Learning Neural Templates for Text Generation

Sam Wiseman, Stuart M. Shieber, and Alexander M. Rush

Template-style Text Generation

A classical NLG template:

<restaurant_name> is a <food_type> <restaurant_type> with a <num_stars> star rating. It is located in <neighborhood>, and its price range is <price_range>.

Why Templates?

Template-based generation addresses certain deficiencies in encoder/decoder style generation.

- More interpretable
- More controllable

Where We're Going

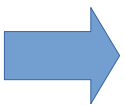
An end-to-end, encoder/decoder style model that allows for:

- Induction of discrete, template-like objects from text
- Interpretable and controllable generation with these induced templates
- Good, but not quite SOTA performance on automatic metrics

Data-to-Text Generation

[c.f., Lebret et al., 2016]

Frederick Parker-Rhodes	
Born	21 November 1914 Newington, Yorkshire
Died	2 March 1987 (aged 72)
Residence	UK
Nationality	British
Known for	Contributions to computational linguistics, combinatorial physics, bit-string physics, plant pathology, and mycology
Scientific career	
Fields	Mycology, Plant Pathology, Mathematics, Linguistics, Computer Science
Author abbrev. (botany)	Park.-Rhodes

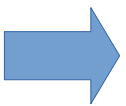


"Frederick Parker-Rhodes (21 November 1914 – 2 March 1987) was an English linguist, plant pathologist, computer scientist, mathematician, mystic, and mycologist."

Data-to-Text Generation

[c.f., Novikova et al., 2017]

Name	The Eagle
Eat Type	coffee shop
Food	French
Price Range	moderate
Customer Rating	3/5
Area	riverside
Kids Friendly	yes
Near	Burger King

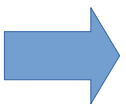


“The three star coffee shop, The Eagle, gives families a mid-priced dining experience featuring a variety of wines and cheeses. Find The Eagle near Burger King.”

Argument for Templates #1: Interpretability

Frederick Parker-Rhodes

Born	21 November 1914 Newington, Yorkshire
Died	2 March 1987 (aged 72)
Residence	UK
Nationality	British
Known for	Contributions to computational linguistics, combinatorial physics, bit-string physics, plant pathology, and mycology
	Scientific career
Fields	Mycology, Plant Pathology, Mathematics, Linguistics, Computer Science
Author abbrev. (botany)	Park.-Rhodes

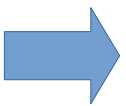


Frederick Parker-Rhodes (21 November 1914 – 2 March 1987) was an English mycology and plant pathology, mathematics at the University of UK."

Argument for Templates #1: Interpretability

Frederick Parker-Rhodes

Born	21 November 1914 Newington, Yorkshire
Died	2 March 1987 (aged 72)
Residence	UK
Nationality	British
Known for	Contributions to computational linguistics, combinatorial physics, bit-string physics, plant pathology, and mycology
	Scientific career
Fields	Mycology, Plant Pathology, Mathematics, Linguistics, Computer Science
Author abbrev. (botany)	Park.-Rhodes

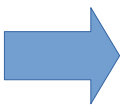


Frederick Parker-Rhodes (21 November 1914 – 2 March 1987) was an English mycology and plant pathology, mathematics at the University of UK."

`<name>` (born `<born>`) was a `<nationality>` `<occupation>`, who lived in the `<residence>`. He was known for contributions to `<known_for>`.

Argument for Templates #2: Controllability

Name	The Eagle
Eat Type	coffee shop
Food	French
Price Range	moderate
Customer Rating	3/5
Area	riverside
Kids Friendly	yes
Near	Burger King



`<name>` is a kid-friendly `<eat_type>` serving `<food>` cuisine in the `<area>` area.

The `<customer_rating>` star rated `<name>` serves `<food>` food at a `<price_range>` price.

Near `<near>` is a `<food>` `<eat_type>` with a `<customer_rating>` star rating. It is family friendly, and its price range is `<price_range>`.

Goal: Learned Template-style Generation

- **Idea:** use a Hidden Semi-Markov Model (HSMM) decoder
  - Preserve most of the encoder/decoder setup
  - Learn template-like representations jointly with learning to generate

Hidden Semi-Markov Models

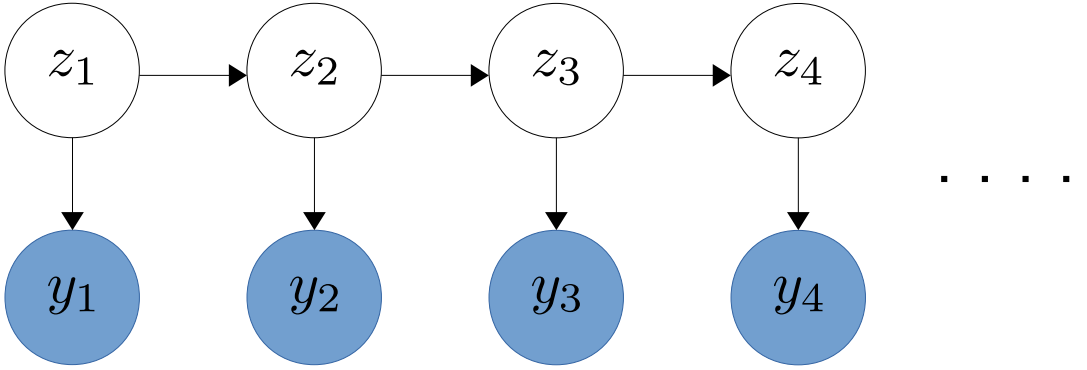
[Gales and Young, 1993; Ostendorf et al., 1996]

- Give a joint distribution over observations $y_{1:T}$ and discrete latents $z_{1:S}$
  - Like HMMs, but observations can last multiple time-steps:

Hidden Semi-Markov Models

[Gales and Young, 1993; Ostendorf et al., 1996]

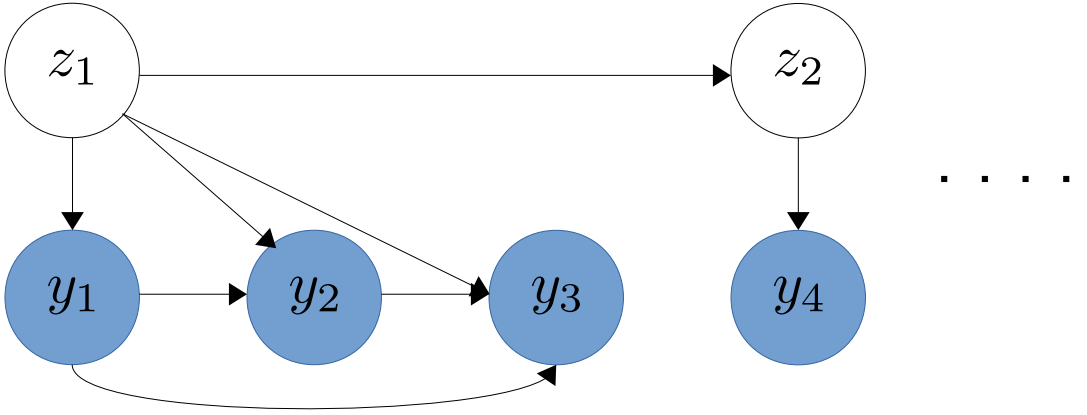
- Give a joint distribution over observations $y_{1:T}$ and discrete latents $z_{1:S}$
  - Like HMMs, but observations can last multiple time-steps:



Hidden Semi-Markov Models

[Gales and Young, 1993; Ostendorf et al., 1996]

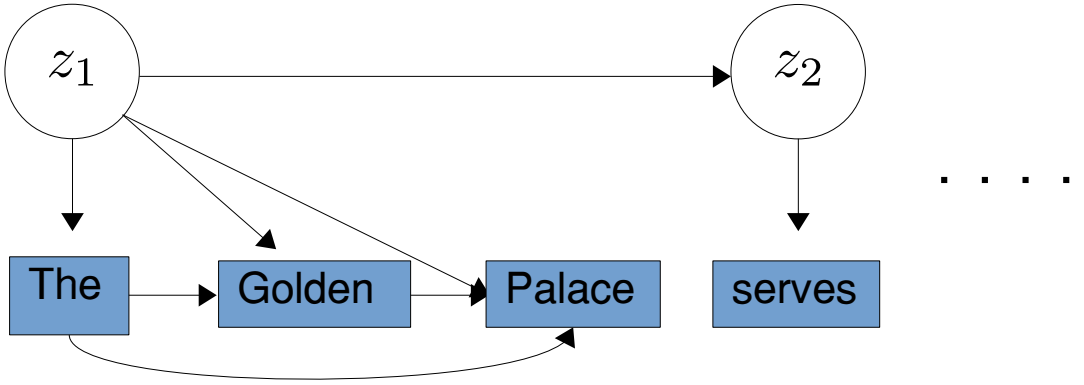
- Give a joint distribution over observations $y_{1:T}$ and discrete latents $z_{1:S}$
  - Like HMMs, but observations can last multiple time-steps:



Hidden Semi-Markov Models

[Gales and Young, 1993; Ostendorf et al., 1996]

- Give a joint distribution over observations $y_{1:T}$ and discrete latents $z_{1:S}$
  - Like HMMs, but observations can last multiple time-steps:

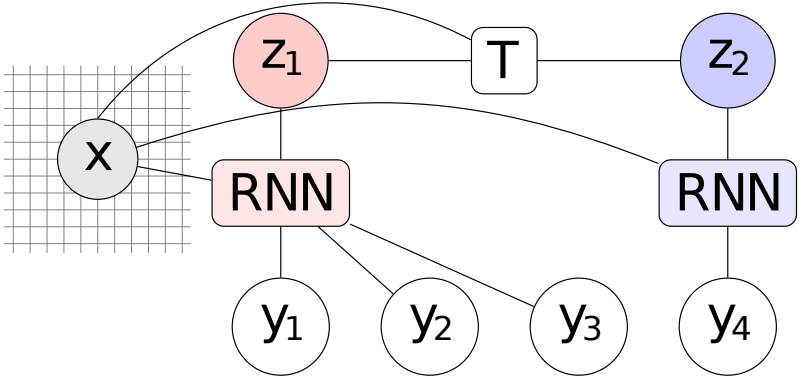


Upshot: HSMMs give us typed segmentations.

A Conditional (Neural) HSMM

$$p(y, z \mid x) = \prod_{s=1}^{S} \underbrace{p(z_s \mid z_{s-1}, x)}_{\text{transition prob}} \underbrace{p(l_s \mid z_s)}_{\text{length prob}} \underbrace{p(y_{t_0(s):t_1(s)} \mid z_s, l_s, x)}_{\text{segment prob}}$$

- We parameterize probabilities with neural components:
  - Segment probabilities are given by an RNN + attention + copy attention



Learning

- We're given a dataset of x, y pairs
- Segmentations z are unobserved at training time
- Maximize $\ln p(y_{1:T} \mid x) = \ln \sum_z p(y_{1:T}, z \mid x)$
  - Can use a dynamic program analogous to the forward or backward algorithm used in learning HMMs [c.f., Murphy 2002]
  - Can simply backprop through the dynamic program
    - Easy with pytorch!

Generation

- Given an input x, we could generate by approximating $\arg\max_{y,z} p(y, z \mid x)$
- Instead, we'll first extract "templates":

  1) Viterbi-segment the training data

Generation

- Given an input x, we could generate by approximating $\arg\max_{y,z} p(y, z \mid x)$
- Instead, we'll first extract "templates":

  1) Viterbi-segment the training data

```
[The Golden Palace]55[is a]59[coffee shop]12[providing]3[Indian]50
[food]1[in the]17[£20-25]125[price range]16[.]2 [It is]8[located
in the]25[riverside]40[.]53[Its customer rating is]19[high]23[.]2
```

Generation

- Given an input x, we could generate by approximating $\arg\max_{y,z} p(y, z \mid x)$
- Instead, we'll first extract "templates":

  1) Viterbi-segment the training data

```
[The Golden Palace]55[is a]59[coffee shop]12[providing]3[Indian]50
[food]1[in the]17[£20-25]125[price range]16[.]2 [It is]8[located
in the]25[riverside]40[.]53[Its customer rating is]19[high]23[.]2
```

Note: each segment gets a latent-state label.

- We'll call a sequence of labels $z^{(i)}$ a "template."
- E.g., $z^{(i)}$ = 55, 59, 12, 3, ...

What's Good about these "Templates"?

1) Dim. reduction: latent states correspond to functional categories.

1. | The Eagle
The Golden Curry
Zizzi
… | provides
providing
serves
… | Indian
Chinese
English
… | food
cuisine
Food
… | in the
with a
and has a
… | high
moderate
average
… | price range
customer rating
rating
… | . | It is
They are
It's
… | near
located in the
located near
… |
   | riverside
city centre
Cafe Sicilia
… | . | Its customer rating is
It has a
The price range is
… | 1 out of 5
average
high
… | . |

2. | Located near
Located in the
Near
… | The Portland Arms
riverside
city centre
… | is an
is a family friendly
there is a
… | Italian
fast food
French
… | restaurant called
place called
restaurant named
… | The Waterman
Cocum
Loch Fyne
… | .

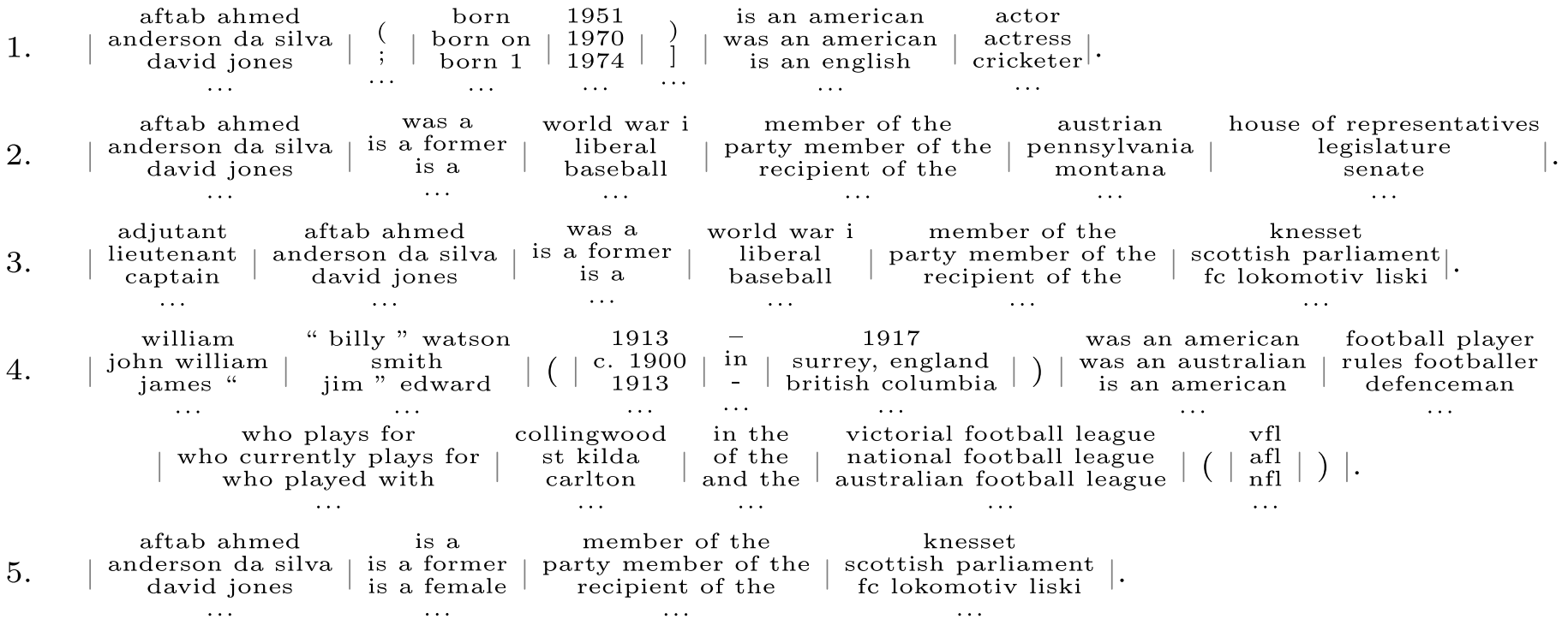
3. | A
An
A family friendly
… | Italian
fast food
French
… | restaurant
pub
coffee shop
… | is
called
named
… | The Waterman
Cocum
Loch Fyne
… | .

4. | Located near
Located in the
Near
… | The Portland Arms
riverside
city centre
… | , | The Eagle
The Golden Curry
Zizzi
… | is a
is a family friendly
is an
… | cheap
family-friendly
family friendly
… | Italian
fast food
French
… | restaurant
pub
coffee shop
… | .

5. | A
An
A family friendly
… | Italian
fast food
French
… | restaurant
pub
coffee shop
… | near
located in the
located near
… | riverside
city centre
Cafe Sicilia
… | is
called
named
… | The Waterman
Cocum
Loch Fyne
… | .

What's Good about these "Templates"?

1) Dim. reduction: latent states correspond to functional categories.



What's Good about these "Templates"?

2) We can use them to control generation:

- Select a template $z^{(i)} = z_1^{(i)}, \ldots, z_S^{(i)}$
- Generate by computing $\arg\max_y p(y, z = z^{(i)} \mid x)$
- Gives a different generation for each $z^{(i)}$
  - (Examples in a few slides...)

Generation Recap

1) Viterbi-segment the training data

2) Collect frequent “templates” $z^{(i)} = z_1^{(i)}, \ldots, z_S^{(i)}$

3) Given a new input x, generate by finding $\arg\max_y p(y, z^{(i)} \mid x)$ for a chosen template $z^{(i)}$

Methods

1) Condition RNNs on latent state by concatenating state-embedding to RNN input

2) Helpful to train with hard constraints: disallow splitting up segments appearing in tables

3) Segment RNNs can condition on all preceding *tokens*

E2E Validation Results

(Val)	BLEU	NIST	ROUGE	CIDEr	METEOR
D&J (2017)	69.25	8.48	72.57	2.40	47.03
Substitution BL					
Neural Template					

- D&J (2017) is an enc/dec + reranker system used in the E2E Challenge
- Substitution BL finds maximally similar training table and performs substitution in corresponding description
- K=60; 1x300 LSTM as segment models
- Used 100 most common $z^{(i)}$ and selected highest overall scorer

E2E Validation Results

(Val)	BLEU	NIST	ROUGE	CIDEr	METEOR
D&J (2017)	69.25	8.48	72.57	2.40	47.03
Substitution BL	43.71	6.72	55.35	1.41	37.87
Neural Template					

- D&J (2017) is an enc/dec + reranker system used in the E2E Challenge
- Substitution BL finds maximally similar training table and performs substitution in corresponding description
- K=60; 1x300 LSTM as segment models
- Used 100 most common $z^{(i)}$ and selected highest overall scorer

E2E Validation Results

(Val)	BLEU	NIST	ROUGE	CIDEr	METEOR
D&J (2017)	69.25	8.48	72.57	2.40	47.03
Substitution BL	43.71	6.72	55.35	1.41	37.87
Neural Template	67.07	7.98	69.50	2.29	43.07

- D&J (2017) is an enc/dec + reranker system used in the E2E Challenge
- Substitution BL finds maximally similar training table and performs substitution in corresponding description
- K=60; 1x300 LSTM as segment models
- Used 100 most common $z^{(i)}$ and selected highest overall scorer

E2E Test Results

(Val)	BLEU	NIST	ROUGE	CIDEr	METEOR
D&J (2017)	65.93	8.59	68.50	2.23	44.83
Substitution BL	43.78	6.88	54.64	1.39	37.35
Neural Template	59.80	7.56	65.01	1.95	38.75

- D&J (2017) is an enc/dec + reranker system used in the E2E Challenge
- Substitution BL finds maximally similar training table and performs substitution in corresponding description
- K=60; 1x300 LSTM as segment models
- Used 100 most common $z^{(i)}$ and selected highest overall scorer

WikiBio Results

	BLEU	NIST	ROUGE-4
Template KN	19.8	5.19	10.7
NNLM (field)	33.4	7.52	23.9
NNLM (field & word)	34.7	7.98	25.8
Neural Template	34.8	7.59	38.6

- Encoder/decoder and template-style baselines from Lebret et al. (2016)
- K=45; 1x300 LSTMs as segment/history models
- Used 100 most common $z^{(i)}$ and selected highest overall scorer

WikiBio Results

	BLEU	NIST	ROUGE-4
Template KN	19.8	5.19	10.7
NNLM (field)	33.4	7.52	23.9
NNLM (field & word)	34.7	7.98	25.8
Neural Template	34.8	7.59	38.6
Liu et al. (2018)	43.7	-	40.3

- Encoder/decoder and template-style baselines from Lebret et al. (2016)
- K=45; 1x300 LSTMs as segment/history models
- Used 100 most common $z^{(i)}$ and selected highest overall scorer

Controllability Example (E2E)

z = 55, 59, 43, 11, 25, 50, 53

Name	Travellers Rest Beefeater
Customer Rating	3 out of 5
Area	riverside
Near	Raja Indian Cuisine

Travellers Rest Beefeater$_{55}$ is a$_{59}$ 3 star$_{43}$ restaurant$_{11}$ located near$_{25}$ Raja Indian Cuisine$_{40}$.$_{53}$

Controllability Example (E2E)

z = 31, 29, 44, 55, 3, 50, 1, 2

Name	Travellers Rest Beefeater
Customer Rating	3 out of 5
Area	riverside
Near	Raja Indian Cuisine

Travellers Rest Beefeater$_{55}$ is a$_{59}$ 3 star$_{43}$ restaurant$_{11}$ located near$_{25}$ Raja Indian Cuisine$_{40}$.$_{53}$

Near$_{31}$ riverside$_{29}$,$_{44}$ Travellers Rest Beefeater$_{55}$ serves$_{3}$ 3 star$_{50}$ food$_{1}$.$_{2}$

Controllability Example (E2E)

z = 55, 59, 12, 3, 50, 1, 17, 26, 16, 2, 8, 25, 40, 53

Name	Travellers Rest Beefeater
Customer Rating	3 out of 5
Area	riverside
Near	Raja Indian Cuisine

Travellers Rest Beefeater$_{55}$ is a$_{59}$ 3 star$_{43}$ restaurant$_{11}$ located near$_{25}$ Raja Indian Cuisine$_{40}$.$_{53}$

Near$_{31}$ riverside$_{29}$,$_{44}$ Travellers Rest Beefeater$_{55}$ serves$_{3}$ 3 star$_{50}$ food$_{1}$.$_{2}$

Travellers Rest Beefeater$_{55}$ is a$_{59}$ restaurant$_{12}$ providing$_{3}$ riverside$_{50}$ food$_{1}$ and has a$_{17}$ 3 out of 5$_{26}$ customer rating$_{16}$.$_{2}$ It is$_{8}$ near$_{25}$ Raja Indian Cuisine$_{40}$.$_{53}$

Controllability Example (E2E)

Name	Travellers Rest Beefeater
Customer Rating	3 out of 5
Area	riverside
Near	Raja Indian Cuisine

Travellers Rest Beefeater$_{55}$ is a$_{59}$ 3 star$_{43}$ restaurant$_{11}$ located near$_{25}$ Raja Indian Cuisine$_{40}$.$_{53}$

Near$_{31}$ riverside$_{29}$,$_{44}$ Travellers Rest Beefeater$_{55}$ serves$_{3}$ 3 star$_{50}$ food$_{1}$.$_{2}$

Travellers Rest Beefeater$_{55}$ is a$_{59}$ restaurant$_{12}$ providing$_{3}$ riverside$_{50}$ food$_{1}$ and has a$_{17}$ 3 out of 5$_{26}$ customer rating$_{16}$.$_{2}$ It is$_{8}$ near$_{25}$ Raja Indian Cuisine$_{40}$.$_{53}$

Travellers Rest Beefeater$_{55}$ is a$_{59}$ place to eat$_{12}$ located near$_{25}$ Raja Indian Cuisine$_{40}$.$_{53}$

Travellers Rest Beefeater$_{55}$ is a$_{59}$ 3 out of 5$_{5}$ rated$_{32}$ riverside$_{43}$ restaurant$_{11}$ near$_{25}$ Raja Indian Cuisine$_{40}$.$_{53}$

Interpretability Example (WikiBio)

Jimmy Deacon

Personal information			
Full name	James Deacon		
Date of birth	23 January 1906		
Place of birth	Glasgow, Scotland		
Date of death	1976 (aged 69–70)		
Height	5 ft 7 in (1.70 m)		
Playing position	Forward		
Senior career*			
Years	**Team**	**Apps**	**(Gls)**
	Darlington	2	(-)
1929–1934	Wolverhampton Wanderers	149	(52)
1934–1939	Southend United	100	(3)
1939–1940	Hartlepool	-	(-)

* Senior club appearances and goals counted for the domestic league only

james deacon$_{42}$ ($_{30}$ born$_{44}$ 23 january 1906$_{11}$)$_{22}$ was a$_{14}$ scottish$_{8}$ football$_{19}$ forward$_{24}$.$_{43}$

Yang Sung-chul

Born	20 November 1939 (age 78) Gokseong County, Jeollanam-do
Citizenship	South Korea
Alma mater	Seoul National University University of Hawaii at Manoa University of Kentucky
Occupation	Political scientist
Employer	Graduate School of International Studies, Korea University
Known for	Member of the National Assembly Ambassador to the United States
Political party	National Congress for New Politics
Children	Two

yang sung-chul$_{42}$ ($_{30}$ november 20, 1939$_{11}$ in$_{21}$ gokseong county, jeollanam-do$_{39}$)$_{22}$ is a$_{14}$ south korean$_{8}$ political scientist$_{24}$.$_{43}$

Conclusion

An encoder/decoder style approach to generation that:

- Allows for the induction of discrete template-like objects
- Allows for generation in a more interpretable way
- Allows for generation in a more controllable way

Thanks!

Code at github.com/harvardnlp/neural-template-gen

Backward Recurrence

$$\begin{aligned}\beta_t(j) &= p(y_{t+1:T} \mid z_t = j, f_t = 1, x) \\ &= \sum_{k=1}^{K} \beta_t^*(k)\, p(z_{t+1} = k \mid z_t = j)\end{aligned}$$

$$\begin{aligned}\beta_t^*(k) &= p(y_{t+1:T} \mid z_{t+1} = k, f_t = 1, x) \\ &= \sum_{l=1}^{L} \Big[\beta_{t+l}(k)\, p(l_{t+1} = l \mid z_{t+1} = k) p(y_{t+1:t+l} \mid z_{t+1} = k, l_{t+1} = l)\Big]\end{aligned}$$

An Aside about Latent Variables

There has been much recent work on learning real-valued latent variable models for text (c.f., Bowman et al. (2016)).

- However, not especially clear what these variables are supposed to represent.

- Templates offer a discrete alternative, which are more interpretable, allow for controllable generation, and are tractable to learn!

Another WikiBio Example

kenny warren

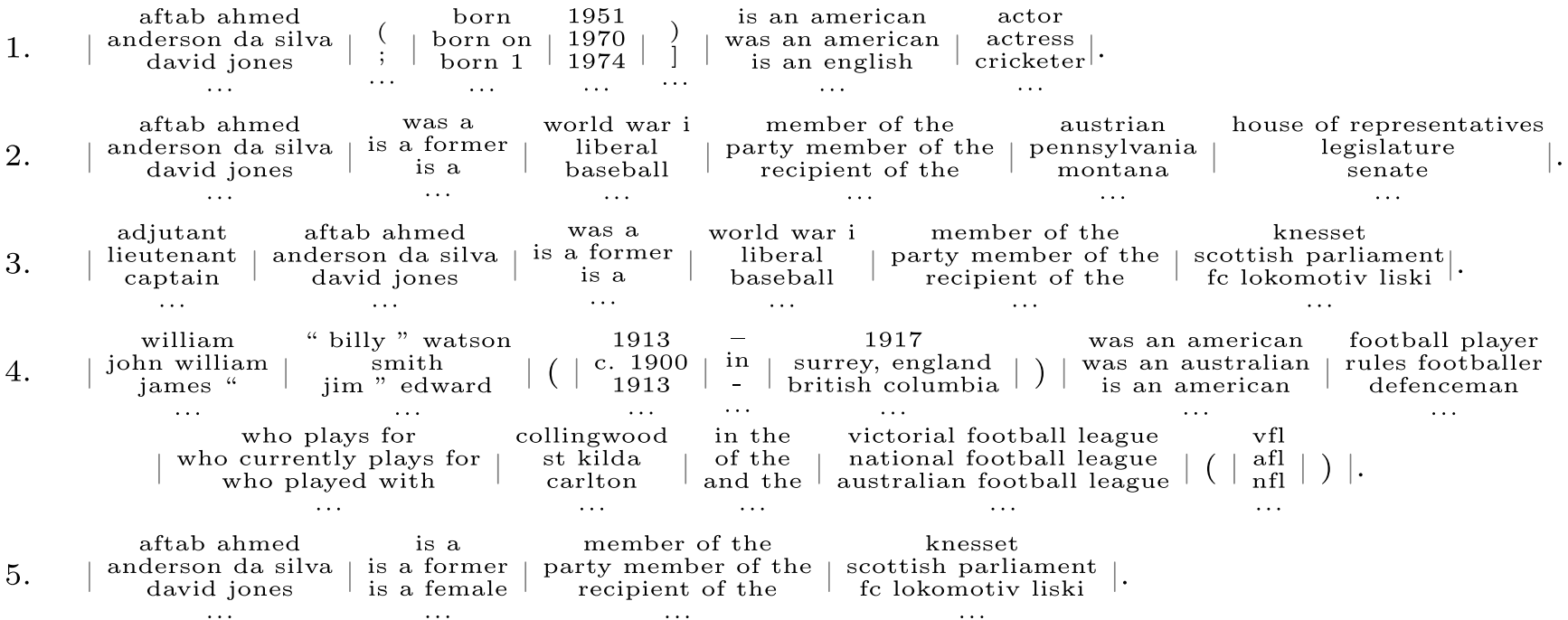
name: kenny warren, **birth date:** 1 april 1946, **birth name:** kenneth warren deutscher, **birth place:** brooklyn, new york, **occupation:** ventriloquist, comedian, author, **notable work:** book **-** the revival of ventriloquism in america

1. [kenneth warren deutscher]$_{132}$ [( ]$_{75}$ [born]$_{89}$ [april 1, 1946]$_{101}$ [ )]$_{67}$ [is an american]$_{82}$ [author]$_{20}$ [and]$_{1}$ [ventriloquist and comedian]$_{69}$ [.]$_{88}$
2. [kenneth warren deutscher]$_{132}$ [( ]$_{75}$ [born]$_{89}$ [april 1, 1946]$_{101}$ [ )]$_{67}$ [is an american]$_{82}$ [author]$_{20}$ [best known for his]$_{95}$ [the revival of ventriloquism]$_{96}$ [.]$_{88}$
3. [kenneth warren]$_{16}$ [“kenny” warren]$_{117}$ [( ]$_{75}$ [born]$_{89}$ [april 1, 1946]$_{101}$ [ )]$_{67}$ [is an american]$_{127}$ [ventriloquist, comedian]$_{28}$ [.]$_{133}$
4. [kenneth warren]$_{16}$ [“kenny” warren]$_{117}$ [( ]$_{75}$ [born]$_{89}$ [april 1, 1946]$_{101}$ [ )]$_{67}$ [is a]$_{104}$ [new york]$_{98}$ [author]$_{20}$ [.]$_{133}$
5. [kenneth warren deutscher]$_{42}$ [is an american]$_{82}$ [ventriloquist, comedian]$_{118}$ [based in]$_{15}$ [brooklyn, new york]$_{84}$ [.]$_{88}$

E2E Learned Template Visualization

1. | The Eagle
The Golden Curry
Zizzi
… | provides
providing
serves
… | Indian
Chinese
English
… | food
cuisine
Food
… | in the
with a
and has a
… | high
moderate
average
… | price range
customer rating
rating
… | . | It is
They are
It's
… | near
located in the
located near
… |
   | riverside
city centre
Cafe Sicilia
… | . | Its customer rating is
It has a
The price range is
… | 1 out of 5
average
high
… | . |
2. | Located near
Located in the
Near
… | The Portland Arms
riverside
city centre
… | is an
is a family friendly
there is a
… | Italian
fast food
French
… | restaurant called
place called
restaurant named
… | The Waterman
Cocum
Loch Fyne
… | . |
3. | A
An
A family friendly
… | Italian
fast food
French
… | restaurant
pub
coffee shop
… | is
called
named
… | The Waterman
Cocum
Loch Fyne
… | . |
4. | Located near
Located in the
Near
… | The Portland Arms
riverside
city centre
… | , | The Eagle
The Golden Curry
Zizzi
… | is a
is a family friendly
is an
… | cheap
family-friendly
family friendly
… | Italian
fast food
French
… | restaurant
pub
coffee shop
… | . |
5. | A
An
A family friendly
… | Italian
fast food
French
… | restaurant
pub
coffee shop
… | near
located in the
located near
… | riverside
city centre
Cafe Sicilia
… | is
called
named
… | The Waterman
Cocum
Loch Fyne
… | . |

WB Learned Template Visualization



State Purity

	NTemp	NTemp+AR
E2E	89.2 (17.4)	85.4 (18.6)
WikiBio	43.2 (19.7)	39.9 (17.9)