

*‘Life is translation, mostly failed by contempt’*

- Craycray

---

# CRAYBOT

How to say dumb things  
inefficiently.

---

---

# Goal: General Purpose Bot

---

- ❖ Automate customer responses
- ❖ Generate positive reviews for your product / page.
- ❖ Launch massive trolling campaign to destabilize foreign powers.

---

# Goal: General Purpose Bot

---

```
In [15]: craycray=CRAYBOT(sent_length=10,n_samples=10,hidden_size=100,n_layers=4,max_iter=10,encoder=w2v_bk)
```

Let's feed the bot some training data.

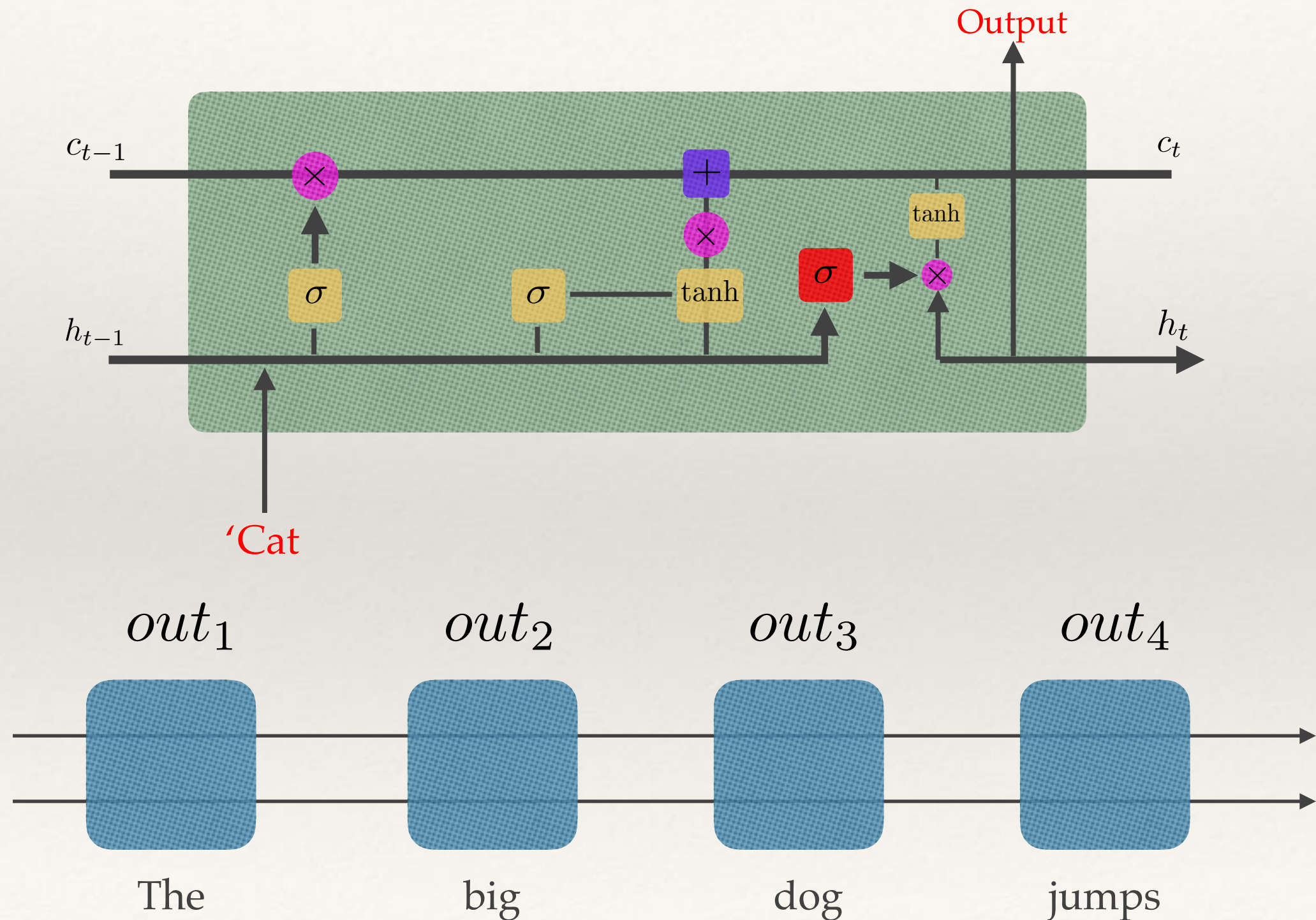
```
In [16]: craycray.fit(string)
```

```
In [40]: craycray.predict(['You are'],7)
```

```
/Users/williamcottrell172/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:145: UserWarning: Implicit dimension choice for softmax has been deprecated. Change the call to include dim=X as an argument.
```

```
Out[40]: 'You are mammy babyish deuce knocked benevolently snub scorn'
```

# LSTM: Long Short Term Memory





---

# Target: Minimize Cross Entropy

---

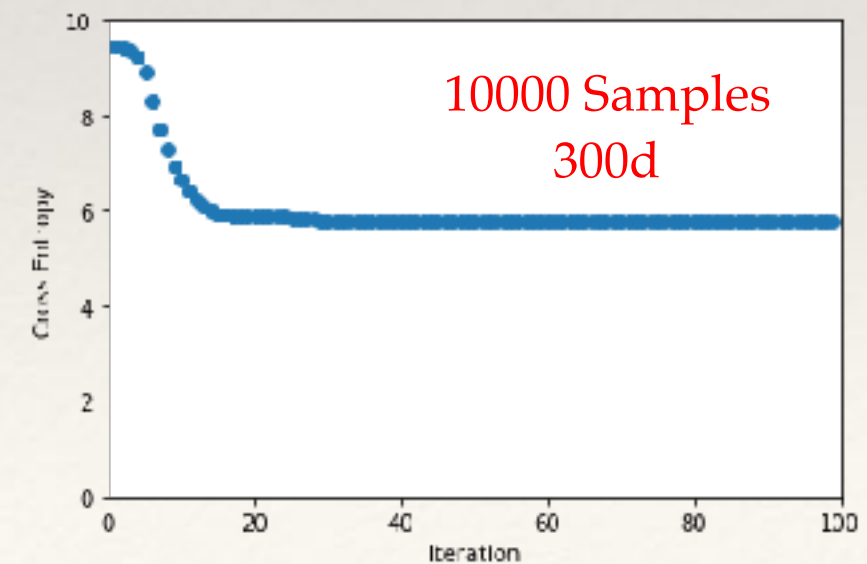
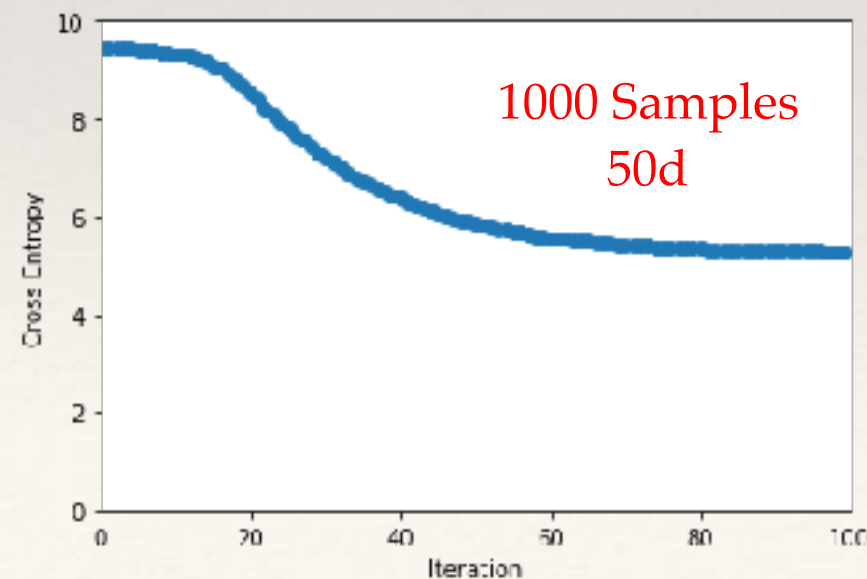
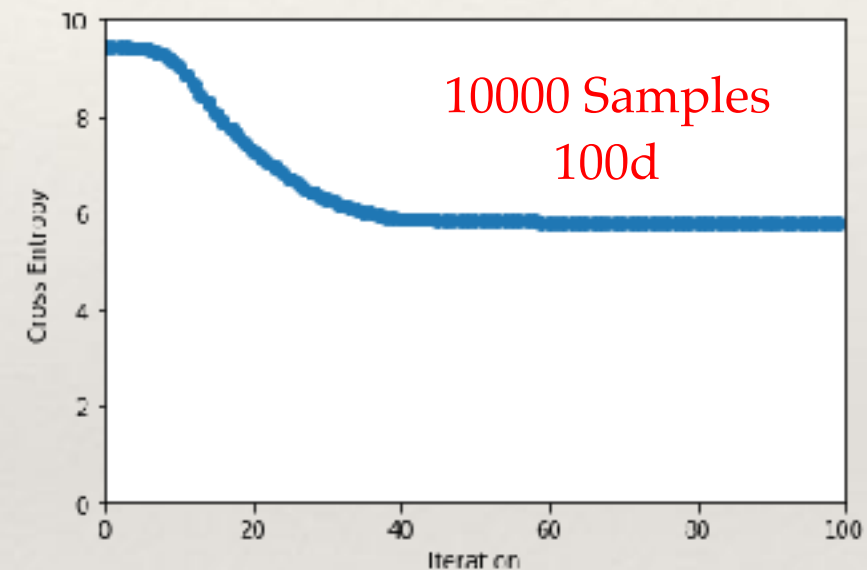
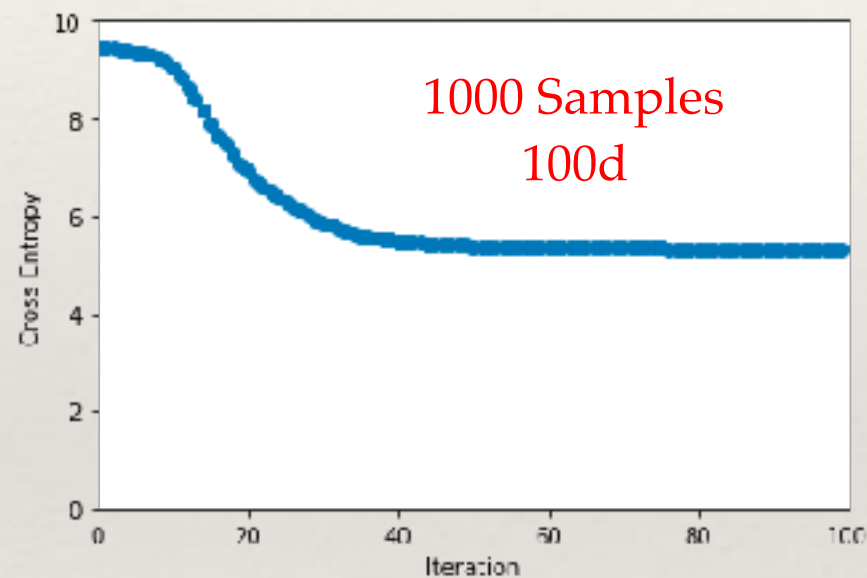
$$\begin{aligned} H(p, q) &= - \sum_x p(x) \ln (q(x)) \\ &= S(p) + D_{KL}(p||q) \end{aligned}$$

- ❖  $p$  = true probability distribution
- ❖  $q$  = distribution that is predicted.
- ❖  $H(p, q)$  = average number of additional bits needed to find find the right  $x$ .

# Typical Training Run

Now we instantiate a bot.

```
In [9]: craycray=CRAYBOT(sent_length=10,n_samples=1000,hidden_size=100,n_layers=3,max_iter=100,encoder=w2v_bk)
```



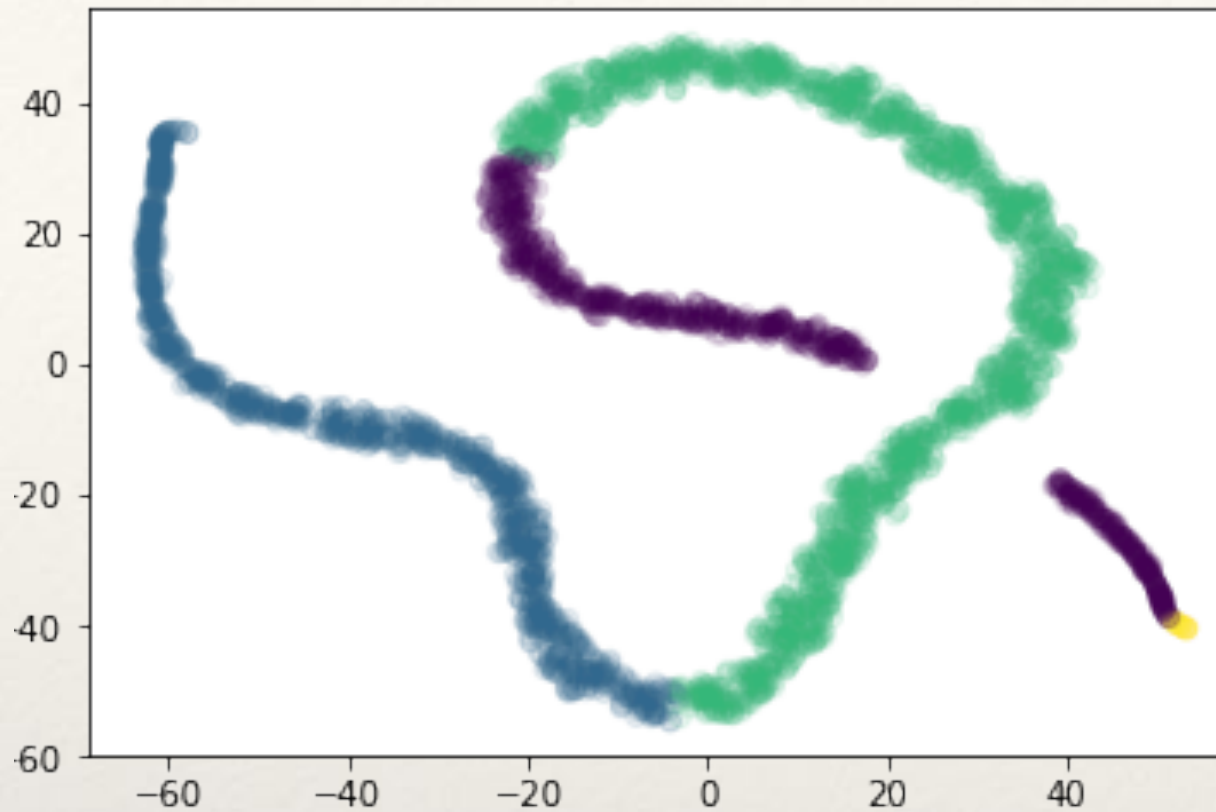
---

# Meet Team CRAYBOT

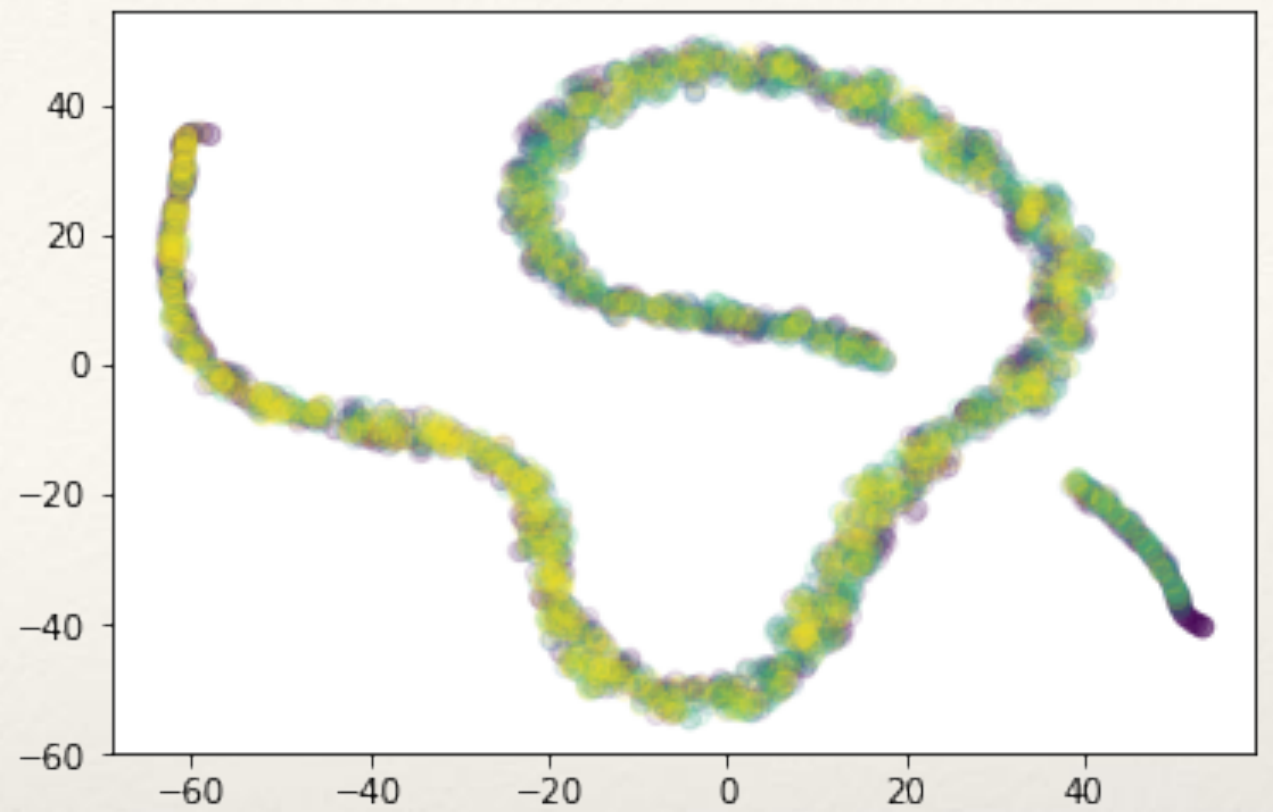
---

- ❖ **craycray** (Brothers Karamazov):
  - ❖ *'You are rouble on associated exalted'*
- ❖ **bot\_bot** (Humans talking to other chatbots):
  - ❖ 所以佛教，道教，各宗教皆教你靜靜靜
- ❖ **conspiro\_bot** (Chronicles of the Illuminati):
  - ❖ *'He is back-all discerned - fertile with Gutenberg'.*
- ❖ **wonder\_bot** (Alice in Wonderland):
  - ❖ *'He said with the hatter, "Before prize, but yes, forgotten" '*
- ❖ **markov\_bot** (... not really RNN but likes to pretend....):
  - ❖ *'Will never be fulfilled, if your happiness depends on money you.'*





TSNE Plot. Coloring via KMeans



TSNE Plot - Labeled by bot.

*Lorem Ipsum Dolor - (craybot could have said this!)*

# Clustering Bots

Why? Because we can.

---

# Future Improvements

---

- ❖ More texts, more training, more hyper-parameter tuning.
- ❖ Better training / cooling schedule.
- ❖ Better text pre-processing.
- ❖ Include more effective embedding layer to take advantage of pre-trained vectors without throwing away new words.
- ❖ Make tone of CRAYBOT tunable, e.g., include a parameter which specifies how much of each text is sampled.