# Scan – A Bayesian Model of Diachronic Meaning Change

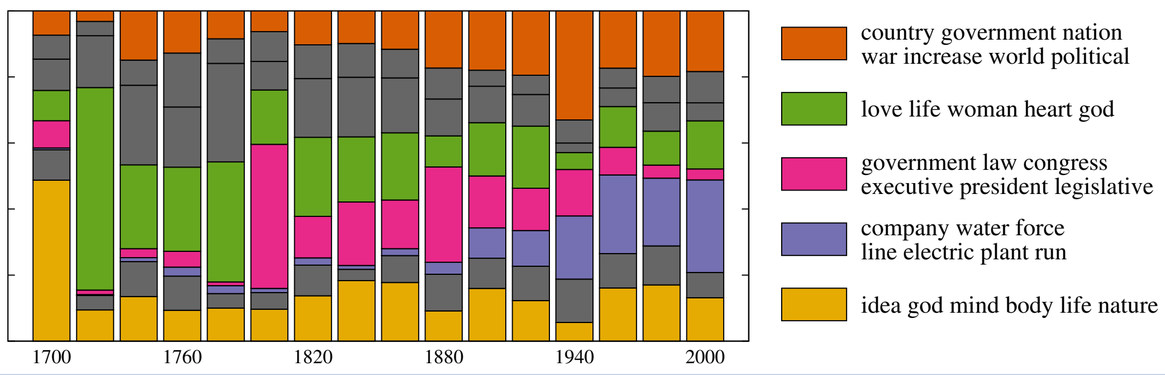
Lea Frermann
l.frermann@ed.ac.uk
Institute for Language, Cognition, and Computation, School of Informatics, University of Edinburgh

## 1. Motivation

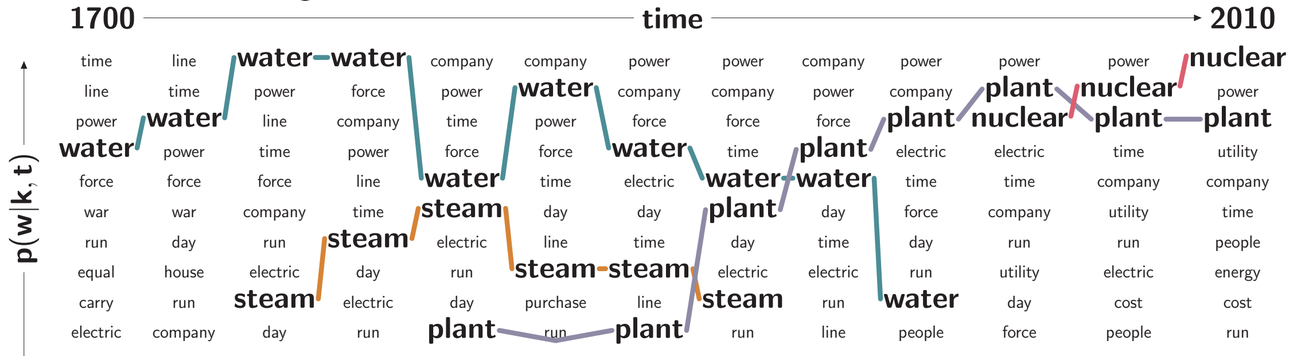
- Meaning is constantly shaped by language users and their environment: across (  ⇝  ) and within (  ⇝  ) senses.
- Large scale historical corpora are useful for socio-linguistic research
  Modeling sense change is useful for QA and IR
- We model **meaning change** as a **smooth process**
- In contrast to previous work, we
  (a) model meaning change at **sense** level
  (b) explicitly capture the **gradual** nature of language **change**
  (c) **infer** meaning representations and **detect** changes with one coherent model

## 2. Intuition: Temporal Dynamics of Word Meaning (example 'power')

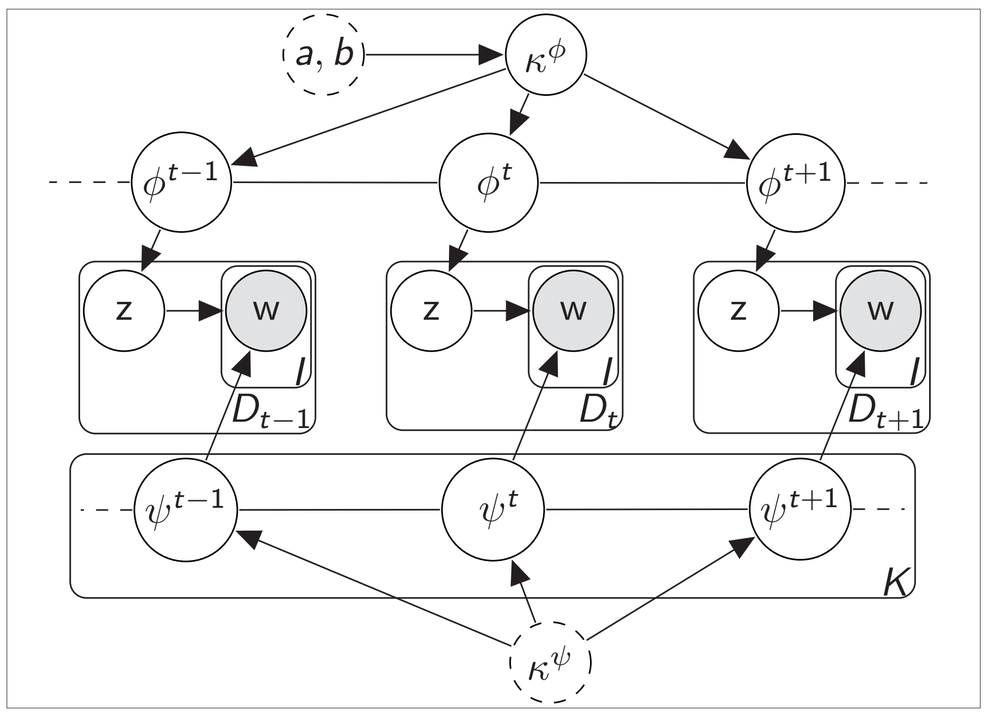
Change in sense prevalence




Sense-internal change




## 3a. Model Description




### Input

- Target word $c$ and corpus of time-tagged documents $[w_{-5} \dots w_{-1} \quad c \quad w_{+1} \dots w_{+5}]$
- Assumption: local context predicts word sense

### Modeling

- Time $t \in [1..T]$ discrete contiguous intervals of fixed size $\Delta t$
- Meaning $k \in [1..K]$ senses (distributions over context words $w$)

### Data Distributions

- Multinomial $\{$time,sense$\}$-specific word distributions $\{\psi\}^{K \times T}$ $p(w|k,t)$
- Multinomial time-specific sense distributions $\{\phi\}^{T}$ $p(k|t)$

### Prior Distributions

- Logistic normal priors on the parameters $\phi$ and each $\psi^k$

$$\beta \sim \mathcal{N}(\mu, \Sigma) \qquad \phi_k = exp(\beta_k) / \sum_{\beta_{k'}} exp(\beta_{k'})$$

- Gamma prior on precision parameter $\kappa^\phi$ (flexibility of sense prevalence)

## 3b. iGMRF priors (Rue and Held, 2005)

- Multivrt Normal distribution $\mathcal{N}(\beta; \mu, \mathcal{Q}^{-1})$ || $\mathcal{G}$: undir graphichal model over $\beta$
- $\mathcal{Q}$ is tridiagonal (and not full rank) || 1st oder random walk on the line
- (i) assume independent *local increments*
  (ii) the value of r.v. $x_i$ is centered around the mean of its neighbors (full cond.)

$$(i) \;\; \Delta\beta^i \sim \mathcal{N}(0, \kappa^{-1}) \qquad (ii) \;\; \beta^i|\beta^{-i} \sim \mathcal{N}\Big(\frac{1}{2}(\beta^{i-1} + \beta^{i+1}), \frac{1}{2\kappa}\Big)$$

- Level of allowed flexibility controlled through precision parameter $\kappa$
- Invariant to global trend / addition of constant

## 3c. Inference: Blocked Gibbs Sampling (Mimno et al, 2008)

**1 Resample sense-assignments** (given logistic normal parameters)

$$p(k^d = i|\mathbf{w}^d, t^d, \phi, \psi) \;\propto\; \phi^i_{t^d} \prod_{w \in \mathbf{w}^d} \psi^{i,t^d}_w$$

**2 Resample logistic-normal parameters** (given sense-assignments)

- Auxiliary variable sampler
- Shift parameter in weighted and bounded region
  → bounds given by # assignments of parameter to data (likelihood)
  → weights and variance (precision $\kappa$) through iGMRF prior

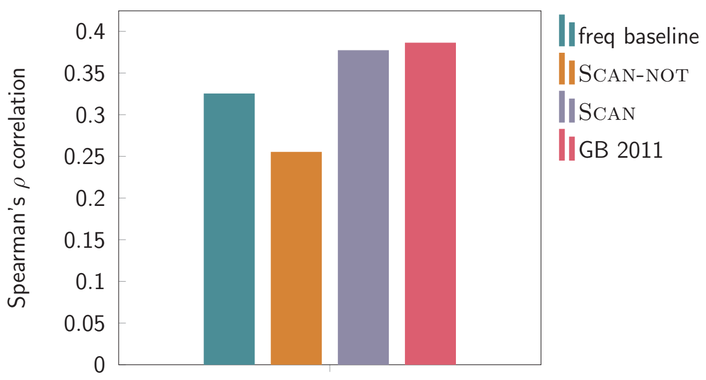
**3 Resample sense prevalence flexibility** $\kappa^\phi$ (from Gamma prior)

## 4a. Experiment 1: Word Meaning Change (Gulordava and Baroni, 2011)

**Task** Rank words by perceived novelty (1990s vs 1960s)

**Data** 100 target words (+human ranking) ; training documents (1960s – 1990s)

**Results**




| most novel words in 1990s | |
|---|---|
| users | computer window information software |
| virtual | reality virtual computer center experience |
| disk | hard disk drive program computer file store |

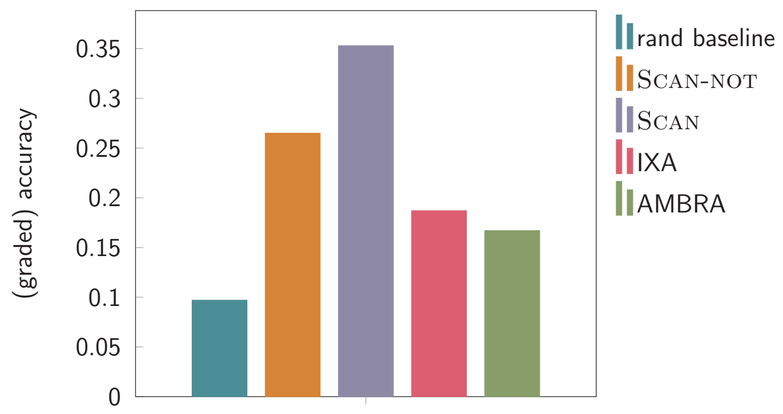
SCAN-NOT → SCAN model without temporal dependencies (iGMRF)
learns independent representation for each time interval $t$

## 4b. Experiment 2: Temporal Text Classification (Semeval 2015)

**Task** Classify a snippet of news text wrt its year of origin

**Data** *President de Gaulle favors an independent European striking force [...]*

**Results**




**Intervals**

1700–1702
1703–1705
1706–1708
...
1958-1960
**1961–1963**
1964–1966
...
2007-2009
2009–2011
2012–2014