Learning Language from Characters

Wang Ling

Why go Character-level?

Standardization	The act of making something conform to a standard.
Diversification	The action of making or becoming more diverse or varied.
Frenchification	???

Why go Character-level?

Standardization	The act of making something conform to a standard.
Diversification	The action of making or becoming more diverse or varied.
Frenchification	To make something more French in appearance or character.

Why go Character-level?

tweetroduce	???
attwaction	???
twalking	???

Why go Character-level?

tweetroduce	Slang term used to describe the situation in which a user introduces one follower to another.
attwaction	???
twalking	???

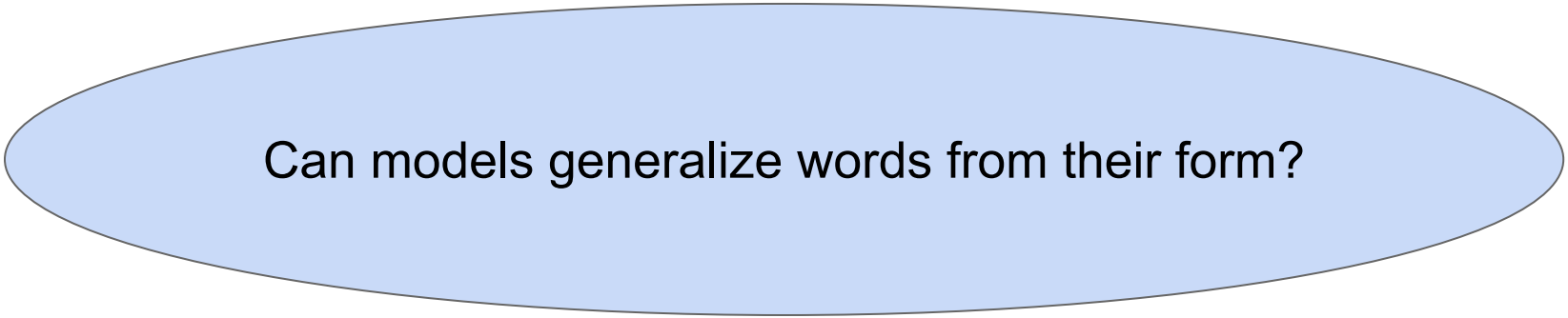
Why go Character-level?

tweetroduce	Slang term used to describe the situation in which a user introduces one follower to another.
attwaction	Slang term used to describe an attraction between two users.
twalking	???

Why go Character-level?

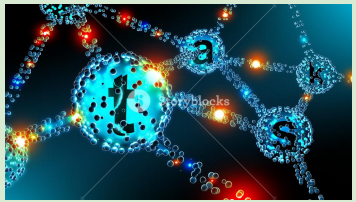
tweetroduce	Slang term used to describe the situation in which a user introduces one follower to another.
attwaction	Slang term used to describe an attraction between two users.
twalking	Someone who is walking while they tweet, using a mobile device.

Why go Character-level?



Character-level
Language Understanding

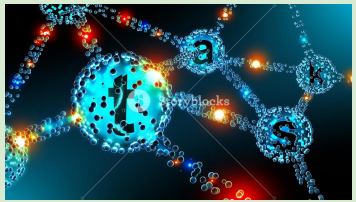
B	a	z	i	n	g	a	!



Character-level
Language Generation

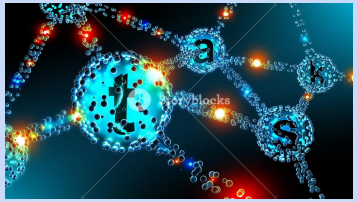
Character-level Language Understanding

B	a	z	i	n	g	a	!



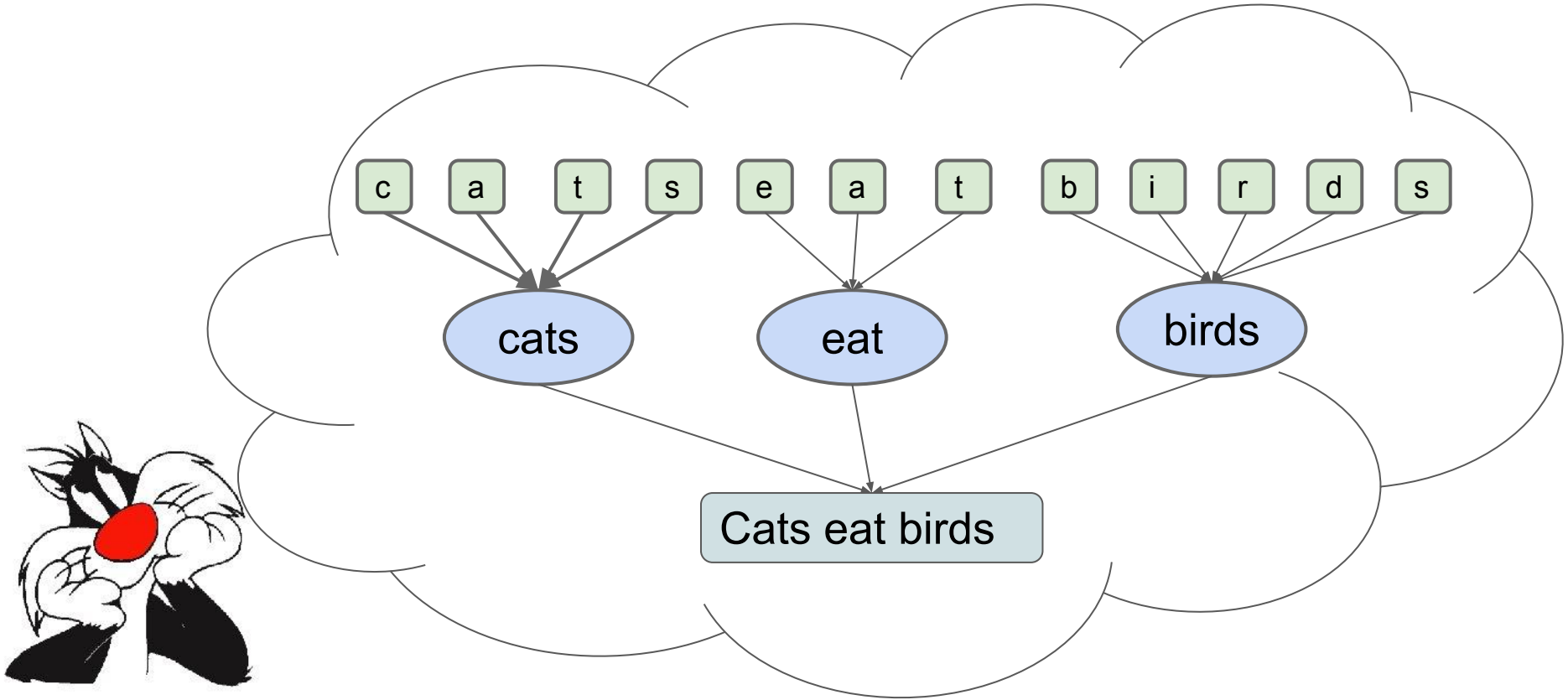
Character-level Language Generation

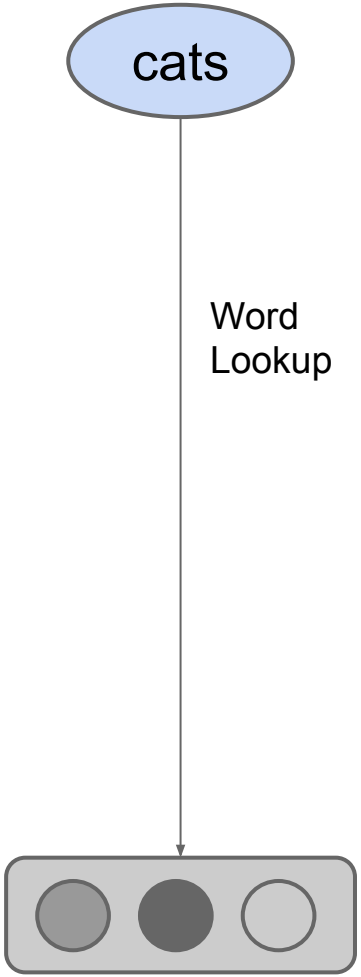
B	a	z	i	n	g	a	!



巴	津	加	!

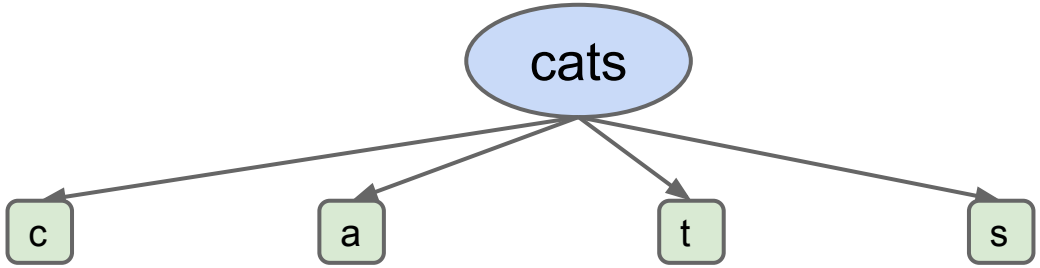
Character-based Language Understanding

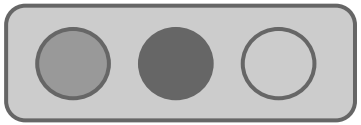




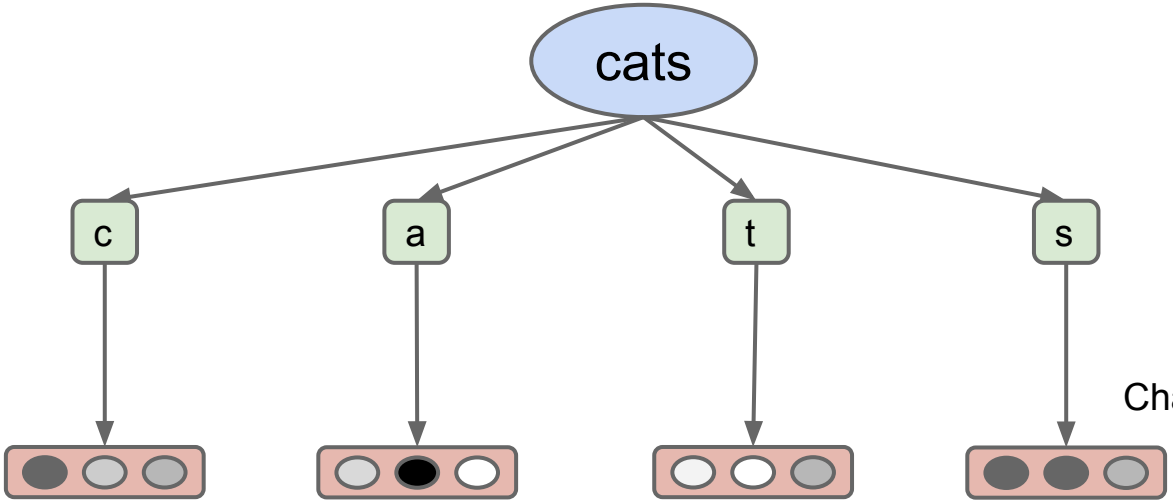
id	word	θ_1	θ_2	…	θ_K
1	the	0.2	0.1	...	0.3
2	cat	0.3	-0.2	...	-0.1
3	bite	-0.1	0.3	...	0.3
...	...	...	...	...	….
V	dog	-0.2	-0.3	...	0.4

Ling et al. 2015

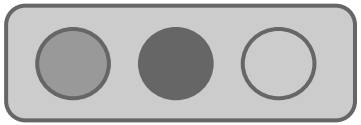




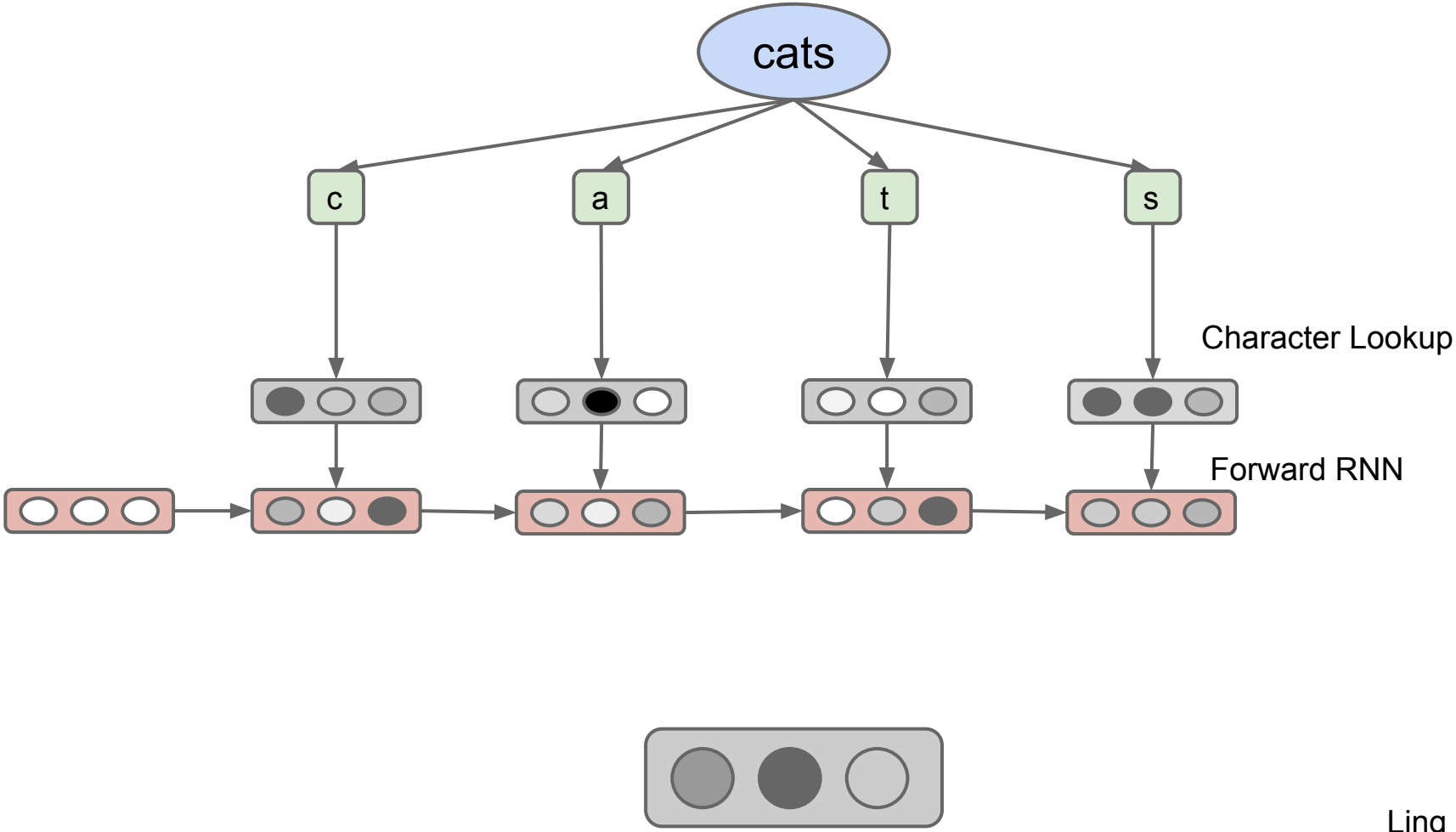
Ling et al. 2015



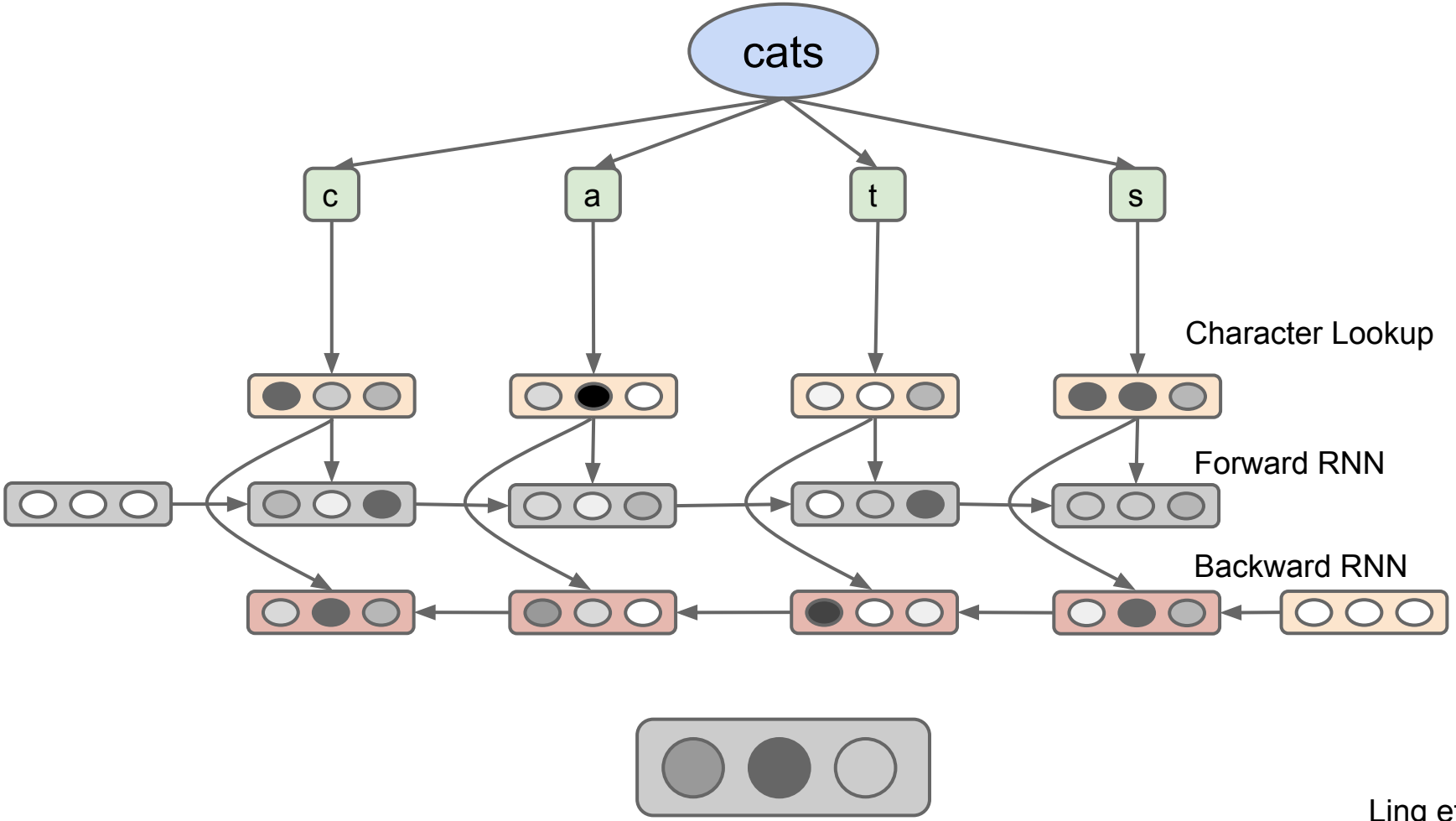
Character Lookup



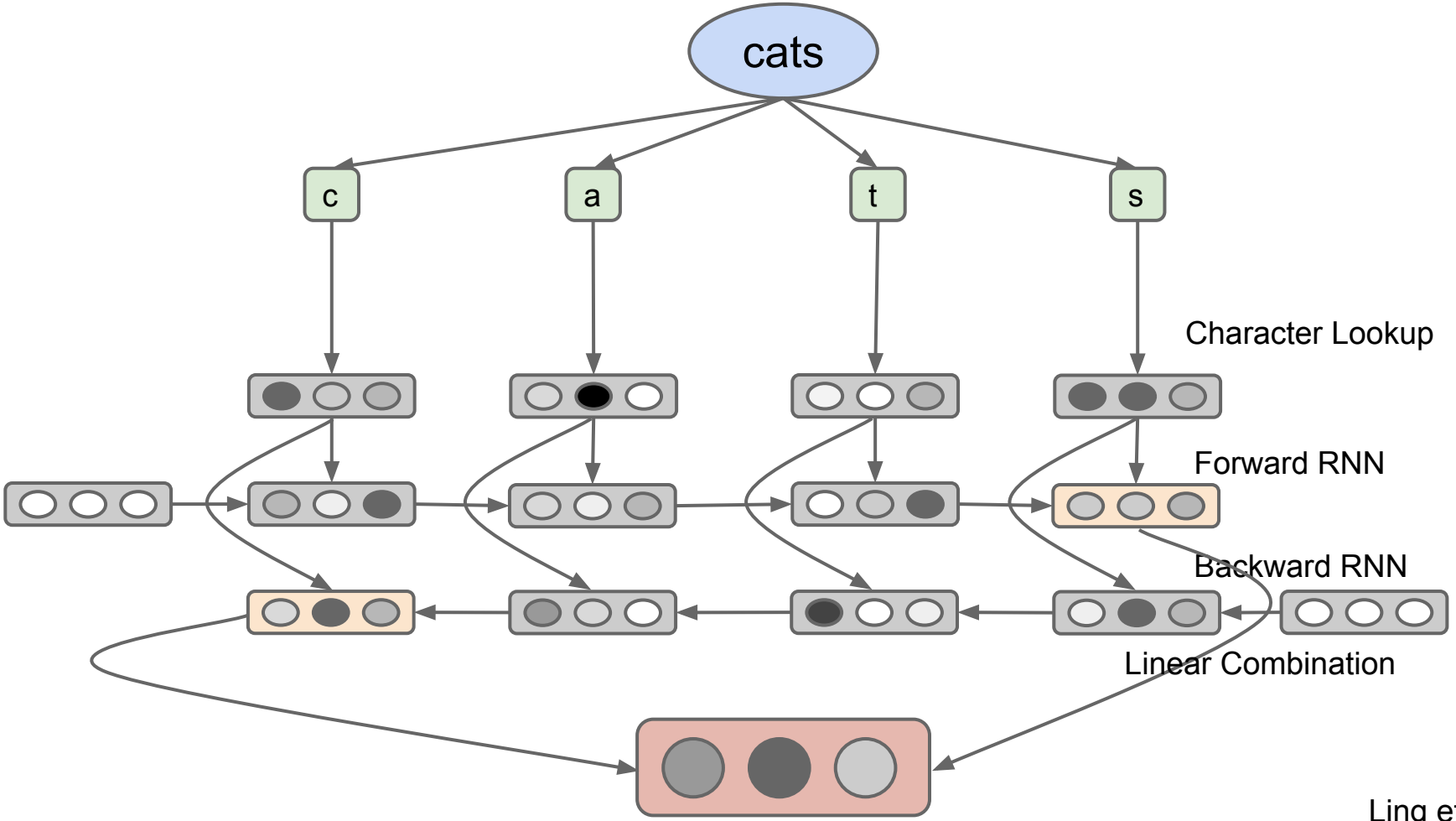
Ling et al. 2015



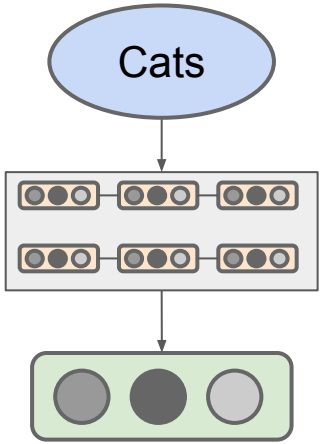
Ling et al. 2015

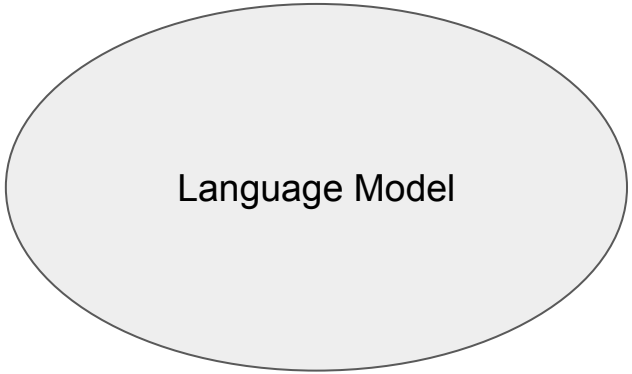


Ling et al. 2015

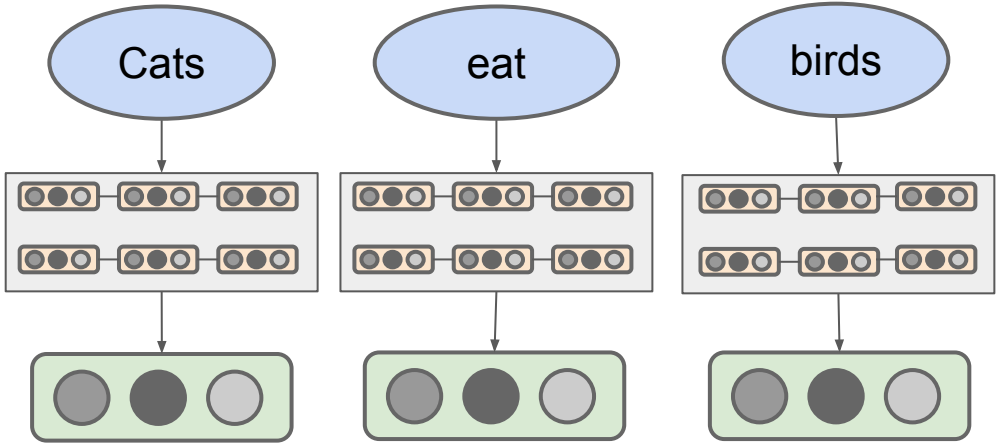


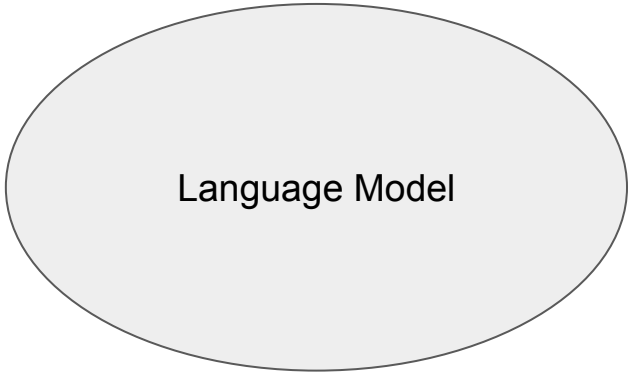
Ling et al. 2015



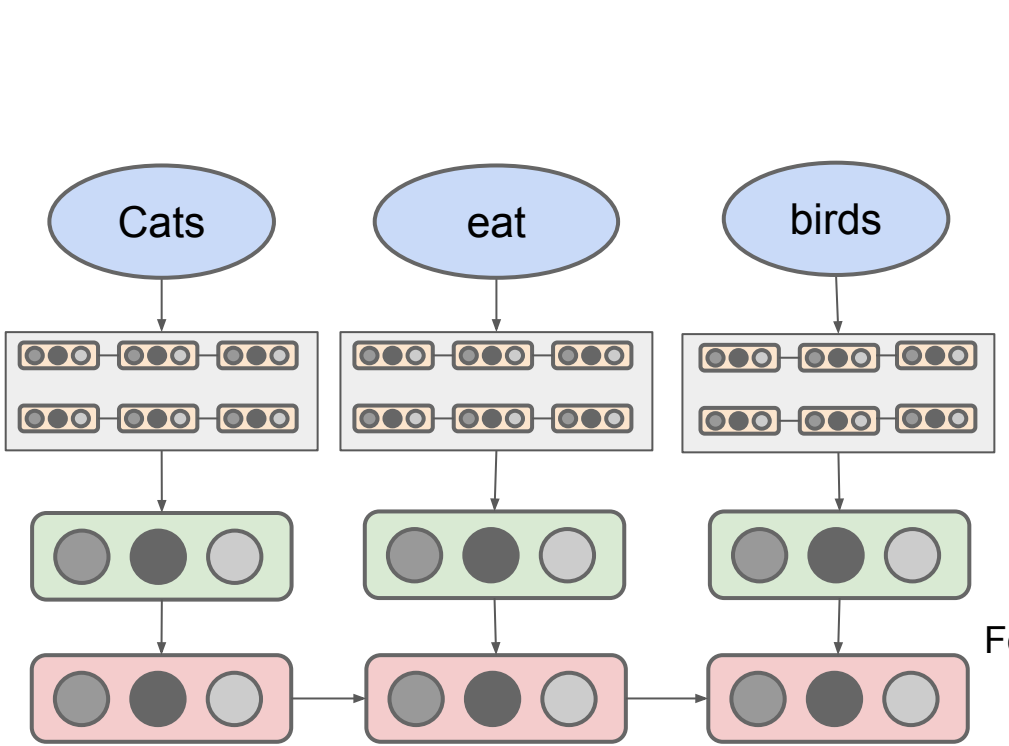


Ling et al. 2015



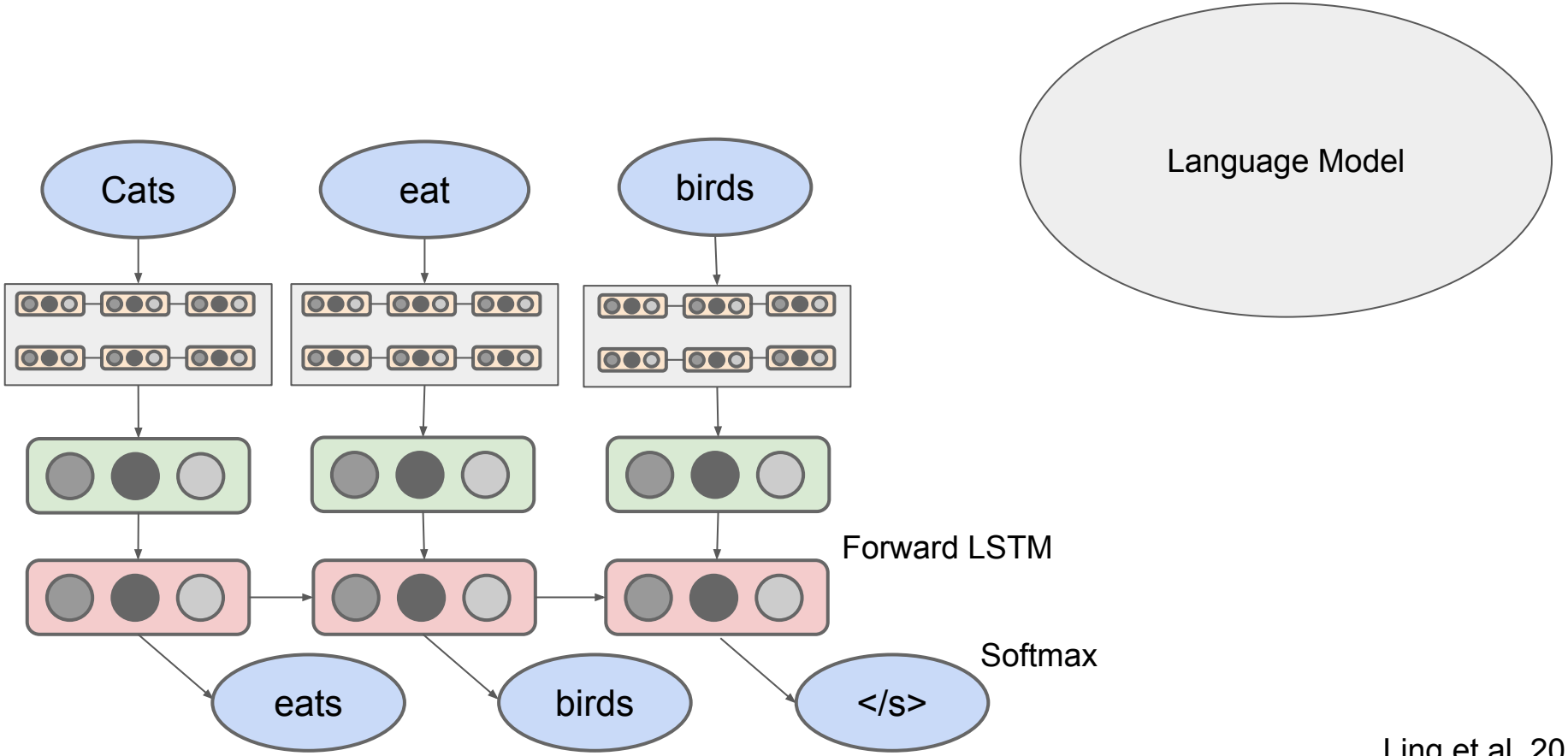


Ling et al. 2015

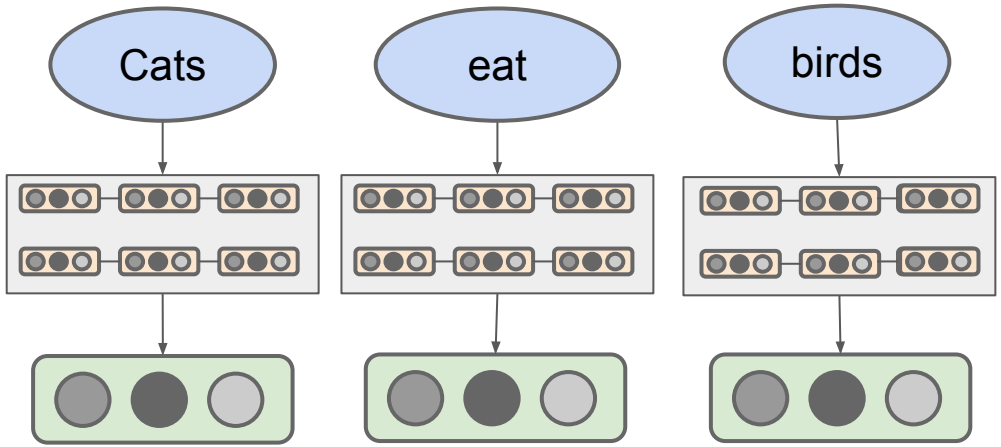


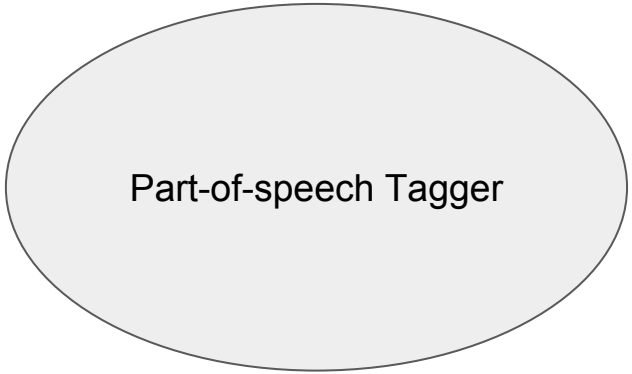
Language Model

Ling et al. 2015

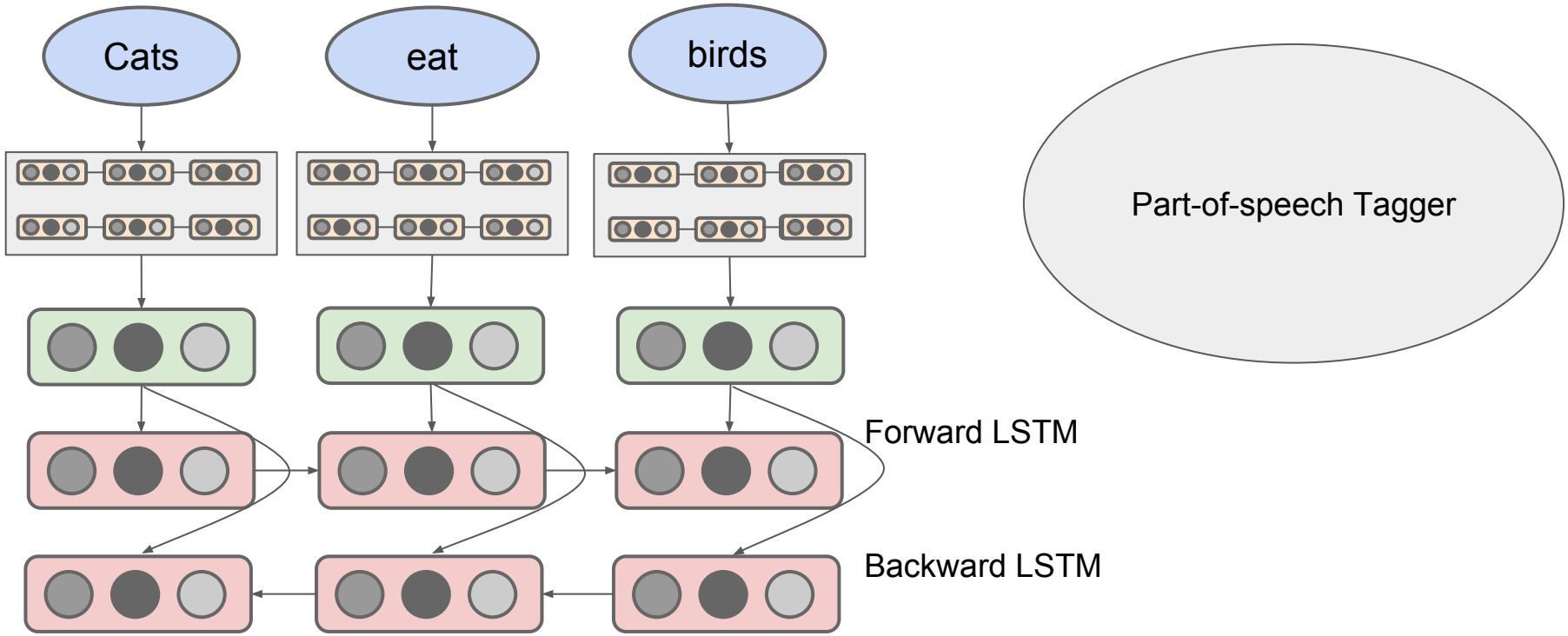


Ling et al. 2015

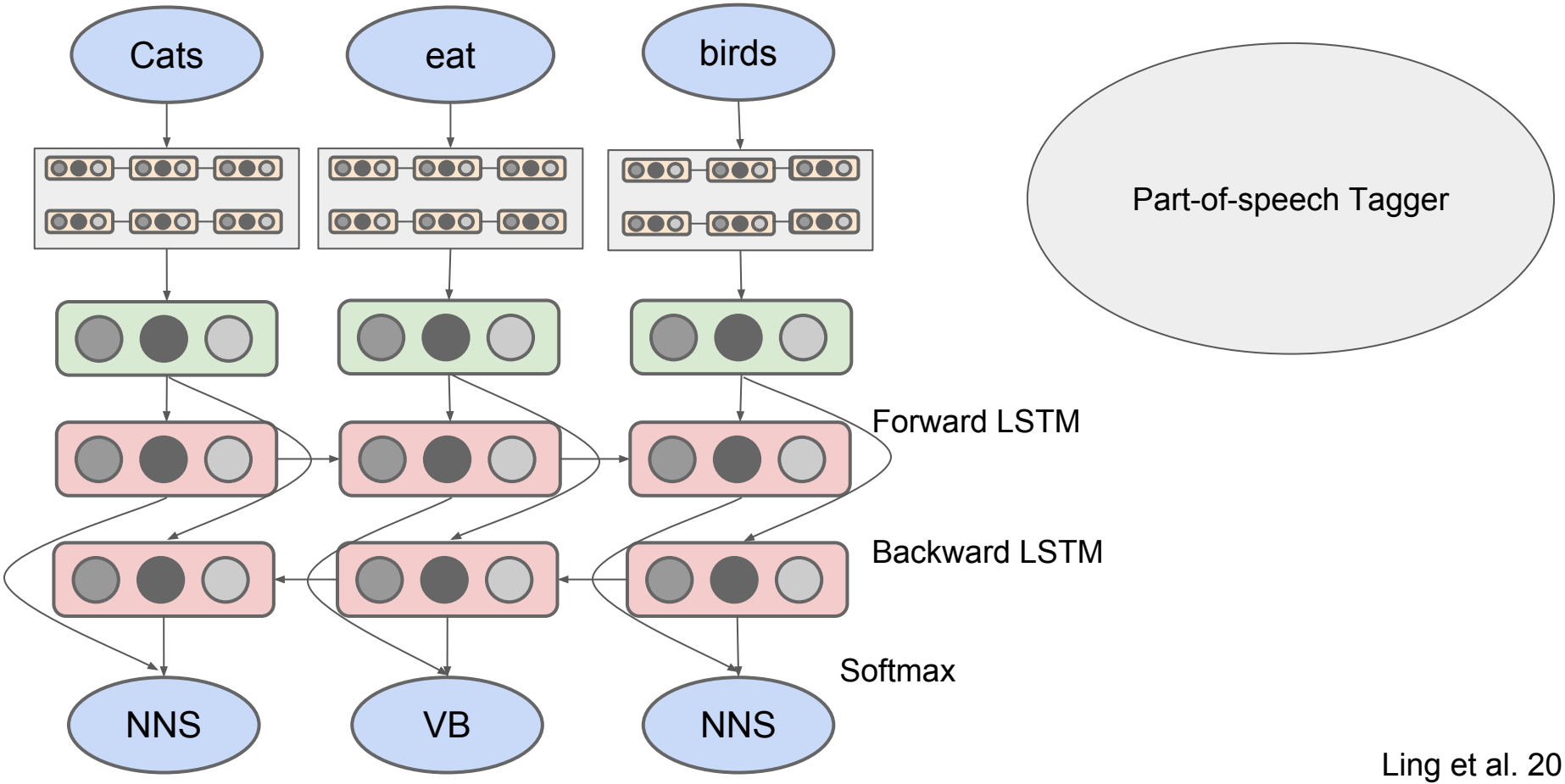




Ling et al. 2015



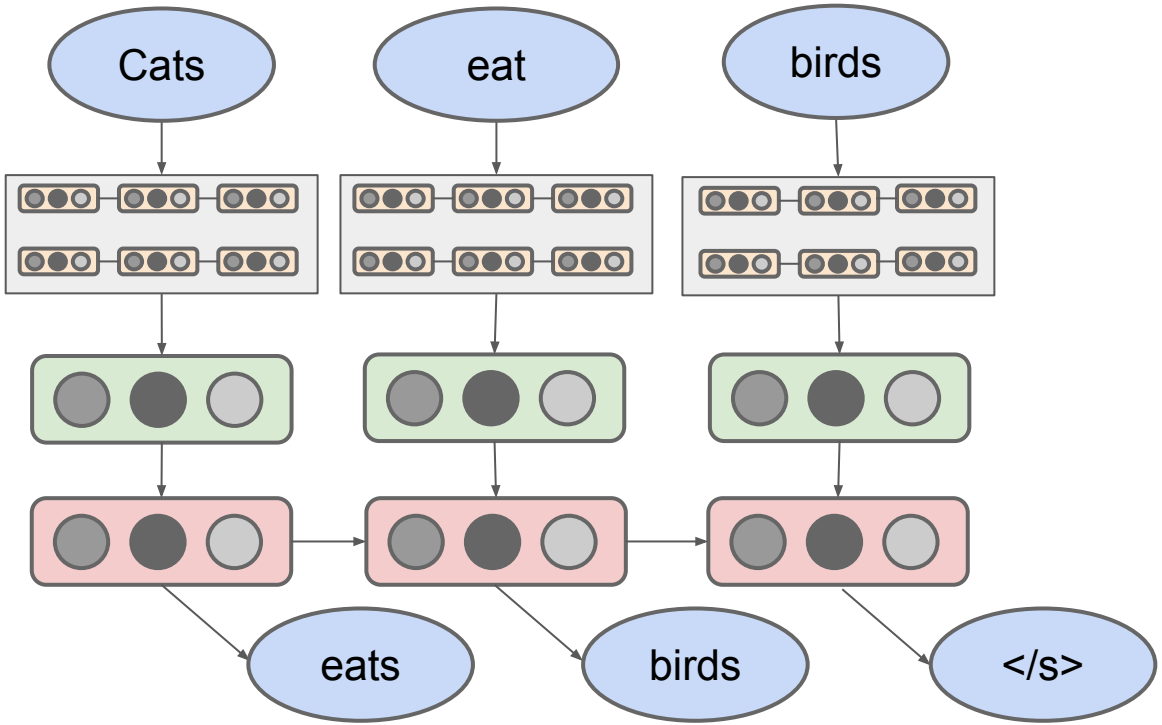
Ling et al. 2015



Ling et al. 2015

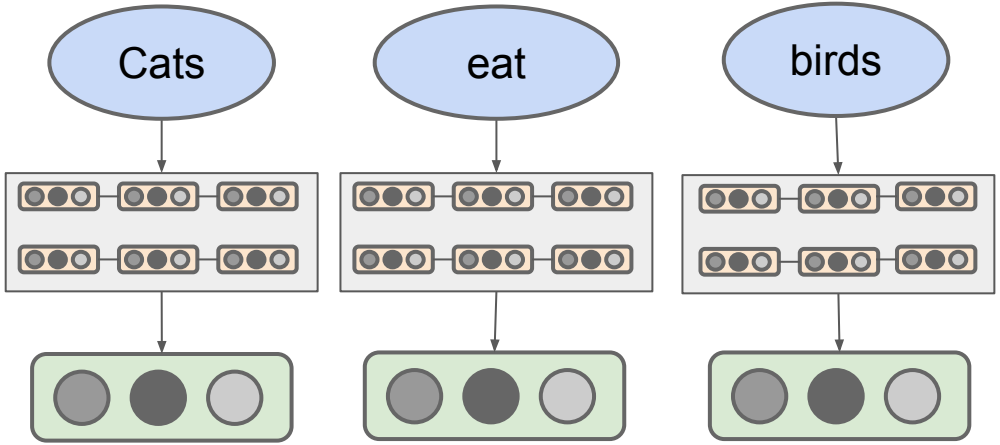
	POS (WSJ)	LM (PTB)
Word Lookup	96.97	115.17
Char-RNN	95.66	136.62
Char-Bi-RNN	95.93	129.95
Char-LSTM	97.12	94.41
Char-Bi-LSTM	**97.36**	**91.89**

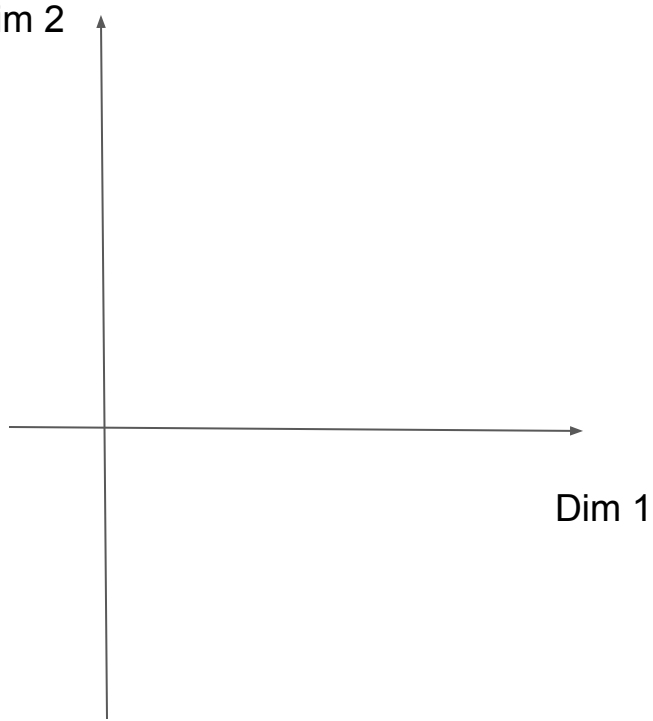
Ling et al. 2015



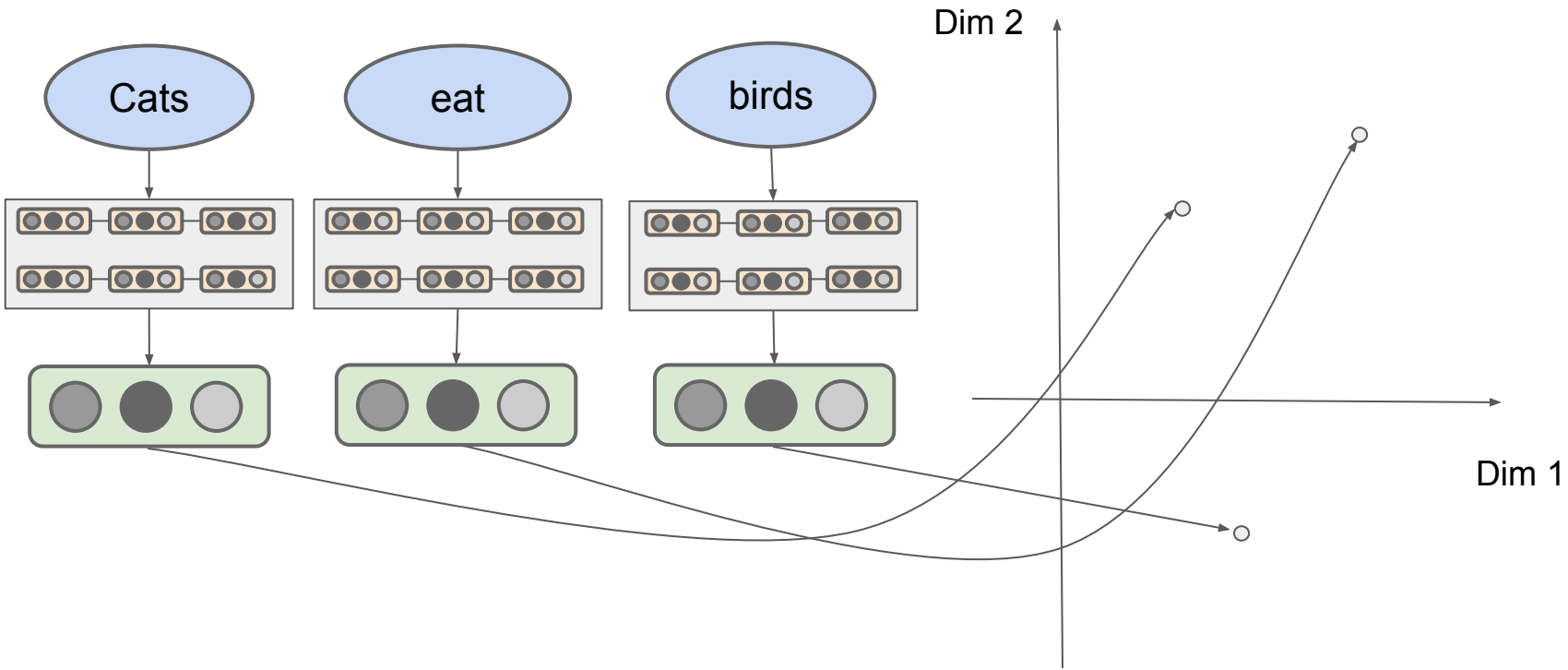
Train Large LM in Wikipedia

Ling et al. 2015

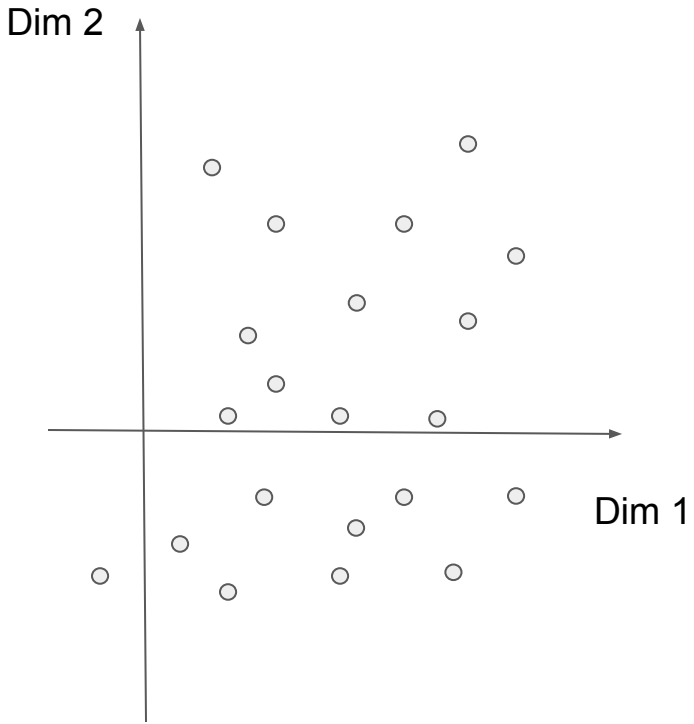




Ling et al. 2015

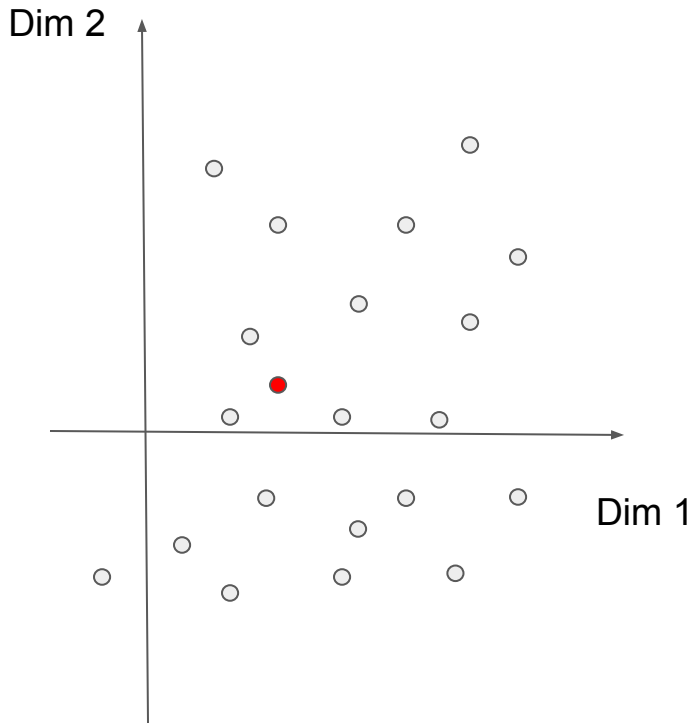


Ling et al. 2015



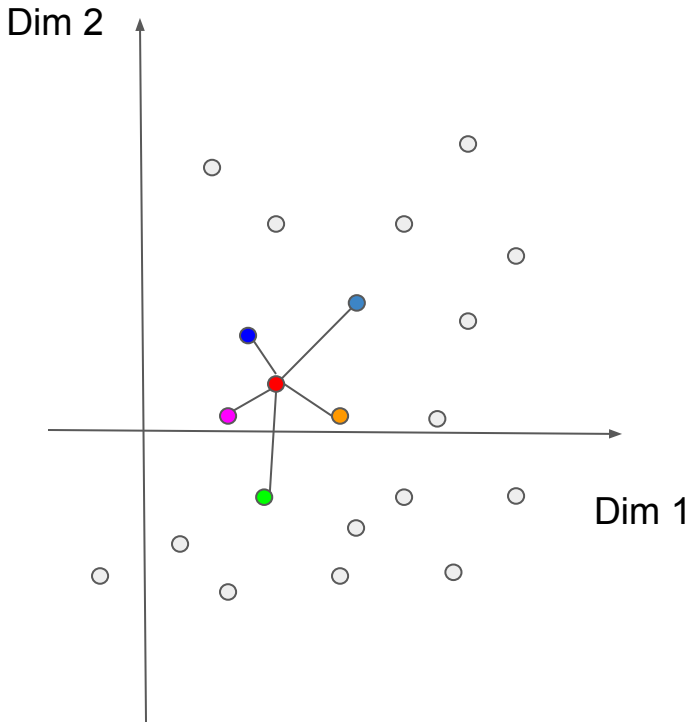
Ling et al. 2015

Input Word	**cats**



Ling et al. 2015

Input Word	**cats**
Top-5	dogs animals rats leopards cat



Ling et al. 2015

Input Word	**Increased**	**John**
Top-5	increasing improved increase decreased decrease	George Richard James Edward Charles

Ling et al. 2015

Word Lookup

Input Word	**Increased**	**John**
Top-5	increasing improved increase decreased decrease	George Richard James Edward Charles

Char-RNN

Input Word	**Increased**	**John**
Top-5	decr**eased** rel**eased** **eased** pl**eased** c**eased**	James Lewis**ohn** Mendelss**ohn** Johnnie Johnny

Ling et al. 2015

Word Lookup

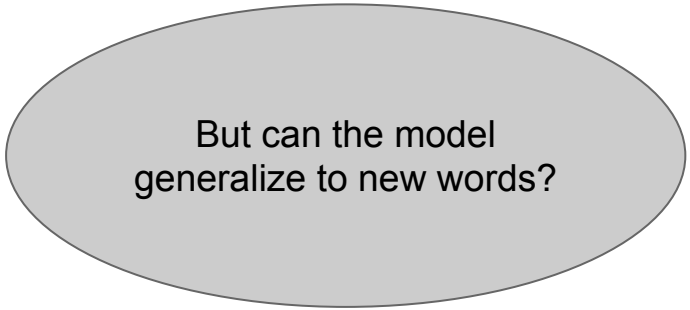
Input Word	**Increased**	**John**
Top-5	increasing improved increase decreased decrease	George Richard James Edward Charles

Char LSTM

Input Word	**Increased**	**John**
Top-5	increasing improved increase decreased decrease	Richard George James Robert Edward

Ling et al. 2015

Standardization	The act of making something conform to a standard.
Diversification	The action of making or becoming more diverse or varied.
Frenchification	???



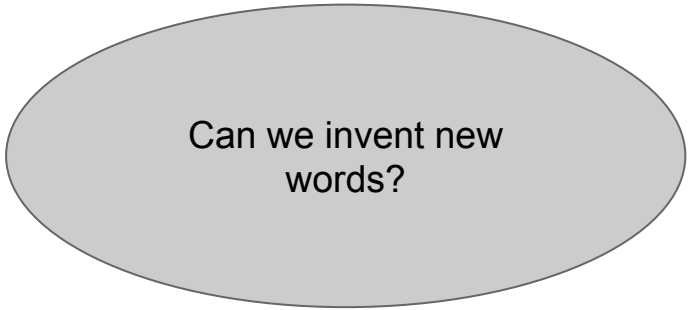
Ling et al. 2015

Input Word	frenchification
Top-5	collectivization stagnation liberalization globalization internationalization

Ling et al. 2015

Input Word	frenchification	attwaction
Top-5	collectivization stagnation liberalization globalization internationalization	attraction proximity attractions beauty nature

Ling et al. 2015



Ling et al. 2015

Noah

Ling et al. 2015

+

Noah

Shire

Ling et al. 2015

+

=

Noah

Shire

Noahshire

Ling et al. 2015

Noah

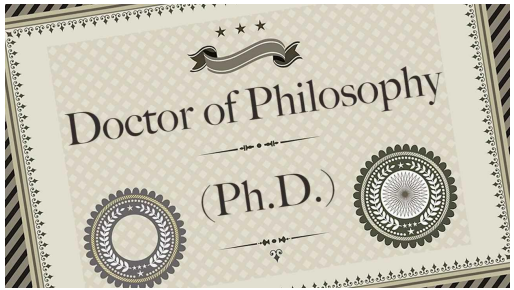
+

Shire

=

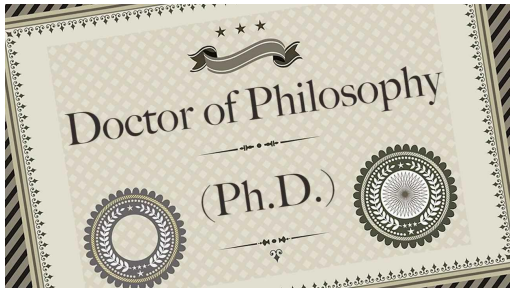
Noahshire
Nottinghamshire Bucharest Saxony Johannesburg Gloucestershire

Ling et al. 2015



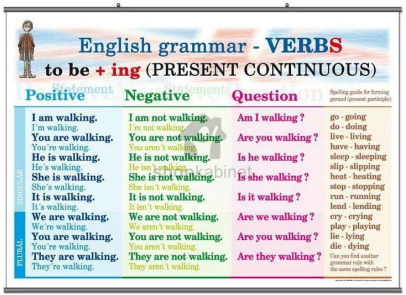
phd

Ling et al. 2015



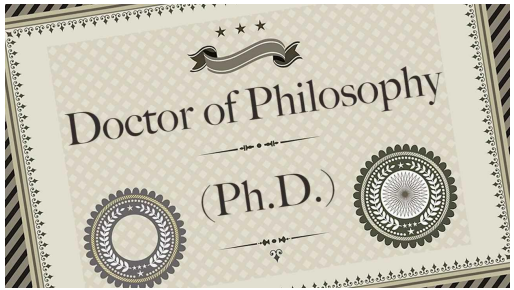
phd

+



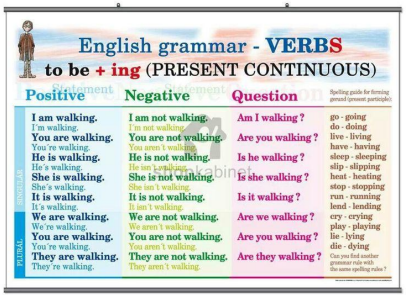
ing

Ling et al. 2015



phd

+

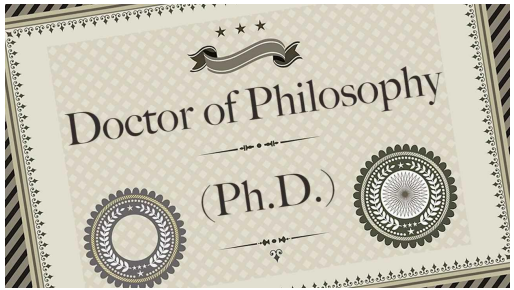


ing

=

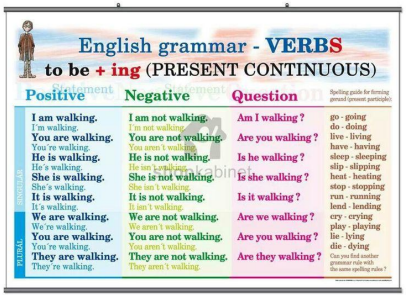
phding

Ling et al. 2015



phd

+

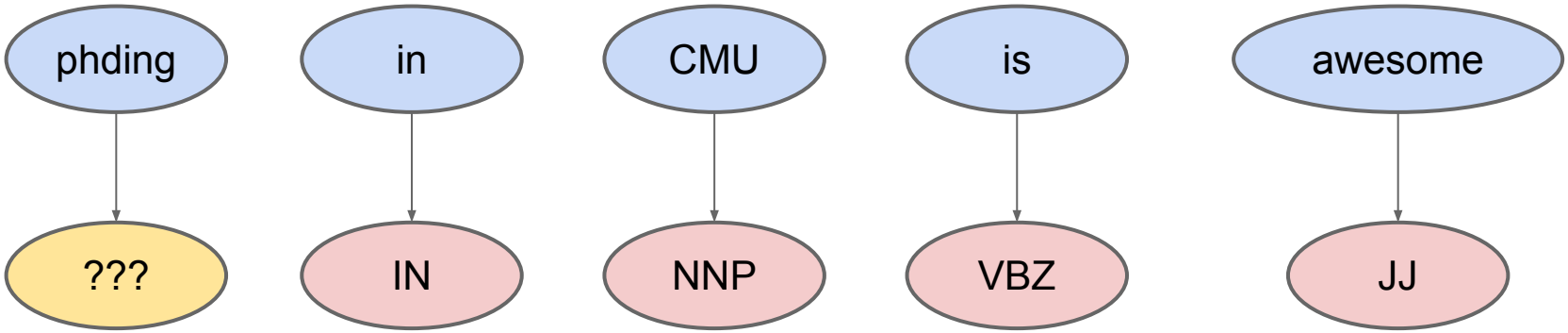


ing

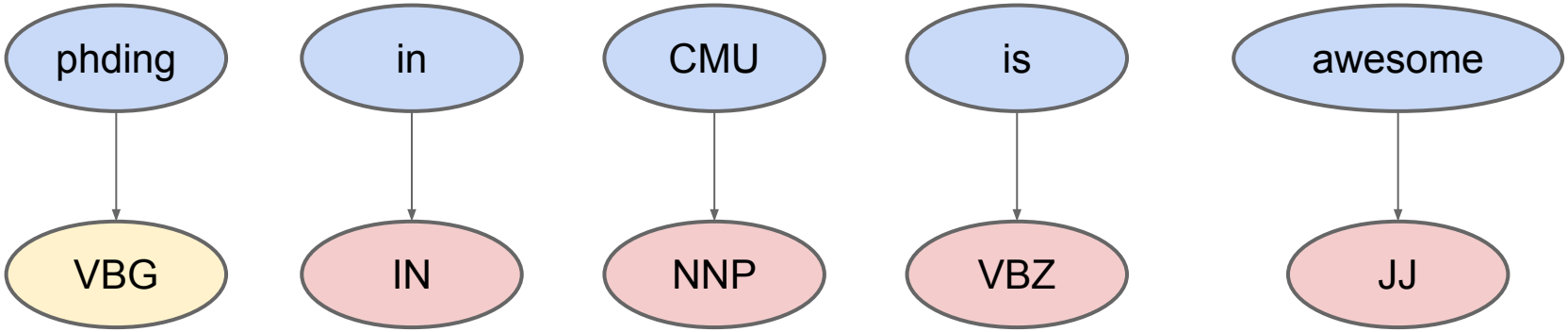
=

phding
mixing modelling styling blaming christening

Ling et al. 2015

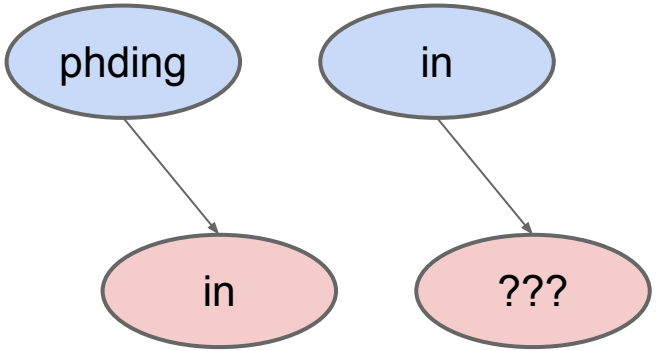


Ling et al. 2015



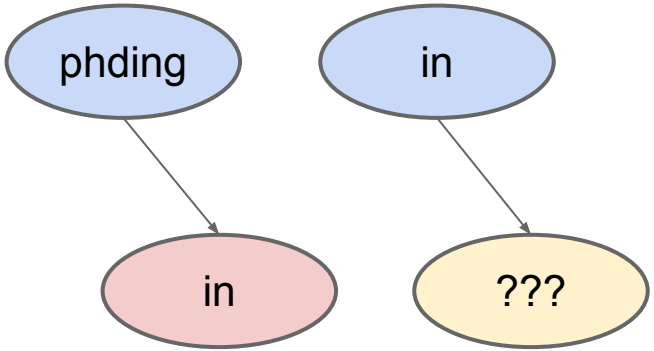
phding
mixing modelling styling blaming christening

Ling et al. 2015



phding
mixing modelling styling blaming christening

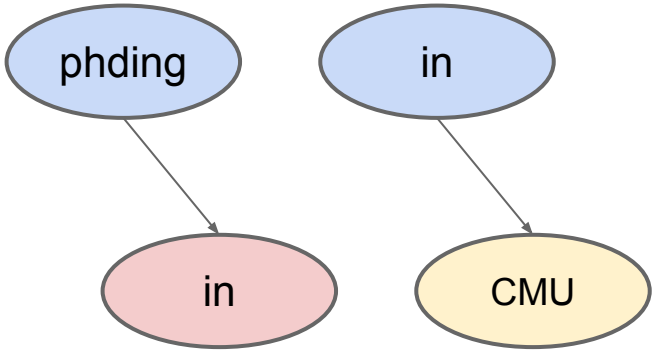
Ling et al. 2015



phding
mixing modelling styling blaming christening

1st	college
4th	univ
6th	Princeton
11th	Yale
12th	Tsinghua
48th	CMU

Ling et al. 2015

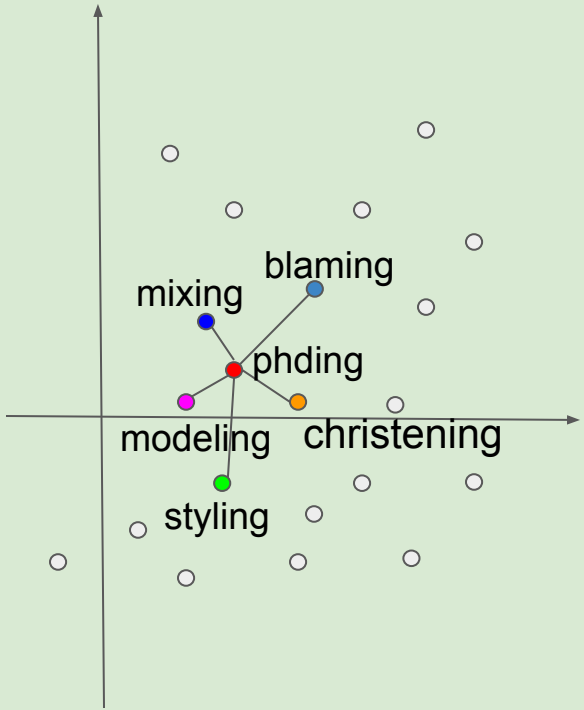


phding
mixing modelling styling blaming christening

1st	college
4th	univ
6th	Princeton
11th	Yale
12th	Tsinghua
48th	CMU

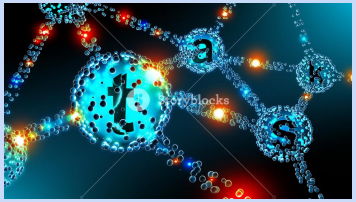
Ling et al. 2015

Character-level Language Understanding



Character-level Language Generation

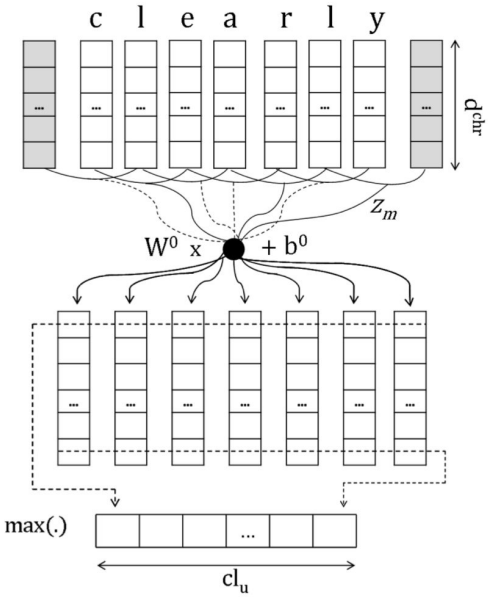
B	a	z	i	n	g	a	!



巴	津	加	!

Related Work

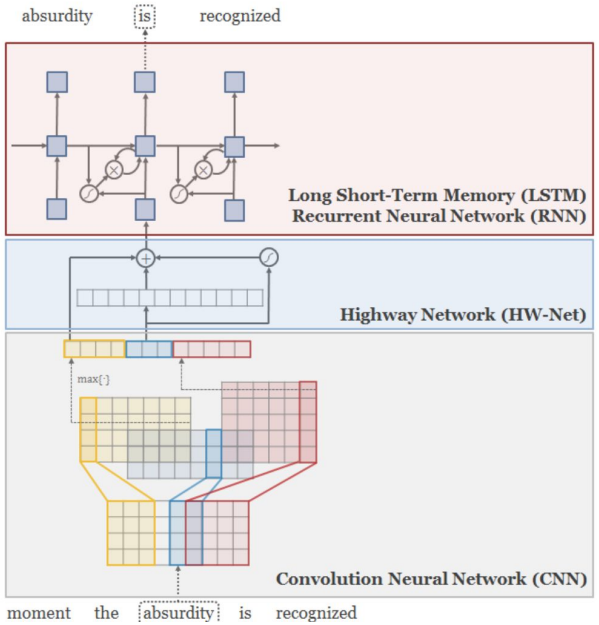
- First proposed character composition model for POS tagging



Dos Santos et al. 2014

Related Work

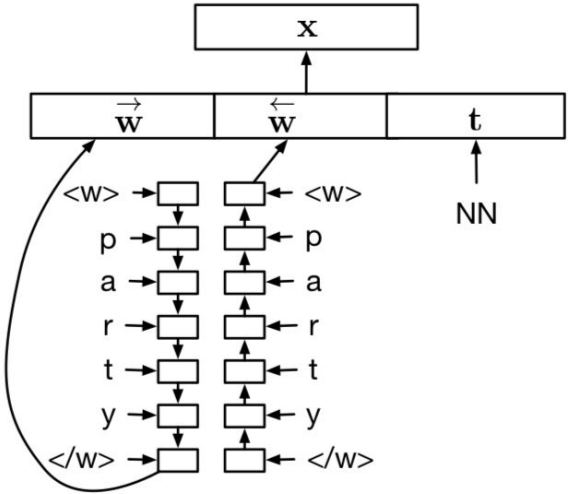
- Language model using CNN-based character composition



Kim et al. 2015

Related Work

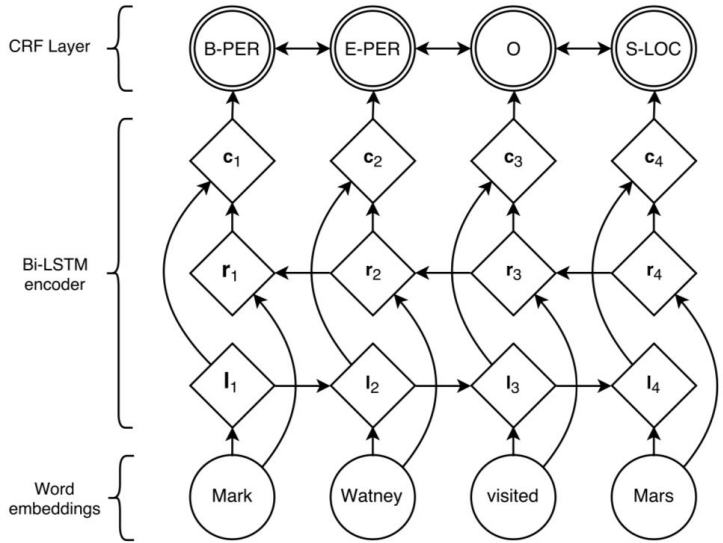
- LSTM-based character model for stack-based Dependency Parsing



Ballesteros et al. 2015

Related Work

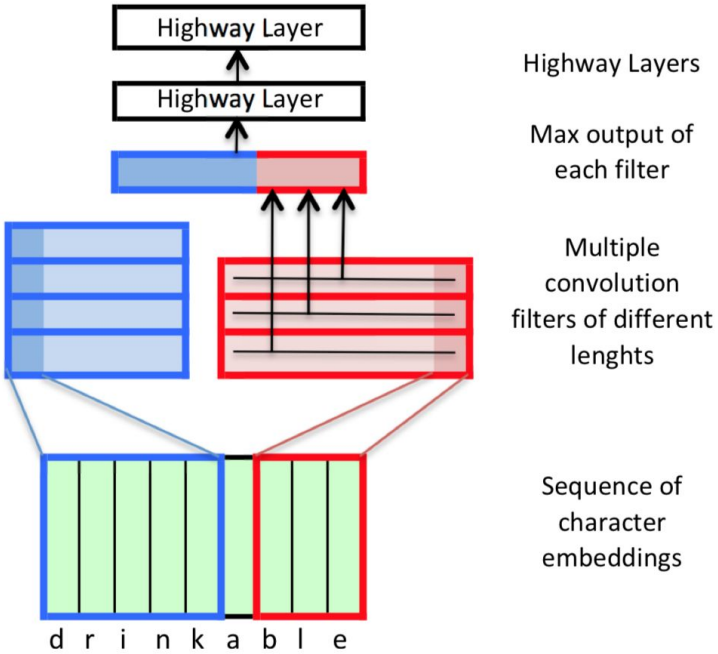
- LSTM-based character model for a LSTM-CRF labelled for NER



Lample et al. 2016

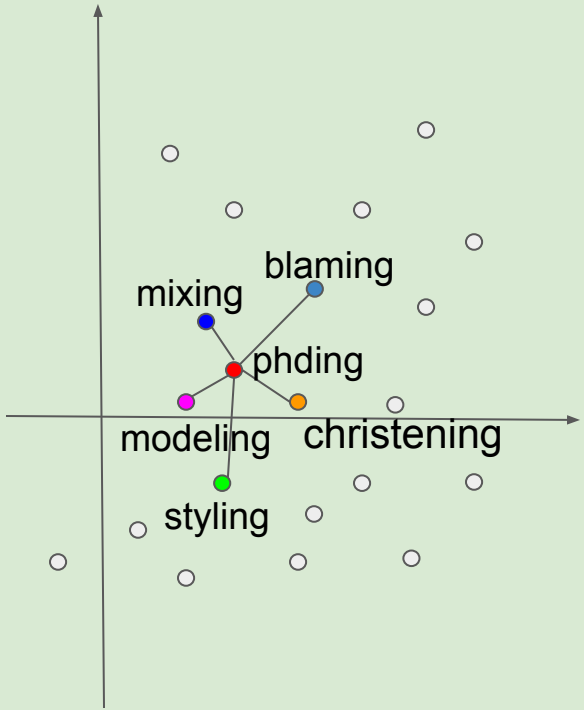
Related Work

- CNN-based Encoder for Machine Translation



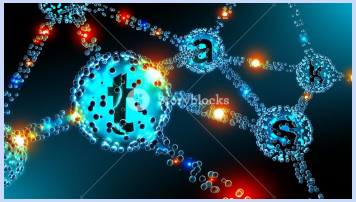
Costa-jussa et al. 2016

Character-level Language Understanding



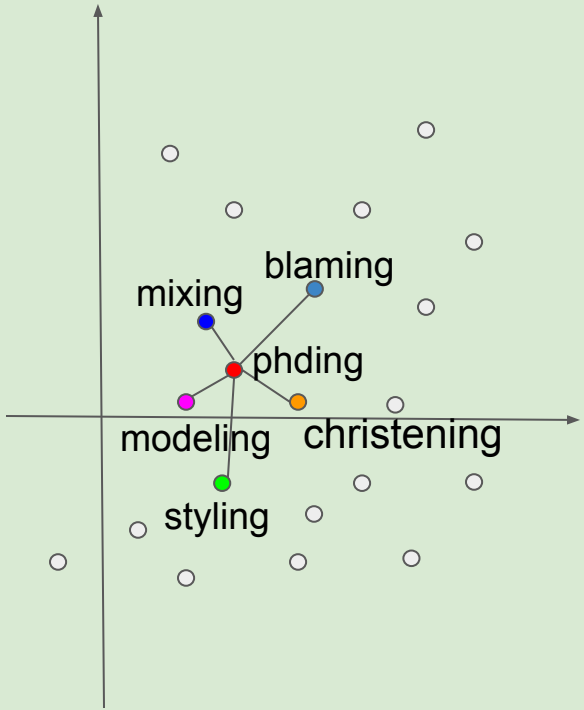
Character-level Language Generation

B	a	z	i	n	g	a	!



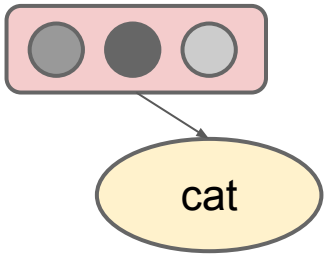
巴	津	加	!

Character-level
Language Understanding



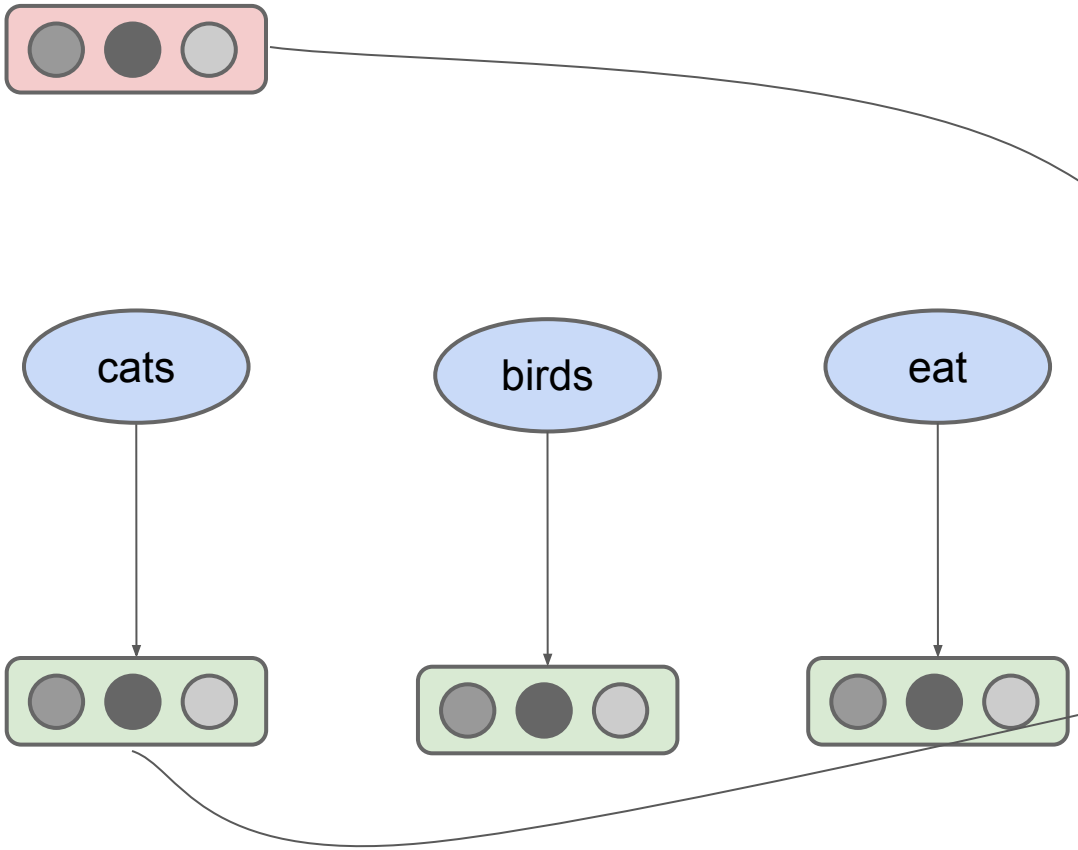
Character-level
Language Generation

Can we do the same here?

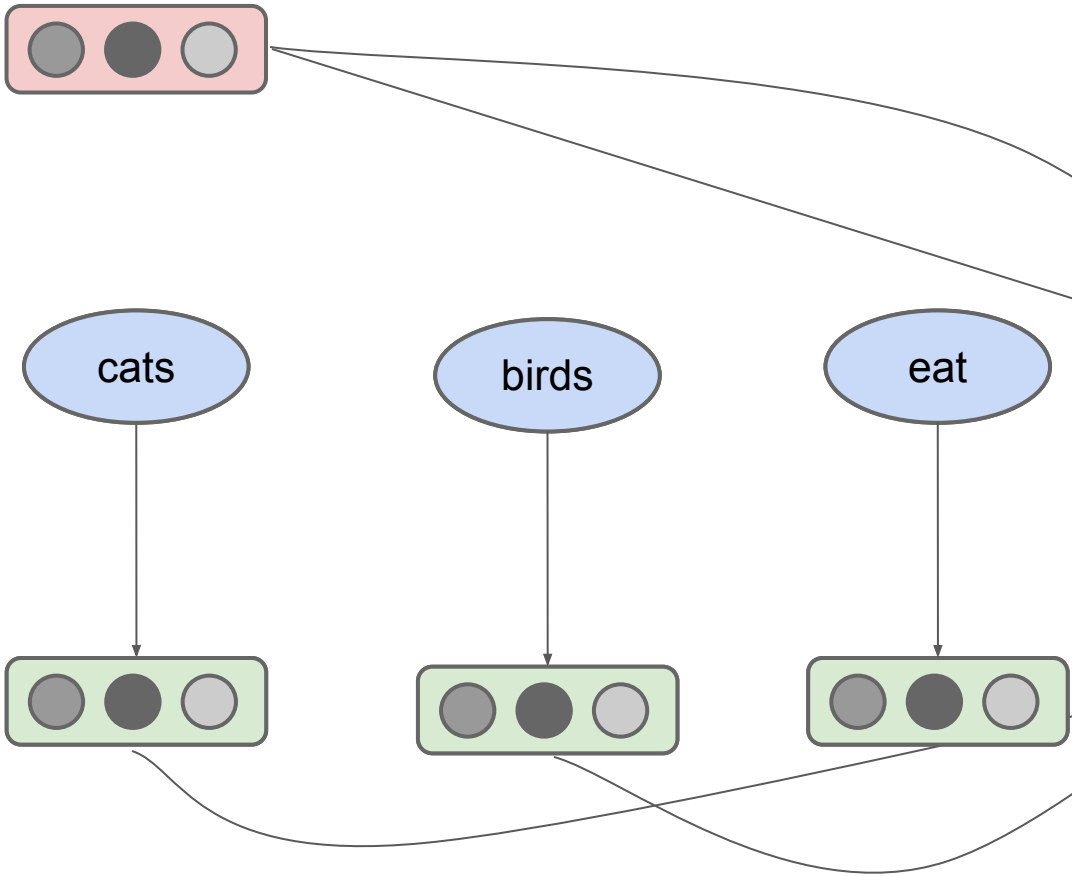


id	word	θ_1	θ_2	…	θ_p
1	birds	0.2	0.1	...	0.3
2	eat	0.3	-0.2	...	-0.1
3	cat	-0.1	0.3	...	0.3
...	...	...	...	...	….
V	dog	-0.2	-0.3	...	0.4

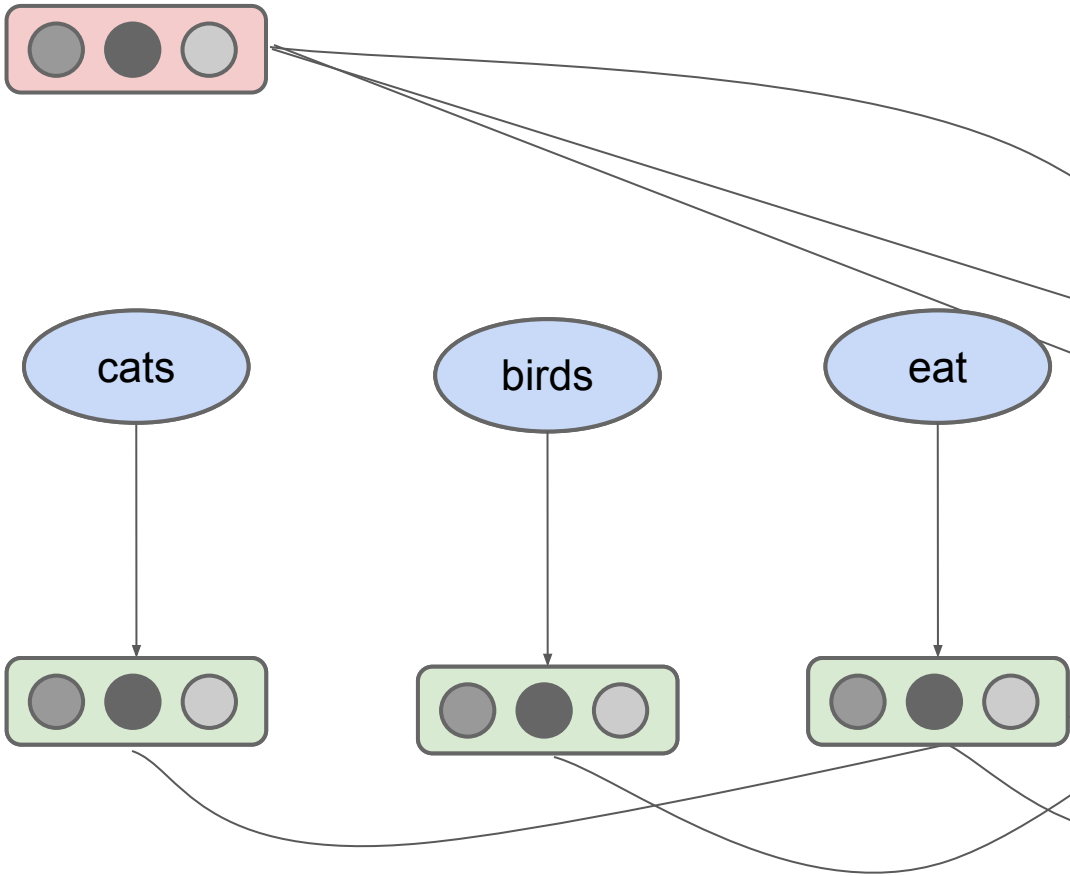
id	word	logit	prob
1	birds	2	0.1
2	eat	1	0.3
3	cat	9	0.1
...	...	1	...
V	dog	3	0.2



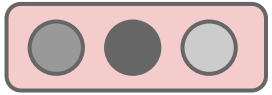
id	word	logit	prob
1	birds	2	0.1
2	eat	1	0.3
3	cat	9	0.1
...	...	1	...
V	dog	3	0.2

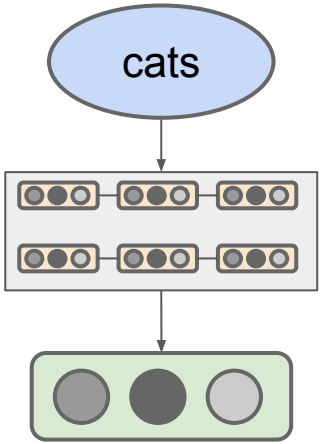


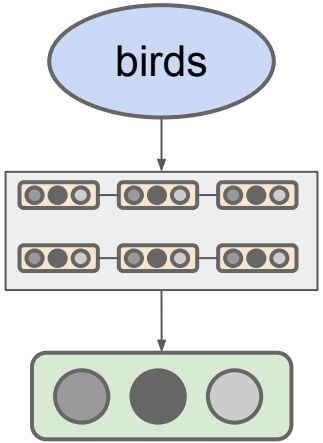
id	word	logit	prob
1	birds	2	0.1
2	eat	1	0.3
3	cat	9	0.1
...	...	1	...
V	dog	3	0.2

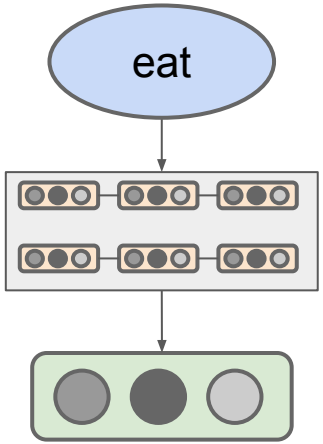


id	word	logit	prob
1	birds	2	0.1
2	eat	1	0.3
3	cat	9	0.1
...	...	1	...
V	dog	3	0.2



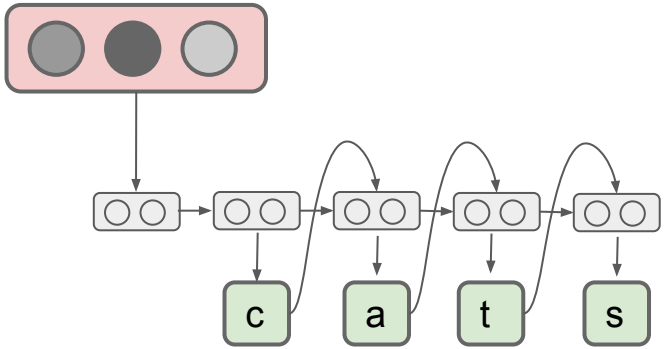


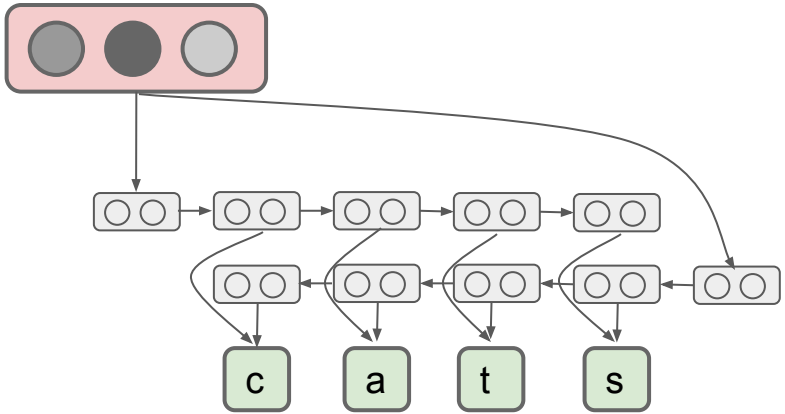


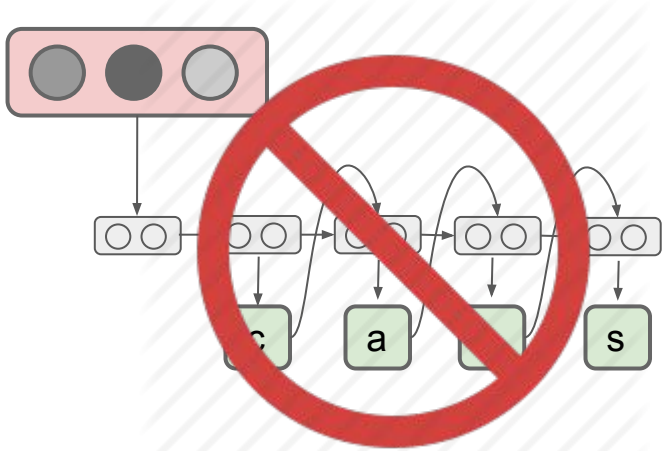


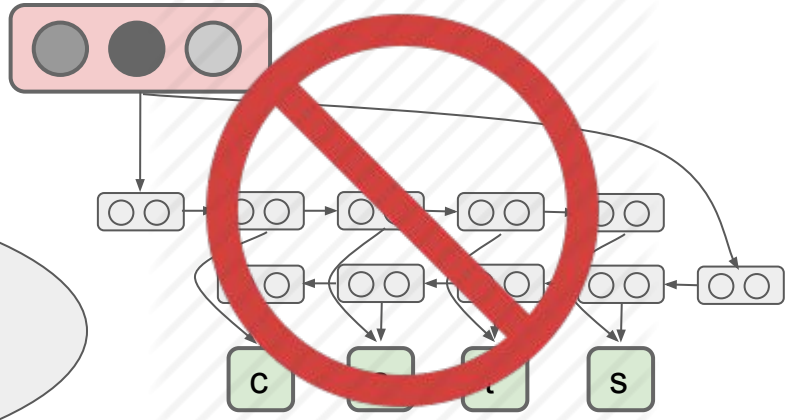
id	word	logit	prob
1	birds	2	0.1
2	eat	1	0.3
3	cat	9	0.1
...	...	1	...
V	dog	3	0.2

	LM (PTB)
Word Softmax	**115.17**
Char Softmax	116.91


c
a
t
s


c
a
t
s





Might as well just use character LSTM

Character-based Language Generation

Closed Vocabulary Perplexity

cats | eat | birds

cat | s | eat | bird | s

c | a | t | s | e | a | t | b | i | r | d | s

Ling et al. 2016

Character-based Language Generation

Closed Vocabulary Perplexity

cats eat birds

cat s eat bird s

c a t s e a t b i r d s

Ling et al. 2016

Character-based Language Generation

Closed Vocabulary Perplexity

cats eat birds

cat s eat bird s

c a t s e a t b i r d s

Ling et al. 2016

Character-based Language Generation

Closed Vocabulary Perplexity

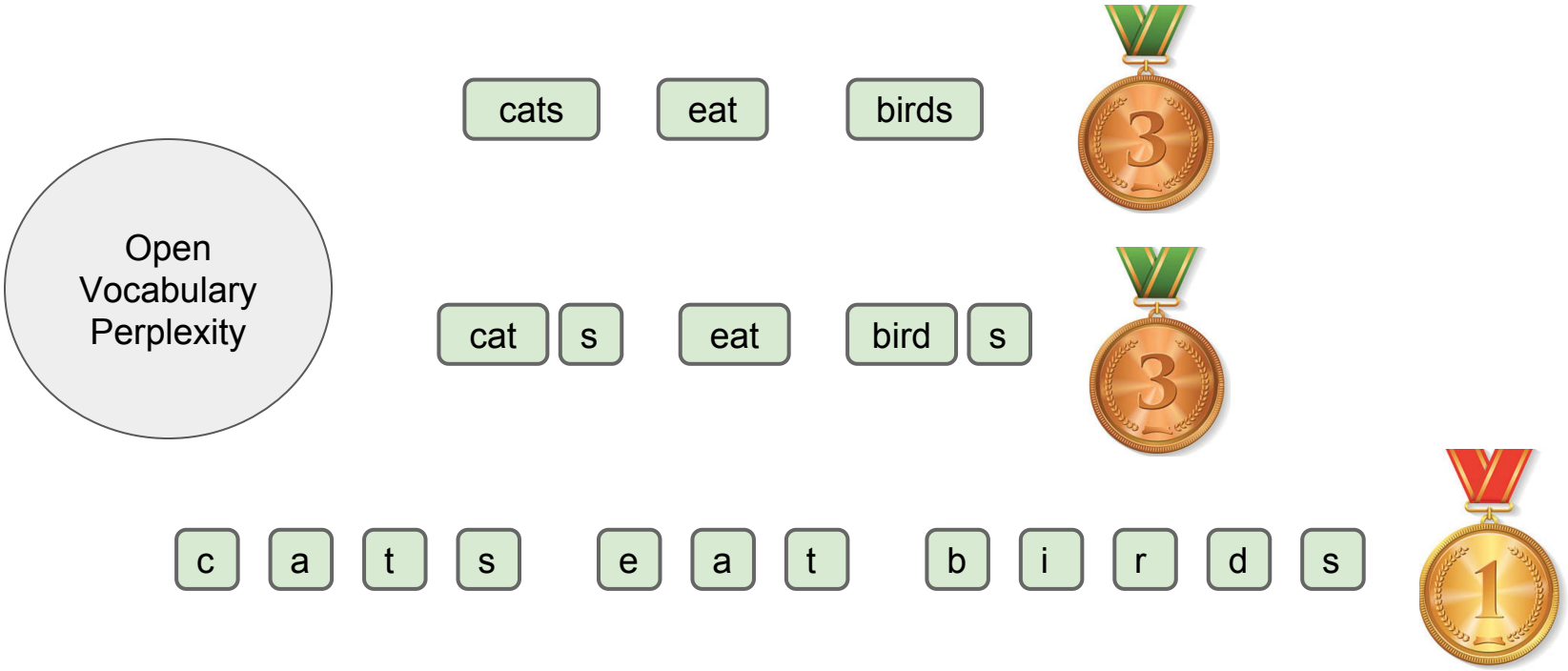
cats eat birds

cat s eat bird s

c a t s e a t b i r d s

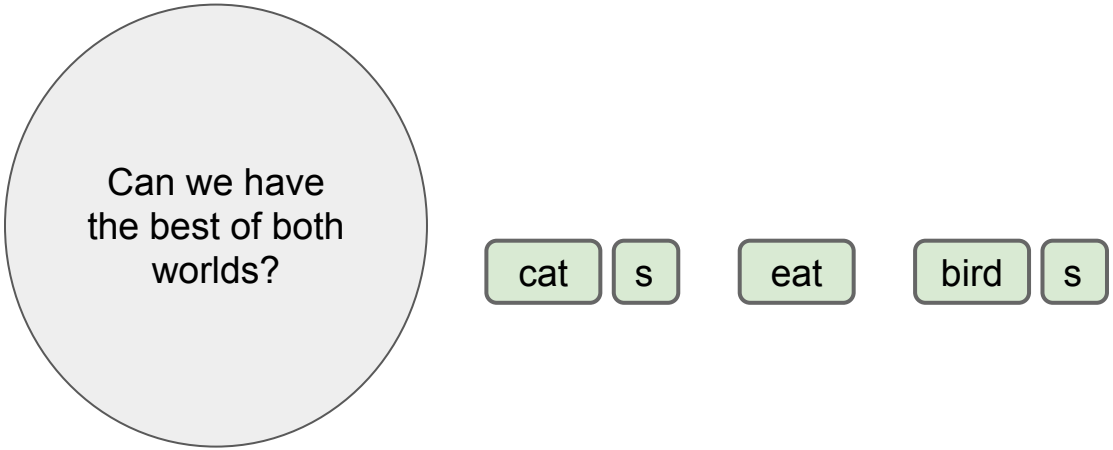
Ling et al. 2016

Character-based Language Generation



Ling et al. 2016

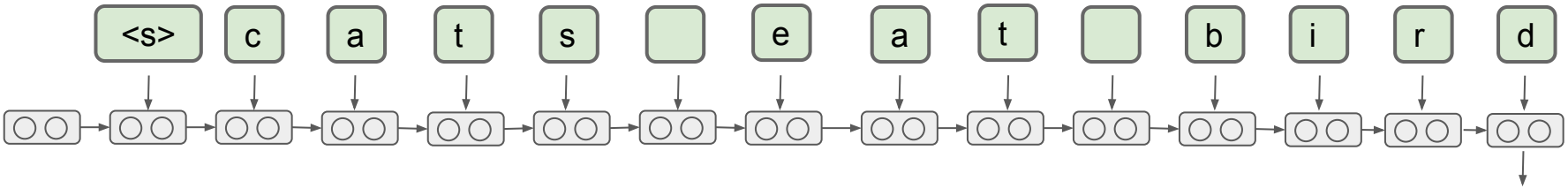
Character-based Language Generation



Ling et al. 2016

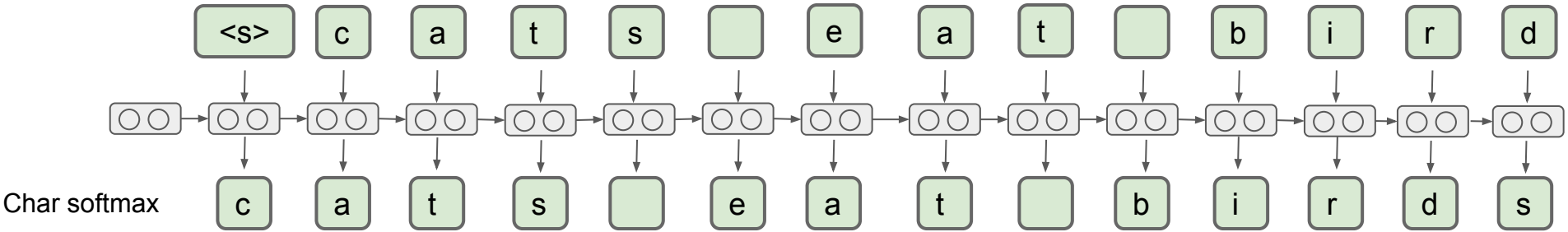
How many different ways to generate this sequence of characters?

Ling et al. 2016



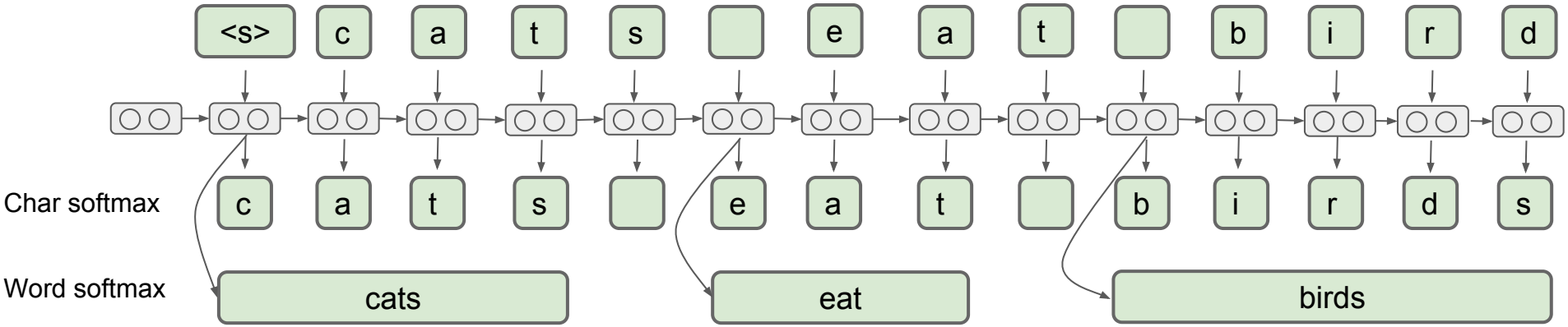
Character by character?

Ling et al. 2016

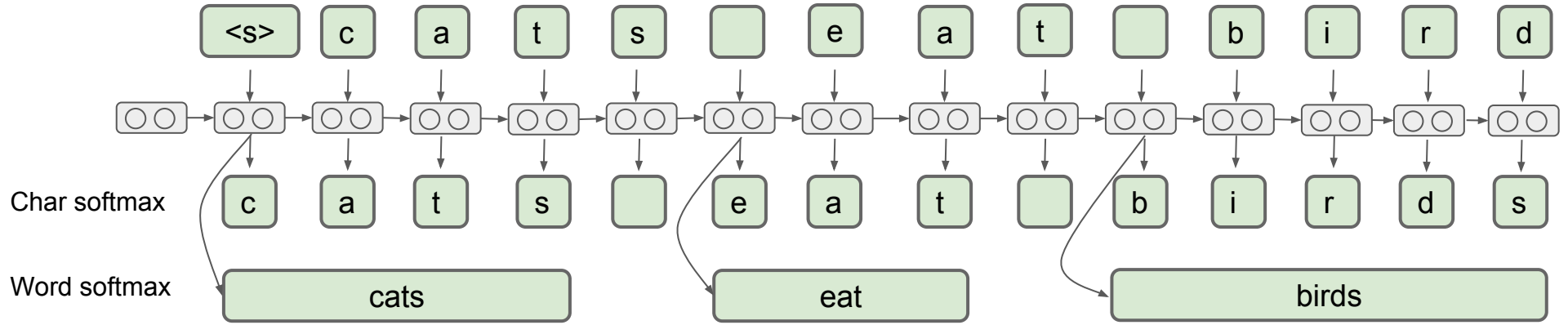


Word by word?

Ling et al. 2016

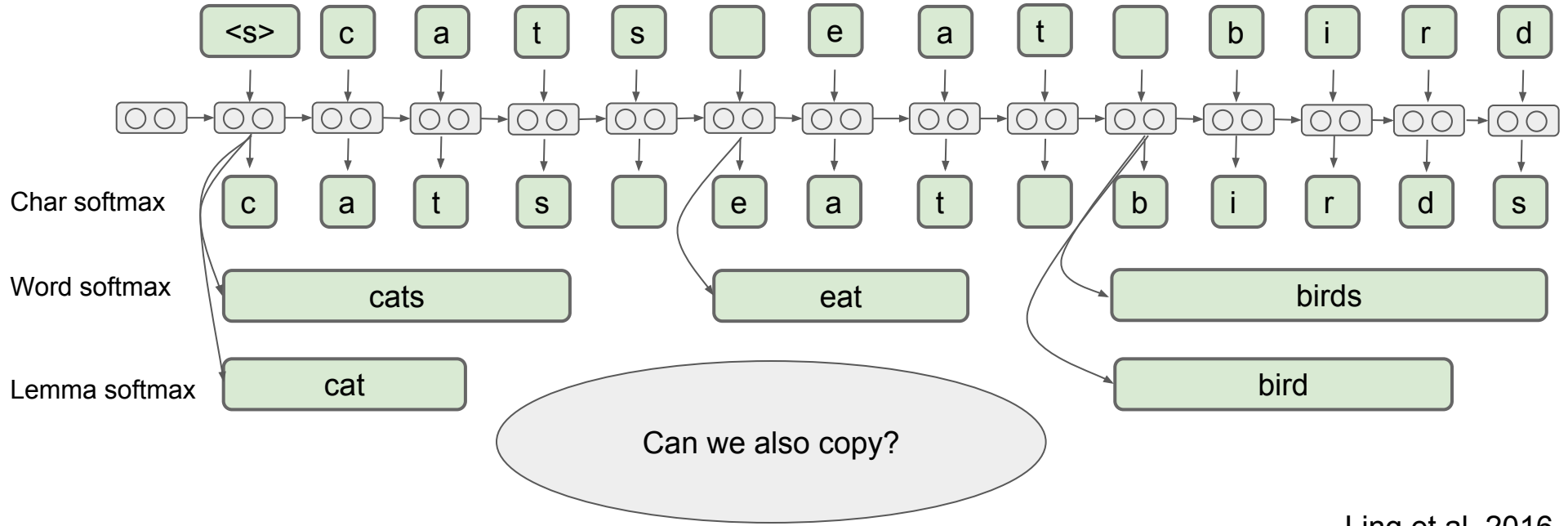


Ling et al. 2016

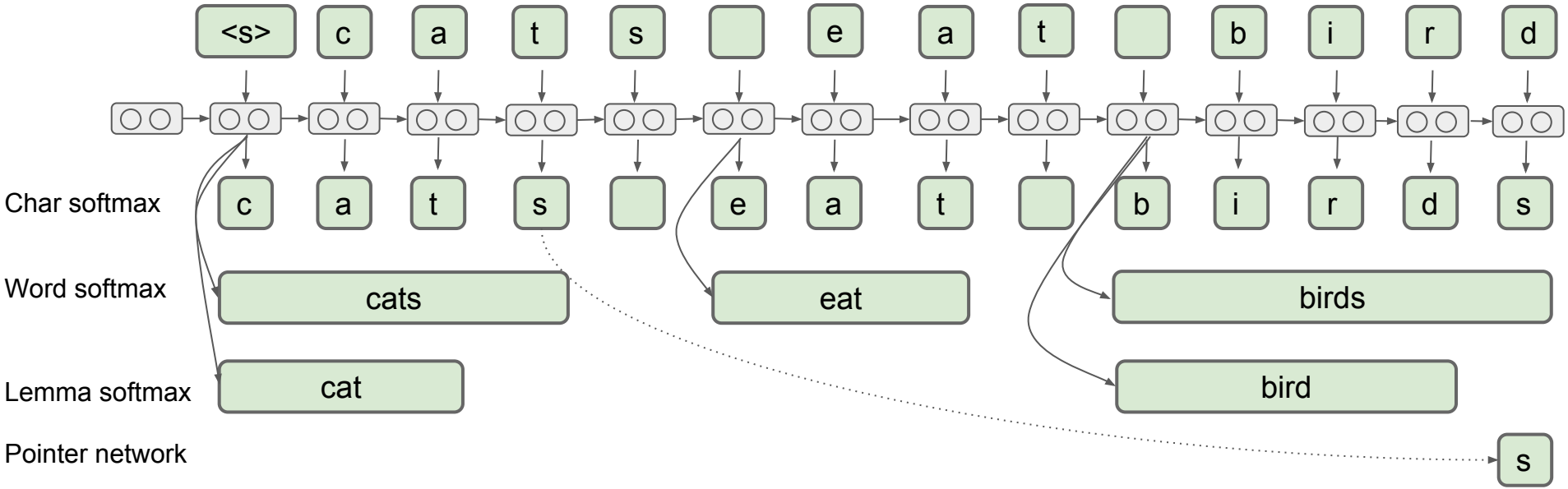


What about lemmas?

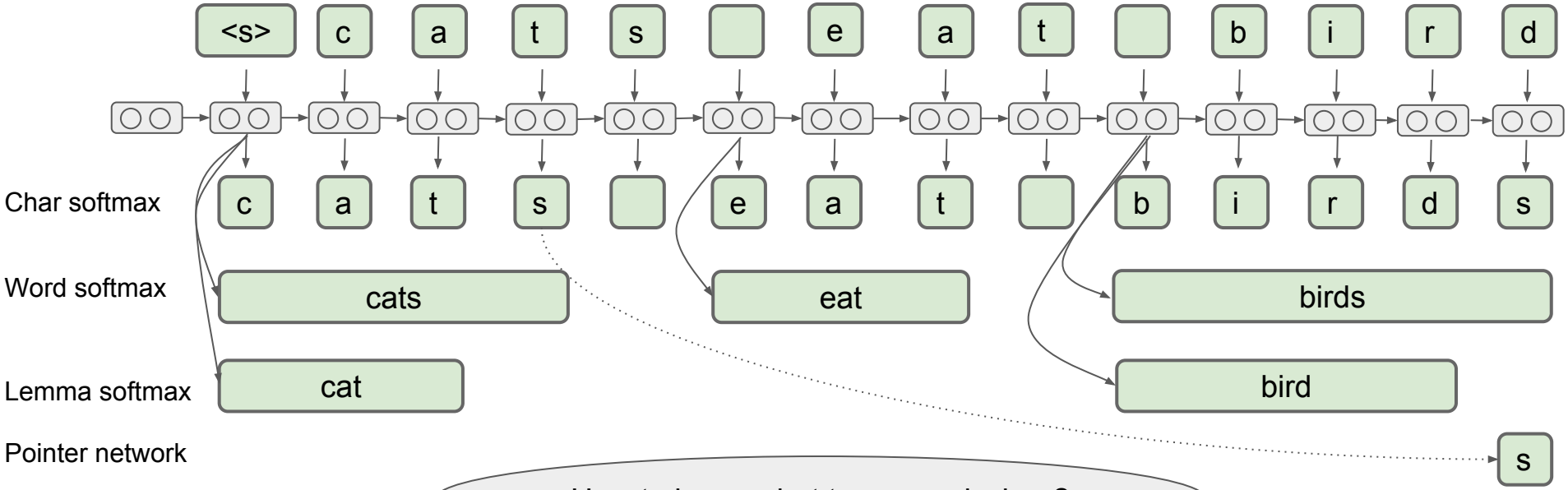
Ling et al. 2016



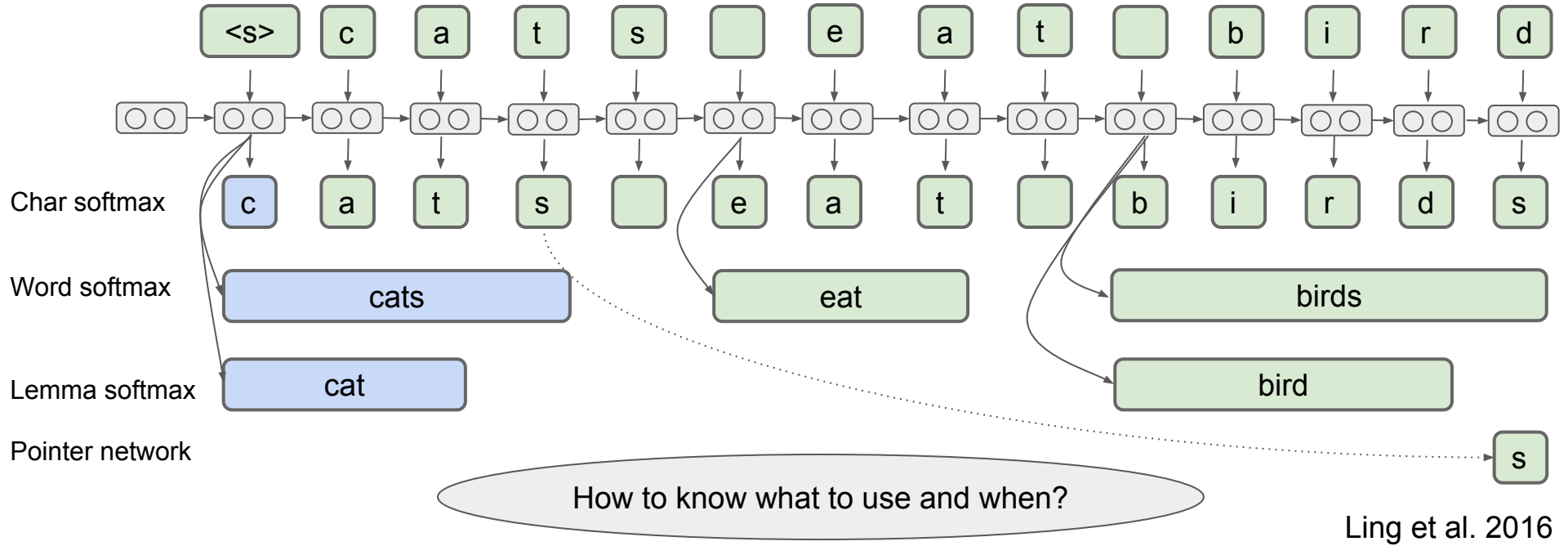
Ling et al. 2016



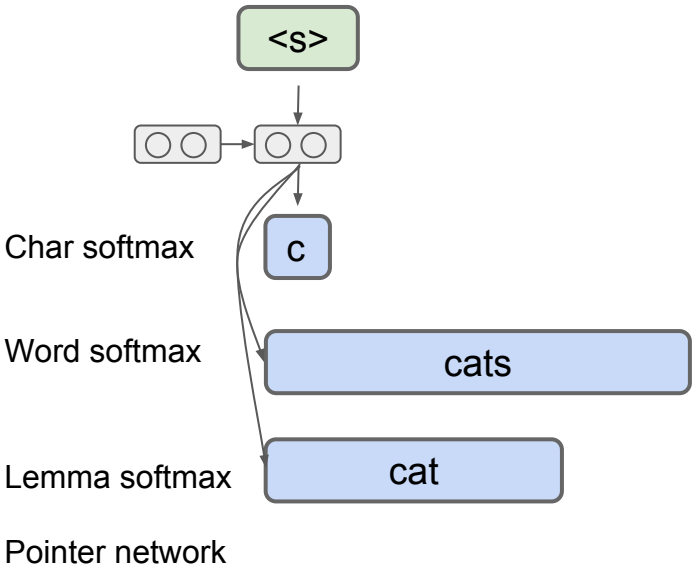
Ling et al. 2016



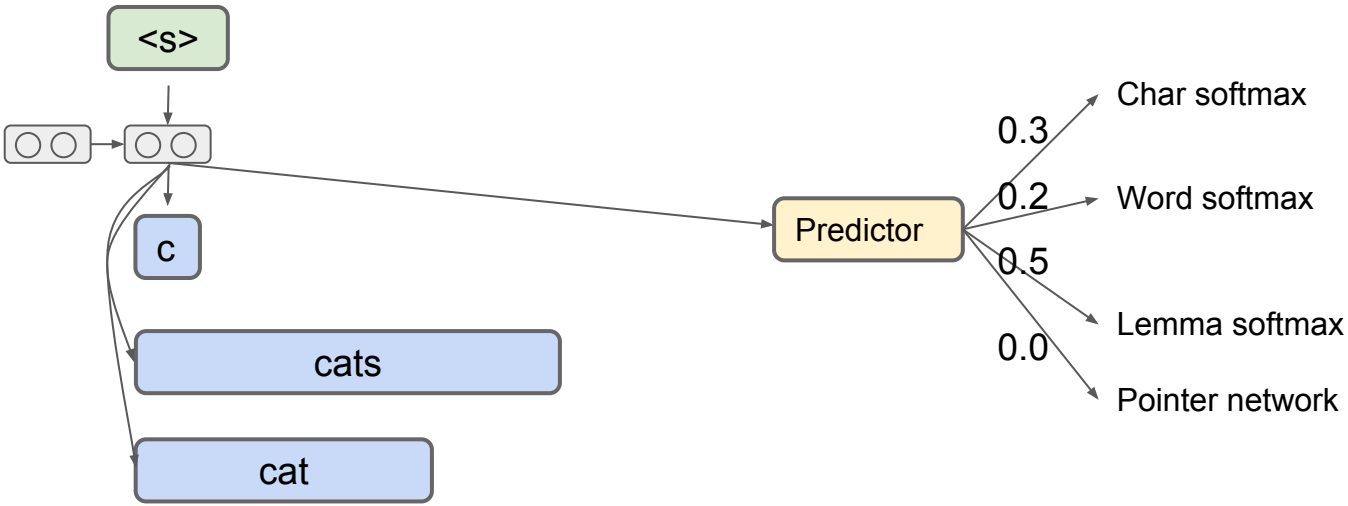
Ling et al. 2016



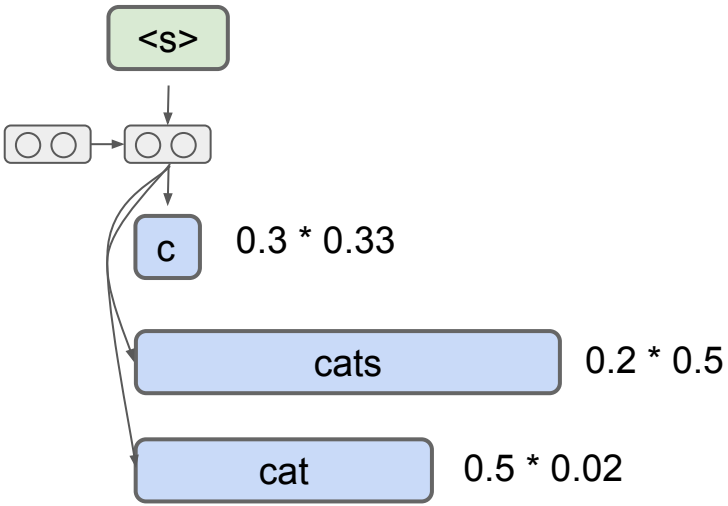
Ling et al. 2016



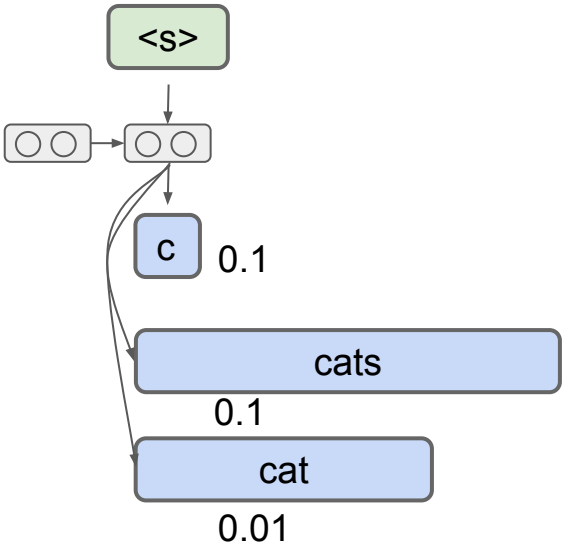
Ling et al. 2016



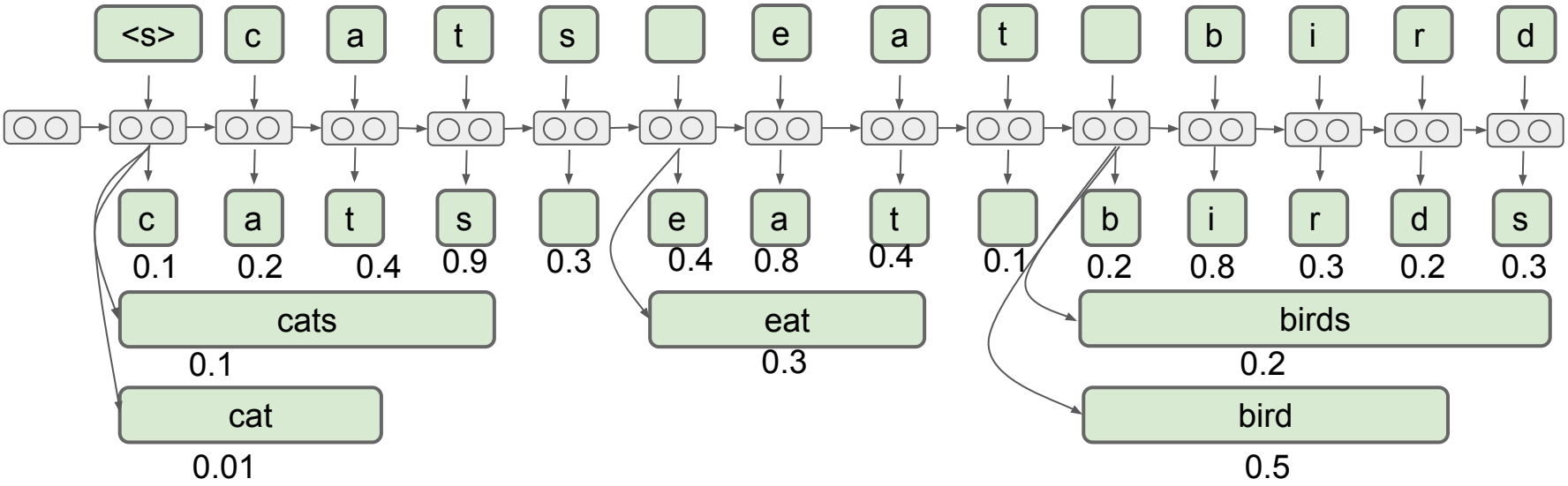
Ling et al. 2016



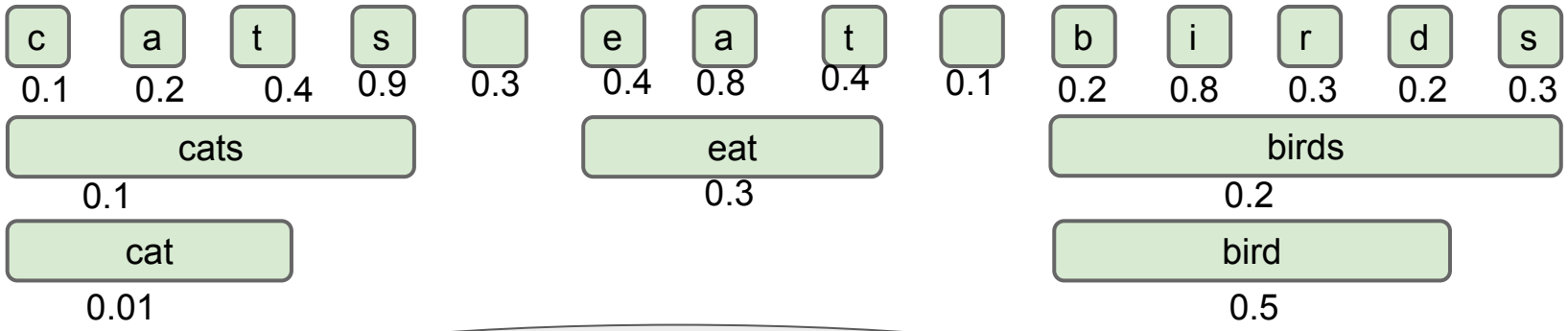
Ling et al. 2016



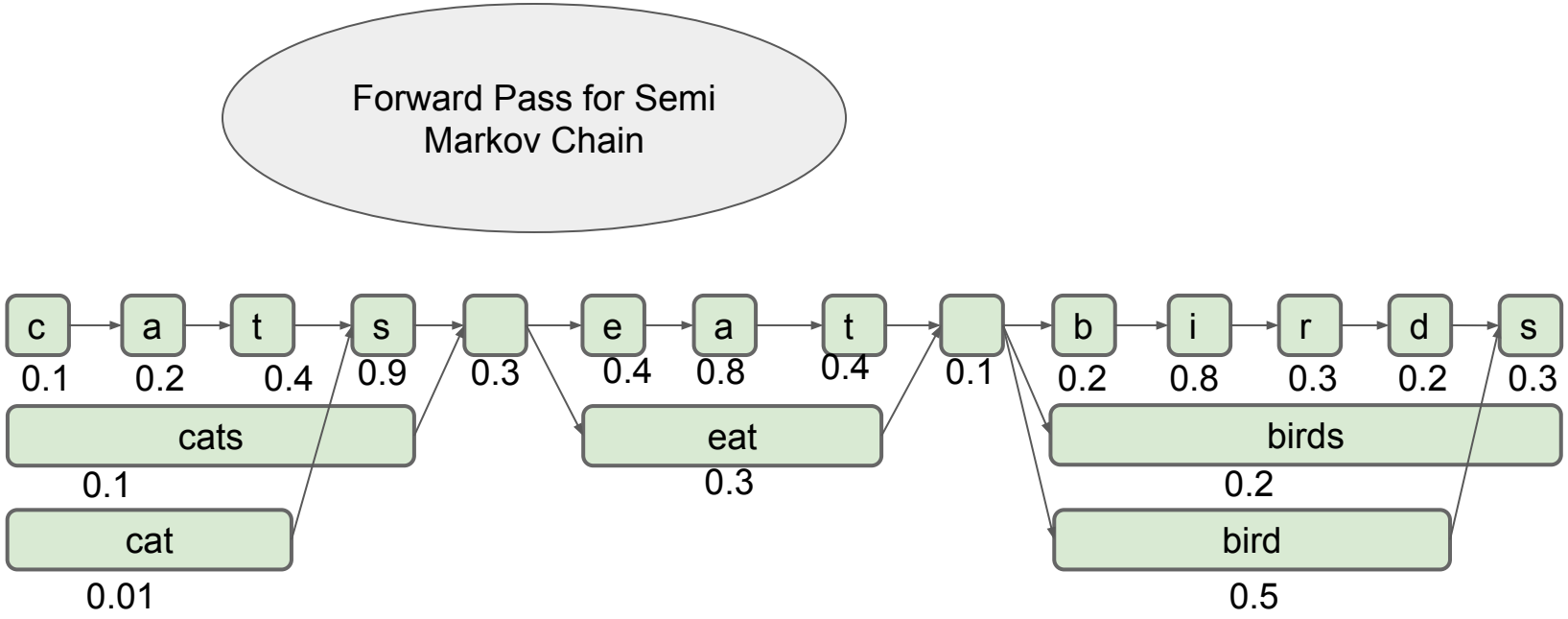
Ling et al. 2016



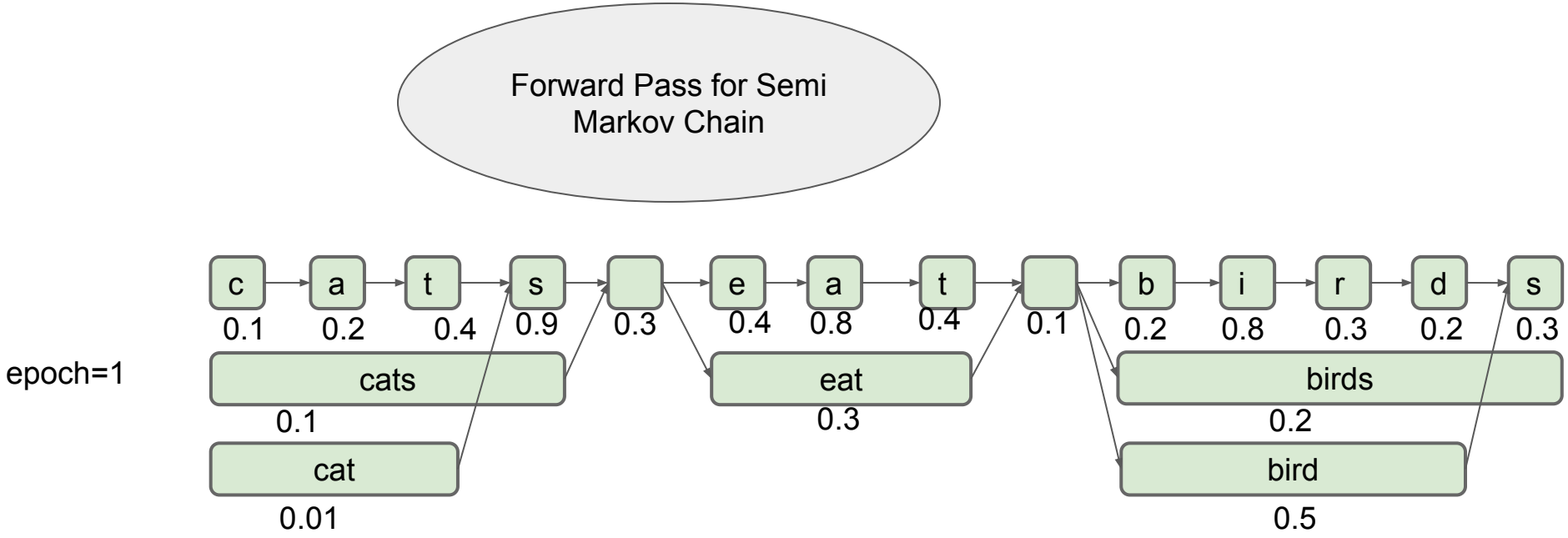
Ling et al. 2016



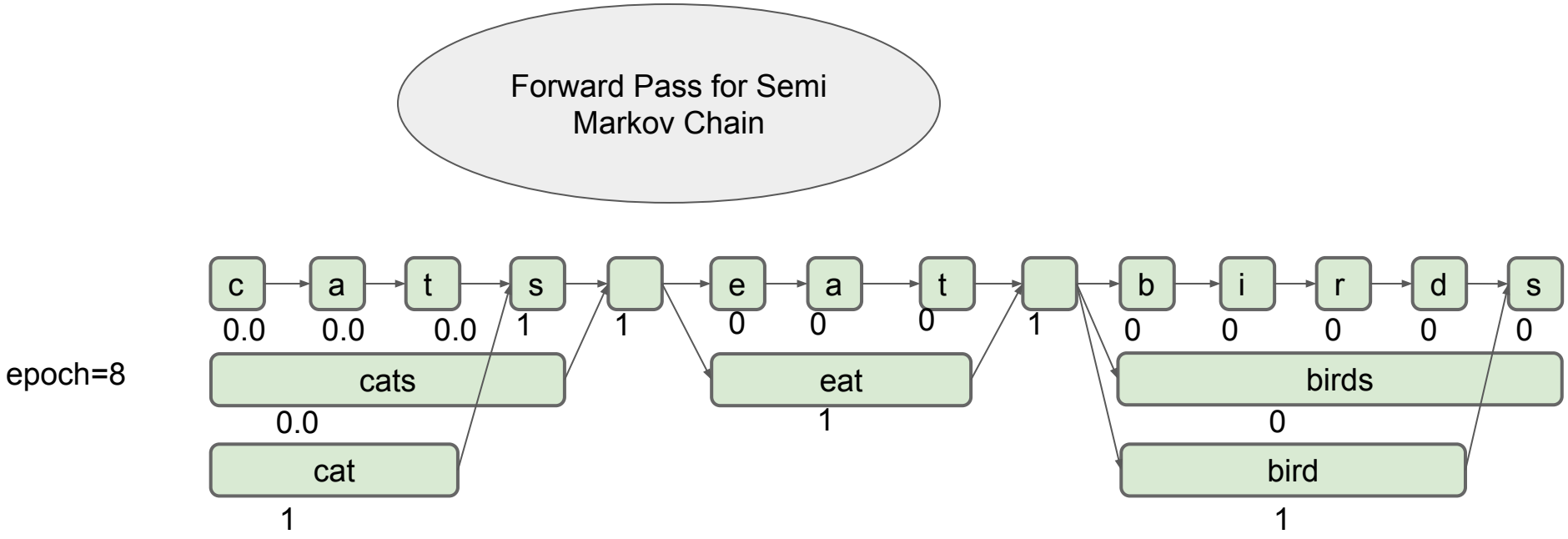
Ling et al. 2016



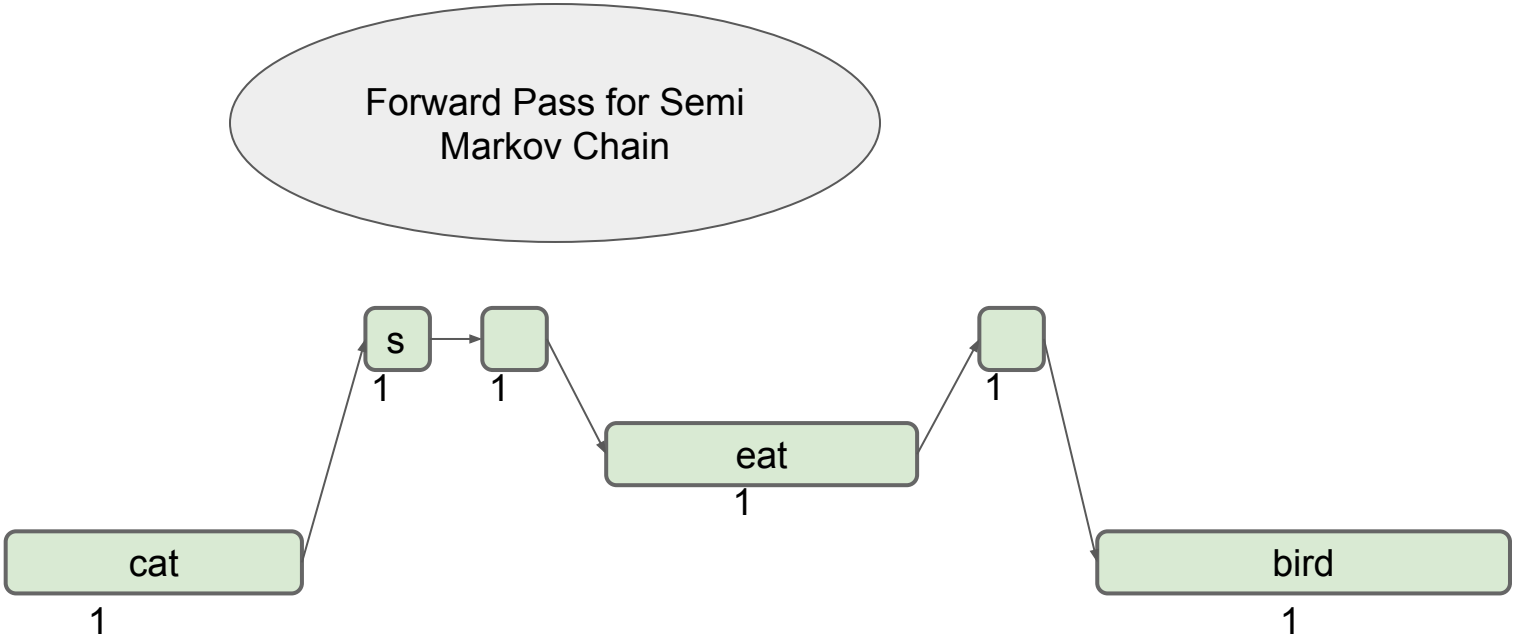
Ling et al. 2016



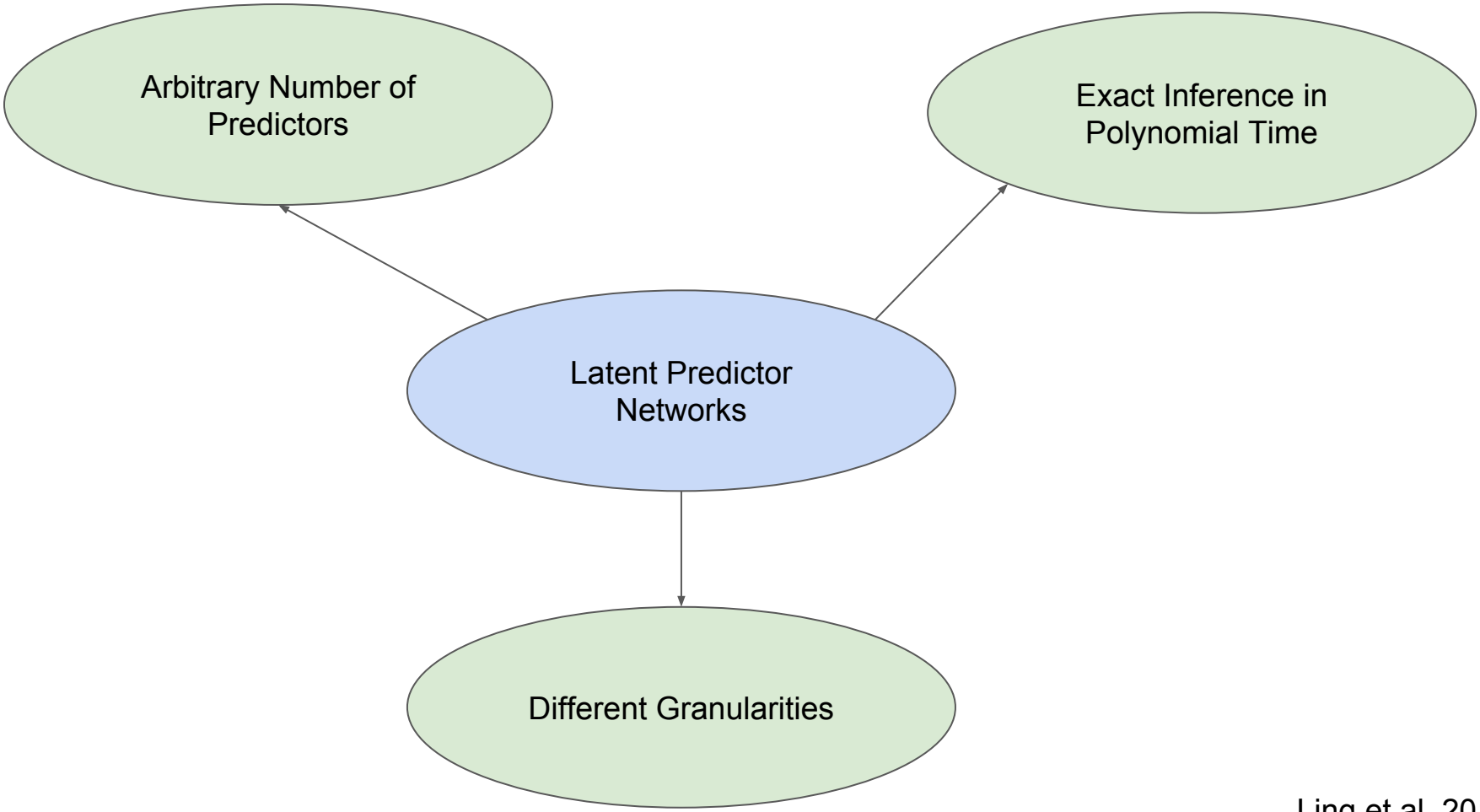
Ling et al. 2016



Ling et al. 2016



Ling et al. 2016



Ling et al. 2016

Can we automatically implement trading card game effects?

Ling et al. 2016


```
class ArcaneExplosion(SpellCard):
    def __init__(self):
        super().__init__("Arcane Explosion", 2,
CHARACTER_CLASS.MAGE, CARD_RARITY.FREE)

    def use(self, player, game):
        super().use(player, game)
        for minion in copy.copy(game.other_player.minions):
            minion.damage(player.effective_spell_damage(1), self)
```

Ling et al. 2016

Card Description

Tirion Fordring	Paladin
8	Divine Shield. Taunt. Deathrattle: Equip a 5/3 Ashbringer.
Minion	6
Legendary	6
NIL	-1

Code

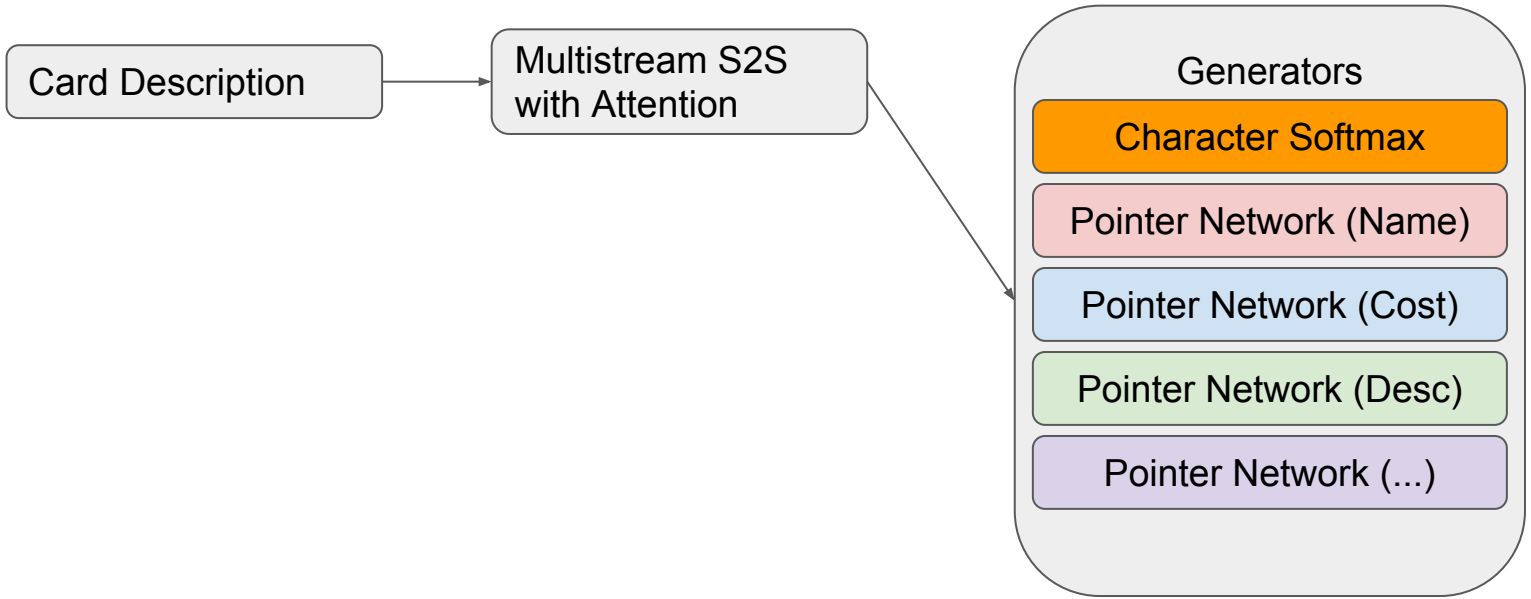
```
class TirionFordring(MinionCard):
    def __init__(self):
        super().__init__("Tirion Fordring", 8, CHARACTER_CLASS.PALADIN, CARD_RARITY.LEGENDARY)

    def create_minion(self, player):
        return Minion(6, 6, divine_shield=True, taunt=True, deathrattle=Deathrattle(Equip(Ashbringer()),
PlayerSelector()))
```

Ling et al. 2016

- Baselines
  - Retrieval - Find the most similar input and retrieve output (Quirk et al, 2015)
  - Machine Translation Baselines (Collapsed input)
    - Phrase-based (Koehn et al, 2007)
    - Hierarchical (Chiang et al, 2005)
  - Neural Attention Model (Bahdanau et al, 2014)

Ling et al. 2016



Ling et al. 2016

- Evaluation
  - BLEU - N-gram matching between reference and hypothesis + brevity penalty
  - Hypothesis and reference are tokenized prior to evaluation

for minion in copy.copy(game.other_player.minions):

for minion in copy . copy (game . other_player . minions) :

Ling et al. 2016

	Magic	Hearthstone
Retrieval	**54.9**	**62.5**
Phrase-based	49.5	34.1
Hierarchical	50.6	43.2
Baseline NN	50.1	43.9

Ling et al. 2016

	Magic	Hearthstone
Retrieval	54.9	62.5
Phrase-based	49.5	34.1
Hierarchical	50.6	43.2
Baseline NN	50.1	43.9
Our Model	**61.4**	**65.6**

Ling et al. 2016

	Magic	Hearthstone	Django (Oda et al, 2015)
Retrieval	54.9	62.5	18.4
Phrase-based	49.5	34.1	47.6
Hierarchical	50.6	43.2	35.9
Baseline NN	50.1	43.9	58.9
Our Model	**61.4**	**65.6**	**77.6**

Ling et al. 2016

- Generated Card Example (Test Set)


```
class MadderBomber(MinionCard):        BLEU = 100.0
  def __init__(self):
    super().__init__("Madder Bomber", 5,
    CHARACTER_CLASS.ALL, CARD_RARITY.RARE,
    battlecry=Battlecry(Damage(1),
    CharacterSelector(players=BothPlayer(),
       picker= RandomPicker(6))))

def create_minion(self, player):§
  return Minion(5, 4)§
```

Ling et al. 2016

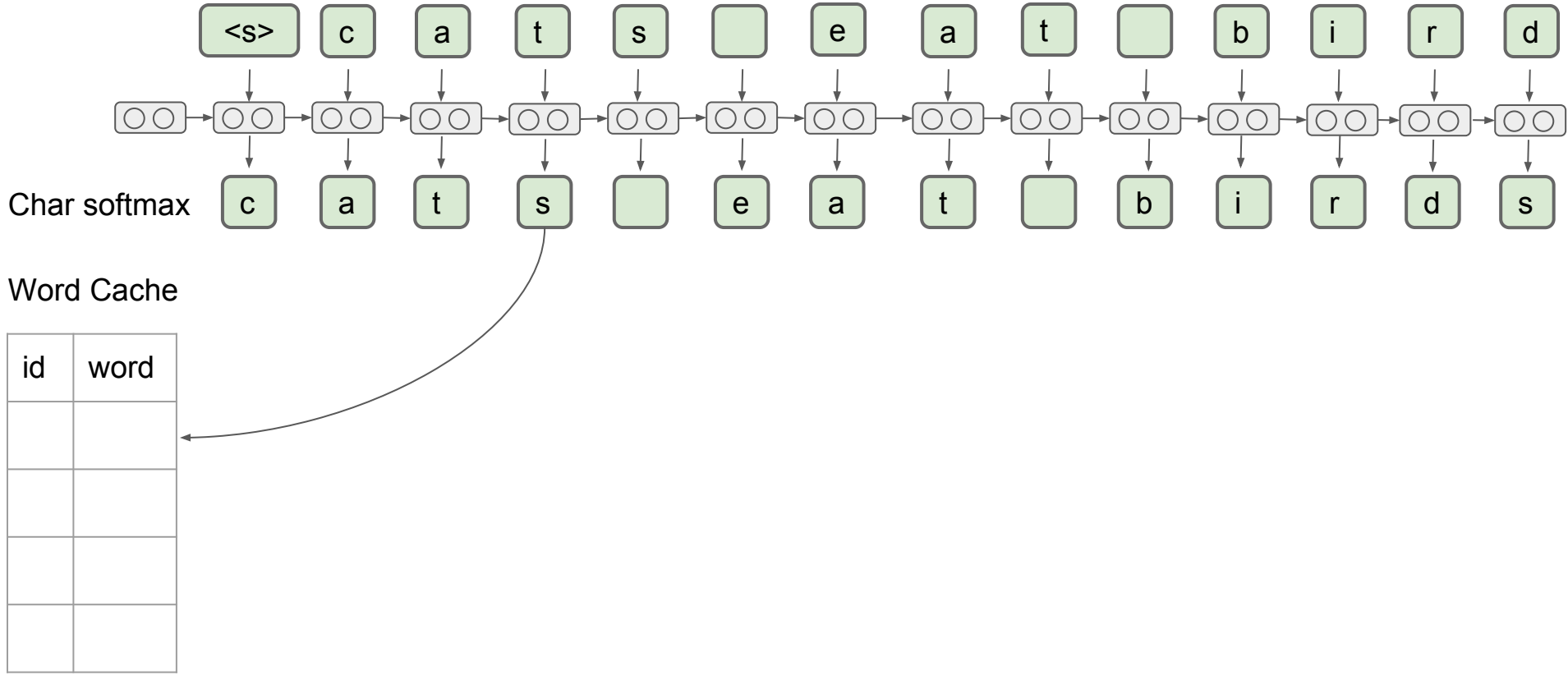
- Generated Card Example (Test Set)


```
class MadderBomber(MinionCard):        BLEU = 100.0
  def __init__(self):
    super().__init__("Madder Bomber", 5,
    CHARACTER_CLASS.ALL, CARD_RARITY.RARE,
    battlecry=Battlecry(Damage(1),
    CharacterSelector(players=BothPlayer(),
       picker= RandomPicker(6))))

def create_minion(self, player):§
  return Minion(5, 4)§
```

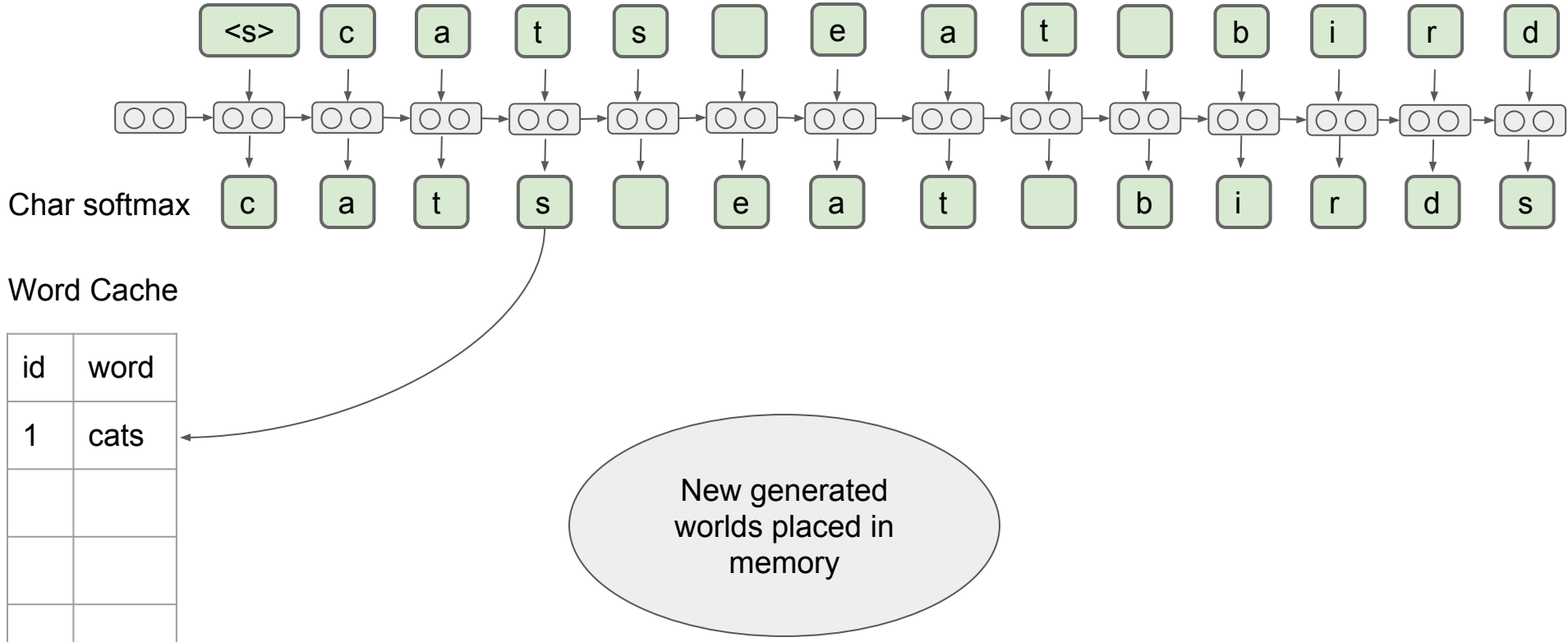
Ling et al. 2016

Related Work



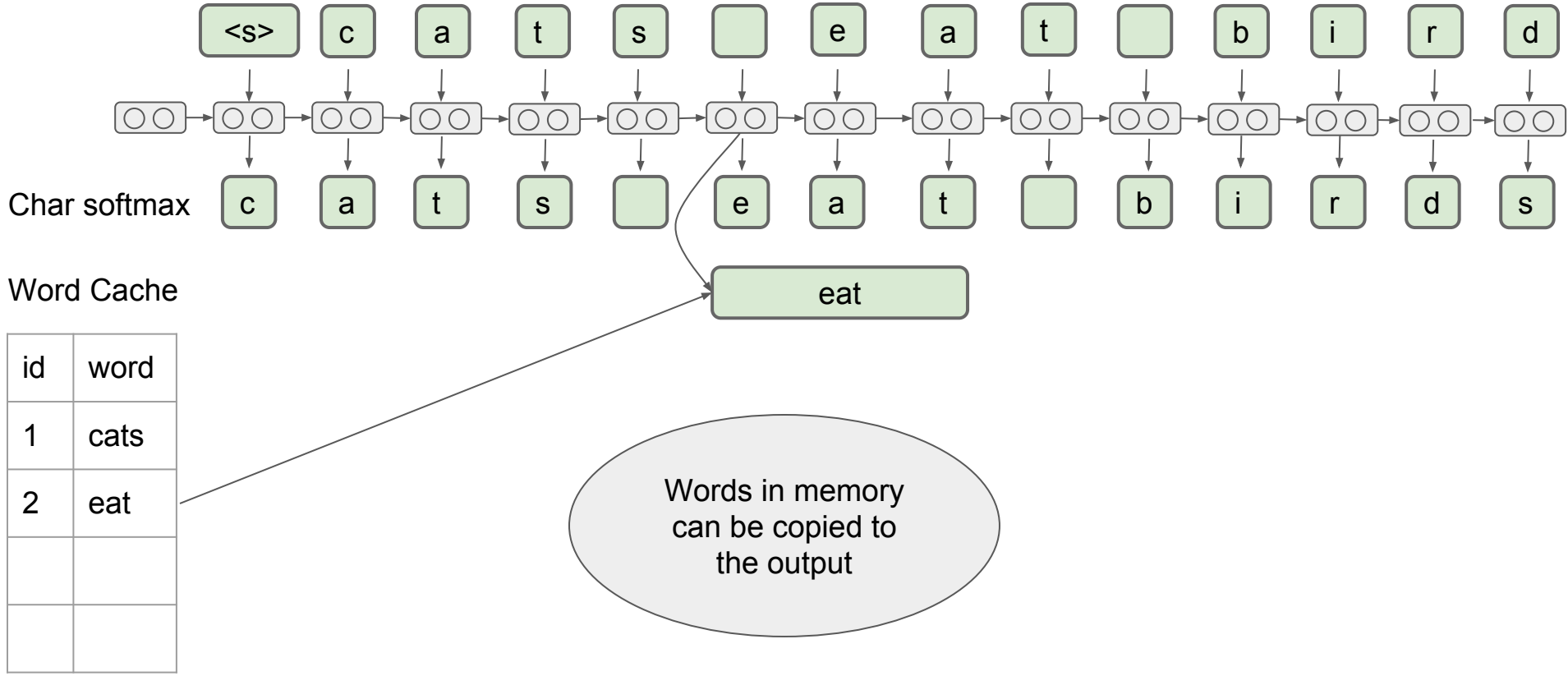
Kawakami et al. 2017

Related Work



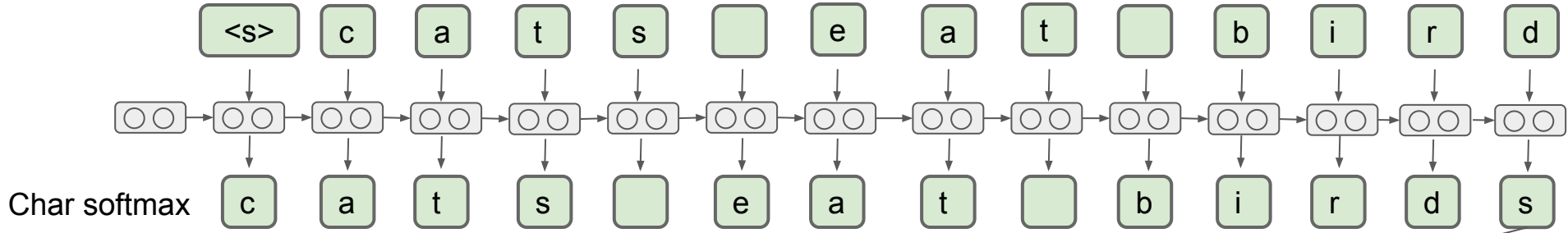
Kawakami et al. 2017

Related Work



Kawakami et al. 2017

Related Work



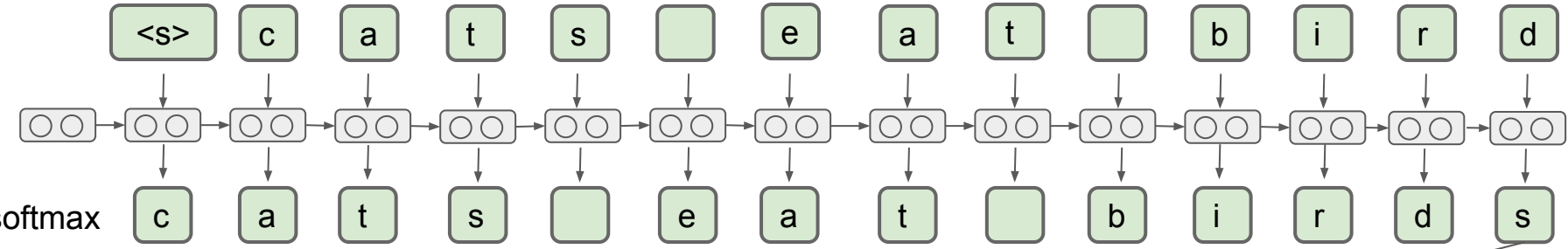
Word Cache

id	word
1	cats
2	eat
3	dog
4	drink

If cache is full,
remove oldest word

Kawakami et al. 2017

Related Work



Word Cache

id	word
1	cats
2	eat
3	birds
4	drink

If cache is full,
remove oldest word

Kawakami et al. 2017

Related Work

Word	$\overline{p(z \mid \boldsymbol{w})} \downarrow$	Word	$\overline{p(z \mid w)} \uparrow$
.	0.997	300	0.000
Lesnar	0.991	act	0.001
the	0.988	however	0.002
NY	0.985	770	0.003
Gore	0.977	put	0.003
Bintulu	0.976	sounds	0.004
Nerva	0.976	instead	0.005
,	0.974	440	0.005
UB	0.972	similar	0.006
Nero	0.967	27	0.009
Osbert	0.967	help	0.009
Kershaw	0.962	few	0.010
Manila	0.962	110	0.010
Boulter	0.958	Jersey	0.011
Stevens	0.956	even	0.011
Rifenburg	0.952	y	0.012
Arjona	0.952	though	0.012
of	0.945	becoming	0.013
31B	0.941	An	0.013
Olympics	0.941	unable	0.014

Kawakami et al. 2017

Related Work

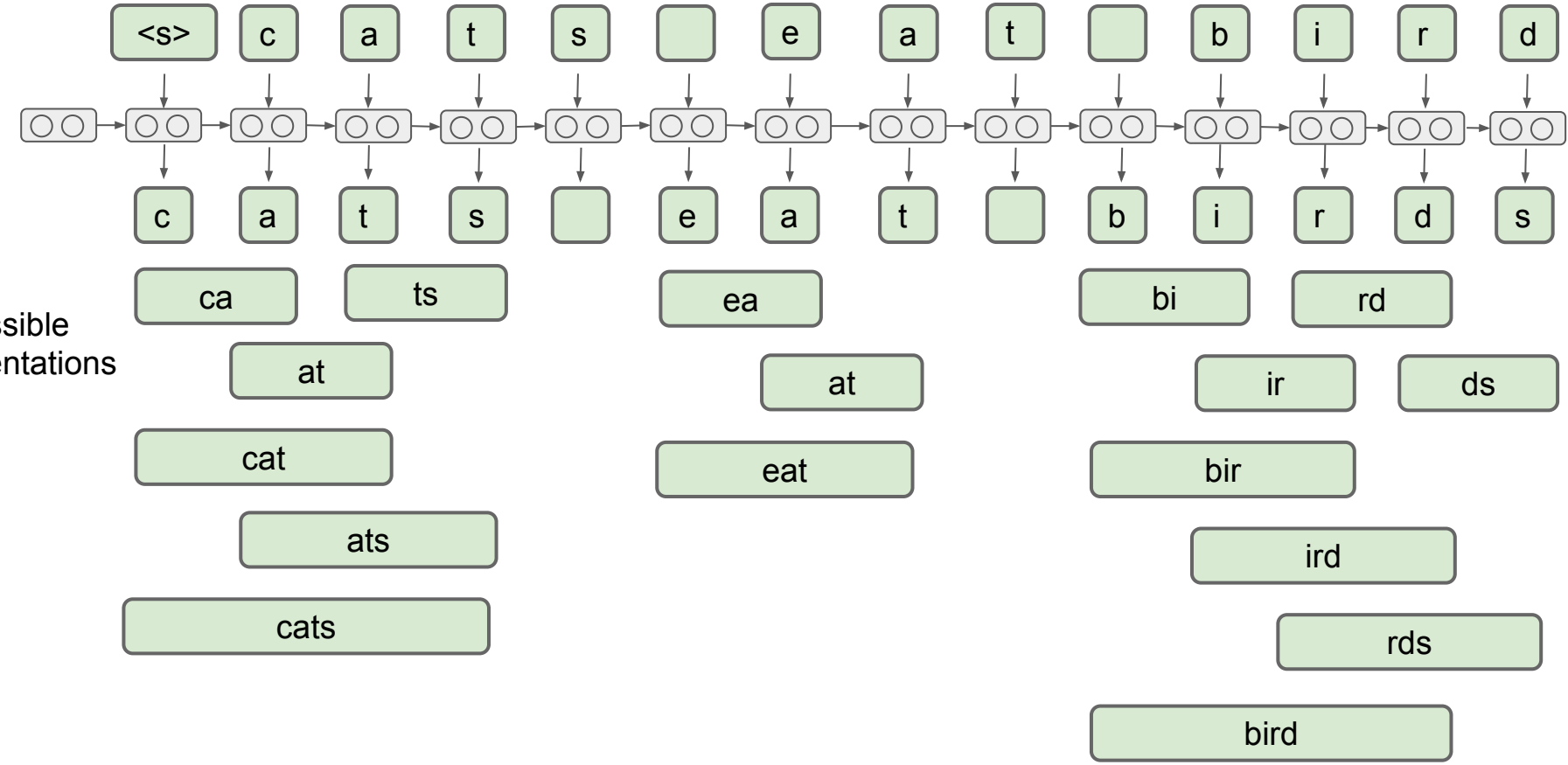
Proper nouns and frequent words

Word	$\overline{p(z \mid \boldsymbol{w})} \downarrow$	Word	$\overline{p(z \mid w)} \uparrow$
.	0.997	300	0.000
Lesnar	0.991	act	0.001
the	0.988	however	0.002
NY	0.985	770	0.003
Gore	0.977	put	0.003
Bintulu	0.976	sounds	0.004
Nerva	0.976	instead	0.005
,	0.974	440	0.005
UB	0.972	similar	0.006
Nero	0.967	27	0.009
Osbert	0.967	help	0.009
Kershaw	0.962	few	0.010
Manila	0.962	110	0.010
Boulter	0.958	Jersey	0.011
Stevens	0.956	even	0.011
Rifenburg	0.952	y	0.012
Arjona	0.952	though	0.012
of	0.945	becoming	0.013
31B	0.941	An	0.013
Olympics	0.941	unable	0.014

Numbers and basic words

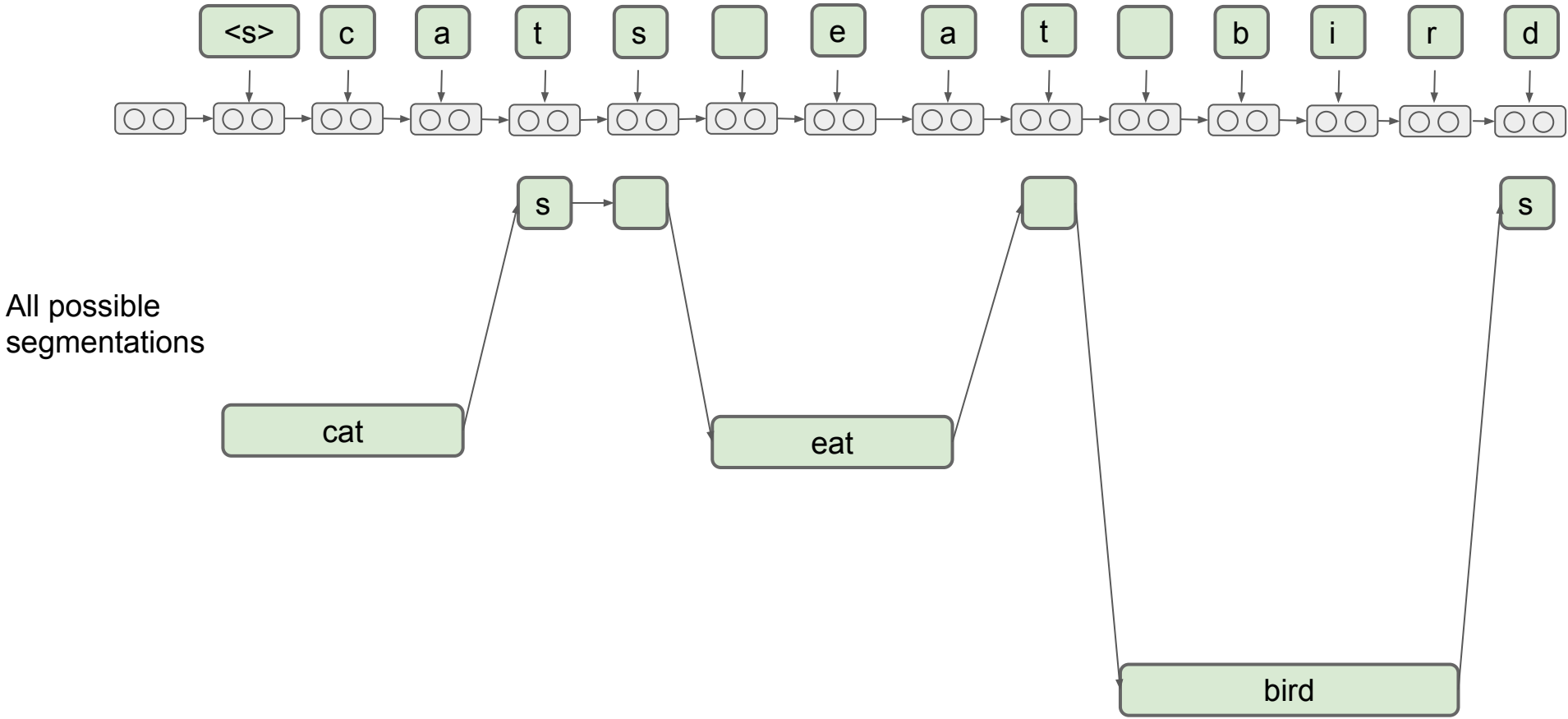
Kawakami et al. 2017

Related Work



Kawakami et al. 2018

Related Work



Kawakami et al. 2018

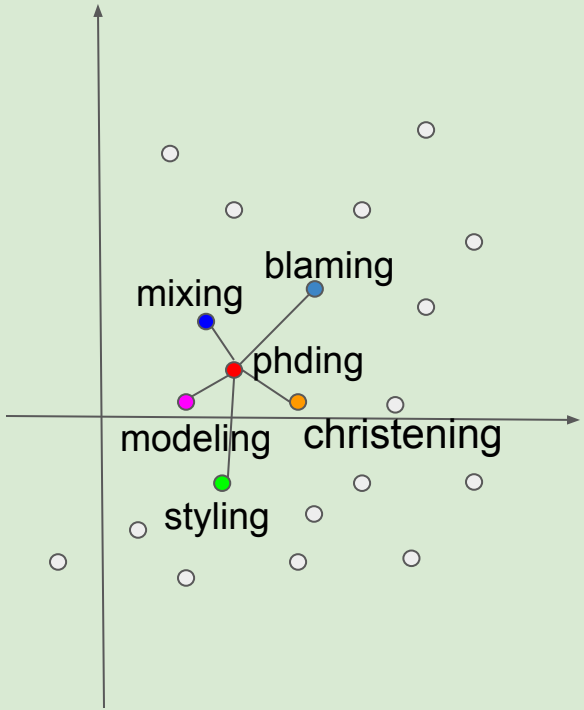
Related Work

wemergelongitudinaldataonchildandyoungadultoutcomeswithinformationonlocalhousepricesandmarketrents

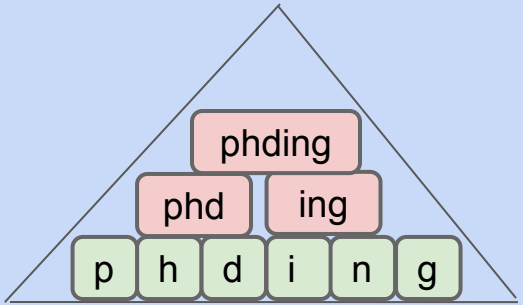
we merge longitudin al data on child and young adult outcome s with informat ion on local house price s and market rent s

Kawakami et al. 2018

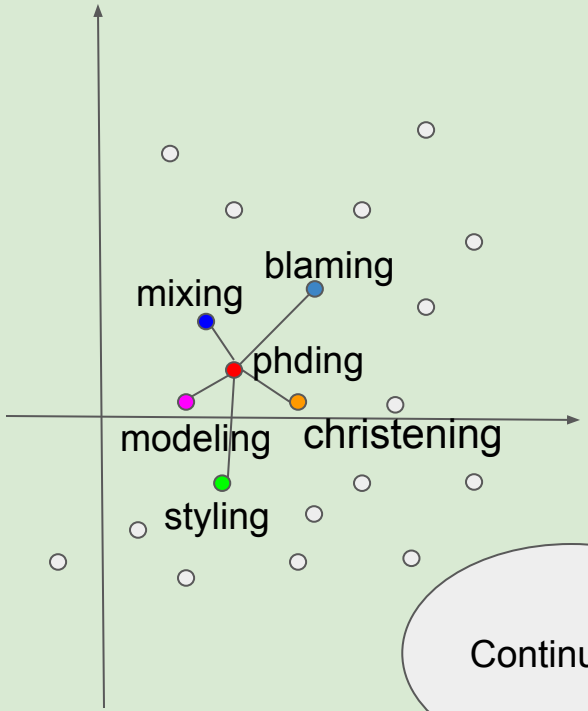
Character-level
Language Understanding



Character-level
Language Generation

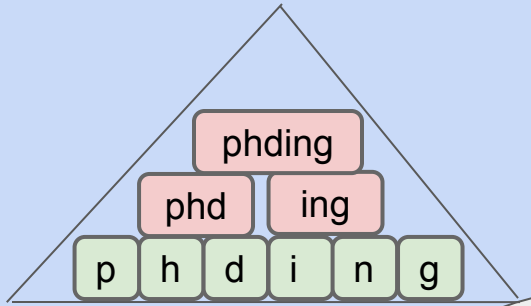


Character-level
Language Understanding



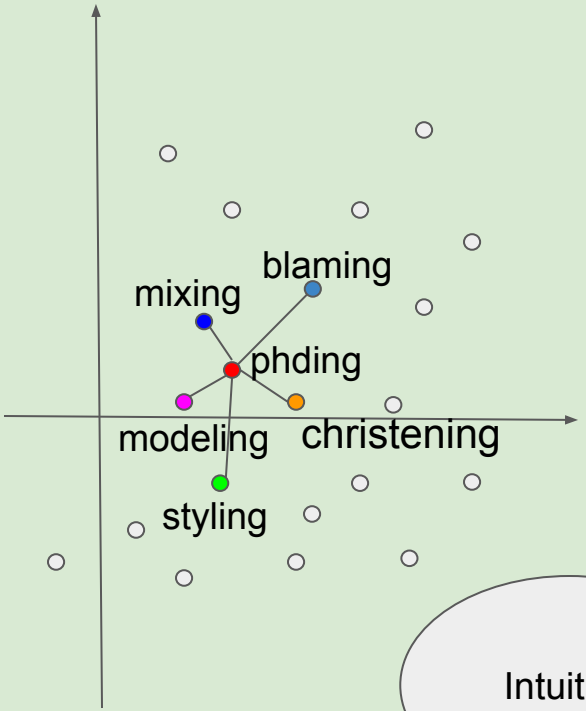
Continuous

Character-level
Language Generation



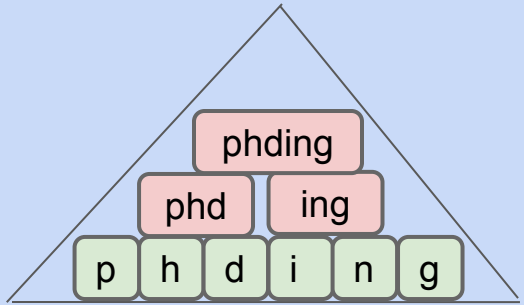
Discrete

Character-level
Language Understanding

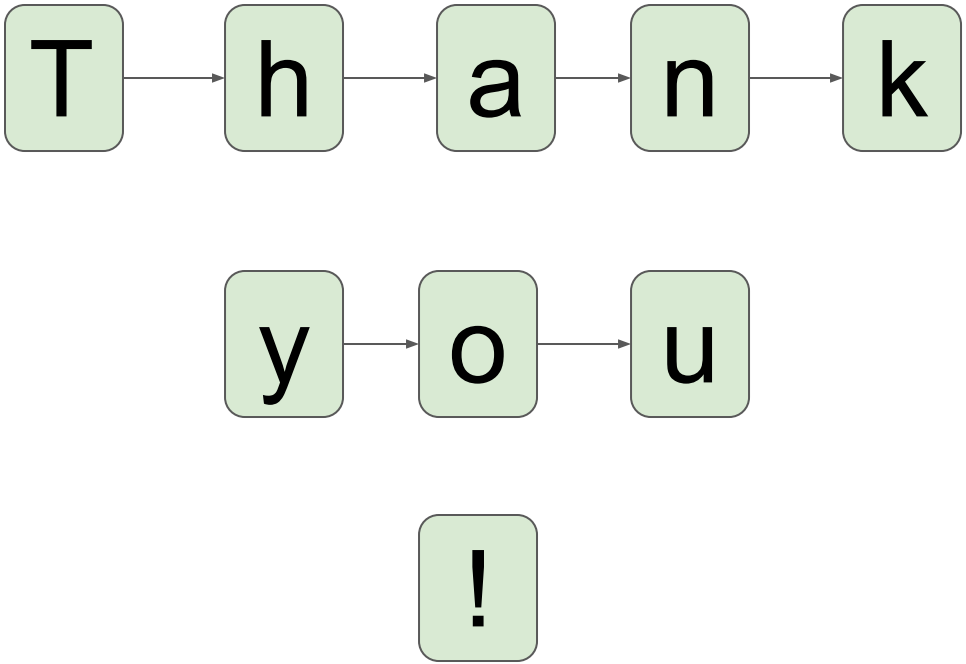


Intuition

Character-level
Language Generation



Rationalization


T h a n k
y o u
!