Robust Morphological Tagging with Word Representations

Thomas Müller and Hinrich Schütze

Center for Information and Language Processing, University of Munich

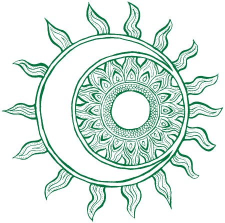
- Gender prediction:

  La ignorancia es la noche de la mente: pero una noche sin luna y sin estrellas.

  "Ignorance is the night of the mind, but a night without moon and star." [Confucius]

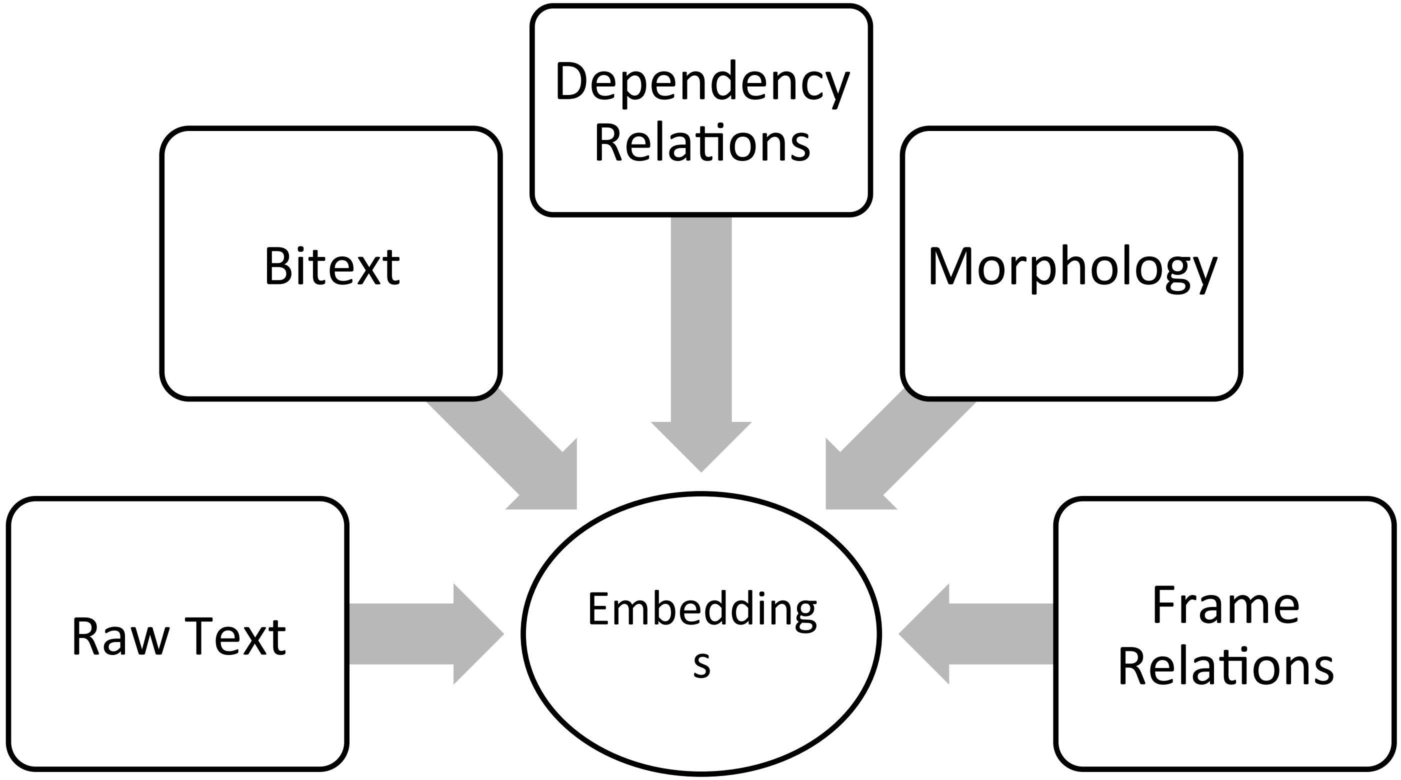
- Frequent contexts:

3373	**la**	luna
600	luna	llen**a**
487	medi**a**	luna
285	un**a**	luna



Multiview LSA

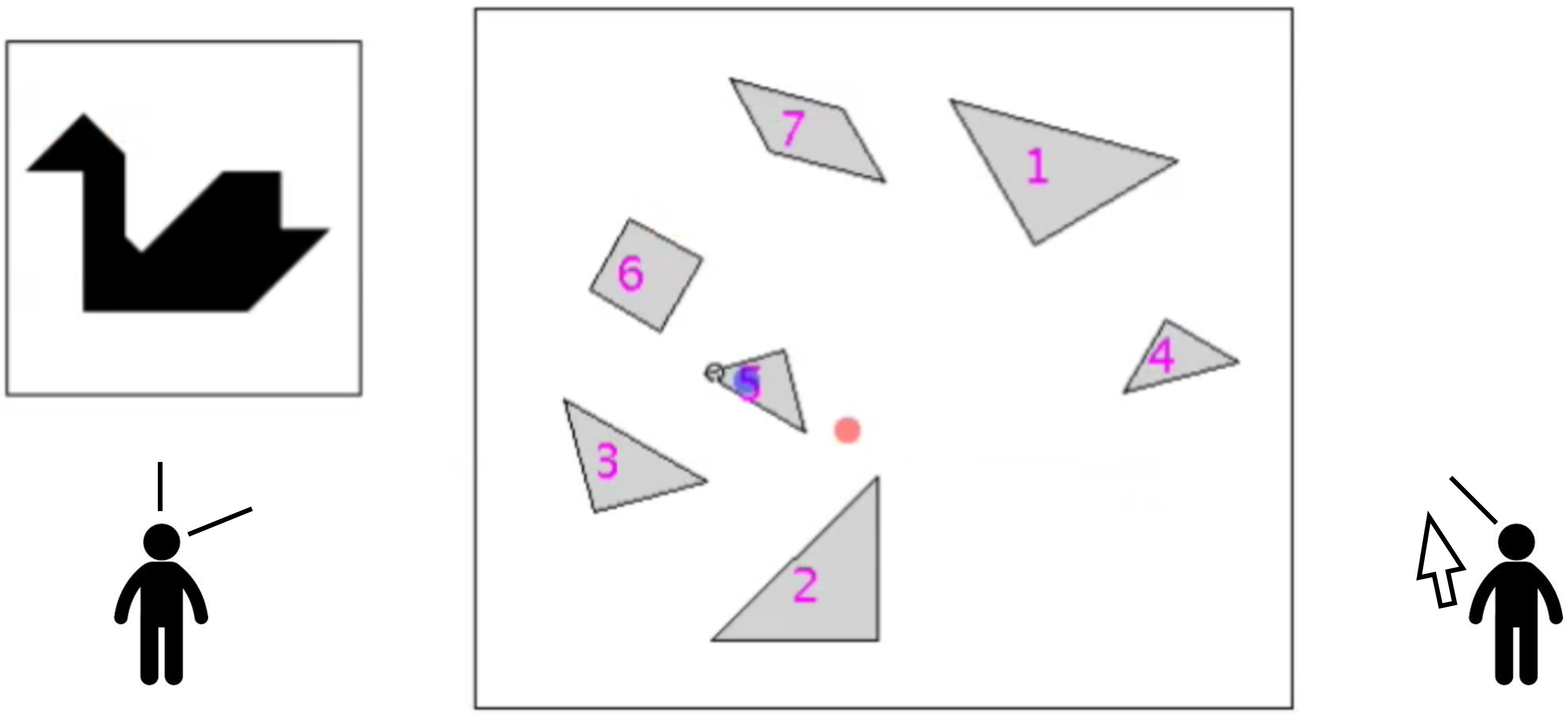
Pushpendre Rastogi, Benjamin Van Durme, Raman Arora
Center for Language and Speech Processing, JHU



- Let's stew some embeddings. *Better than Word2Vec, Glove, Retrofitting**
- Gather lots of co-occurrence counts, other embeddings.
- Add a generalization of PCA/CCA called GCCA.
- Cook using Incremental PCA.
- Season with regularization to handle sparsity.
- Test on 13 test sets to make sure the embeddings are ready to serve.

125: Incrementally Tracking Reference in Human/Human Dialogue Using Linguistic and Extra-Linguistic Information

Casey Kennington, Ryu Iida, Takenobu Tokunaga, David Schlangen*



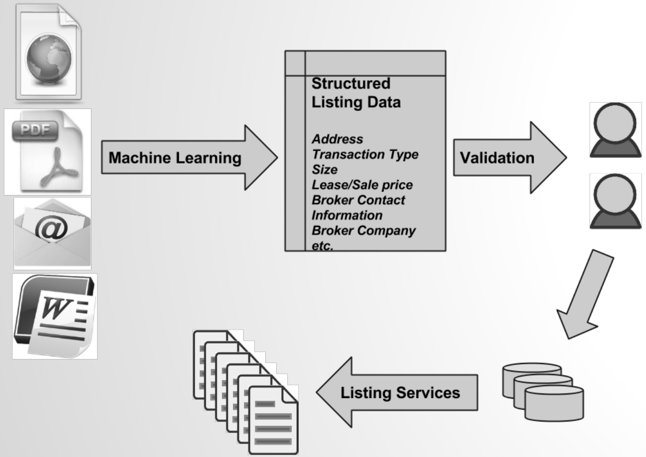
そのちっちゃい三角形を左に置いて。右に回転して。

Put [that] [little triangle] on the left. Rotate [it] right.

Digital Leafleting: Extracting Structured Data from Multimedia Online Flyers

Emilia Apostolova & Jeffrey Sack & Payam Pourashraf
emilia@brokersavant.com jeff@brokersavant.com ppourash@cdm.depaul.edu

BrokerSavant



Business Objective: develop an automated approach to the task of identifying listing information from commercial real estate flyers.

An example of a commercial real estate flyer © Kudan Group Real Estate.

- Information in visually rich formats such as PDF and HTML is often conveyed by a combination of textual and visual features.
- Genres such as marketing flyers and info-graphics often augment textual information by its color, size, positioning, etc.
- Traditional text-based approaches to information extraction (IE) could underperform.

Find us in poster session 1A-169

UNIVERSITY OF CAMBRIDGE

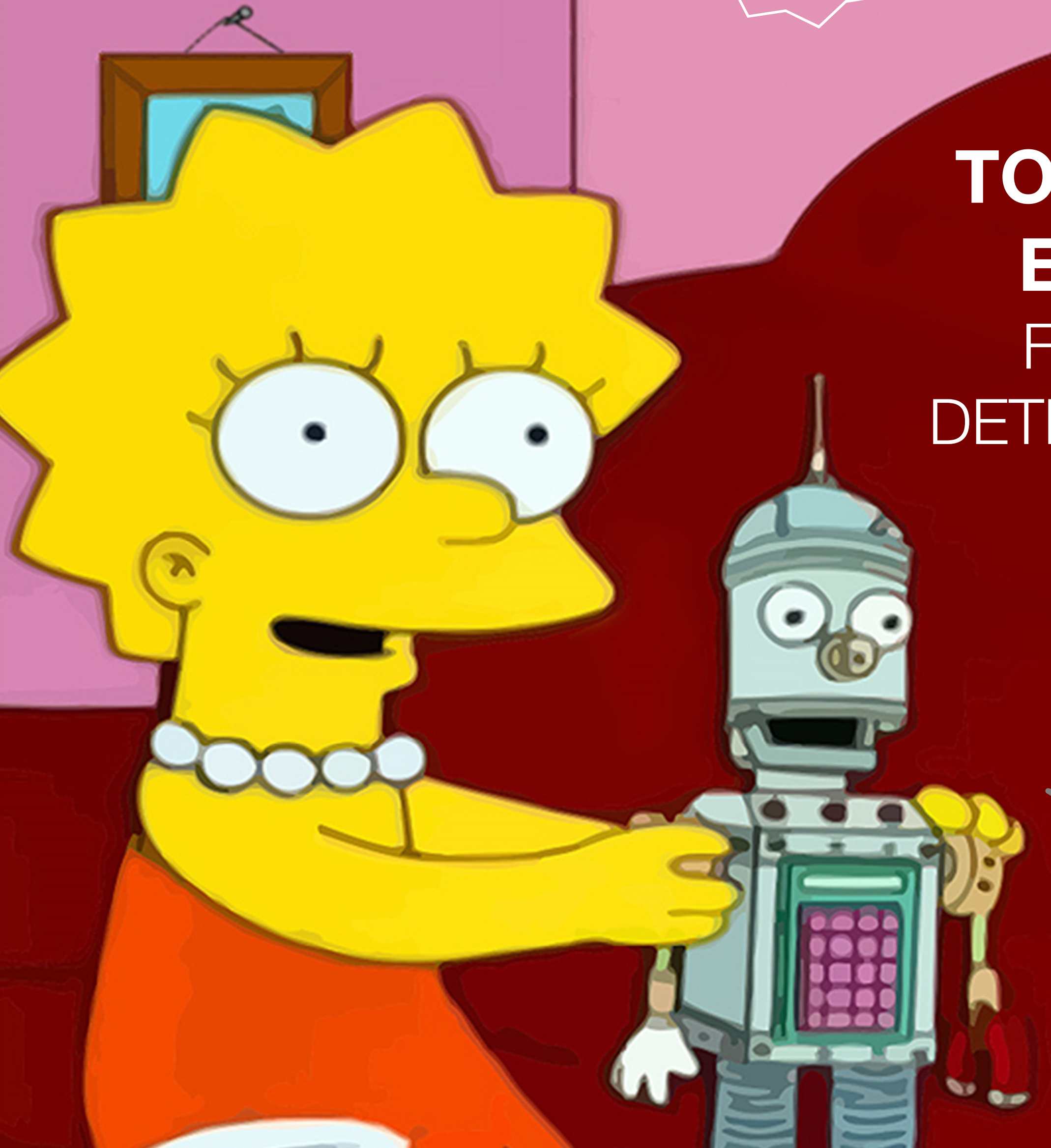
ALTA Institute Computer Laboratory

TOWARDS A STANDARD EVALUATION METHOD

FOR GRAMMATICAL ERROR DETECTION AND CORRECTION

Mariano Felice
mf501@cl.cam.ac.uk

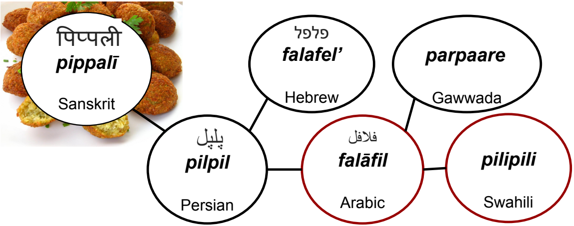
Ted Briscoe
ejb@cl.cam.ac.uk



Constraint-Based Models of Lexical Borrowing

Yulia Tsvetkov Waleed Ammar Chris Dyer

Carnegie Mellon University

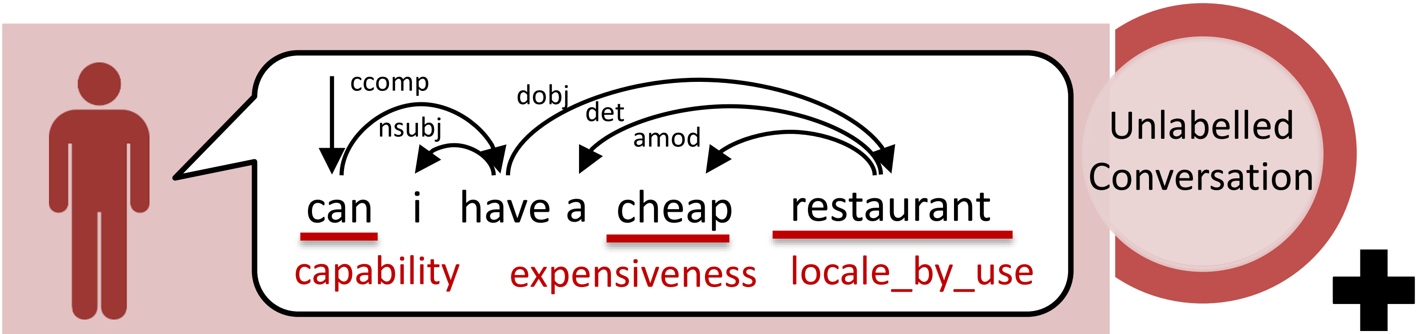


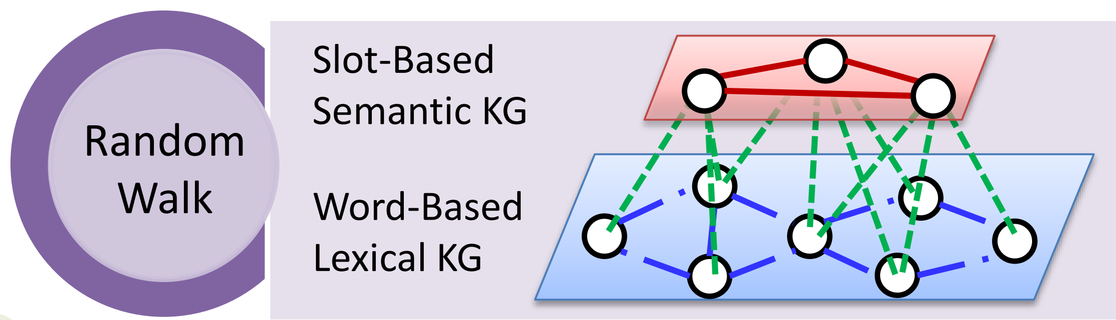
- ✓ Cross-lingual model of lexical borrowing
- ✓ Linguistically informed, with Optimality-Theoretic features
- ✓ Good performance with only a few dozen training examples

Jointly Modeling Inter-Slot Relations by Random Walk on Knowledge Graphs for Unsupervised Spoken Language Understanding

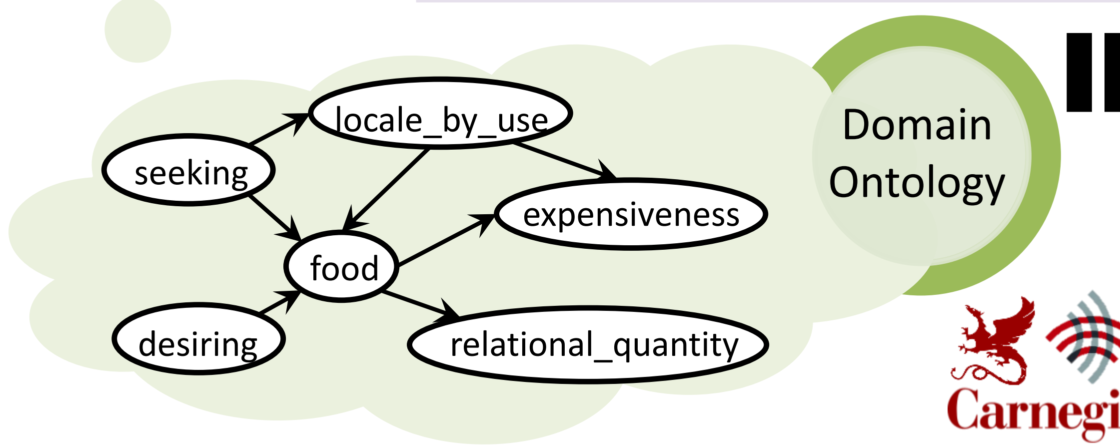
Yun-Nung (Vivian) Chen, William Yang Wang, and Alexander I. Rudnicky

Can a dialogue system automatically learn open domain knowledge?





Spoken Dialogue System / Intelligent Assistant



Carnegie Mellon

Expanding Paraphrase Lexicons by Exploiting Lexical Variants

Atsushi FUJITA (NICT, Japan)
Pierre ISABELLE (NRC, Canada)

airports in Europe ⇔ European airports
amendment of regulation ⇔ amending regulation
should be noted that ⇔ is worth noting that

economy in Uruguay ⇔ Uruguayan economy
recruitment of engineers ⇔ recruiting engineers
should be highlighted that ⇔ is worth highlighting that

<INPUT>
Paraphrase
Lexicon

42x – 206x
in the English
experiments

<OUTPUT>
Expanded
Paraphrase Lexicon
(large & still clean)

- Unsupervised acquisition
  - Paraphrase patterns
  - Lexical correspondences
- Use of monolingual corpus

Lexicon-Free Conversational Speech Recognition with Neural Networks

Andrew Maas*, Ziang Xie*, Dan Jurafsky, & Andrew Ng

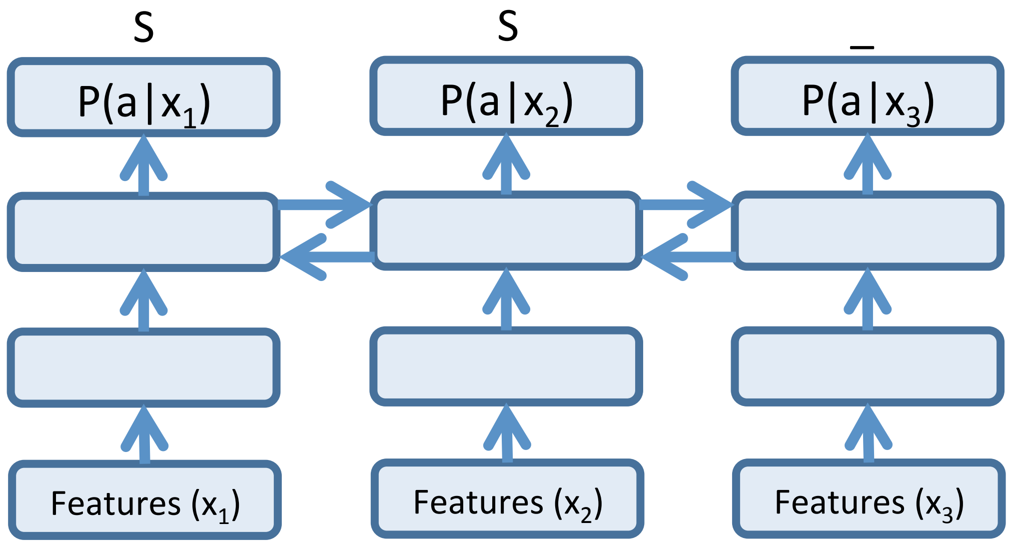
Characters: SPEECH

Character Language Model

$P(a_N | a_1, \ldots, a_{N-1})$

Collapsing function: SS__P_EEE_EE______CCHHH

Acoustic Model:



Audio Input:

RNN estimates $P(a|x)$

distribution over *characters* (a)

Trained with CTC loss function (Graves & Jaitly, 2014)

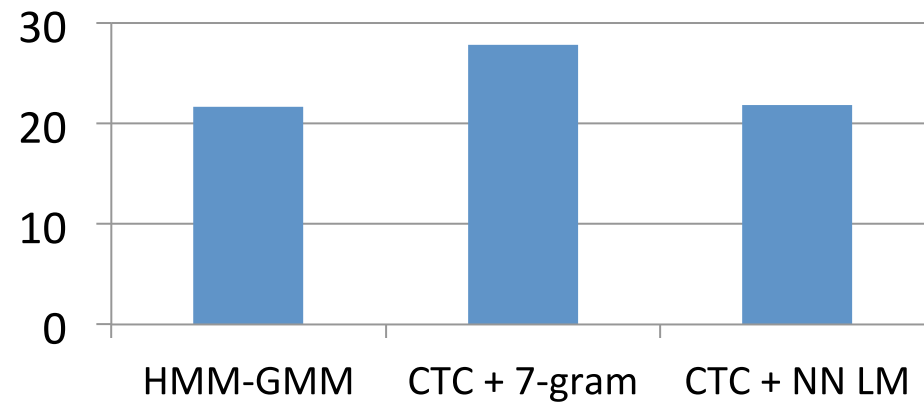
Transcribing Out of Vocabulary Words

Truth: yeah i went into the i do not know what you think of *fidelity* but

HMM-GMM: yeah when the i don't know what you think of **fidel it even them**

CTC-CLM: yeah i went to i don't know what you think of **fidelity but um**

Switchboard WER

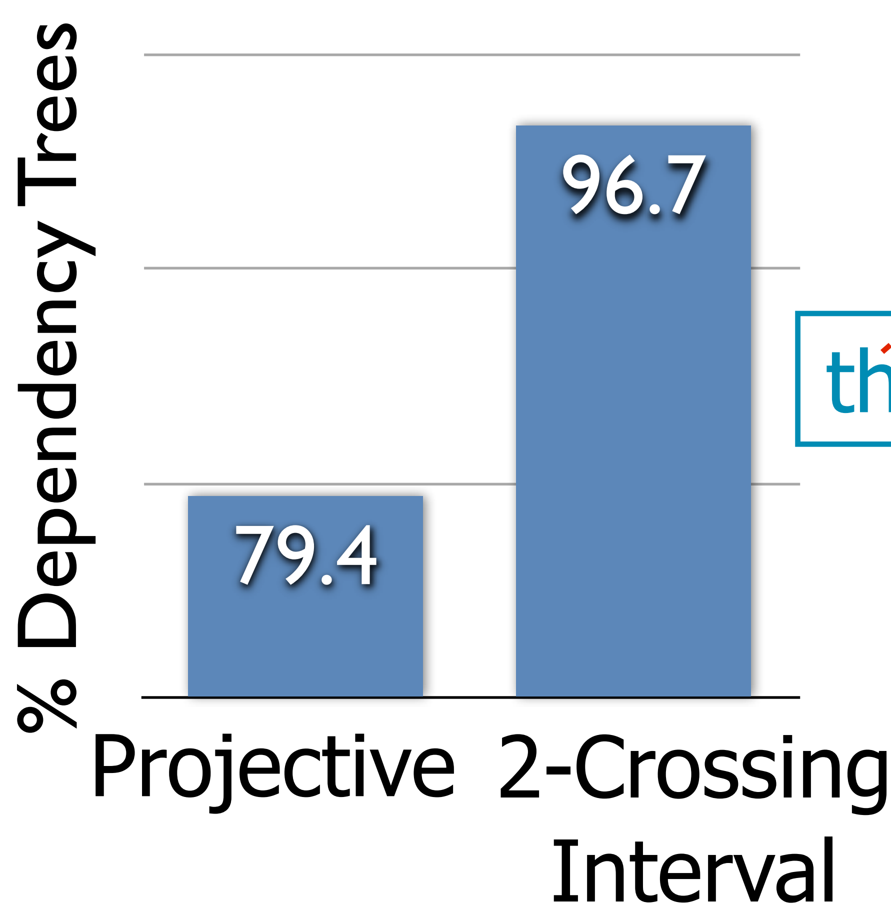


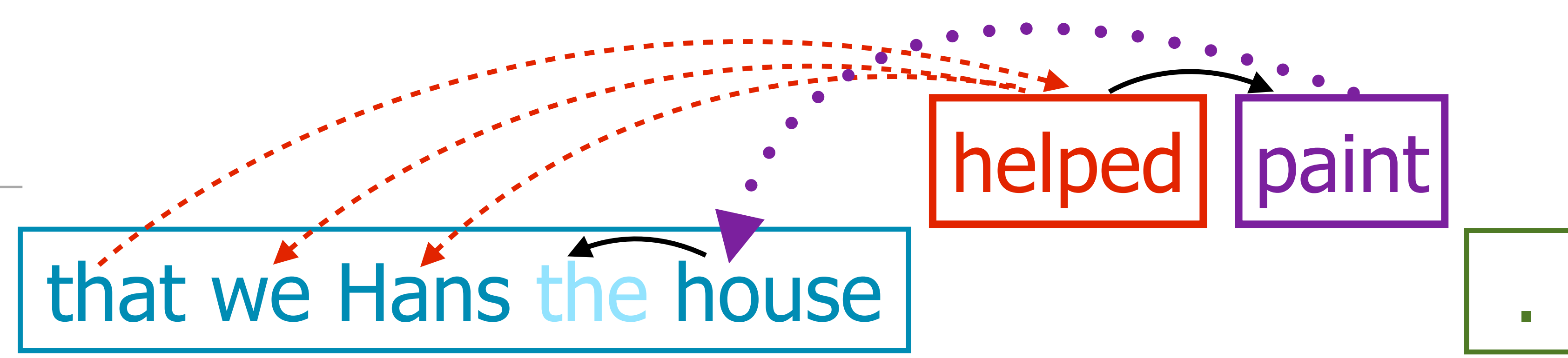
A Linear-Time Transition System for Crossing Interval Trees

Emily Pitler and Ryan McDonald

Existing: Arc-Eager	Stack Buffer	Projective	O(n)
New: Two-Registers	R1 R2 Stack Buffer	2-Crossing Interval	O(n)

Coverage Across 10 Languages



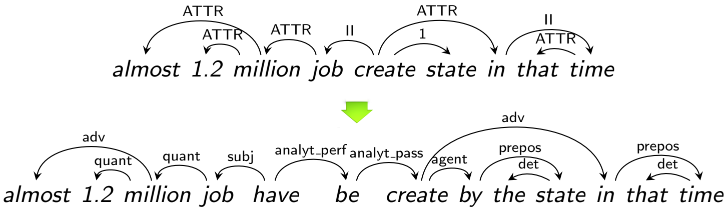


More accurate than arc-eager and swap (all trees, $O(n^2)$)

Data-driven sentence generation with non-isomorphic trees

Miguel Ballesteros **Bernd Bohnet** **Simon Mille** **Leo Wanner**

- Statistical sentence generator that handles the non-isomorphism between PropBank-like structures and sentences.
- 77 BLEU for English.
- 54 BLEU for Spanish.

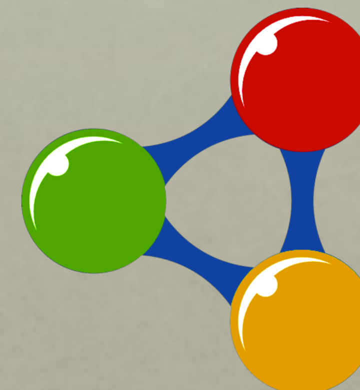


Ask us for English and Spanish Deep-Syntactic corpora!

Ontologically Grounded Multi-sense Representation Learning

Sujay Kumar Jauhar, Chris Dyer & Eduard Hovy

bank



Words have more than one meaning

Problematic when learning word vectors

Our secret sauce: ontologies

Fast

Effective

Flexible

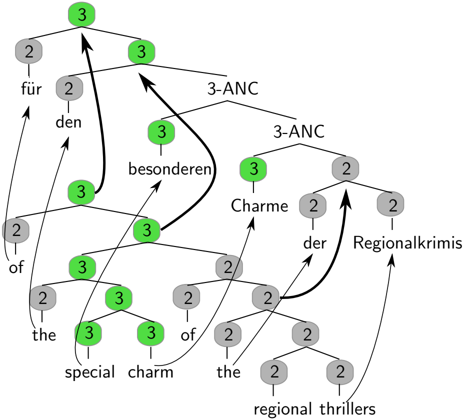
Interpretable

Subsentential Sentiment on a Shoestring

Michael Haas   Yannick Versley

Heidelberg University

- Can we use **English data** to bootstrap **compositional sentiment classification** in **another language**? (yes!)
- Are fancy **Recursive Neural Tensor Network** models always the best solution? (no!)
- Can we make them better suited for **sparse data** cases? (maybe!)



Using Summarization to Discover Argument Facets in Online Idealogical Dialog

Amita Misra, Pranav Anand, Jean Fox Tree, and Marilyn Walker

What are the arguments that are repeated across many dialogues on a topic?

Two Steps:

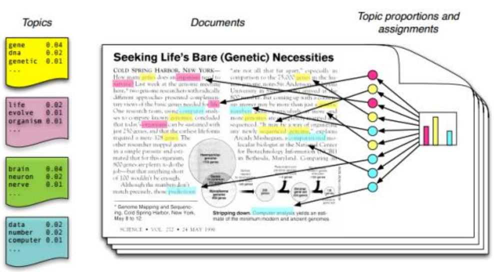
- Can we find them?
- Can we recognize when two arguments are paraphrases of each other?

Row	Feature Set	R	MAE	RMS
1	NGRAM (N)	0.39	0.90	1.09
2	UMBC (U)	0.46	0.86	1.06
3	LIWC (L)	0.32	0.92	1.13
4	DISCO (D)	0.33	0.93	1.12
5	ROUGE (R)	0.34	0.91	1.12
6	N-U	0.47	0.85	1.05
7	N-L	0.45	0.86	1.06
8	N-R	0.42	0.88	1.08
9	N-D	0.41	0.89	1.08
10	U-R	0.48	0.84	1.04
11	U-L	0.51	0.83	1.02
12	U-D	0.45	0.86	1.06
13	N-L-R	0.48	0.84	1.04
14	U-L-R	0.53	0.81	1.00
15	N-L-R-D	0.50	0.83	1.03
16	N-L-R-U	0.54	0.80	1.00
17	N-L-R-D-U	0.54	0.80	1.00

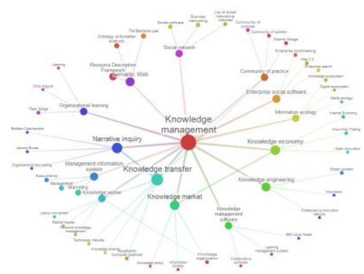
Incorporating Word Correlation Knowledge into Topic Modeling

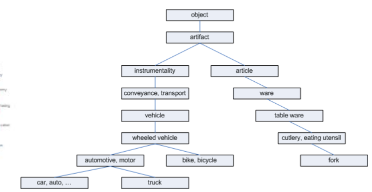
Pengtao Xie, Diyi Yang and Eric Xing
Carnegie Mellon University

Topic Modeling

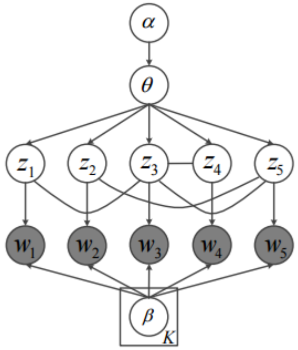


Word Correlation Knowledge





MRF-LDA



Greatly Improve Topic Coherence

Method	A1	A2	A3	A4	Mean	Std
LDA	30	33	22	29	28.5	4.7
DF-LDA	35	41	35	27	36.8	2.9
Quad-LDA	32	36	33	26	31.8	4.2
MRF-LDA	**60**	**60**	**63**	**60**	**60.8**	**1.5**

ILLINOIS INSTITUTE OF TECHNOLOGY

Manali Sharma, Di Zhuang, and Mustafa Bilgic

ACTIVE LEARNING WITH RATIONALES FOR TEXT CLASSIFICATION

Expert

Superhero or villain?

What is the RATIONALE?

Label: Superhero

Rationale: Saves the New York City

Active Learner

Question: How to incorporate rationales to speed-up the learning?

Answer: We provide a simple approach to incorporate rationales into training of any off-the-shelf classifier

The Unreasonable Effectiveness of Word Representations for Twitter Named Entity Recognition

Colin Cherry and Hongyu Guo

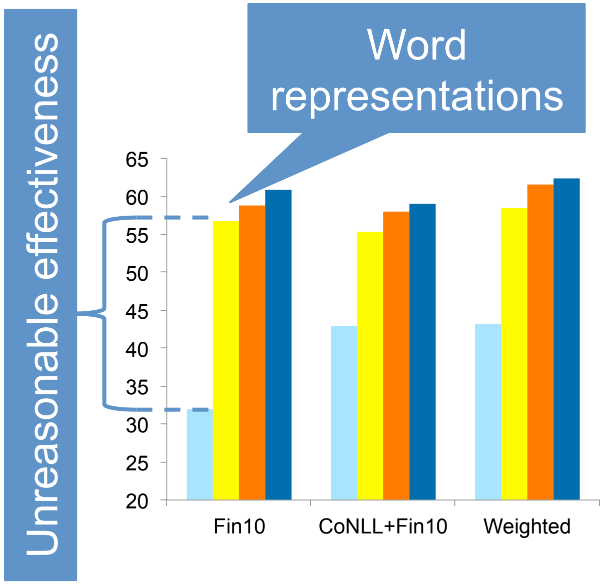
ORG | PER | ORG

Ducks sign LW Beleskey to 2-year extension - San Jose Mercury News http://dlvr.it/5RcvP #ANADucks

A state-of-the-art system for Twitter NER, using just 1,000 annotated tweets

We analyze the impact of:

- Brown clusters and word2vec
- in- and out-of domain training data
- data weighting
- POS tags and gazetteers



Inferring Temporally-Anchored Spatial Knowledge from Semantic Roles

Eduardo Blanco and Alakananda Vempala

- Semantic roles tells you who did what to whom, how, when and where
  - *Today, FBI agents and divers were collecting evidence at Lake Logan*
    - Who? *FBI agents and divers*
    - What? *evidence*
    - When? *Today*
    - Where? *at Lake Logan*
- Given the above semantic roles …
  - Can we infer whether
    - *FBI agents and divers* have LOCATION *Lake Logan*?
    - *evidence* has LOCATION *Lake Logan*?
  - Can we temporally-anchor the LOCATIONs?
    - before *collecting?*
    - during *collecting?*
    - after *collecting?*

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
College of Engineering
UNT

Is Your Anchor Going Up or Down? Fast and Accurate Supervised Topic Models

Thang Nguyen, Jordan Boyd-Graber, Jeff Lund, Kevin Seppi, and Eric Ringger
daithang@umiacs.umd.edu, Jordan.Boyd.Graber@colorado.edu, {jefflund, kseppi}@byu.edu, ringger@cs.byu.edu
University of Maryland, College Park / University of Colorado, Boulder / Brigham Young University

Motivation

- Supervised topic models leverage latent document-level themes to capture nuanced sentiment, create sentiment-specific topics and improve sentiment prediction.
- *The downside for Supervised LDA is that it is slow, which this work addresses.*

Contribution

- We create a supervised version of the anchor word algorithm (**ANCHOR**) (Arora et al., 2013).

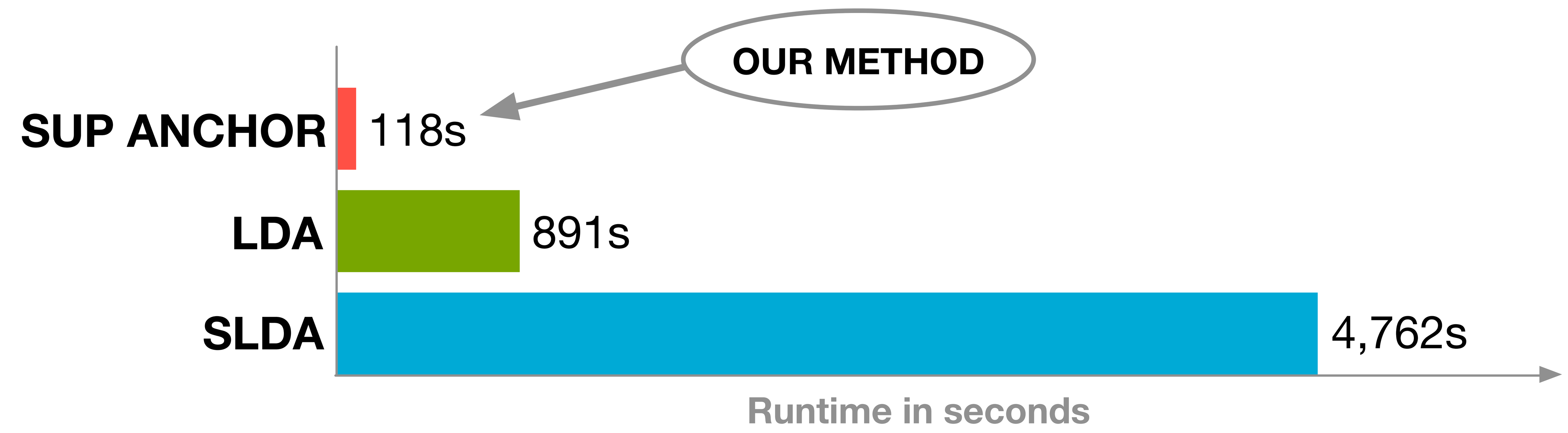
$$\bar{Q} \equiv \begin{bmatrix} p(w_1|w_1) \dots & & \\ & \vdots & \\ & & p(w_j|w_i) \end{bmatrix}$$

$$S \equiv \begin{bmatrix} p(w_1|w_1) \dots & & & p(y^{(l)}|w_1) \\ & \vdots & & \vdots \\ & & p(w_j|w_i) & p(y^{(l)}|w_i) \end{bmatrix}$$

New column(s) encoding word-sentiment relationship

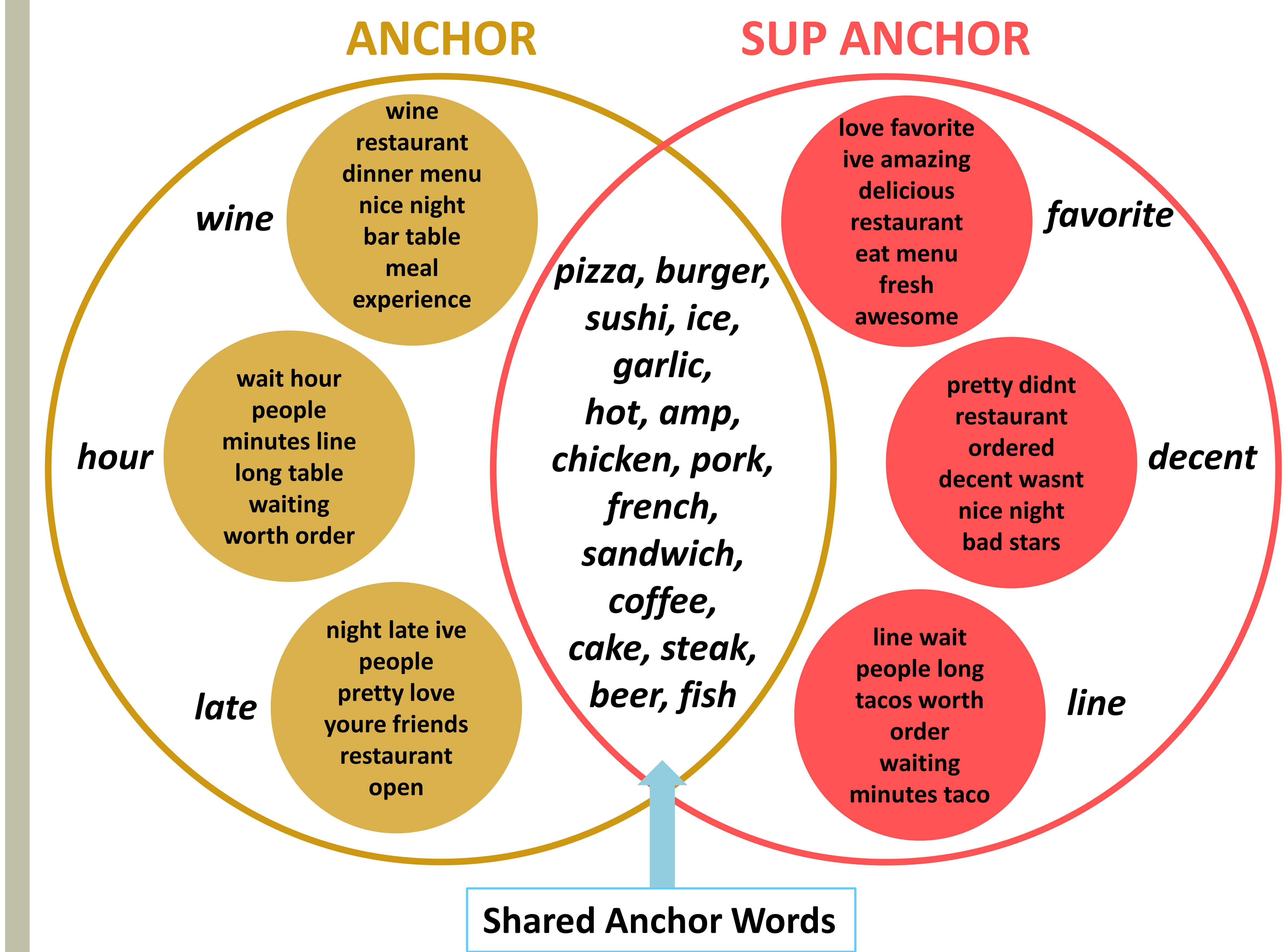
Runtime Analysis

- **SUP ANCHOR** takes much less time than **SLDA**.



Anchor Words and Their Topics

- **SUP ANCHOR** produces anchor words around the same strong lexical cues that could discover better sentiment topics (e.g. positive reviews mentioning a *favorite* restaurant or negative reviews complaining about *long waits*).



Grounded Semantic Parsing for Complex Knowledge Extraction

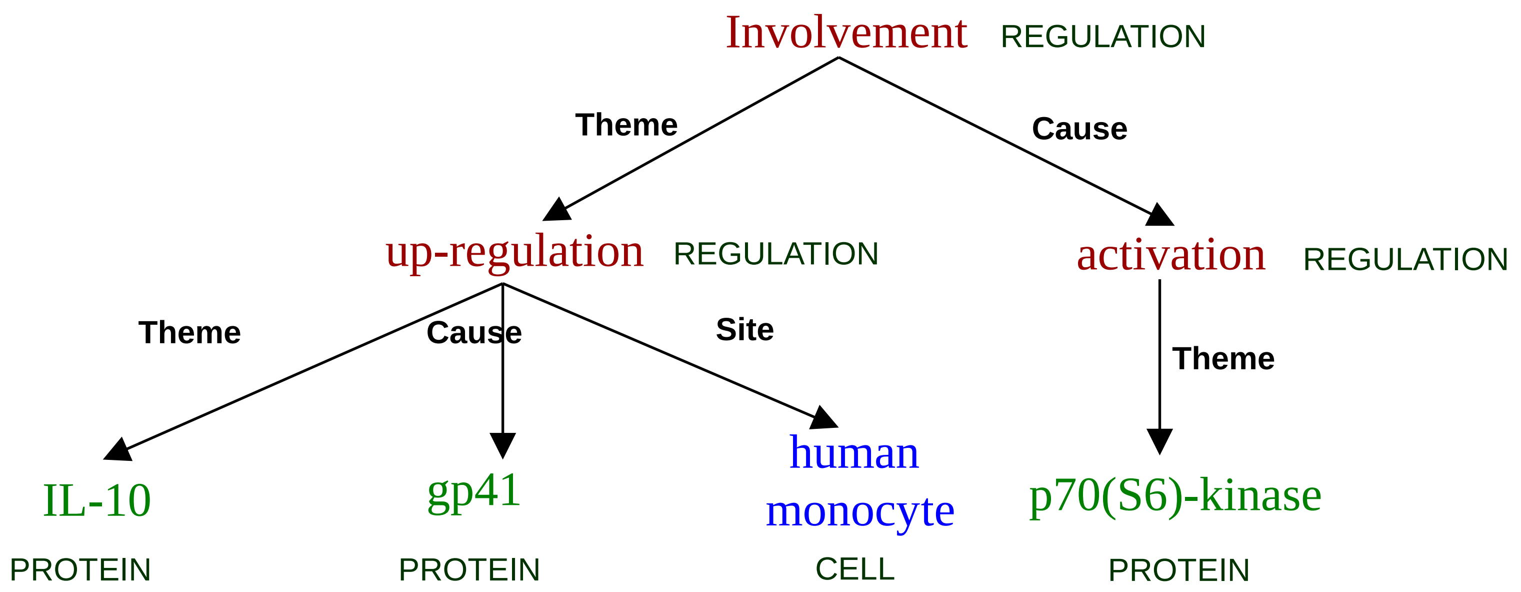
Ankur Parikh, Hoifung Poon, Kristina Toutanova A-557

Generalized distant supervision to extracting nested events

Involvement of p70(S6)-kinase activation in IL-10 up-regulation in human monocytes by gp41 envelope protein of human immunodeficiency virus type 1

Extract complex knowledge from scientific literature

PubMed
2 new papers / minute
1 million / year



Outperformed 19 out of 24 supervised systems in GENIA Shared Task

UT DALLAS

BOSCH

589: Using External Resources and Joint Learning for Bigram Weighting in ILP-Based Multi-Document Summarization

Chen Li, Yang Liu, Lin Zhao

$$\max \sum_i w_i c_i \;\Rightarrow\; \max \sum_i (\theta \cdot f(b_i)) c_i$$

LDC
LINGUISTIC DATA CONSORTIUM

WIKIPEDIA
The Free Encyclopedia

DBpedia

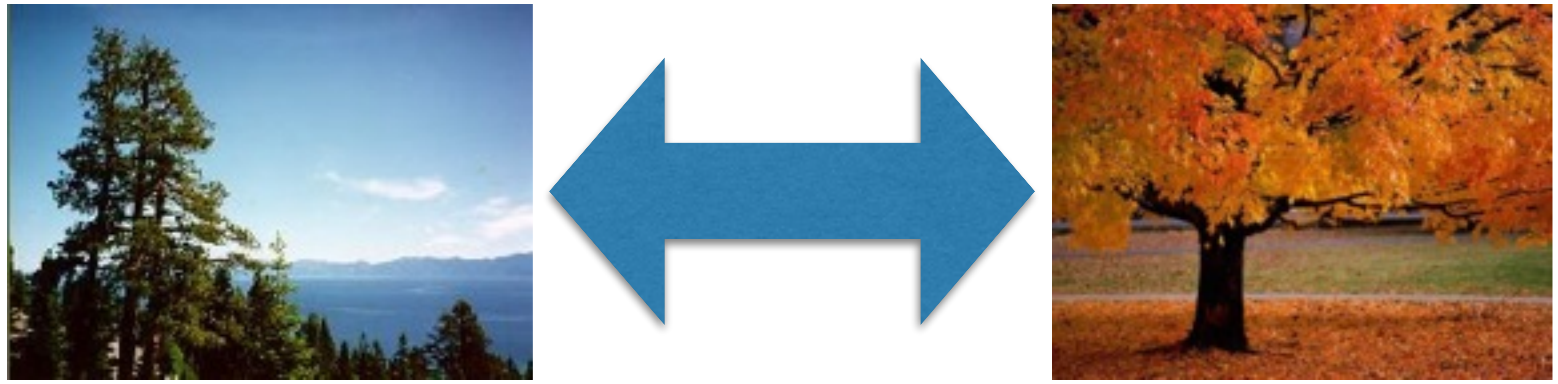
SentiWordNet

WordNet
A lexical database for English

Transforming Dependencies into Phrase Structures

Lingpeng Kong, Alexander M. Rush, Noah A. Smith

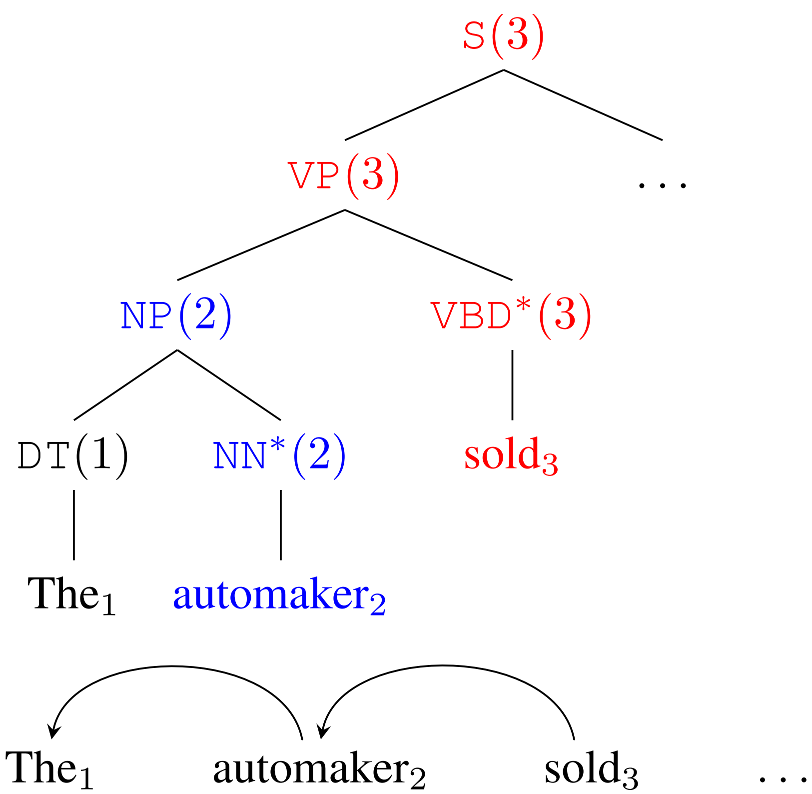
It takes some time to grow so many high-quality phrase structure trees…

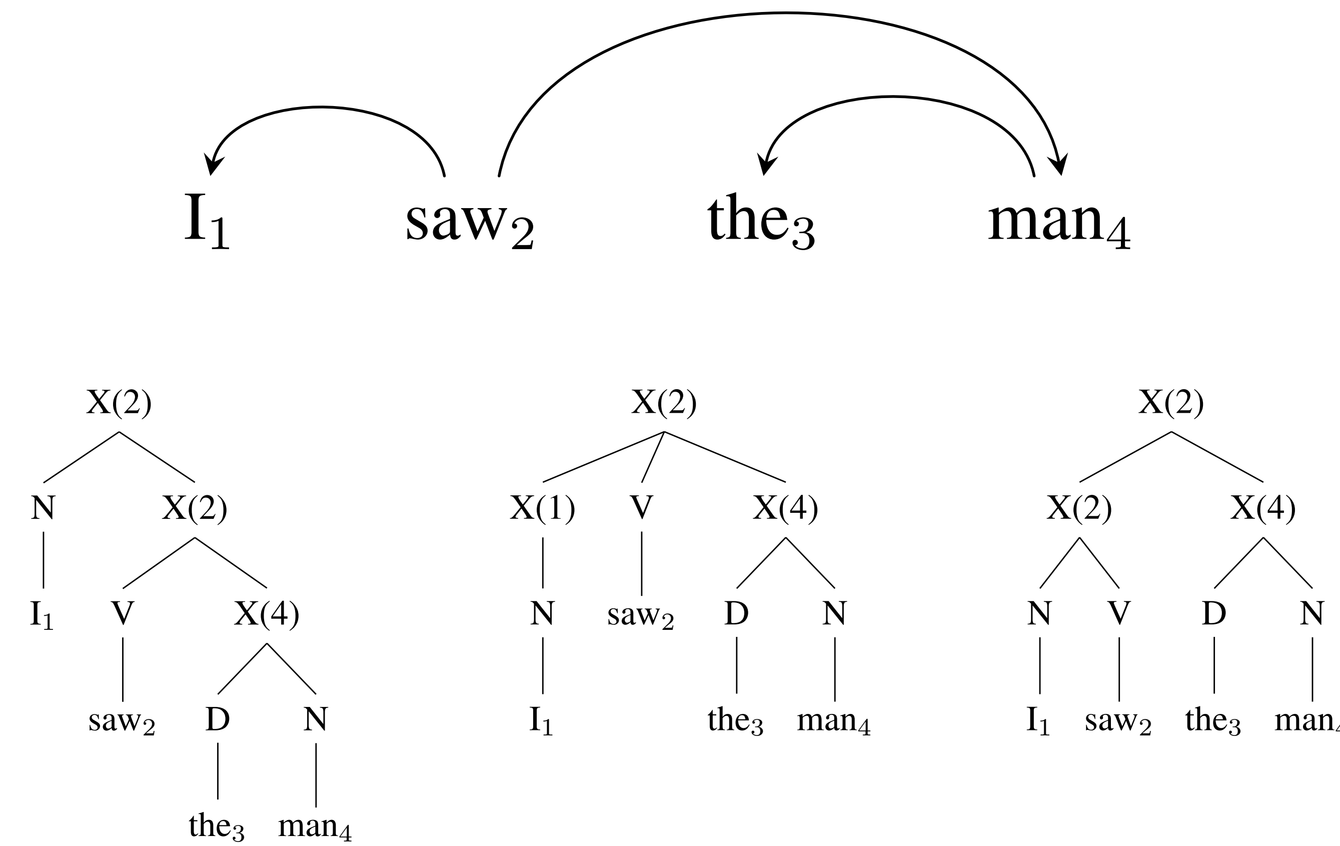


What if we grew these dependency trees first?

Phrase Structure Trees

Dependency Trees





- A linear observable-time structured model that accurately predicts phrase-structure parse trees based on dependency trees!
- Our phrase-structure parser, PAD (Phrase-After-Dependencies) is available as open-source software at — https://github.com/ikekonglp/PAD.

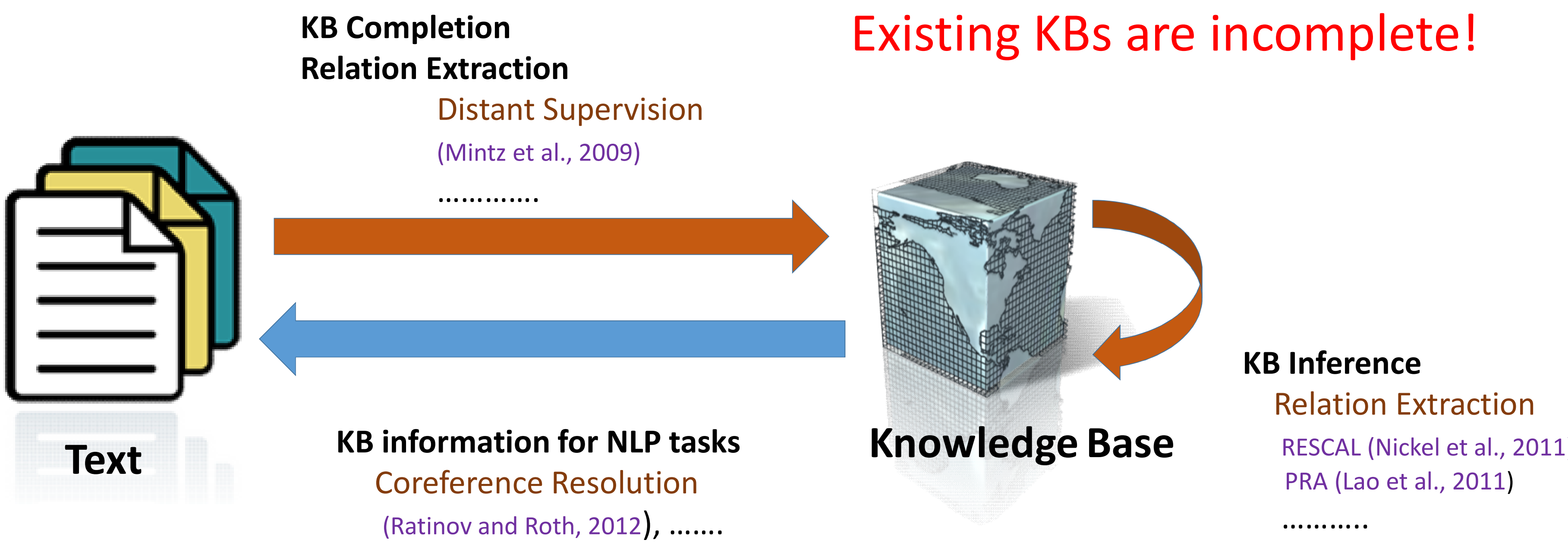
Improving the Inference of **Implicit Discourse Relations** via Classifying Explicit **Discourse Connectives**

Attapol T. Rutherford & Nianwen Xue
Brandeis University

Not All Discourse Connectives are Created Equal.

Because	vs	In sum
Furthermore		Further
Therefore		Nevertheless
In other words		On the other hand
…		…

Inferring Missing Entity Type Instances for Knowledge Base Completion: New Dataset and Methods



Our Work: Text + KB to *infer* missing KB entity type instances

Example: Jim Mahon

Pragmatic Neural Language Modelling for MT

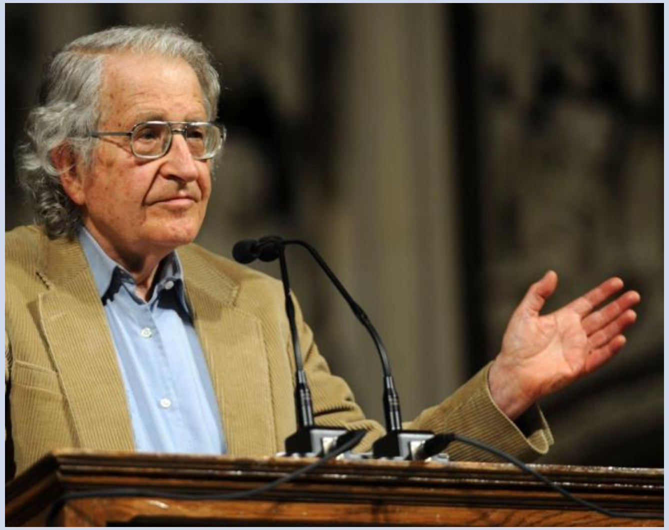
Paul Baltescu and Phil Blunsom, University of Oxford

- Comparison of popular optimization tricks for scaling neural language models in the context of machine translation:
  - Class factorisation, tree hierarchies and ignoring normalisation for speeding up the softmax over the target vocabulary
  - Brown clustering vs. frequency binning for class factorisation
  - Noise contrastive estimation vs. maximum likelihood on very large corpora
  - Diagonal context matrices
- Comparison of neural and back-off n-gram models with and without memory constraints

The Chaos

"Dearest ***creature*** in ***creation***
Studying English pronunciation
I will teach you in my ***verse***
Sounds like corpse, corps, ***horse***, and ***worse***."

Gerard Nolst Trenité



Conventional orthography is a near optimal system for the lexical representation of English words. (Chomsky and Halle, 1968)

English Orthography is **not** "close to optimal"

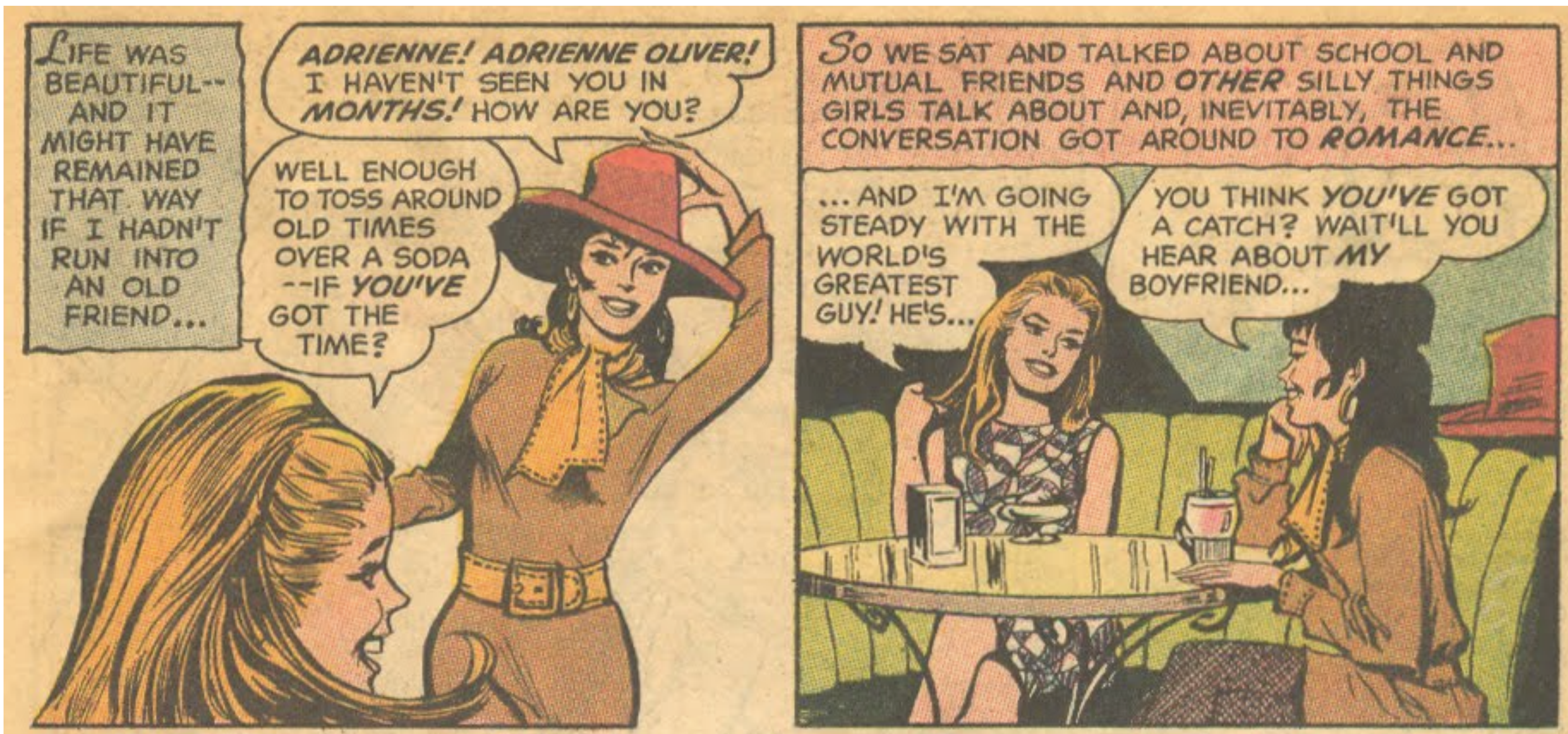
Garrett Nicolai and Greg Kondrak

UNIVERSITY OF ALBERTA

Key Female Characters in Film Have More to Talk About Besides Men: Automating the Bechdel Test

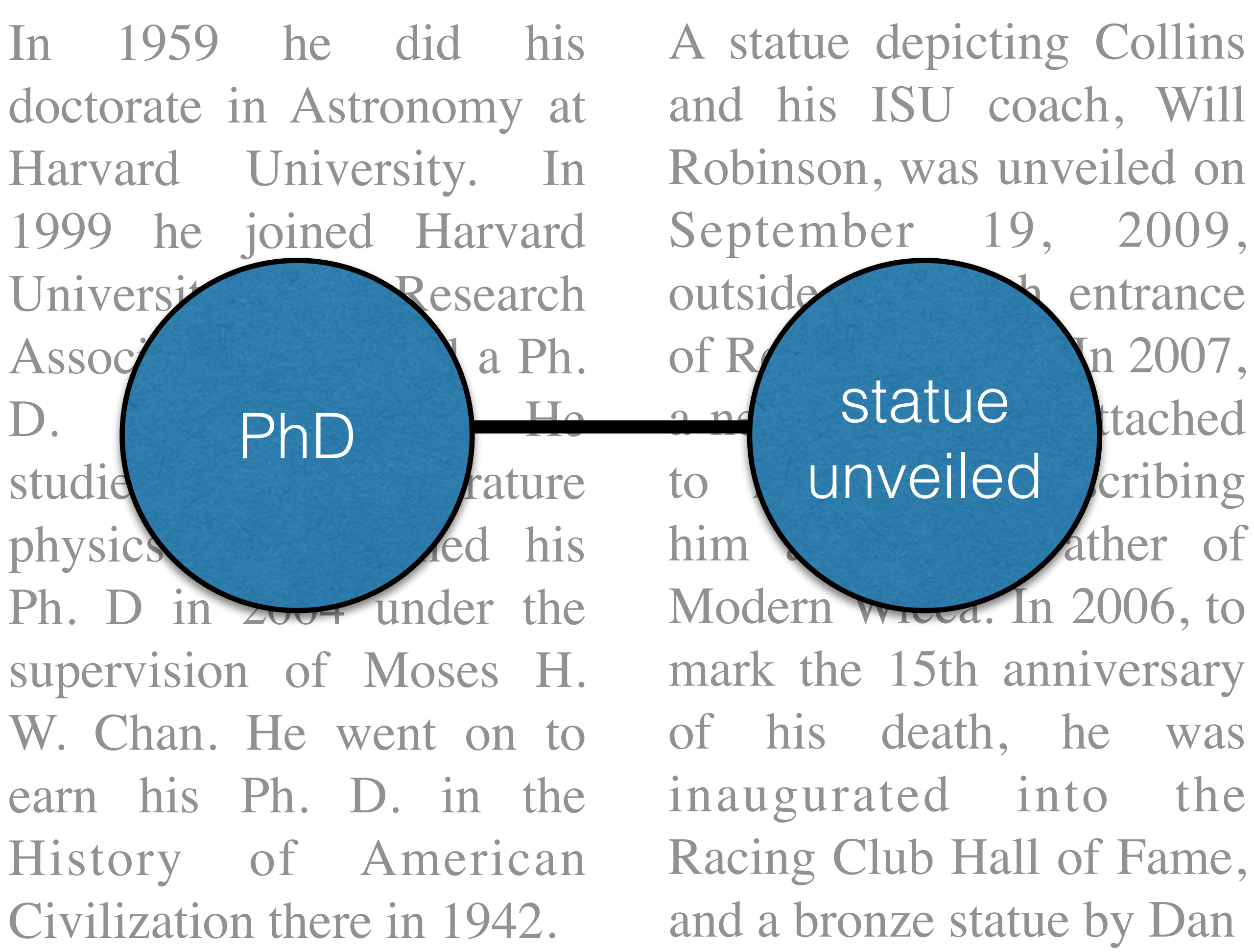
Apoorv Agarwal Jiehan Zheng
Shruti Kamath Sriram Balasubramanian Shirin Dey

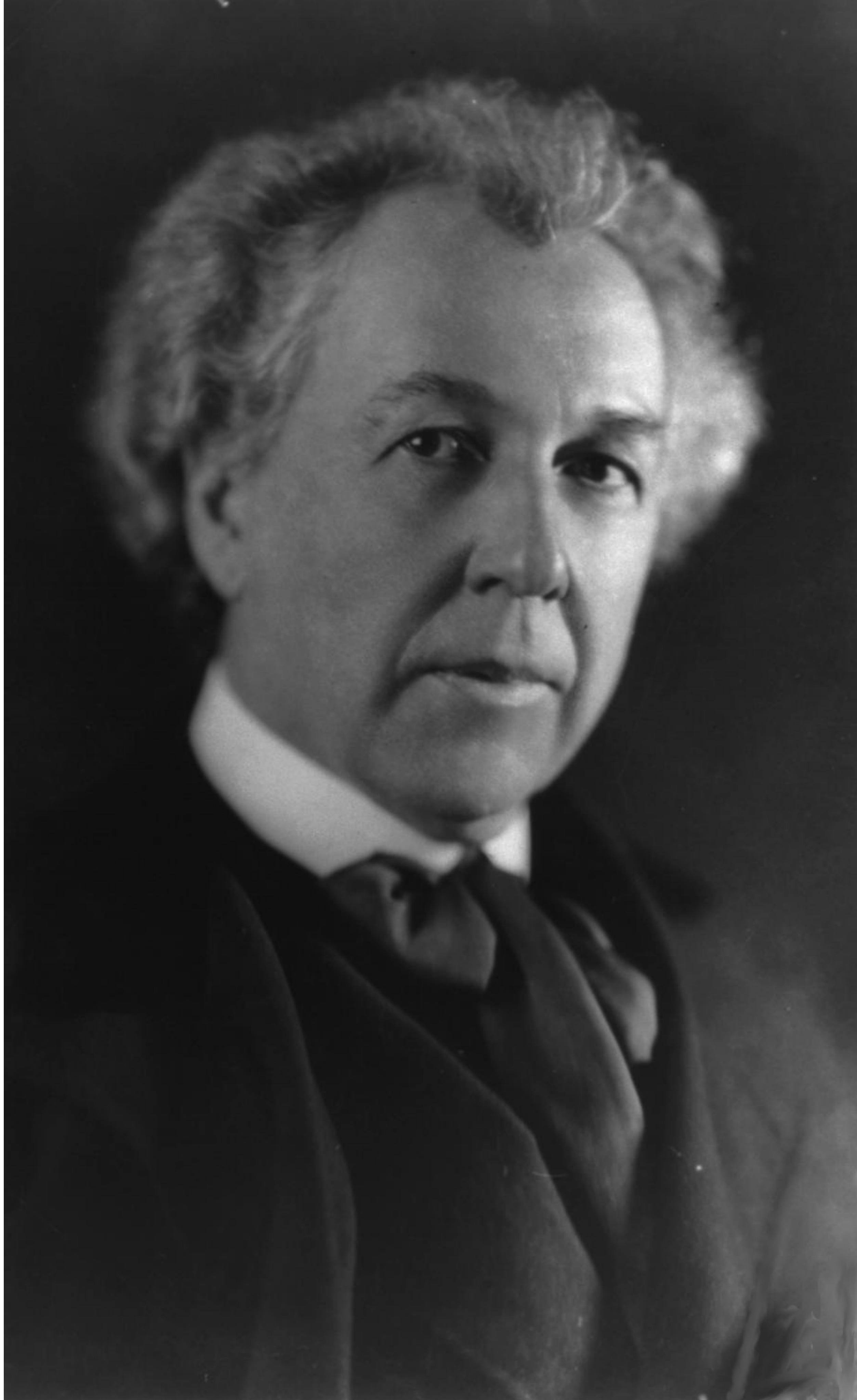
[1] two named women?
[2] do they talk to each other?
[3] talk about something other than a man?



UNSUPERVISED DISCOVERY OF BIOGRAPHICAL STRUCTURE FROM TEXT

DAVID BAMMAN AND NOAH SMITH
CARNEGIE MELLON UNIVERSITY





Locally Non-Linear Learning via Discretization and Structured Regularization

Jonathan Clark, Chris Dyer, Alon Lavie

Slicing fruit wastes time.

Slicing (discretizing) features improves translation quality...

...when combined with structured regularization.

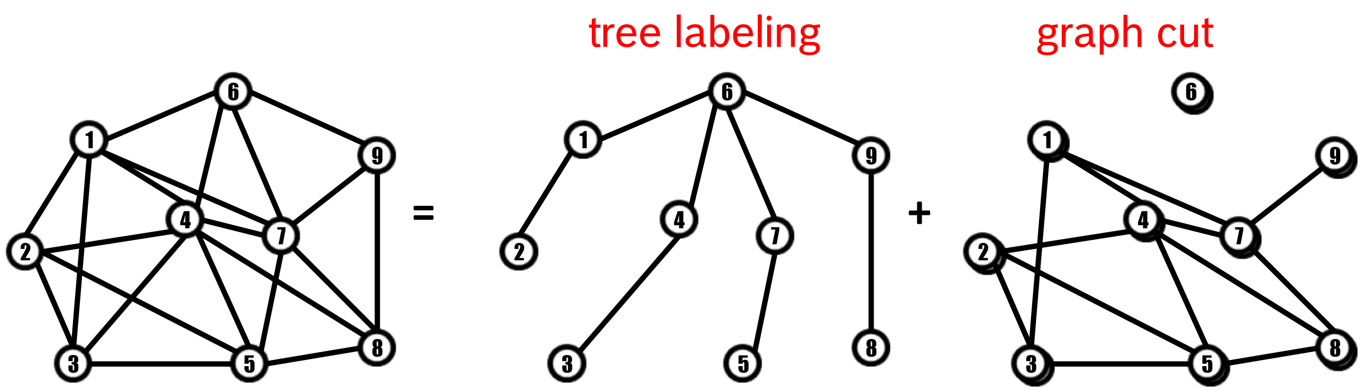
Transform your new features to avoid throwing away good work!

Non-Linearity:
Not just for neural networks.

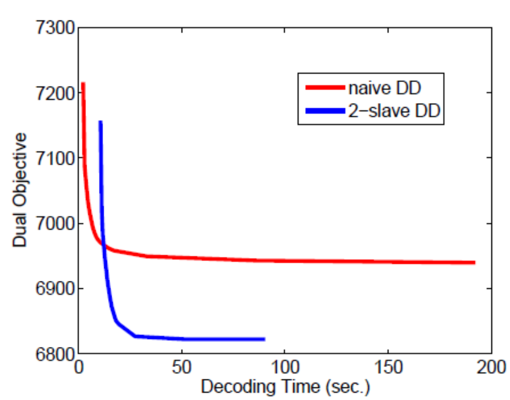
2 Slave Dual Decomposition for Generalized Higher Order CRFs

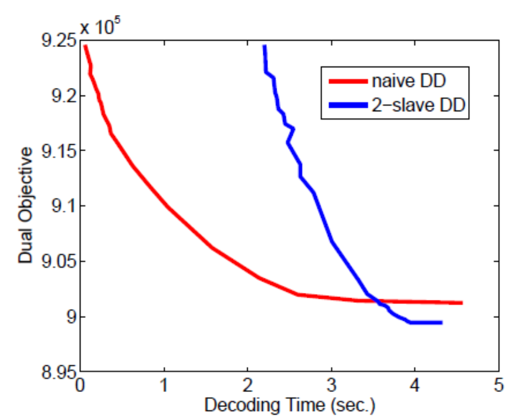
Xian Qian, Yang Liu
The University of Texas at Dallas

Fast dual decomposition based decoding algorithm for general higher order Conditional Random Fields using only **TWO** slaves.



2-slave DD empirically achieves tighter dual objectives than naive DD in less time.





Sprite: Generalizing Topic Models with Structured Priors

Michael J. Paul   Mark Dredze
Johns Hopkins University

LDA

Factorial LDA

Dirichlet Multinomial Regression

Pachinko Allocation

SAGE

Shared Components Topic Models

One model to generalize them all

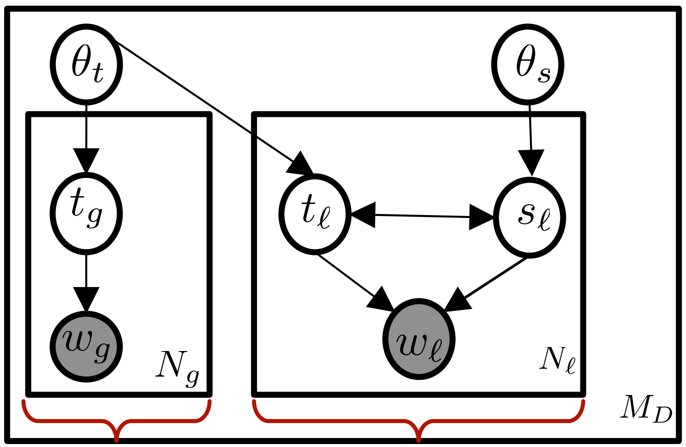
A Sense-Topic Model for Word Sense Induction with Unsupervised Data Enrichment

Jing Wang[1], Mohit Bansal[2], Kevin Gimpel[2], Brian Ziebart[1], Clement Yu[1]

[1]University of Illinois at Chicago

[2]Toyota Technological Institute at Chicago

Sense-Topic Model



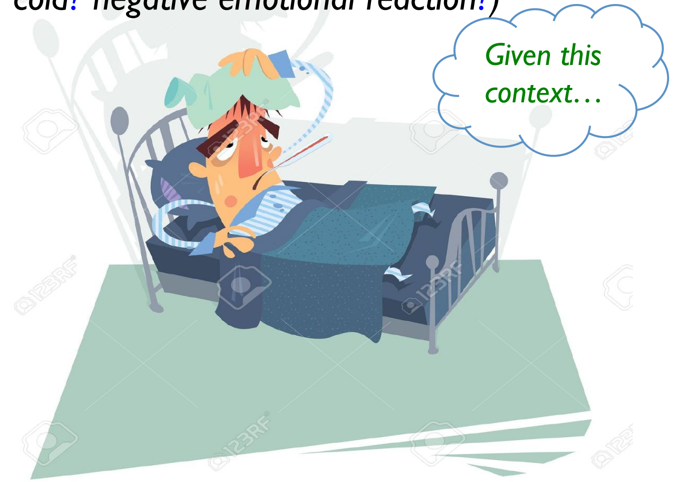
global context words generated from topics

local context words generated from topics and senses

Unsupervised Data Enrichment

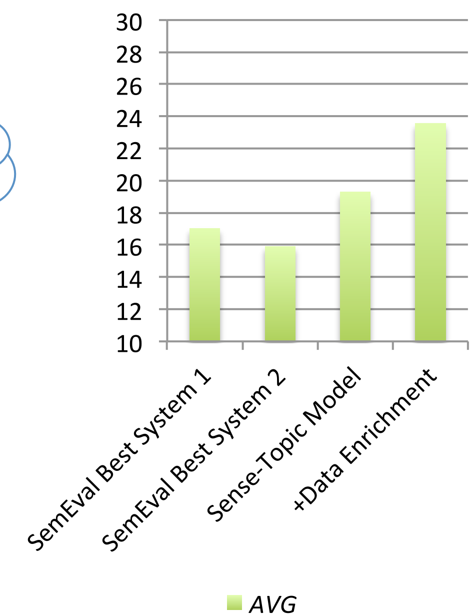
- Adding context

His reaction to the experiment was cold.
(cold temperature? cold sensation? common cold? negative emotional reaction?)



Experiments

SemEval 2013 Dataset



…body temperature… → *His reaction to the experiment was cold* ← *…sore throat…fever…*