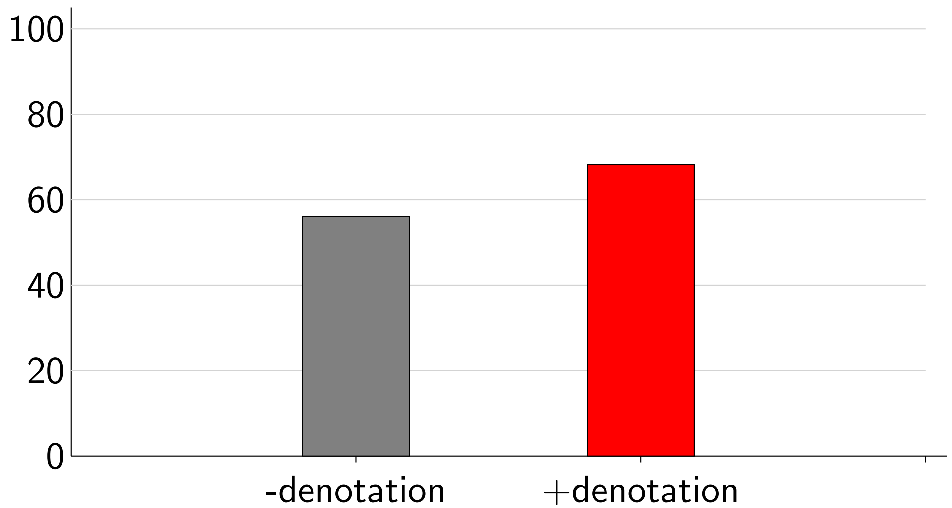
/html[1]/body[1]/table[2]/tr/td[1]		/html[1]/body[1]/div[2]/a
hiking trails near Baltimore		*hiking trails near Baltimore*
Avalon Super Loop		Home
Patapsco Valley State Park		About Baltimore Tour
Gunpowder Falls State Park	$>$	Pricing
Rachel Carson Conservation Park		Contact
Union Mills Hike		Online Support
...		...

Impact of denotation features

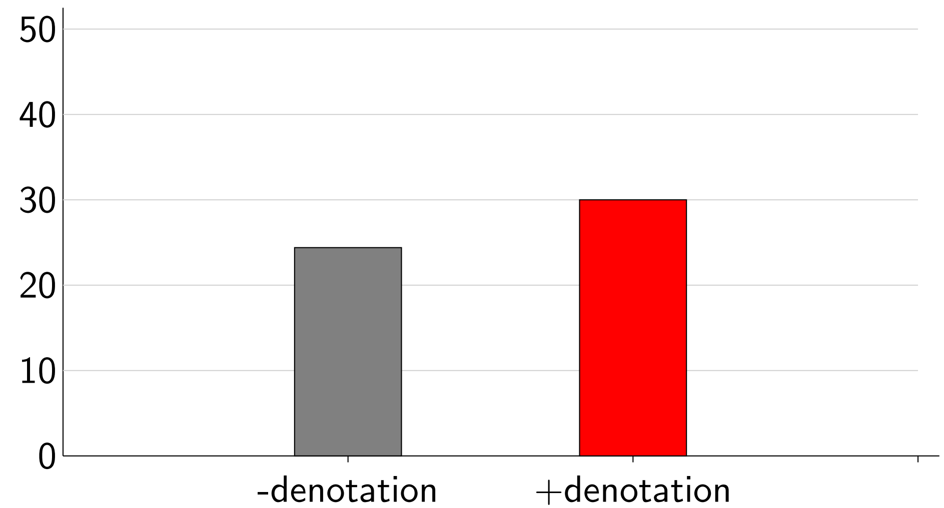


Free917

Impact of denotation features

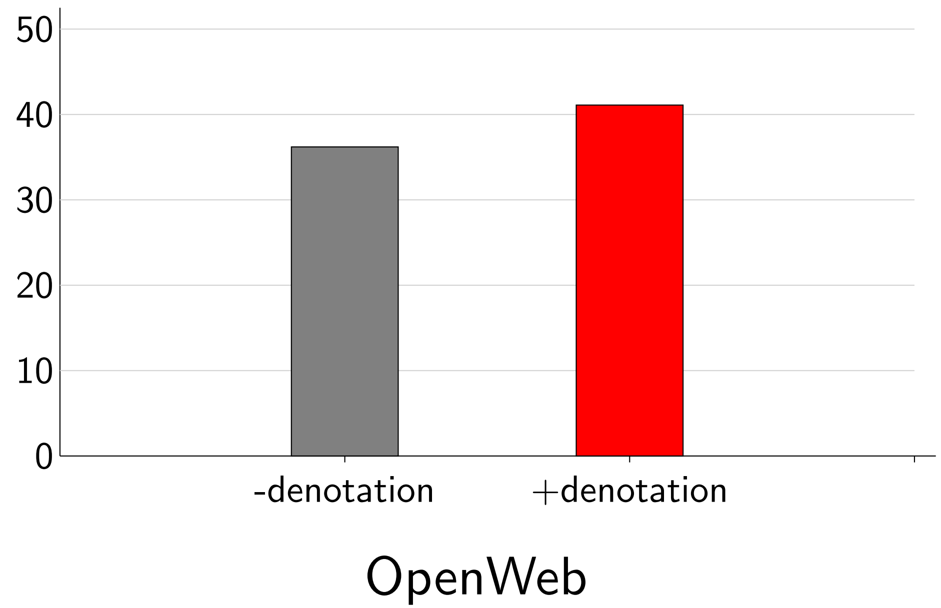


Free917

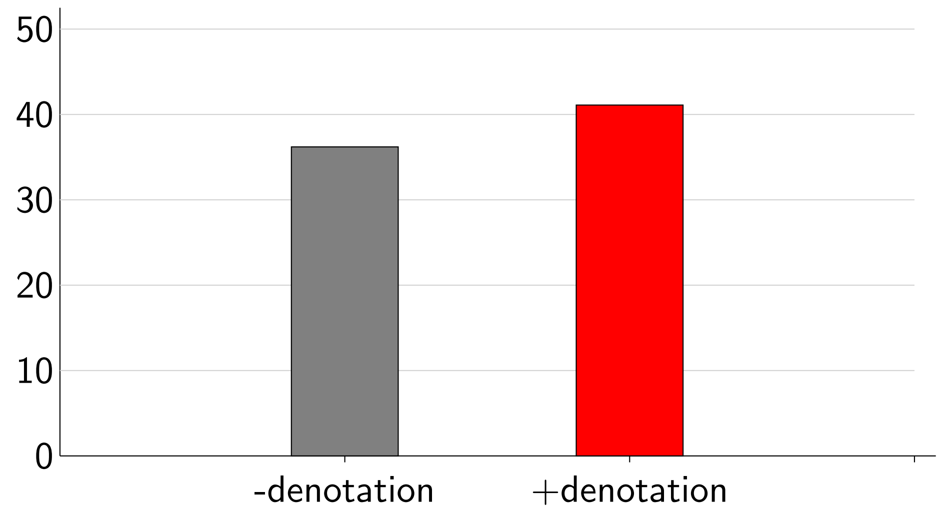


WebQuestions

Impact of denotation features



Impact of denotation features



OpenWeb

Working with denotations actually provides more information than just logical forms

Outline

- Semantic parsing in 5 minutes
- A closer look at the elements
  - Knowledge base incompleteness
  - Lexical coverage
  - Search over logical forms
  - Learning via bootstrapping
  - Leveraging denotations ("grounding")
  - **Datasets**
- Final remarks

Dataset collection

Obtain naturally occurring questions (inputs)

Dataset collection

Obtain naturally occurring questions (inputs)

Strategy: breadth-first search over Google Suggest graph

Dataset collection

Obtain naturally occurring questions (inputs)

Strategy: breadth-first search over Google Suggest graph

Where was Barack Obama born?

Dataset collection

Obtain naturally occurring questions (inputs)

Strategy: breadth-first search over Google Suggest graph

Where was Barack Obama born?

Where was _ born? —**Google Suggest**→ Barack Obama / Lady Gaga / Steve Jobs

Dataset collection

Obtain naturally occurring questions (inputs)

Strategy: breadth-first search over Google Suggest graph

Where was Barack Obama born?

Where was _ born? **Google Suggest** → Barack Obama, Lady Gaga, Steve Jobs

Where was Steve Jobs born?

Dataset collection

Obtain naturally occurring questions (inputs)

Strategy: breadth-first search over Google Suggest graph

Where was Barack Obama born?

Where was _ born? —**Google Suggest**→ Barack Obama / Lady Gaga / Steve Jobs

Where was Steve Jobs born?

Where was Steve Jobs _? —**Google Suggest**→ born / raised / on the Forbes list

Dataset collection

Obtain naturally occurring questions (inputs)

Strategy: breadth-first search over Google Suggest graph

Where was Barack Obama born?

Where was _ born? **Google Suggest** → Barack Obama, Lady Gaga, Steve Jobs

Where was Steve Jobs born?

Where was Steve Jobs _? **Google Suggest** → born, raised, on the Forbes list

Where was Steve Jobs raised?

Dataset collection

Obtain naturally occurring questions (inputs)

Strategy: breadth-first search over Google Suggest graph

Where was Barack Obama born?

Where was _ born? **Google Suggest** → Barack Obama, Lady Gaga, Steve Jobs

Where was Steve Jobs born?

Where was Steve Jobs _? **Google Suggest** → born, raised, on the Forbes list

Where was Steve Jobs raised?

...

AMT annotation ⇒ 6.6K question/answer pairs

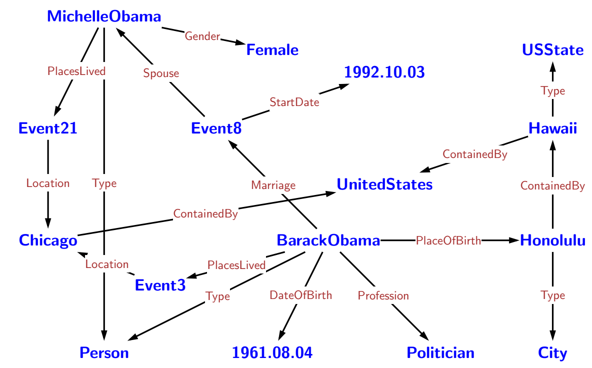
Question answering on Freebase

WebQuestions dataset (6K questions) [Berant et al., 2013]

what did obama study in school

where to fly into bali

what was tupac name in juice



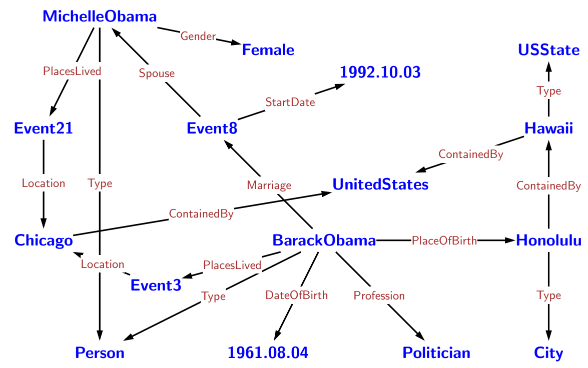
Question answering on Freebase

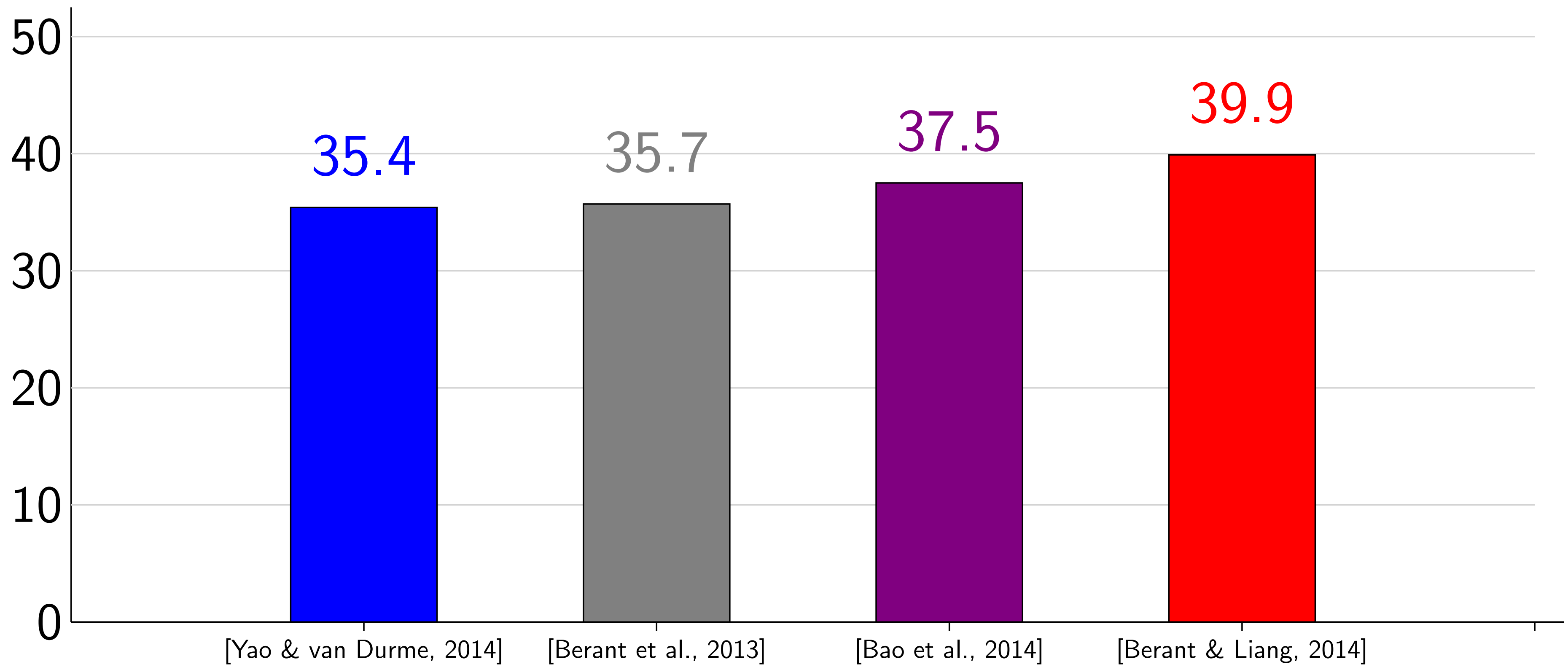
WebQuestions dataset (6K questions) [Berant et al., 2013]

what did obama study in school

where to fly into bali

what was tupac name in juice





OpenWeb dataset

airlines of italy

natural causes of global warming

lsu football coaches

bf3 submachine guns

badminton tournaments

foods high in dha

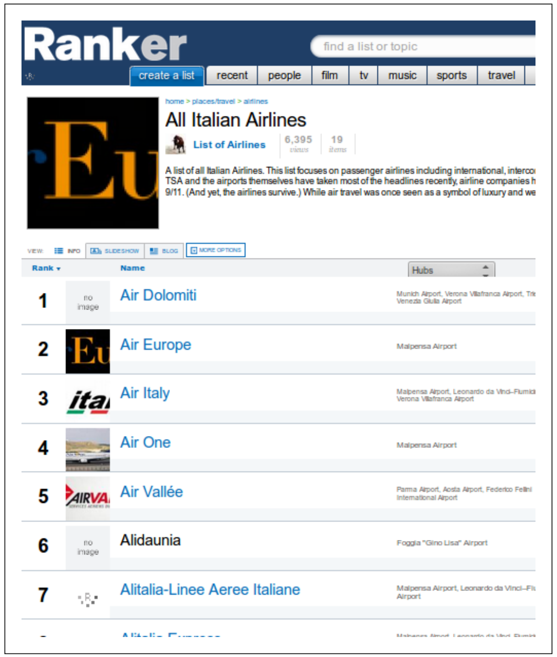
technical colleges in south carolina

songs on glee season 5

singers who use auto tune

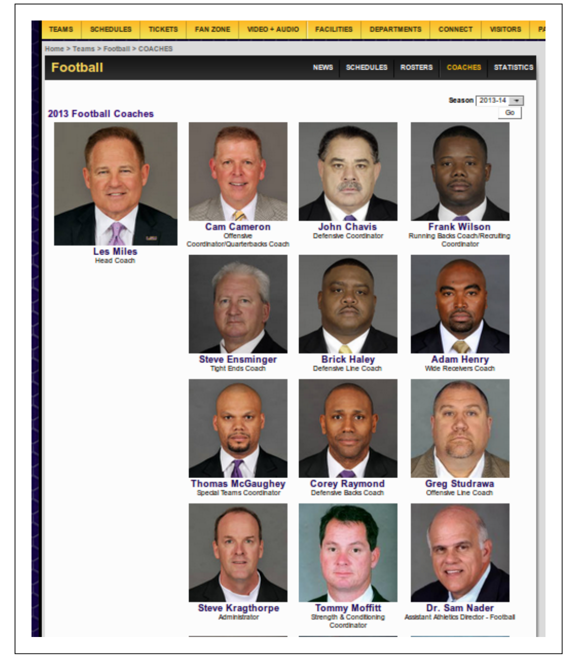
san francisco radio stations

OpenWeb dataset



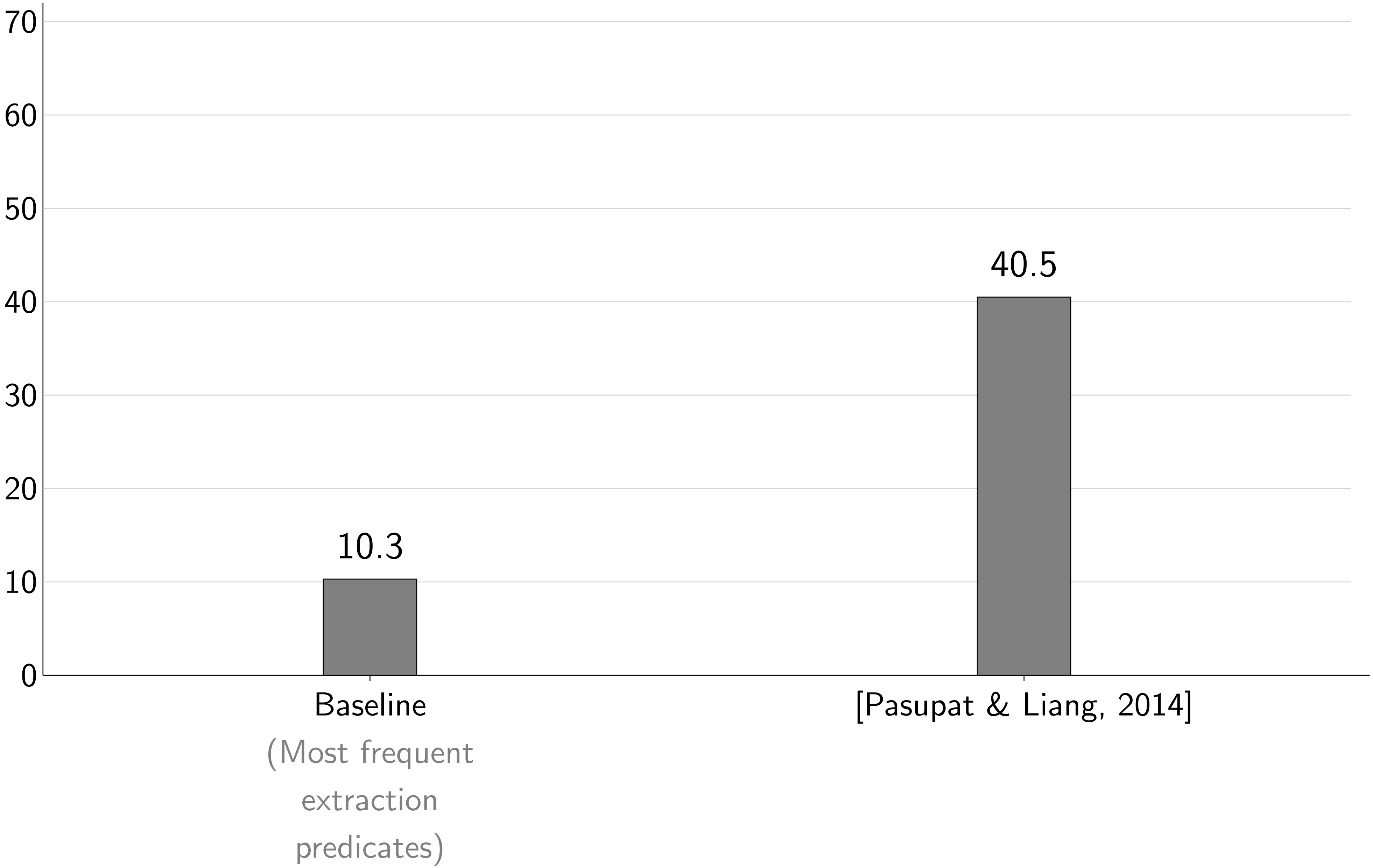
airlines of italy

natural causes of global warming



lsu football coaches

Results on OpenWeb



A new dataset?

compositional AND **open-domain**

A new dataset?

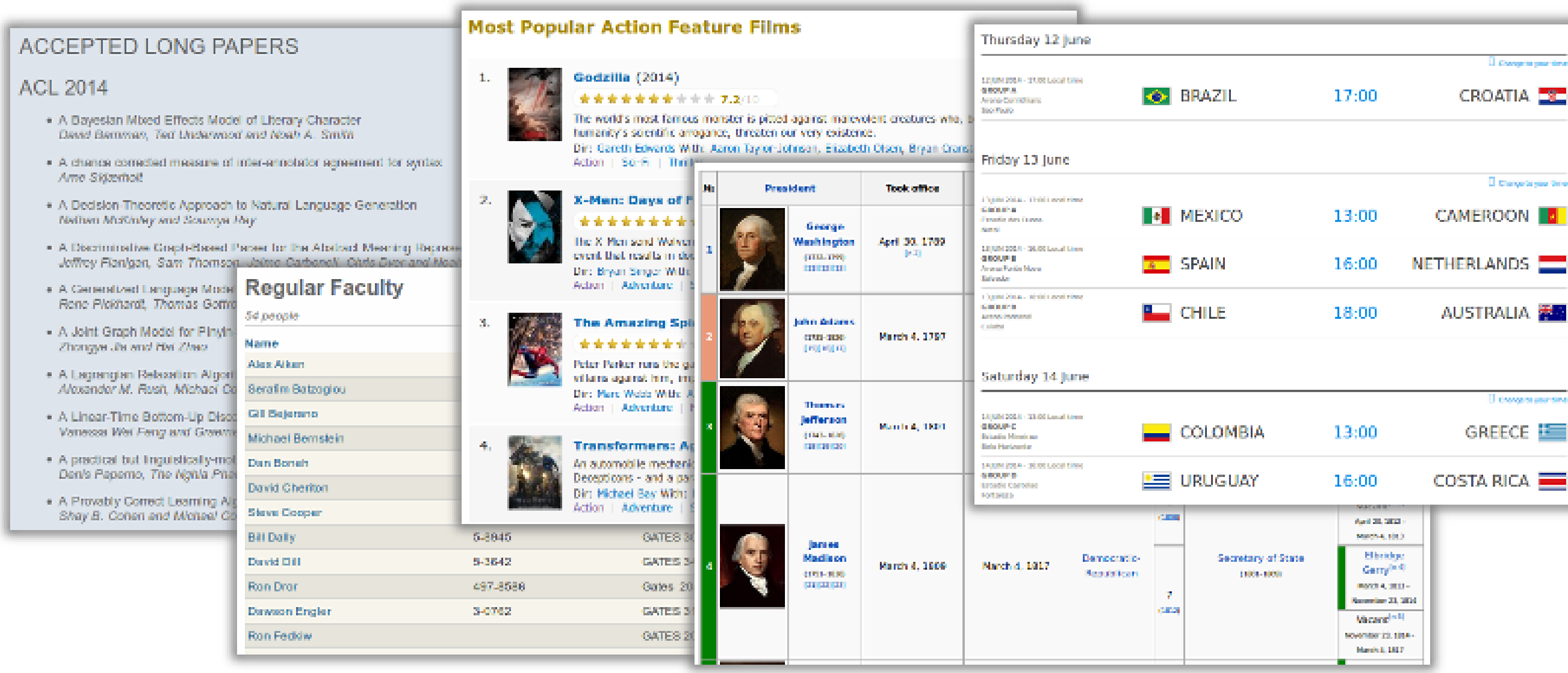
compositional AND **open-domain**

How old are presidents when they take office on average?

A new dataset?

compositional AND **open-domain**

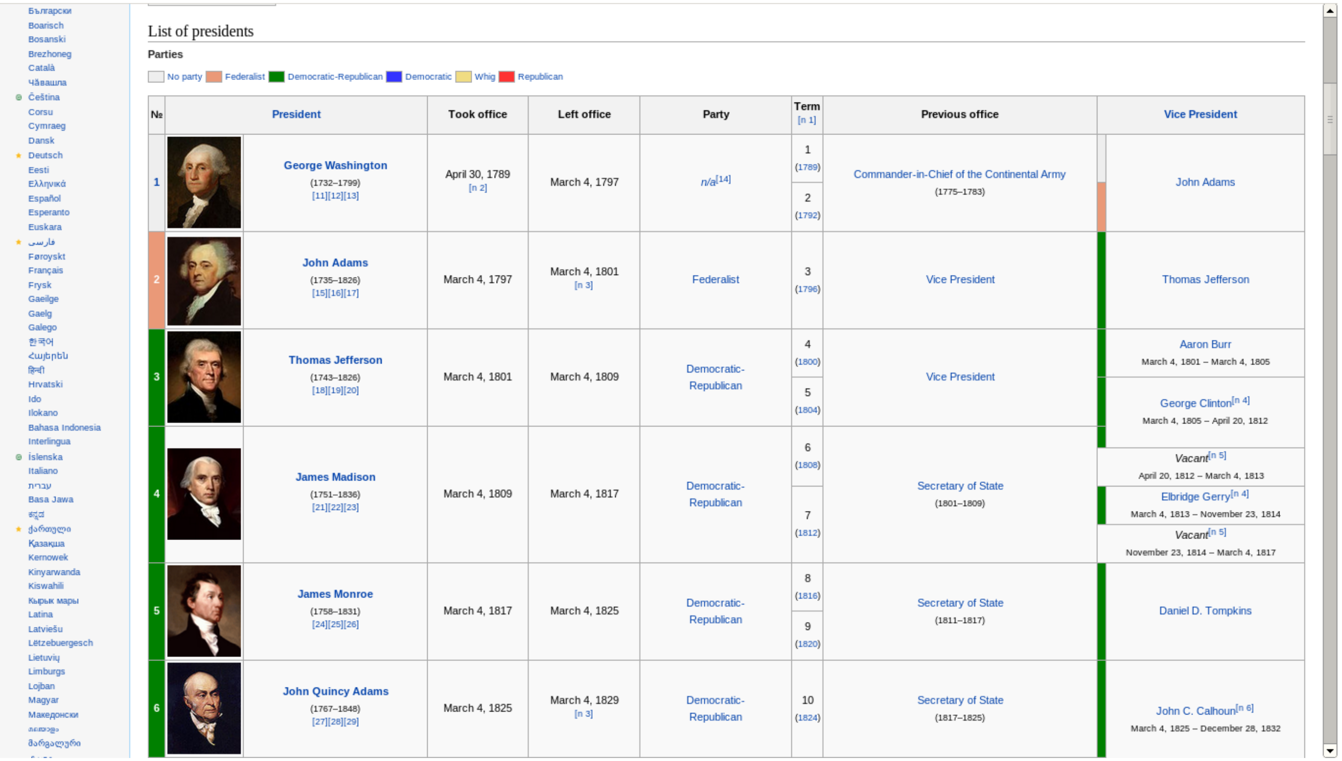
How old are presidents when they take office on average?



A new dataset?

compositional AND **open-domain**

How old are presidents when they take office on average?



List of presidents

Parties

No party | Federalist | Democratic-Republican | Democratic | Whig | Republican

№	President	Took office	Left office	Party	Term [n 1]	Previous office	Vice President
1	George Washington (1732–1799) [11][12][13]	April 30, 1789 [n 2]	March 4, 1797	n/a[14]	1 (1789) 2 (1792)	Commander-in-Chief of the Continental Army (1775–1783)	John Adams
2	John Adams (1735–1826) [15][16][17]	March 4, 1797	March 4, 1801 [n 3]	Federalist	3 (1796)	Vice President	Thomas Jefferson
3	Thomas Jefferson (1743–1826) [18][19][20]	March 4, 1801	March 4, 1809	Democratic-Republican	4 (1800) 5 (1804)	Vice President	Aaron Burr March 4, 1801 – March 4, 1805 George Clinton[n 4] March 4, 1805 – April 20, 1812
4	James Madison (1751–1836) [21][22][23]	March 4, 1809	March 4, 1817	Democratic-Republican	6 (1808) 7 (1812)	Secretary of State (1801–1809)	Vacant[n 5] April 20, 1812 – March 4, 1813 Elbridge Gerry[n 4] March 4, 1813 – November 23, 1814 Vacant[n 5] November 23, 1814 – March 4, 1817
5	James Monroe (1758–1831) [24][25][26]	March 4, 1817	March 4, 1825	Democratic-Republican	8 (1816) 9 (1820)	Secretary of State (1811–1817)	Daniel D. Tompkins
6	John Quincy Adams (1767–1848) [27][28][29]	March 4, 1825	March 4, 1829 [n 3]	Democratic-Republican	10 (1824)	Secretary of State (1817–1825)	John C. Calhoun[n 6] March 4, 1825 – December 28, 1832

Other tasks

Playing computer games [Branavan et al., 2010, 2011]

Following navigational instructions [Tellex et. al 2011; Chen et. al 2012; Artzi et. al 2013]

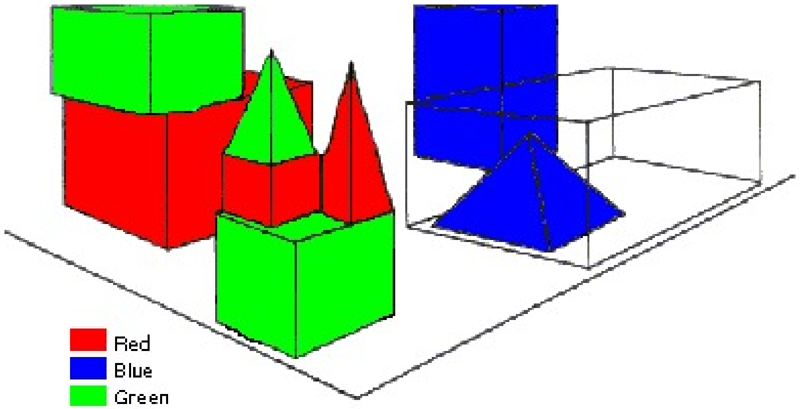
Understanding visual world [Matuszek et. al; 2012, Krishnamurthy & Kollar, 2013]

Solving algebra word problems [Kushman et. al, 2013]

Outline

- Semantic parsing in 5 minutes
- A closer look at the elements
  - Knowledge base incompleteness
  - Lexical coverage
  - Search over logical forms
  - Learning via bootstrapping
  - Leveraging denotations ("grounding")
  - Datasets
- **Final remarks**

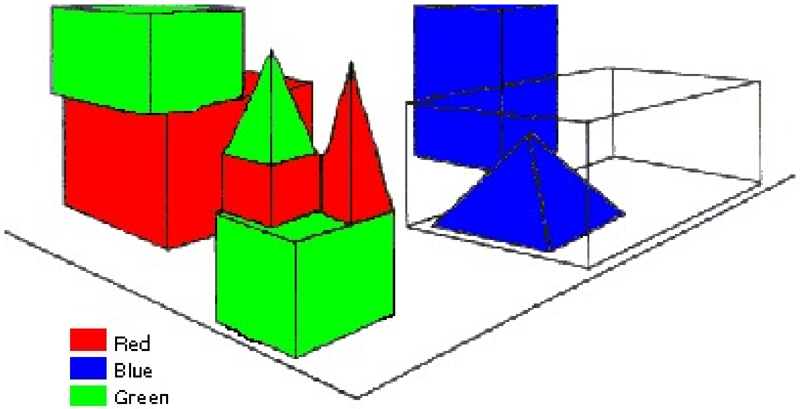
SHRDLU [1971]



SHRDLU [1971]

Person: Pick up a big red block.

Computer: OK.



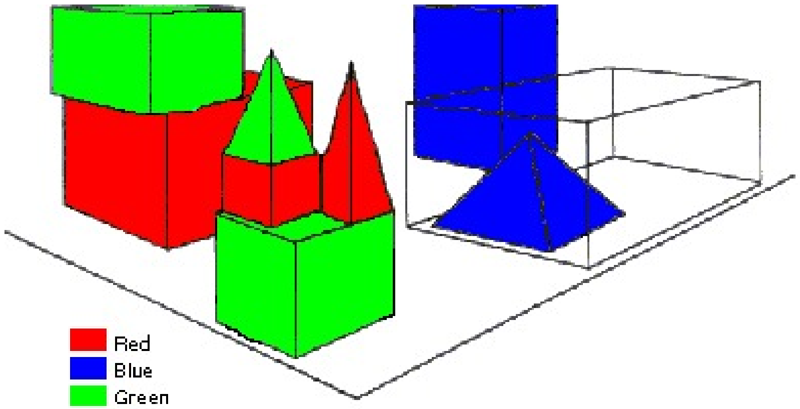
SHRDLU [1971]

Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.



SHRDLU [1971]

Person: Pick up a big red block.

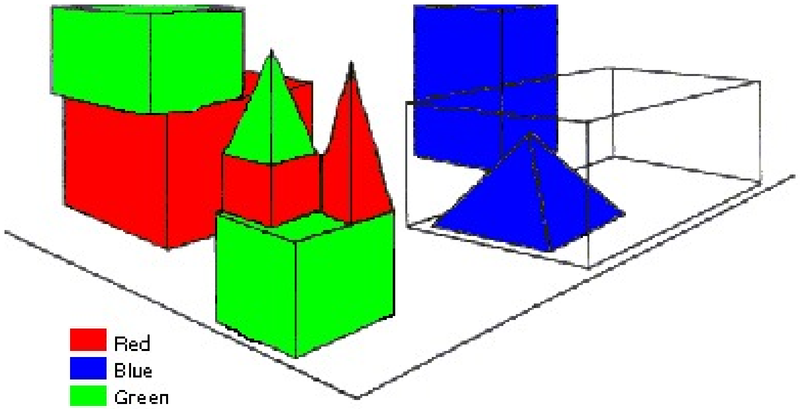
Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.

Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.

Computer: By "it", I assume you mean the block which is taller than the one I am holding.



SHRDLU [1971]

Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.

Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.

Computer: By "it", I assume you mean the block which is taller than the one I am holding.

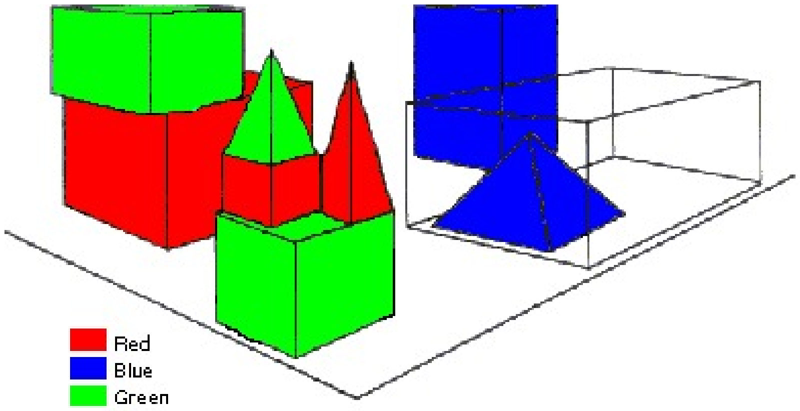
Computer: OK.

Person: What does the box contain?

Computer: The blue pyramid and the blue block.

Person: What is the pyramid supported by?

Computer: The box.



Takeaway 1/3: denotations

[utterance: user input]

↓ semantic parsing

[intermediate semantic representations (text? logical forms? vectors?)]

↓ execute

`[denotation: user output]`

Semantic representations as a means to an end

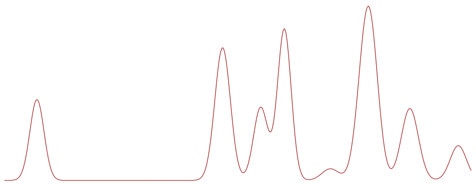
Takeaway 2/3: learning and search

Challenging search / learning problem

Takeaway 2/3: learning and search

Challenging search / learning problem

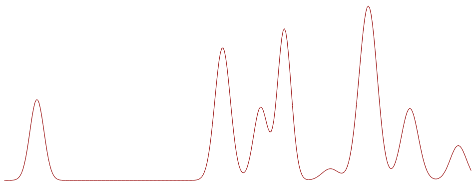
Non-convex optimization



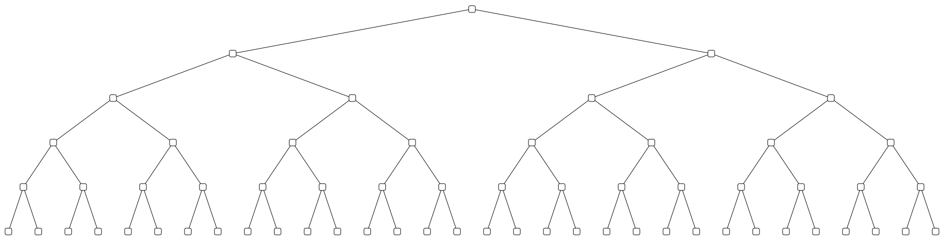
Takeaway 2/3: learning and search

Challenging search / learning problem

- **Non-convex optimization**



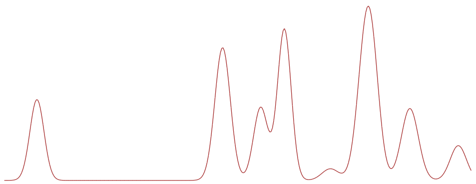
- **Exponential search space**



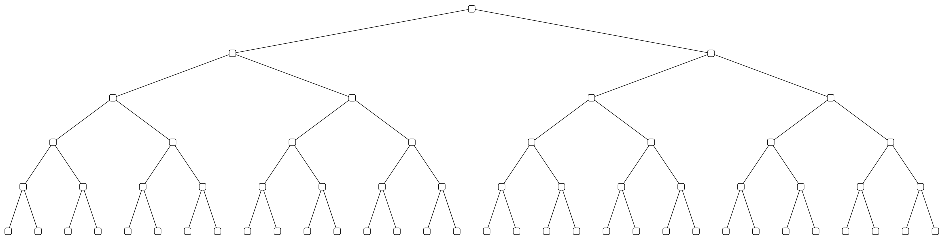
Takeaway 2/3: learning and search

Challenging search / learning problem

- **Non-convex optimization**



- **Exponential search space**



Need to create better abstractions for people to work on the core search/learning issues

Takeaway 3/3: data and users

Semantic parsing provides utility to users

Users provide get back realistic datasets

The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period[1]. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations[2–5], but the mechanisms responsible for those perturbations remain uncertain[1,2]. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–480 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity[6]. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this palaeogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the 'Cambrian explosion' of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally tilted sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup[6]. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana[7,8], Baltica[8], Avalonia[9] and Siberia[10], making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first skeletonized crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution[1,6].

Here we use stratigraphic and lithologic data for 21,521 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence[11,12] that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences[11,12], but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth's surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth's crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreface system, often called the 'wave-base razor'[13], as well as adjacent transitional backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors[7] was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regolith[14,15], and

472
488
542
Myr ago
Ordovician
Cambrian
Early
Late
Middle
Early

Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

[1]Department of Geoscience, University of Wisconsin, Madison, Wisconsin 53706, USA. [2]Geology Department, Pomona College, Claremont, California 91711, USA.

How long do species tend to exist before going extinct?

Semantic parsing is useful

Code and data online

http://www-nlp.stanford.edu/software/sempre/

http://www-nlp.stanford.edu/software/web-entity-extractor-ACL2014/

Code and data online

http://www-nlp.stanford.edu/software/sempre/
http://www-nlp.stanford.edu/software/web-entity-extractor-ACL2014/

Collaborators

Jonathan Berant (post-doc)
Andrew Chou (masters)
Roy Frostig (Ph.D.)
Panupong Pasupat (Ph.D.)

Code and data online

http://www-nlp.stanford.edu/software/sempre/
http://www-nlp.stanford.edu/software/web-entity-extractor-ACL2014/

Collaborators

Jonathan Berant (post-doc)
Andrew Chou (masters)
Roy Frostig (Ph.D.)
Panupong Pasupat (Ph.D.)

Thank you!