Infinite SCAN: An Infinite Model of Diachronic Semantic Change

Seiichi Inoue[1], Mamoru Komachi[1], Toshinobu Ogiso[2], Hiroya Takamura[3], and Daichi Mochihashi[4]

[1] Tokyo Metropolitan University
[2] The National Institute for Japanese Language and Linguistics
[3] The National Institute of Advanced Industrial Science and Technology
[4] The Institute of Statistical Mathematics

Sense Modeling with a Topic Model

1

Creating Inputs

target word: "*coach*"

Diachronic Corpus

Usage with Time Stamp

```
1853  The driver made room for the trunk on the top of the coach.
1900  The chair passed the coach, the horses proceeding at a walk.
1949  Tell him if I start coaching, it'll be as a head coach at a top school.
2003  Football coach and other top school officials have been interviewed.
                                ⋮
```

Snippets of Contextual Information

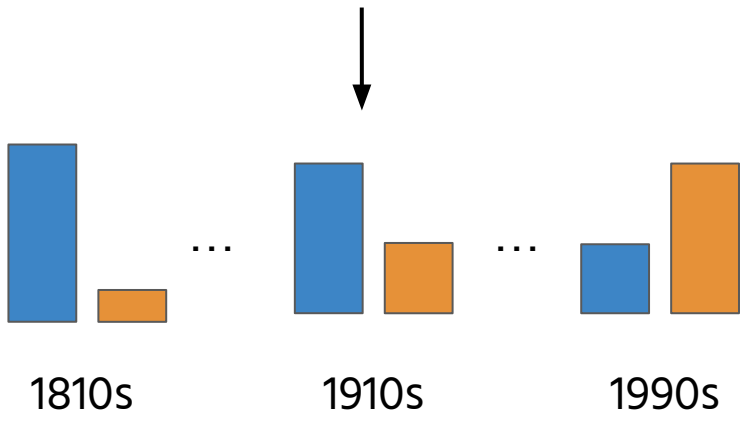
```
1853  [driver, make, room, trunk]
1900  [chair, pass, horse, proceed, walk]
1949  [tell, start, coach, head, top, school]
2003  [football, top, school, official, interview]
                        ⋮
```

Sense Modeling with a Topic Model

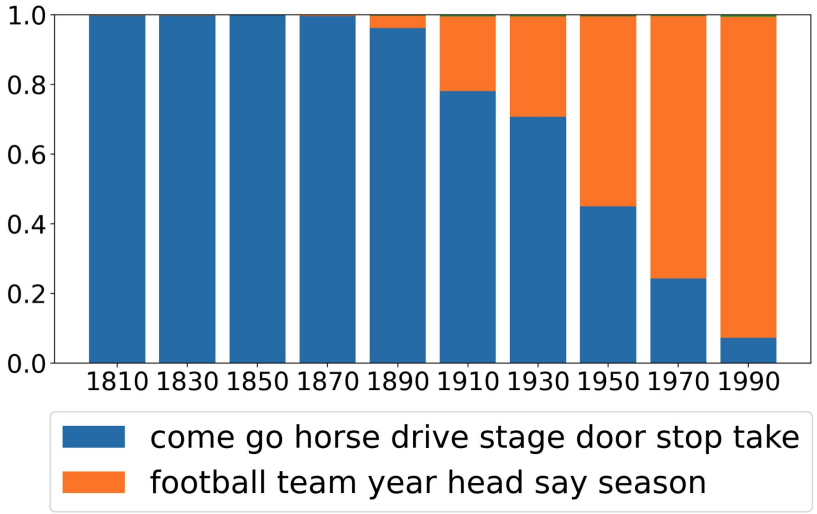
2 Sense Modeling

Snippets (inputs)

```
1853  [driver, make, room, trunk]
1900  [chair, pass, horse, proceed, walk]
1949  [tell, start, coach, head, top, school]
2003  [football, top, school, official, interview]
                        ⋮
```



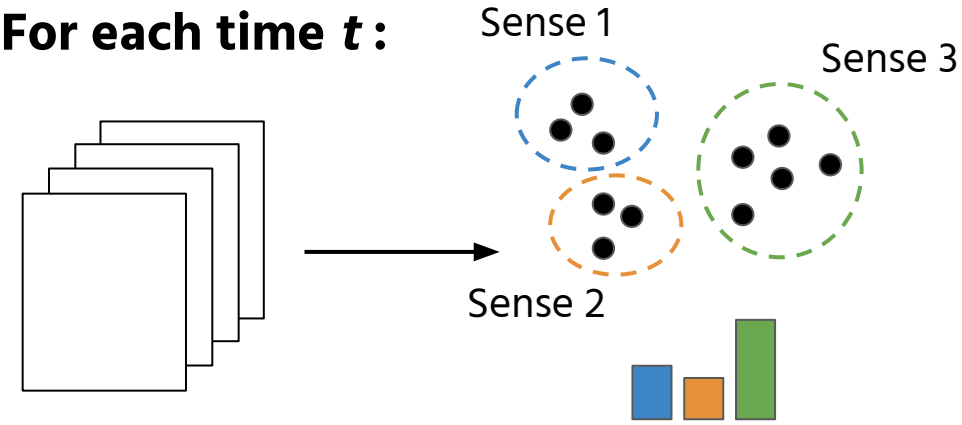
3 Analysis



Preliminaries: SCAN

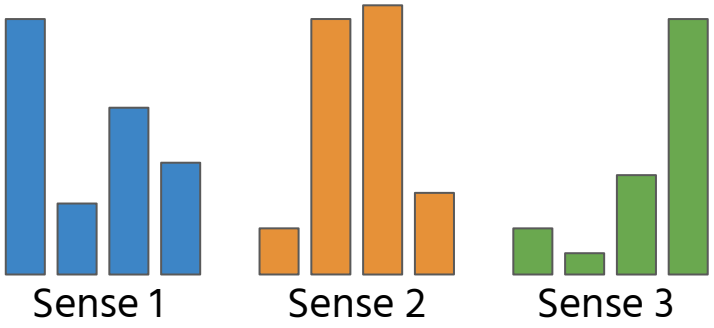
SCAN: Bayesian Model of Sense Change [Frermann+, 2016]

For each time *t* :

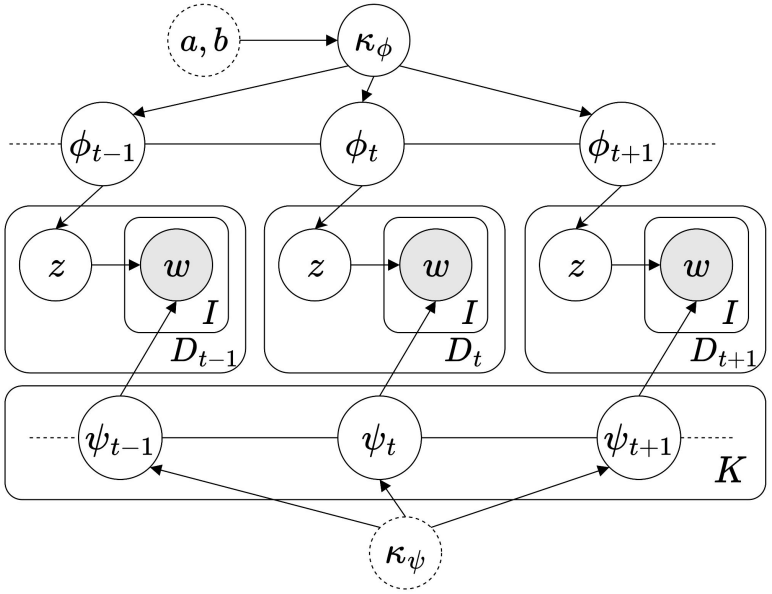


Snippets at time *t*

φ: Sense Distribution



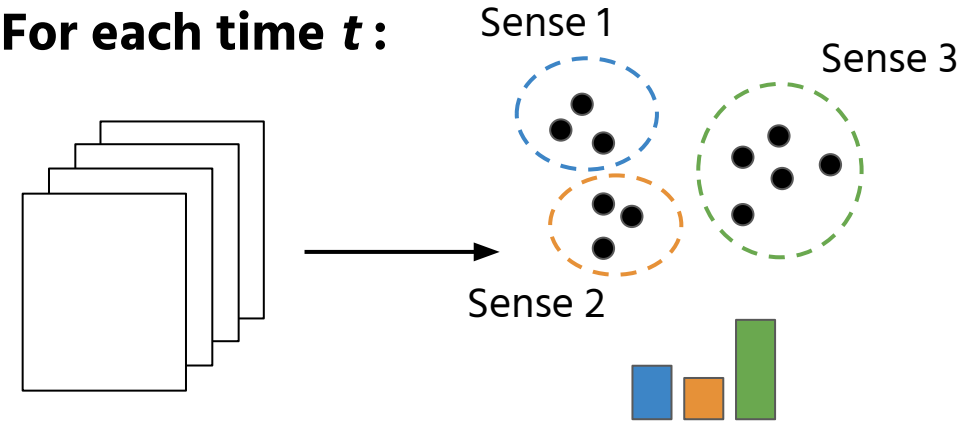
ψ: Word Distribution per Sense



Preliminaries: SCAN

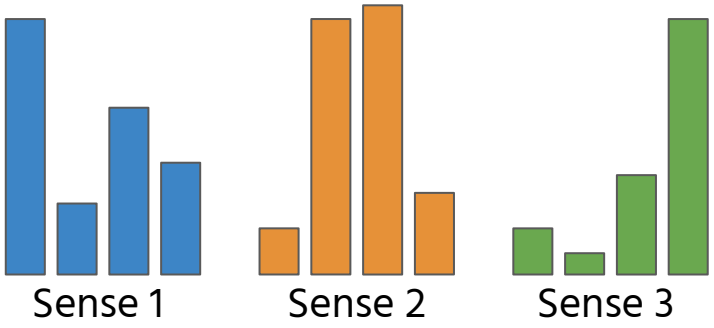
SCAN: Bayesian Model of Sense Change [Frermann+, 2016]

For each time *t* :

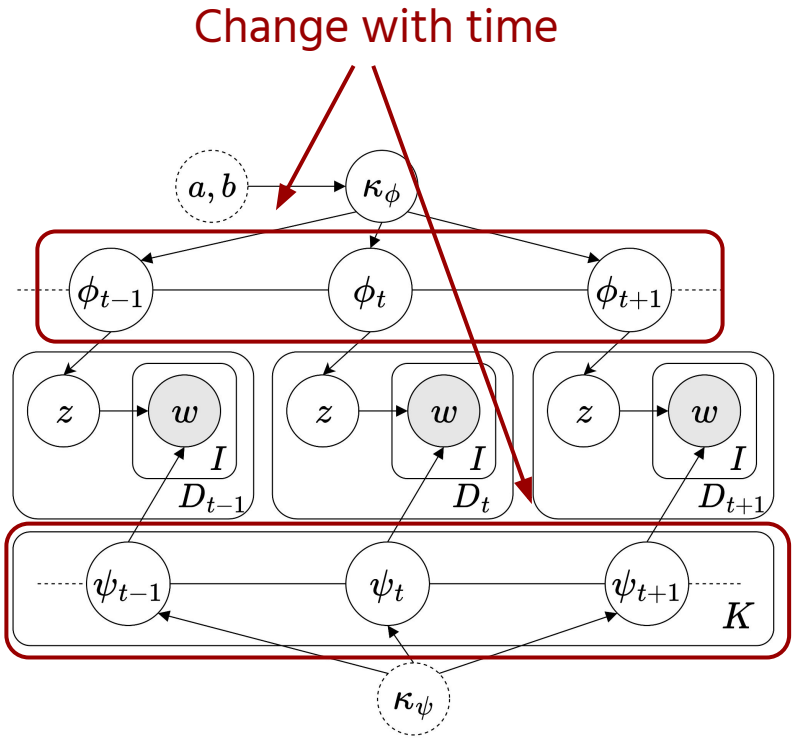


Snippets at time *t*

φ: Sense Distribution



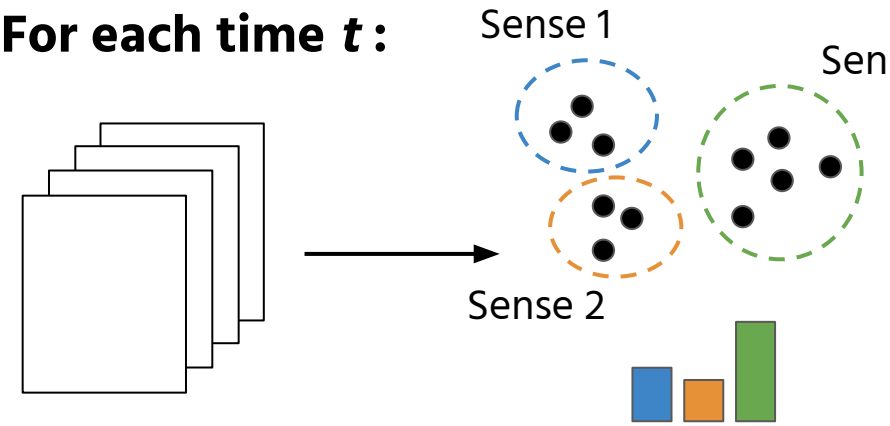
ψ: Word Distribution per Sense



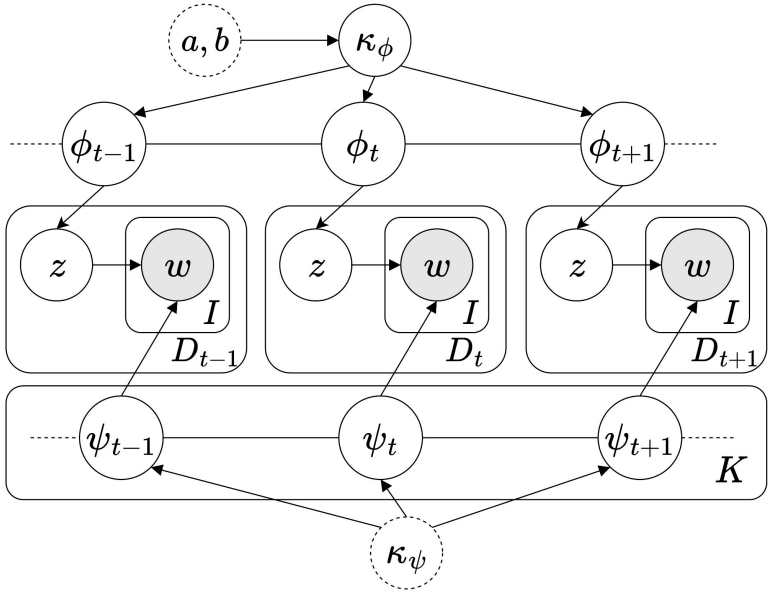
Preliminaries: SCAN

SCAN: Bayesian Model of Sense Change [Frermann+, 2016]

For each time *t* :

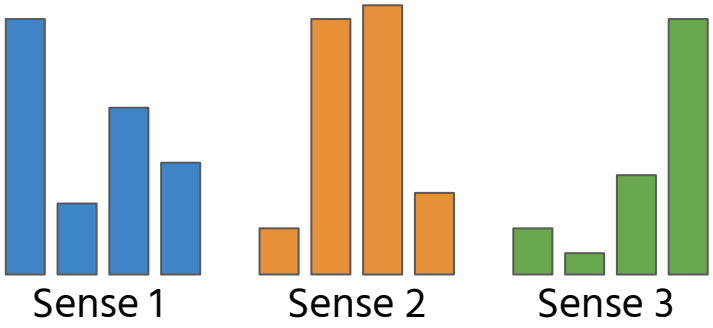


How many senses ?



Snippets at time *t*

φ: Sense Distribution

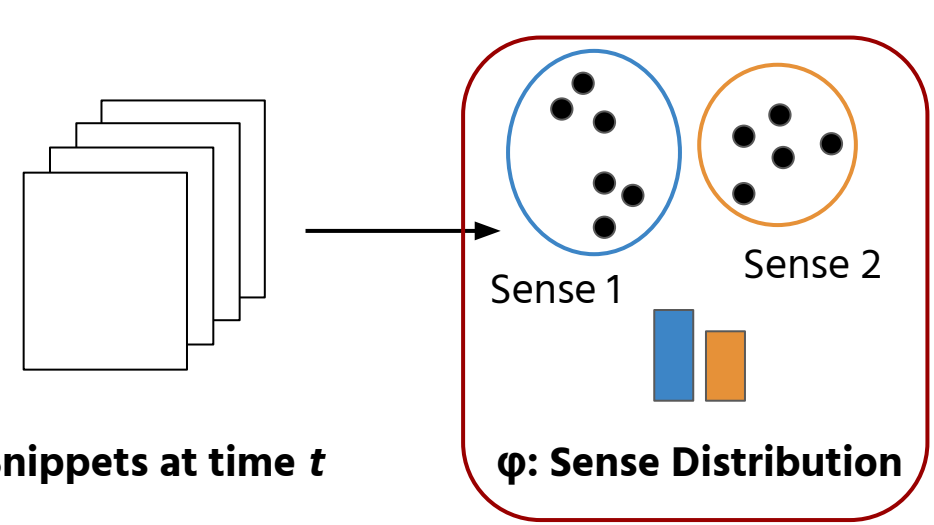


ψ: Word Distribution per Sense

Proposal: Infinite SCAN

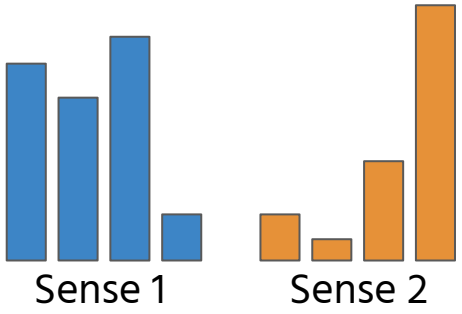
Infinite SCAN: Infinite Model of Diachronic Semantic Change

Automatically estimate the number of senses using Dirichlet process (Stick-breaking process)

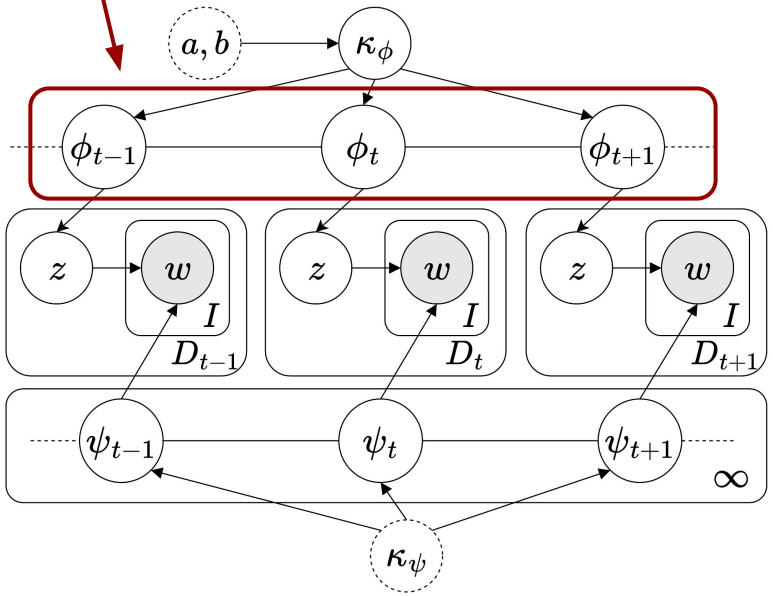


Snippets at time *t*

φ: Sense Distribution



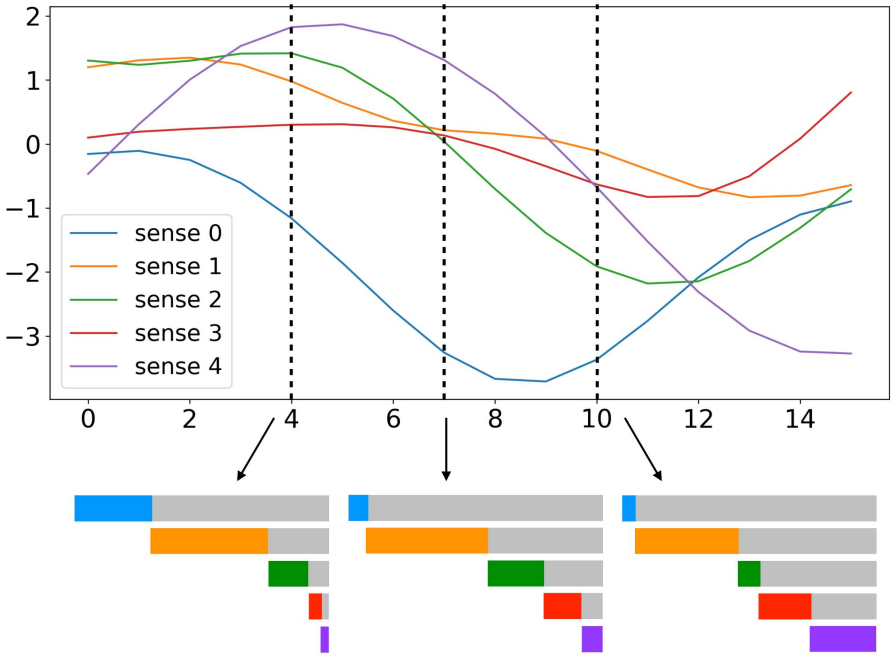
ψ: Word Distribution per Sense



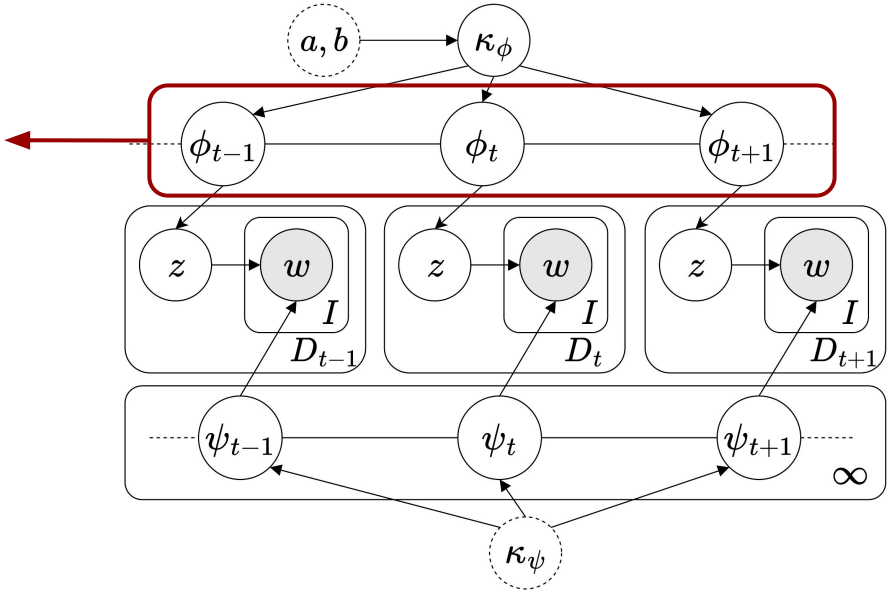
Proposal: Infinite SCAN

Stick-breaking transformation of Gaussian sense distribution

Logistic normal priors on
Gaussian Markov Random Field

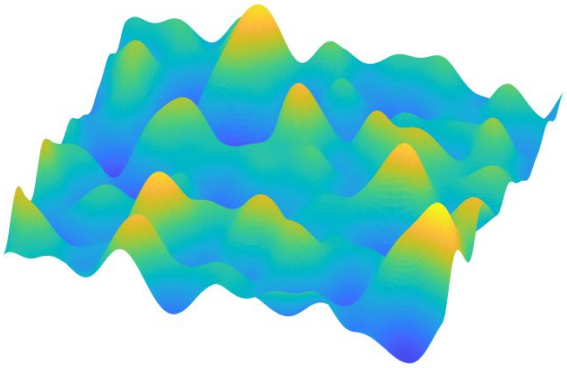


Projecting to K-1 simplex using Logistic
stick-breaking process (not softmax)



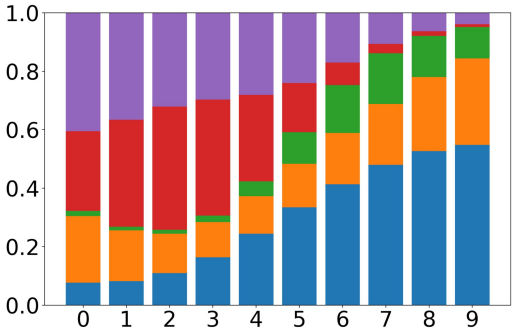
Setup: Experiments on Artificial Data

Random prior



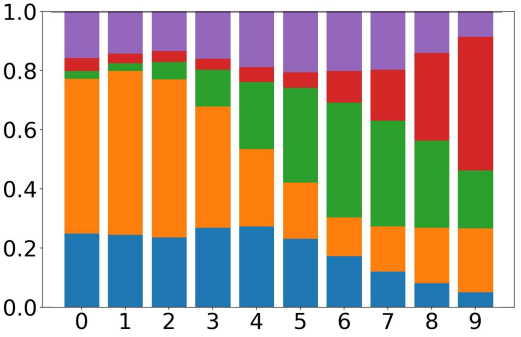
Draw

"Ideal" distribution



Estimate & Compare

Estimated distribution



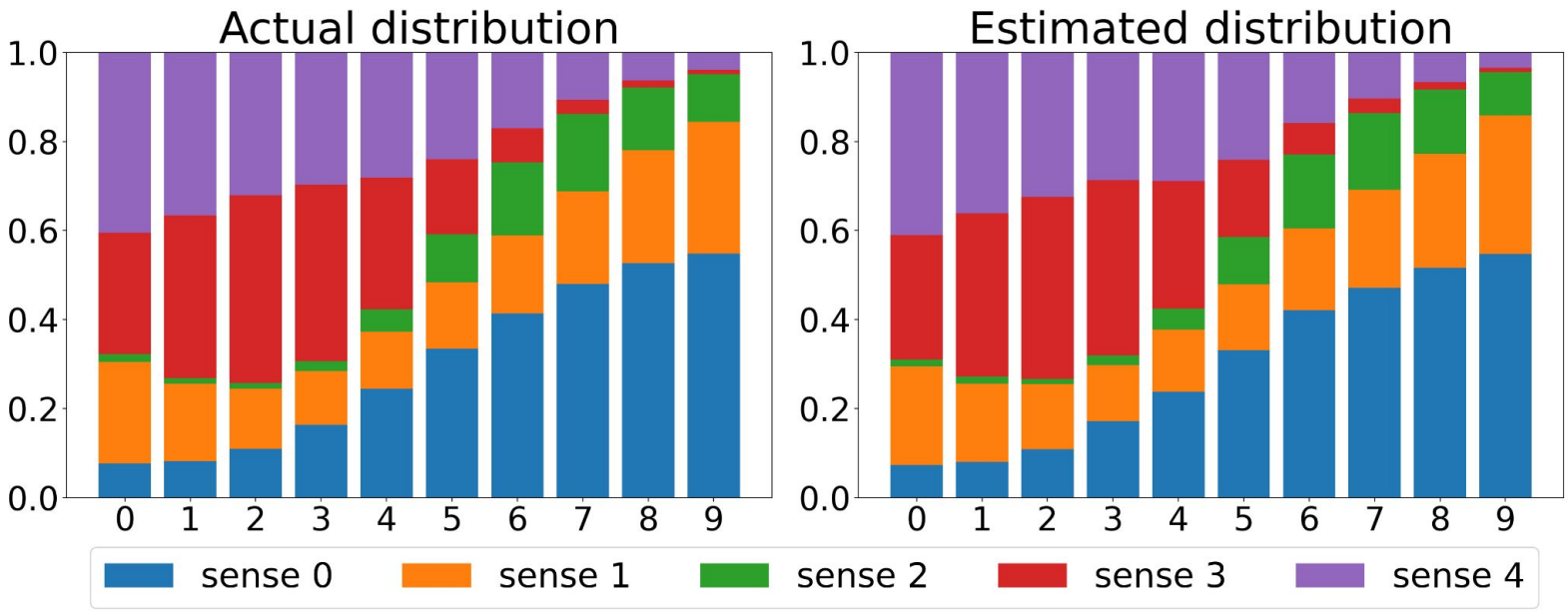
- Sense distribution: Draw from Gaussian process
- Sense-word distribution: Draw from Zipfian distribution

Experiments on Artificial Data

For artificial data with ranging the number of senses 1 to 5:

- Correctly estimate the number of senses except for $S = 1$
- Properly estimate the actual distribution: *KL (Actual || Estimated)* ≈ 0.0

Example for the number of senses: $S = 5$

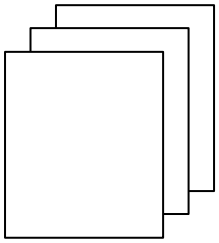


Kullback-Leibler divergence (Actual || Estimated) = **0.004**

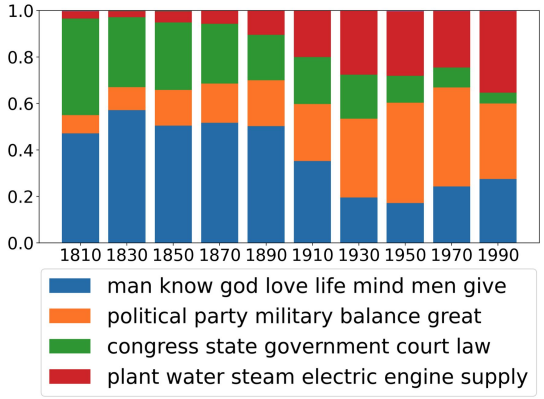
Setup: Experiments on Real Data

Diachronic Corpus
CCOHA [Alatrash+, 2020]

Extract



Snippets



Estimated distribution

Evaluate

Gold number of senses

Sense inventory of
OntoNotes [Hovy+, 2006]

Known
Sense change

Previous work
[Giulianelli+, 2020;
Frermann+, 2016]

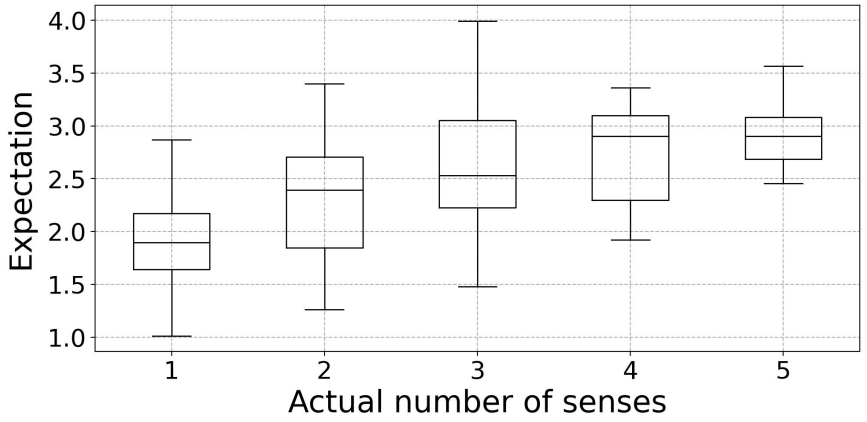
Experiments on Real Data

Can the model correctly estimate the number of senses?

For 120 target words randomly selected from OntoNotes:

Model	Accuracy	PCC
HDP-LDA	0.258	0.019
BERT + K-means	0.217	0.026
BERT + DBSCAN	0.125	-0.070
SCAN ($K = 5$)	0.158	0.141
SCAN ($K = 8$)	0.000	0.087
Infinite SCAN	**0.358**	**0.474**

Prediction results for the number of senses.

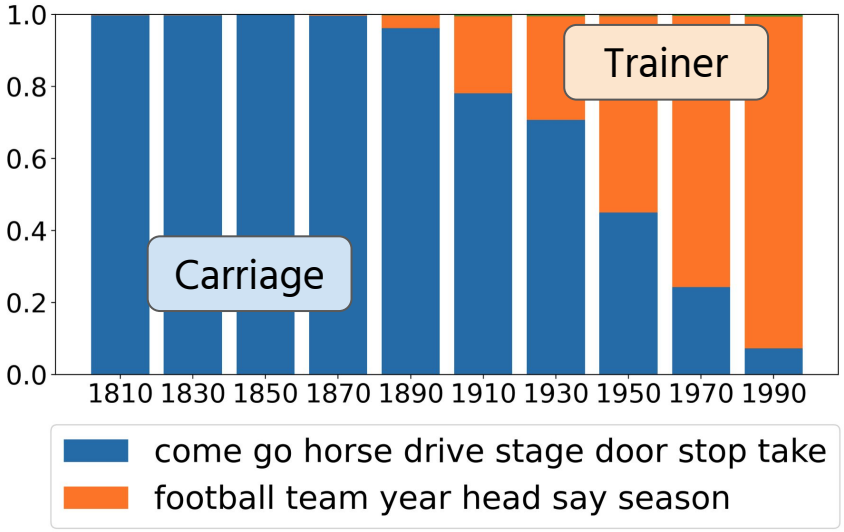


Correlation between actual and estimated number of senses by Infinite SCAN.

Experiments on Real Data: *"coach"*

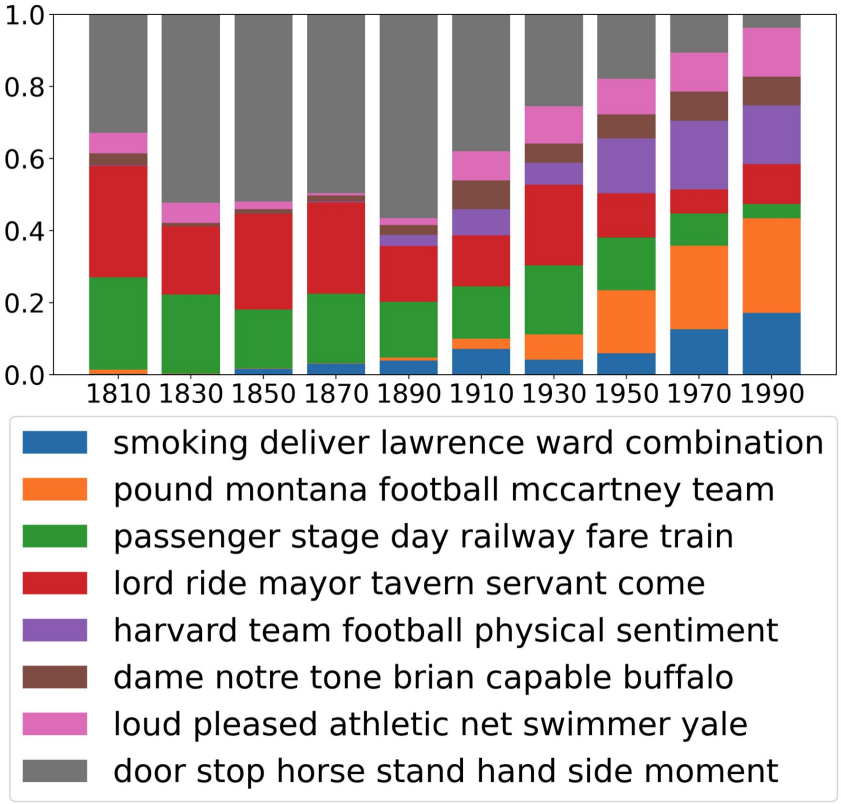
Can the model properly describe the semantic change?

Infinite SCAN



Semantic shift: "carriage" → "trainer"

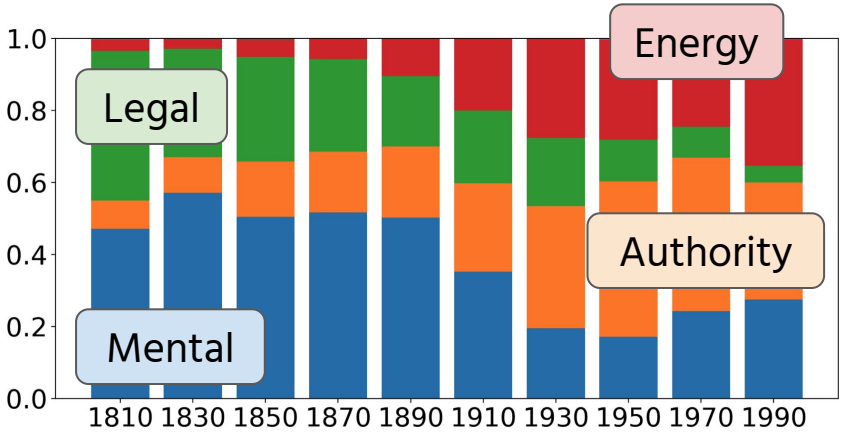
SCAN



Experiments on Real Data: *"power"*

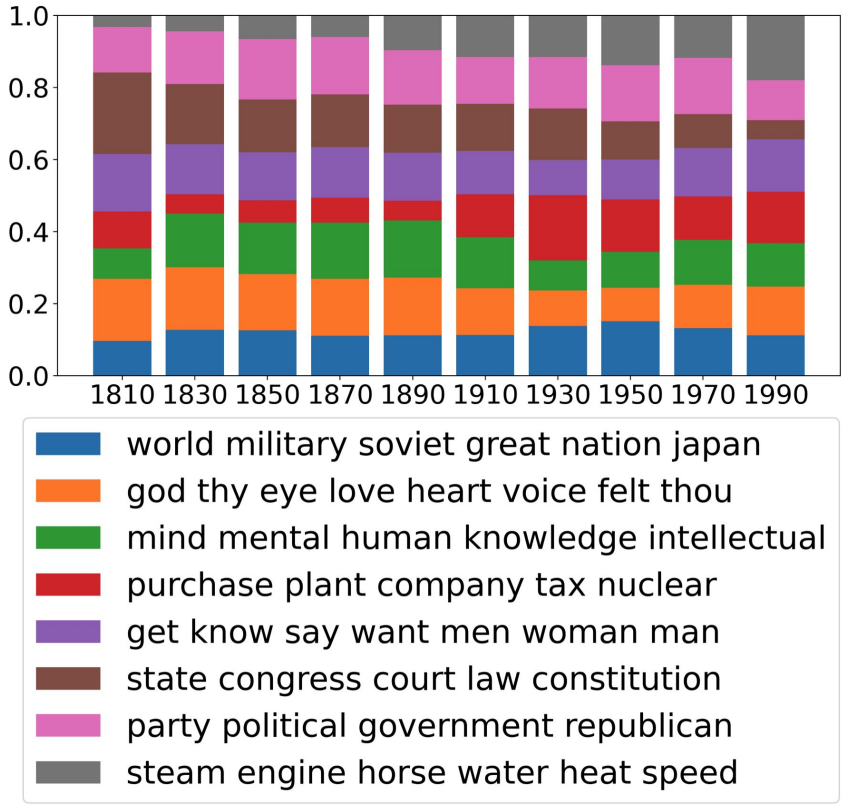
Can the model properly describe the semantic change?

Infinite SCAN



Sense birth: "energy"

SCAN



Conclusions / Future Work

- Infinite SCAN: An infinite model of diachronic semantic change
  - based on dynamic topic model
  - combining Gaussian Markov random fields with a Dirichlet process
- Outperformed baseline models
  - interpretability of model output
  - estimation of the number of senses
- Captured semantic changes in line with previous studies
- In future, working on classification of change patterns

Appendices: Setup of experiments on Artificial Data

Generation of artificial data

- Sense distribution: Draw from Gaussian process with the number of senses K
- Sense-word distribution: Draw from Zipfian distribution with the vocabulary size V

Parameter settings

- Number of time points: $T = 10$
- Original vocabulary size: $V = 5{,}000$
- Snippet length: $l = 10$
- Number of senses: $S = 1 \sim 5$
- Sample size: $N = 10{,}000$

Metric

- Kullback-Leibler distance between actual and estimated distribution

#Senses	SCAN $K=5$	SCAN $K=8$	Infinite SCAN
1	1.523	1.997	**0.468**
2	0.335	0.578	**0.039**
3	0.216	0.735	**0.030**
4	0.212	0.150	**0.061**
5	0.004	0.017	**0.004**

Table 1: Kullback-Leibler divergence (lower is better) between actual sense distribution and sense distribution estimated by each model for the artificial data.

Appendices: Setup of experiments on Real Data

Historical corpus

Clean Corpus of Historical American English (CCOHA) [Alatrash+, 2020]

Target words with gold number of senses

randomly selected 120 words (noun, verbs) with $S \leqq 5$ from OntoNotes [Hovy+, 2006]

Target words with semantic change description

- "*coach*" [Giulianelli+, 2020; Aida+, 2021]
- "*power*" [Frermann+, 2016]

Baseline models

- HDP-LDA [Teh+, 2010]
- BERT (base-uncased) [Devlin+, 2019] + K-means, DBSCAN
- SCAN [Frermann+, 2016]