Learning Sentence Planning Rules with Bayesian Methods

David M. Howcroft

Department of Language Science and Technology
Saarland Informatics Campus, Saarland University, Germany

24 October 2018

Natural Language Generation

How do we transform **non-linguistic input**...

name	price	cuisine
Due Fratelli	\$\$	Italian
Andalucia	\$\$\$	Spanish, Seafood

...into a **natural language text**?

Due Fratelli is an Italian restaurant, while Andalucia is a Spanish seafood restaurant. However, Due Fratelli's price is average, while Andalucia's price is more expensive.

Templates and traditional approaches to NLG

name	price	cuisine
Due Fratelli	$$	Italian
Andalucia	$$$	Spanish, Seafood

Templates and traditional approaches to NLG

name	price	cuisine
Due Fratelli	$$	Italian
Andalucia	$$$	Spanish, Seafood

Simplest case: use templates!

Templates and traditional approaches to NLG

name	price	cuisine
Due Fratelli	$$	Italian
Andalucia	$$$	Spanish, Seafood

Simplest case: use templates!

`inform(NAME, CUISINE)`

Templates and traditional approaches to NLG

name	price	cuisine
Due Fratelli	$$	Italian
Andalucia	$$$	Spanish, Seafood

Simplest case: use templates!

`inform(NAME, CUISINE)`

→ "NAME is a CUISINE restaurant"

Templates and traditional approaches to NLG

name	price	cuisine
Due Fratelli	$$	Italian
Andalucia	$$$	Spanish, Seafood

Simplest case: use templates!

`inform(NAME, CUISINE)`

$\rightarrow$ "NAME is a CUISINE restaurant"

$\rightarrow$ "NAME serves CUISINE food"

Templates and traditional approaches to NLG

name	price	cuisine
Due Fratelli	$$	Italian
Andalucia	$$$	Spanish, Seafood

Simplest case: use templates!

`inform(NAME, CUISINE)`

→ "NAME is a CUISINE restaurant"

→ "NAME serves CUISINE food"

But this doesn't generalize; only good for limited applications!

Templates and traditional approaches to NLG

name	price	cuisine
Due Fratelli	$$	Italian
Andalucia	$$$	Spanish, Seafood

Simplest case: use templates!

`inform(NAME, CUISINE)`

$\rightarrow$ "NAME is a CUISINE restaurant"

$\rightarrow$ "NAME serves CUISINE food"

But this doesn't generalize; only good for limited applications!
Solution: modularity for reusability

Templates and traditional approaches to NLG

name	price	cuisine
Due Fratelli	$$	Italian
Andalucia	$$$	Spanish, Seafood

Simplest case: use templates!

`inform(NAME, CUISINE)`

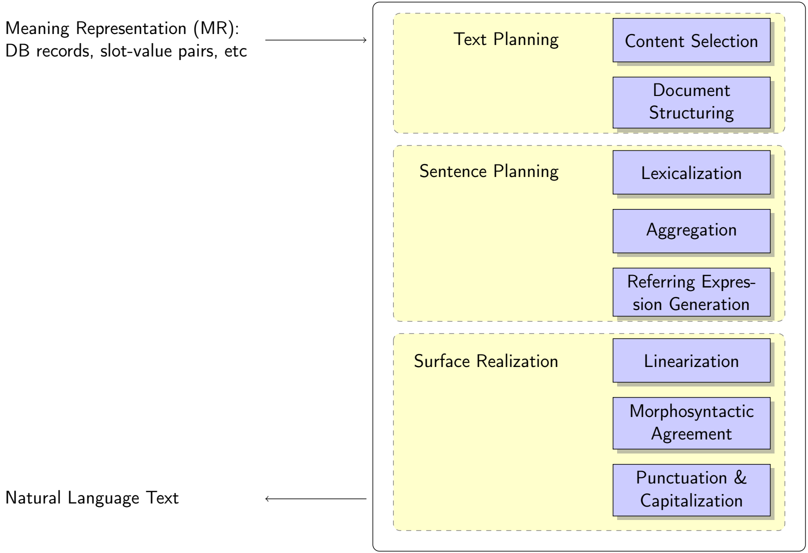
$\rightarrow$ "NAME is a CUISINE restaurant"

$\rightarrow$ "NAME serves CUISINE food"

But this doesn't generalize; only good for limited applications!
Solution: modularity for reusability

- Factor out language-general elements

'The' NLG Pipeline



More reusable components, but still requires a lot of human attention.

End-to-end machine learning approaches to NLG

End-to-end machine learning approaches to NLG

Represent meanings as collections of slot-value pairs...

End-to-end machine learning approaches to NLG

Represent meanings as collections of slot-value pairs...

```
inform(name=DueFratelli, cuisine=Italian)
inform(name=Andalucia, cuisine=Spanish,Seafood)
```

End-to-end machine learning approaches to NLG

Represent meanings as collections of slot-value pairs...

```
inform(name=DueFratelli, cuisine=Italian)
inform(name=Andalucia, cuisine=Spanish,Seafood)
```

Represent texts as sequences of words...

> Due Fratelli serves Italian food.
> Andalucia is a Spanish, seafood restaurant.

End-to-end machine learning approaches to NLG

Represent meanings as collections of slot-value pairs...

```
inform(name=DueFratelli, cuisine=Italian)
inform(name=Andalucia, cuisine=Spanish,Seafood)
```

Represent texts as sequences of words...

Due Fratelli serves Italian food.
Andalucia is a Spanish, seafood restaurant.

Apply your favorite sequence-to-sequence model

End-to-end machine learning approaches to NLG

Represent meanings as collections of slot-value pairs...

```
inform(name=DueFratelli, cuisine=Italian)
inform(name=Andalucia, cuisine=Spanish,Seafood)
```

Represent texts as sequences of words...

Due Fratelli serves Italian food.
Andalucia is a Spanish, seafood restaurant.

Apply your favorite sequence-to-sequence model

- Bayesian networks for gen. w/active learning (Mairesse et al. 2010; BAGEL)

End-to-end machine learning approaches to NLG

Represent meanings as collections of slot-value pairs...

```
inform(name=DueFratelli, cuisine=Italian)
inform(name=Andalucia, cuisine=Spanish,Seafood)
```

Represent texts as sequences of words...

Due Fratelli serves Italian food.
Andalucia is a Spanish, seafood restaurant.

Apply your favorite sequence-to-sequence model

- Bayesian networks for gen. w/active learning (Mairesse et al. 2010; BAGEL)
- Semantically-Conditioned LSTM (Wen et al. 2015)
- TGen (Dušek & Jurčíček 2016)
- Neural Checklist Model (Kiddon et al. 2016)

End-to-end machine learning approaches to NLG

Represent meanings as collections of slot-value pairs...

```
inform(name=DueFratelli, cuisine=Italian)
inform(name=Andalucia, cuisine=Spanish,Seafood)
```

Represent texts as sequences of words...

Due Fratelli serves Italian food.
Andalucia is a Spanish, seafood restaurant.

Apply your favorite sequence-to-sequence model

- Bayesian networks for gen. w/active learning (Mairesse et al. 2010; BAGEL)
- Semantically-Conditioned LSTM (Wen et al. 2015)
- TGen (Dušek & Jurčíček 2016)
- Neural Checklist Model (Kiddon et al. 2016)

Problem: shallow representations and relative opacity

Can we have the best of both worlds?

From the rule-based approach:

- existing resources for NLG;
- richer semantic & syntactic structure;
- inspectability; and
- modularity where it's helpful.

From the ML approach:

- reduced development effort

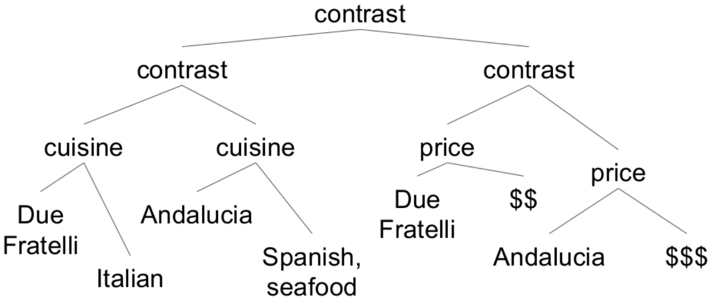
Let's find out!

Narrowing our focus:

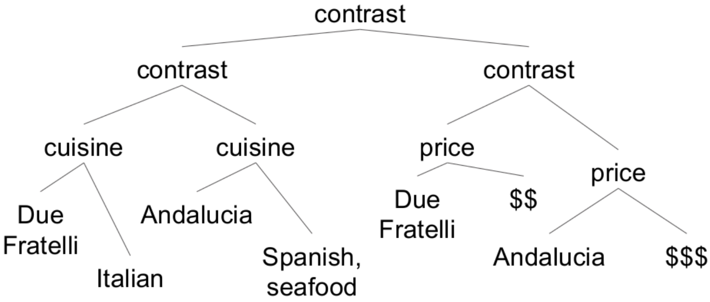
- assume text planning already done
- **learn sentence planning rules**
- use existing systems for surface realization

Text Plans with Discourse Structure

Text Plans with Discourse Structure



Text Plans with Discourse Structure

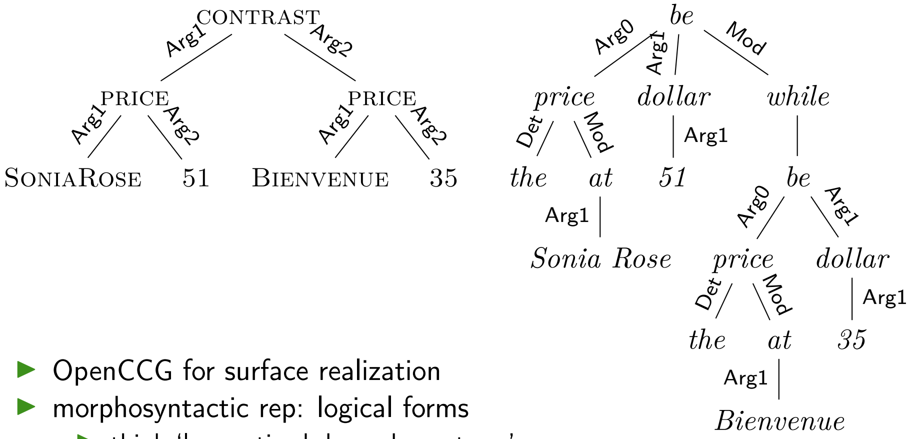


Due Fratelli is an Italian restaurant, while Andalucia is a Spanish seafood restaurant. However, Due Fratelli's price is average, while Andalucia's price is more expensive.

Morphosyntactic Reps for Surface Realization

Text Plan

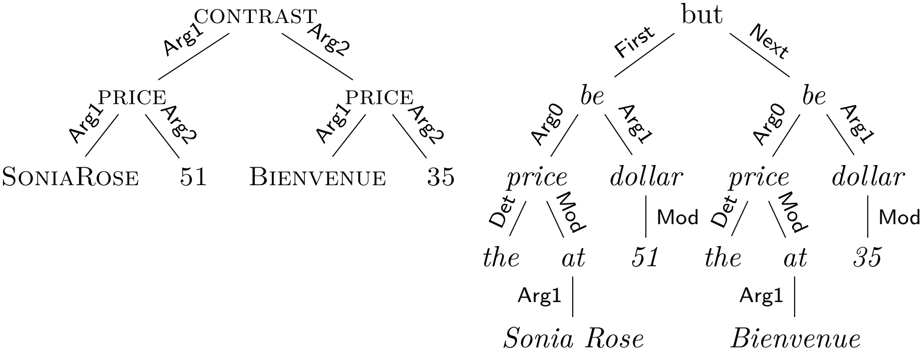
'Logical Form'

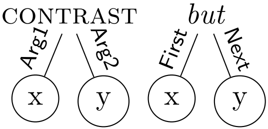


- OpenCCG for surface realization
- morphosyntactic rep: logical forms
  - think 'lemmatized dependency trees'
- based on CCGbank (Hockenmaier 2006) $\Longrightarrow$ WSJ coverage

How to represent tree-to-tree mappings?

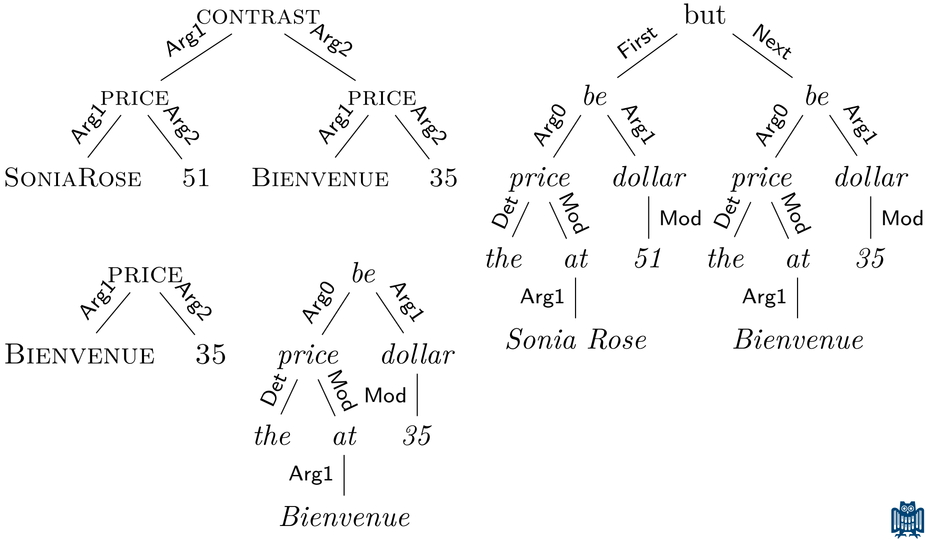
Synchronous Tree Substitution Grammars





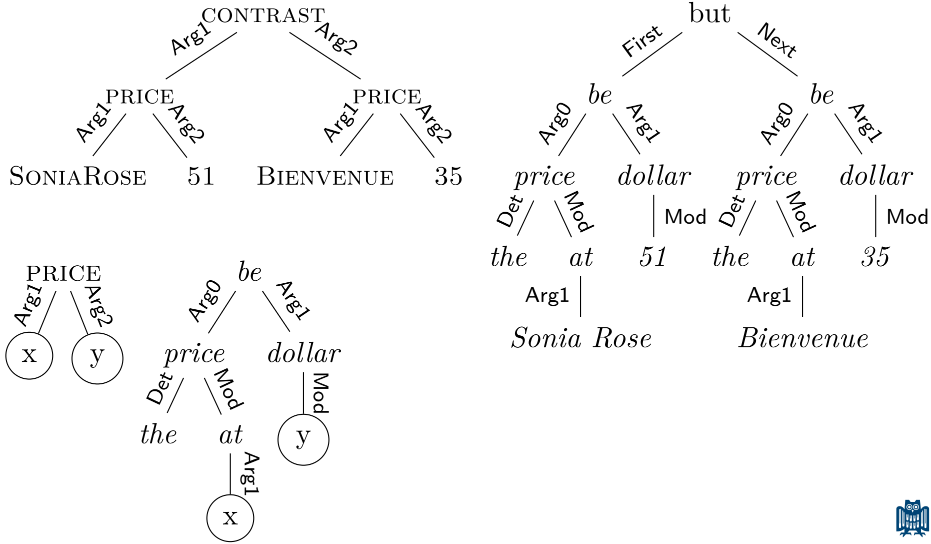
How to represent tree-to-tree mappings?

Synchronous Tree Substitution Grammars

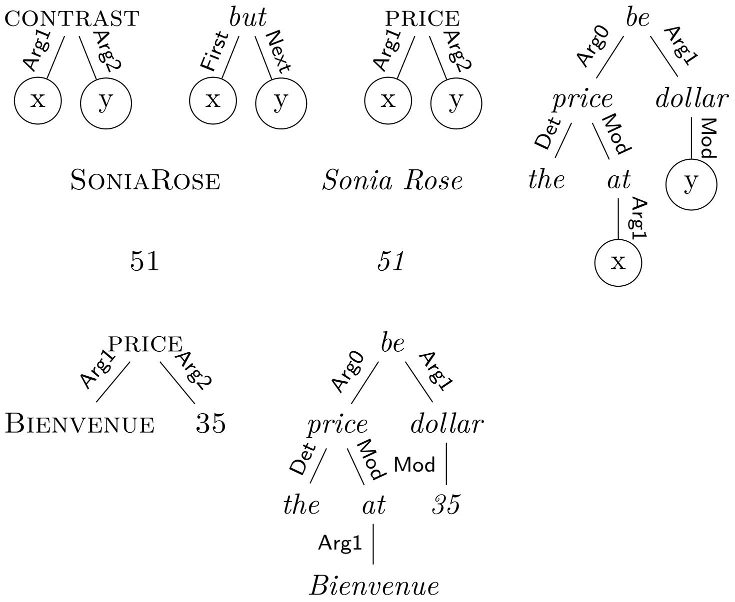


How to represent tree-to-tree mappings?

Synchronous Tree Substitution Grammars



Synchronous Derivation with TSGs



How to learn tree-to-tree mappings?

Hierarchical Dirichlet Processes

- prior biases towards small, reusable trees
- observations provide evidence for larger trees

How to learn tree-to-tree mappings?

Hierarchical Dirichlet Processes

- prior biases towards small, reusable trees
- observations provide evidence for larger trees

How to learn tree-to-tree mappings?

Hierarchical Dirichlet Processes

- prior biases towards small, reusable trees
- observations provide evidence for larger trees

Chinese Restaurant Process

$$P(e) = \frac{freq(e)}{\#obs + \alpha} + \frac{\alpha}{\#obs + \alpha} P_{prior}(e), \tag{1}$$

where α is the concentration parameter, $freq(e)$ is the number of times we have observed the elementary tree e, $\#obs$ is the total number of observations, and P_{prior} is our prior.

Statistical Model for sTSGs for SP

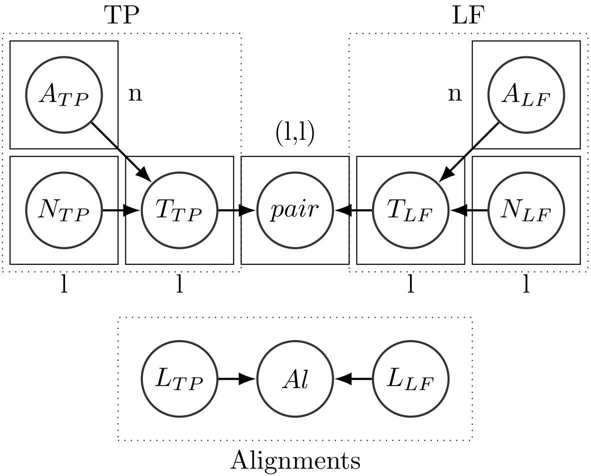


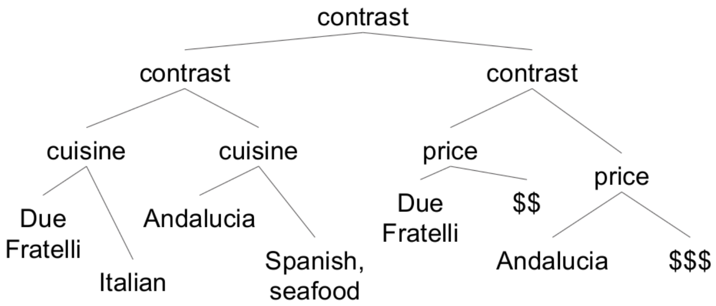
Figure: Dependencies in our statistical model, omitting parameters for clarity. Each node represents a Dirichlet process over base distributions with $\alpha = 1$. n here indexes node labels, while l similarly represents tree locations.

Some example texts

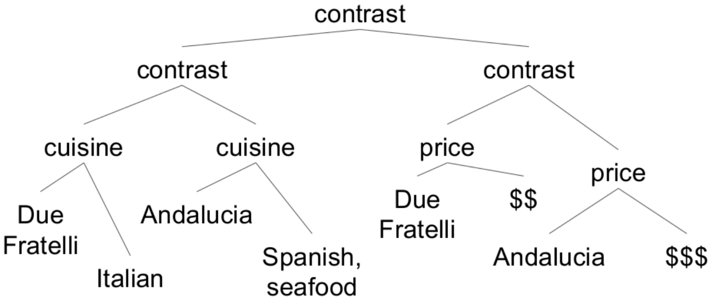
1. Chanpen Thai has the best overall quality among the selected restaurants. Its price is 24 dollars and it has good service. **This Thai restaurant** has good food quality, with decent decor.
2. **Since** Komodo's price is 29 dollars and it has good decor, it has the best overall quality among the selected restaurants.
3. Azuri Cafe, **which** is a **Vegetarian** restaurant has very good food quality. Its price is 14 dollars. It has the best overall quality among the selected restaurants.
4. Komodo has very good service. It has **food food quality, with very good food quality, it has very good food quality** and its price is 29 dollars.

Training data?

Training data?

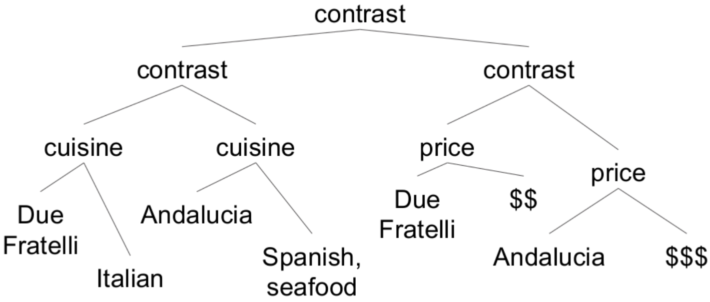


Training data?



Due Fratelli is an Italian restaurant, while Andalucia is a Spanish seafood restaurant. However, Due Fratelli's price is average, while Andalucia's price is more expensive.

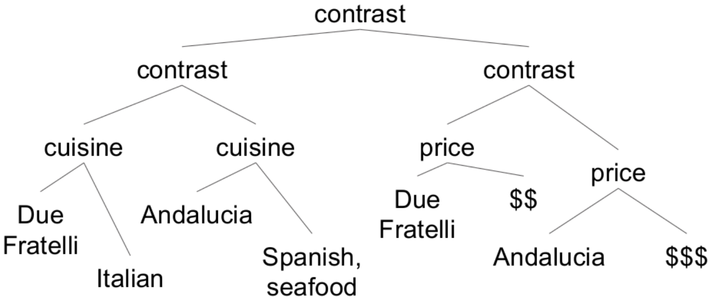
Training data?



Due Fratelli is an Italian restaurant, while Andalucia is a Spanish seafood restaurant. However, Due Fratelli's price is average, while Andalucia's price is more expensive.

- The SPaRKy Restaurant Corpus (Walker et al. 2007)

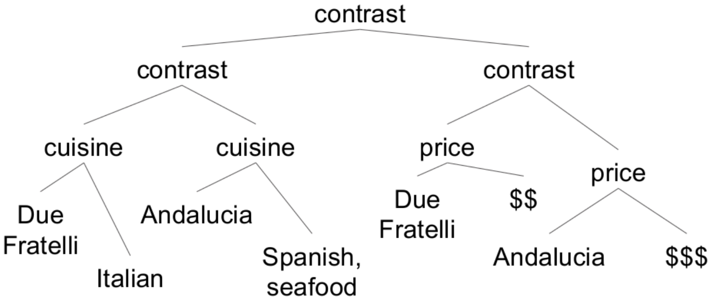
Training data?



Due Fratelli is an Italian restaurant, while Andalucia is a Spanish seafood restaurant. However, Due Fratelli's price is average, while Andalucia's price is more expensive.

- The SPaRKy Restaurant Corpus (Walker et al. 2007)
  - 1800 texts from an NLG system

Training data?



Due Fratelli is an Italian restaurant, while Andalucia is a Spanish seafood restaurant. However, Due Fratelli's price is average, while Andalucia's price is more expensive.

- The SPaRKy Restaurant Corpus (Walker et al. 2007)
  - 1800 texts from an NLG system
  - discourse semantics, but limited variation

Crowdsourced Corpora

Crowdsourced Corpora

BAGEL Corpus (Mairesse et al. 2010)

- 404 utterances for 202 dialogue acts
- e.g. `inform(name=DueFratelli;price=$$;cuisine=Italian)`

Crowdsourced Corpora

BAGEL Corpus (Mairesse et al. 2010)

- 404 utterances for 202 dialogue acts
- e.g. `inform(name=DueFratelli;price=$$;cuisine=Italian)`

SFX-restaurants (Wen et al. 2015)

- 5k utterances+DAs

Crowdsourced Corpora

BAGEL Corpus (Mairesse et al. 2010)

- 404 utterances for 202 dialogue acts
- e.g. `inform(name=DueFratelli;price=$$;cuisine=Italian)`

SFX-restaurants (Wen et al. 2015)

- 5k utterances+DAs

E2E Challenge Dataset (Novikova et al. 2016, 2017)

- 50k utterances+DAs
- increased variation (image-based elicitation)

Building the corpus

Objective: the best of both worlds

Building the corpus

Objective: the best of both worlds

1. discourse-level semantic representation

Building the corpus

Objective: the best of both worlds

1. discourse-level semantic representation
2. with a good amount of variation
   - esp. with respect to *information density*

Building the corpus

Objective: the best of both worlds

1. discourse-level semantic representation
2. with a good amount of variation
   - esp. with respect to *information density*

Method: collect paraphrases

Building the corpus

Objective: the best of both worlds

1. discourse-level semantic representation
2. with a good amount of variation
   - esp. with respect to *information density*

Method: collect paraphrases

- already have discourse-level semantics

Building the corpus

Objective: the best of both worlds

1. discourse-level semantic representation
2. with a good amount of variation
   - esp. with respect to *information density*

Method: collect paraphrases

- already have discourse-level semantics
- more variation than in the original SPaRKy corpus

Building the corpus

Objective: the best of both worlds

1. discourse-level semantic representation
2. with a good amount of variation
   - esp. with respect to *information density*

Method: collect paraphrases

- already have discourse-level semantics
- more variation than in the original SPaRKy corpus

2 conditions: default vs. elderly audience

We are adding variety to an existing dialogue system and we need your help!

In this task, you will be given a text about one or more restaurants written by our existing system.

Your job is to express the same facts, describing the restaurant(s) as you would describe them to your...
default: **...friends or family.**
elderly: **...85-year-old grandmother.**

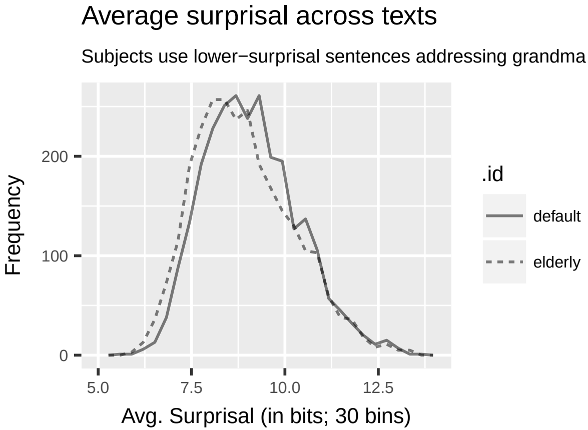
Corpus Statistics

Corpus Statistics

- about 5k texts, with discourse-level semantic annotations

Corpus Statistics

- about 5k texts, with discourse-level semantic annotations
- significantly lower information density in the elderly condition



Examples (1)

Examples (1)

One Italian restaurant is called Caffe Buon Gusto. However, John's Pizzeria is an Italian pizza restaurant.

Choose Caffe Buon Gusto if you desire a traditional Italian restaurant. Otherwise, try out John's Pizzeria.

Examples (1)

One Italian restaurant is called Caffe Buon Gusto. However, John's Pizzeria is an Italian pizza restaurant.

Choose Caffe Buon Gusto if you desire a traditional Italian restaurant. Otherwise, try out John's Pizzeria.

cf. Caffe Buon Gusto is an Italian restaurant. John's Pizzeria, on the other hand, is an Italian, Pizza restaurant.

Examples (2)

Chez Joesphine is the best choice because of food quality, service and decor.

Hands down, Chez Josephine has the best quality food out of all of these restaurants. Employees are always happy to help you and the atmosphere is fantastic.

Examples (2)

Chez Joesphine is the best choice because of food quality, service and decor.

Hands down, Chez Josephine has the best quality food out of all of these restaurants. Employees are always happy to help you and the atmosphere is fantastic.

cf. Chez Josephine has the best overall quality among the selected restaurants. It has very good service, with very good decor. It has very good food quality.

Summary

We have **motivated a new approach to ML for NLG**:

We built a **corpus** which includes:

Summary

We have **motivated a new approach to ML for NLG**:

We built a **corpus** which includes:

Summary

We have **motivated a new approach to ML for NLG**:

- with a focus on learning **sentence planning rules**

We built a **corpus** which includes:

Summary

We have **motivated a new approach to ML for NLG**:

- with a focus on learning **sentence planning rules**
- represented as **synchronous tree-substitution grammars**

We built a **corpus** which includes:

Summary

We have **motivated a new approach to ML for NLG**:

- with a focus on learning **sentence planning rules**
- represented as **synchronous tree-substitution grammars**
- using **hierarchical Dirichlet Processes**.

We built a **corpus** which includes:

Summary

We have **motivated a new approach to ML for NLG**:

- with a focus on learning **sentence planning rules**
- represented as **synchronous tree-substitution grammars**
- using **hierarchical Dirichlet Processes**.

We built a **corpus** which includes:

Summary

We have **motivated a new approach to ML for NLG**:

- with a focus on learning **sentence planning rules**
- represented as **synchronous tree-substitution grammars**
- using **hierarchical Dirichlet Processes**.

We built a **corpus** which includes:

- variation with respect to **information density**, and

Summary

We have **motivated a new approach to ML for NLG**:

- with a focus on learning **sentence planning rules**
- represented as **synchronous tree-substitution grammars**
- using **hierarchical Dirichlet Processes**.

We built a **corpus** which includes:

- variation with respect to **information density**, and
- a hierarchical **semantic annotation**.

Summary

We have **motivated a new approach to ML for NLG**:

- with a focus on learning **sentence planning rules**
- represented as **synchronous tree-substitution grammars**
- using **hierarchical Dirichlet Processes**.

We built a **corpus** which includes:

- variation with respect to **information density**, and
- a hierarchical **semantic annotation**.

This work was supported by the DFG through
SFB 1102 'Information Density and Linguistic Encoding'.

Summary

We have **motivated a new approach to ML for NLG**:

- with a focus on learning **sentence planning rules**
- represented as **synchronous tree-substitution grammars**
- using **hierarchical Dirichlet Processes**.

We built a **corpus** which includes:

- variation with respect to **information density**, and
- a hierarchical **semantic annotation**.

This work was supported by the DFG through
SFB 1102 'Information Density and Linguistic Encoding'.

Thank you!

Did we achieve greater lexical variety?

corpus	# texts	Vocabulary
BAGEL	404	74
SFX-restaurant	5192	353
Novikova et al.	1243	238
Original SRC	1760	99
Extended SRC	**5356**	**577**

Table: Vocabulary diversity and corpus size