

Computing with Natural Language

Percy Liang

ACL Workshop on Semantic Parsing - June 15, 2014

Stanford University



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 period¹. 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^{2–4}, but the mechanisms responsible for these perturbations remain uncertain⁵. Here we use new stratigraphic and geochemical data to show that early Paleozoic 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 Unconformity⁶. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this paleogeomorphic 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 filled sedimentary rocks of the 1,200–740-Myr-old Grand Canyon Supergroup⁷. 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^{8,9}, Baltica¹⁰, Avalonia¹¹ and Siberia¹², 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 adductor-limb crustacean animals, a fact that some paleontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution¹³.

Here we use stratigraphic and lithologic data for 21,531 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^{14,15} 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^{16,17}, 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 shoreline system, often called the ‘wave-base razor’¹⁸, 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 interiors¹⁹ 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 regoliths^{20,21} and

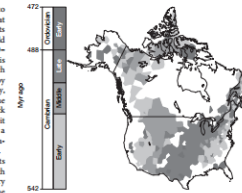


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology 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 period¹. 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^{2,3}, but the mechanisms responsible for these perturbations remain uncertain⁴. 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 Unconformity⁵. 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 filled sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup⁶. 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}, Baltica⁹, Avalonia¹⁰ and Siberia¹¹, 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 additional crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution¹².

Here we use stratigraphic and lithologic data for 21,531 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^{13,14} 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^{15,16}, 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 shoreline system, often called the ‘wave-base razor’¹⁷, 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 interiors¹⁸ 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^{19,20}, and

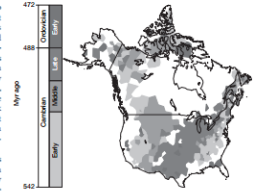


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology 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 period¹. 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^{2,3}, but the mechanisms responsible for these perturbations remain uncertain⁴. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–600 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 Unconformity⁵. 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 filled sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup⁶. 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}, Baltica⁹, Avalonia¹⁰ and Siberia¹¹, 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 eukaryotic crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic hiatus and an incomplete record of early animal evolution¹².

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^{13,14} 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^{15,16}, 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 shoreline system, often called the ‘wave-base razor’¹⁷, 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 interiors¹⁸ 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^{19,20}, and

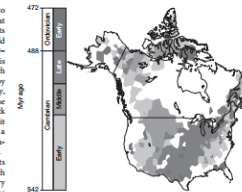


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology 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 period¹. 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^{2,3}, but the mechanisms responsible for these perturbations remain uncertain⁴. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–600 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. 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 filled sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup⁵. 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^{6,7}, Baltica⁸, Avalonia⁹ and Siberia¹⁰, 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 eukaryotic crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic hiatus and an incomplete record of early animal evolution¹¹.

Here we use stratigraphic and lithologic data for 21,531 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^{12,13} 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^{14,15}, 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 shoreline system, often called the ‘wave-base razor’¹⁶, 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 interiors¹⁷ 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^{18,19}, and

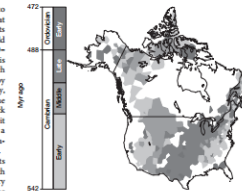


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology 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?

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 period¹. 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^{2,3}, but the mechanisms responsible for these perturbations remains uncertain⁴. Here we use new stratigraphic and geochemical data to show that early Paleozoic marine sediments deposited approximately 540-600 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. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this paleogeomorphic 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 Tapesats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally filled sedimentary rocks of the 1,200-740 Myr-old Grand Canyon Supergroup⁵. 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^{6,7}, Baltica⁸, Avalonia⁹ and Siberia¹⁰, 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 eukaryotic crown-group animals, a fact that some paleontologists have interpreted as evidence for stratigraphic hiatus and an incomplete record of early animal evolution¹¹.

Here we use stratigraphic and lithologic data for 21,531 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^{12,13} 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^{14,15}, 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 shoreline system, often called the 'wave-base razor'¹⁶, 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 interiors¹⁷ 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^{18,19}, and

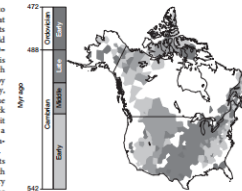


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology 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

LocationOf.argmax(Type.Occurrence \sqcap Genus.Mammut, 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

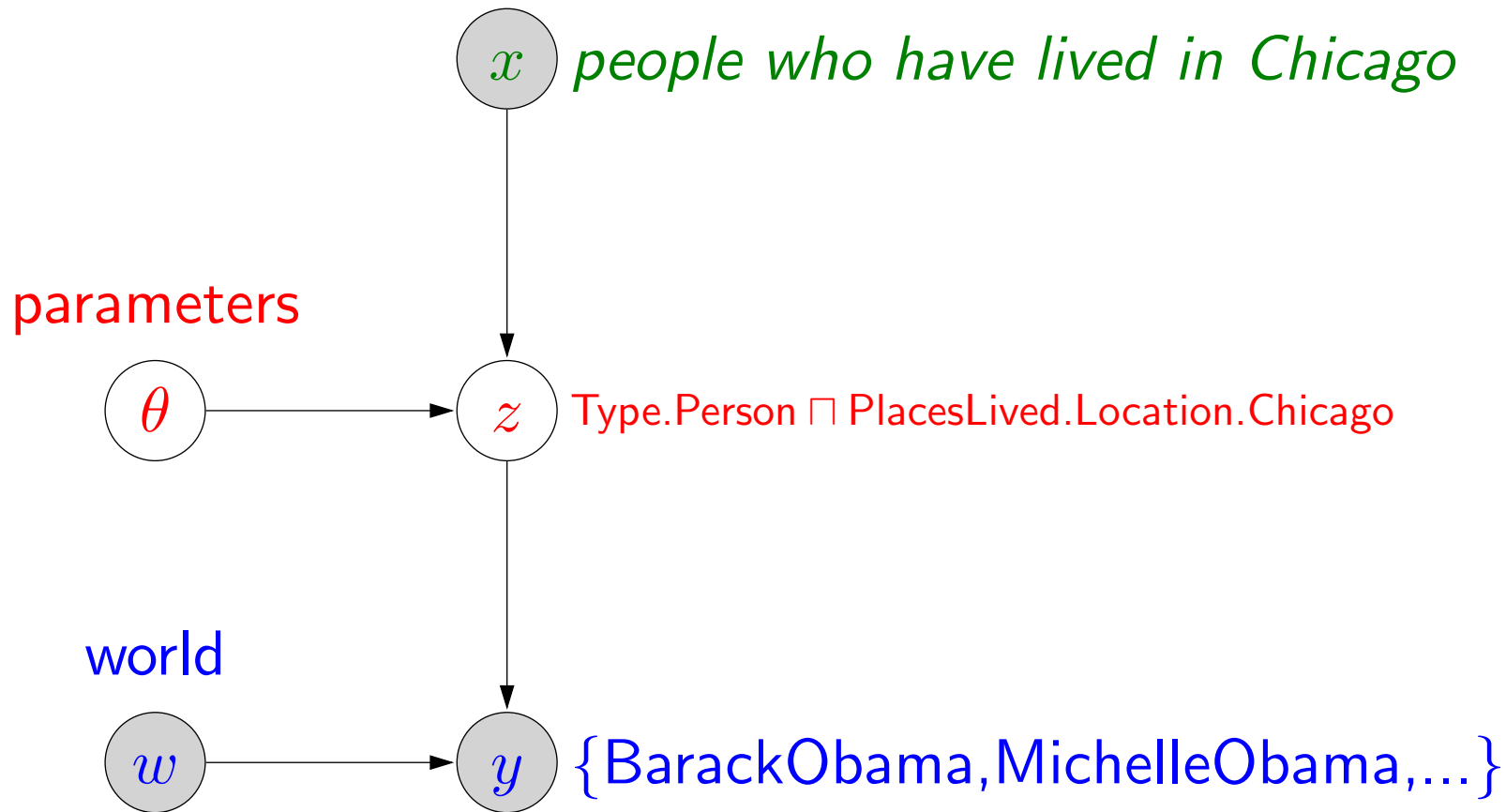
[behavior: user output]

Programs affect the world

Outline

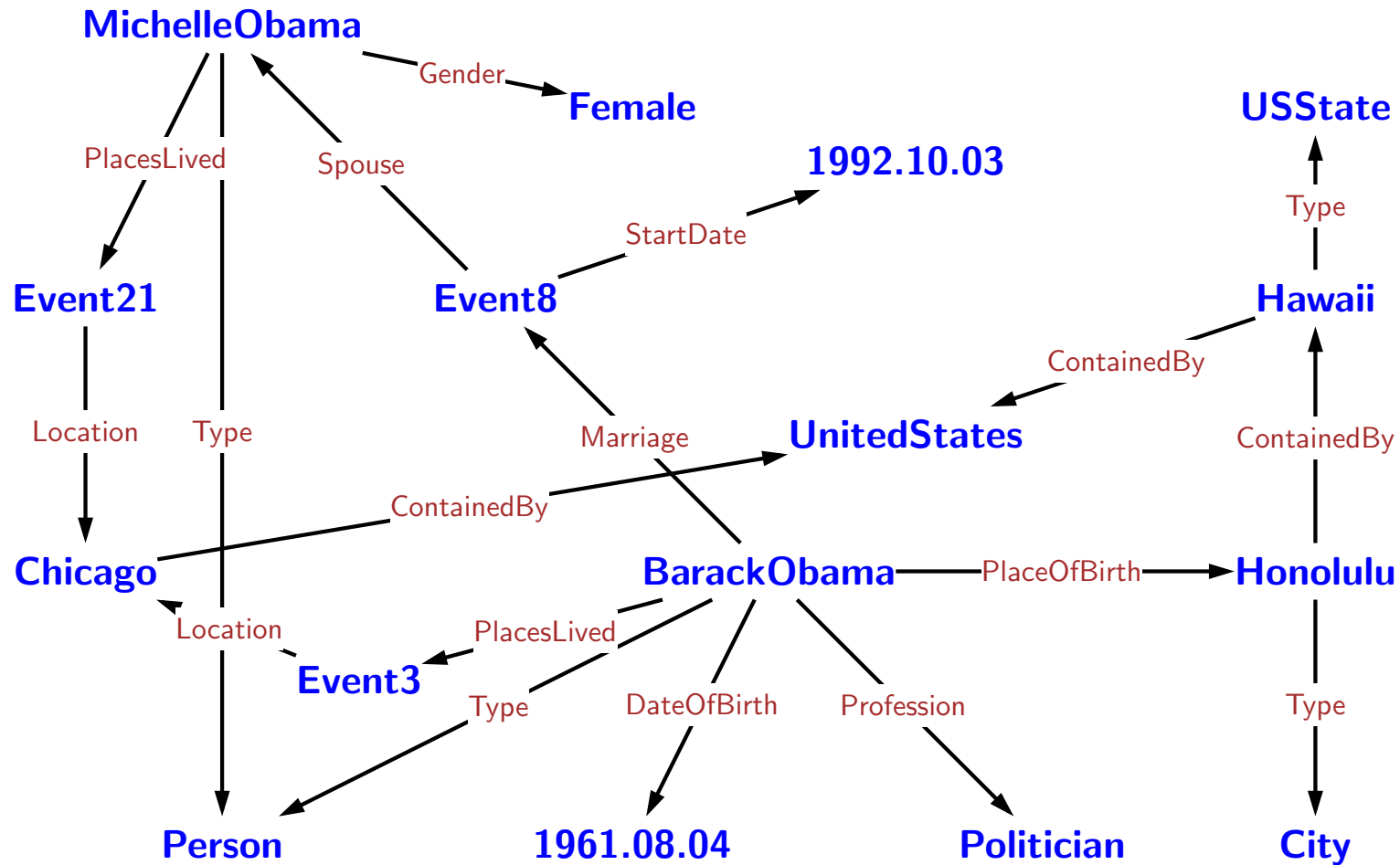
- **Semantic parsing in 5 minutes**
- A closer look at the elements
 - Knowledge base incompleteness
 - Lexical coverage
 - Search over logical forms
 - Learning via bootstrapping
 - Leveraging denotations ("grounding")
 - Datasets
- Final remarks

Framework



World: Freebase

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

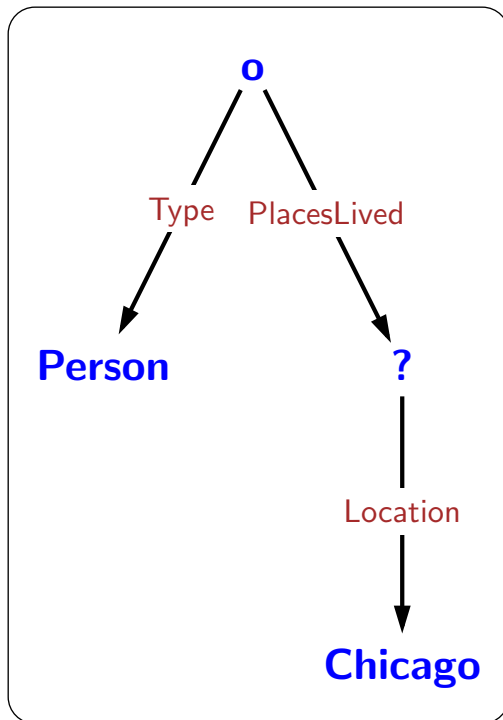


Logical forms

Type.Person \sqcap PlacesLived.Location.Chicago

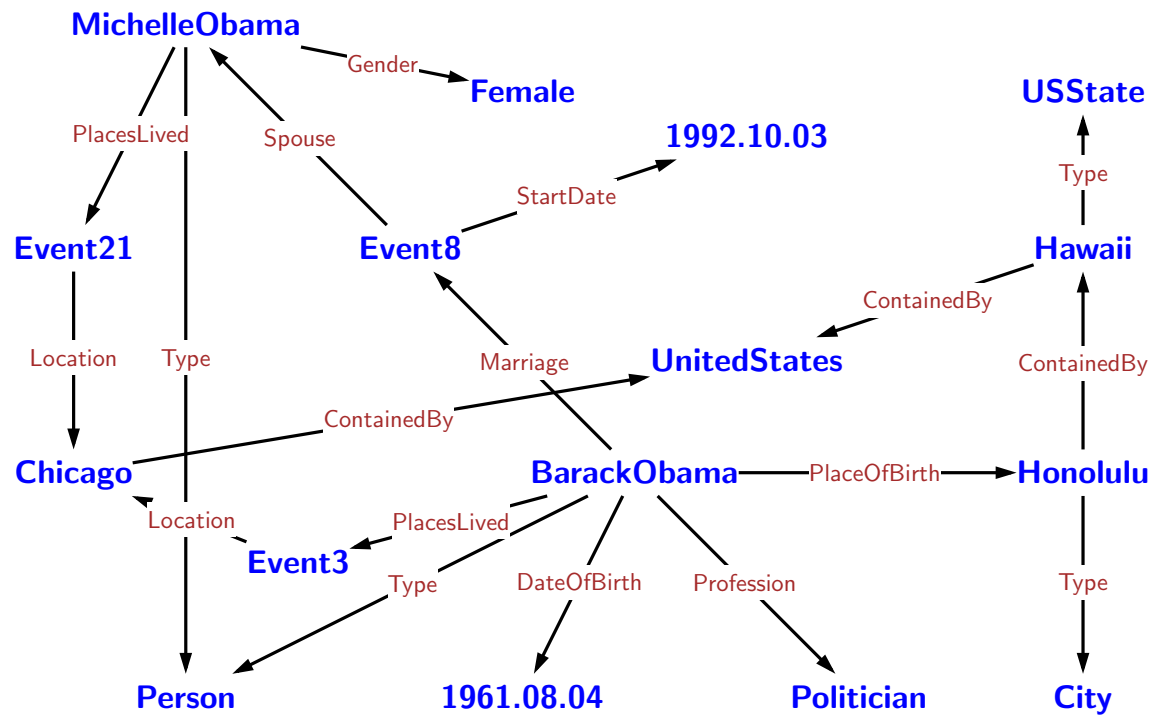
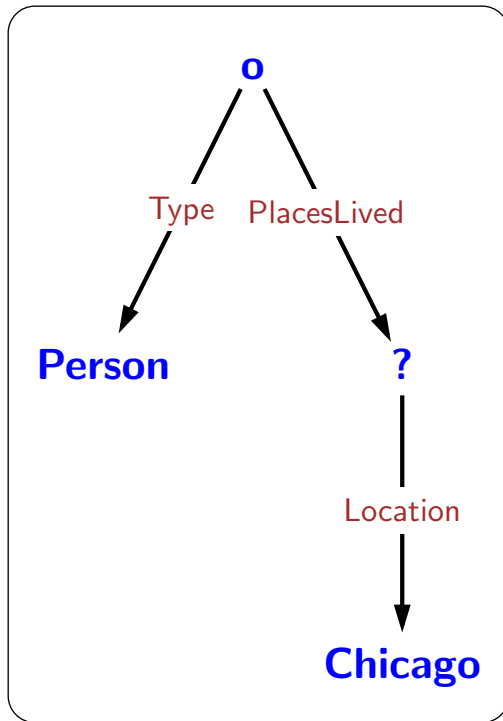
Logical forms

Type.Person \sqcap PlacesLived.Location.Chicago



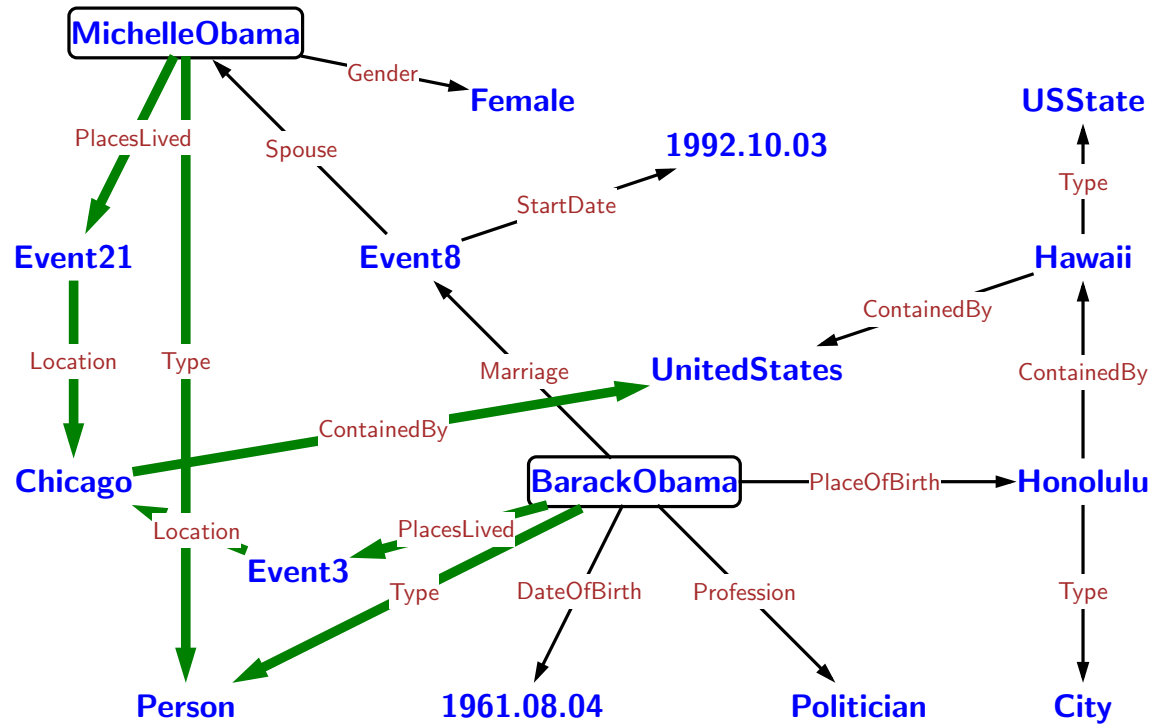
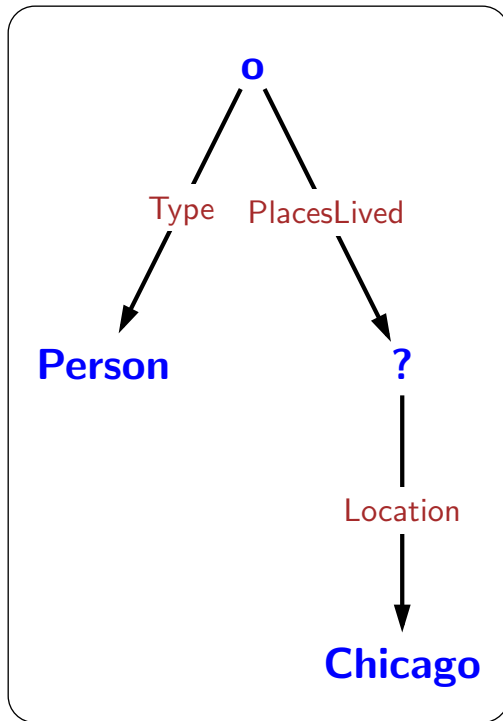
Logical forms

Type.Person \sqcap PlacesLived.Location.Chicago

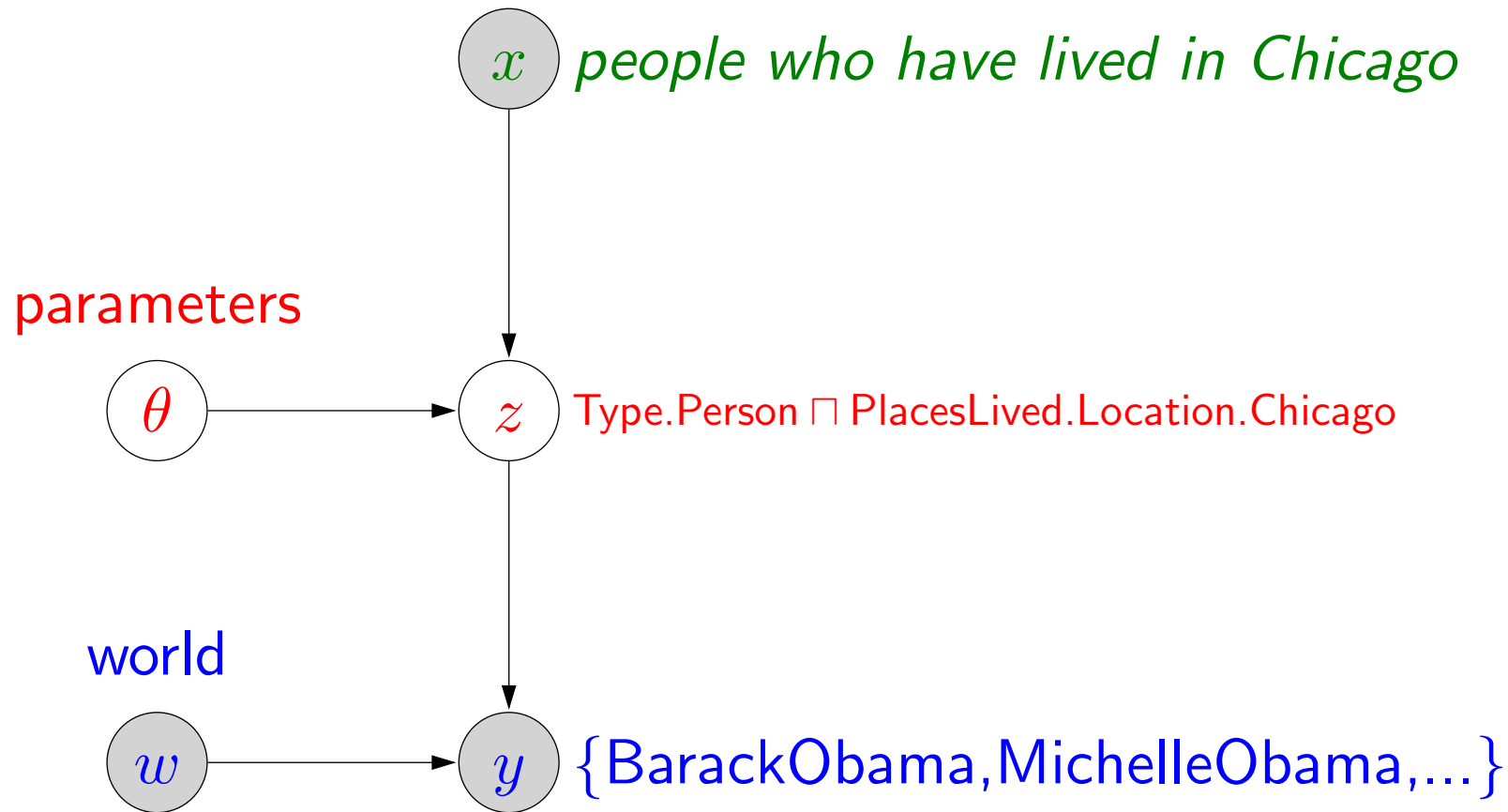


Logical forms

Type.Person \sqcap PlacesLived.Location.Chicago

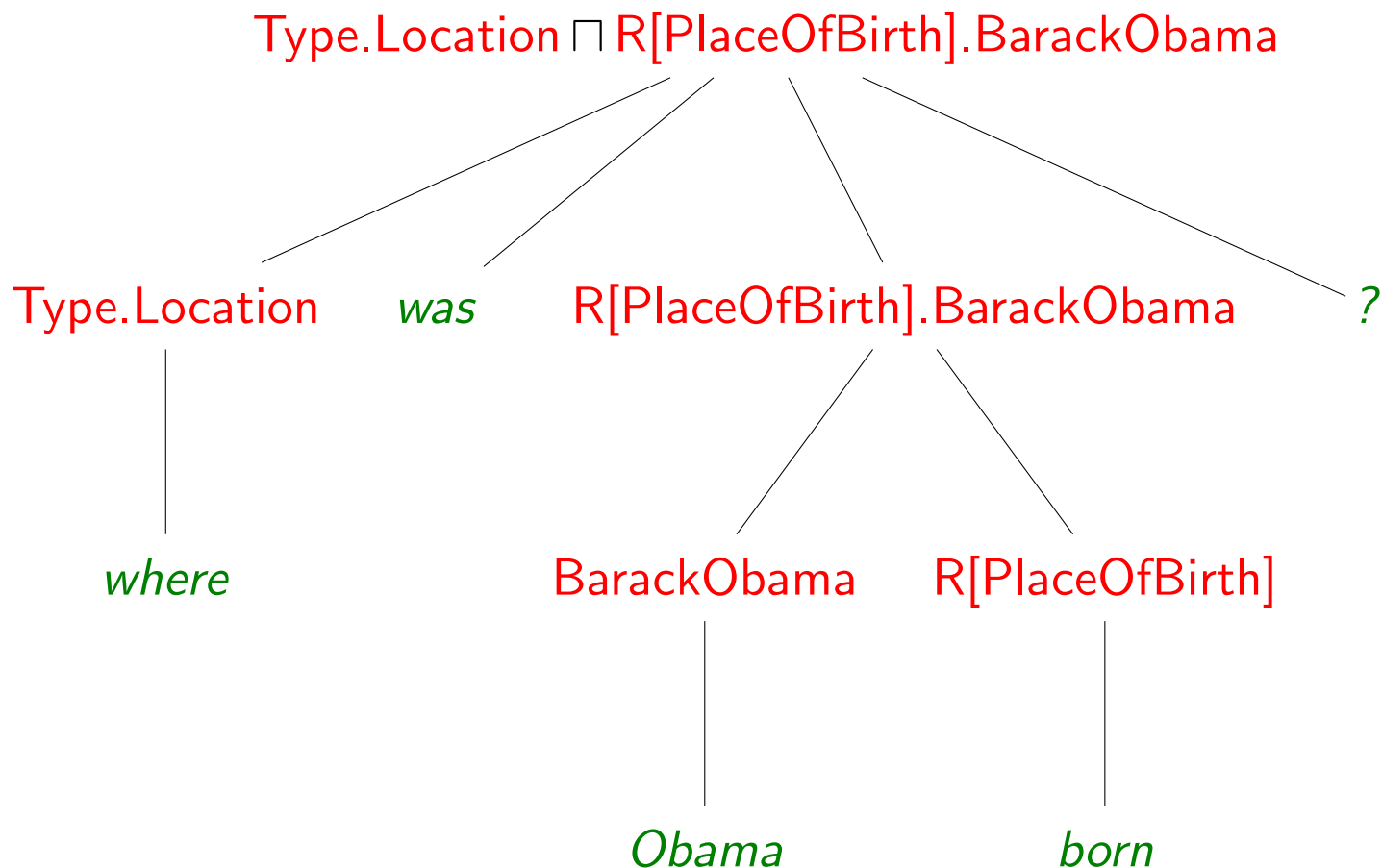


Framework



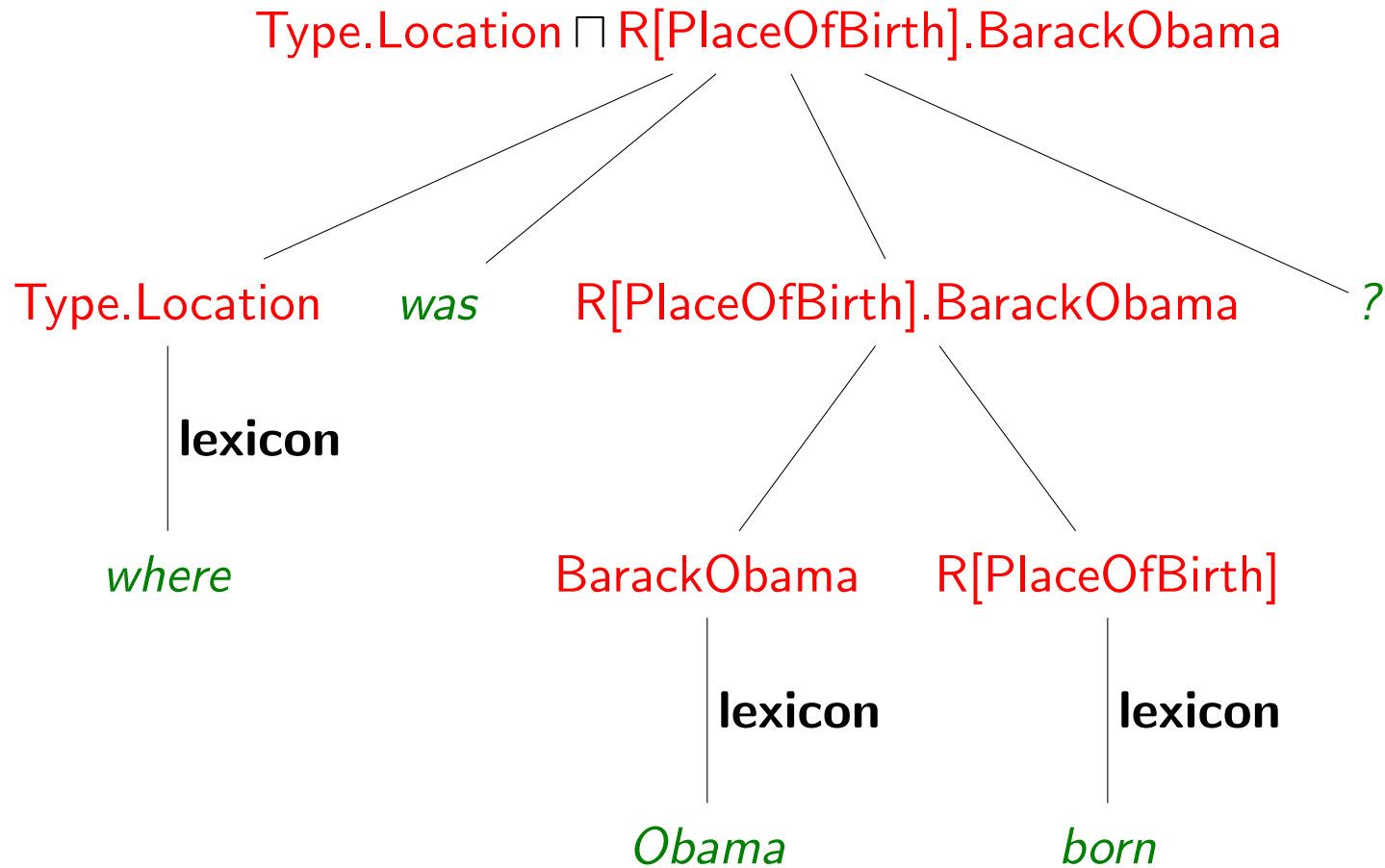
Derivations

Derivation: construction of logical form given utterance



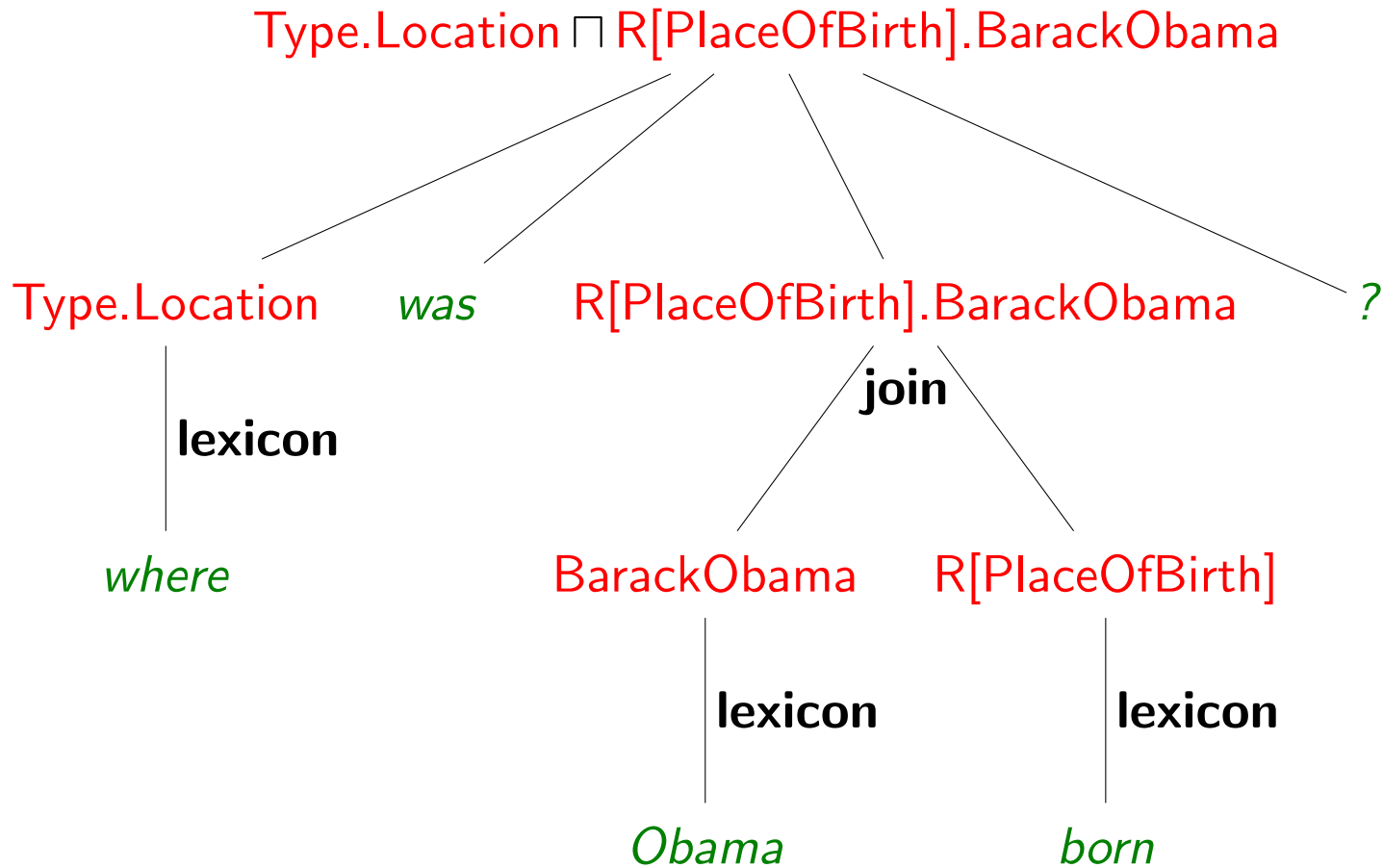
Derivations

Derivation: construction of logical form given utterance



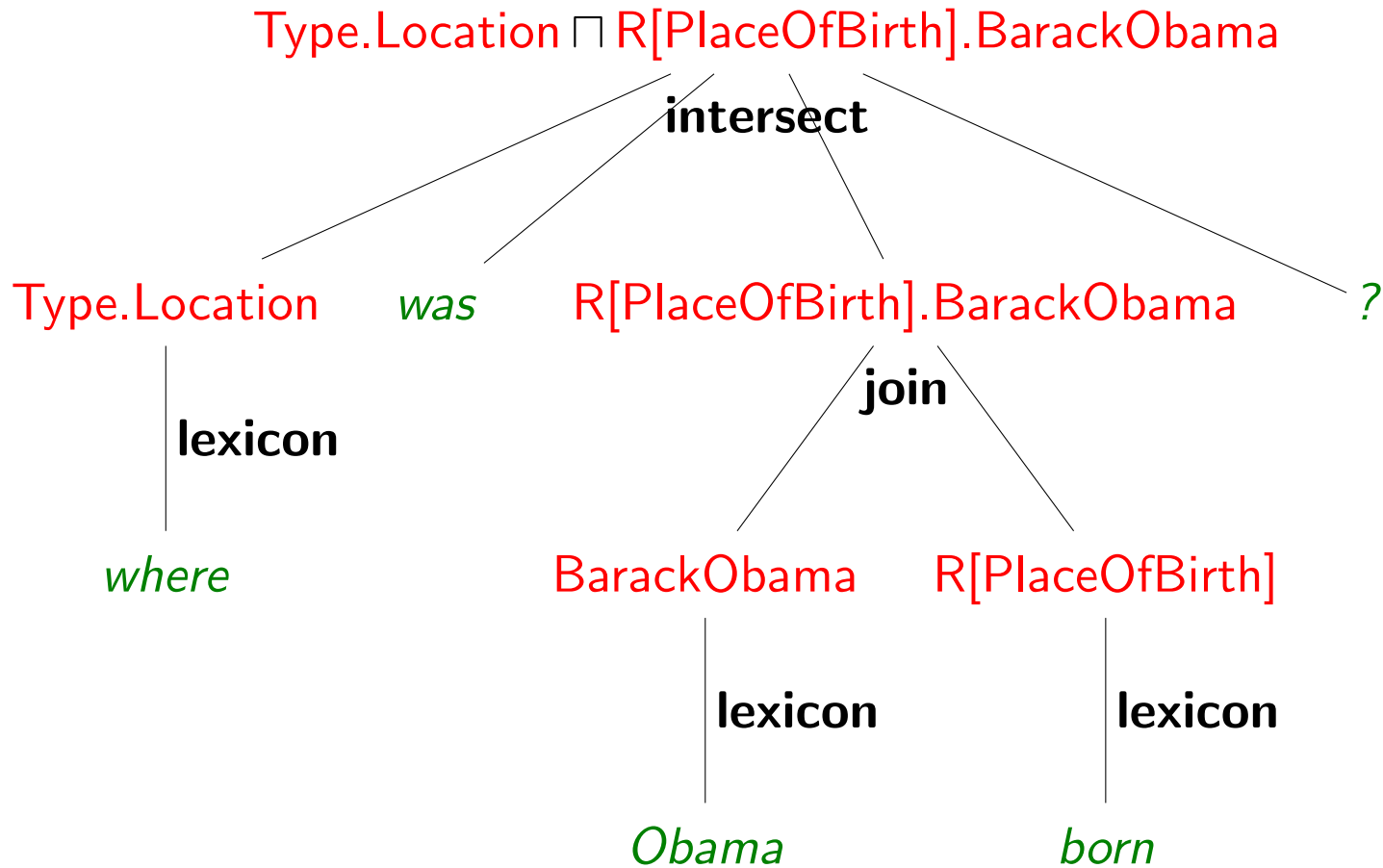
Derivations

Derivation: construction of logical form given utterance

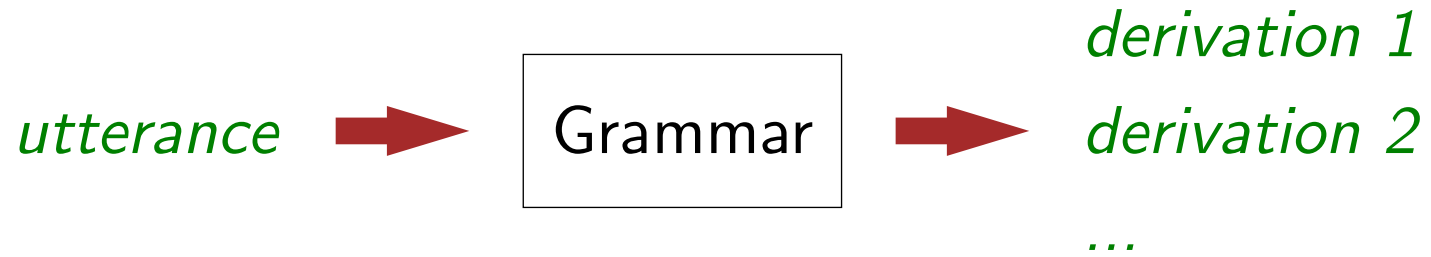


Derivations

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 : $b.u$

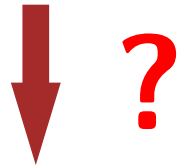
(intersect) Unary : u Unary : v \Rightarrow Unary : $u \sqcap v$

Many possible derivations!

Where was Obama born?

Many possible derivations!

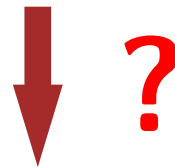
Where was Obama born?



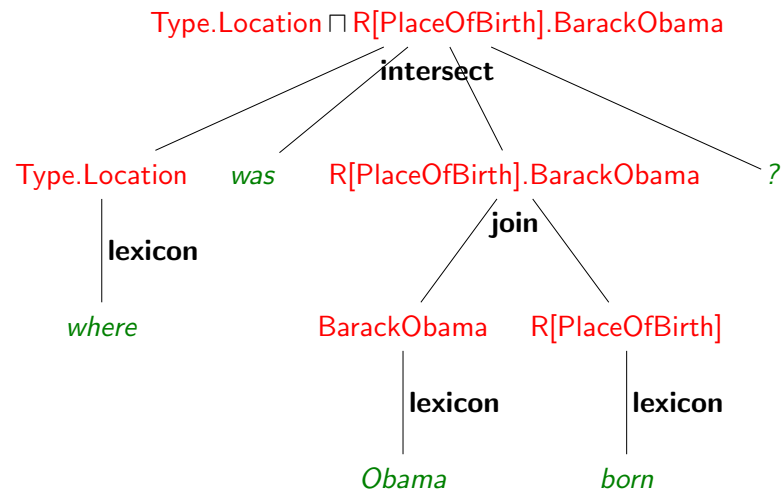
set of candidate derivations $\mathcal{D}(x)$

Many possible derivations!

Where was Obama born?

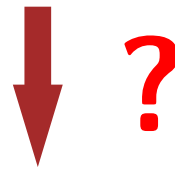


set of candidate derivations $\mathcal{D}(x)$

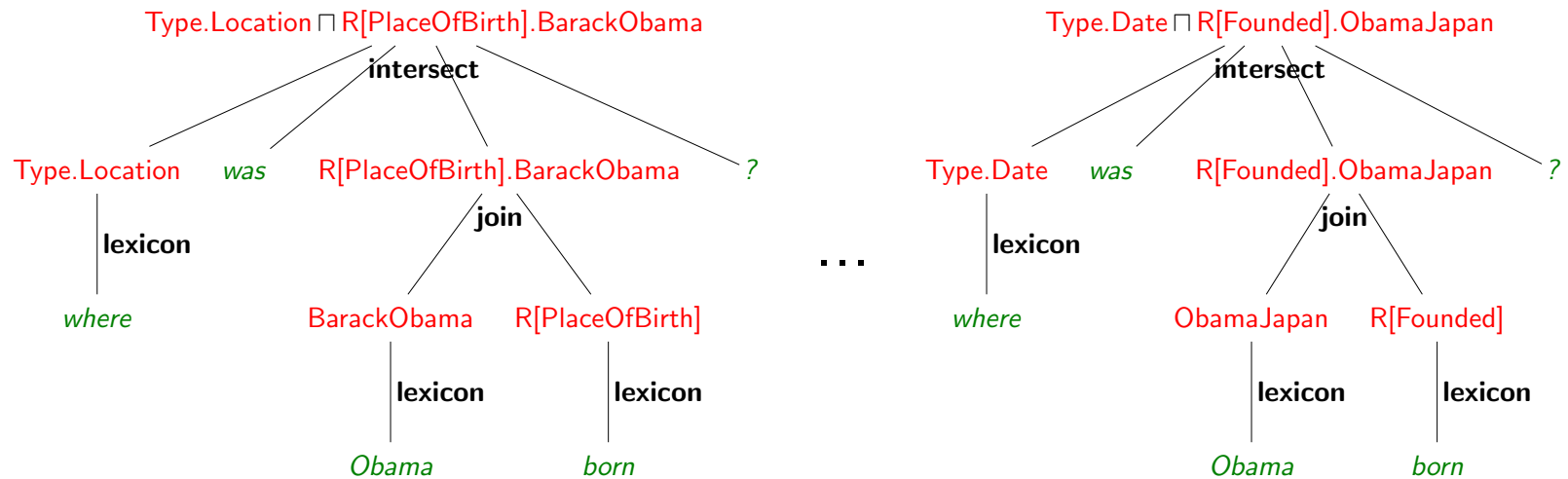


Many possible derivations!

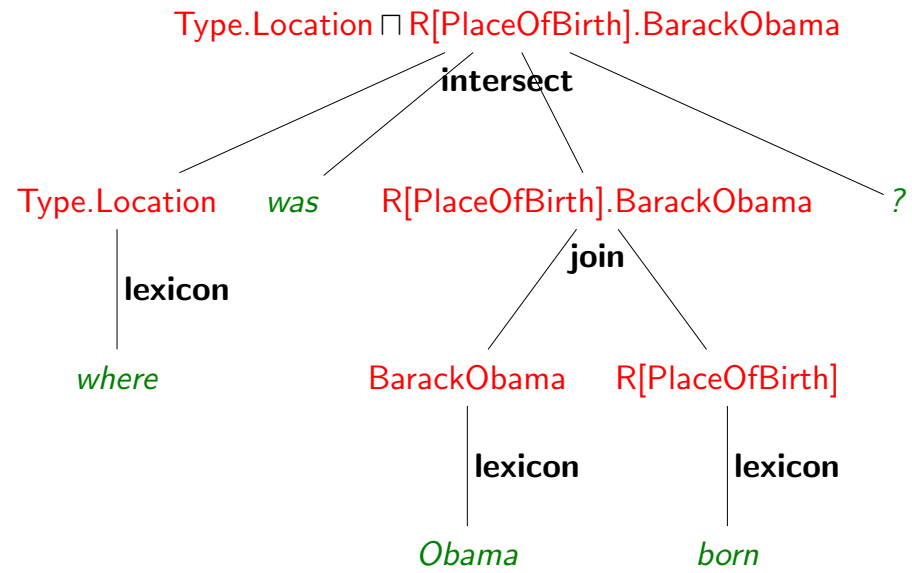
Where was Obama born?



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

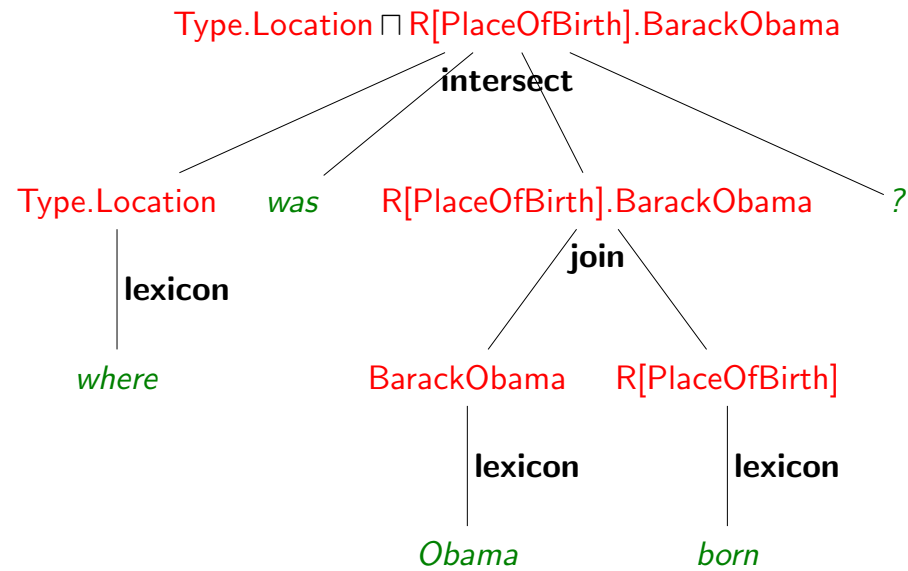


x : utterance
 d : derivation



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
<i>born</i> maps to PlaceOfBirth	1
<i>born</i> maps to PlacesLived.Location	0
alignmentScore	1.52
denotation-size=1	1
...	...

Scoring derivations

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

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

Scoring function:

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

Log-linear model

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

Log-linear model

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

Model: distribution over derivations d given utterance x

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

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

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

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

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

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 ⇒ Salzburg

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Training intuition

Where did Mozart tuppess?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tuppess?

Training intuition

Where did Mozart tuppess?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tuppess?

PlaceOfBirth.WilliamHogarth

PlaceOfDeath.WilliamHogarth

PlaceOfMarriage.WilliamHogarth

London

Training intuition

Where did Mozart tuppess?

~~PlaceOfBirth.Mozart~~ → ~~Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tuppess?

PlaceOfBirth.WilliamHogarth ⇒ London

PlaceOfDeath.WilliamHogarth ⇒ London

PlaceOfMarriage.WilliamHogarth ⇒ Paddington

London

Training intuition

Where did Mozart tuppess?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tuppess?

PlaceOfBirth.WilliamHogarth ⇒ London

PlaceOfDeath.WilliamHogarth ⇒ London

~~PlaceOfMarriage.WilliamHogarth → Paddington~~

London

Training intuition

Where did Mozart tupsress?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tupsress?

PlaceOfBirth.WilliamHogarth ⇒ London

PlaceOfDeath.WilliamHogarth ⇒ London

~~PlaceOfMarriage.WilliamHogarth → Paddington~~

London

Outline

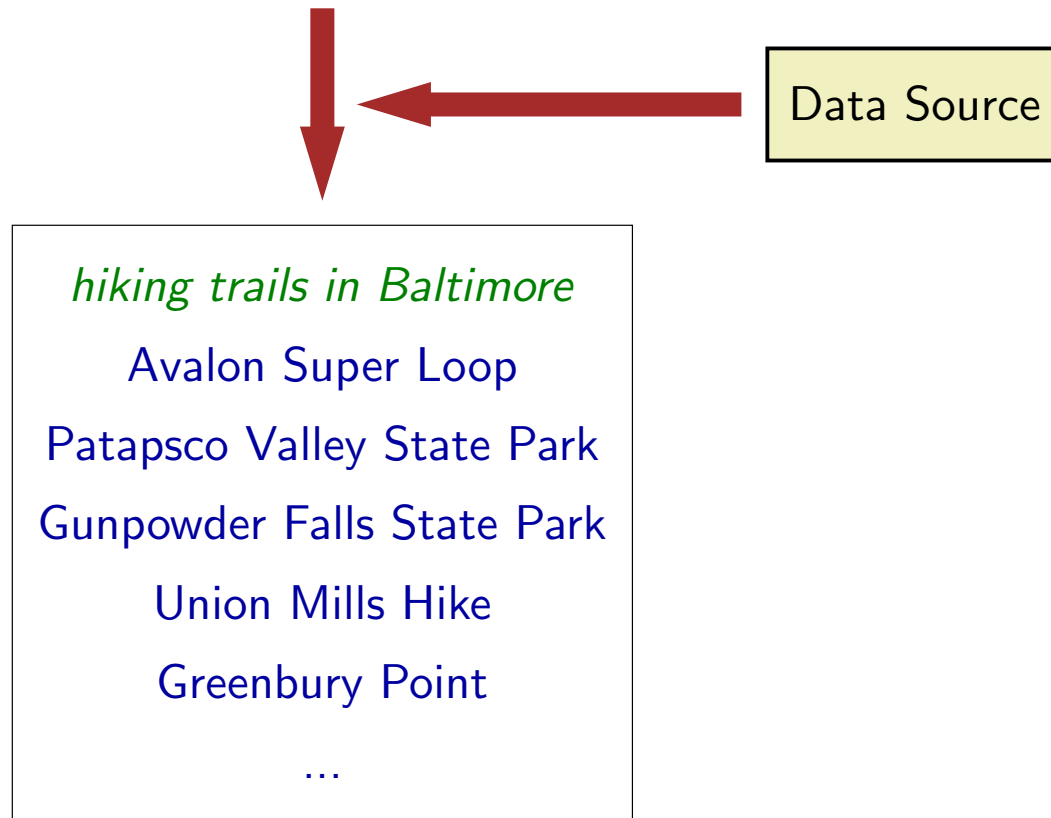
- Semantic parsing in 5 minutes
- **A closer look at the elements**
 - Knowledge base incompleteness
 - Lexical coverage
 - Search over logical forms
 - Learning via bootstrapping
 - Leveraging denotations ("grounding")
 - Datasets
- Final remarks

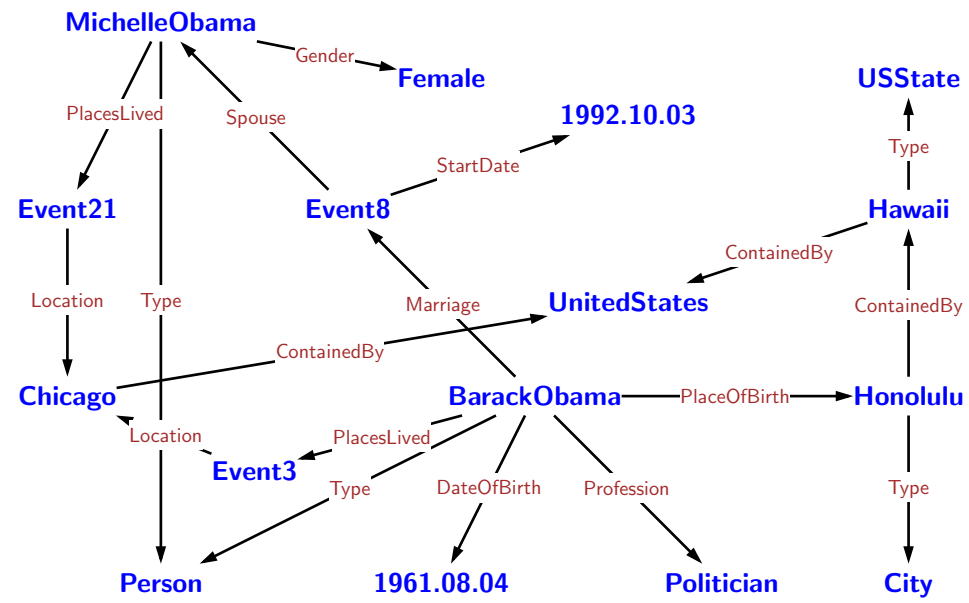
Outline

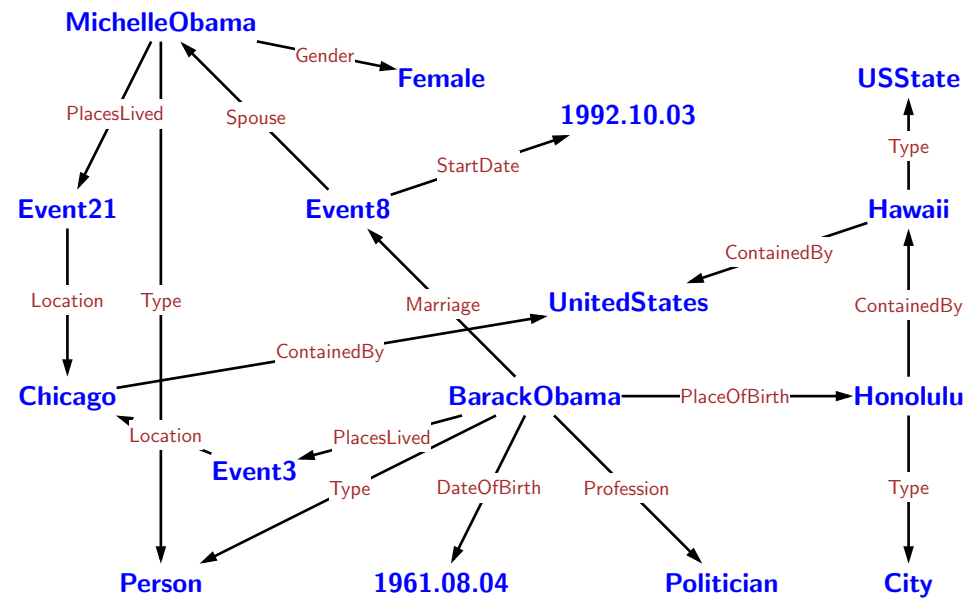
- Semantic parsing in 5 minutes
- A closer look at the elements
 - **Knowledge base incompleteness**
 - Lexical coverage
 - Search over logical forms
 - Learning via bootstrapping
 - Leveraging denotations ("grounding")
 - Datasets
- Final remarks

Challenge: incomplete knowledge base

What are the longest hiking trails in Baltimore?







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

ACCEPTED LONG PAPERS

ACL 2014

- A Bayesian Mixed Effects Model of Literary Character
David Bamman, Ted Underwood and Noah A. Smith
- A chance-constrained measure of inter-annotator agreement for syntax
Arne Sippel
- A Decision Theoretic Approach to Natural Language Generation
Nathan McKinley and Saumya May
- A Descriptive Graph-Based Parser for the Abstract Meaning Representations
Jeffrey Flanigan, Sam Thomson, John Carbonell, Goro Doi and Mostafa Dehghani
- A Generalized Language Model
Rene Plathard, Thomas Gatterbauer and Gero Grottel
- A Joint Graph Model for Phylogenetic Tree Reconstruction
Zongyi Jin and Hui Zhao
- A Learning-Based Algorithm for Text Classification
Alex Atkeson, Alexander M. Rush, Michael Collins and John D. Elson
- A Linear-Time Bottom-Up Disambiguation Algorithm
Vanesa Wei Feng and Goro Doi
- A practical but linguistically-motivated approach to word segmentation
Denis Paperno, Theodor A. G. Fischer and Gero Grottel
- A Provably Correct Learning Algorithm for Text Classification
Shay B. Cohen and Michael Collins

Regular Faculty

54 people

Name

Alex Atkeson

Berfin Batozoglou

Gil Bejerman

Michael Bernstein

Dan Borah

David Cherton

Steve Cooper

Bill Dally

David Diez

Ron Dror













Dawson Engler

Ron Fedkiw

Most Popular Action Feature Films

- 
Godzilla (2014)
 ★★★★★★ 7.2/10
 The world's most famous monster is pitted against invulnerable creatures who, in humanity's scientific arrogance, threaten our very existence.
 Dir: Gareth Edwards With: Aaron Taylor-Johnson, Elizabeth Olsen, Bryan Cranston
 Action | Sci-Fi | Thriller
- 
X-Men: Days of Future Past
 ★★★★★★
 The X-Men send Wolverine to the past in a desperate effort to change history and prevent a future that results in the extinction of his kind.
 Dir: Bryan Singer With: Hugh Jackman, James McAvoy, Michael Fassbender
 Action | Adventure
- 
The Amazing Spider-Man
 ★★★★★★
 Peter Parker runs the gauntlet against Iron Man, the Hulk, and the Vulture.
 Dir: Marc Webb With: Tobey Maguire, Kirsten Dunst, Andrew Garfield
 Action | Adventure
- 
Transformers: Age of Extinction
 An automobile mechanic saves the world from a deadly alien invasion.
 Dir: Michael Bay With: Mark Wahlberg, Will Bridges, Anthony Mackie
 Action | Adventure

#	President	Took office
1	 George Washington (1732-1799) (1789-1797)	April 30, 1789 (3-3)
2	 John Adams (1735-1826) (1797-1801)	March 4, 1797
3	 Thomas Jefferson (1743-1826) (1801-1809)	March 4, 1801
4	 James Madison (1751-1836) (1809-1817)	March 4, 1809

Thursday 12 June			
12:00 (2014 - 2016 Local Time)	 BRAZIL	17:00	CROATIA 
Friday 13 June			
13:00 (2014 - 2016 Local Time)	 MEXICO	13:00	CAMEROON 
16:00 (2014 - 2016 Local Time)	 SPAIN	16:00	NETHERLANDS 
18:00 (2014 - 2016 Local Time)	 CHILE	18:00	AUSTRALIA 
Saturday 14 June			
13:00 (2014 - 2016 Local Time)	 COLOMBIA	13:00	GREECE 
16:00 (2014 - 2016 Local Time)	 URUGUAY	16:00	COSTA RICA 

March 4, 1817	Democratic-Republican	Secretary of State (1801-1809)	March 4, 1817 April 28, 1822 Henry Clay March 4, 1823 - November 23, 1824 Vice President November 23, 1824 - March 4, 1827
---------------	-----------------------	-----------------------------------	---

Semantic parsing on the web

Input:

- query x

hiking trails near Baltimore

- web page w

Semantic parsing on the web

Input:

HOME | EXPLORE | MOBILE APPS | CREATE TRIP | MY EVERYTRAIL GO

(Update Current Location) Login | Signup

Hiking near Baltimore, Maryland

Like 49 people like this. Tweet 1

This list shows the most popular Hiking near Baltimore, Maryland based on user reviews, votes, and mobile downloads. Plan your next trip with EveryTrail guides by downloading a guide to your mobile phone with the EveryTrail iPhone or Android app.

Sort: show community trips

Filter Trails

Guides

Avalon Super Loop - Patapsco State Park
 Patapsco State Park, Maryland, United States (7.5 miles away)
 ★★★★★
Difficult: 12.7 miles, Full day
 lots of ruins, waterfalls, trains, and river views

Do the entire Avalon Patapsco state park in 1 day! This loop covers the majority of the Avalon area, with multiple ruins, waterfalls and other artifacts to find along the way. Starting at the parking lot, you hike up the road a ways to the Ridge trail sign. The next leg is the maintenance loop which has an old old tractor to look at and some...

Patapsco Valley State Park - Hilton Area 8 Miles/Moderate
 Catonsville, Maryland, United States (7.7 miles away)
Moderate: 7.8 miles, Half day
 8 mile circuit hike including sections in the Avalon, Orange Grove and Glen Artney areas of PVSP.

OVERVIEW: One of the more scenic routes in the Patapsco Valley State Park in the Hilton Area which includes multiple stream crossings, viewings and waterfalls including Cascade waterfalls, two swinging bridge crossings, Ilchester Overlook, and Bloedes Dam. This is a moderate hike and can be hiked in either direction. Counterclockwise is an easier hike...

Popular places for Hiking

- Hiking in Maryland
- Hiking in Patapsco Valley State Park
- Hiking in Calvert Cliffs State Park
- Hiking in Patuxent River State Park

Semantic parsing on the web

Input:

HOME | EXPLORE | MOBILE APPS | CREATE TRIP | MY EVERYTRAIL GO

(Update Current Location) Login | Signup

Hiking near Baltimore, Maryland


Like 49 people like this. Tweet 1

This list shows the most popular Hiking near Baltimore, Maryland based on user reviews, votes, and mobile downloads. Plan your next trip with EveryTrail guides by downloading a guide to your mobile phone with the EveryTrail iPhone or Android app.

Sort: show community trips

Filter Trails


Guides



Avalon Super Loop - Patapsco State Park
Patapsco State Park, Maryland, United States (7.5 miles away)
★★★★★

Difficult: 12.7 miles, Full day
 lots of ruins, waterfalls, trains, and river views

Do the entire Avalon Patapsco state park in 1 day! This loop covers the majority of the Avalon area, with multiple ruins, waterfalls and other artifacts to find along the way. Starting at the parking lot, you hike up the road a ways to the Ridge trail sign. The next leg is the maintenance loop which has an old old tractor to look at and some...



Patapsco Valley State Park - Hilton Area 8 Miles/Moderate
Patapsco Valley State Park, Maryland, United States (8 miles away)

Moderate: 7.8 miles, Half day
 8 mile circuit hike including sections in the Avalon, Orange Grove and Glen Artney areas of PVSP.

OVERVIEW: One of the more scenic routes in the Patapsco Valley State Park in the Hilton Area which includes multiple stream crossings, viewings and waterfalls including Cascade waterfalls, two swinging bridge crossings, Ilchester Overlook, and Bloedes Dam. This is a moderate hike and can be hiked in either direction. Counterclockwise is an easier hike...

Popular places for Hiking

- Hiking in Maryland
- Hiking in Patapsco Valley State Park
- Hiking in Calvert Cliffs State Park
- Hiking in Patuxent River State Park

Semantic parsing on the web

Input:

- query x

hiking trails near Baltimore

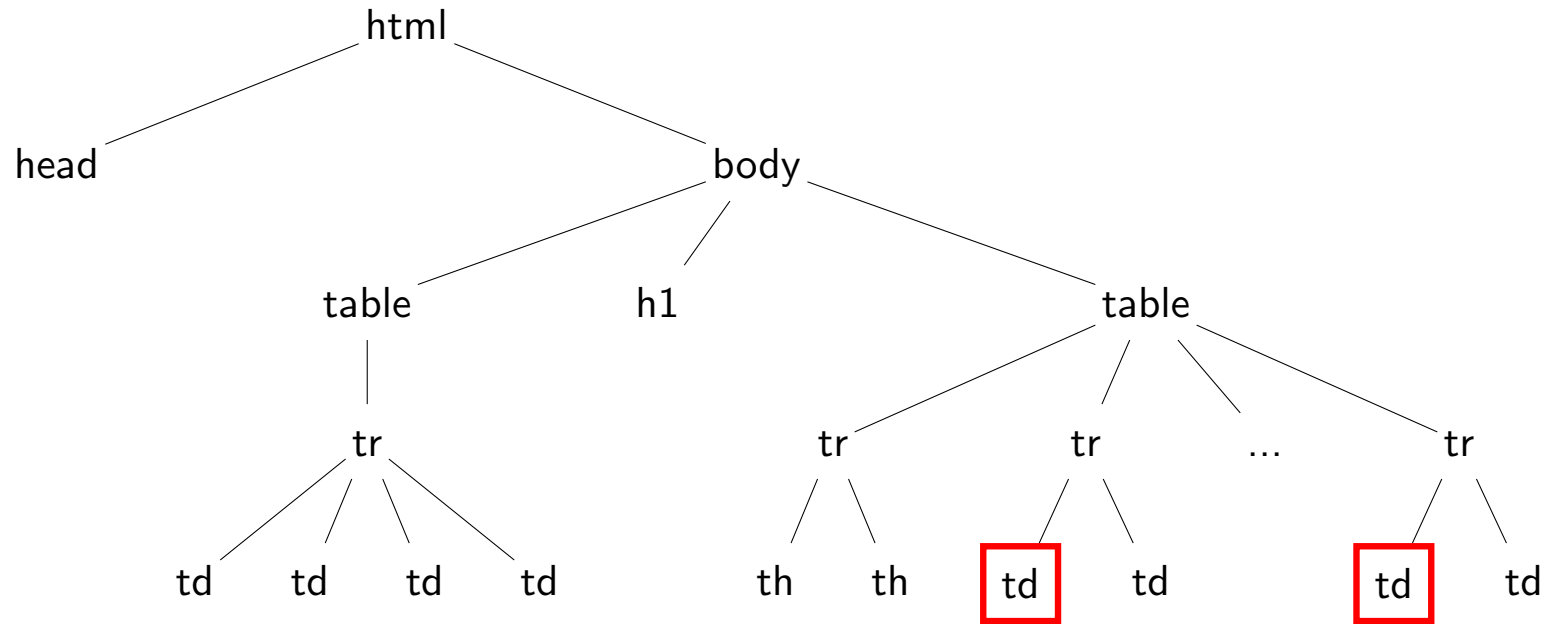
- web page w

Output:

- list of entities y

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

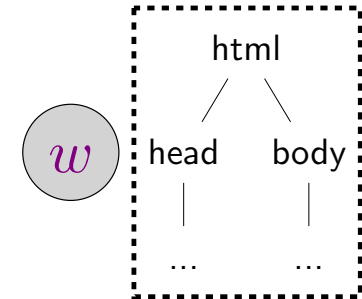
Logical forms: XPath expressions



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

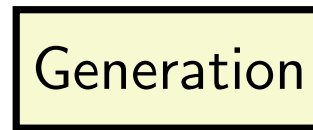
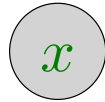
Framework

*hiking trails
near Baltimore* x

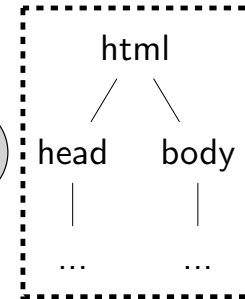
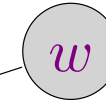


Framework

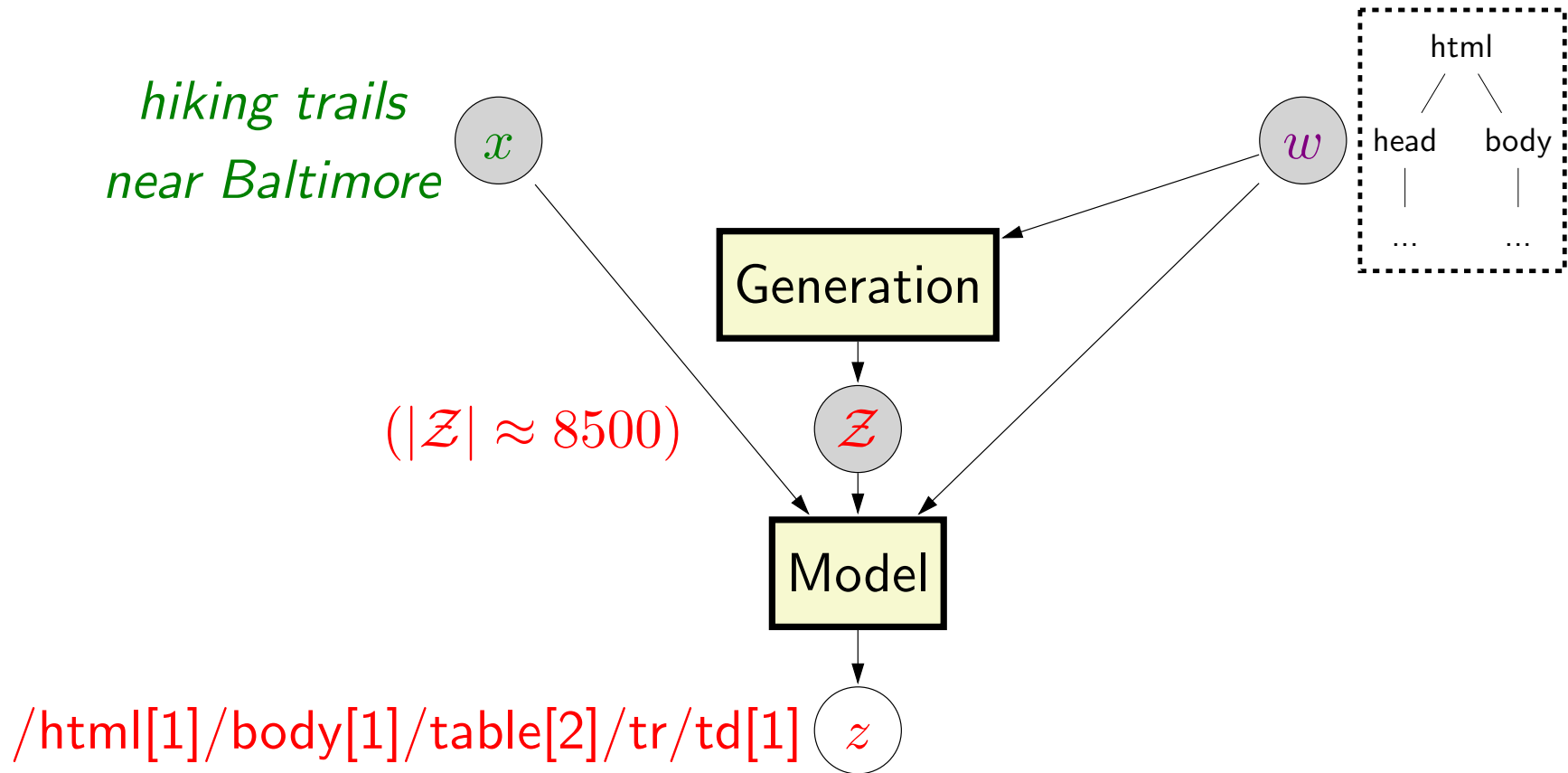
*hiking trails
near Baltimore*



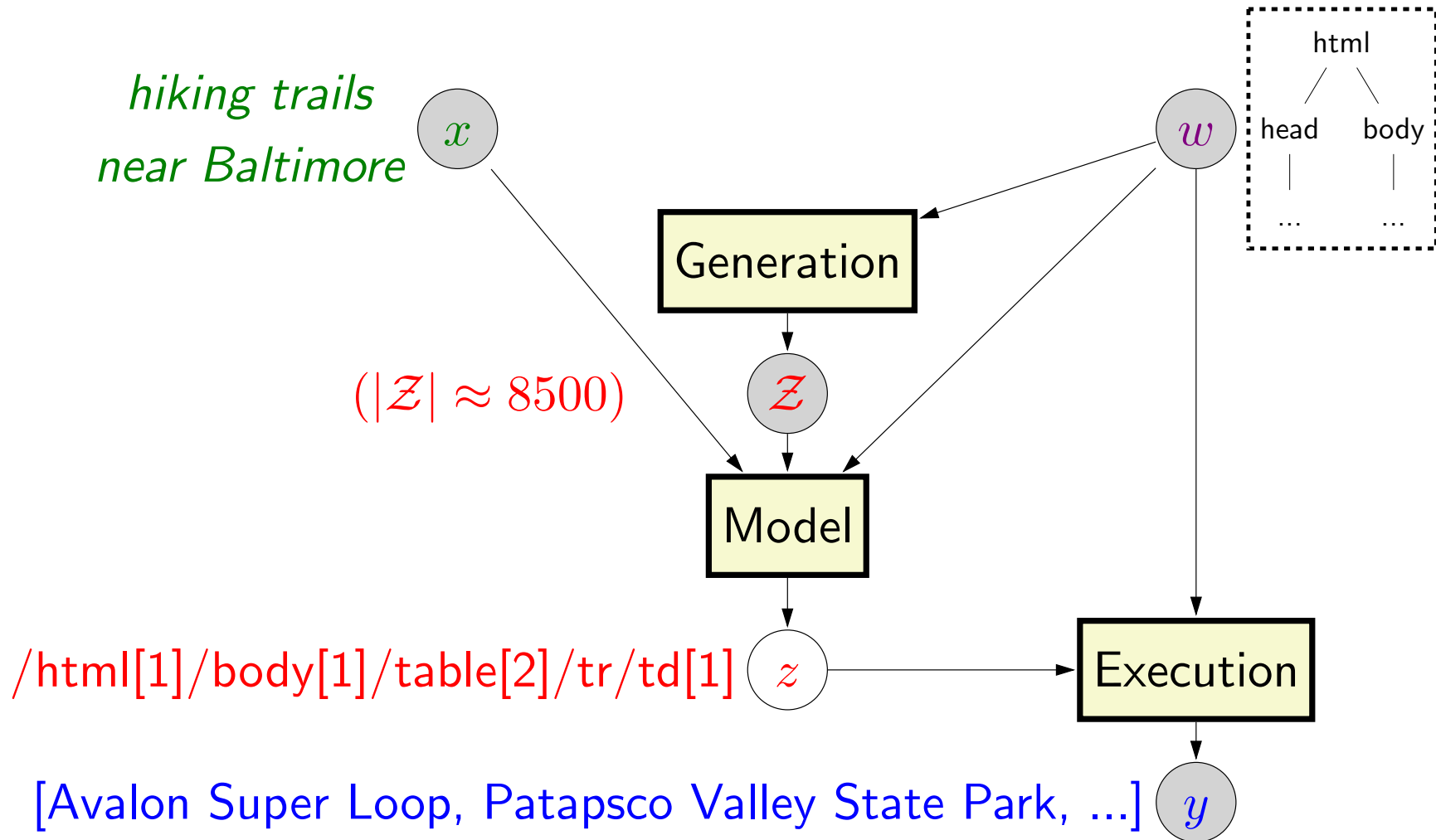
$(|\mathcal{Z}| \approx 8500)$



Framework



Framework



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

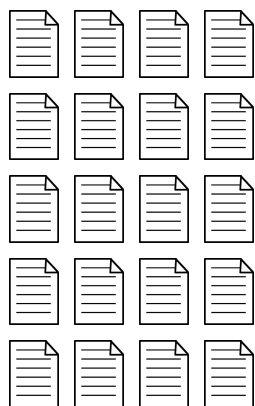
Challenge: lexical coverage

born ⇒ Type.City, PeopleBornHere, Profession.Lawyer, ...

?

Solution: alignment

Open information extraction on ClueWeb09:



(Barack Obama, was born in, Honolulu)

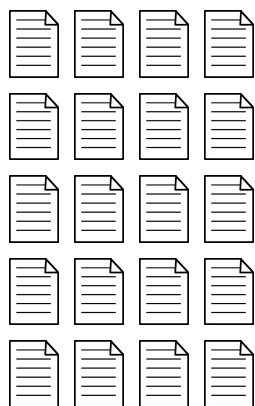
(Albert Einstein, was born in, Ulm)

(Barack Obama, lived in, Chicago)

... 15M triples ...

Solution: alignment

Open information extraction on ClueWeb09:



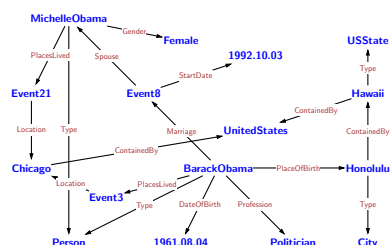
(Barack Obama, was born in, Honolulu)

(Albert Einstein, was born in, Ulm)

(Barack Obama, lived in, Chicago)

... 15M triples ...

Freebase:



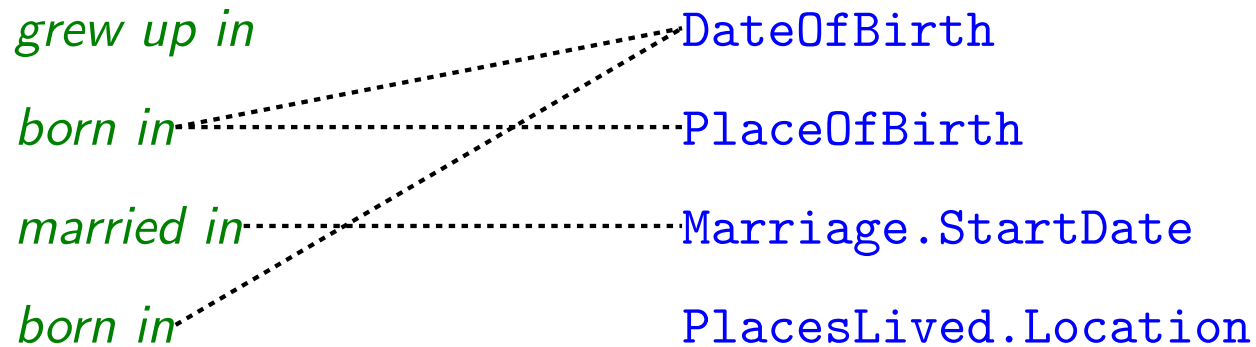
(BarackObama, PlaceOfBirth, Honolulu)

(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!

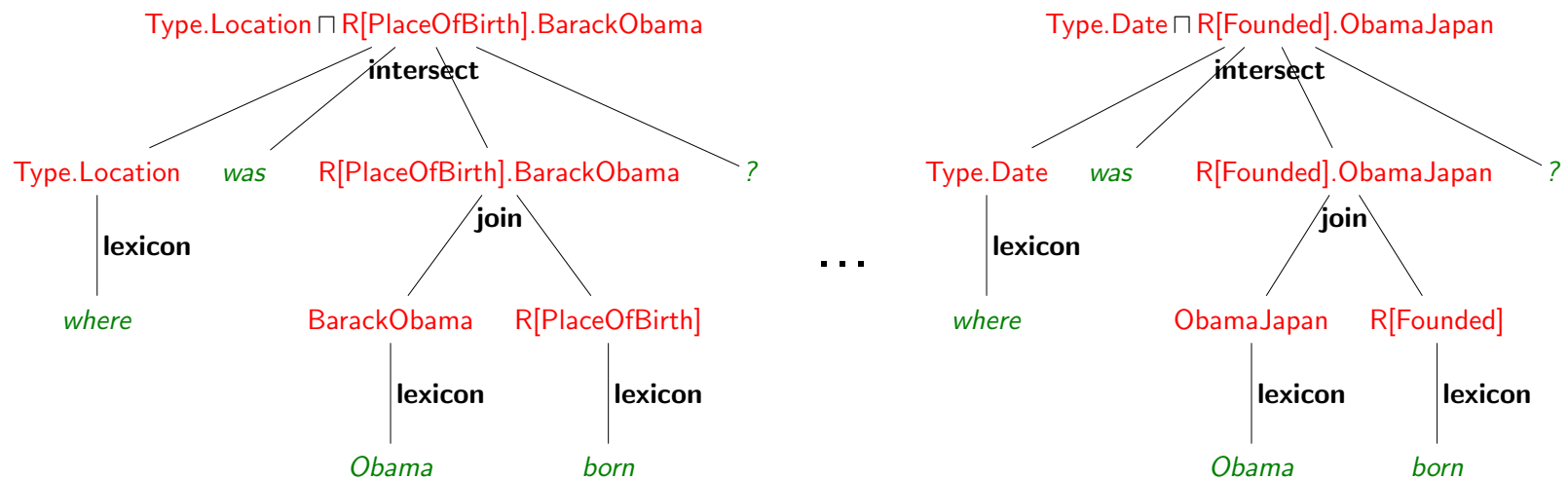
Where was Obama born?



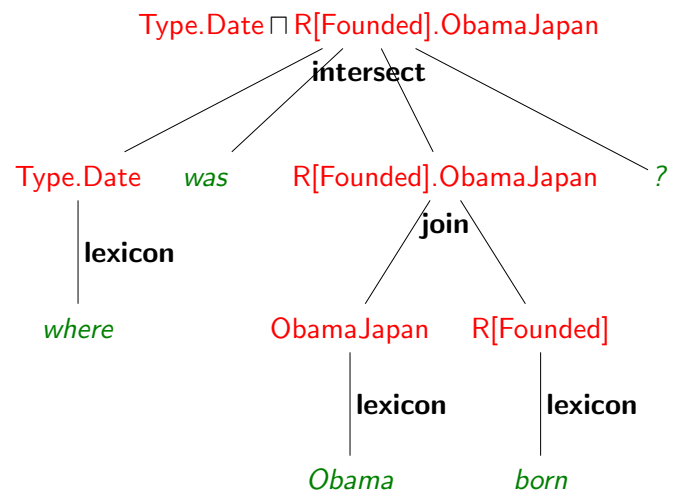
A Really Dumb Grammar

- (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*

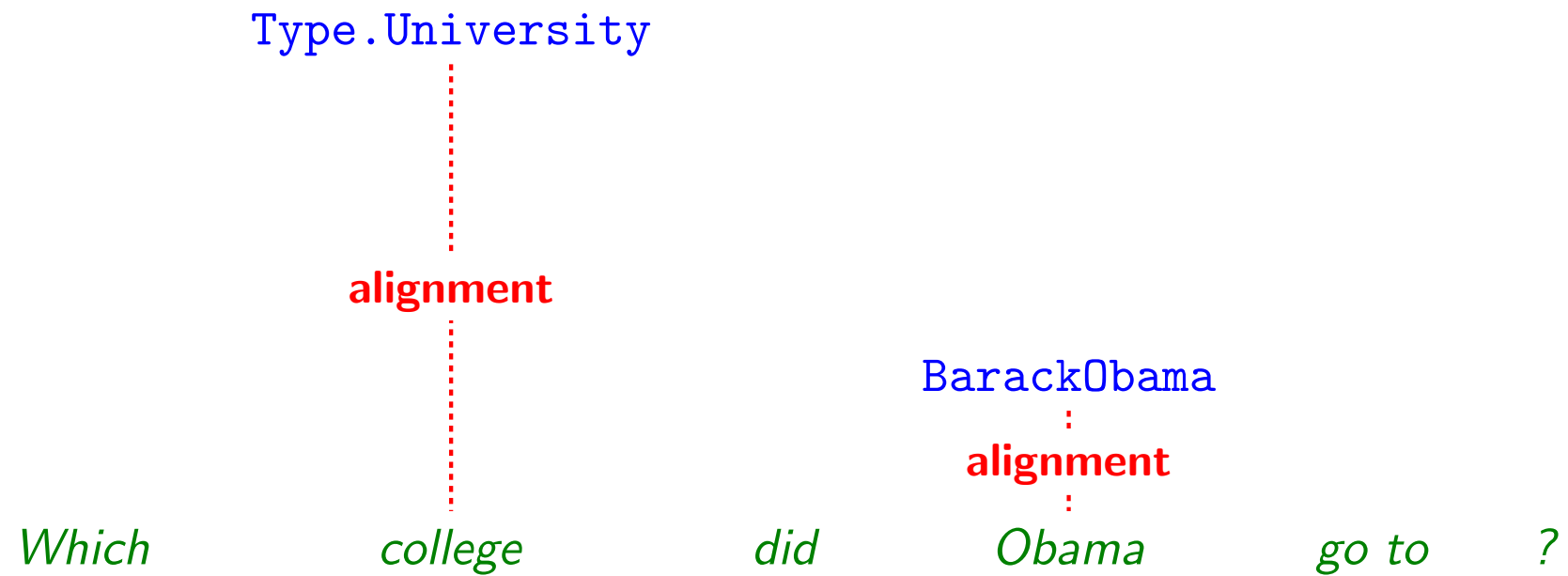
set of candidate derivations $\mathcal{D}(x)$



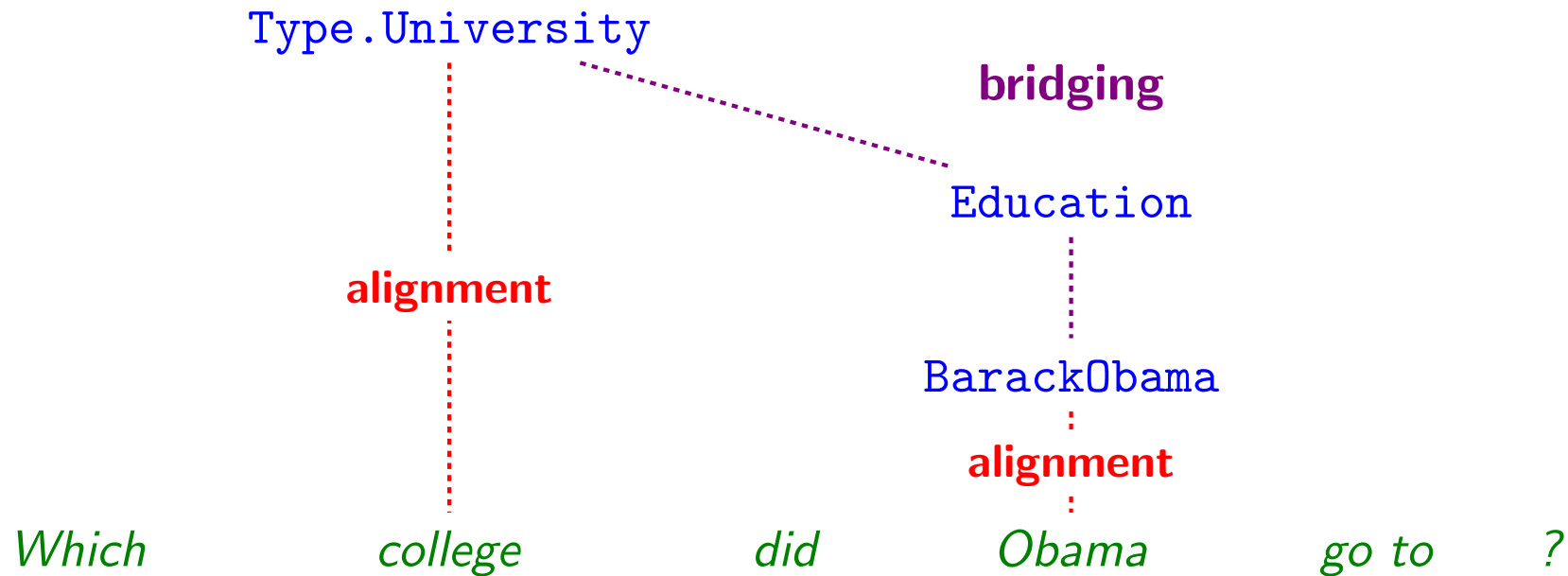
...



Bridging

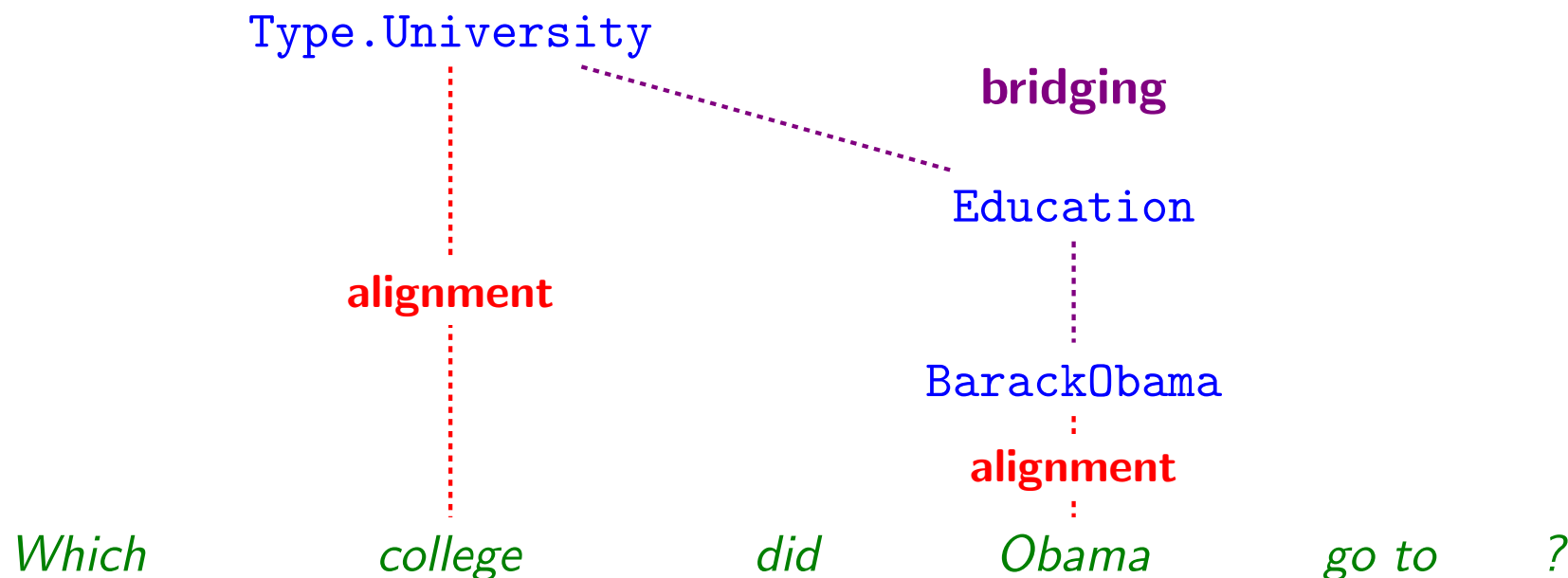


Bridging



Bridging: use neighboring predicates / type constraints

Bridging



Bridging: use neighboring predicates / type constraints

Start building from parts with more certainty

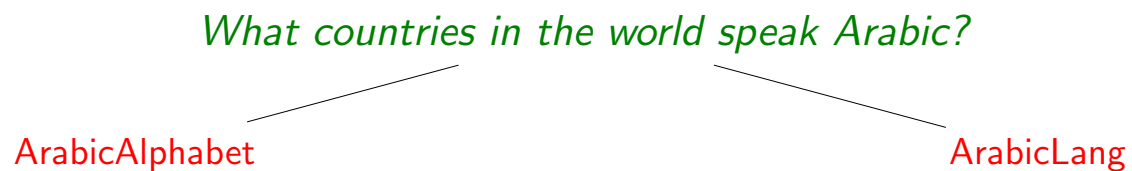
Bridging to nowhere

Search logical forms based on "prior":

What countries in the world speak Arabic?

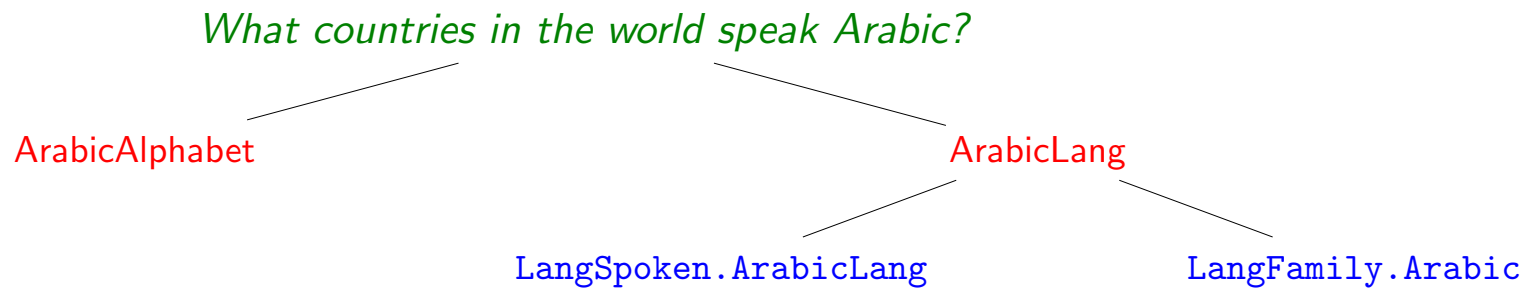
Bridging to nowhere

Search logical forms based on "prior":



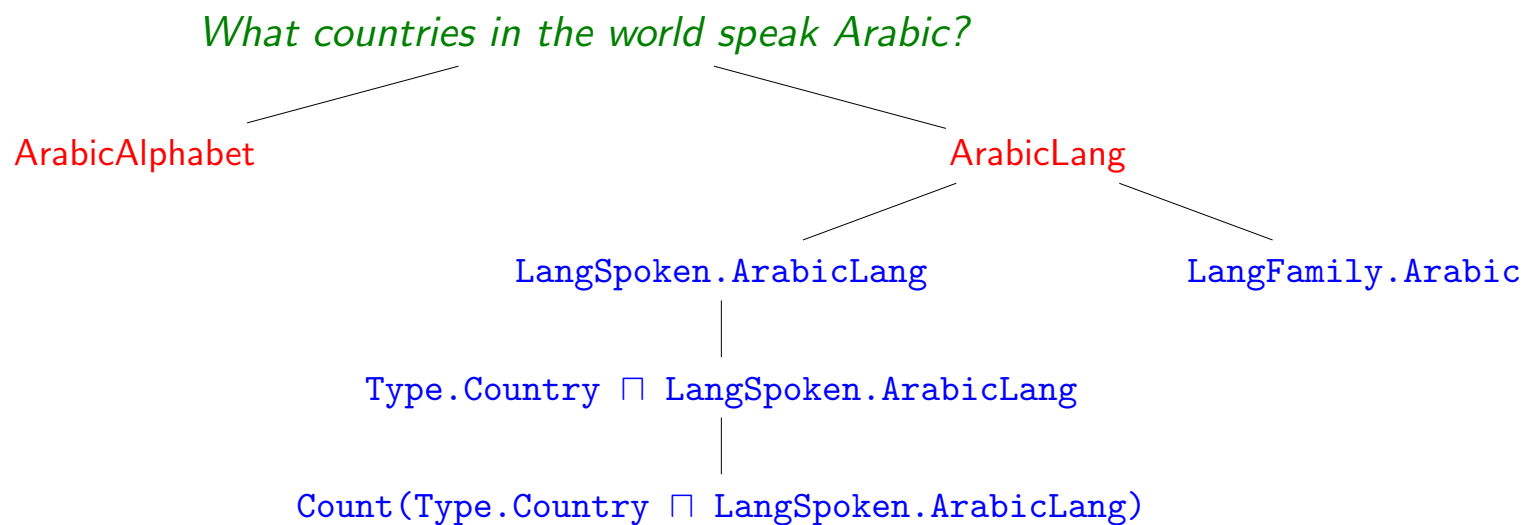
Bridging to nowhere

Search logical forms based on "prior":



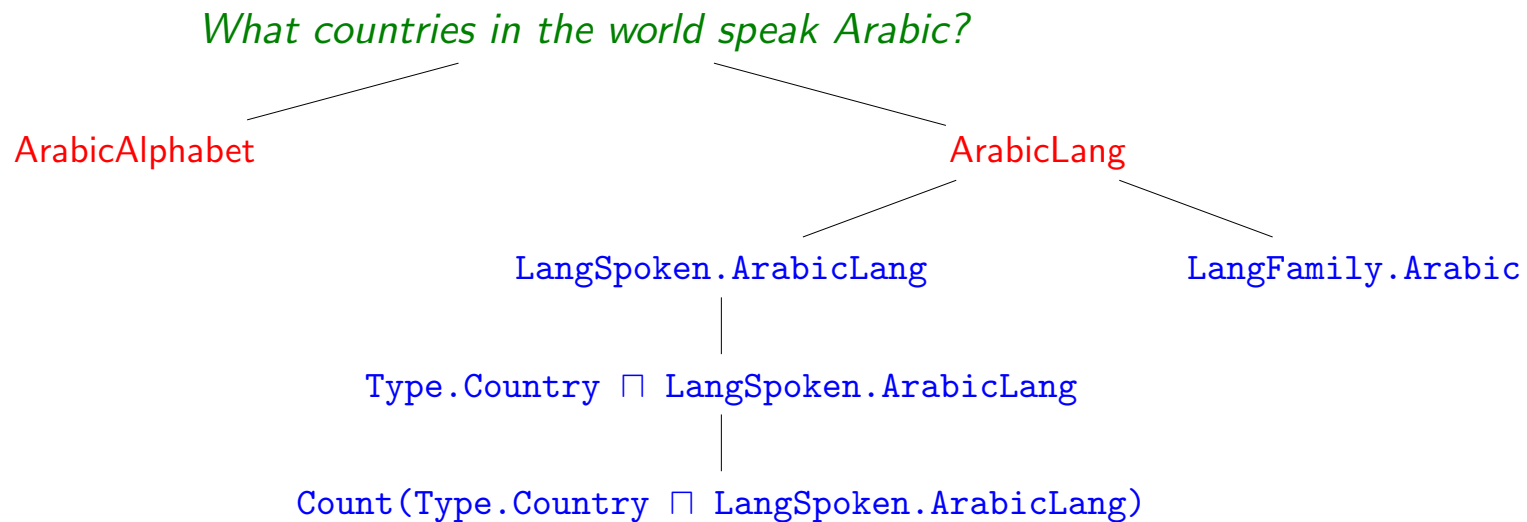
Bridging to nowhere

Search logical forms based on "prior":



Bridging to nowhere

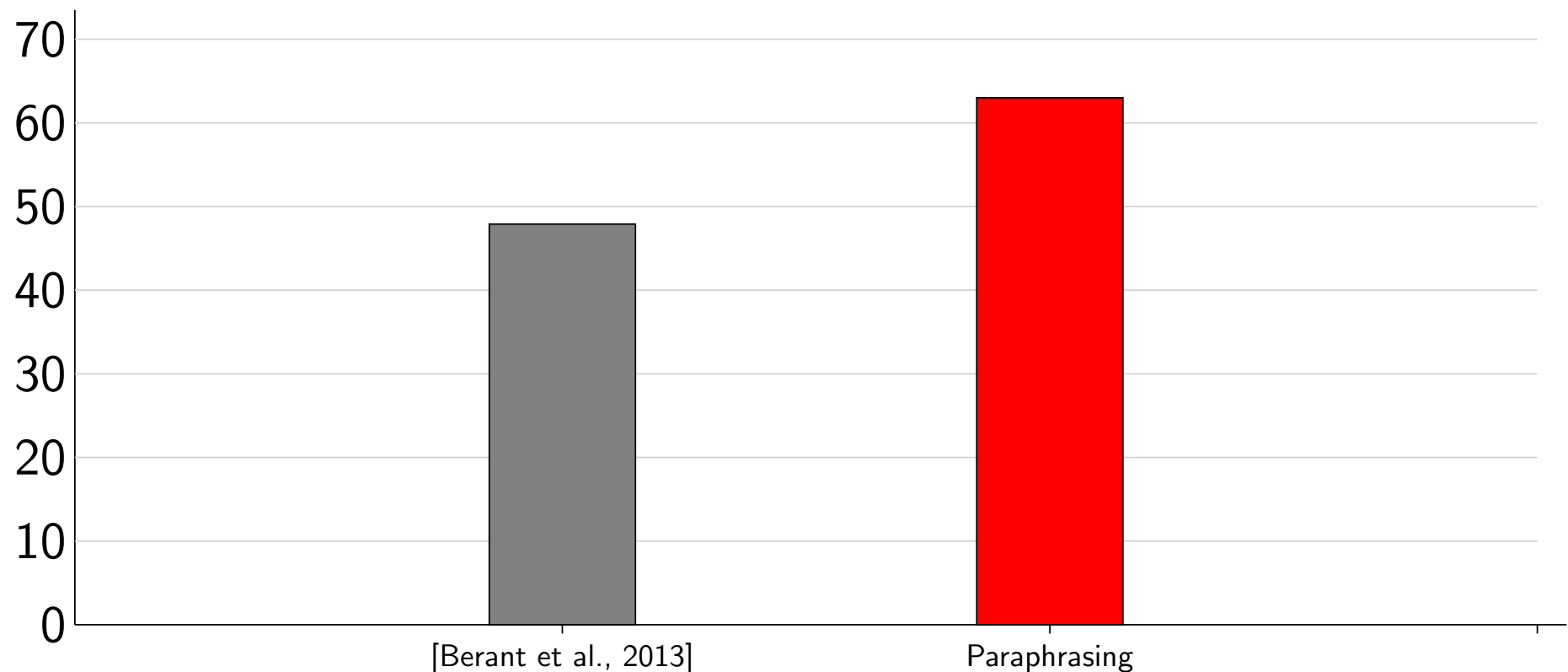
Search logical forms based on "prior":



Start building from parts with more certainty

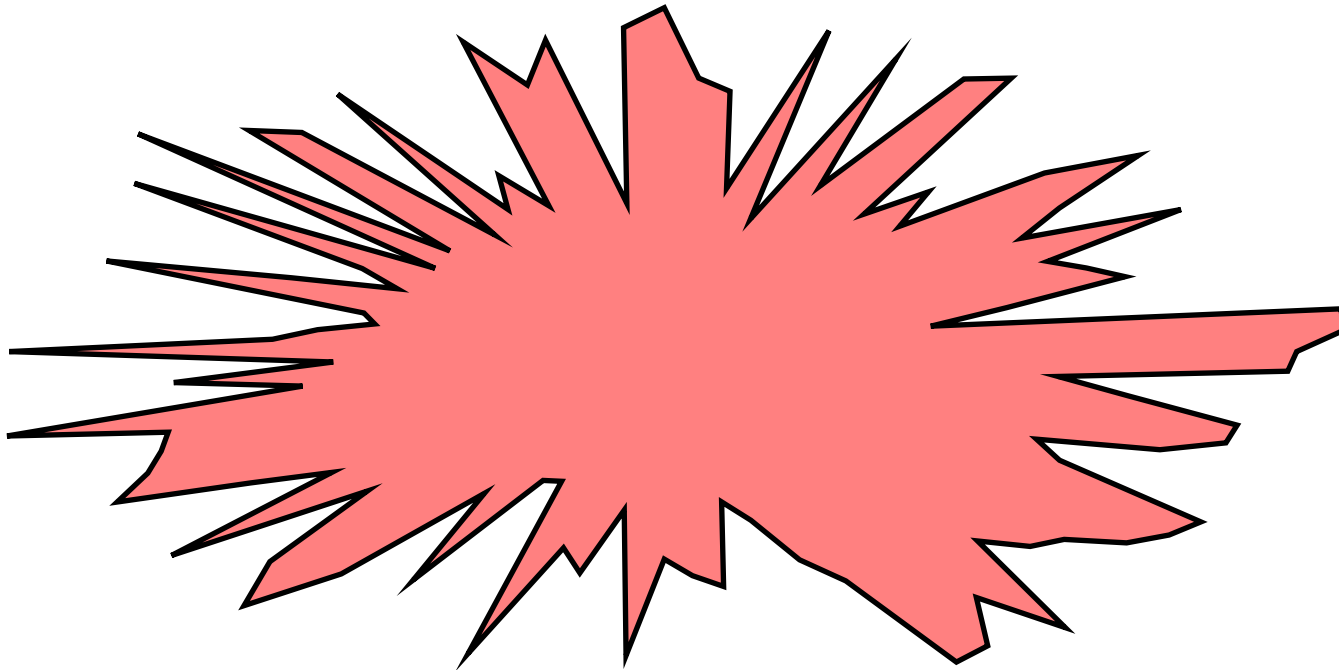
Oracle on WebQuestions

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



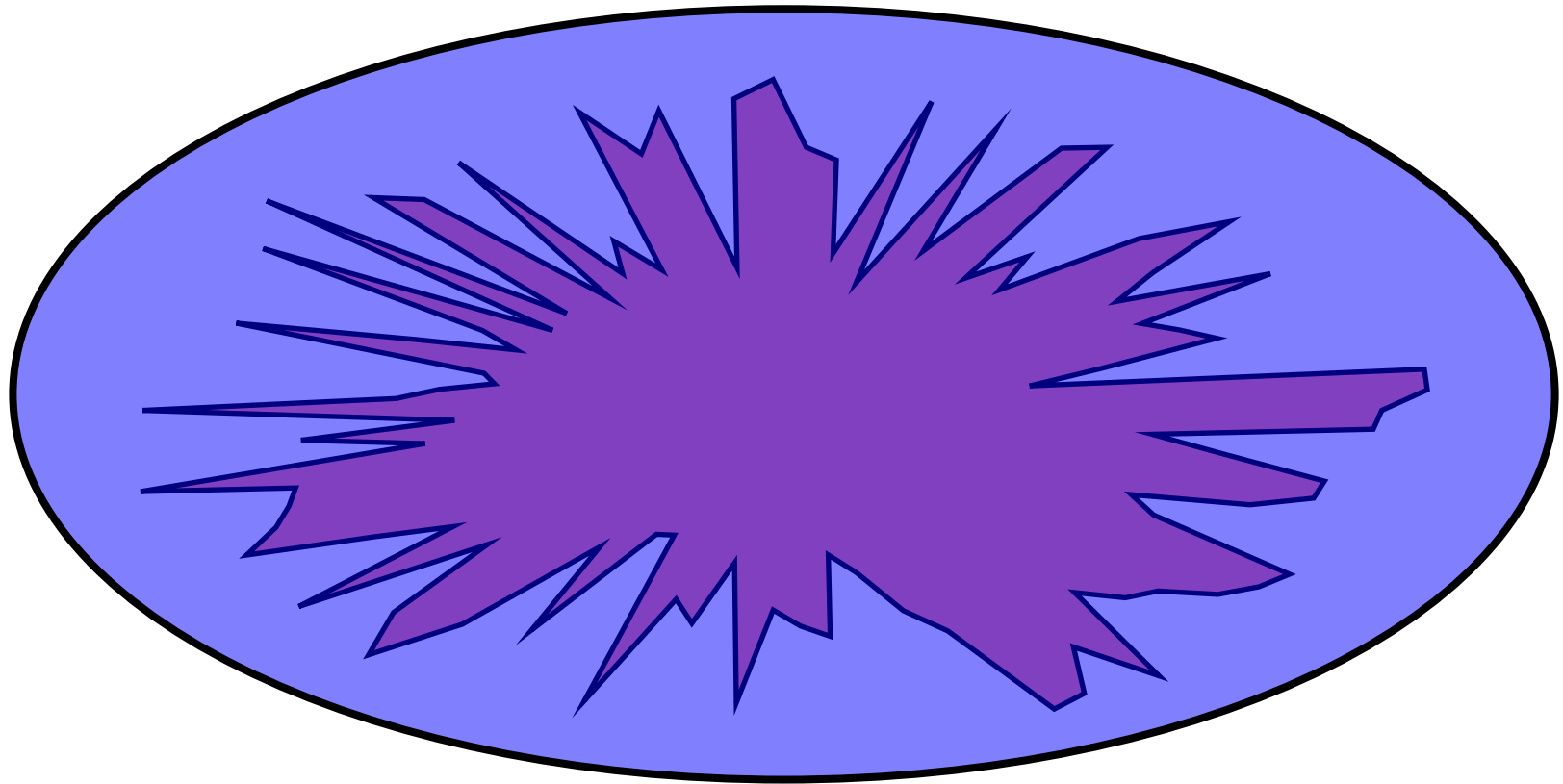
Overapproximation via simple grammars

- Modeling correct derivations requires complex rules



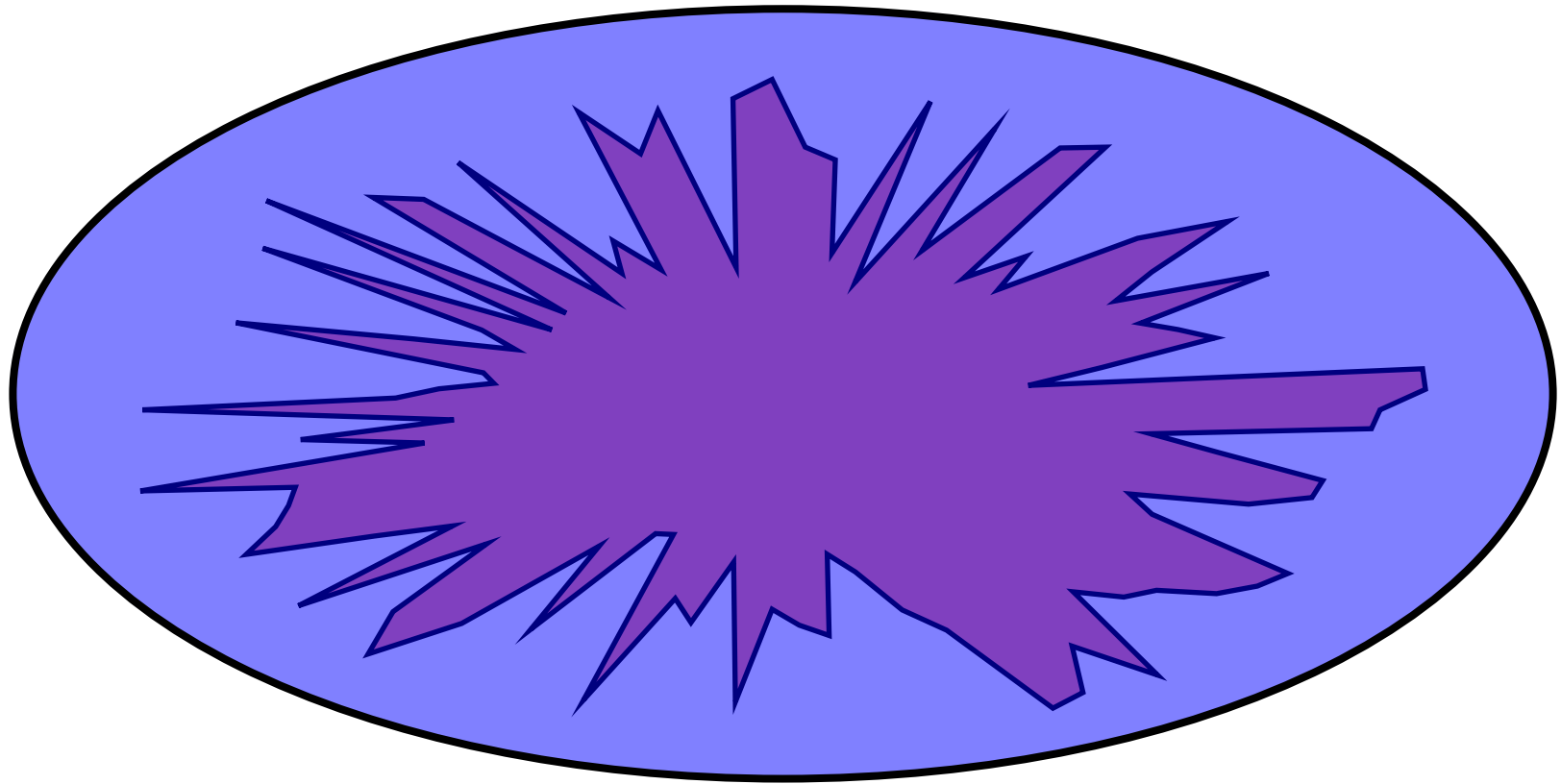
Overapproximation via simple grammars

- 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 \Rightarrow soft/overlapping features

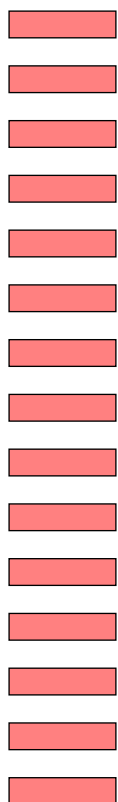
Outline

- 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

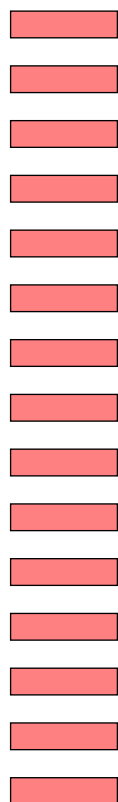
Iteration 1

Example 1



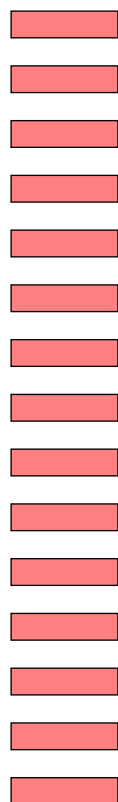
...

Example 2



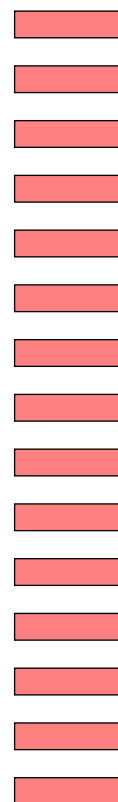
...

Example 3



...

Example 4



...

Example 5

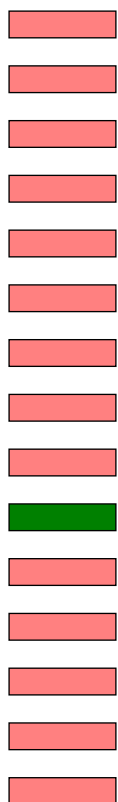


...

Bootstrapping from easy examples

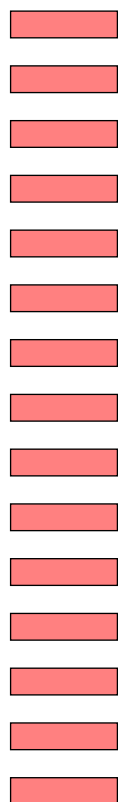
Iteration 2

Example 1



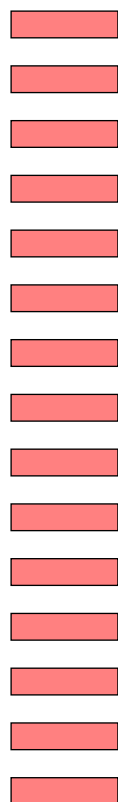
...

Example 2



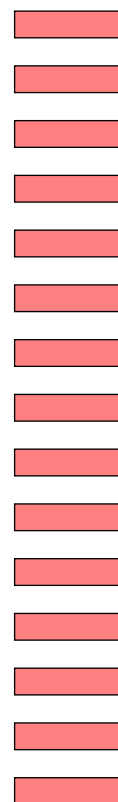
...

Example 3



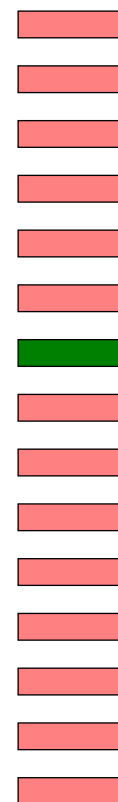
...

Example 4



...

Example 5

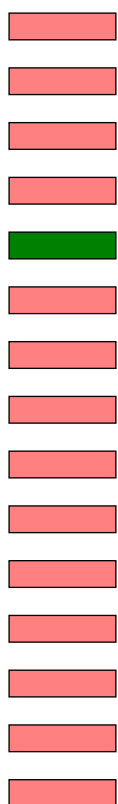


...

Bootstrapping from easy examples

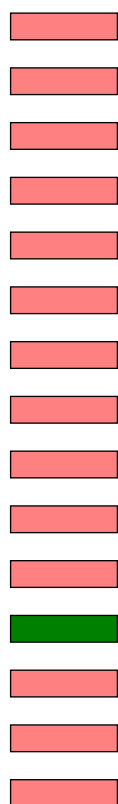
Iteration 3

Example 1



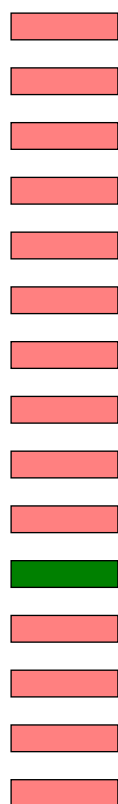
...

Example 2



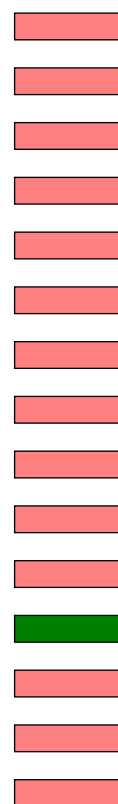
...

Example 3



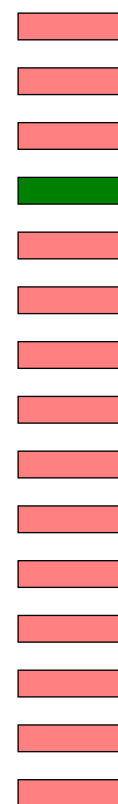
...

Example 4



...

Example 5

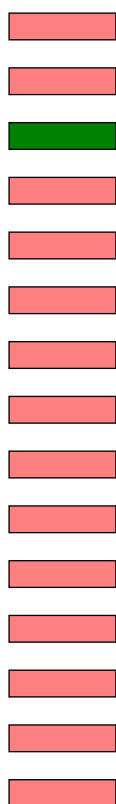


...

Bootstrapping from easy examples

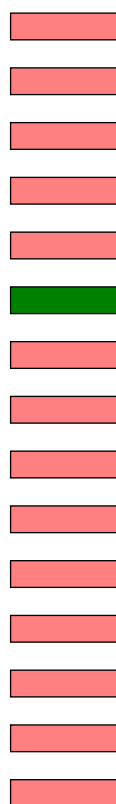
Iteration 4

Example 1



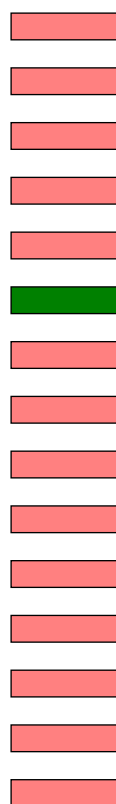
...

Example 2



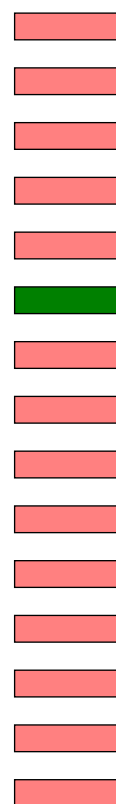
...

Example 3



...

Example 4



...

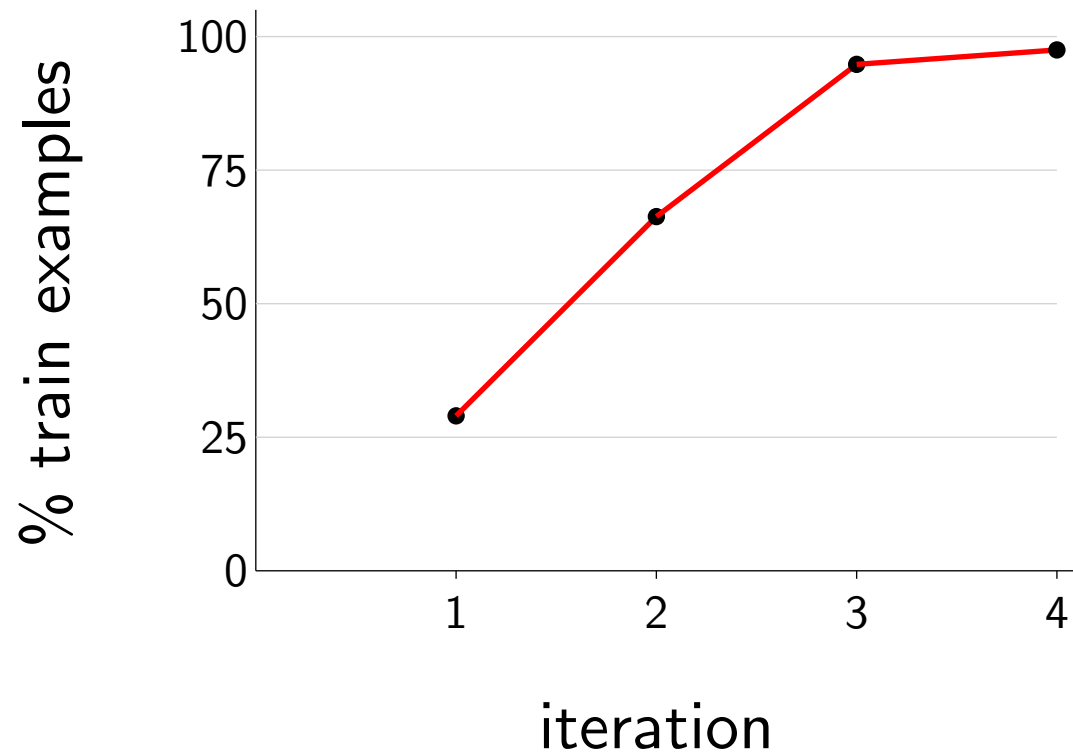
Example 5



...

Bootstrapping from easy examples

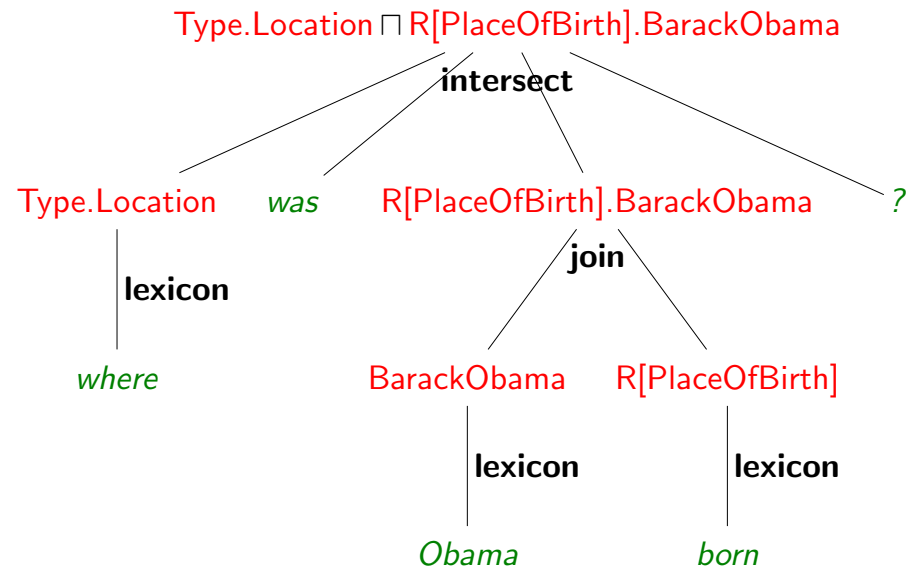
On GeoQuery [Liang et al., 2011]:



Outline

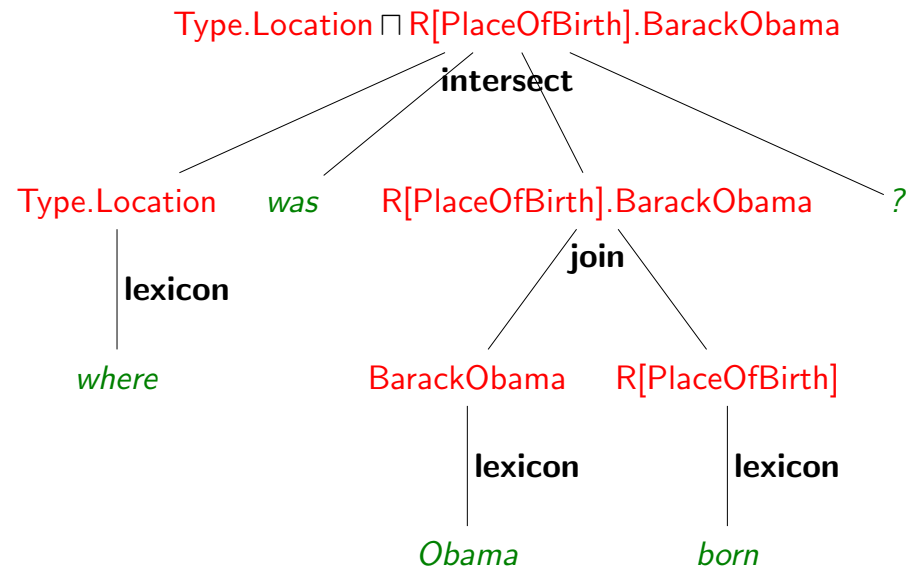
- Semantic parsing in 5 minutes
- A closer look at the elements
 - Knowledge base incompleteness
 - Lexical coverage
 - Search over logical forms
 - Learning via bootstrapping
 - **Leveraging denotations ("grounding")**
 - Datasets
- Final remarks

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
<i>born</i> maps to PlaceOfBirth	1
<i>born</i> 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]`

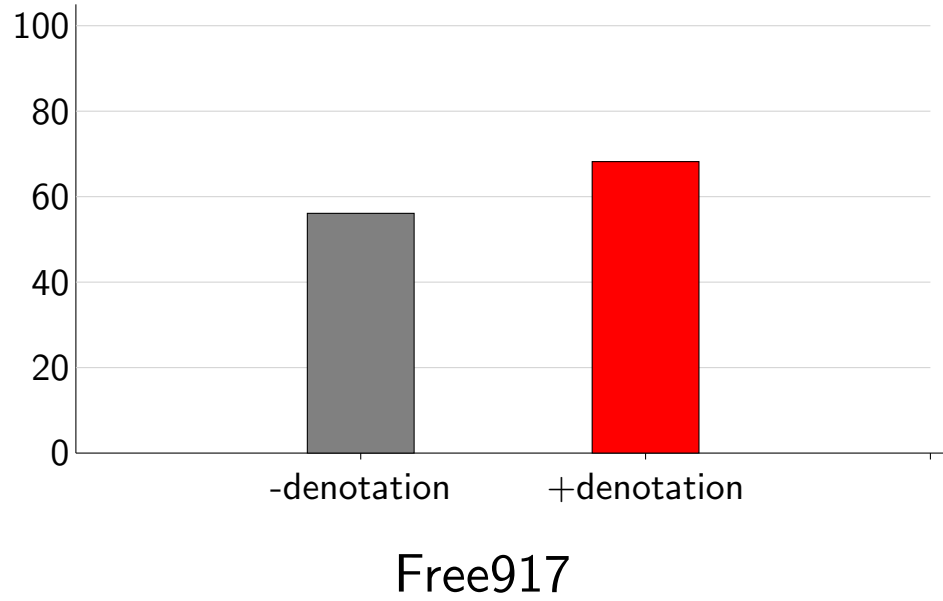
hiking trails near Baltimore
Avalon Super Loop
Patapsco Valley State Park
Gunpowder Falls State Park
Rachel Carson Conservation Park
Union Mills Hike
...

`/html[1]/body[1]/div[2]/a`

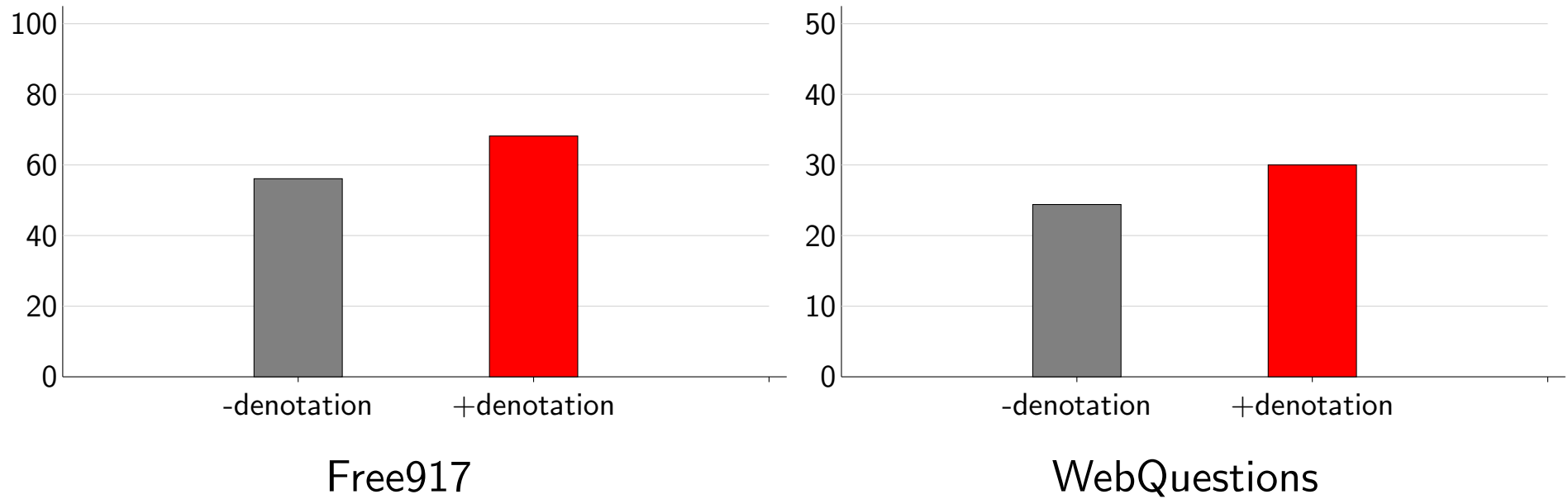


hiking trails near Baltimore
Home
About Baltimore Tour
Pricing
Contact
Online Support
...

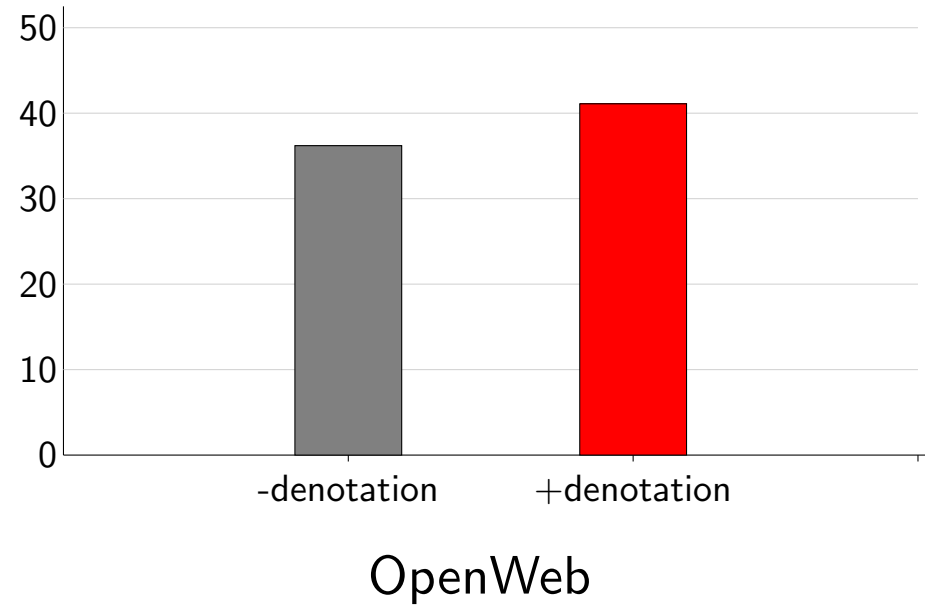
Impact of denotation features



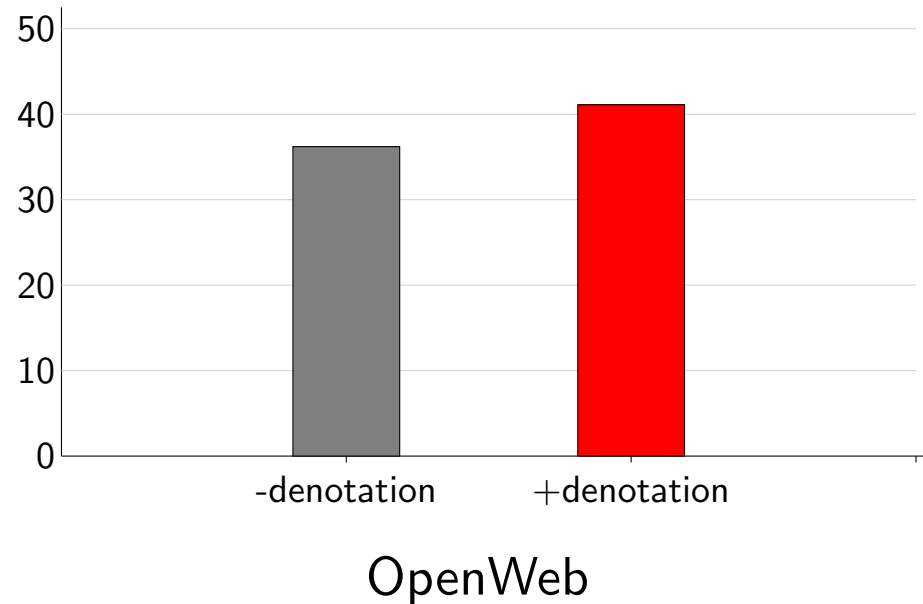
Impact of denotation features



Impact of denotation features



Impact of denotation features



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?  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? $\xrightarrow{\text{Google Suggest}}$ Barack Obama
Lady Gaga
Steve Jobs

Where was *Steve Jobs* born?

Where was *Steve Jobs* *_*? $\xrightarrow{\text{Google Suggest}}$ born
raised
on the Forbes list

Where was *Steve Jobs* raised?

...

AMT annotation \Rightarrow 6.6K question/answer pairs

Question answering on

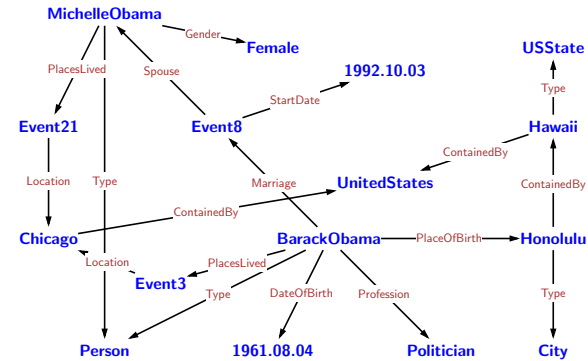


WebQuestions dataset (6K questions) [Berant et al., 2013]

what did obama study in school

where to fly into bali

what was tupac name in juice



Question answering on

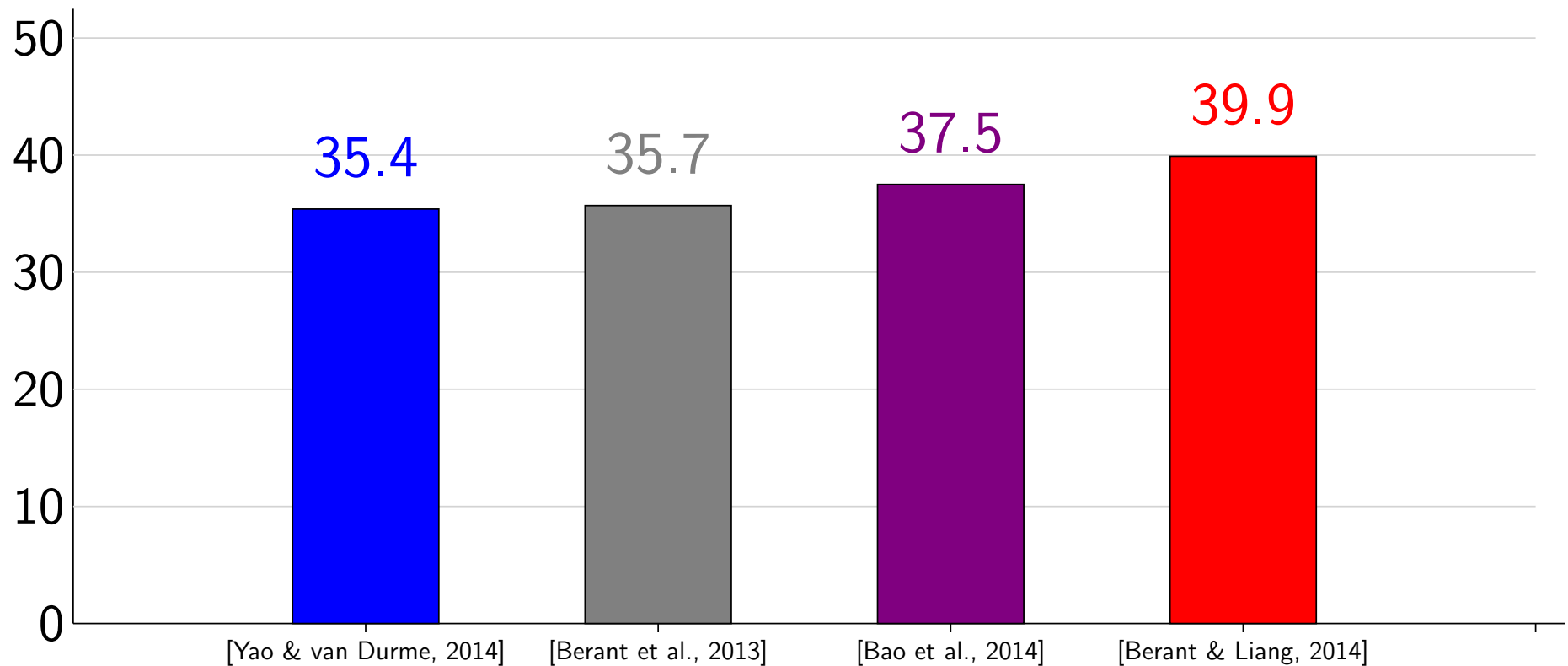
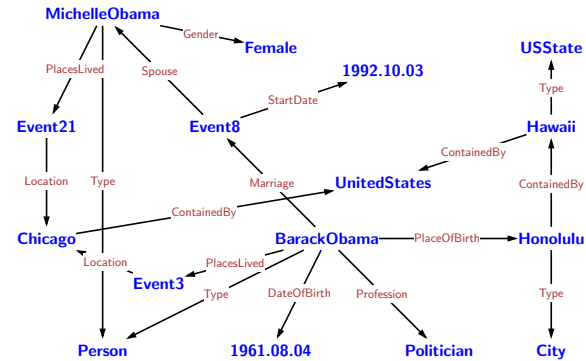


WebQuestions dataset (6K questions) [Berant et al., 2013]

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

Ranker find a list or topic

create a list recent people film tv music sports travel

home > places/travel > airlines

All Italian Airlines

List of Airlines 6,395 votes 19 items

A list of all Italian Airlines. This list focuses on passenger airlines including international, interco... TSA and the airports themselves have taken most of the headlines recently, airline companies h... 9/11. (And yet, the airlines survive.) While air travel was once seen as a symbol of luxury and we

Rank	Name	Hubs
1	Air Dolomiti	Munich Airport, Verona Villafranca Airport, Tre Venezie Glisè Airport
2	Air Europe	Malpensa Airport
3	Air Italy	Malpensa Airport, Leonardo da Vinci-Fiumi... Verona Villafranca Airport
4	Air One	Malpensa Airport
5	Air Vallée	Prima Airport, Aosta Airport, Federico Fellini International Airport
6	Alidaunia	Foggia "Gino Lisa" Airport
7	Alitalia-Linee Aeree Italiane	Malpensa Airport, Leonardo da Vinci-Fli... Airport

airlines of italy

10. Greenhouse Effect

Greenhouse effect is the process in which the atmosphere of the Earth trap some of the heat coming from the sun, making the Earth warm but due to burning fuels, cutting trees, the concentration of heat on Earth is increased to abnormal levels making greenhouse effect as one of the major causes of global warming. Carbon Dioxide, methane, nitrous oxide are the greenhouse gases which helps to keep the Earth warm. It is a natural phenomenon that takes place with the adequate concentrations of the greenhouse gases. But when the concentration of these gases rises, they disturb the climatic conditions, making the Earth more warm. These gases are not able to escape, which is the cause of world-wide increase in temperature. So the balance of carbon dioxide and other gases should be maintained so that it does not become the major reason of global warming.

9. Air Pollution

The harmful gases emitted from the vehicles and factories and the greenhouse gases cause pollution in the air and these gases not remain in the atmosphere. The smoke, rather, in the atmosphere forms a thick full of harmful

natural causes of global warming

TEAMS SCHEDULES TICKETS FAN ZONE VIDEO + AUDIO FACILITIES DEPARTMENTS CONNECT VISITORS

Football

NEWS SCHEDULES ROSTERS COACHES STATISTICS

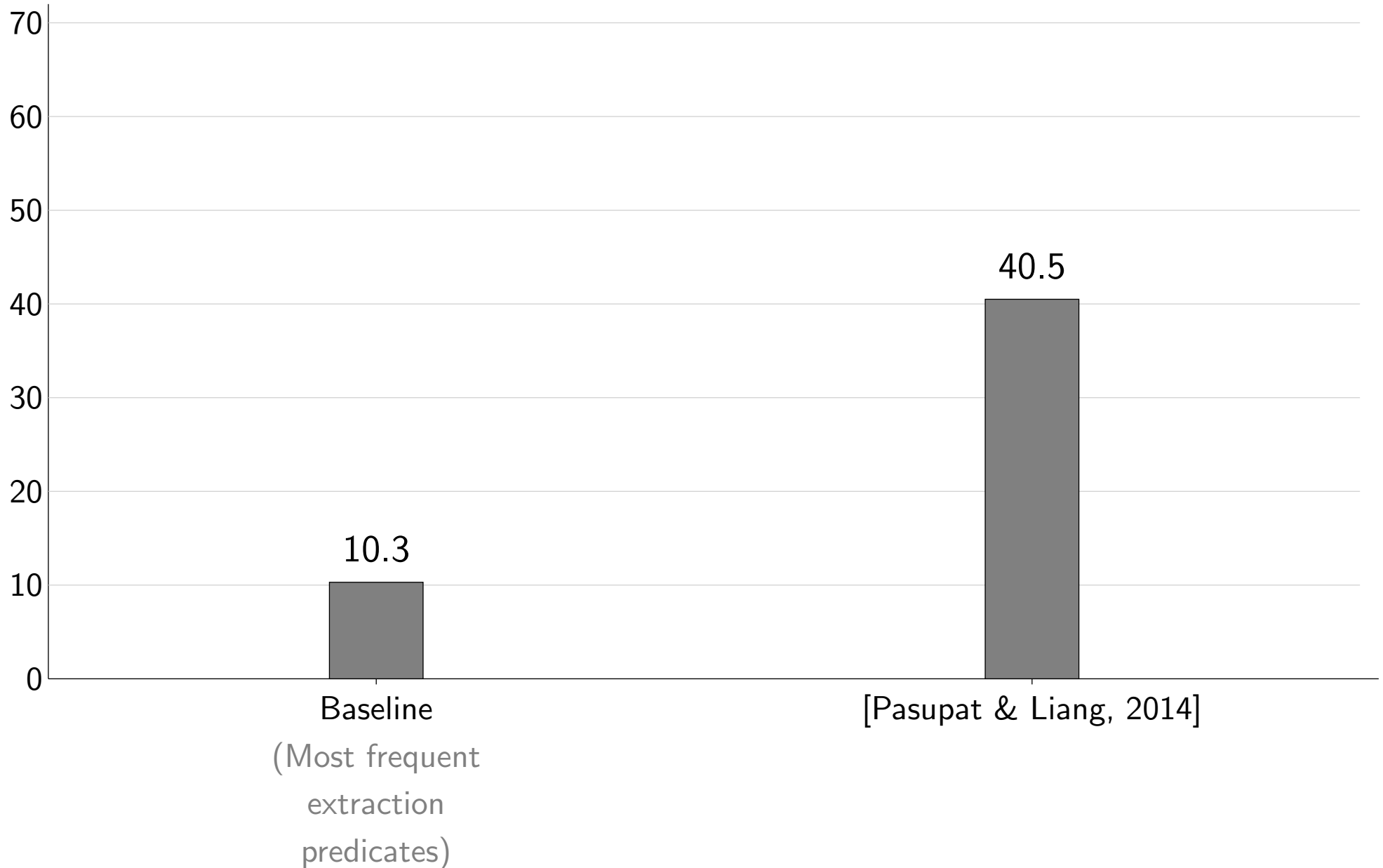
Season 2013-14

2013 Football Coaches

 Les Miles Head Coach	 Cam Cameron Offense Coordinator/Quarterbacks Coach	 John Chavis Defensive Coordinator	 Frank Wilson Running Backs Coach/Receiving Coordinator
 Steve Ensminger Tight Ends Coach	 Brick Haley Defensive Line Coach	 Adam Henry Wide Receivers Coach	
 Thomas McGaughey Special Teams Coordinator	 Corey Raymond Defensive Backs Coach	 Greg Studrawa Offensive Line Coach	
 Steve Kragthorpe Administrator	 Tommy Moffitt Strength & Conditioning Coordinator	 Dr. Sam Nader Assistant Athletics Director - Football	

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?

The image displays four overlapping screenshots from different datasets, illustrating the concept of a new dataset for compositional and open-domain tasks.

- ACCEPTED LONG PAPERS ACL 2014:** A list of research papers, including "A Bayesian Mixed Effects Model of Literary Character" and "A Decision Theoretic Approach to Natural Language Generation".
- Most Popular Action Feature Films:** A list of movies such as "Godzilla (2014)", "X-Men: Days of Future Past", "The Amazing Spider-Man", and "Transformers: Age of Extinction".
- President Took office:** A table showing the names and inauguration dates of US Presidents.
- Thursday 12 June:** A flight schedule for Thursday, June 12, showing flights from Brazil to Croatia, Mexico to Cameroon, Spain to Netherlands, and Chile to Australia.

No.	President	Took office
1	 George Washington (1732-1799) (1789-1800)	April 30, 1789 (1789)
2	 John Adams (1735-1826) (1797-1801)	March 4, 1797 (1797)
3	 Thomas Jefferson (1743-1826) (1801-1809)	March 4, 1801 (1801)
4	 James Madison (1751-1836) (1809-1817)	March 4, 1809 (1809)

Group A	Group B	Group C
 BRAZIL	17:00	 CROATIA
 MEXICO	13:00	 CAMEROON
 SPAIN	16:00	 NETHERLANDS
 CHILE	18:00	 AUSTRALIA
 COLOMBIA	13:00	 GREECE
 URUGUAY	16:00	 COSTA RICA

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

No	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)	Commander-in-Chief of the Continental Army (1775–1783)	John Adams
					2 (1792)		
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)	Vice President	Aaron Burr March 4, 1801 – March 4, 1805
					5 (1804)		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)	Secretary of State (1801–1809)	Vacant ^[n 5] April 20, 1812 – March 4, 1813
					7 (1812)		Ebbridge Gerry ^[n 4] March 4, 1813 – November 23, 1814
					8 (1816)		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	9 (1820)	Secretary of State (1811–1817)	Daniel D. Tompkins
					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]

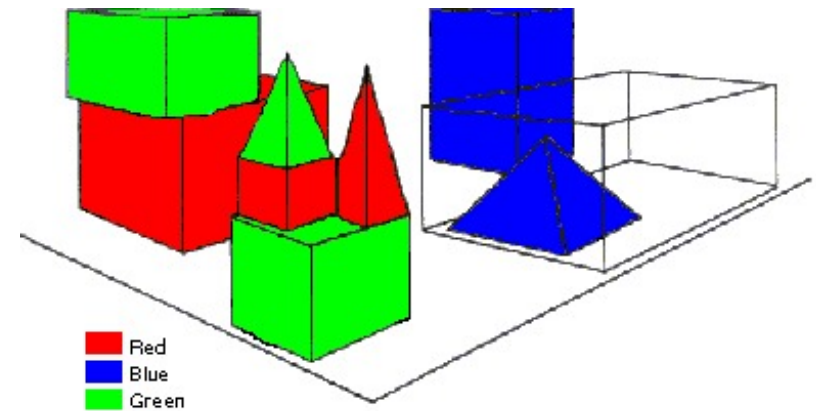
Understanding visual world [Matuszek et. al; 2012, Krishnamurthy & Kollar, 2013]

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**

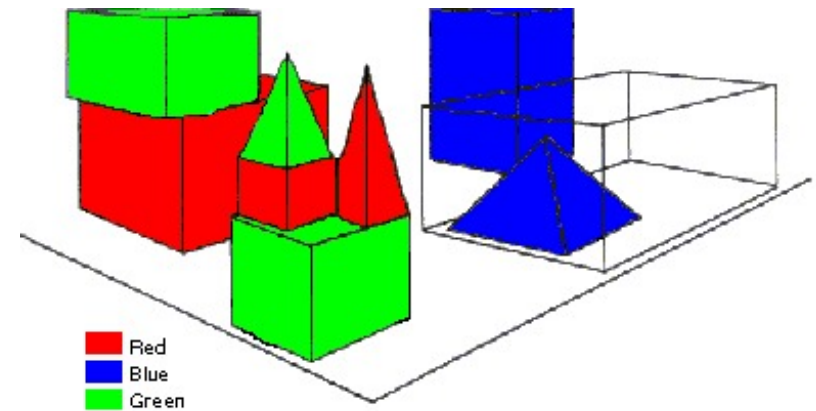
SHRDLU [1971]



SHRDLU [1971]

Person: Pick up a big red block.

Computer: OK.



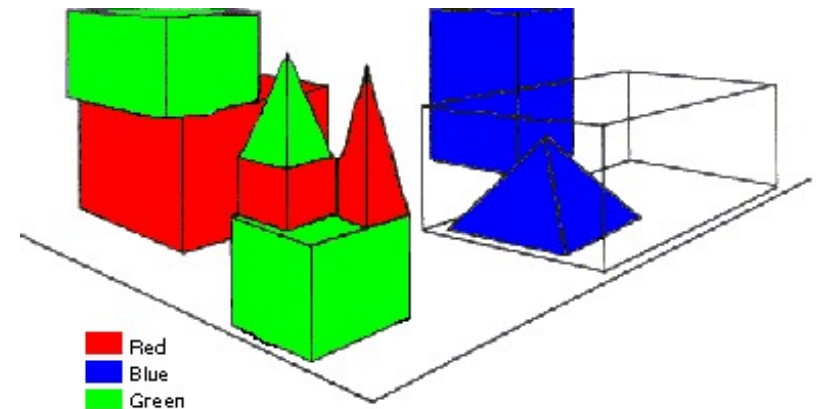
SHRDLU [1971]

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.

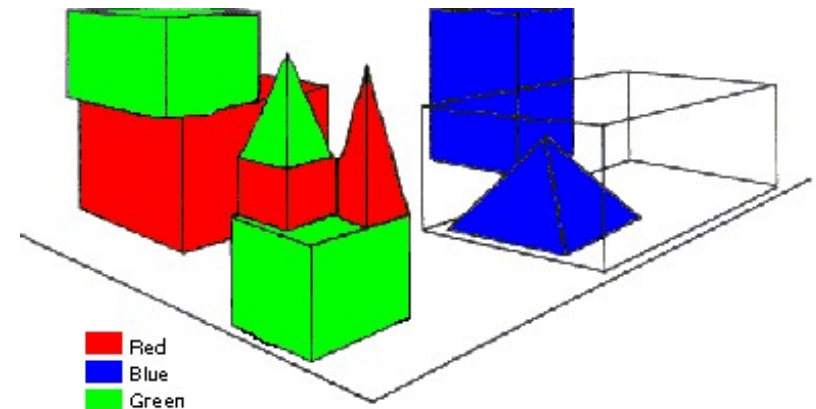
Computer: OK.

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.

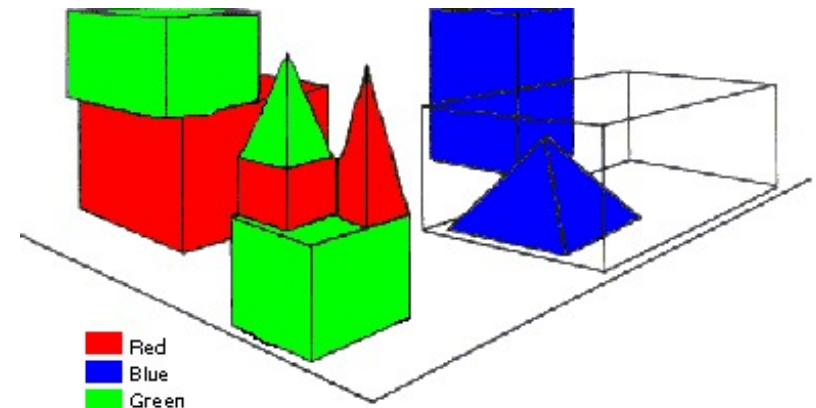
Computer: OK.

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

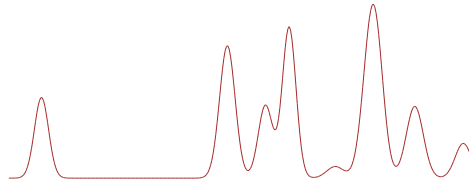
Takeaway 2/3: learning and search

Challenging search / learning problem

Takeaway 2/3: learning and search

Challenging search / learning problem

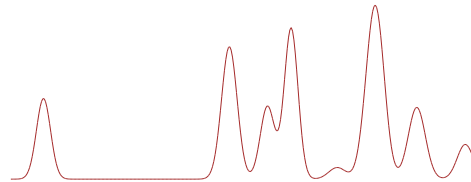
 **Non-convex optimization**



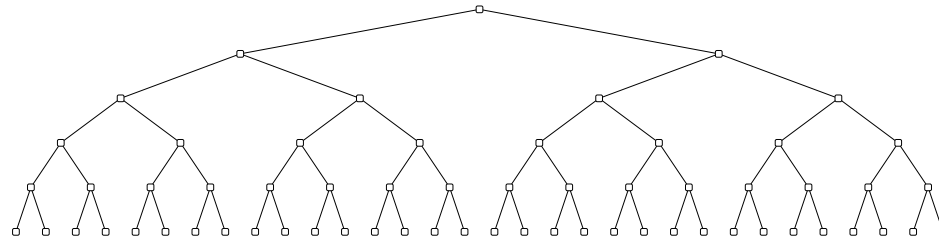
Takeaway 2/3: learning and search

Challenging search / learning problem

 **Non-convex optimization**



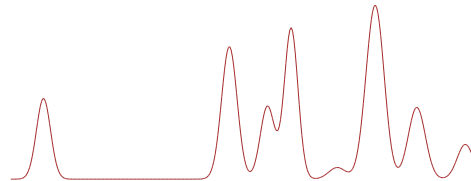
 **Exponential search space**



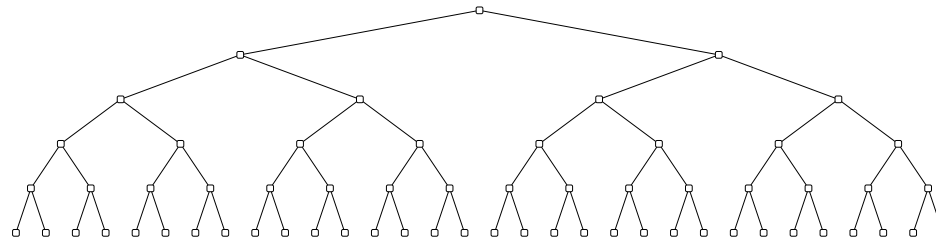
Takeaway 2/3: learning and search

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 period¹. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic-Cambrian fossil record with geochemical and physical environmental perturbations²⁻⁴, but the mechanisms responsible for those perturbations remains uncertain^{5,6}. Here we use new stratigraphic and geochemical data to show that early Paleozoic marine sediments deposited approximately 540-480 Myr ago record 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, regoliths 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 Unconformity⁷. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this paleogeographic 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 Supergroup⁸. 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 Goodwin⁹, Bahig¹⁰, Avdeyeva¹¹ and Siberia¹², 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 skeletal record of crown-group animals, a fact that some paleontologists have interpreted as evidence for stratigraphic hiatus and an incomplete record of early animal evolution¹³.

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 Saak Sequence^{14,15} that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the paleocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Saak Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences^{16,17}, but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth's surface rest nonconformably 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 substantial 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'¹⁸, 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 interiors¹⁹ was eroded and mobilized during the transgression, thereby expositing 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^{20,21}, and

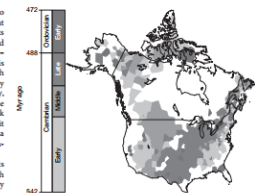


Figure 1 | Saak Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology 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/sempr/>

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

Code and data online

<http://www-nlp.stanford.edu/software/sempr/>

<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/sempr/>

<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!