

# **Unsupervised Discrete Pattern Mining for Text Generation**

Deng Cai

# References

- AAAI17 Mechanism-Aware Neural Machine for Dialogue Response Generation
- ACL18 Unsupervised Discrete Sentence Representation Learning for Interpretable Neural Dialog Generation
- EMNLP18 Learning Neural Templates for Text Generation

# Neural Text Generation

- **Black-box nature** of generic encoder-decoder models.
- Two Lost desiderata
  - Interpretable
  - Controllable

# Discrete Pattern

- Discrete features (categories) that characterize a piece of text.
  - E.g., sentiments, topics, acts, templates.
- Discrete pattern recognition often brings about **interpretability** and **controllability**, which are in desperate need.
- It also solves diversity and informativeness

Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

## Mechanism-Aware Neural Machine for Dialogue Response Generation

**Ganbin Zhou,<sup>1,2</sup> Ping Luo,<sup>1</sup> Rongyu Cao,<sup>1,2</sup> Fen Lin,<sup>3</sup> Bo Chen,<sup>3</sup> Qing He<sup>1</sup>**

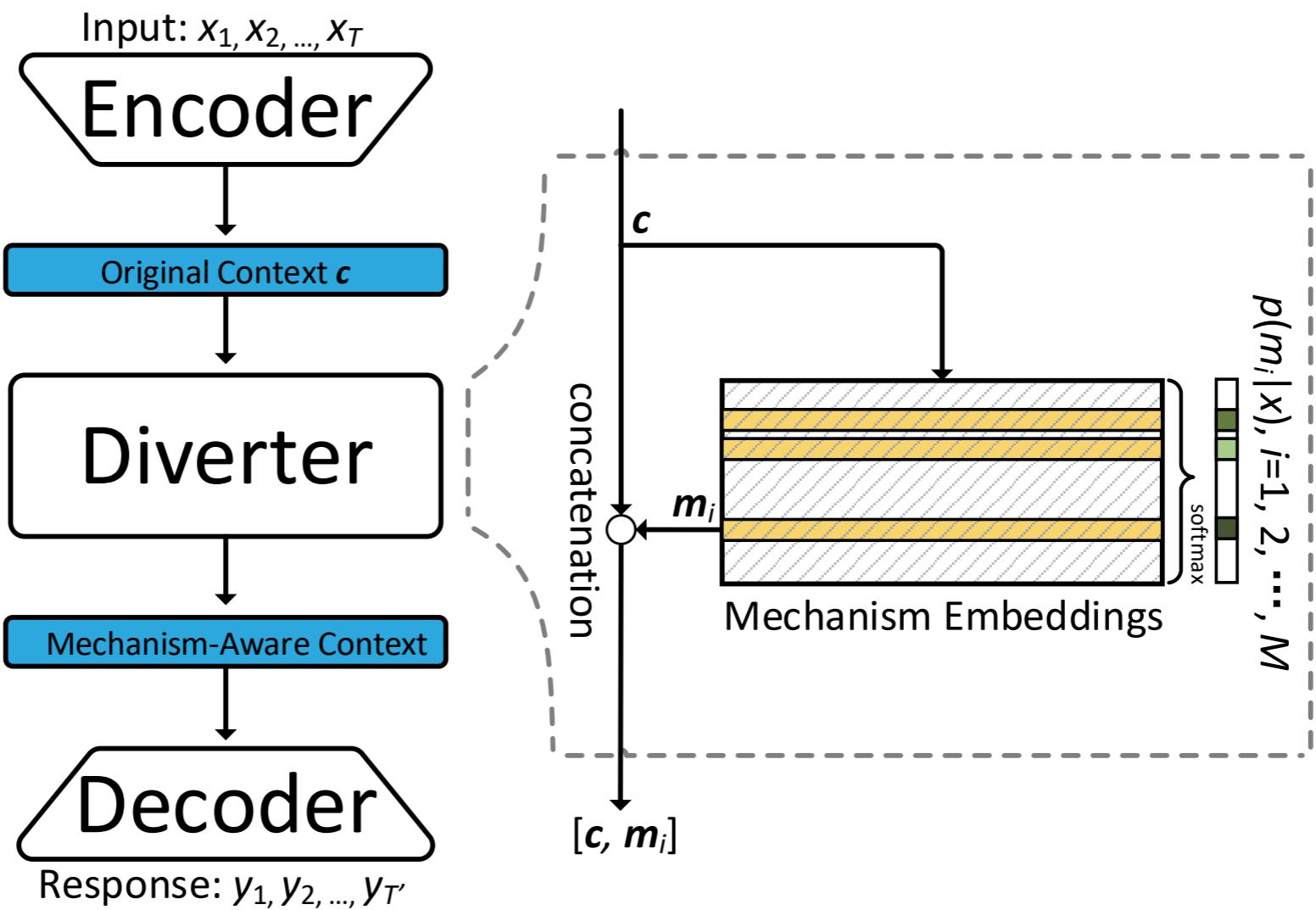
<sup>1</sup>Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),  
Institute of Computing Technology, CAS, Beijing 100190, China.

{zhouganbin, caory}@ics.ict.ac.cn, {luop, heqing}@ict.ac.cn

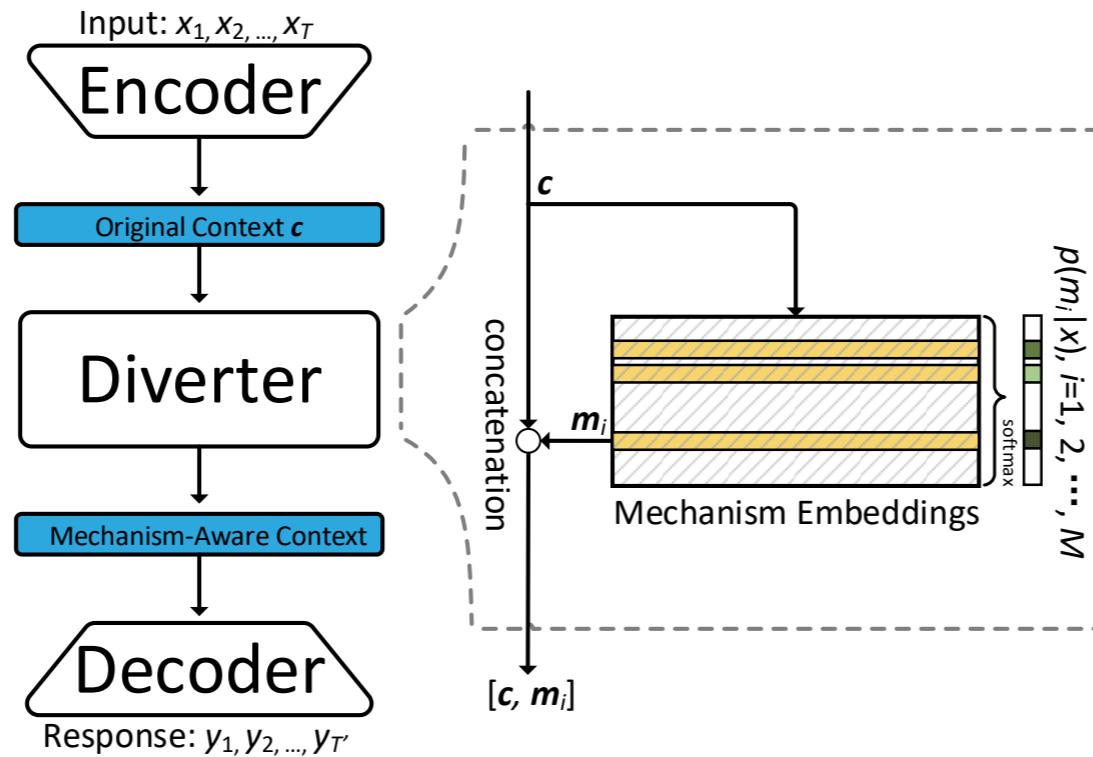
<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China.

<sup>3</sup>Pattern Recognition Center, WeChat Technical Architecture Department, Tencent, China.

# A Latent Variable



# A Latent Variable



$$\sum_{(\mathbf{x}, \mathbf{y}) \in \mathcal{D}^c} \log p(\mathbf{y}|\mathbf{x}) = \sum_{(\mathbf{x}, \mathbf{y}) \in \mathcal{D}^c} \log \sum_{i=1}^M p(m_i|\mathbf{x}) p(\mathbf{y}|m_i, \mathbf{x})$$

$$p(\mathbf{y}, m_l | \mathbf{x}) = p(m_l | \mathbf{x}) p(\mathbf{y} | m_l, \mathbf{x})$$

# Results

Questions	Mechanism 1	Mechanism 2	Mechanism 3	Mechanism 4
我要坐地铁回家 I take subway back home	好的，路上注意安全 Ok, be safe on the subway	真是个不错的选择 That's a good choice	坐车去哪里 Where are you going on the subway	坐车一定要注意安全 Be careful on the subway
生活中怎样感到开心快乐 How do I feel happy in my life	开心就来和我聊天 Have a talk with me if you are happy	看来你今天心情很不错 You seem to be happy today	有什么开心的事情 What are the happy things you have experienced	和我聊天就很开心 Talk with me and you will be happy
明天考试、求安慰 I have a test tomorrow, and I need some solace	好好复习就好了 Just work hard on the reviews	我会为你加油的 I will stand by you	你要去哪里试试 Where are you taking the test	我会一直安慰你的 I will always be on your side
上班好累啊 The work is really hard	那就好好休息 Take a good rest	要懂得劳逸结合 You need to balance work and life	你是不是累了 Are you tired	上班就是这样 That is work

Keyword in m <sub>1</sub>		Keyword in m <sub>2</sub>		Keyword in m <sub>3</sub>		Keyword in m <sub>4</sub>	
Chinese	English	Chinese	English	Chinese	English	Chinese	English
看	look	个	a	哪里	where	不会	cannot
好好	ok	看来	seem	?	?	一直	always
还是	still	不错	nice	呀	oh	可是	however
那	that	小	little	么	why	一定	must
注意	attention	还	yet	怎么	how	都	all
自己	self	这样	this	什么	what	陪	company
知道	know	人	person	在	stay	也	also
就	at once	可	may	样子	appearance	会	can
可以	can	微	tiny	想	think	认错	admit
去	go to	对	yes	说	say	很	very

# **Unsupervised Discrete Sentence Representation Learning for Interpretable Neural Dialog Generation**

**Tiancheng Zhao, Kyusong Lee and Maxine Eskenazi**

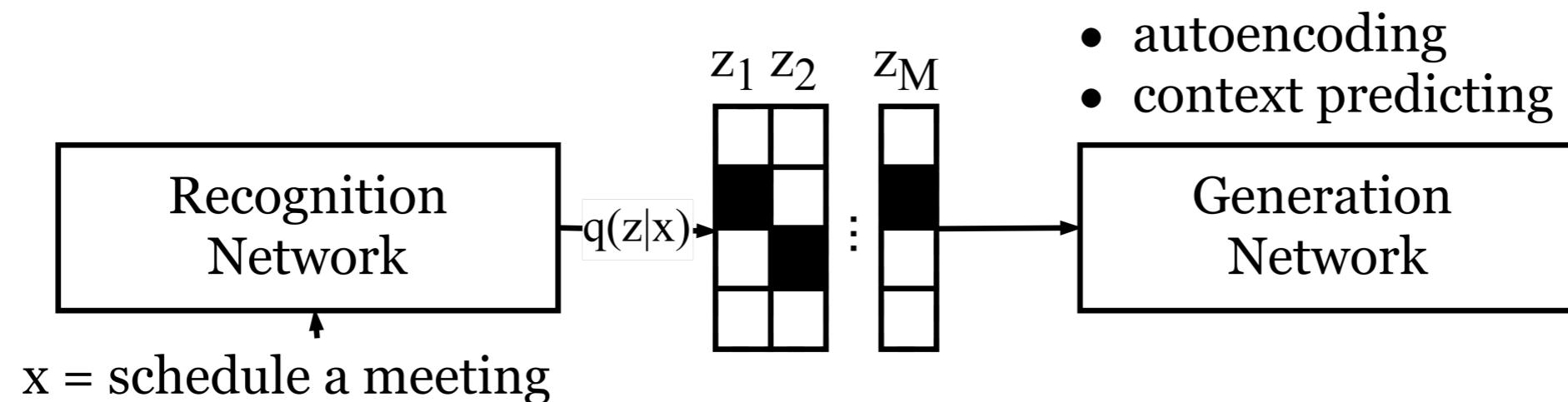
Language Technologies Institute

Carnegie Mellon University

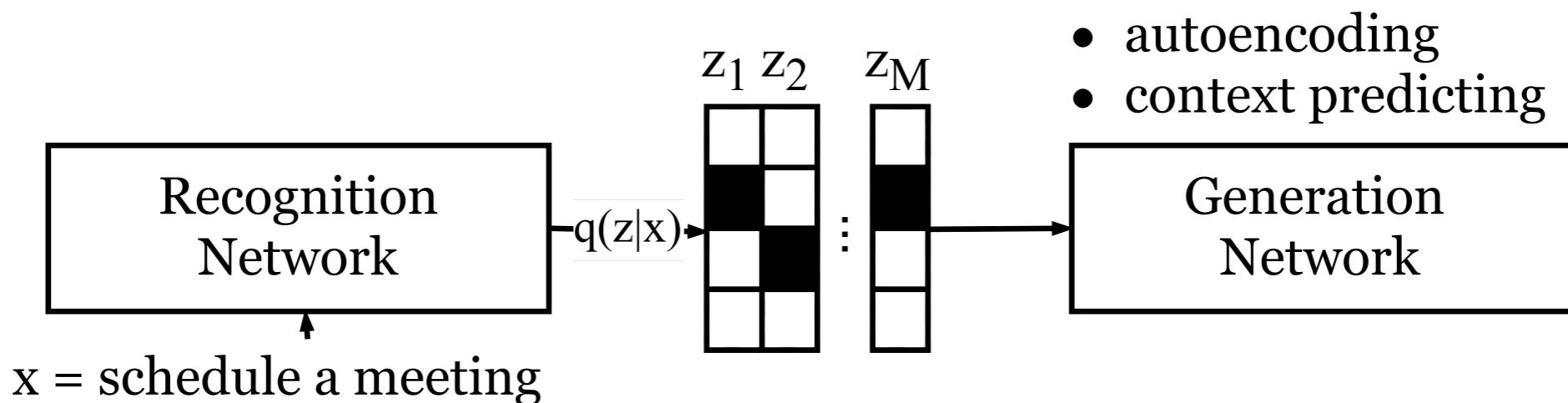
Pittsburgh, Pennsylvania, USA

{tianchez, kyusongl, max+}@cs.cmu.edu

# a set of Latent Variables



# a set of Latent Variables



$$\mathcal{L}_{\text{LAED}}(\theta_{\mathcal{F}}, \theta_{\pi}) =$$

$$\mathbb{E}_{q_{\mathcal{R}}(\mathbf{z}|\mathbf{x})p(\mathbf{x}, \mathbf{c})} [\log p_{\pi}(\mathbf{z}|\mathbf{c}) + \log p_{\mathcal{F}}(\mathbf{x}|\mathbf{z}, \mathbf{c})]$$

$$\mathcal{L}_{\text{Attr}}(\theta_{\mathcal{F}}) = \mathbb{E}_{q_{\mathcal{R}}(\mathbf{z}|\mathbf{x})p(\mathbf{c}, \mathbf{x})} [\log q_{\mathcal{R}}(\mathbf{z}|\mathcal{F}(\mathbf{c}, \mathbf{z}))]$$

# Results

<b>Model</b>	<b>Action</b>	<b>Sample utterance</b>
DI-VAE	scheduling	<ul style="list-style-type: none"><li>- sys: okay, scheduling a yoga activity with Tom for the 8th at 2pm.</li><li>- sys: okay, scheduling a meeting for 6 pm on Tuesday with your boss to go over the quarterly report.</li></ul>
	requests	<ul style="list-style-type: none"><li>- usr: find out if it's supposed to rain</li><li>- usr: find nearest coffee shop</li></ul>
DI-VST	ask schedule info	<ul style="list-style-type: none"><li>- usr: when is my football activity and who is going with me?</li><li>- usr: tell me when my dentist appointment is?</li></ul>
	requests	<ul style="list-style-type: none"><li>- usr: how about other coffee?</li><li>- usr: 11 am please</li></ul>

# **Learning Neural Templates for Text Generation**

**Sam Wiseman**

**Stuart M. Shieber**

**Alexander M. Rush**

School of Engineering and Applied Sciences

Harvard University

Cambridge, MA, USA

{swiseman, shieber, srush}@seas.harvard.edu

# a Sequence of Latent Variables

## Source Entity: Cotto

type[coffee shop], rating[3 out of 5],  
food[English], area[city centre],  
price[moderate], near[The Portland Arms]

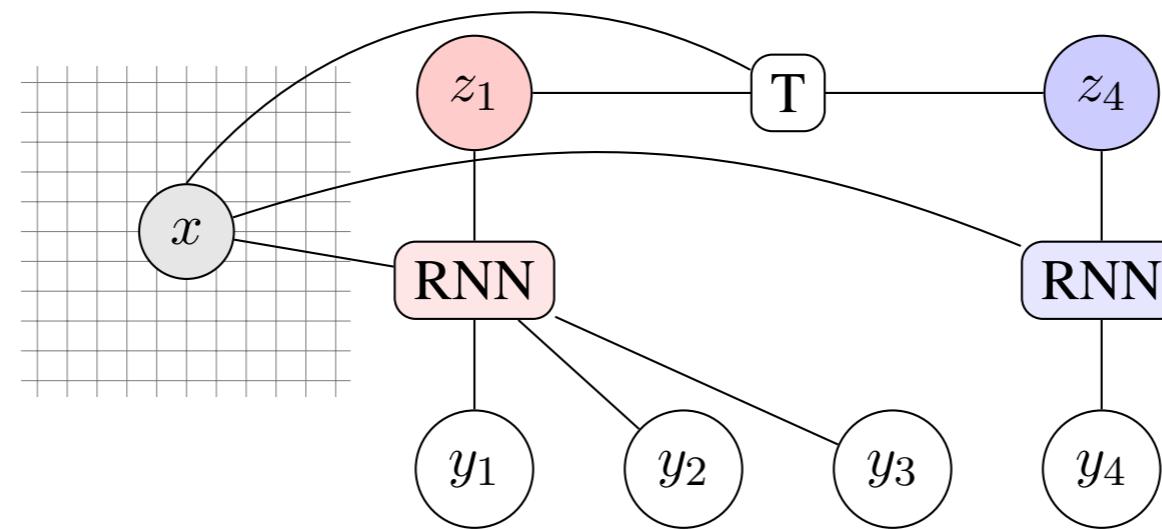
## System Generation:

Cotto is a coffee shop serving English food  
in the moderate price range. It is located  
near The Portland Arms. Its customer rating is  
3 out of 5.

## Neural Template:

The	—	is a	providing	
—		is an	serving	
...		expensive	offering	—
		...		
food	in the	price range	...	
cuisine	with a	price bracket	.	It's
foods	and has a	pricing	.	It is
...	...			The place is
located in the	Its customer rating is	...		
located near	Their customer rating is	...		
near	Customers have rated it	—	.	
...				

# a Sequence of Latent Variables



$$\begin{aligned} p(y, z, l, f \mid x; \theta) &= \prod_{t=0}^{T-1} p(z_{t+1}, l_{t+1} \mid z_t, l_t, x)^{f_t} \\ &\times \prod_{t=1}^T p(y_{t-l_t+1:t} \mid z_t, l_t, x)^{f_t}, \end{aligned}$$

# Model Details (state transition)

$$\begin{array}{c} p(z_{t+1}, l_{t+1} | z_t, l_t, x) \\ \downarrow \\ p(z_{t+1} | z_t, x) \times p(l_{t+1} | z_{t+1}) \\ \searrow \quad \swarrow \\ p(z_{t+1} | z_t, x) \propto AB + C(x_u)D(x_u) \end{array}$$

**Length Distribution** We simply fix all length probabilities  $p(l_{t+1} | z_{t+1})$  to be uniform up to a maximum length  $L$ .<sup>1</sup>

# Model Details (Emission)

$$\begin{aligned} p(y_{t-l_t+1:t} \mid z_t = k, l_t = l, x) &= \\ \prod_{i=1}^{l_t} p(y_{t-l_t+i} \mid y_{t-l_t+1:t-l_t+i-1}, z_t = k, x) & \\ \times p(</\text{seg}> \mid y_{t-l_t+1:t}, z_t = k, x) \times \mathbf{1}_{\{l_t = l\}} & \end{aligned}$$

An Autoregressive Variant

$$p(y_{t-l_t+i} = w \mid y_{1:t-l_t+i-1}, z_t = k, x)$$

# Results

---

## Travellers Rest Beefeater

---

name[Travellers Rest Beefeater], customerRating[3 out of 5],  
area[riverside], near[Raja Indian Cuisine]

---

1. [Travellers Rest Beefeater]<sub>55</sub> [is a]<sub>59</sub> [3 star]<sub>43</sub>  
[restaurant]<sub>11</sub> [located near]<sub>25</sub> [Raja Indian Cuisine]<sub>40</sub> [.]<sub>53</sub>
  2. [Near]<sub>31</sub> [riverside]<sub>29</sub> [,]<sub>44</sub> [Travellers Rest Beefeater]<sub>55</sub>  
[serves]<sub>3</sub> [3 star]<sub>50</sub> [food]<sub>1</sub> [.]<sub>2</sub>
  3. [Travellers Rest Beefeater]<sub>55</sub> [is a]<sub>59</sub> [restaurant]<sub>12</sub>  
[providing]<sub>3</sub> [riverside]<sub>50</sub> [food]<sub>1</sub> [and has a]<sub>17</sub>  
[3 out of 5]<sub>26</sub> [customer rating]<sub>16</sub> [.]<sub>2</sub> [It is]<sub>8</sub> [near]<sub>25</sub>  
[Raja Indian Cuisine]<sub>40</sub> [.]<sub>53</sub>
  4. [Travellers Rest Beefeater]<sub>55</sub> [is a]<sub>59</sub> [place to eat]<sub>12</sub>  
[located near]<sub>25</sub> [Raja Indian Cuisine]<sub>40</sub> [.]<sub>53</sub>
  5. [Travellers Rest Beefeater]<sub>55</sub> [is a]<sub>59</sub> [3 out of 5]<sub>5</sub>  
[rated]<sub>32</sub> [riverside]<sub>43</sub> [restaurant]<sub>11</sub> [near]<sub>25</sub>  
[Raja Indian Cuisine]<sub>40</sub> [.]<sub>53</sub>
- 

---

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

# Discrete Patterns Mining for Text Generation

unsupervised / semi-supervised

discrete / semi-discrete

single/ multiple latent variable(s)

sequential / hierarchical combinations

Determined before/during generation