Learning Semantic Correspondences with Less Supervision

ACL 2009 – Singapore

Percy Liang Michael I. Jordan Dan Klein

Grounded semantics

Grounded semantics

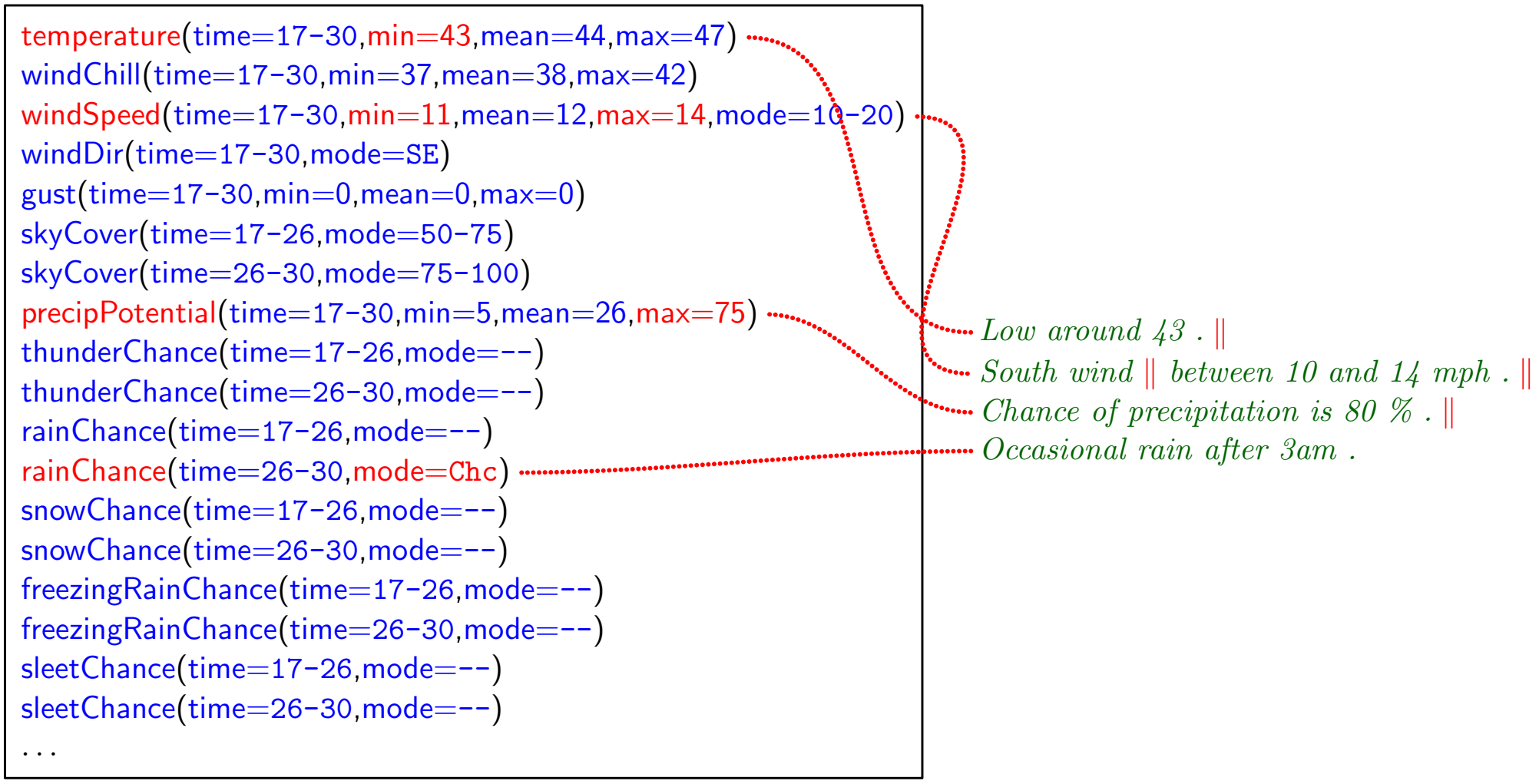
Low around 43 .
South wind between 10 and 14 mph .
Chance of precipitation is 80 % .
Occasional rain after 3am .

Grounded semantics

temperature(time=17-30,min=43,mean=44,max=47)
windChill(time=17-30,min=37,mean=38,max=42)
windSpeed(time=17-30,min=11,mean=12,max=14,mode=10-20)
windDir(time=17-30,mode=SE)
gust(time=17-30,min=0,mean=0,max=0)
skyCover(time=17-26,mode=50-75)
skyCover(time=26-30,mode=75-100)
precipPotential(time=17-30,min=5,mean=26,max=75)
thunderChance(time=17-26,mode=--)
thunderChance(time=26-30,mode=--)
rainChance(time=17-26,mode=--)
rainChance(time=26-30,mode=Chc)
snowChance(time=17-26,mode=--)
snowChance(time=26-30,mode=--)
freezingRainChance(time=17-26,mode=--)
freezingRainChance(time=26-30,mode=--)
sleetChance(time=17-26,mode=--)
sleetChance(time=26-30,mode=--)
...

Low around 43 .
South wind between 10 and 14 mph .
Chance of precipitation is 80 % .
Occasional rain after 3am .

Grounded semantics



Big question:

Which (tiny) part of the world is referenced by the text?

Related work

Supervised [Ge & Mooney, 2005; Zettlemoyer & Collins, 2005; Lu et al., 2008]

$\lambda x.\mathsf{flight}(x) \wedge \mathsf{dest}(x, \mathtt{Boston})$ ..... *Show me all flights to Boston*

Related work

Supervised [Ge & Mooney, 2005; Zettlemoyer & Collins, 2005; Lu et al., 2008]

$\lambda x.\text{flight}(x) \wedge \text{dest}(x, \texttt{Boston})$ ..... *Show me all flights to Boston*

Partially-supervised [Kate & Mooney, 2007; Chen & Mooney, 2008]

ballstopped()
badPass(arg1=pink11,arg2=purple3) ..... *pink11 makes a bad pass and was picked off by purple3*
kick(arg1=pink11)
turnover(arg1=pink11,arg2=purple3)

Related work

Supervised [Ge & Mooney, 2005; Zettlemoyer & Collins, 2005; Lu et al., 2008]

$\lambda x.\text{flight}(x) \wedge \text{dest}(x, \texttt{Boston})$ ..... *Show me all flights to Boston*

Partially-supervised [Kate & Mooney, 2007; Chen & Mooney, 2008]

ballstopped()
badPass(arg1=pink11,arg2=purple3) ..... *pink11 makes a bad pass and was picked off by purple3*
kick(arg1=pink11)
turnover(arg1=pink11,arg2=purple3)

Language acquisition [Siskind, 1996; Yu & Ballard, 2007; Frank et al., 2009]

How do children connect language with perception?

Related work

Supervised [Ge & Mooney, 2005; Zettlemoyer & Collins, 2005; Lu et al., 2008]

$\lambda x.\text{flight}(x) \wedge \text{dest}(x, \texttt{Boston})$ ..... *Show me all flights to Boston*

Partially-supervised [Kate & Mooney, 2007; Chen & Mooney, 2008]

ballstopped()
badPass(arg1=pink11,arg2=purple3) ..... *pink11 makes a bad pass and was picked off by purple3*
kick(arg1=pink11)
turnover(arg1=pink11,arg2=purple3)

Language acquisition [Siskind, 1996; Yu & Ballard, 2007; Frank et al., 2009]

How do children connect language with perception?

Reinforcement learning from rewards [Branavan et al., 2009]

Clustering-based [Poon & Domingos, 2009]

...

Problem statement

```
rainChance(time=26-30,mode=Chc)
temperature(time=17-30,min=43,mean=44,max=47)
windDir(time=17-30,mode=SE)
windSpeed(time=17-30,min=11,mean=12,max=14,mode=10-20)
precipPotential(time=17-30,min=5,mean=26,max=75)
windChill(time=17-30,min=37,mean=38,max=42)
skyCover(time=17-30,mode=50-75)
...
```

s: world state

set of records;
each record is a set of
(field=value) pairs

Problem statement

rainChance(time=26-30,mode=Chc)
temperature(time=17-30,min=43,mean=44,max=47)
windDir(time=17-30,mode=SE)
windSpeed(time=17-30,min=11,mean=12,max=14,mode=10-20)
precipPotential(time=17-30,min=5,mean=26,max=75)
windChill(time=17-30,min=37,mean=38,max=42)
skyCover(time=17-30,mode=50-75)
...

Occasional rain after 3am .
Low around 43 .
South wind between 10 and 14 mph .
Chance of precipitation is 80 % .

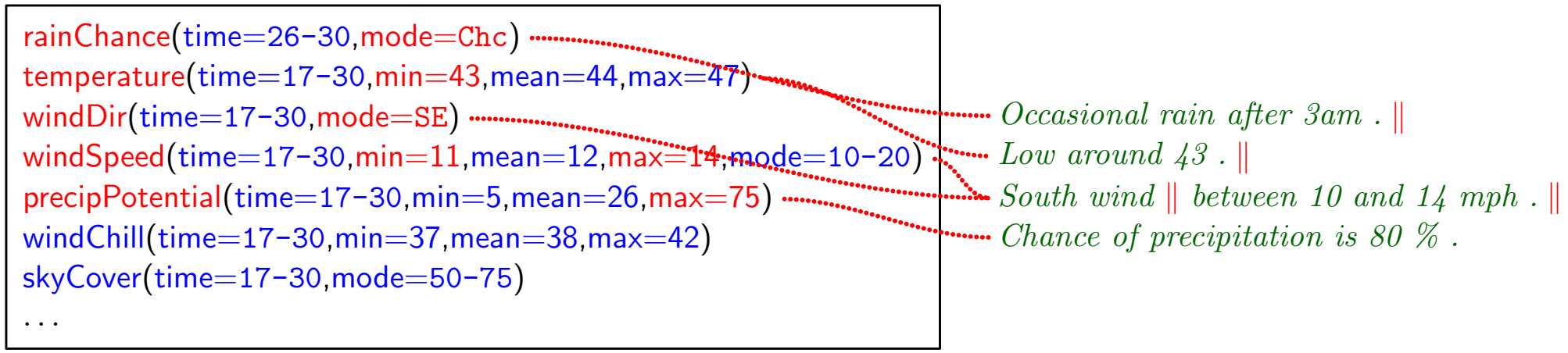
s: world state

set of records;
each record is a set of
(field=value) pairs

w: text

sequence of words

Problem statement



s: world state

set of records;
each record is a set of
(field=value) pairs

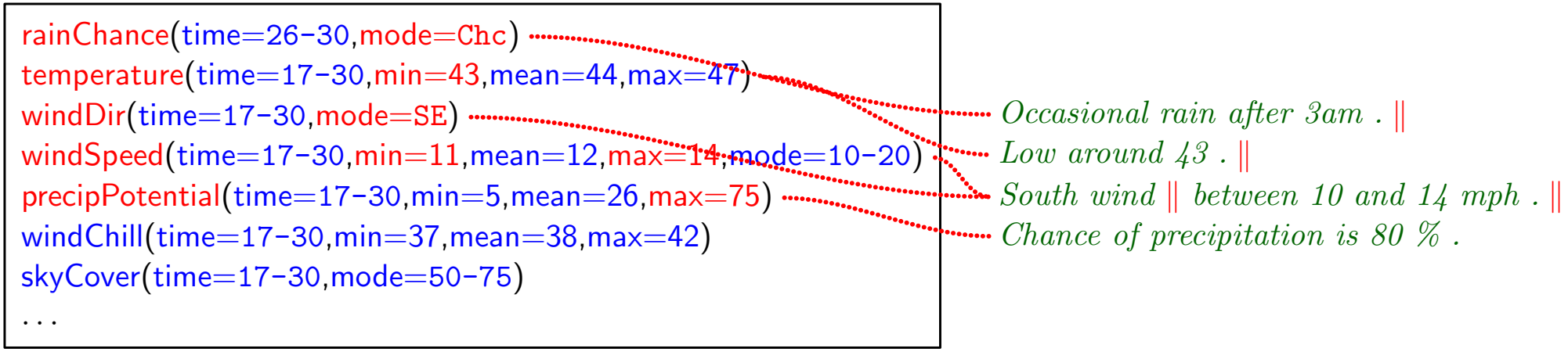
a: correspondence

segmentation of **w**;
for each segment,
which record and fields

w: text

sequence of words

Problem statement



s: world state

set of records;
each record is a set of
(field=value) pairs

a: correspondence

segmentation of **w**;
for each segment,
which record and fields

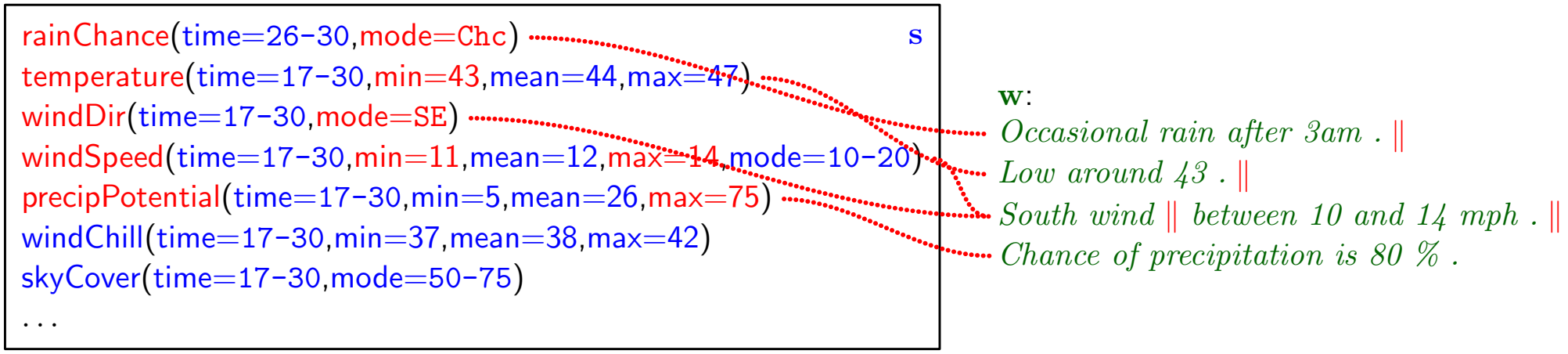
w: text

sequence of words

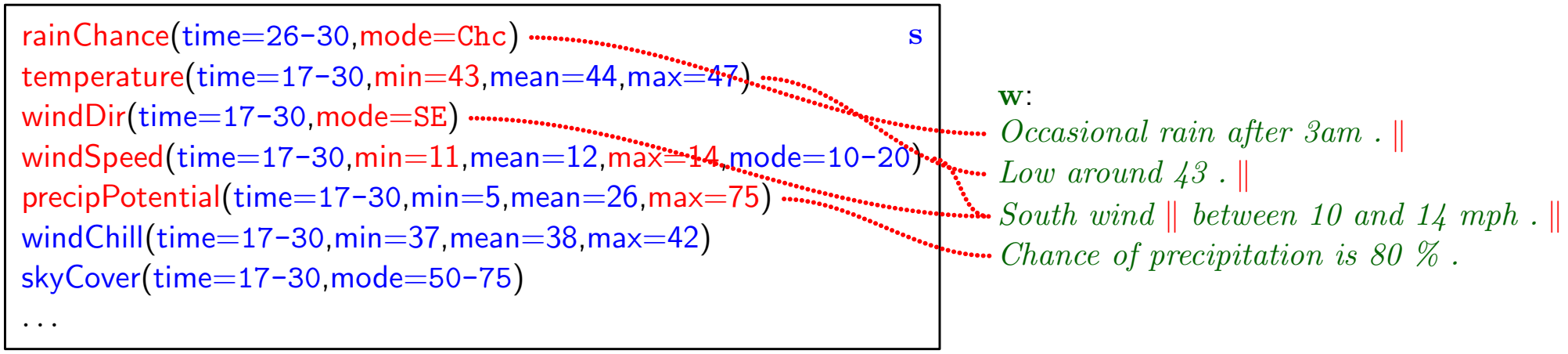
The task (unsupervised induction):

scenarios $\{(\mathbf{s}^{(i)}, \mathbf{w}^{(i)})\}_{i=1}^{n}$ $\longrightarrow$ correspondences $\{\mathbf{a}^{(i)}\}_{i=1}^{n}$

Data characteristics



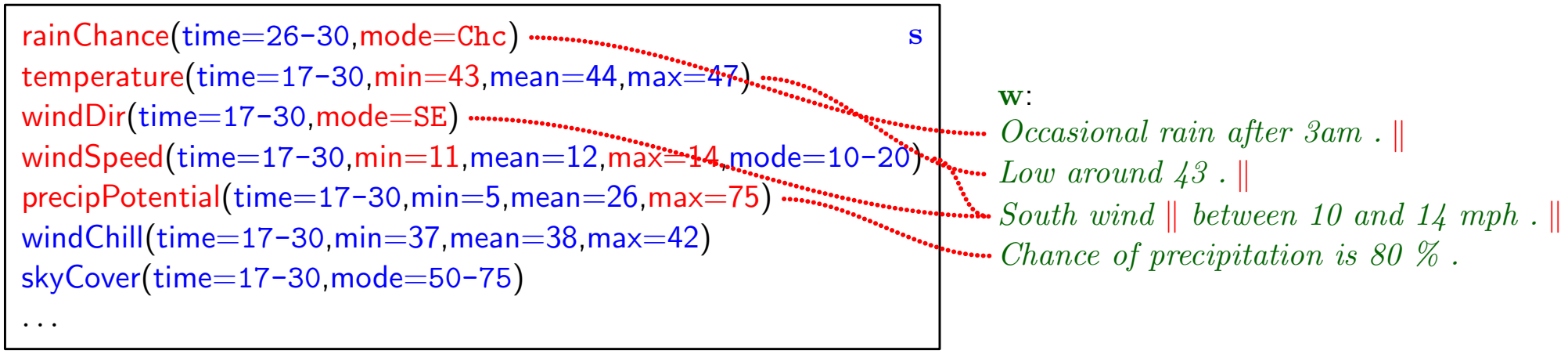
Data characteristics



Referential ambiguity: what is being talked about?

(e.g., *low* is about temperature or wind speed?)

Data characteristics



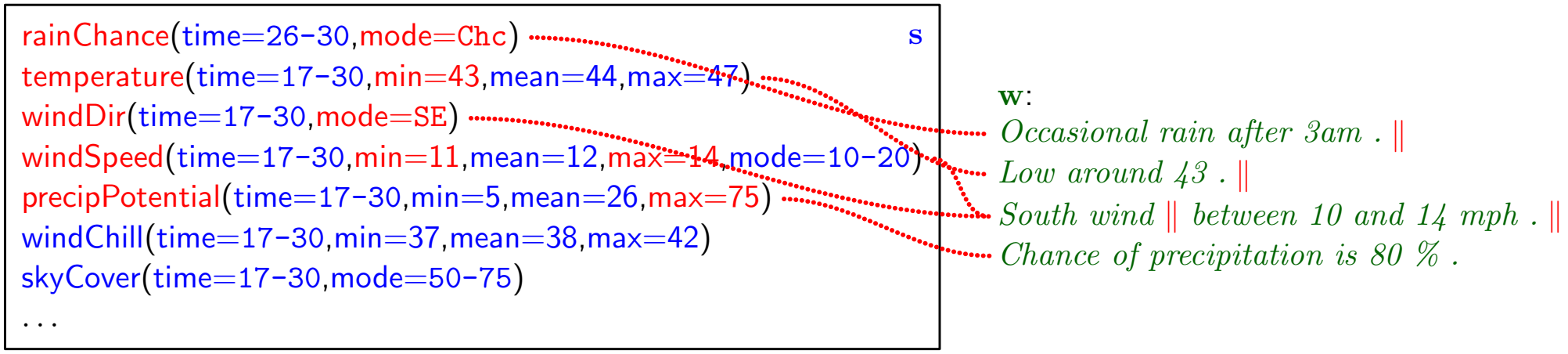
Referential ambiguity: what is being talked about?

(e.g., *low* is about temperature or wind speed?)

Lexical variety: a field value can be realized in multiple ways

(e.g., SE maps to *southeast* or *south* or *east*)

Data characteristics

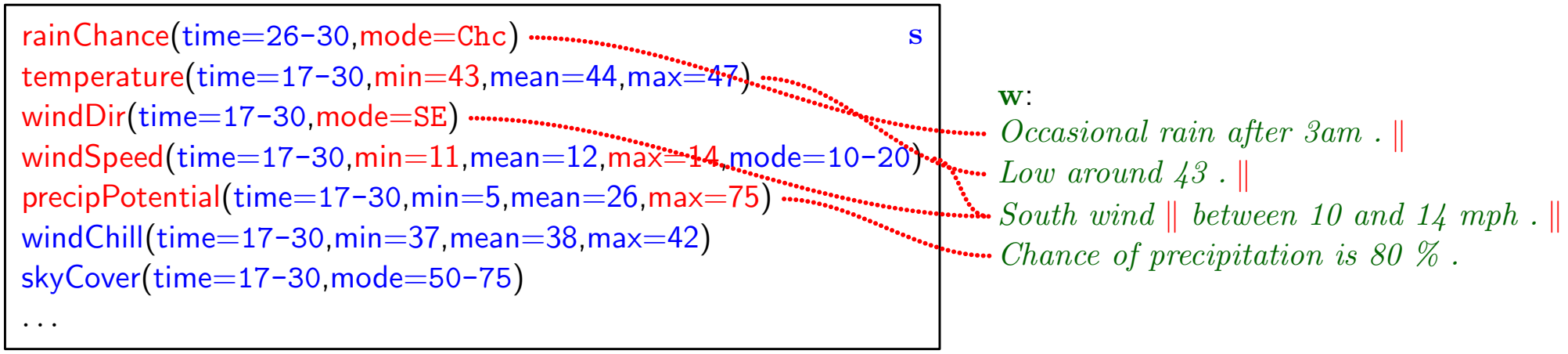


Referential ambiguity: what is being talked about?
(e.g., *low* is about temperature or wind speed?)

Lexical variety: a field value can be realized in multiple ways
(e.g., SE maps to *southeast* or *south* or *east*)

Segmentation: if start a new record, persist a while
(e.g., *low around 43 .* references a single record)

Data characteristics



Referential ambiguity: what is being talked about?
(e.g., *low* is about temperature or wind speed?)

Lexical variety: a field value can be realized in multiple ways
(e.g., SE maps to *southeast* or *south* or *east*)

Segmentation: if start a new record, persist a while
(e.g., *low around 43 .* references a single record)

Coherence: talk about records/fields in a certain order
(e.g., windSpeed typically follows windDir)

Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

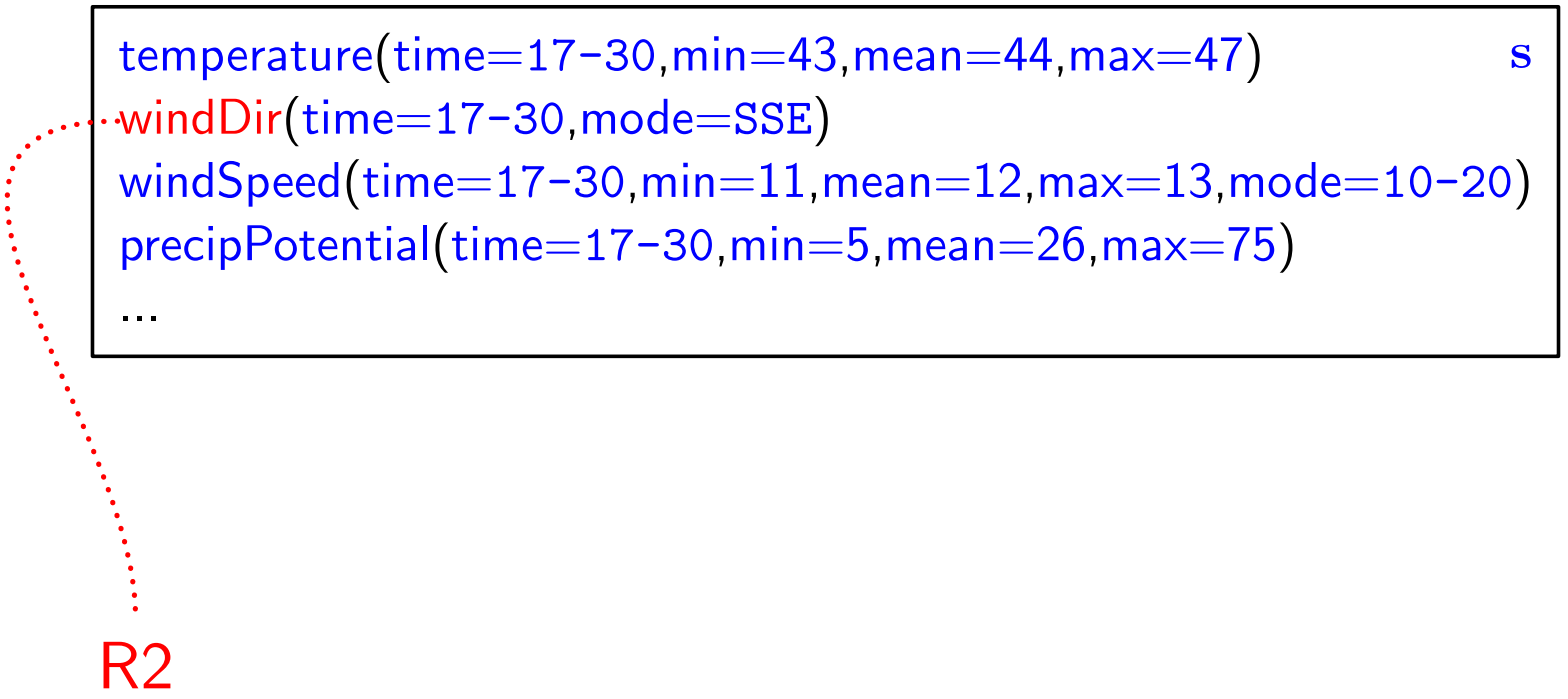
An example:

temperature(time=17-30,min=43,mean=44,max=47) **s**
windDir(time=17-30,mode=SSE)
windSpeed(time=17-30,min=11,mean=12,max=13,mode=10-20)
precipPotential(time=17-30,min=5,mean=26,max=75)
...

Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

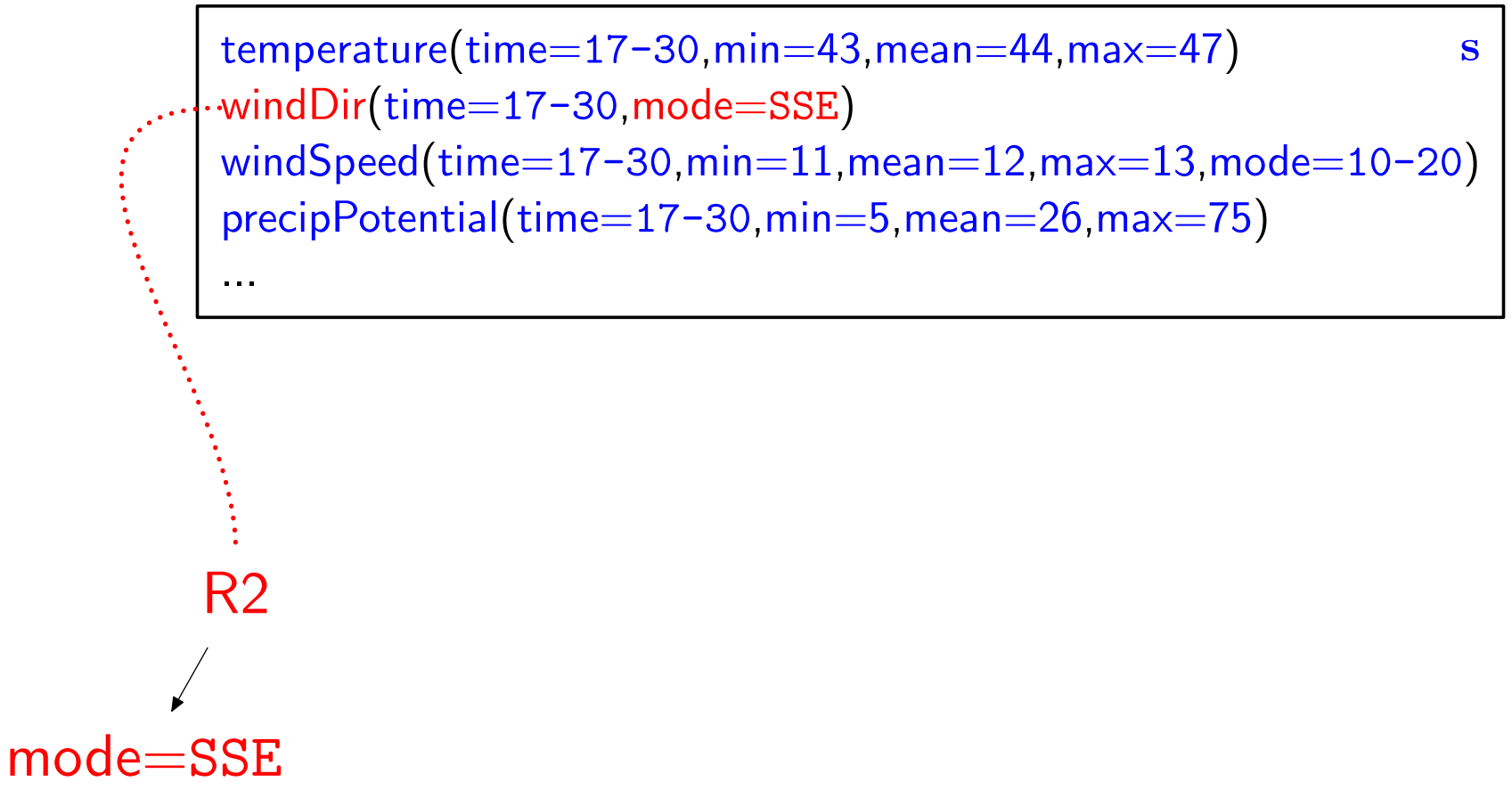
An example:



Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

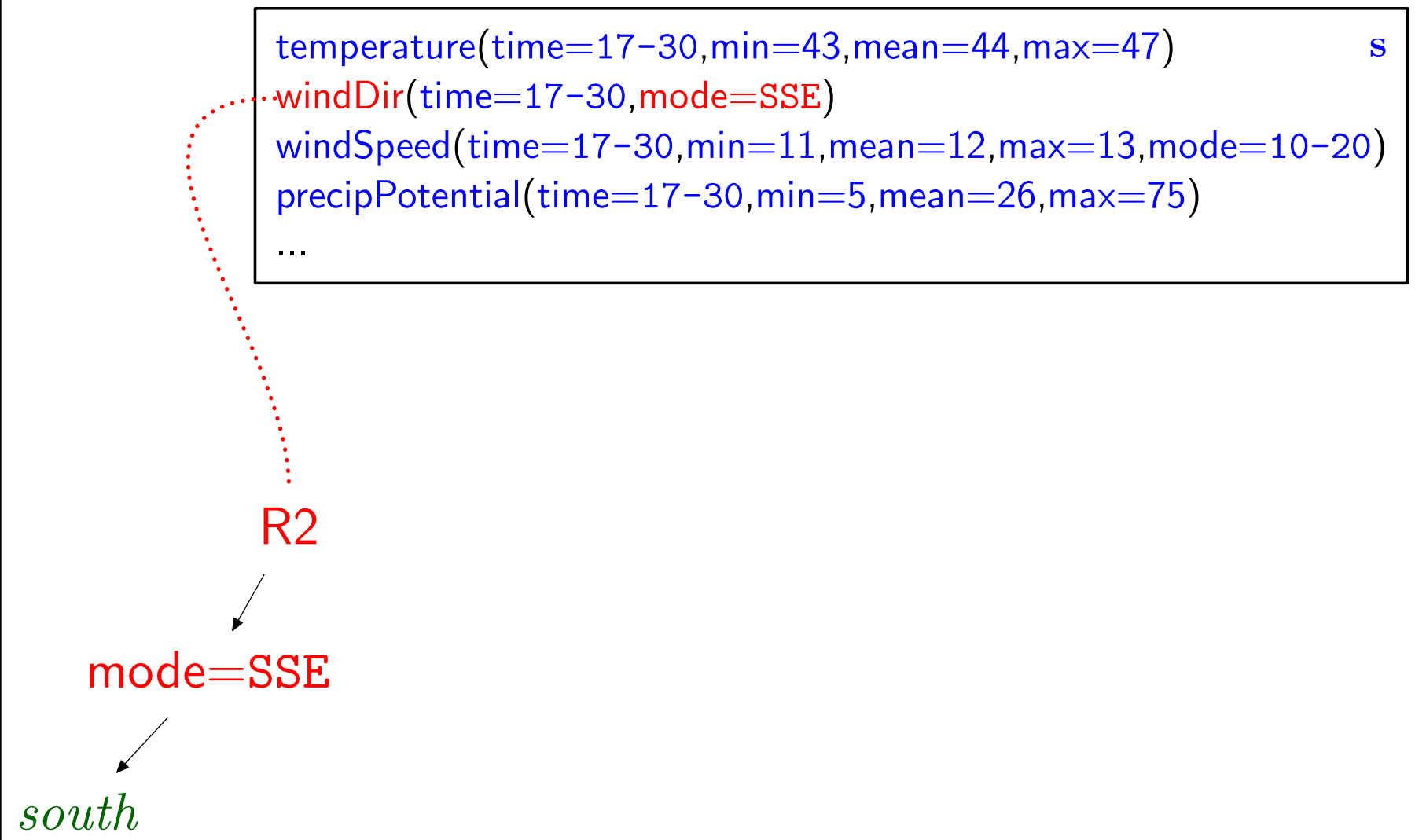
An example:



Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

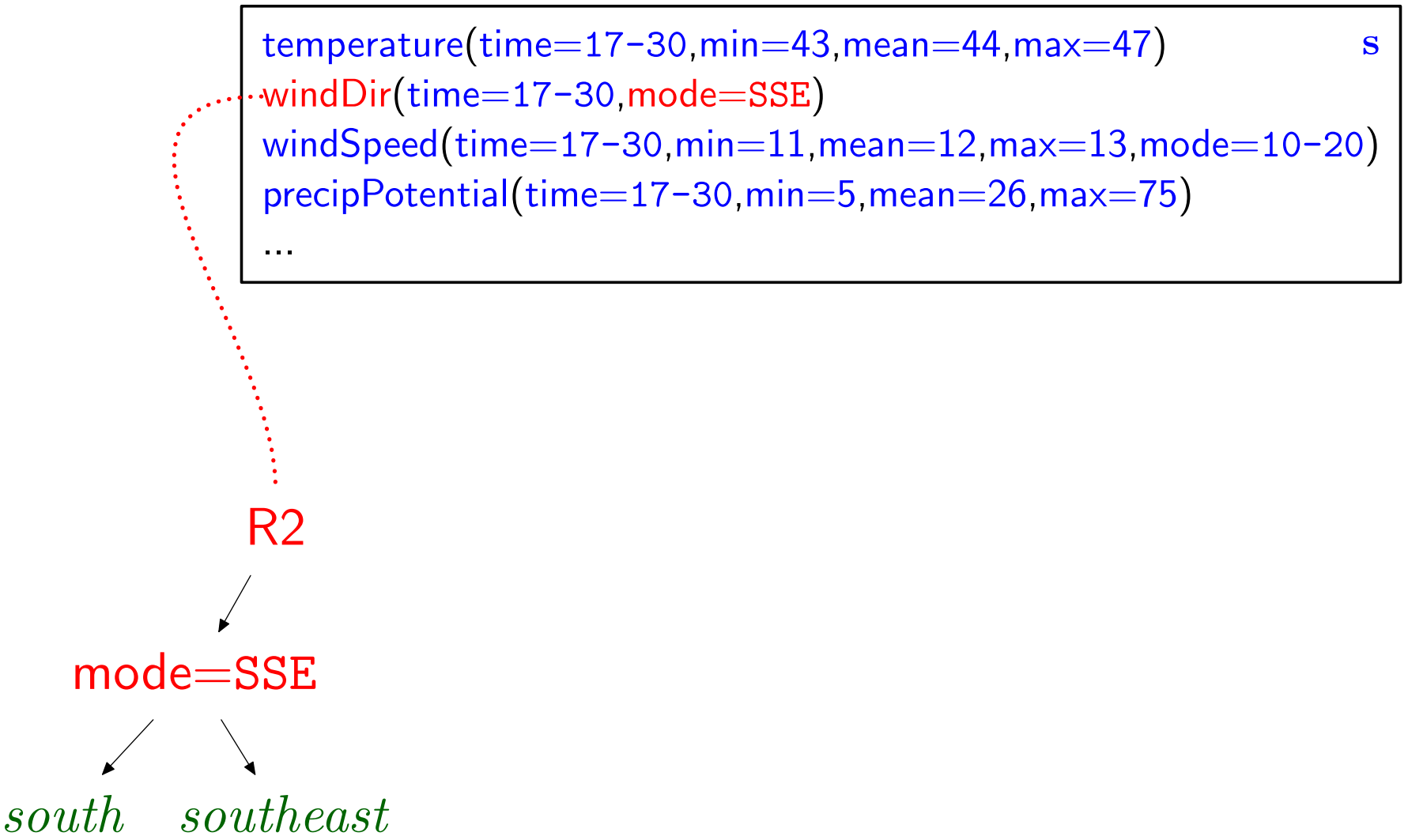
An example:



Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

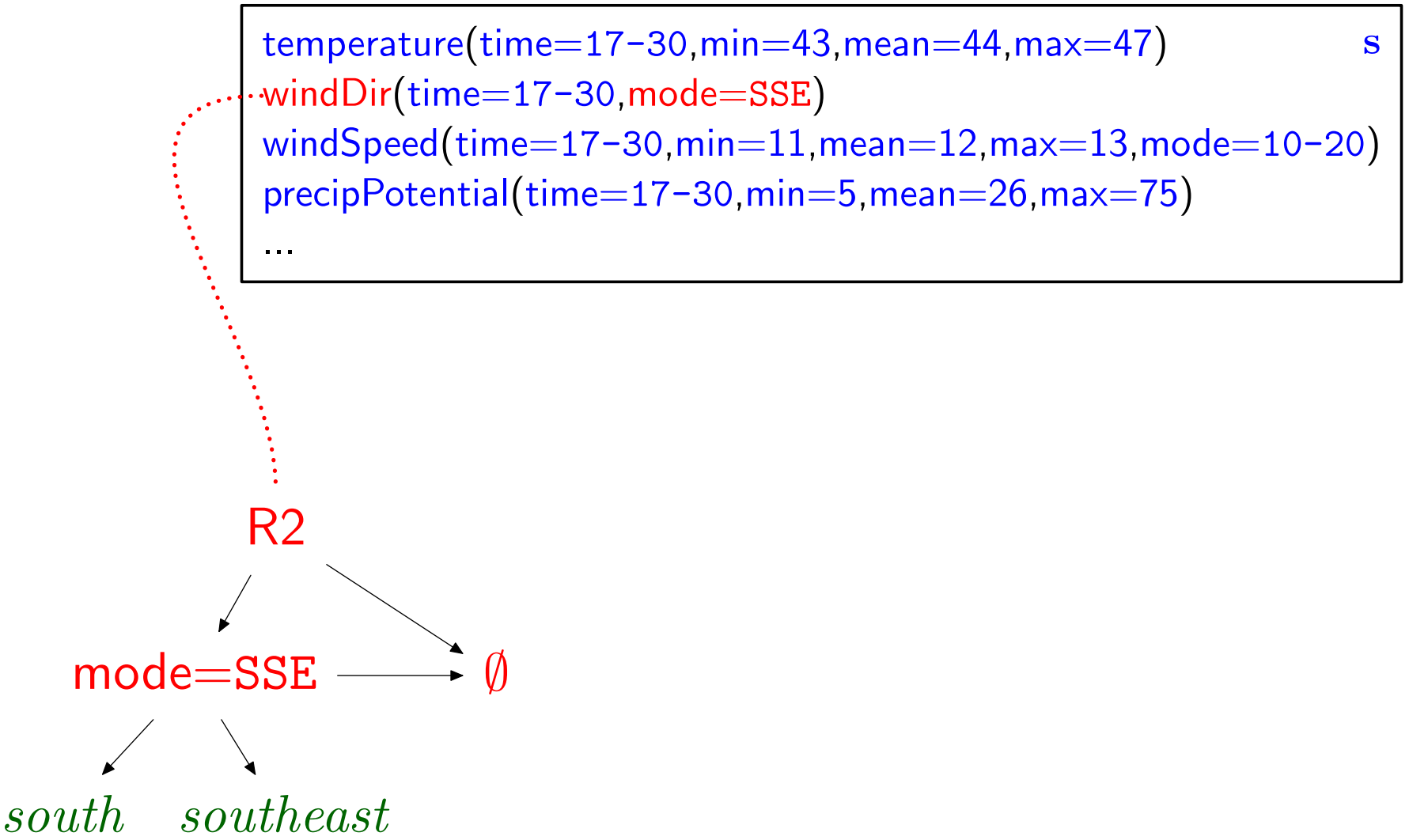
An example:



Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

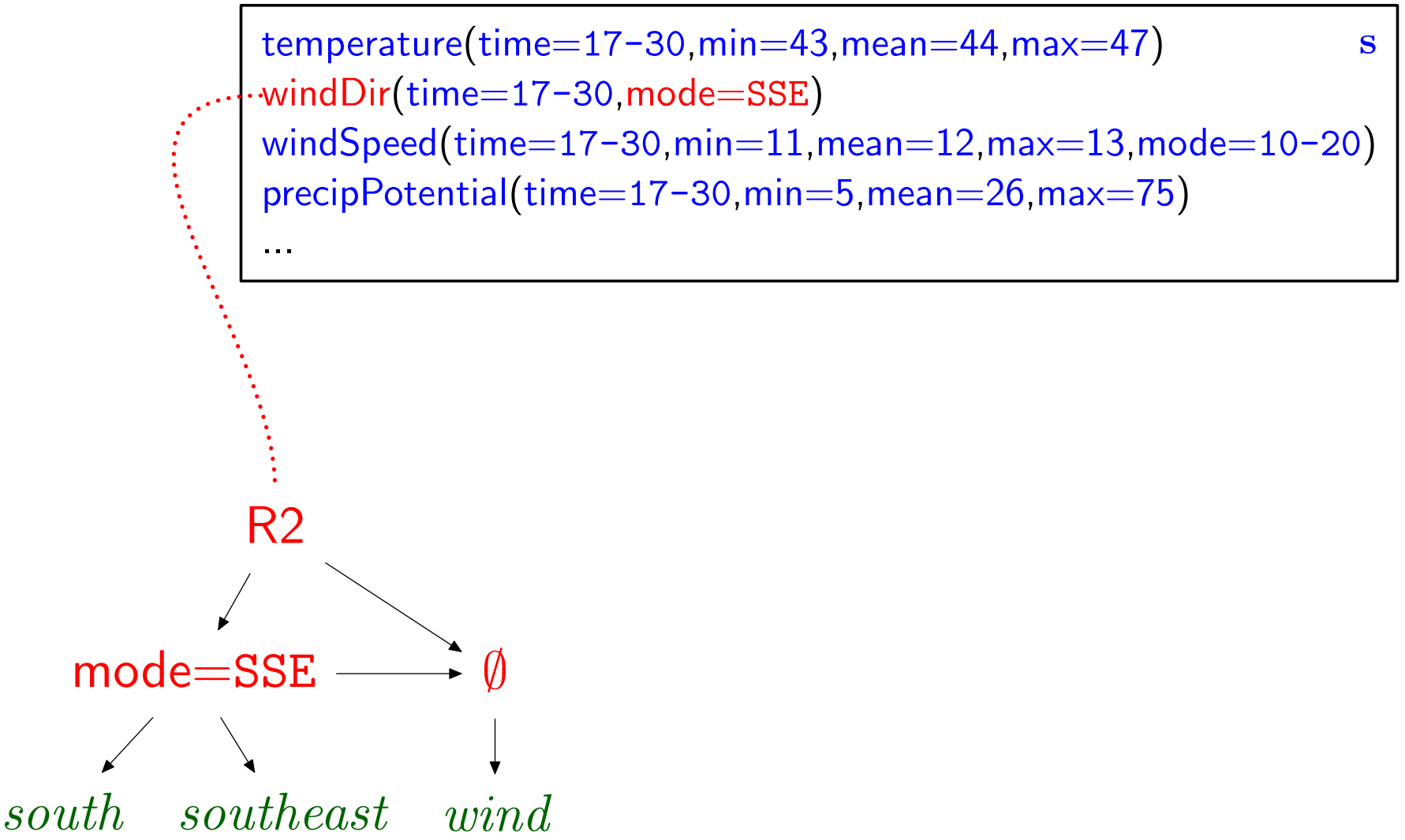
An example:



Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

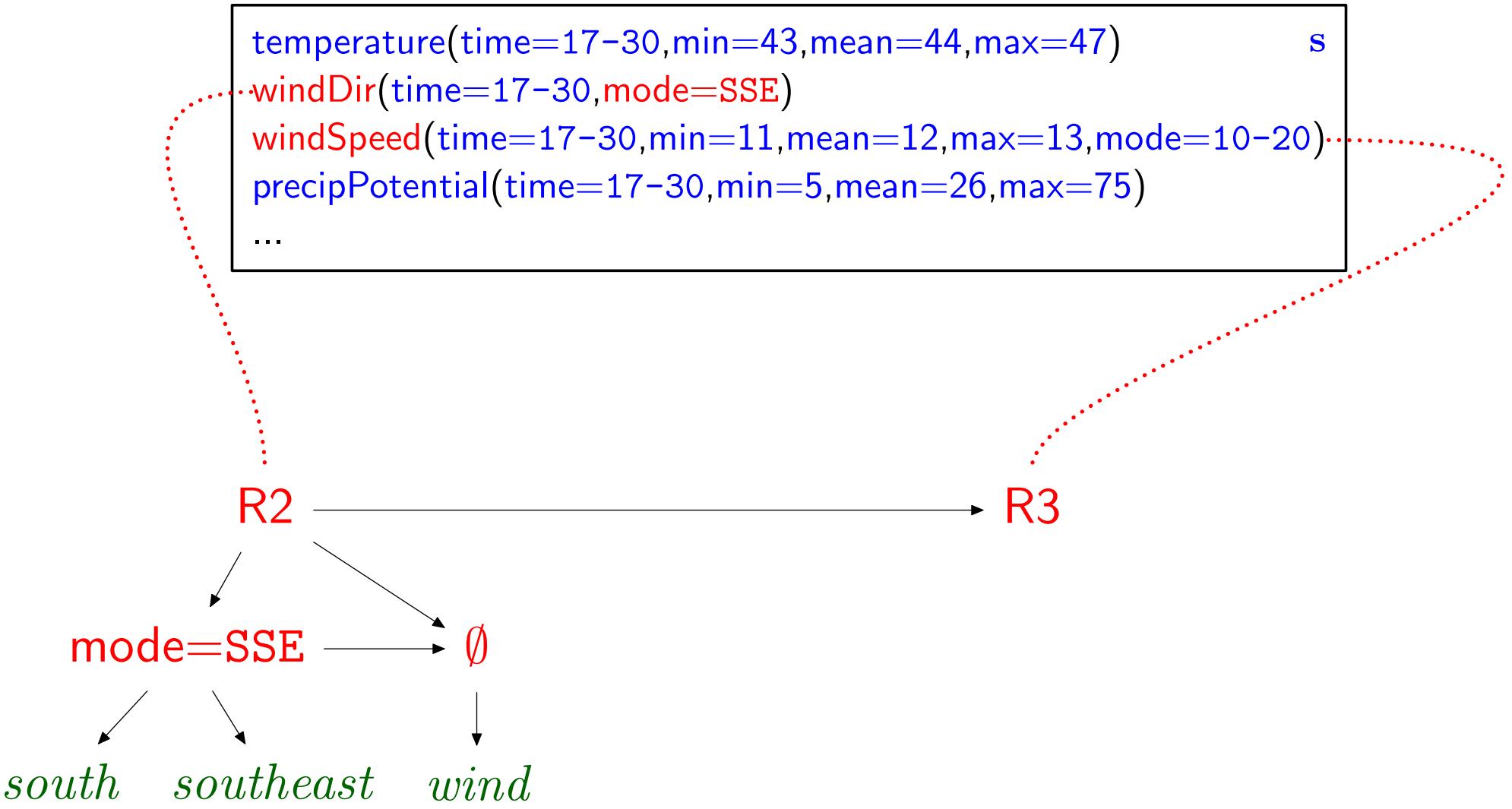
An example:



Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

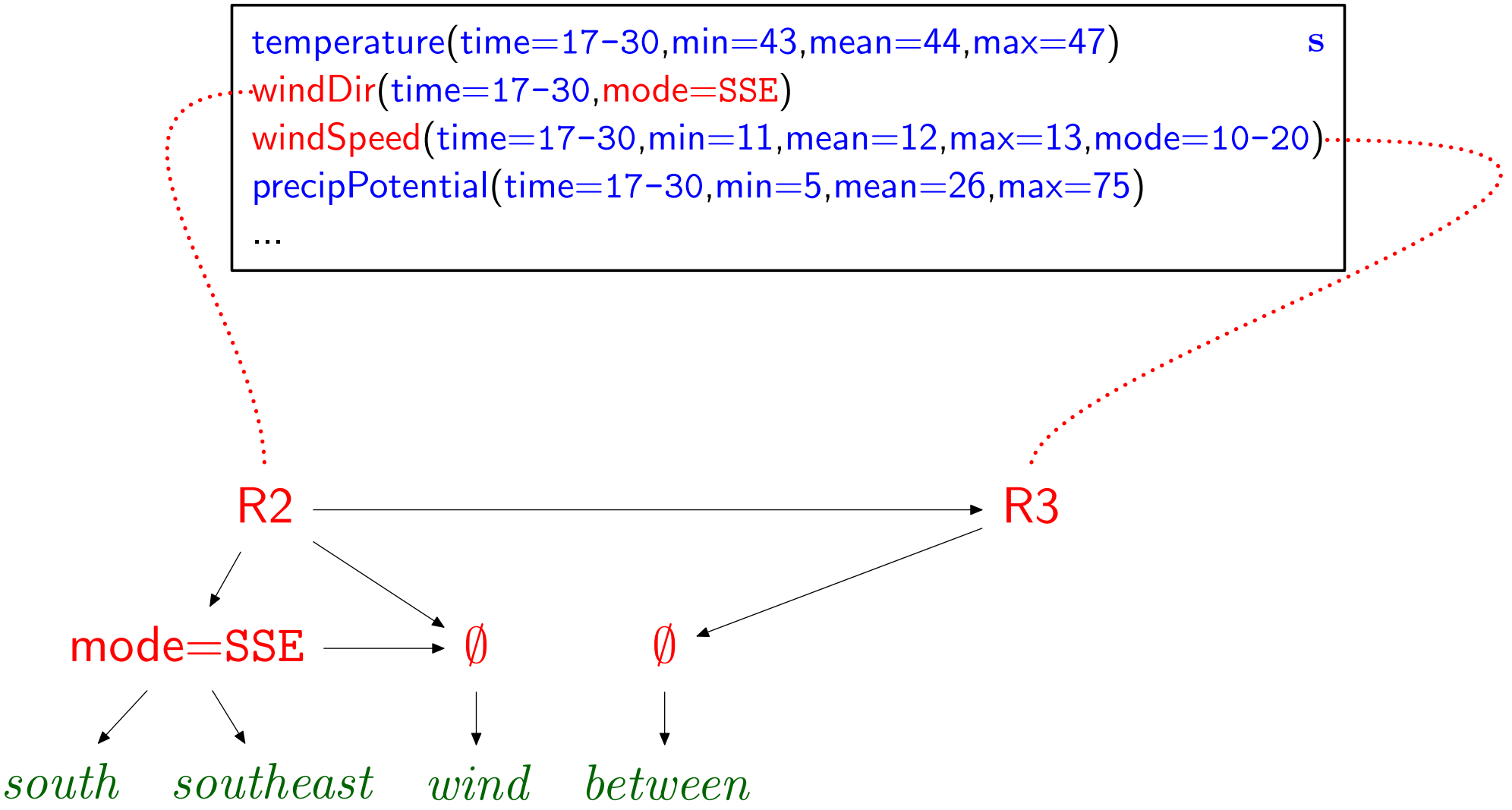
An example:



Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

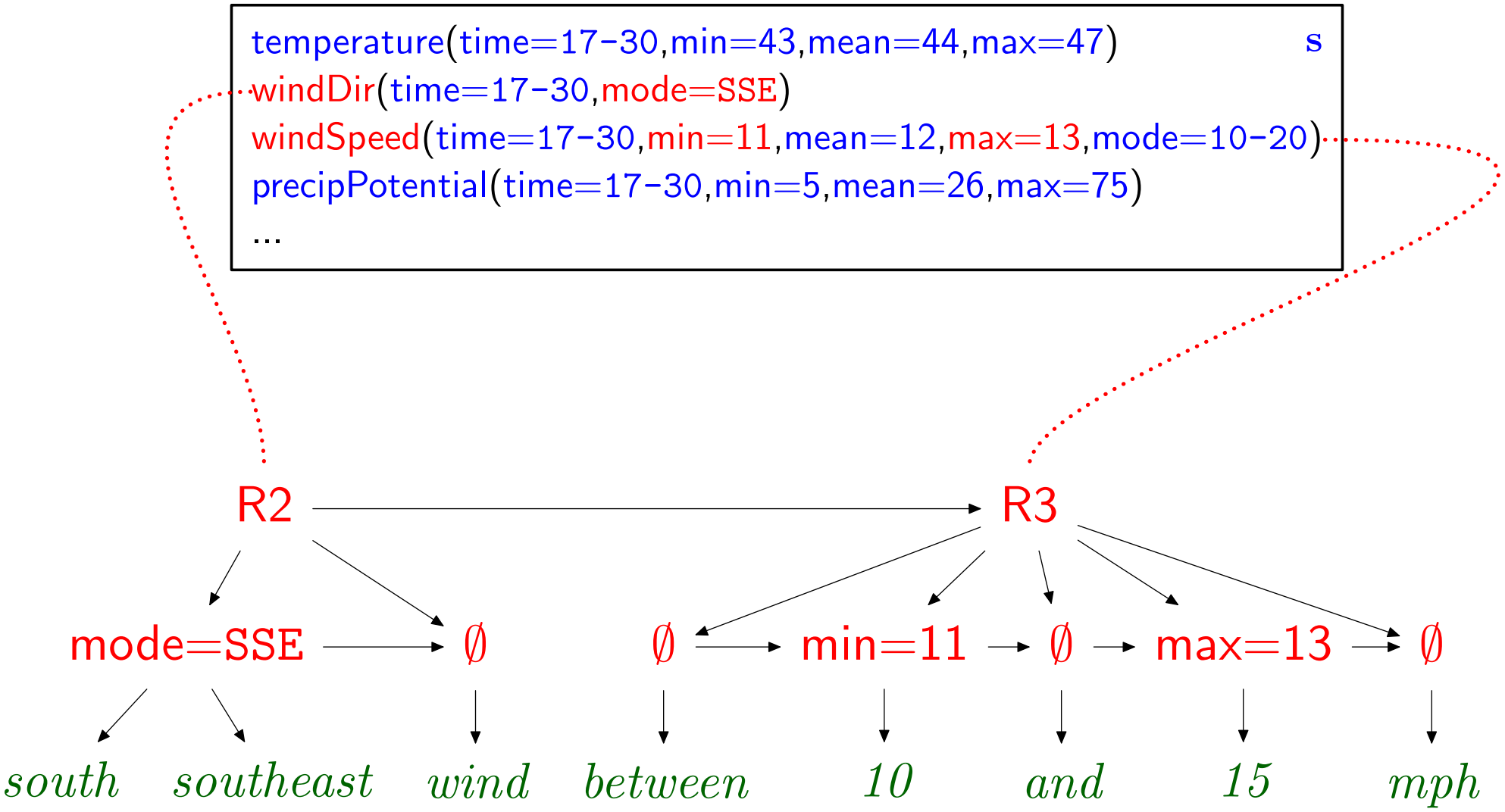
An example:



Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

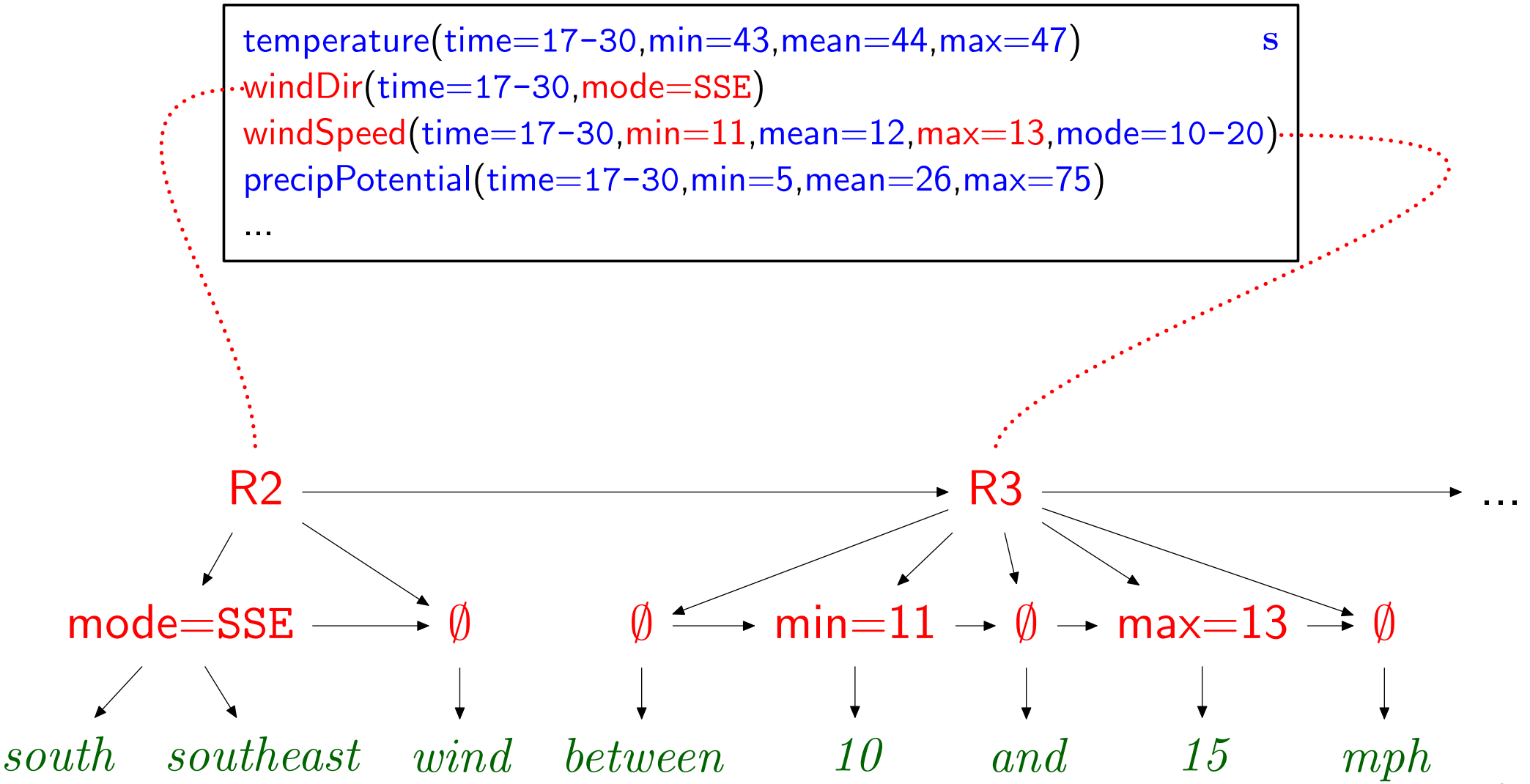
An example:



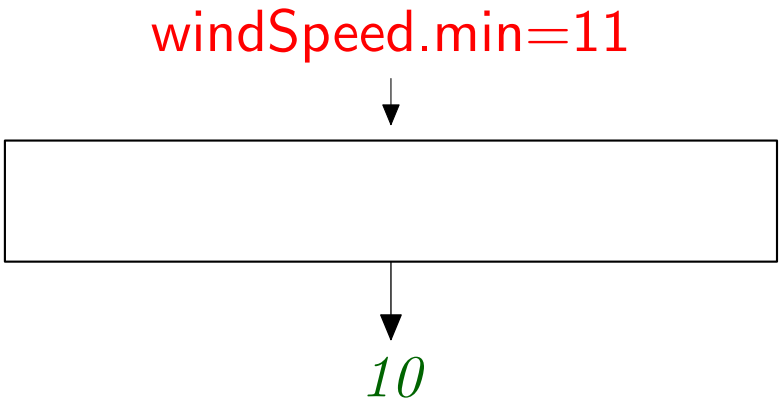
Proposed solution

Define a probabilistic model $p(\mathbf{w}, \mathbf{a} \mid \mathbf{s})$

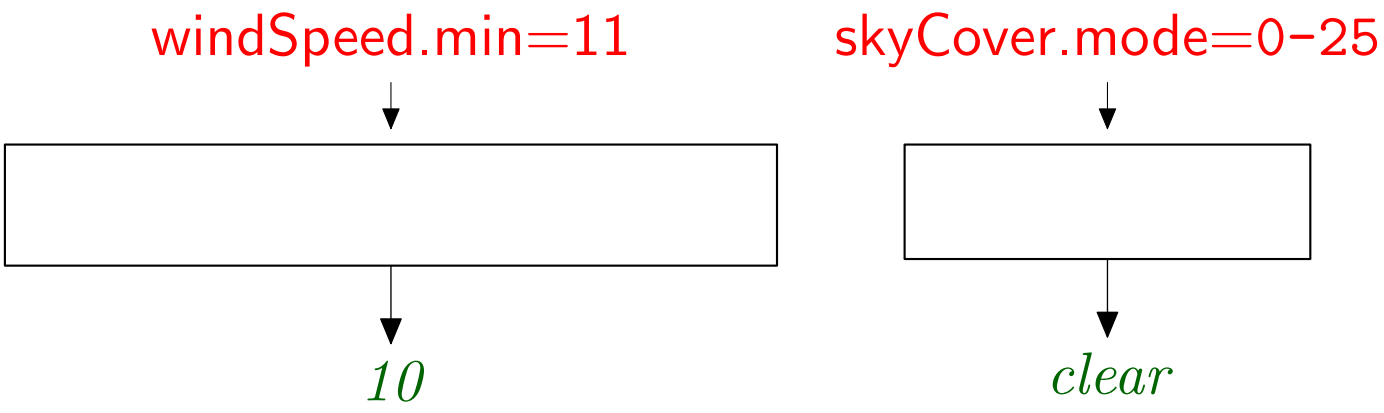
An example:



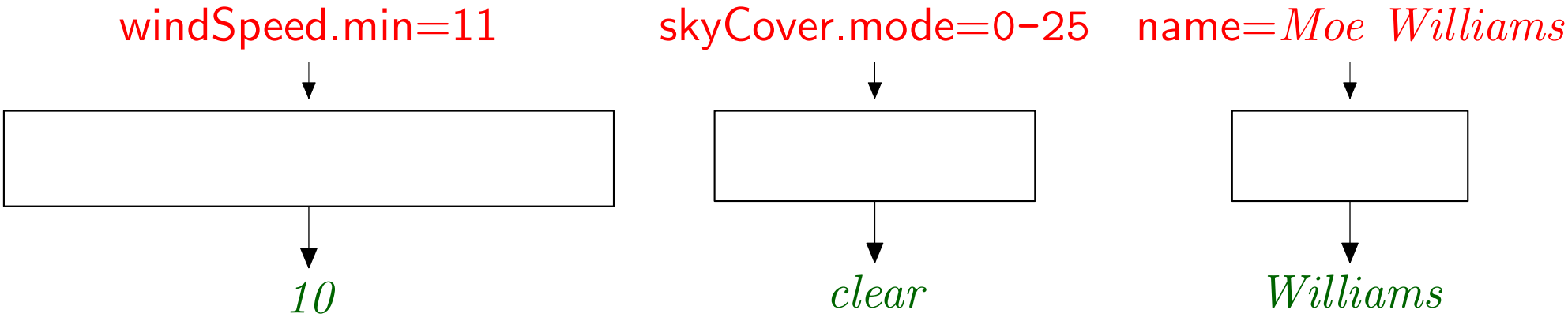
Lexical model



Lexical model



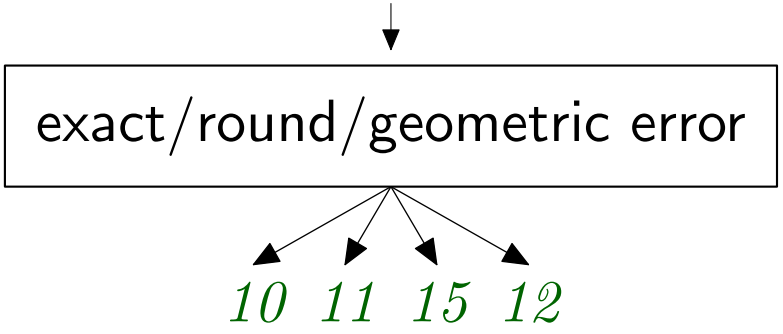
Lexical model



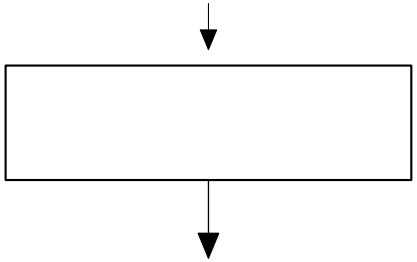
Lexical model

Integer

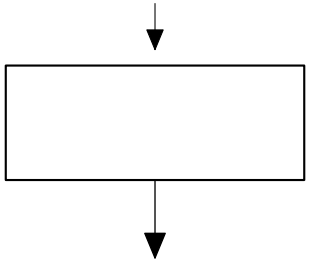
windSpeed.min=11



skyCover.mode=0-25



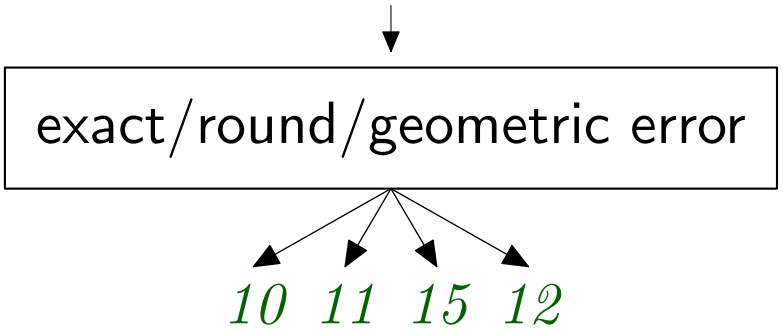
name=*Moe Williams*



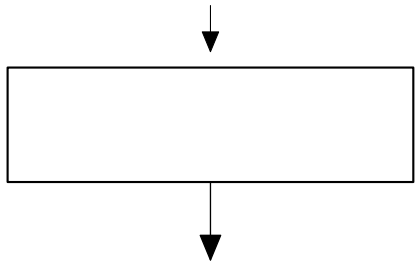
Lexical model

Integer

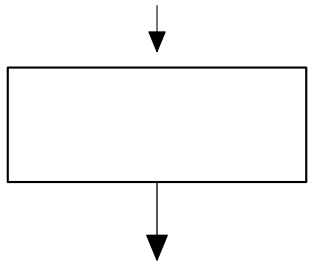
windSpeed.min=11



skyCover.mode=0-25

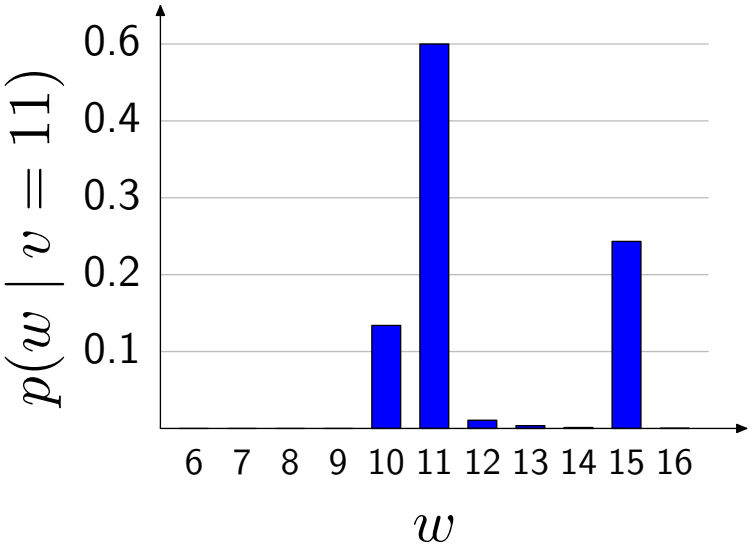


name=*Moe Williams*



Learned distributions:

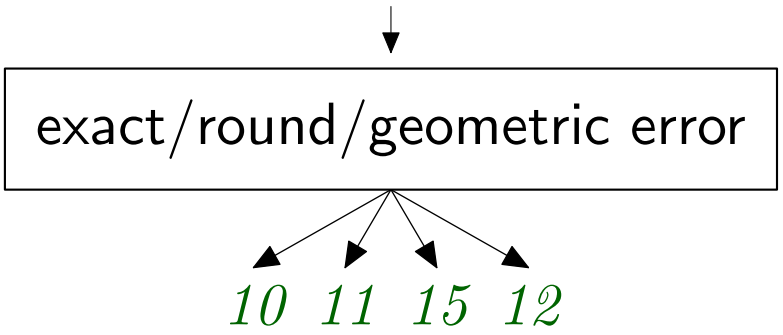
windSpeed.min



Lexical model

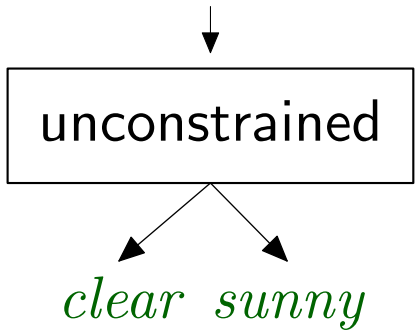
Integer

windSpeed.min=11

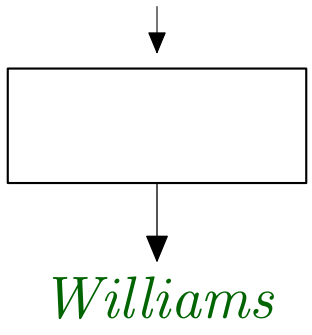


Categorical

skyCover.mode=0-25

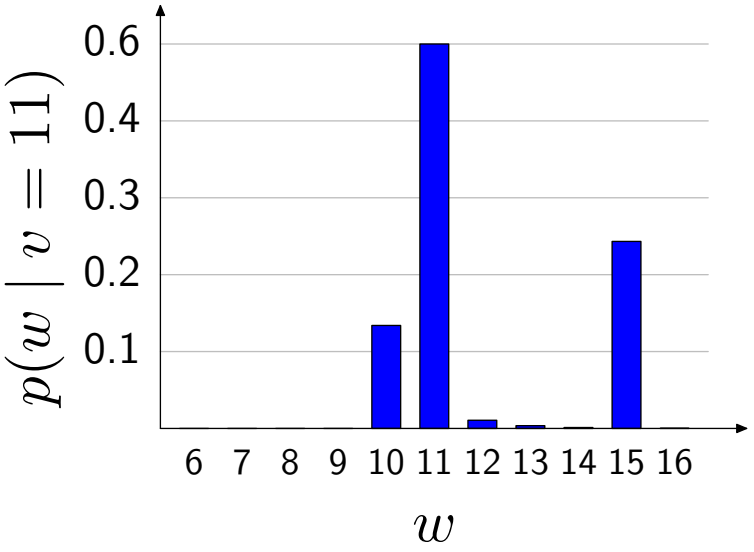


name=*Moe Williams*



Learned distributions:

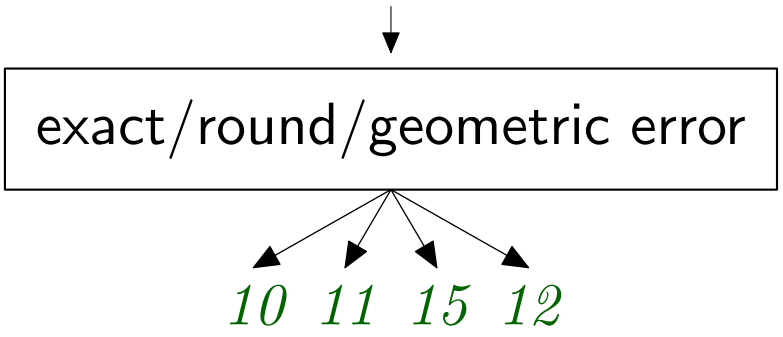
windSpeed.min



Lexical model

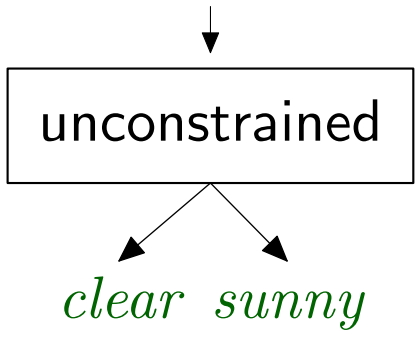
Integer

windSpeed.min=11

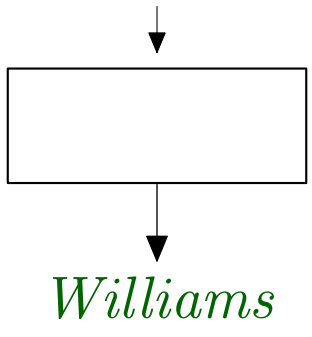


Categorical

skyCover.mode=0-25

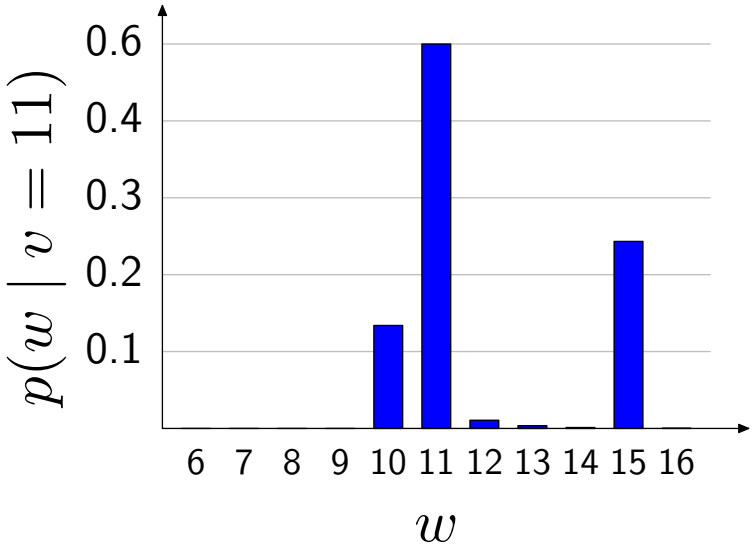


name=*Moe Williams*



Learned distributions:

windSpeed.min



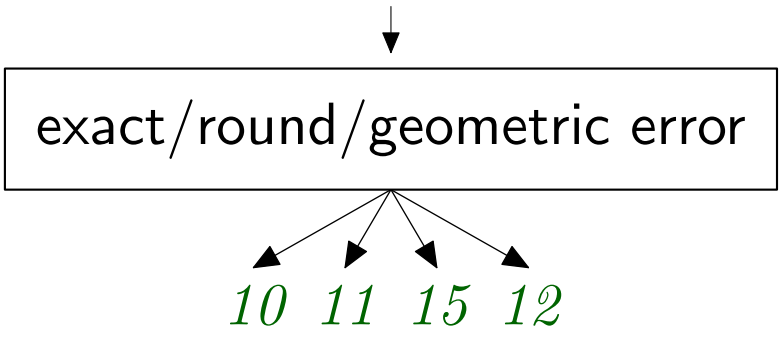
skyCover.mode

0-25	25-50	50-75	75-100
,	partly	mostly	of
clear	,	cloudy	inch
mostly	cloudy	,	an
sunny	increasing	partly	possible
			new
			a
			rainfall

Lexical model

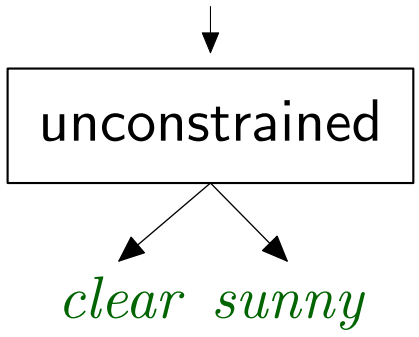
Integer

windSpeed.min=11



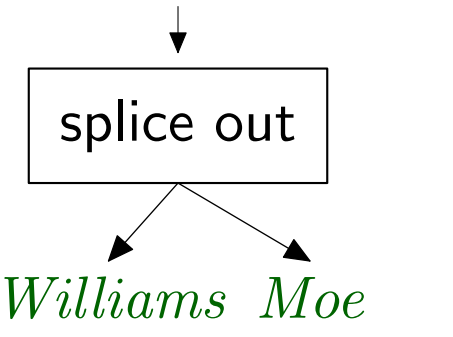
Categorical

skyCover.mode=0-25



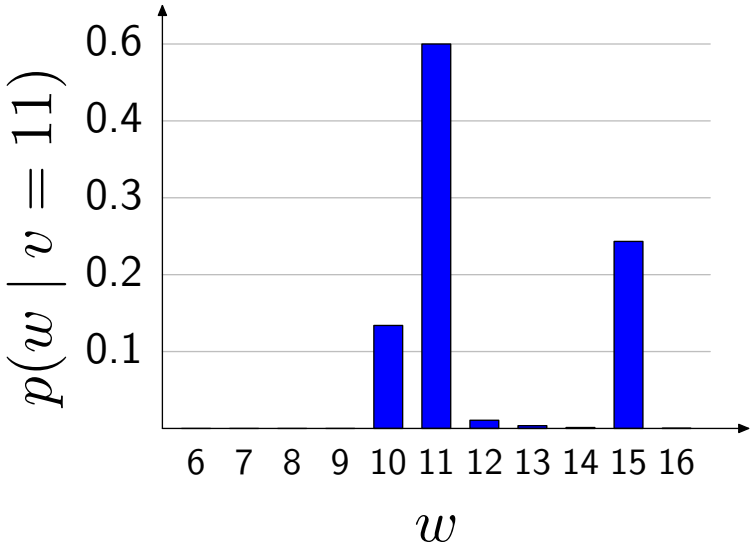
String

name=*Moe Williams*



Learned distributions:

windSpeed.min

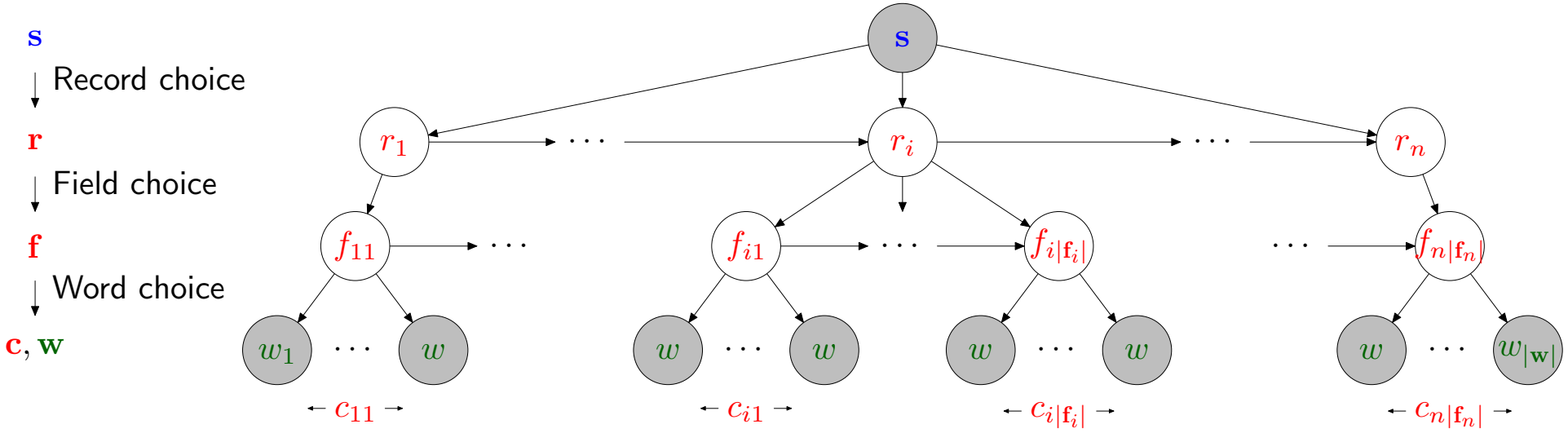


skyCover.mode

0-25	25-50	50-75	75-100
,	partly	mostly	of
clear	,	cloudy	inch
mostly	cloudy	,	an
sunny	increasing	partly	possible
			new
			a
			rainfall

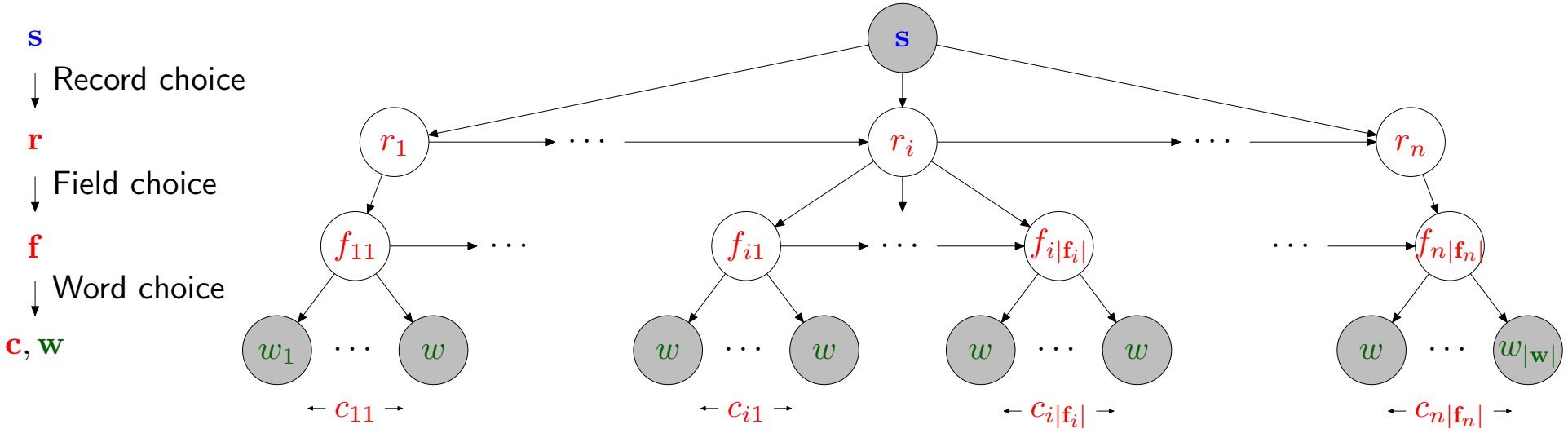
Generative model

Hierarchical semi-Markov model $p(\mathbf{w}, \mathbf{c}, \mathbf{r}, \mathbf{f} \mid \mathbf{s}; \theta)$:



Generative model

Hierarchical semi-Markov model $p(\mathbf{w}, \mathbf{c}, \mathbf{r}, \mathbf{f} \mid \mathbf{s}; \theta)$:

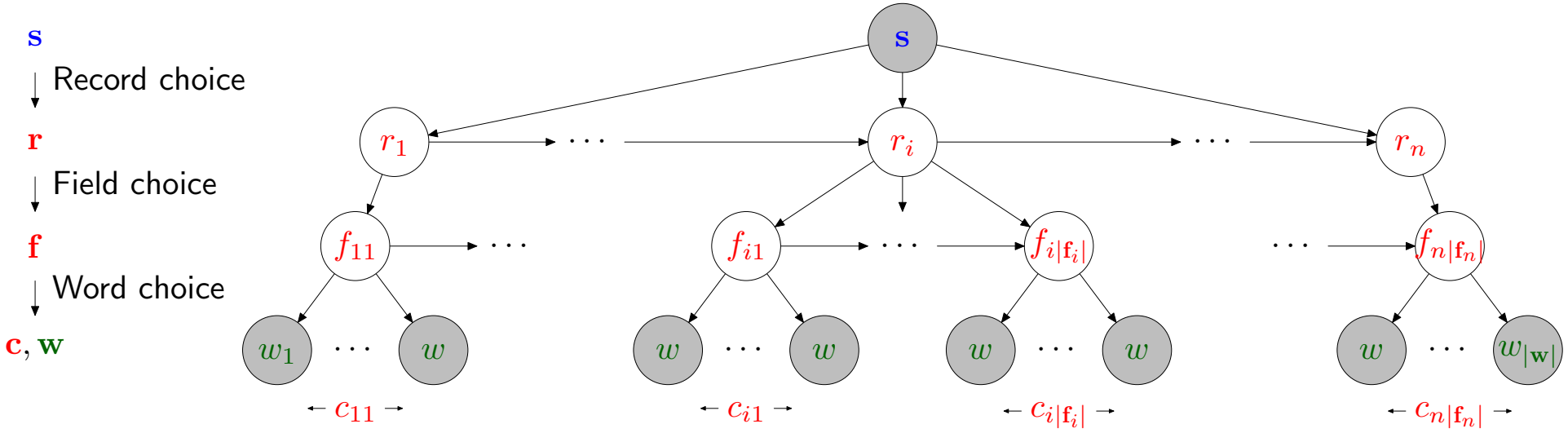


Objective: maximize likelihood (marginalize out $\mathbf{a} = (\mathbf{c}, \mathbf{r}, \mathbf{f})$)

$$\max_\theta \sum_{i=1}^{n} \log p(\mathbf{w}^{(i)} \mid \mathbf{s}^{(i)}; \theta)$$

Generative model

Hierarchical semi-Markov model $p(\mathbf{w}, \mathbf{c}, \mathbf{r}, \mathbf{f} \mid \mathbf{s}; \theta)$:



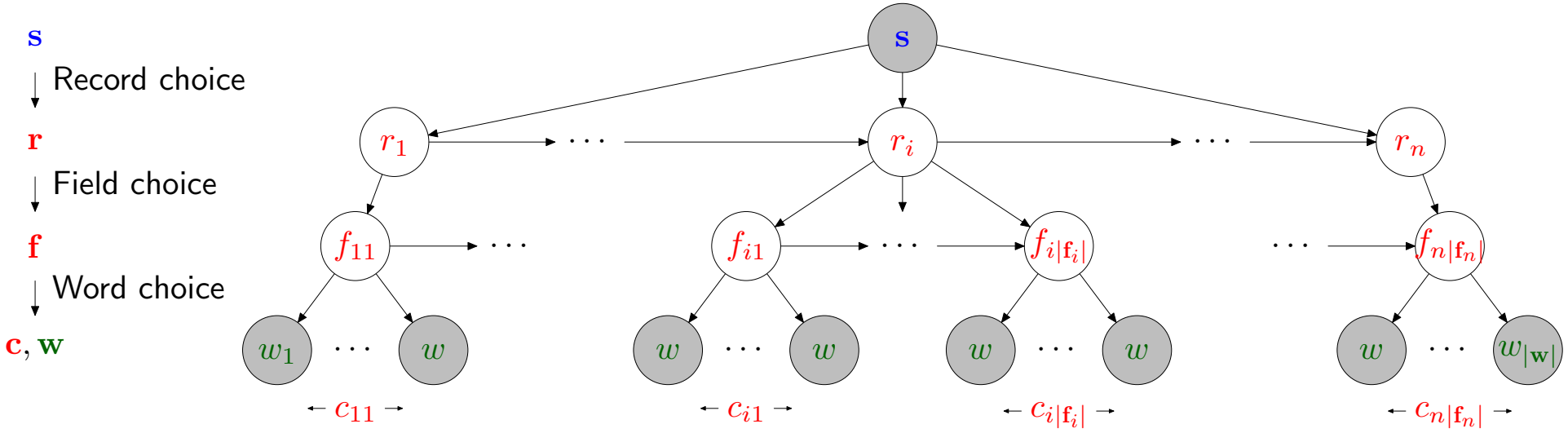
Objective: maximize likelihood (marginalize out $\mathbf{a} = (\mathbf{c}, \mathbf{r}, \mathbf{f})$)

$$\max_\theta \sum_{i=1}^{n} \log p(\mathbf{w}^{(i)} \mid \mathbf{s}^{(i)}; \theta)$$

Why likelihood?

Generative model

Hierarchical semi-Markov model $p(\mathbf{w}, \mathbf{c}, \mathbf{r}, \mathbf{f} \mid \mathbf{s}; \theta)$:



Objective: maximize likelihood (marginalize out $\mathbf{a} = (\mathbf{c}, \mathbf{r}, \mathbf{f})$)

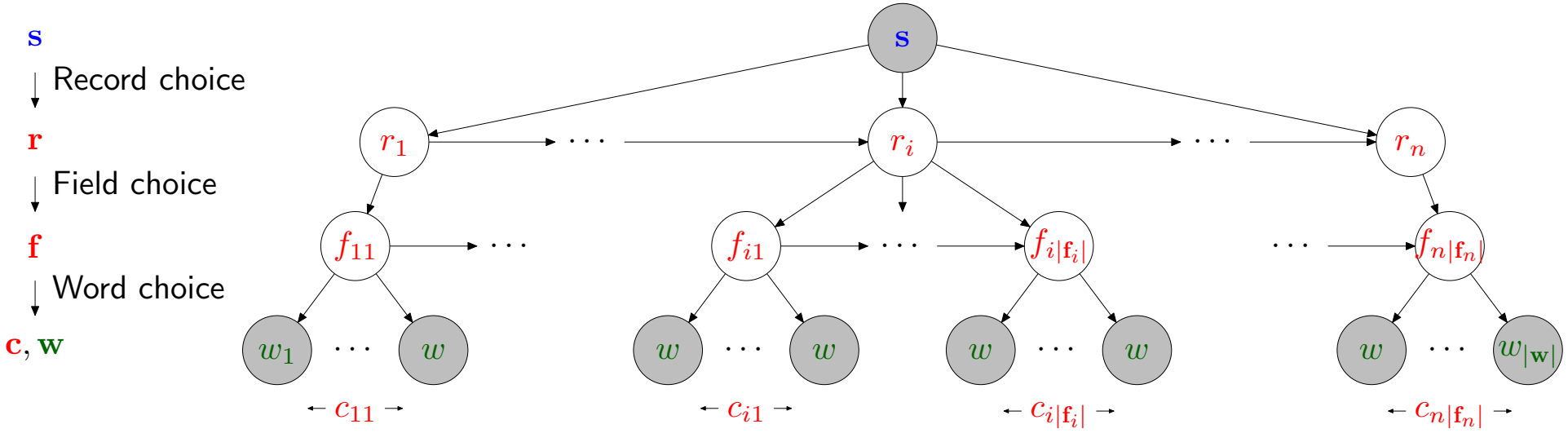
$$\max_\theta \sum_{i=1}^{n} \log p(\mathbf{w}^{(i)} \mid \mathbf{s}^{(i)}; \theta)$$

Why likelihood?

- To make $\mathbf{w}$ likely, use $\mathbf{a}$ to correspond co-occurring things

Generative model

Hierarchical semi-Markov model $p(\mathbf{w}, \mathbf{c}, \mathbf{r}, \mathbf{f} \mid \mathbf{s}; \theta)$:



Objective: maximize likelihood (marginalize out $\mathbf{a} = (\mathbf{c}, \mathbf{r}, \mathbf{f})$)

$$\max_\theta \sum_{i=1}^{n} \log p(\mathbf{w}^{(i)} \mid \mathbf{s}^{(i)}; \theta)$$

Why likelihood?

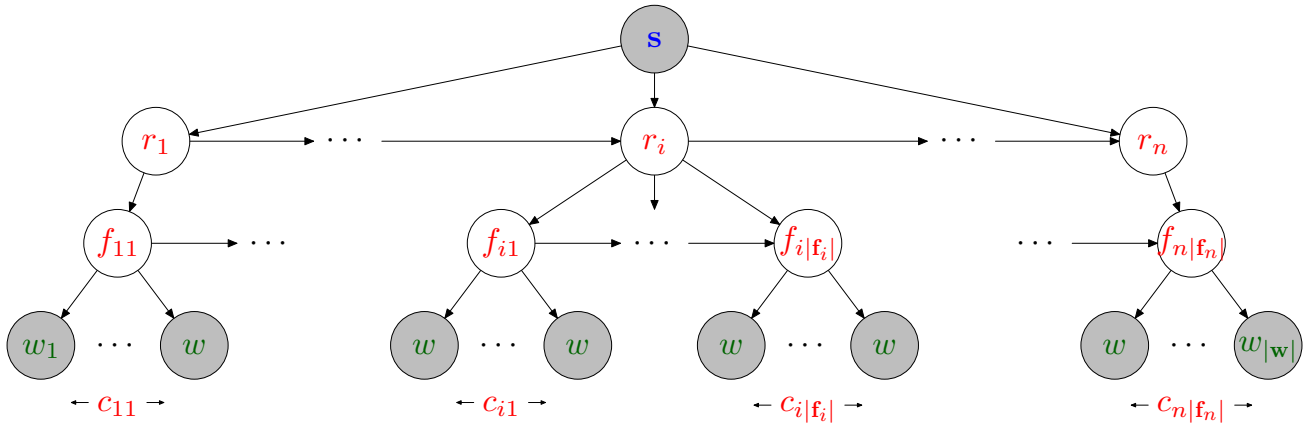
- To make $\mathbf{w}$ likely, use $\mathbf{a}$ to correspond co-occurring things
- $\mathbf{w} \mid \mathbf{s}$ because text only mentions small part of world state

Alternate models

Model 3

coherence: ✔

segmentation: ✔

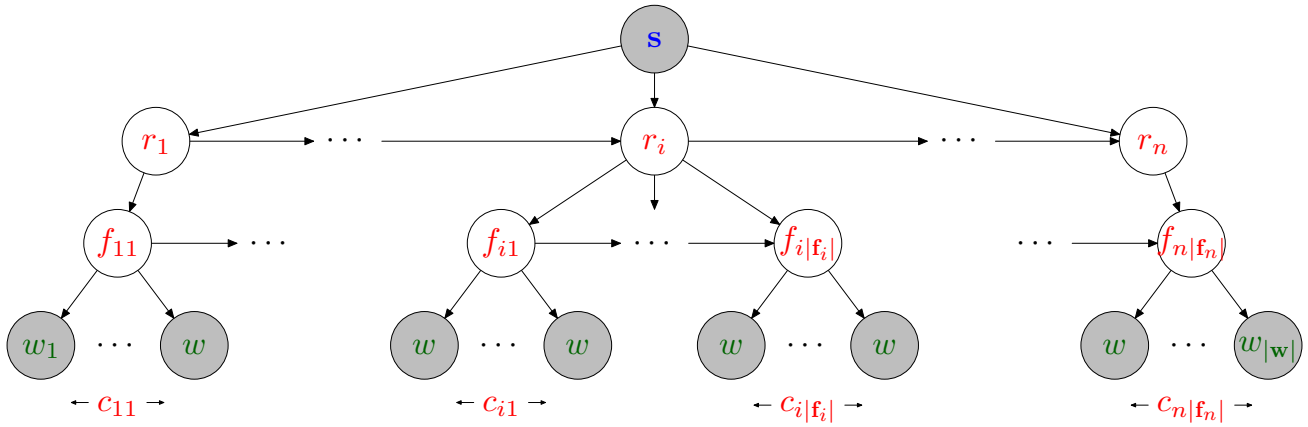


Alternate models

Model 3

coherence: ✔

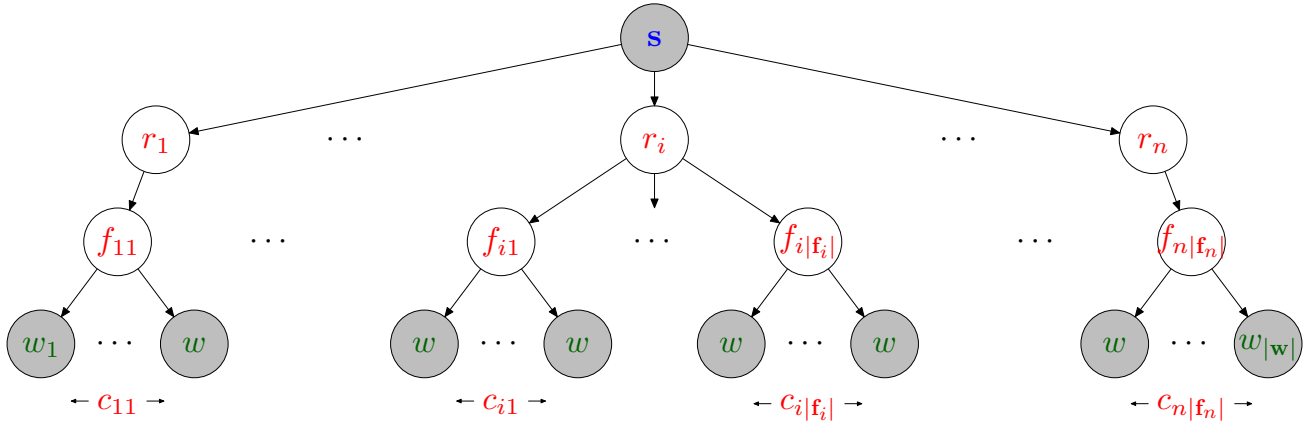
segmentation: ✔



Model 2

coherence: ✘

segmentation: ✔

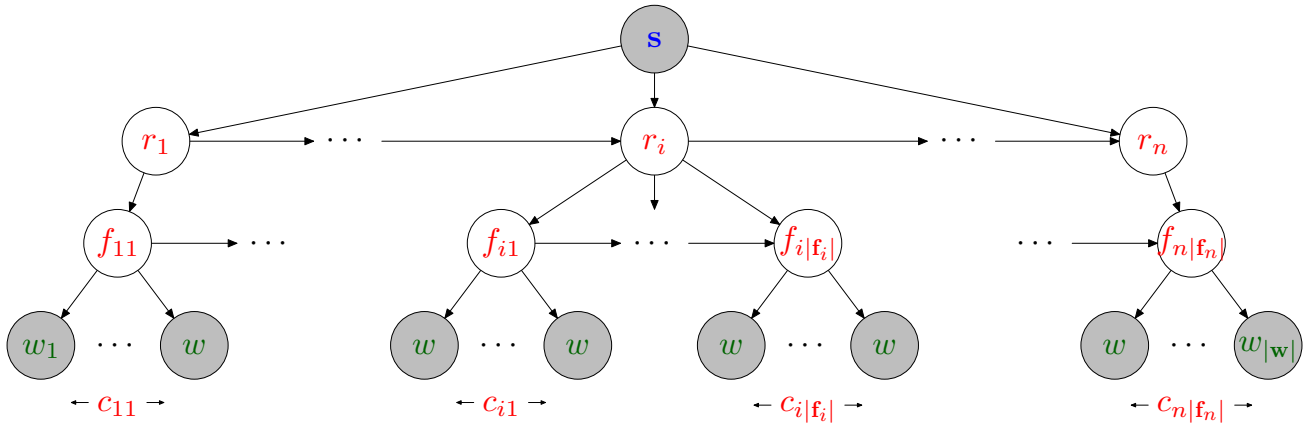


Alternate models

Model 3

coherence: ✔

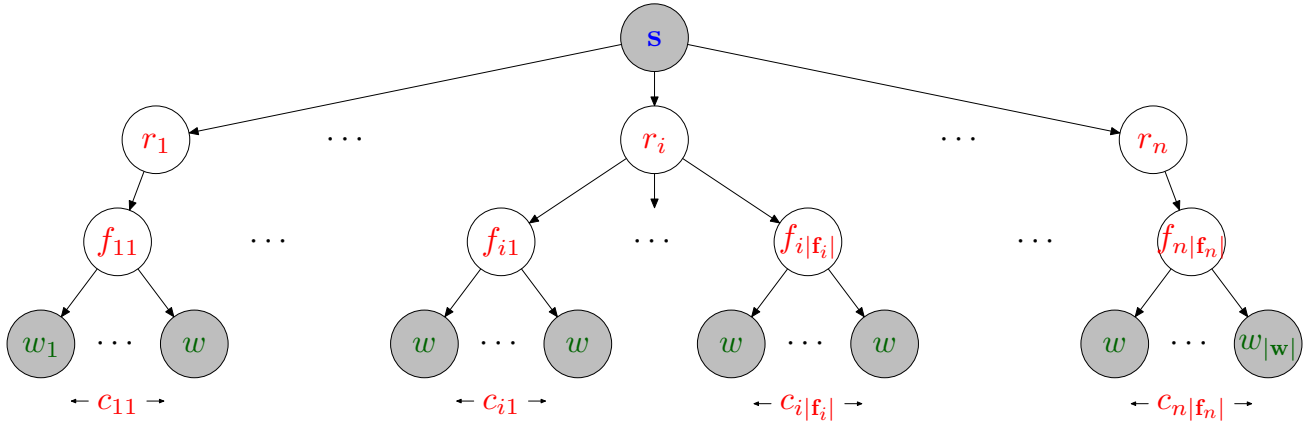
segmentation: ✔



Model 2

coherence: ✗

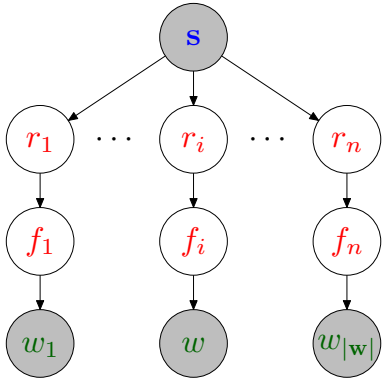
segmentation: ✔



Model 1

coherence: ✗

segmentation: ✗



Experiments overview

Datasets:

- Weather forecasts [Liang et al., 2009]
- Robocup sportscasting [Chen & Mooney, 2008]
- NFL recaps [Barzilay & Lapata, 2005; Snyder & Barzilay, 2007]

Experiments overview

Datasets:

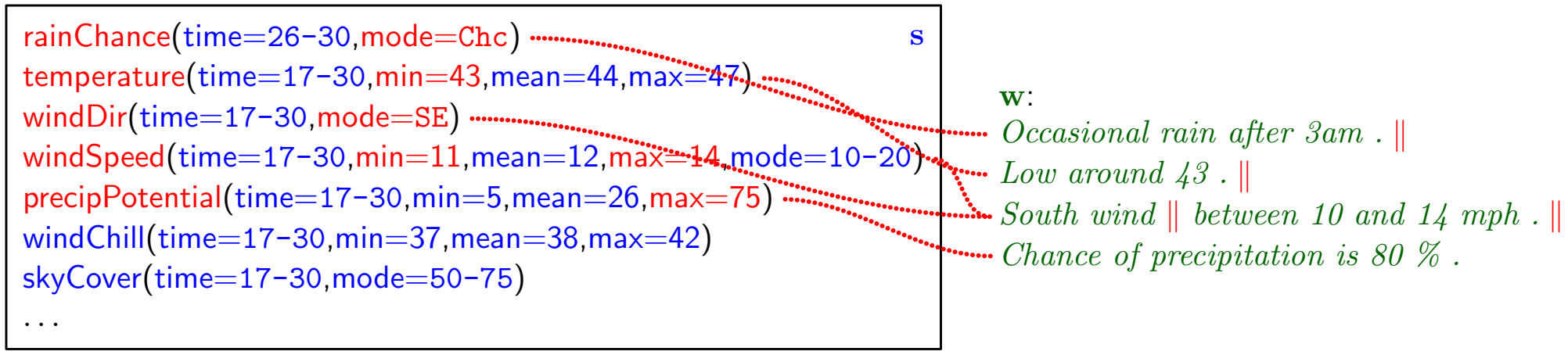
- Weather forecasts [Liang et al., 2009]
- Robocup sportscasting [Chen & Mooney, 2008]
- NFL recaps [Barzilay & Lapata, 2005; Snyder & Barzilay, 2007]

Evaluation:

F_1 on coarse record-level alignments

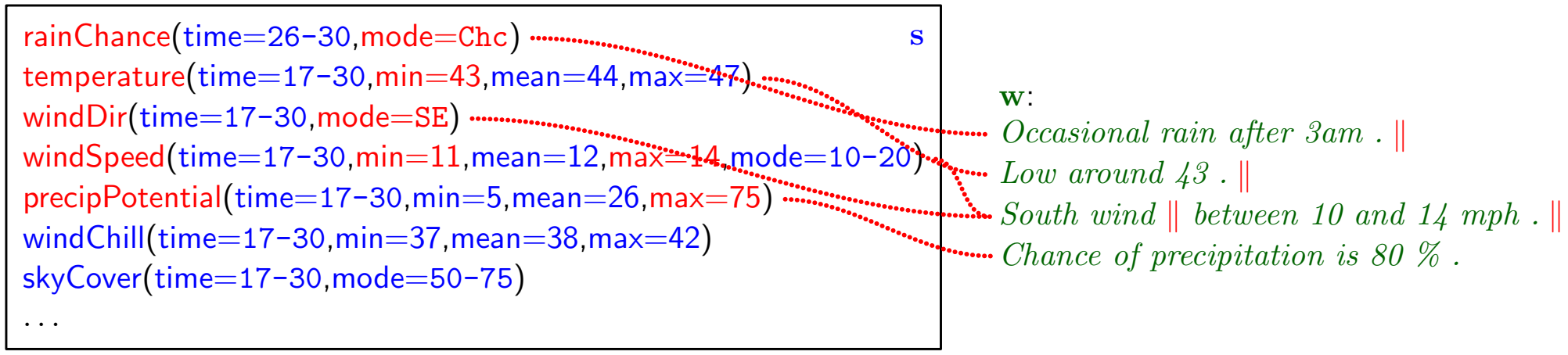
(although model produces finer correspondence)

Experiments: weather forecasts



Dataset (scraped `weather.gov`: 3 days, 3.7K US cities)

Experiments: weather forecasts



Dataset (scraped `weather.gov`: 3 days, 3.7K US cities)

22K scenarios

Each scenario contains

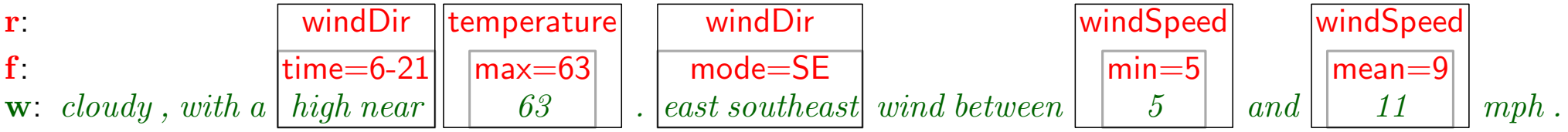
$|\mathbf{s}| = 36$ records, $|\mathbf{w}| = 28.7$ words/text, 5.8 gold alignments

Experiments: weather forecasts (results)

[Model 1] segmentation: ✗ coherence: ✗ (60.0 F_1)

Experiments: weather forecasts (results)

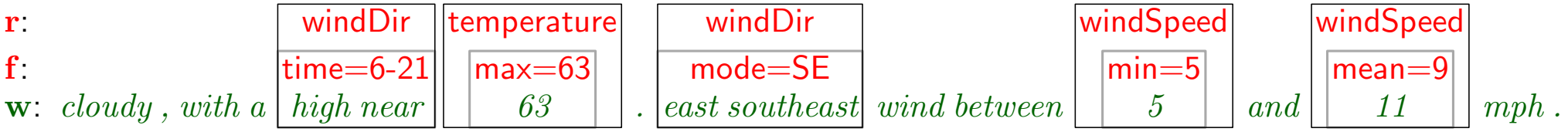
[Model 1] segmentation: ✗ coherence: ✗ (60.0 F_1)



+ aligns many content words

Experiments: weather forecasts (results)

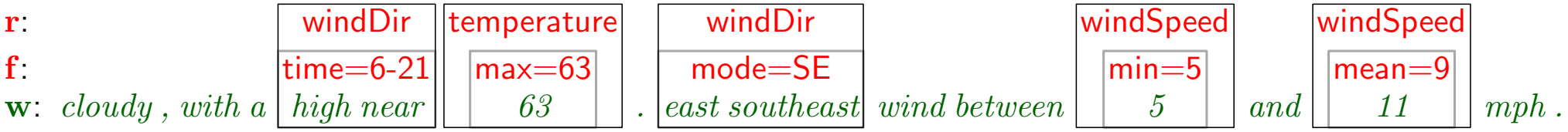
[Model 1] segmentation: ✗ coherence: ✗ (60.0 F_1)



\+ aligns many content words − misses *wind*, *cloudy*

Experiments: weather forecasts (results)

[Model 1] segmentation: ✗ coherence: ✗ (60.0 F_1)



+ aligns many content words − misses *wind*, *cloudy*

[Model 2] segmentation: ✓ coherence: ✗ (68.8 F_1)

Experiments: weather forecasts (results)

[Model 1] segmentation: ✗ coherence: ✗ (60.0 F_1)

r:		windDir		temperature		windDir		windSpeed		windSpeed		
f:		time=6-21		max=63		mode=SE		min=5		mean=9		
w:	*cloudy , with a*	*high near*		*63*	*.*	*east southeast*	*wind between*	*5*	*and*	*11*	*mph .*	

\+ aligns many content words − misses *wind*, *cloudy*

[Model 2] segmentation: ✓ coherence: ✗ (68.8 F_1)

r:	rainChance	temperature			windDir		windSpeed		
f:	mode=–	x	time=6-21	max=63	mode=SE	x	mean=9		
w:	*cloudy ,*	*with a*	*high near*	*63 .*	*east southeast wind*	*between 5 and*	*11 mph .*		

\+ aligns more continuously

Experiments: weather forecasts (results)

[Model 1] segmentation: ✗ coherence: ✗ (60.0 F_1)

r:		windDir		temperature		windDir		windSpeed		windSpeed		
f:		time=6-21		max=63		mode=SE		min=5		mean=9		
w:	*cloudy , with a*	*high near*		*63*	*.*	*east southeast*	*wind between*	*5*	*and*	*11*	*mph .*	

\+ aligns many content words − misses *wind*, *cloudy*

[Model 2] segmentation: ✓ coherence: ✗ (68.8 F_1)

r:	rainChance	temperature			windDir		windSpeed		
f:	mode=–	x	time=6-21	max=63	mode=SE	x	mean=9		
w:	*cloudy ,*	*with a*	*high near*	*63 .*	*east southeast wind*	*between 5 and*	*11 mph .*		

\+ aligns more continuously − wrong field structure for windSpeed

Experiments: weather forecasts (results)

[Model 1] segmentation: ✗ coherence: ✗ (60.0 F_1)

r:		windDir		temperature		windDir		windSpeed		windSpeed		
f:		time=6-21		max=63		mode=SE		min=5		mean=9		
w:	*cloudy , with a*	*high near*		*63*	*.*	*east southeast*	*wind between*	*5*	*and*	*11*	*mph .*	

\+ aligns many content words − misses *wind*, *cloudy*

[Model 2] segmentation: ✓ coherence: ✗ (68.8 F_1)

r:	rainChance	temperature			windDir		windSpeed		
f:	mode=–	x	time=6-21	max=63	mode=SE	x	mean=9		
w:	*cloudy ,*	*with a*	*high near*	*63 .*	*east southeast wind*	*between 5 and*	*11 mph .*		

\+ aligns more continuously − wrong field structure for windSpeed

[Model 3] segmentation: ✓ coherence: ✓ (**75.0** F_1)

Experiments: weather forecasts (results)

[Model 1] segmentation: ✗ coherence: ✗ (60.0 F_1)

r:		windDir		temperature		windDir		windSpeed		windSpeed		
f:		time=6-21		max=63		mode=SE		min=5		mean=9		
w:	*cloudy , with a*	*high near*		*63*	*.*	*east southeast*	*wind between*	*5*	*and*	*11*	*mph .*	

\+ aligns many content words − misses *wind*, *cloudy*

[Model 2] segmentation: ✓ coherence: ✗ (68.8 F_1)

r:	rainChance	temperature			windDir		windSpeed		
f:	mode=–	x	time=6-21	max=63	mode=SE	x	mean=9		
w:	*cloudy ,*	*with a*	*high near*	*63 .*	*east southeast wind*	*between 5 and*	*11 mph .*		

\+ aligns more continuously − wrong field structure for windSpeed

[Model 3] segmentation: ✓ coherence: ✓ (**75.0** F_1)

r:	skyCover	temperature				windDir		windSpeed			
f:	x	x	time=6-21	max=63	mean=56	mode=SE	x	min=5	max=13	x	
w:	*cloudy ,*	*with a*	*high near*	*63*	*.*	*east southeast*	*wind between*	*5*	*and 11*	*mph .*	

\+ fixes field structure

Experiments: weather forecasts (results)

[Model 1] segmentation: ✗ coherence: ✗ (60.0 F_1)

r:		windDir		temperature		windDir		windSpeed		windSpeed		
f:		time=6-21		max=63		mode=SE		min=5		mean=9		
w:	*cloudy , with a*	*high near*		*63*	*.*	*east southeast*	*wind between*	*5*	*and*	*11*	*mph .*	

\+ aligns many content words − misses *wind*, *cloudy*

[Model 2] segmentation: ✓ coherence: ✗ (68.8 F_1)

r:	rainChance	temperature			windDir		windSpeed		
f:	mode=–	x	time=6-21	max=63	mode=SE	x	mean=9		
w:	*cloudy ,*	*with a*	*high near*	*63 .*	*east southeast wind*	*between 5 and*	*11 mph .*		

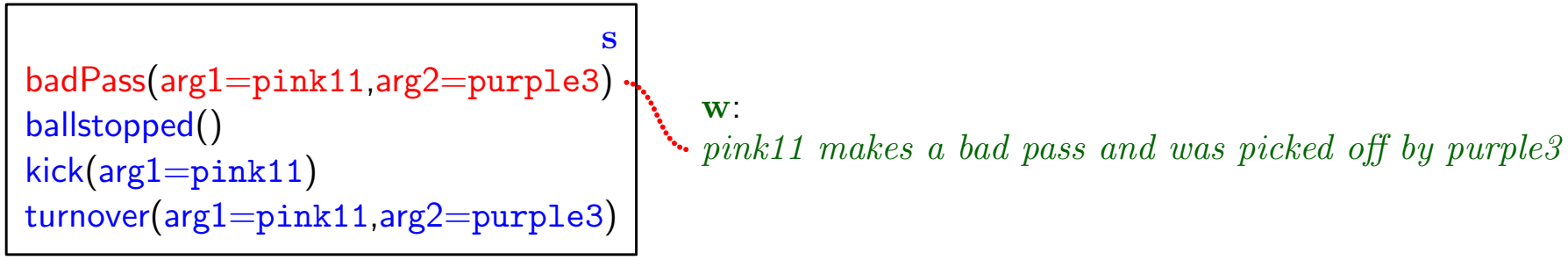
\+ aligns more continuously − wrong field structure for windSpeed

[Model 3] segmentation: ✓ coherence: ✓ (**75.0** F_1)

r:	skyCover	temperature				windDir		windSpeed			
f:	x	x	time=6-21	max=63	mean=56	mode=SE	x	min=5	max=13	x	
w:	*cloudy ,*	*with a*	*high near*	*63*	*.*	*east southeast*	*wind between*	*5*	*and 11*	*mph .*	

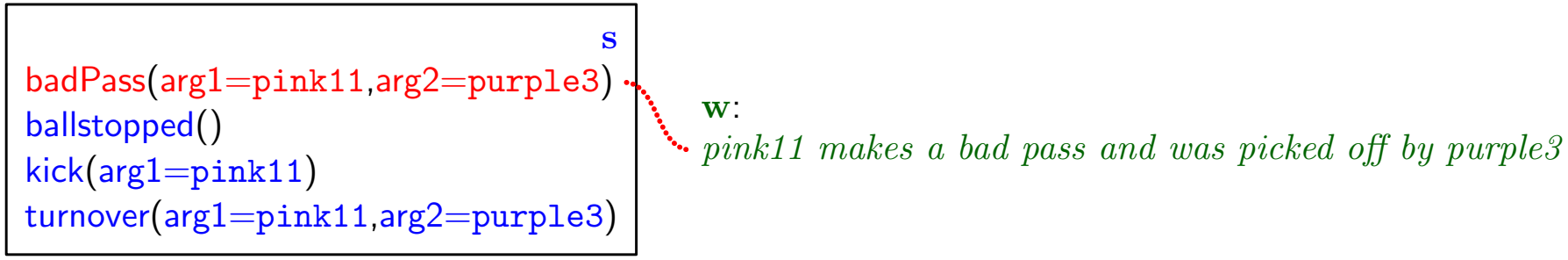
\+ fixes field structure − mean=56 garbage collects .
− time=6-21 correlated with *high*

Experiments: Robocup sportscasting



Dataset [Chen & Mooney, 2008]

Experiments: Robocup sportscasting



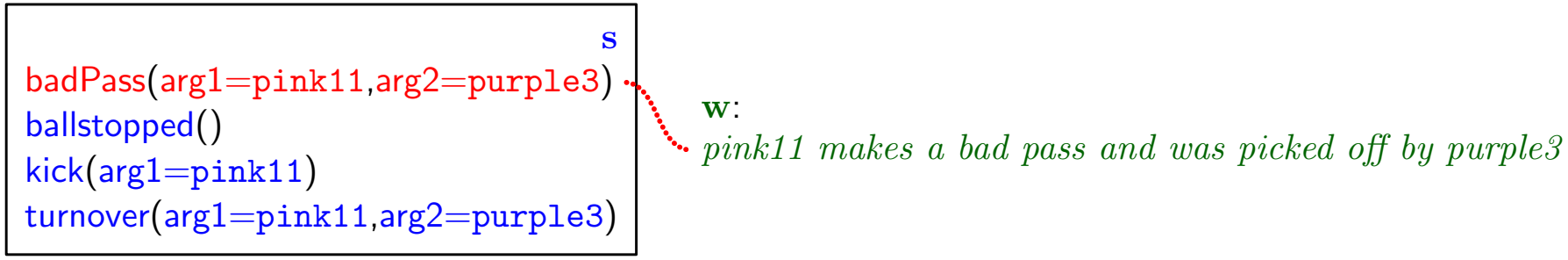
Dataset [Chen & Mooney, 2008]

1.9K scenarios

Each scenario contains

$|\mathbf{s}| = 2.4$ records, $|\mathbf{w}| = 5.7$ words/text, 0.8 gold alignments

Experiments: Robocup sportscasting



Dataset [Chen & Mooney, 2008]

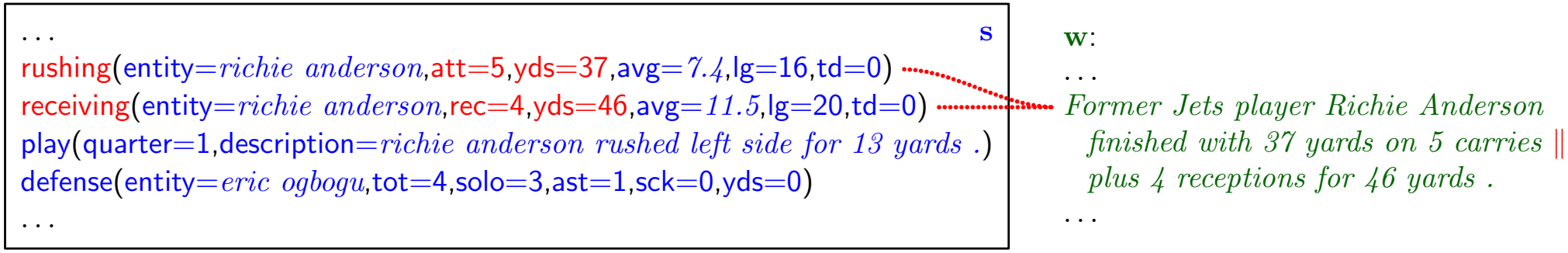
1.9K scenarios

Each scenario contains

$|\mathbf{s}| = 2.4$ records, $|\mathbf{w}| = 5.7$ words/text, 0.8 gold alignments

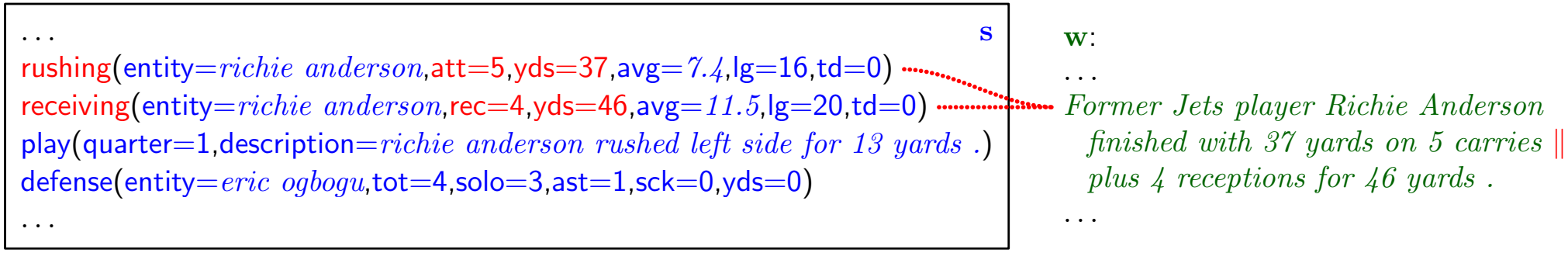
Method	F_1
Random	48.0
Chen & Mooney (2008)	67.0
Model 3	**75.7**

Experiments: NFL recaps



Dataset [Barzilay & Lapata, 2005; Snyder & Barzilay, 2007]

Experiments: NFL recaps



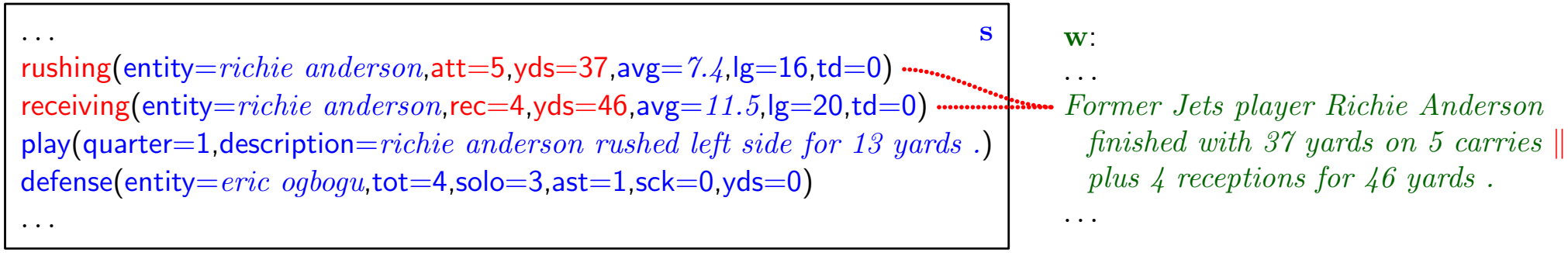
Dataset [Barzilay & Lapata, 2005; Snyder & Barzilay, 2007]

78 scenarios

Each scenario contains

$|\mathbf{s}| = 329$ records, $|\mathbf{w}| = 969$ words/text, 24.3 gold alignments

Experiments: NFL recaps



Dataset [Barzilay & Lapata, 2005; Snyder & Barzilay, 2007]

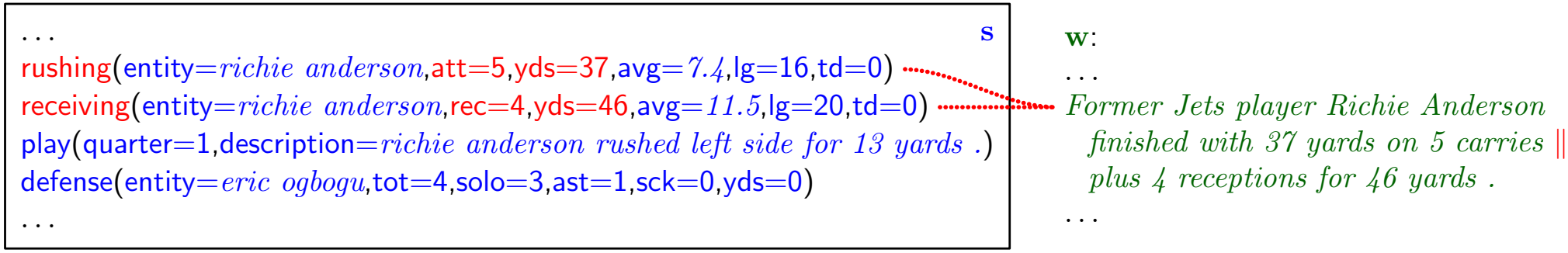
78 scenarios

Each scenario contains

$|\mathbf{s}| = 329$ records, $|\mathbf{w}| = 969$ words/text, 24.3 gold alignments

Method	F_1	
Random	2.4	} unsupervised
Baseline	26.7	
Model 2*	**53.2**	

Experiments: NFL recaps



Dataset [Barzilay & Lapata, 2005; Snyder & Barzilay, 2007]

78 scenarios

Each scenario contains

$|\mathbf{s}| = 329$ records, $|\mathbf{w}| = 969$ words/text, 24.3 gold alignments

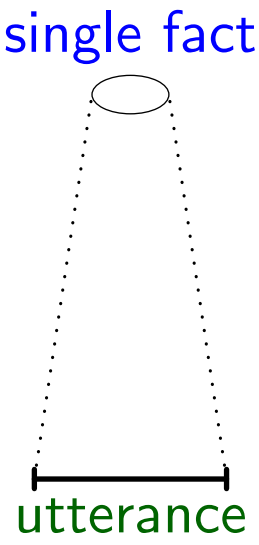
Method	F_1	
Random	2.4	unsupervised
Baseline	26.7	unsupervised
Model 2*	**53.2**	unsupervised
Graph matching	68.6	supervised
Multilabel global	**80.3**	supervised

Summary

more supervision

less supervision

Summary

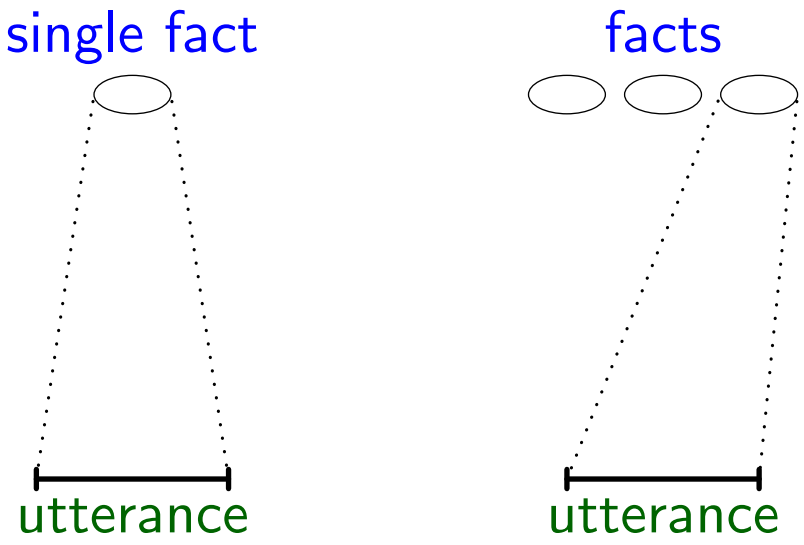


[Ge & Mooney, 2005]
[Zettlemoyer & Collins, 2005]
[Lu et al., 2008]

more supervision

less supervision

Summary



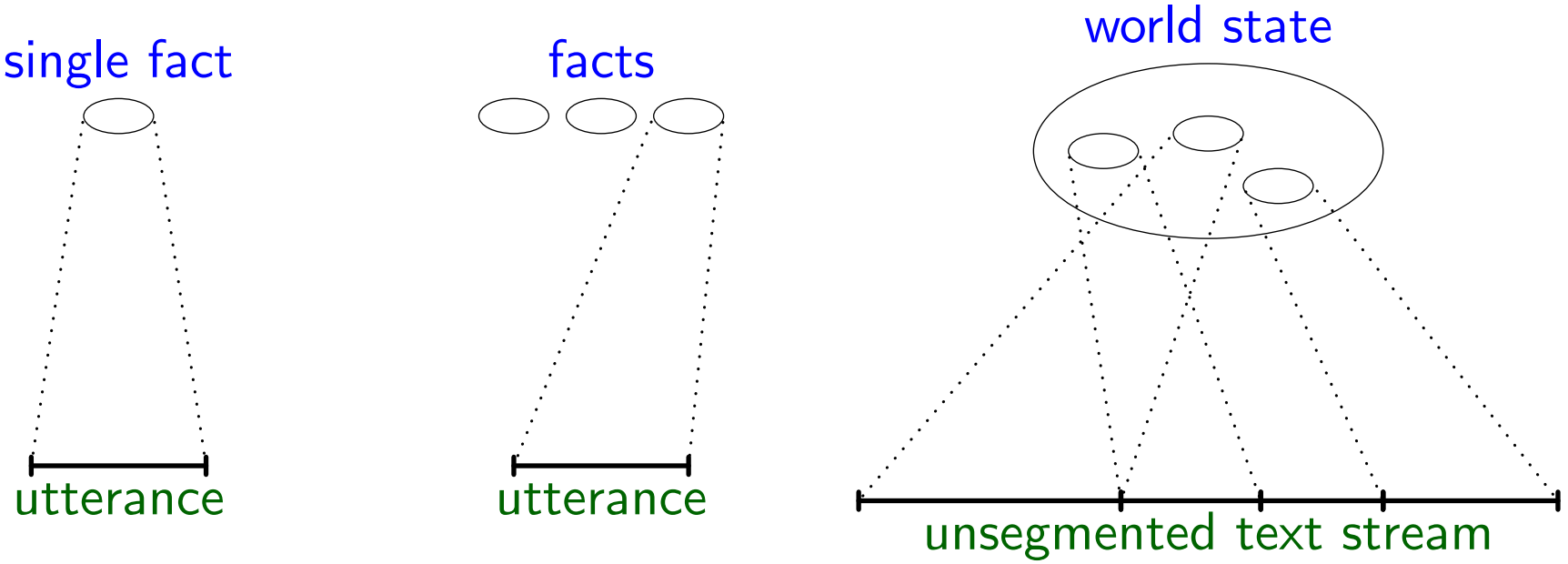
[Ge & Mooney, 2005]
[Zettlemoyer & Collins, 2005]
[Lu et al., 2008]

[Kate & Mooney, 2007]
[Chen & Mooney, 2008]

more supervision

less supervision

Summary



[Ge & Mooney, 2005]
[Zettlemoyer & Collins, 2005]
[Lu et al., 2008]

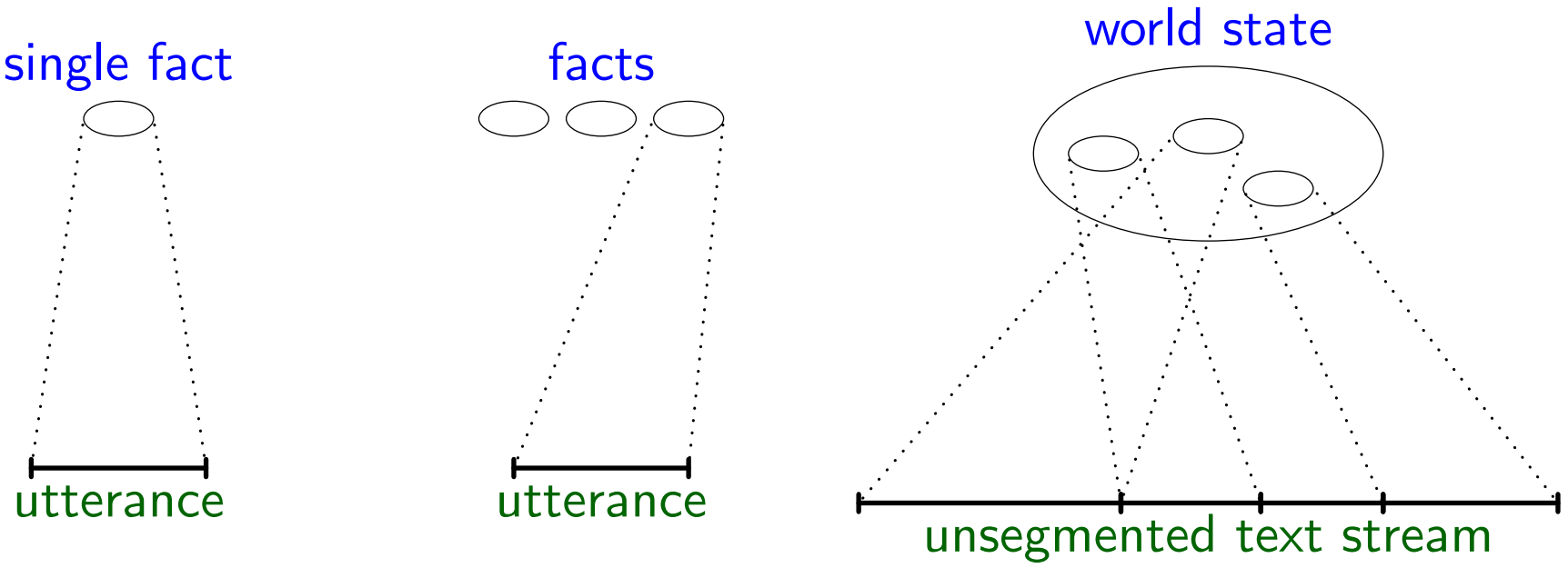
[Kate & Mooney, 2007]
[Chen & Mooney, 2008]

[Liang et al., 2009]

more supervision

less supervision

Summary



Our approach:

joint model capturing segmentation and coherence