Modelling fine-grained Change in Word Meaning over centuries from Large Collections of Unstructured Text

Lea Frermann

Keynote at Drift-a-LOD Workshop, November 20, 2016

Institute for Language, Cognition, and Computation
The University of Edinburgh

lea@frermann.de
www.frermann.de

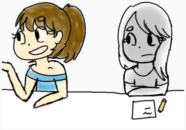
The Dynamic Nature of Meaning I

Language is inherently ambiguous

Words have different meanings (or senses), e.g. `mouse`

animal

shy person

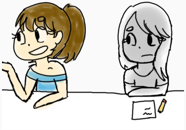


computing device

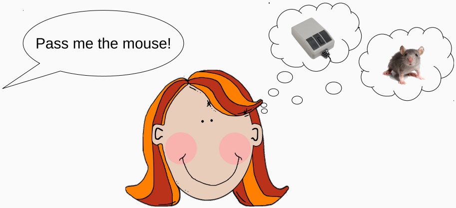
The Dynamic Nature of Meaning I

Language is inherently ambiguous

Words have different meanings (or senses), e.g. `mouse`

animal	shy person	computing device
		

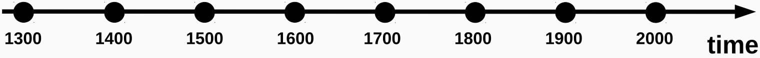
… and the relevant sense depends on the **context** or situation



The dynamic Nature of Meaning II

Language is a dynamic system

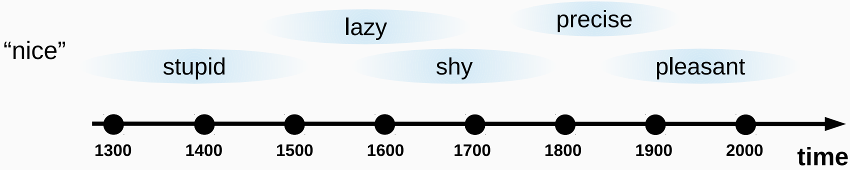
The meaning of words is **constantly shaped** by users and their environment



The dynamic Nature of Meaning II

Language is a dynamic system

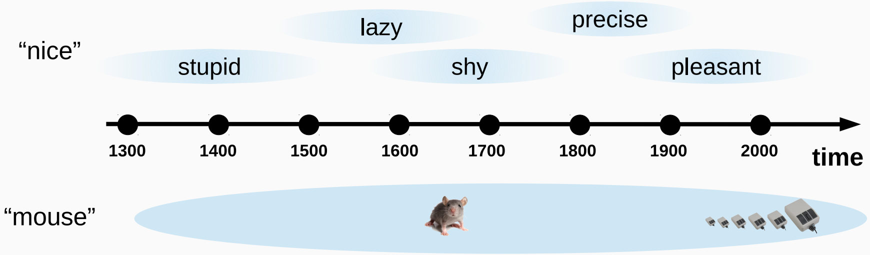
The meaning of words is **constantly shaped** by users and their environment



The dynamic Nature of Meaning II

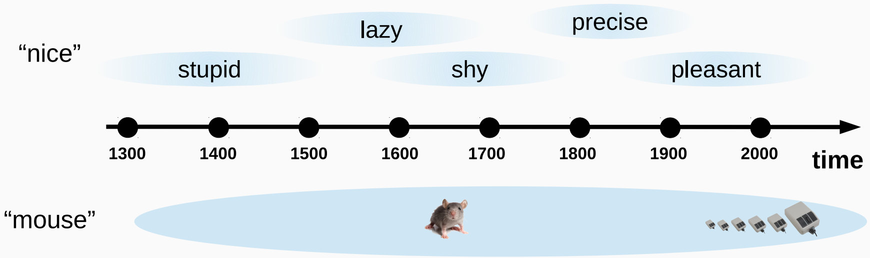
Language is a dynamic system

The meaning of words is **constantly shaped** by users and their environment



Language is a dynamic system

The meaning of words is **constantly shaped** by users and their environment



Meaning changes **smoothly** (in written language, across societies)

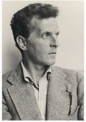
The distributional Hypothesis

"You shall know a word by the company it keeps."

John R. Firth (1957)

"The meaning of a word is its use in the language."

Ludwig Wittgenstein (1953)



The distributional Hypothesis

Distributional Semantics: Take large collections of texts and look at the **contexts** in which a **target word** occurs

left context	target	right context
finance director used the	`mouse`	and expanded a window
nose twitching like a	`mouse`	's , but Douggie 's
There 's been a	`mouse`	in the pantry , " she said
using the	`mouse`	, and learning how to type
She can see the	`mouse`	rolling that pearl to its hole
She was quiet as a	`mouse`	most of the time
...	...	...

→ **characterize senses and their prevalence**

The distributional Hypothesis

Distributional Semantics: Take large collections of texts and look at the **contexts** in which a **target word** occurs

left context	**target**	**right context**
finance director used the		and expanded a window
nose twitching like a		's , but Douggie 's
There 's been a		in the pantry , " she said
using the		, and learning how to type
She can see the		rolling that pearl to its hole
She was quiet as a		most of the time
...	...	...

→ **characterize senses and their prevalence**

The distributional Hypothesis

Distributional Semantics: Take large collections of texts and look at the **contexts** in which a **target word** occurs

left context	target	right context
finance director used the		and expanded a window
nose twitching like a		's , but Douggie 's
There 's been a		in the pantry , " she said
ing the		, and learning how to type
She can see the		rolling that pearl to its hole
She was quiet as a		most of the time
. . .	. . .	. . .

nose tail
roll cheese cat
hole

keyboard expand
file open klick
computer

quiet
shy still
timid

→ **characterize senses and their prevalence**

The distributional Hypothesis

Distributional Semantics: Take large collections of texts and look at the **contexts** in which a **target word** occurs

left context	target	right context	year
What teaches the little		to hide , with its glimmering	1823
you couldn't hide a		here without its being	1849
Laura thinks she sees a		, an' she trembles an' she	1915
caused cancer in a		or a hamster	1972
she 's such a quiet little		and everyone 's in love	1982
finance director used the		and expanded a window	2000
nose twitching like a		's , but Douggie 's	2000
She was quiet as a		most of the time	2000
using the		, and learning how to type	2000
she clicked the		until her fingers tingled	2008

→ **characterize senses and their prevalence <u>over time</u>**

Motivation

We want to **understand, model,** and **predict** word meaning change at scale

Motivation

We want to **understand, model,** and **predict** word meaning change at scale

Why is this an important problem?

- aid historical sociolinguistic research
- improve historical text mining and information retrieval
- aid onthology construction / updating

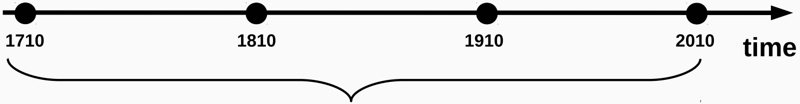
Can we build task-agnostic models?

- learn time-specific meaning representations which
- are interpretable and
- are useful across tasks

Data

DATE – A DiAchronic TExt Corpus

We use **three** historical corpora



DTE training corpus – SemEval 2015 Task

- **D**iachronic **T**ext **E**valuation
- News text
- Snippets between a few words and multiple sentences

DATE – A DiAchronic TExt Corpus

We use **three** historical corpora

COHA – Corpus of Historical American English

- 400 million words
- Texts from multiple genres
  (e.g., fiction, news, other non-fiction, …)

DATE – A DiAchronic TExt Corpus

We use **three** historical corpora

CLMET – Corpus of Late Modern English Texts

- We use a subset (full corpus: 1710 – 1920)
- Full texts form open online archives
  (e.g., Project Gutenberg)

DATE – A DiAchronic TExt Corpus

We use **three** historical corpora

CLMET – Corpus of Late Modern English Texts

- We use a subset (full corpus: 1710 – 1920)
- Full texts form open online archives (e.g., Project Gutenberg)

Why not **Google Books**? → only provides up to 5-grams.

Data Preprocessing

1. Text Processing

original →	she clicked the mouse, until her fingers tickled.
tokenize →	she clicked the mouse , until her fingers tickled .
lemmatize →	she click the mouse , until she finger tickle .
remove stopwords →	click mouse finger tickle
POS-tag →	click_V mouse_N finger_N tickle_V

2. Cluster texts from 3 corpora by year of publication

→ **Create target word-specific training corpora**

DATE – A DiAchronic TExt Corpus

Target word-specific training corpora

All mentions of target word with context of $\pm$ 5 surrounding words tagged with year of origin

text snippet			year
fortitude time woman shrieks	*mouse*	rat capable poisoning husband	1749
rabbit lived hole small grey	*mouse*	made nest pocket coat	1915
ralph nervous hand twitch computer	*mouse*	keyboard pull image file online	1998
scooted chair clicking button wireless	*mouse*	hibernate computer stealthy exit	2009
	...		

Scan: A Dynamic Model of Sense chAnge

Model Input and Assumptions

- target word-specific corpus

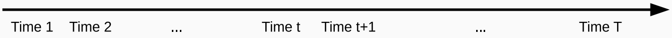
text snippet			year
fortitude time woman shrieks	*mouse*	rat capable poisoning husband	1749
rabbit lived hole small grey	*mouse*	made nest pocket coat	1915
ralph nervous hand twitch computer	*mouse*	keyboard pull image file online	1998
scooted chair clicking button wireless	*mouse*	hibernate computer stealthy exit	2009
	...		

- number of word senses (K)

- granularity of temporal intervals (ΔT)
  (e.g., a year, decade, or century)

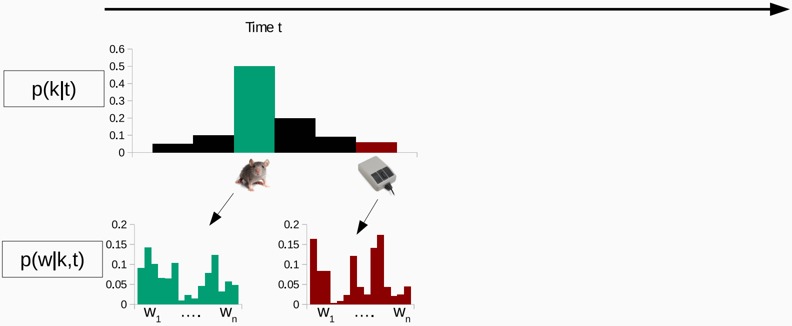
Model Overview

A Bayesian and knowledge-lean model of meaning change of individual words (e.g., "mouse")



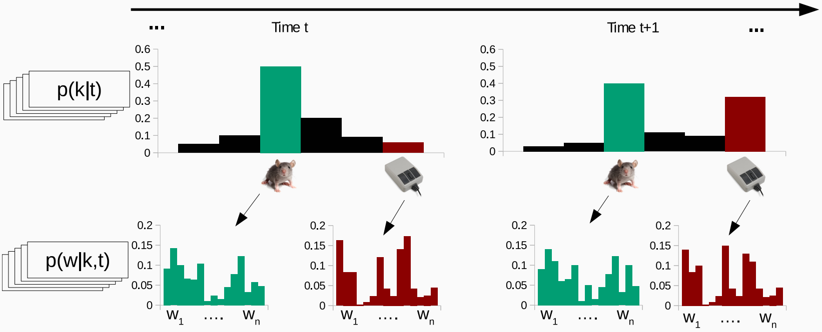
Model Overview

A Bayesian and knowledge-lean model of meaning change of individual words (e.g., “mouse”)



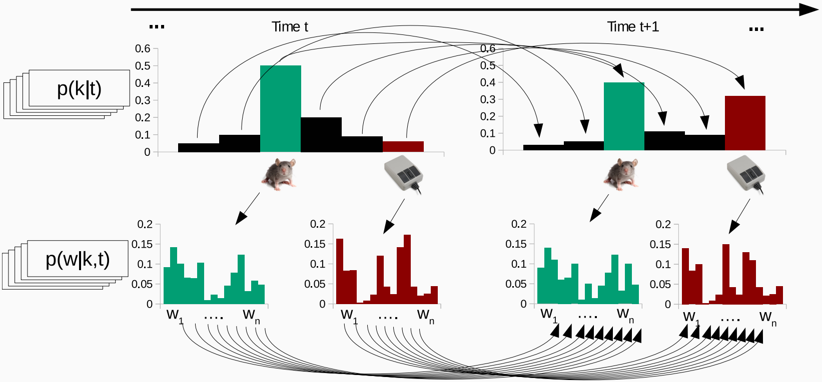
Model Overview

A Bayesian and knowledge-lean model of meaning change of individual words (e.g., "mouse")



Model Overview

A Bayesian and knowledge-lean model of meaning change of individual words (e.g., "mouse")

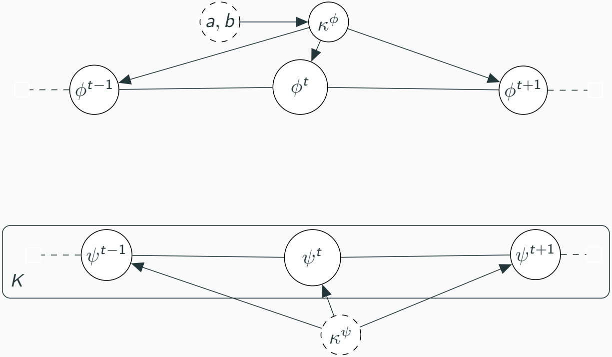


Model Description: Generative Story

Model Description: Generative Story

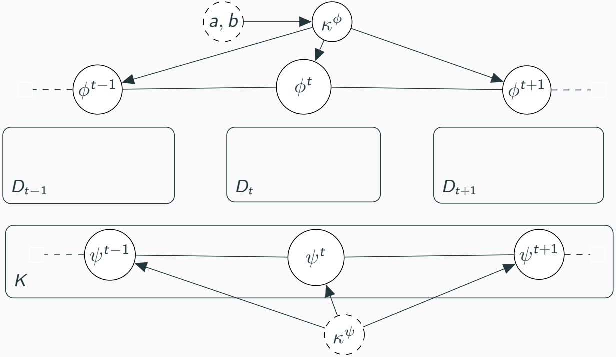
1. Extent of meaning change	Generate temporal sense flexibility parameter $\kappa^{\phi} \sim Gamma(a, b)$

Model Description: Generative Story



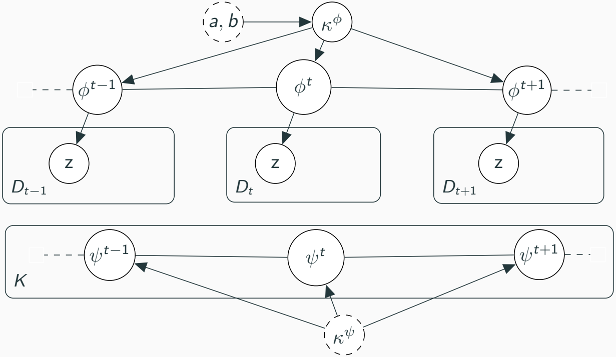
1. Extent of meaning change	Generate temporal sense flexibility parameter $\kappa^\phi \sim \mathit{Gamma}(a, b)$
2. Time-specific representations	Generate sense distributions ϕ^t Generate sense-word distributions $\psi^{k,t}$

Model Description: Generative Story



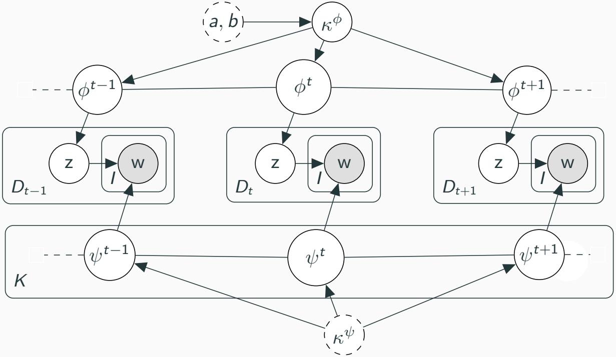
1. Extent of meaning change	Generate temporal sense flexibility parameter $\kappa^{\phi} \sim Gamma(a, b)$
2. Time-specific representations	Generate sense distributions ϕ^{t} Generate sense-word distributions $\psi^{k,t}$
3. Text generation given time t	

Model Description: Generative Story



1. Extent of meaning change	Generate temporal sense flexibility parameter $\kappa^{\phi} \sim Gamma(a, b)$
2. Time-specific representations	Generate sense distributions ϕ^{t} Generate sense-word distributions $\psi^{k,t}$
3. Text generation given time t	Generate sense $z \sim Mult(\phi^{t})$

Model Description: Generative Story

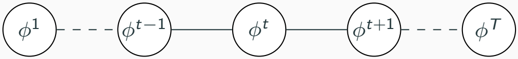


1. Extent of meaning change	Generate temporal sense flexibility parameter $\kappa^\phi \sim Gamma(a, b)$
2. Time-specific representations	Generate sense distributions ϕ^t Generate sense-word distributions $\psi^{k,t}$
3. Text generation given time t	Generate sense $z \sim Mult(\phi^t)$ Generate context words $w_i \sim Mult(\psi^{t,k=z})$

Scan: The Prior

First-order random walk model

intrinsic Gaussian Markov Random Field (Rue, 2005; Mimno, 2009)



draw **local changes** from a normal distribution

mean	temporally neighboring parameters
variance	meaning flexibility parameter κ^ϕ

Blocked Gibbs sampling

Details in Frermann and Lapata (2016)

Related Work

Word meaning change

Gulordava (2011), Popescu (2013), Kim (2014) , Kulkarni (2015)

Word	Neighboring Words in	
	1900	2009
gay	*cheerful* *pleasant* *brilliant*	*lesbian* *bisexual* *lesbians*

- ✗ word-level meaning
- ✗ two time intervals
- ✗ representations are independent
- ✓ knowledge-lean

Related work

Word meaning change

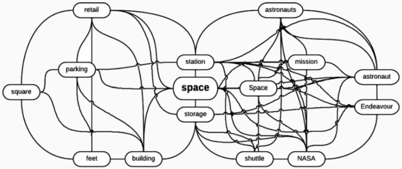
Gulordava (2011), Popescu (2013), Kim (2014) , Kulkarni (2015)

Word	Neighboring Words in	
	1900	2009
gay	*cheerful* *pleasant* *brilliant*	*lesbian* *bisexual* *lesbians*

- ✗ word-level meaning
- ✗ two time intervals
- ✗ representations are independent
- ✓ knowledge-lean

Graph-based tracking of word sense change

Mitra (2014, 2015)



- ✓ sense-level meaning
- ✓ multiple time intervals
- ✗ representations are independent
- ✗ knowledge-heavy

Evaluation

Evaluation: Overview

✗ no gold standard test set or benchmark corpora

✗ small-scale evaluation with hand-picked test examples

DATE: **DiAchronic TExt Corpus** (years 1710 – 2010)

1. COHA Corpus (Davies, 2010)
2. SemEval DTE Task Training Data (Popescu, 2015)
3. parts of the CLMET3.0 corpus (Diller, 2011)

Evaluation: Overview

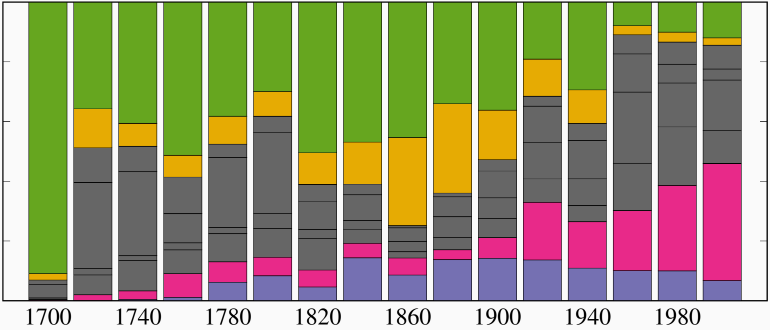
- ✗ no gold standard test set or benchmark corpora
- ✗ small-scale evaluation with hand-picked test examples

We evaluate on various previously proposed tasks and metrics

1. qualitative evaluation
2. perceived word novelty (Gulordava, 2011)
3. temporal text classification SemEval DTE (Popescu, 2015)

1. Qualitative Evaluation

battery



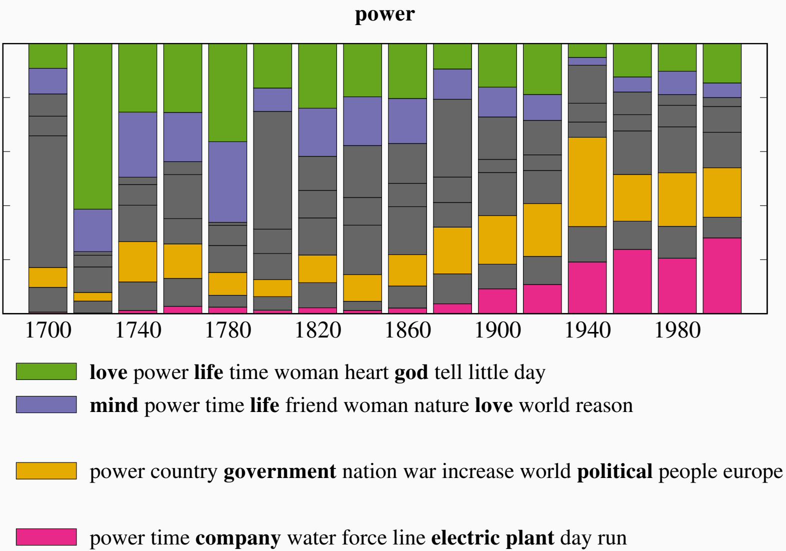
fire gun enemy fort shore time position **shell**

gun artillery battery infantry regiment cavalry **fire**

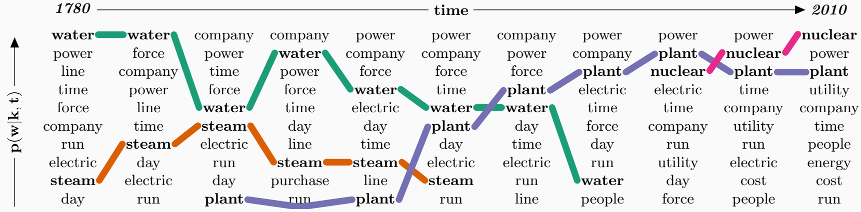
light **power** battery run **charge** hour life

wire battery **current** electric plate **power** cell **electricity**

1. Qualitative Evaluation



1. Qualitative Evaluation



2. Human-perceived Word Meaning Change (Gulordava (2011))

Task: Rank 100 `target words` by meaning change.

How much did $\begin{cases} \texttt{baseball} \\ \texttt{network} \\ \ldots \end{cases}$ change between the 1960s and the 1990s?

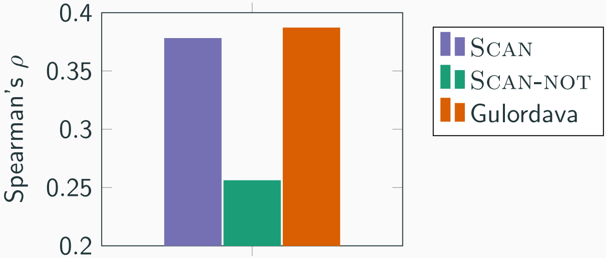
4-point scale 0: no change ... 3: significant change

2. Human-perceived Word Meaning Change (Gulordava (2011))

Task: Rank 100 `target words` by meaning change.

How much did { `baseball`, `network`, ... } change between the 1960s and the 1990s?

4-point scale   0: no change ... 3: significant change



2. Human-perceived Word Meaning Change (Gulordava (2011))

Task: Rank 100 `target words` by meaning change.

How much did $\begin{cases}\texttt{baseball} \\ \texttt{network} \\ \ldots\end{cases}$ change between the 1960s and the 1990s?

4-point scale 0: no change ... 3: significant change

most changed target words according to SCAN	
`environmental`	supra note law protection id agency impact policy factor
`virtual`	reality virtual computer center experience week community
`disk`	hard disk drive program computer file store ram business
`users`	computer window information software system wireless drive web building available

3. Diachronic Text Evaluation (DTE) (SemEval, 2015)

Task: predict the time frame of origin of a given text snippet

President de Gaulle favors an independent European nuclear striking force [...] (**1962**)

Prediction granularity

fine	2-year intervals	{1700–1702, ..., 1961–1963, ..., 2012–2014}
medium	6-year intervals	{1699–1706, ..., 1959–1965, ..., 2008–2014}
coarse	12-year intervals	{1696–1708, ..., 1956–1968, ..., 2008–2020}

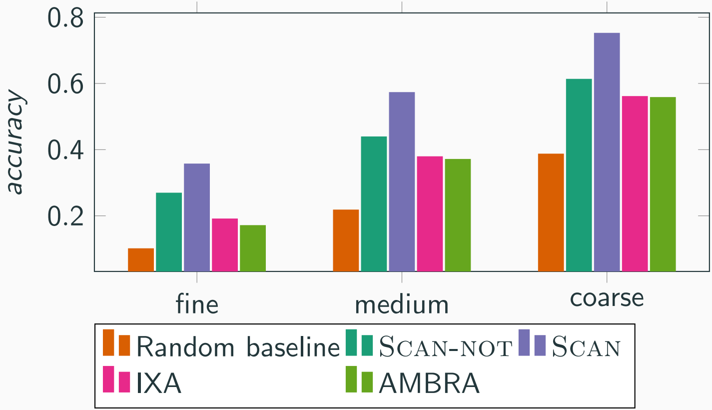
3. Diachronic Text Evaluation (DTE) (SemEval, 2015)

SCAN **temporal word representations**

- 883 nouns and verbs from the DTE development dataset
- $\Delta T = 5$ years
- $K = 8$ senses

$\rightarrow$ predict time of a test snippet using SCAN representations

3. Diachronic Text Evaluation (DTE) (SemEval, 2015)



accuracy: precision measure discounted by distance from true time

Conclusions

A dynamic Bayesian model of diachronic meaning change

- ✓ sense-level meaning change
- ✓ arbitrary time spans and intervals
- ✓ knowledge lean
- ✓ explicit model of smooth temporal dynamics

Conclusions

A dynamic Bayesian model of diachronic meaning change

- ✓ sense-level meaning change
- ✓ arbitrary time spans and intervals
- ✓ knowledge lean
- ✓ explicit model of smooth temporal dynamics

Our work opens up avenues for a variety of applications

- aiding historical text mining or QA
- building and updating onthologies
- modeling short term opinion change from twitter data

Thank you!

lea@frermann.de
www.frermann.de

References

David M. Blei and John D. Lafferty. 2006a. Correlated Topic Models. In *Advances in Neural Information Processing Systems 18*, pages 147–154. Vancouver, BC, Canada.

David M. Blei and John D. Lafferty. 2006b. Dynamic Topic Models. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 113–120. Pittsburgh, PA, USA.

References II

Paul Cook, Jey Han Lau, Diana McCarthy, and Timothy Baldwin. 2014. Novel Word-sense Identification. In *Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1624–1635. Dublin, Ireland. URL http://www.aclweb.org/anthology/C14-1154.

Mark Davies. 2010. The Corpus of Historical American English: 400 million words, 1810-2009. Available online at http://corpus.byu.edu/coha/.

Hans-Jürgen Diller, Hendrik de Smet, and Jukka Tyrkkö. 2011. A European database of descriptors of english electronic texts. *The European English messenger*, 19(2):29–35.

Pieter C. N. Groenewald and Lucky Mokgatlhe. 2005. Bayesian Computation for Logistic Regression. *Computational Statistics & Data Analysis*, 48(4):857–868.

References III

Kristina Gulordava and Marco Baroni. 2011. A Distributional Similarity Approach to the Detection of Semantic Change in the Google Books Ngram Corpus. In *Proceedings of the Workshop on GEometrical Models of Natural Language Semantics*, pages 67–71. Edinburgh, Scotland.

Adam Kilgarriff. 2009. Simple maths for keywords. In *Proceedings of the Corpus Linguistics Conference*.

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal Analysis of Language through Neural Language Models. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 61–65. Baltimore, MD, USA.

References IV

Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. Statistically Significant Detection of Linguistic Change. In *Proceedings of the 24th International Conference on World Wide Web*, pages 625–635. Geneva, Switzerland.

Han Jey Lau, Paul Cook, Diana McCarthy, Spandana Gella, and Timothy Baldwin. 2014. Learning Word Sense Distributions, Detecting Unattested Senses and Identifying Novel Senses using Topic Models. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 259–270. Baltimore, MD, USA.

Jey Han Lau, Paul Cook, Diana McCarthy, David Newman, and Timothy Baldwin. 2012. Word Sense Induction for Novel Sense Detection. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 591–601. Avignon, France.

David Mimno, Hanna Wallach, and Andrew McCallum. 2008. Gibbs Sampling for Logistic Normal Topic Models with Graph-Based Priors. In *NIPS Workshop on Analyzing Graphs*. Vancouver, Canada. URL `http://www.cs.umass.edu/~mimno/papers/sampledlgstnorm.pdf`.

Sunny Mitra, Ritwik Mitra, Suman Kalyan Maity, Martin Riedl, Chris Biemann, Pawan Goyal, and Animesh Mukherjee. 2015. An automatic approach to identify word sense changes in text media across timescales. *Natural Language Engineering*, 21:773–798.

References VI

Sunny Mitra, Ritwik Mitra, Martin Riedl, Chris Biemann, Animesh Mukherjee, and Pawan Goyal. 2014. That's sick dude!: Automatic identification of word sense change across different timescales. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 1020–1029. Baltimore, MD, USA.

Octavian Popescu and Carlo Strapparava. 2013. Behind the Times: Detecting Epoch Changes using Large Corpora. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 347–355. Nagoya, Japan.

Octavian Popescu and Carlo Strapparava. 2015. SemEval 2015, Task 7: Diachronic Text Evaluation. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 869–877. Denver, CO, USA. URL http://www.aclweb.org/anthology/S15-2147.