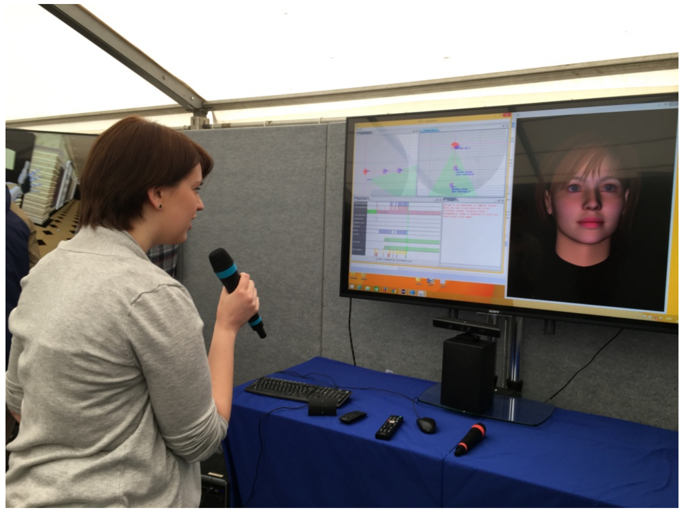
Domain Adaptation using Linguistic Knowledge

Nina Dethlefs
School of Engineering and Computer Science
University of Hull, UK

Let's Discuss: Learning Methods for Dialogue

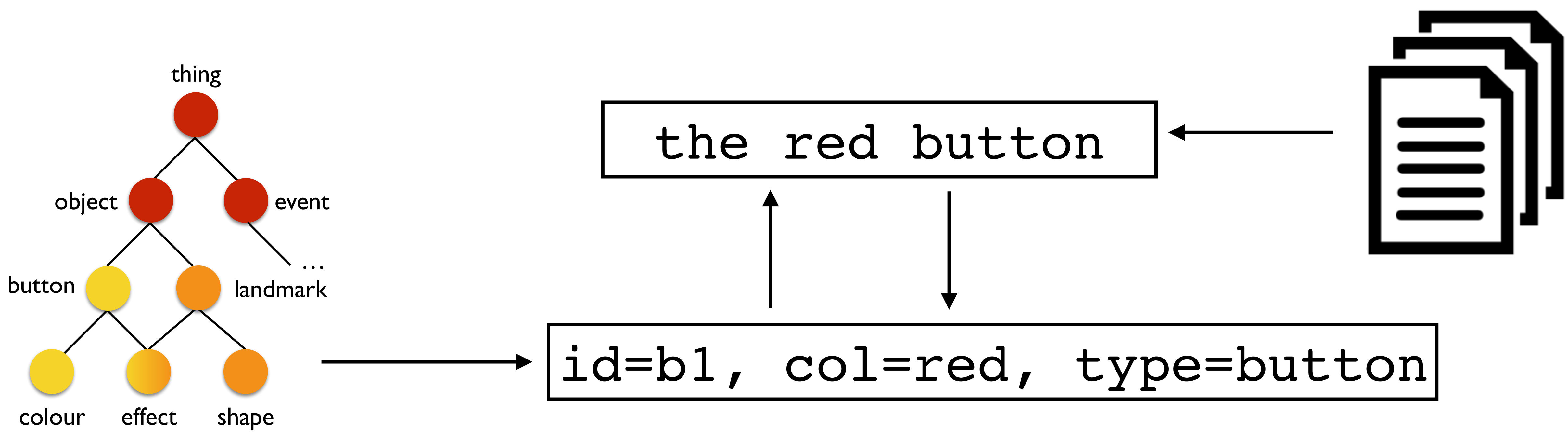
Data-driven NLG

Mostly within spoken dialogue systems



Natural Language Learning

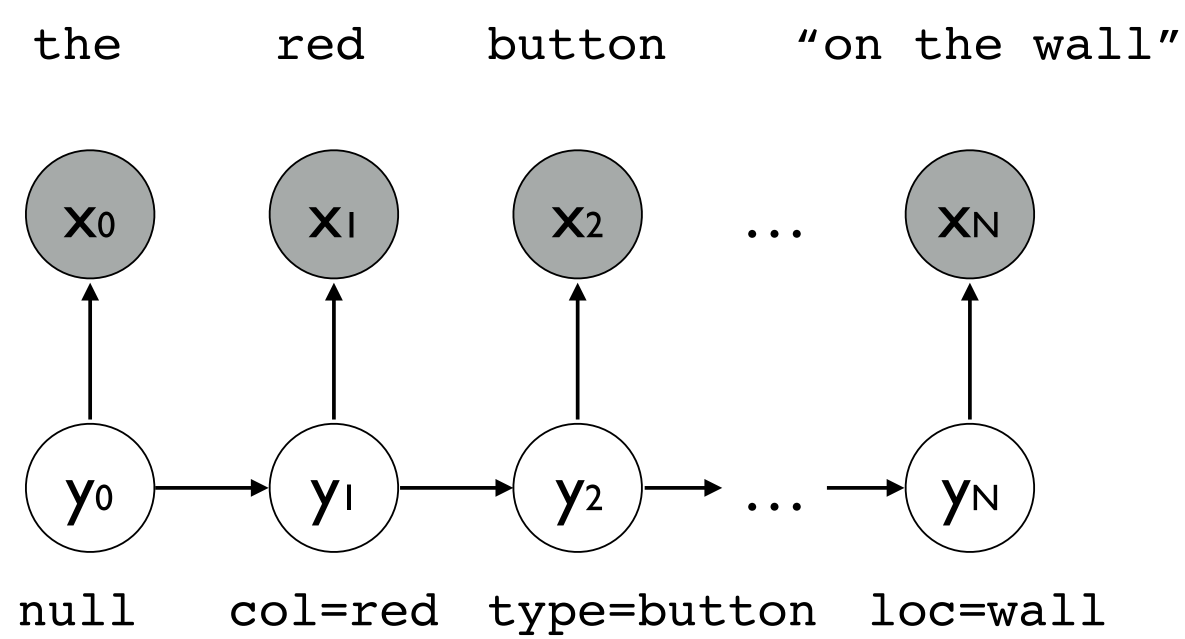
Task: learn a mapping between a sequence of semantic symbols and a sequence of words.



Learnt models can be used for language understanding and language generation (production).

Supervised Natural Language Generation

Task: learn a mapping between a sequence of semantic symbols and a sequence of words.

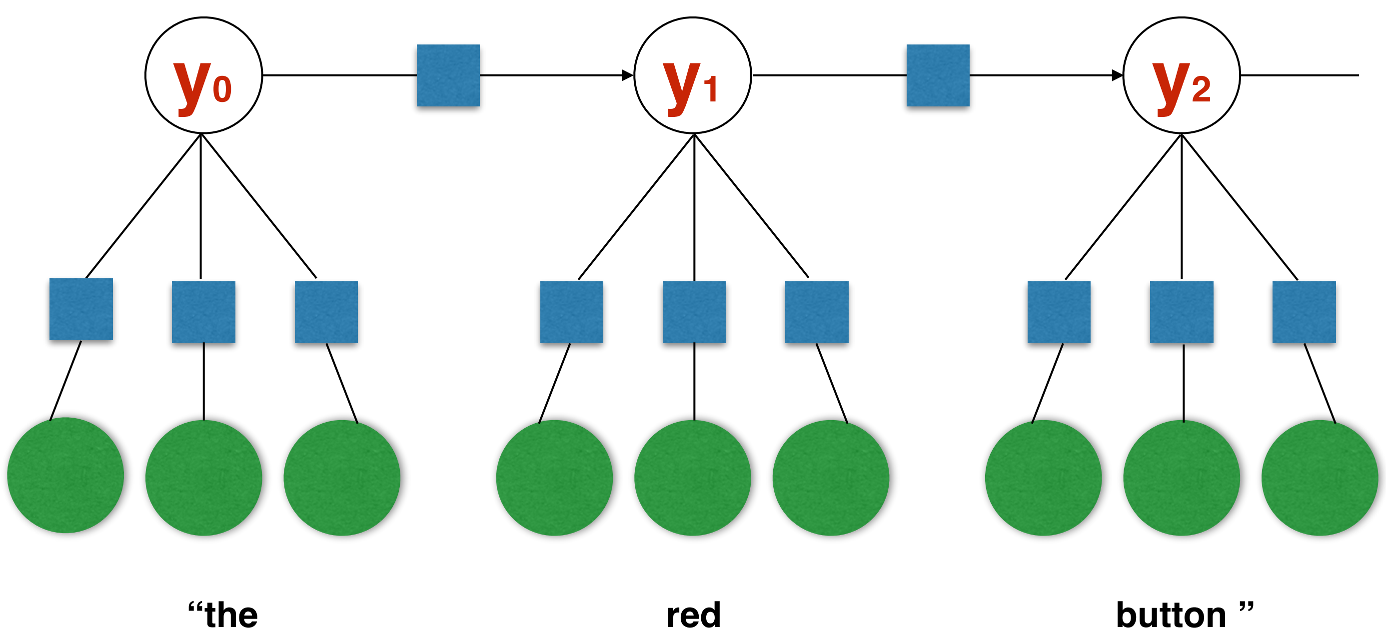


Approach 1: use a graphical model to (1) represent mapping from observed to hidden nodes and (2) model transitions between observed nodes.

Dethlefs and Cuayahuitl, 2011

Supervised Natural Language Generation

Task: learn a mapping between a sequence of semantic symbols and a sequence of words.



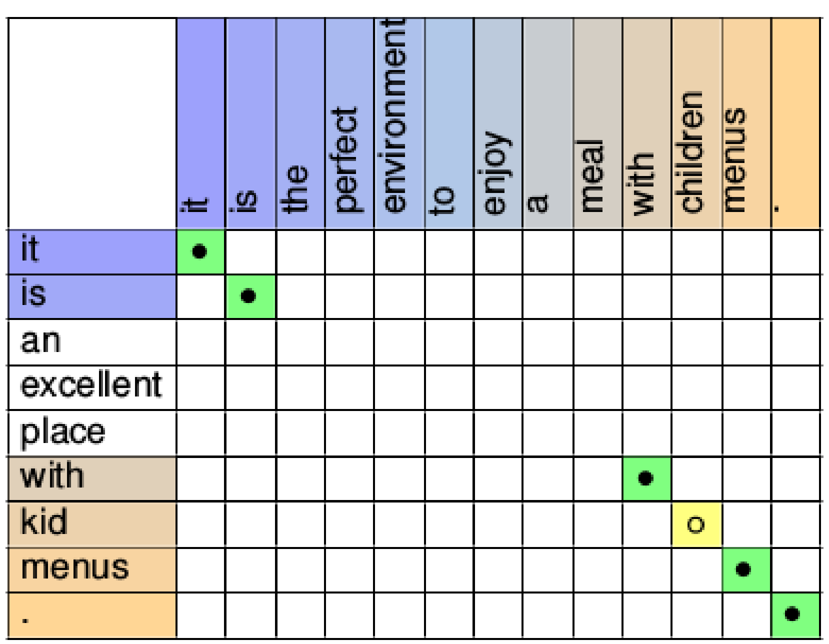
Approach 2: use a different graphical model to capture longer-range dependencies between symbols.

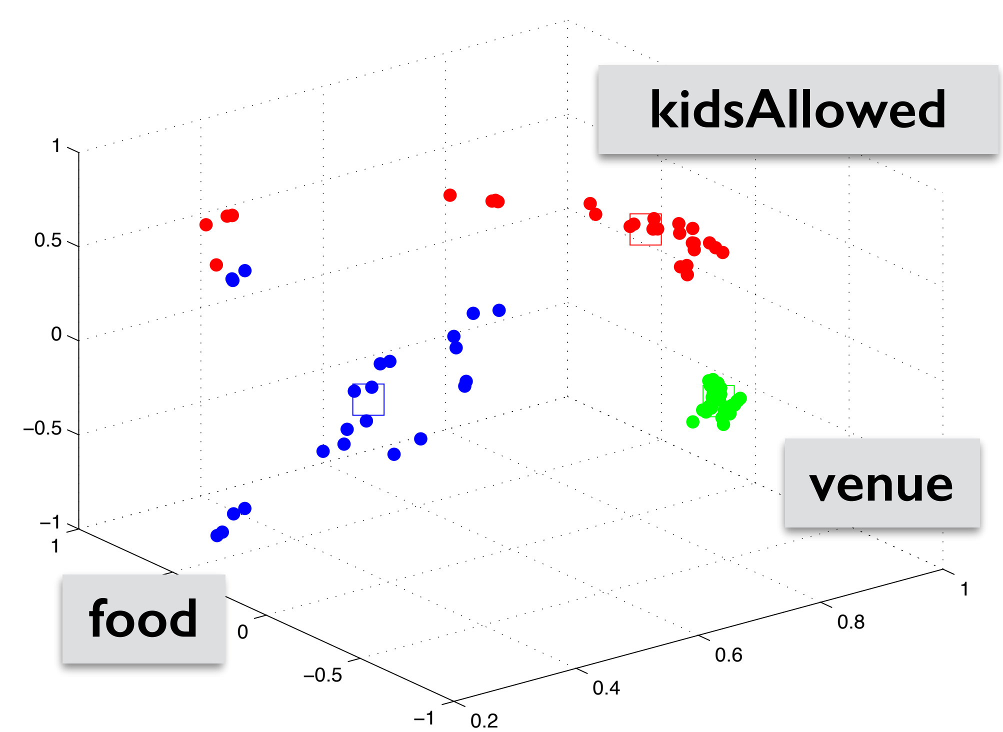
Dethlefs et al, 2013

Unsupervised Natural Language Generation

Various methods to avoid annotations and map raw data to semantic representations

Clustering: mining a structured dataset for similar phrases



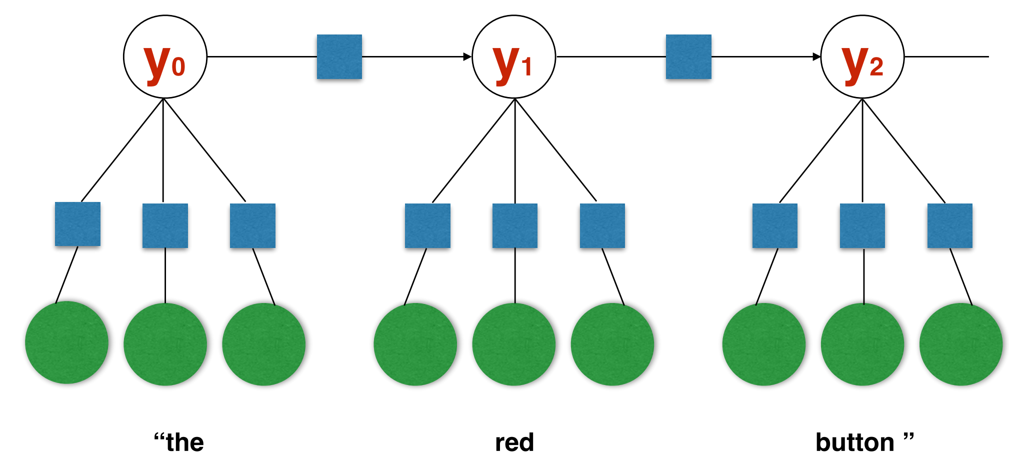


Cuayahuitl et al. 2012

Common observations

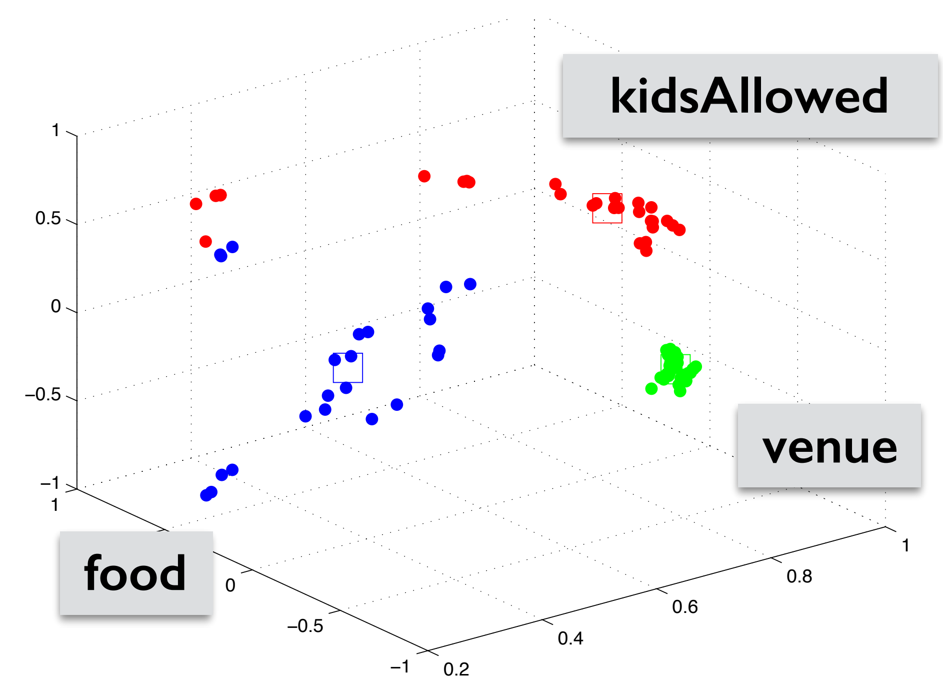
Supervised learning

- no direct mapping between semantics and words
- data sparsity
- domain-specific inputs

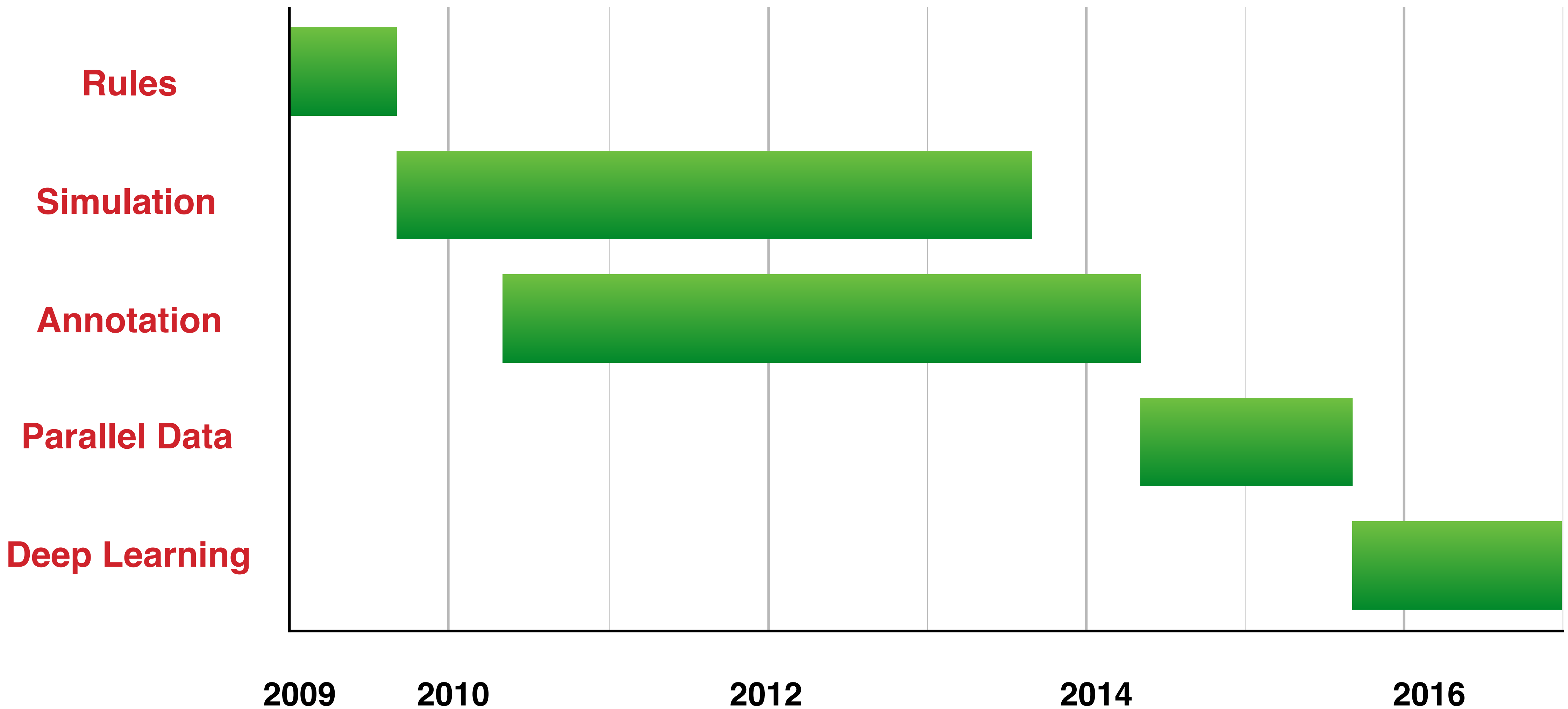


Unsupervised learning

- brittle due to data sparsity
- many models rely on parallel data
- still domain-specific inputs



Minimising Human Input?

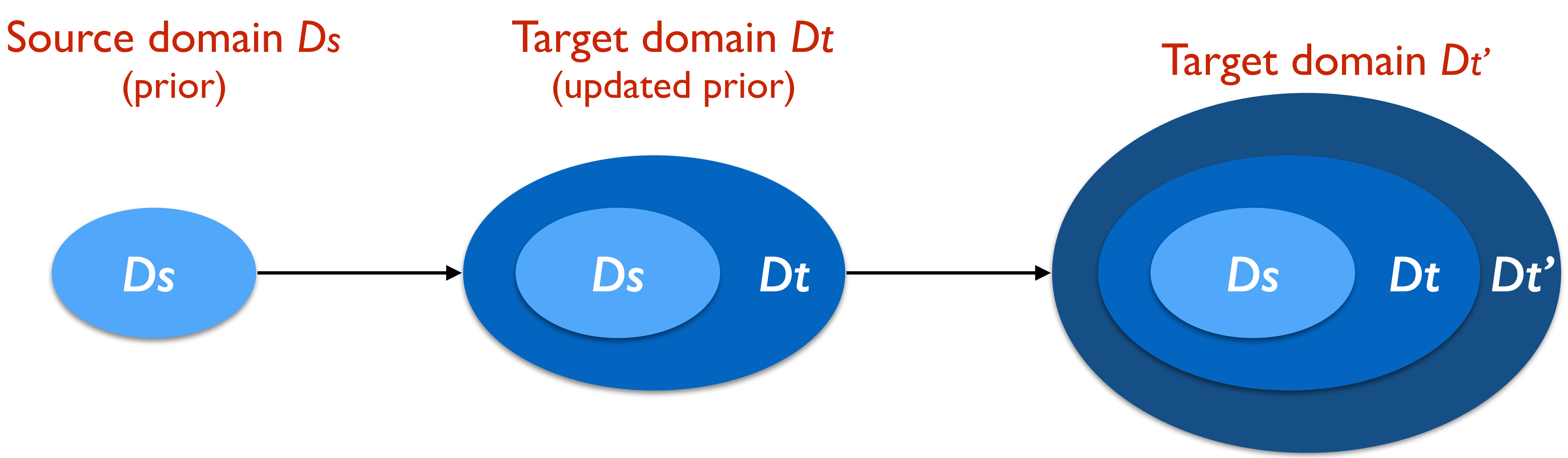


Domain Adaptation in NLG

Now trying some deep learning

Domain adaptation

- Estimate a model for target domain from source domain
- Semantic similarity can identify suitable source domains



Domain Adaptation in NLP

- Not a new idea in other areas of NLP: machine translation, NER, capitalisation, shallow parsing

- Approaches often reuse prior knowledge

- Transfer probs from a source to a target domain, and then adapt them (Chelba and Acero, 2006; Daume 2007, many more since)

- Normally rely on a common feature representation

Data-driven NLG / NLU

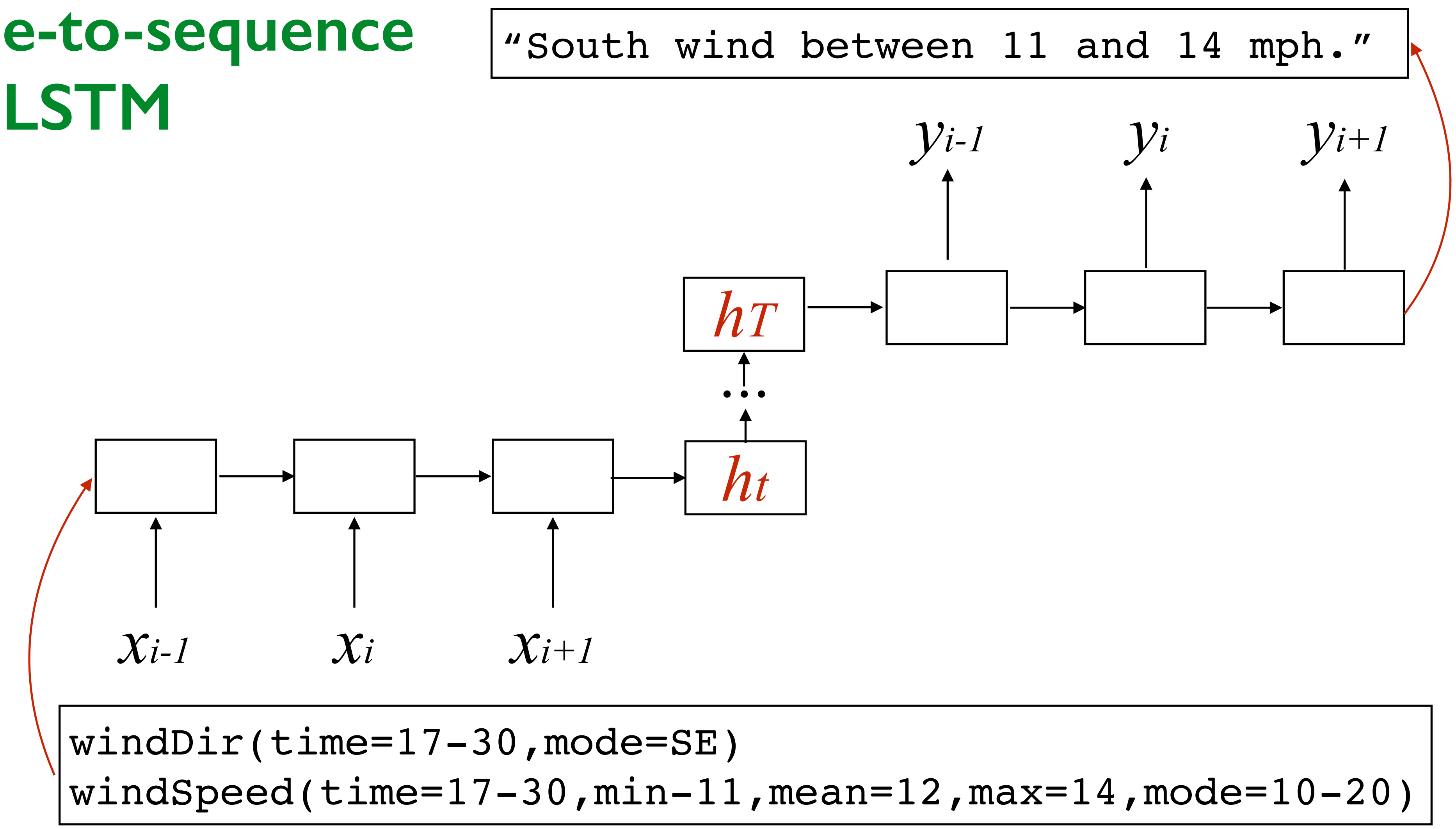
- Recent work on NLG across domains using synthetic data and shared input representations (Wen et al., 2016)

- Natural language understanding based on canonical correlation analysis of labels across domains (Kim et al, 2015; earlier similar ideas)

- Related to work on parallel data: RoboCup (Chen et al. 2010), WeatherGov (Angeli et al., 2010), ATIS (Konstas and Lapata 2012)

Deep Learning and Language Generation

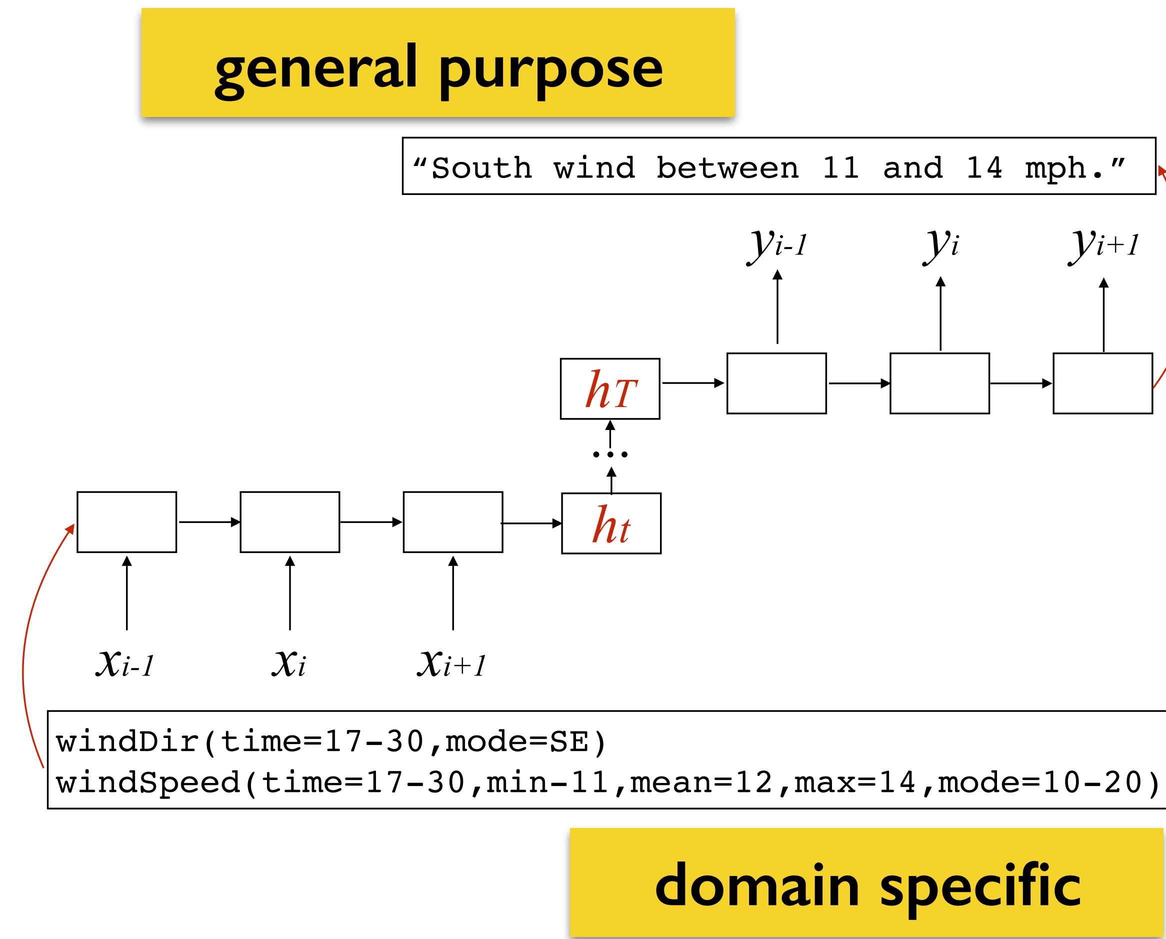
Sequence-to-sequence LSTM



Learn conditional prob. distribution from input to output.

Deep Learning and Language Generation

- Models rely on idiosyncratic input representations
- Often require delexicalised semantic inputs
- Data can be noisy (typos, spelling, etc.)



Comparison of inputs - Flat

Flat input = inform(name='beijing restaurant', food=chinese)

Realisations:

1. Beijing restaurant serves Chinese food.
2. Beijing restaurant is a nice place. It serves Chinese food.
3. Beijing restaurant is a Chinese restaurant.

No similarity with…
The blue cube touches the red sphere.

Learning Constructions

Constructions in cognitive grammar are conventionalised form-meaning pairs at different levels of abstraction: (idioms), partially lexically filled, fully general phrases.

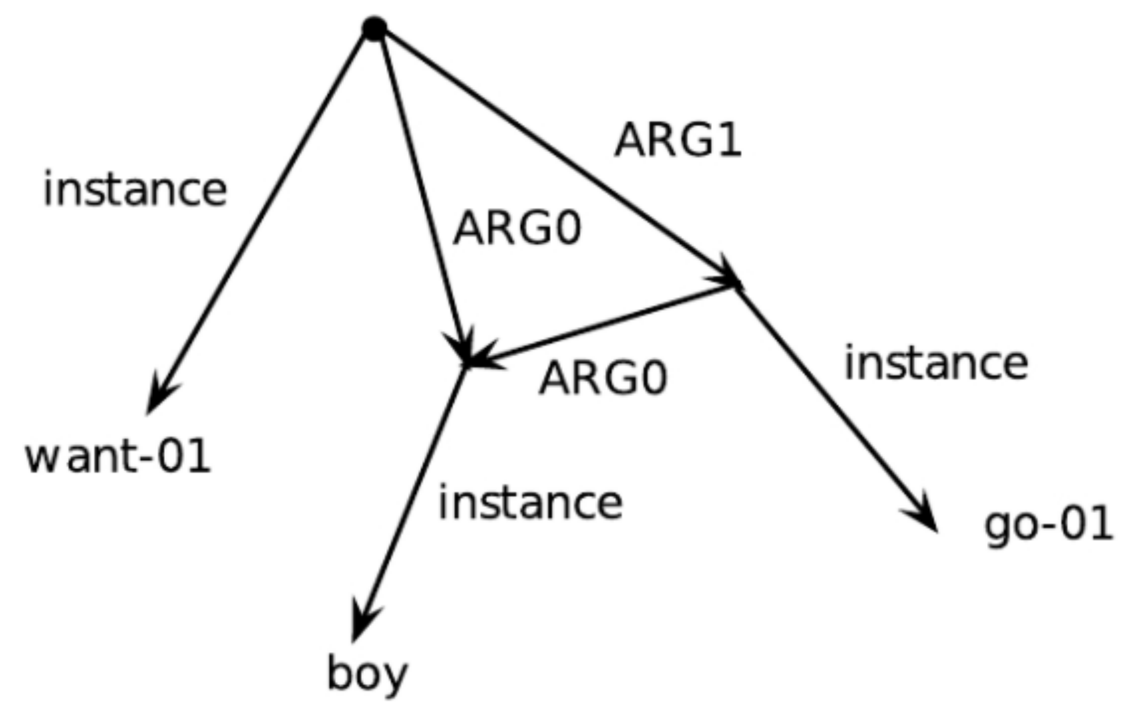
1. Elmo the car gopping. (SOV)
2. Dacking Elmo the car. (VSO)
3. Elmo blicking the car. (SVO)

Can we use constructions in our **x** to **y** mappings?

Goldberg, 2005

Abstract meaning representations

```
(w / want-01
 :ARG0 (b / boy)
 :ARG1 (g / go-01
          :ARG0 b))
```



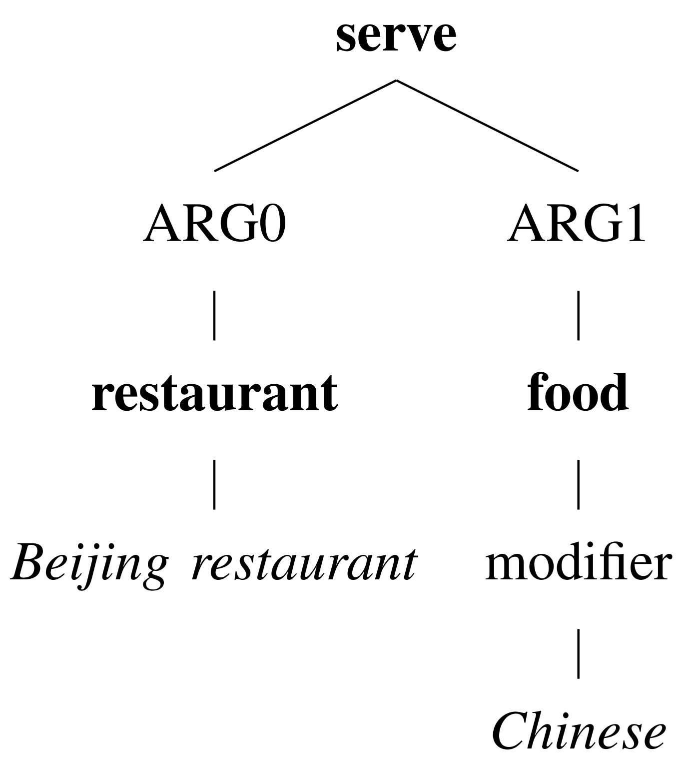
- Semantic representations as directed graphs
- A shared set of semantic roles, named entities, modality, negation, questions, quantities, etc.

Comparison of inputs - AMR

Flat input = inform(name='beijing restaurant', food=chinese)

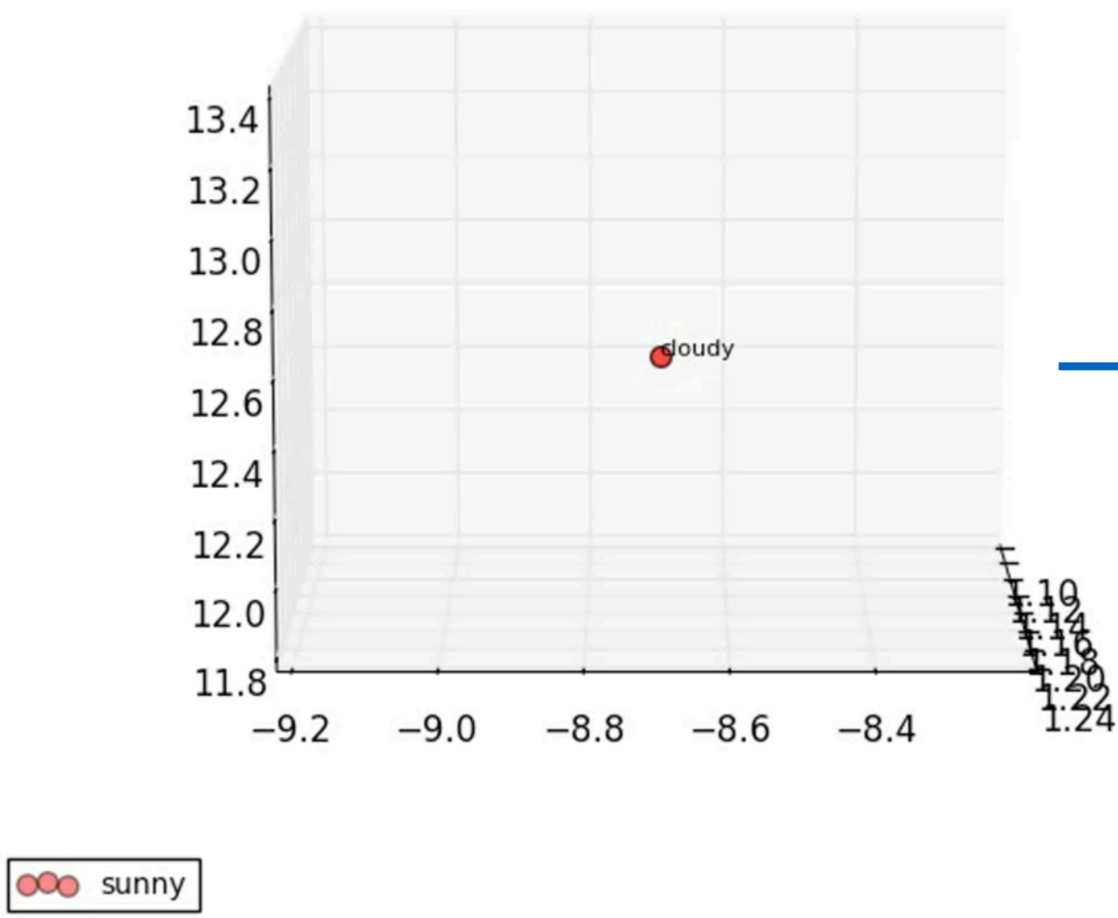
AMR:

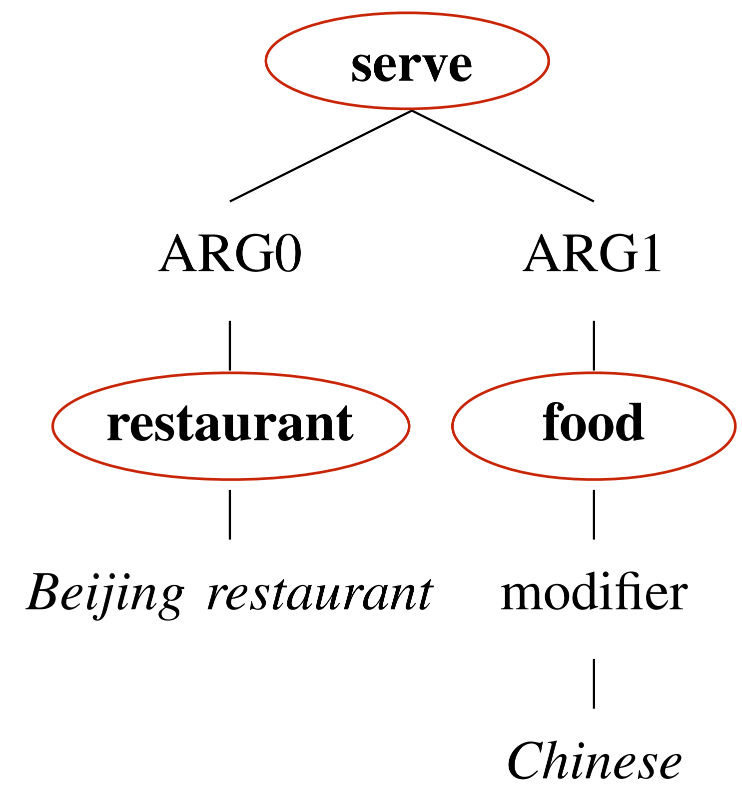
```
(s / serve
     :arg0 (n / restaurant
               :name 'beijing_restaurant')
     :arg1 (f / food :mod chinese)
)
```



Realisations: Beijing restaurant serves Chinese food.

Making AMRs





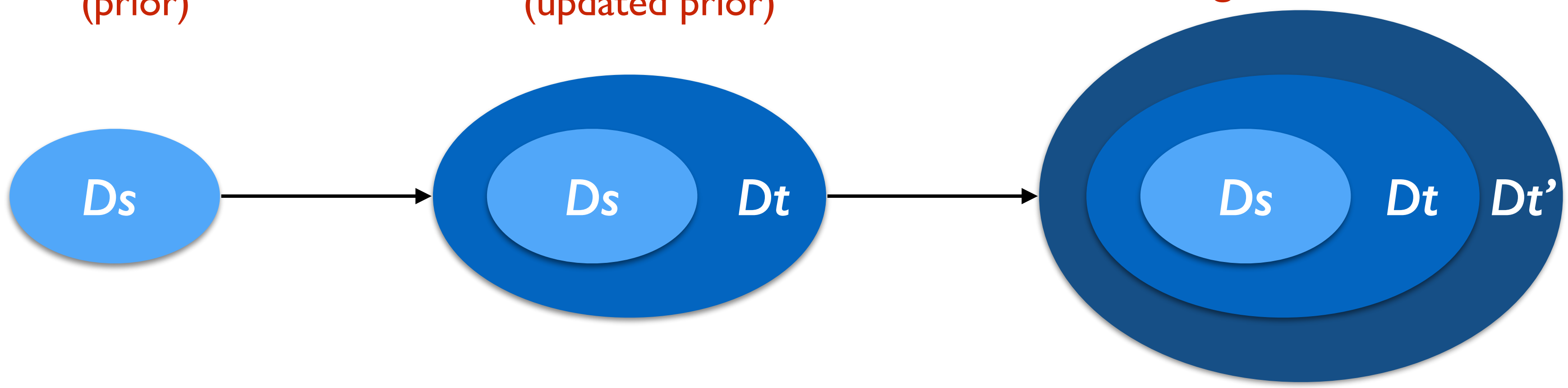
Semantic classes have the same identifier
if they occur in the same context.

Domain adaptation

Source domain *Ds* (prior)

Target domain *Dt* (updated prior)

Target domain *Dt'*



annotate, train, generate → generate, train, generate → generate, train, generate

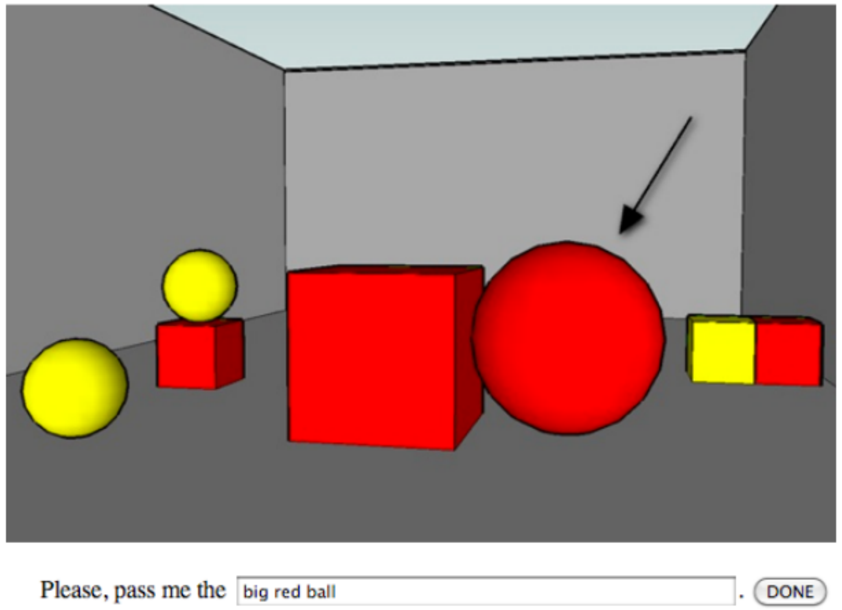
Hypothesis: using AMRs + deep learning, we can learn enough lexical-syntactic patterns to generate language in unseen target domains.

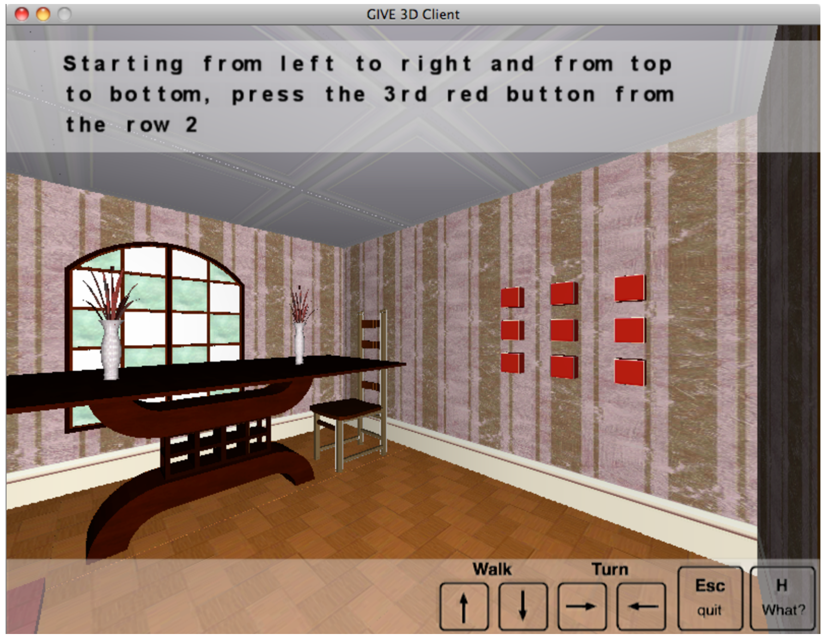
Datasets

	GRE-7	GIVE	RESTAURANTS	HOTELS
	full	full	full	10%
Examples	4,480	1,756	6,198	638
Vocabulary	195	467	2,058	237
Ave expl length	3.50	3.30	12.91	11.51
NPs	4,969	706	6,094	609
Spatial relations	1,084	1,716	956	159
Transitive cl.	45	294	2,597	186
Intransitive cl.	0	17	872	94
Relative cl.	15	37	598	20
Imperatives	0	1,043	3	0
Conjunctions	3	100	905	31

Datasets - GRE-7 and GIVE

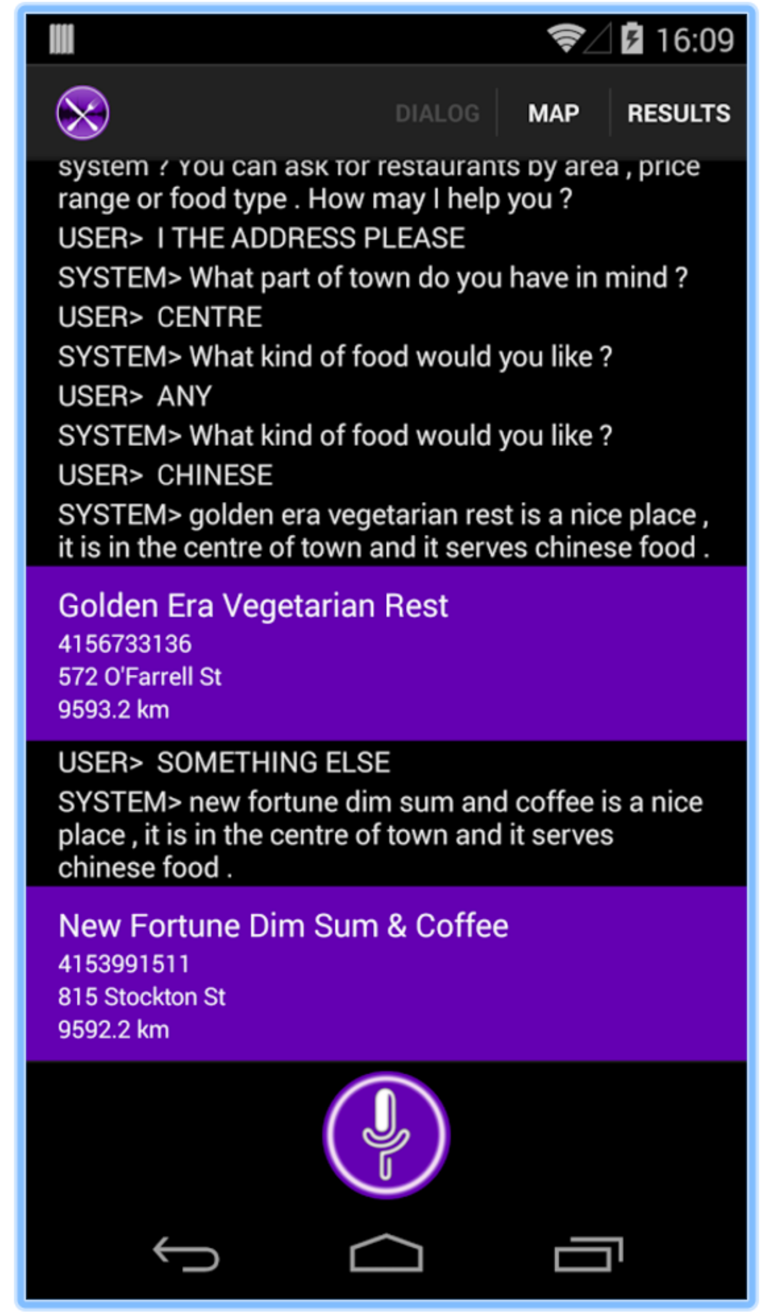
	GRE-7 (full)	GIVE (full)
Examples	4,480	1,756
Vocabulary	195	467
Ave expl length	3.50	3.30
NPs	4,969	706
Spatial relations	1,084	1,716
Transitive cl.	45	294
Intransitive cl.	0	17
Relative cl.	15	37
Imperatives	0	1,043
Conjunctions	3	100





Datasets - SFX-R and SFX-H

	RESTAURANT	HOTELS
Examples	6,198	638
Vocabulary	2,058	237
Ave expl length	12.91	11.51
NPs	6,094	609
Spatial relations	956	159
Transitive cl.	2,597	186
Intransitive cl.	872	94
Relative cl.	598	20
Imperatives	3	0
Conjunctions	905	31



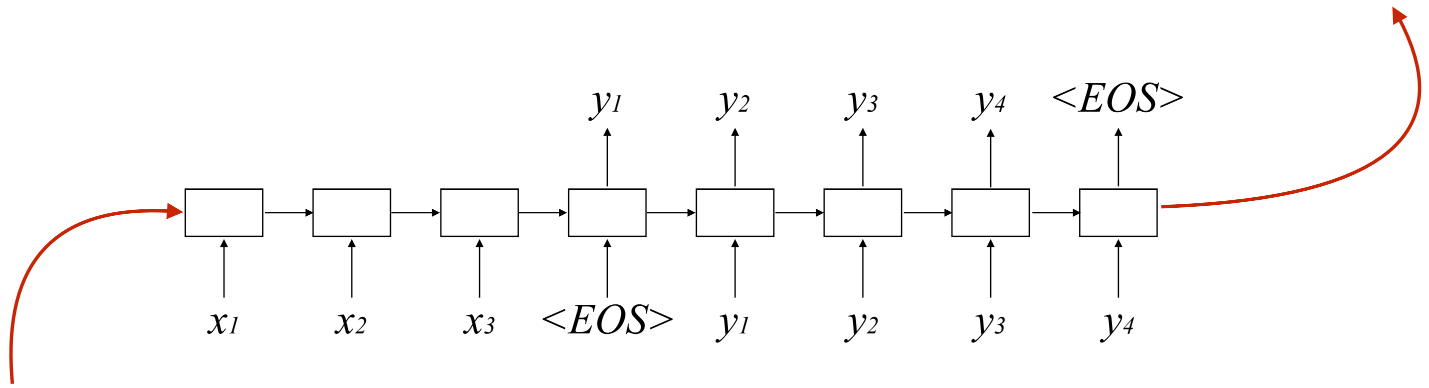
Datasets

	full	full	full	10%
	GRE-7	**GIVE**	**RESTAURANTS**	**HOTELS**
Examples	4,480	1,756	6,198	638
Vocabulary	195	467	2,058	237
Ave expl length	3.50	3.30	12.91	11.51
NPs	4,969	706	6,094	609
Spatial relations	1,084	1,716	956	159
Transitive cl.	45	294	2,597	186
Intransitive cl.	0	17	872	94
Relative cl.	15	37	598	20
Imperatives	0	1,043	3	0
Conjunctions	3	100	905	31

Deep Learning and Language Generation

Idea: improve performance of generators by leveraging general-purpose linguistic knowledge.

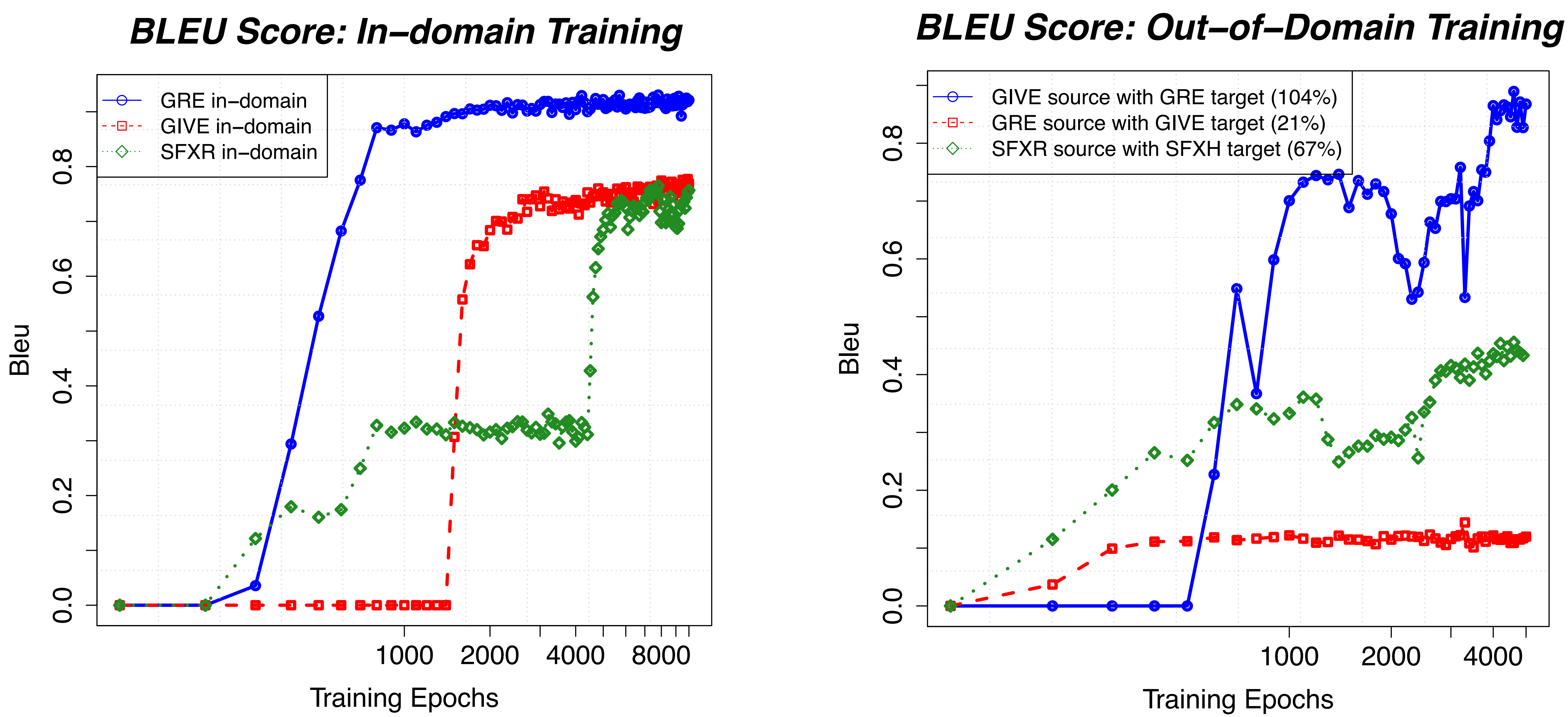
Beijing restaurant serves Chinese food.




```
(s / serve
     :arg0 (n / restaurant
             :name 'beijing_restaurant')
     :arg1 (f / food :mod chinese)
)
```

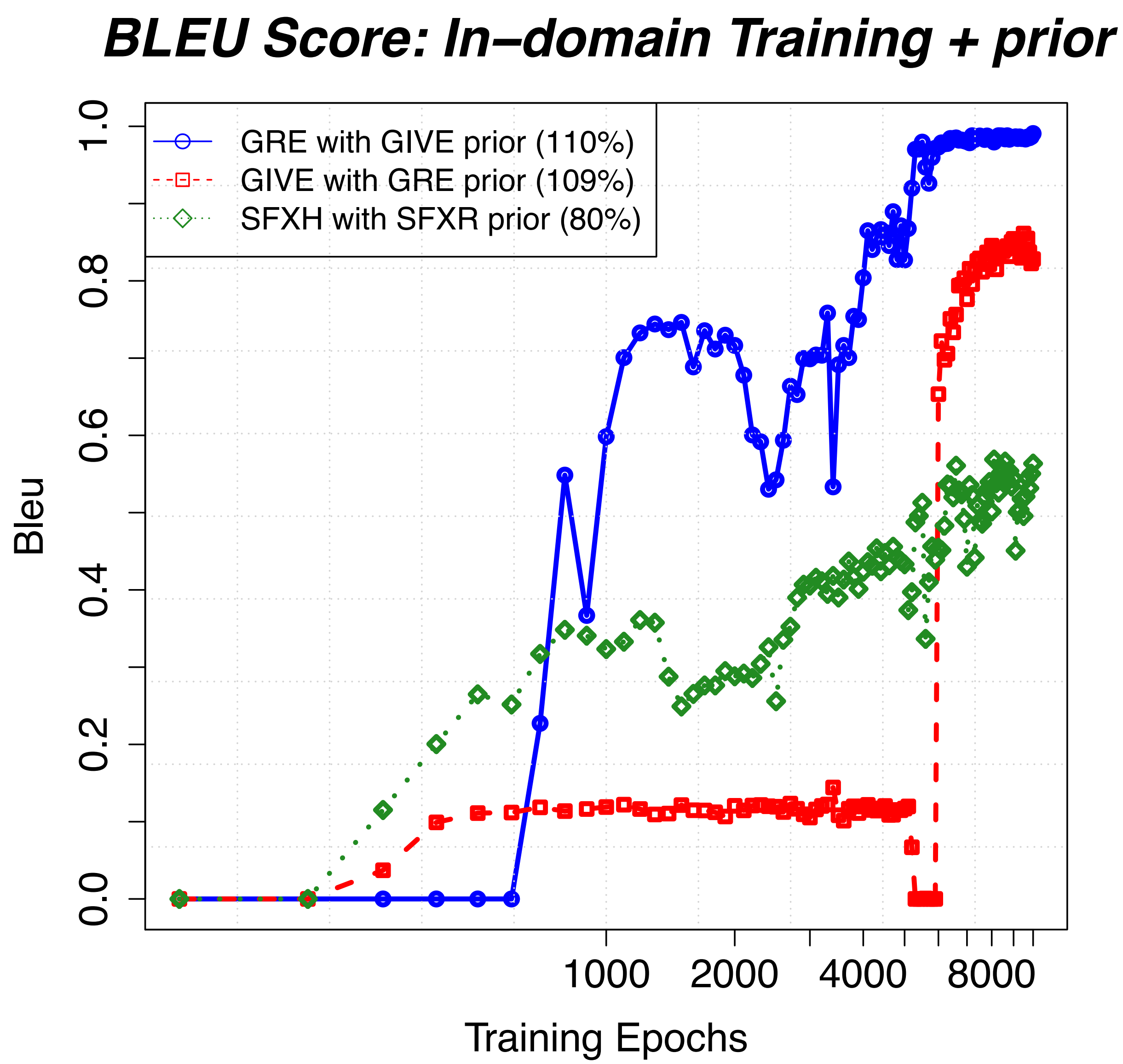
Trained 4-layered LSTMs with 50 hidden units, batch size 128 over 10,000 training epochs.

Results: Objective metrics



- Best SFX-R: 0.71 / Best SFX-H: 0.82 (Wen et al 2015) for in-domain; 0.48 for across domain (Wen et al 2016).

Results: Objective metrics



- Best SFX-R: 0.71 / Best SFX-H: 0.82 (Wen et al 2015) for in-domain; 0.48 for across domain (Wen et al 2016).

Results: Semantic error

- Semantic error computed for missing and superfluous slots

missing slots + extra slots / all slots

- Best SFX-R/H: 0.04 (in different domains); Wen et al. 2015/6
- Error consistently lower with prior than pure in-domain

	ERROR
GRE-HUM	0.00
GIVE-HUM	0.00
SFXR-HUM	0.00
SFX-H-HUM	0.00
GRE-IND	0.02
GIVE-IND	0.09
SFXR-IND	0.10
GRE-OUTD	0.04
GIVE-OUTD	0.29
SFX-H-OUTD	0.12
GRE-PRIOR	**0.01**
GIVE-PRIOR	**0.06**
SFX-H-PRIOR	**0.08**
SHARED-D	0.06

Results: Subjective metrics

- 163 human raters
- 3245 utterances ratings, randomly selected from a pool of 120 per model

"The utterance is natural, i.e. could have been produced by a human."

- Best GIVE: 3.36 (my PhD)
- Best SFX-R/H: 4.18 (rescaled)

	NATURAL
GRE-HUM	2.97 (4)
GIVE-HUM	3.77 (4)
SFXR-HUM	4.17 (4)
SFX-H-HUM	3.93 (4)
GRE-IND	3.12 (3)
GIVE-IND	2.76 (2)
SFXR-IND	**3.67 (4)**
GRE-OUTD	3.11 (3)
GIVE-OUTD	2.03 (2)
SFX-H-OUTD	3.45 (3)
GRE-PRIOR	**3.29 (3)**
GIVE-PRIOR	**3.09 (3)**
SFX-H-PRIOR	3.58 (4)
SHARED-D	2.66 (3)

Results: Linguistic Patterns

Complex NP with spatial relation and temporal adverb

```
(e1 / obj :time (n / now)
      :domain (b1 / obj :mod property
      :location (l / obj
      :op1 ( n / on [near, by]))) )
```

Now, the blue circle on the green sphere.

Now, the green button by the window.

Now, an Indian restaurant near Pacific Heights.

Results: Linguistic Patterns

Imperative clause construction (with relative clause)

```
(e1 / event :arg0 (y / you)
      :arg1 (b1 / obj :mod property
      :location (w /obj :op1 (o1 / on [in]) ))
      :mode imperative )
```

Click the red button (that is) on the wall.

Try Chinese restaurant Kirin in the Pacific Heights area.

Results: Linguistic Patterns

Transitive clause construction

```
(e1 / event :arg0 (b1 / obj :mod property)
:arg1 (b2 / obj :mod property))
```

The yellow sphere (that is) touching the blue box.

Source restaurant serves Italian food.

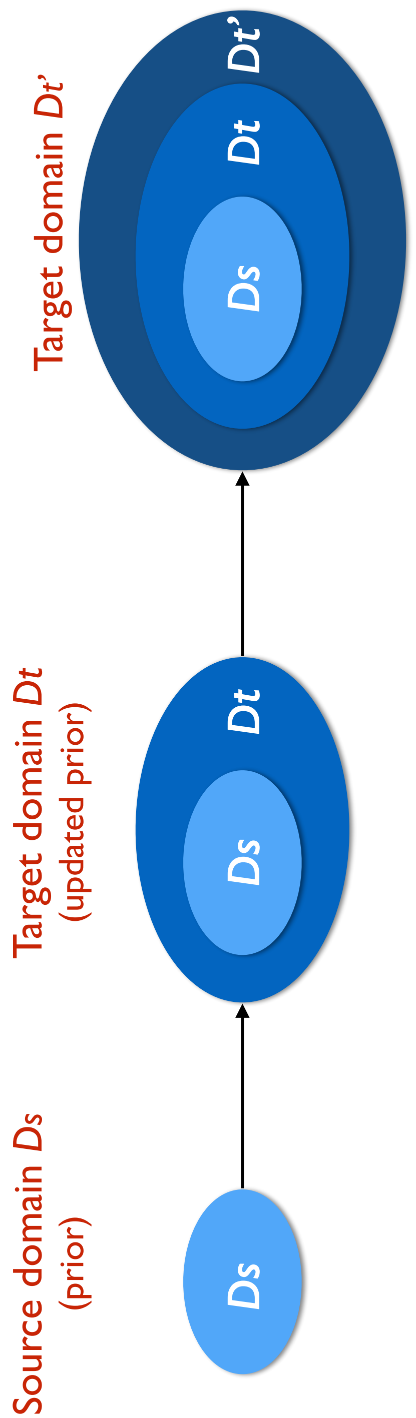
Did it work?

Yes:

- Model learns lexical-syntactic patterns from AMRs that are not domain-specific
- Out-of-domain results are acceptable in some domains, similar to in-domain training in others
- Learning from source domain priors gives us up to 10% extra performance

Limitations:

- AMR authoring is an overhead.



Future work

Improve the deep learning model

Use existing AMR annotations to annotate domains automatically

Find constructions automatically