B-56: Modeling Word Meaning in Context with Substitute Vectors

Oren Melamud, Ido Dagan, Jacob Goldberger

Bar-Ilan University

What's **_new_** ?

different
next
happening

Up to 50% relative improvement in generating *context-sensitive* lexical substitutes

Our **_new_** results are based on a **_new_** way to model contexts

latest
recent
current

unique
novel
different

Try it yourselves with our software toolkit

Empty Category Detection With Joint Context-Label Embedding

To the MT Community:

Bothered by word alignments?

How can a word aligned with a “NULL”?

Using **Empty Category Detection** to make it easier.

Before		After
吃了吗？	—>	你吃了晚饭了吗？
食べた？	—>	あなたは晩ご飯をたべた？

Did you have supper?

Xun Wang, Katsuhito Sudoh, Masaaki Nagata

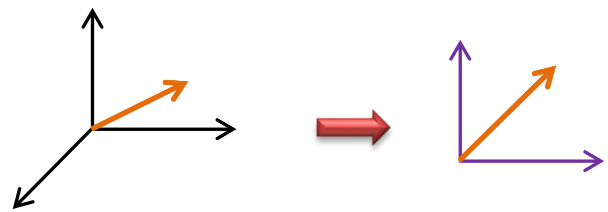
NTT, Japan

NASARI

Unsupervised ✓

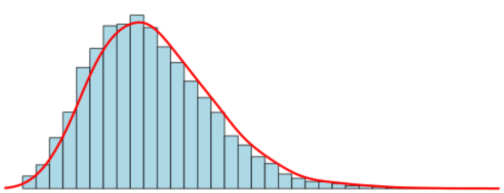
Robust across multiple tasks ✓

a Novel Approach to a Semantically-Aware Representation of Items



Transforming vectors to a semantic space of concepts

➢ *Inherent disambiguation & dimensionality reduction*



Lexical specificity

➢ consistent improvements over *tf-idf*

SoA performance on five datasets in two tasks:

word similarity

sense clustering

SAPIENZA UNIVERSITÀ DI ROMA

José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli

Multi-target Machine Translation with Multi-synchronous Context-free Grammars

Graham Neubig, Kevin Duh, Philip Arthur (NAIST)

Can you pick the correct translation?

اتفاقية كيوتو

京都条约 (a)

京都制裁 (b)

京都议定书 (c)

At the poster, find out more about:

- **Multi-synchronous CFGs**, which can generate strings in multiple languages
- How to **use Spanish to improve English-French translation**

Using Zero-Resource Spoken Term Discovery for Ranked Retrieval

Jerome White, Douglas W. Oard, Jiaul Paik, Rashmi Sankepally, Aren Jansen

Sign constraints on feature weights improve a joint model of word segmentation and phonology

Mark Johnson, Joe Pater, Robert Staubs and Emmanuel Dupoux

- Task: jointly *segment utterances into words* and *learn phonological alternations*

  j △ u ▲ w △ ɑ △ n ▲ t △ u ▲ s △ i ▲ ð △ ə ▲ b △ ʊ △ k

  ju wɑnt tu si ðə bʊk

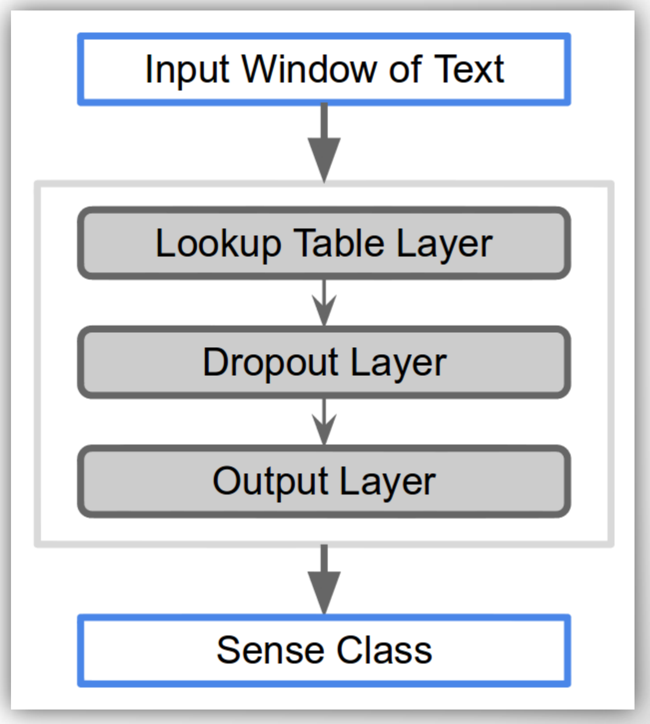
  "you want to see the book"

- Our model combines:
  - a *MaxEnt model of word segmentation* (Berg-Kirkpatrick et al), and
  - a *Harmony Theory/MaxEnt model of phonology* (Smolensky, Goldwater et al)
- Harmony Theory says that *certain feature weights should have specific signs*
- Our results are *significantly better* when we *constrain the weights* as Harmony theory suggests

Semi-Supervised Word Sense Disambiguation Using Word Embeddings in General and Specific Domains

Kaveh Taghipour & Hwee Tou Ng

- A supervised SVM-based *Word Sense Disambiguation* system is improved by incorporating *word embeddings*.
- Moreover, we use a feed-forward neural network to enrich pre-trained embeddings with *task-specific discriminative* information.

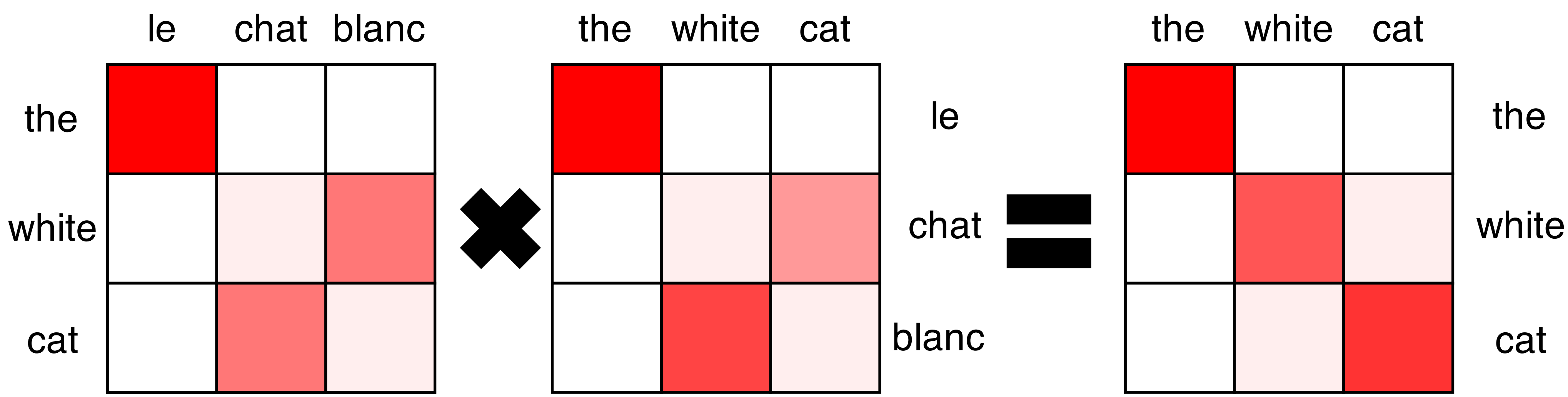


Poster #222

Model Invertibility Regularization:

Sequence Alignment With and Without Parallel Data

$$T_1 \times T_2 \approx I$$



Alignment models trained in reverse directions should be **inverses** of each other

Works even **without parallel data!**

Reviewer 1
"***Ingenious*** *and* ***useful*** *trick*"

Reviewer 2
"***Easier to understand***" *than other methods*

Reviewer 3
"*The idea is very* ***intuitive***"

We can also use **transitivity**! For more information see **Poster #122**

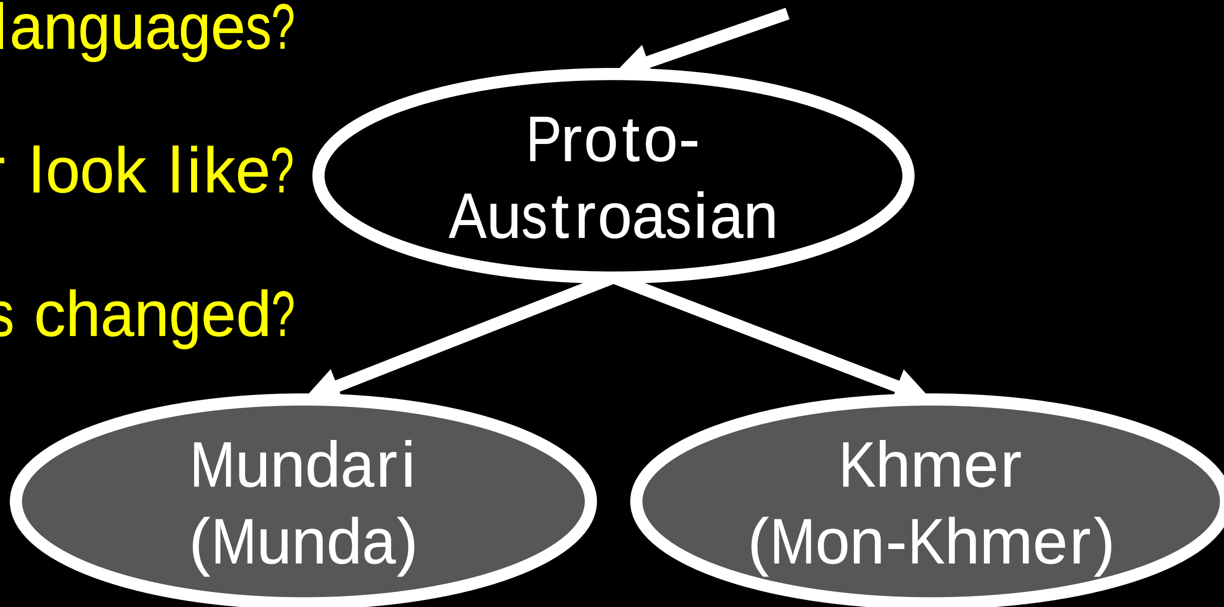
Continuous Space Representations of Linguistic Typology and their Application to Phylogenetic Inference

Yugo Murawaki, Kyushu University

How related to other languages?

How did the ancestor look like?

How have features changed?



Word order:	SOV	SVO
Affixation:	Strongly suffixing	Little affixation
Adposition:	Postpositions	Prepositions

Interpreting Compound Noun Phrases Using Web Search Queries

Marius Paşca - Google Inc. – Mountain View, California

- Goal: Collect normalized lexical interpretations of the semantic roles played by modifiers relative to heads within noun phrases

victorinox knives → knives made by victorinox → knives

kitchen knives → knives used in the kitchen → knives

metal knives → knives made of metal → knives

- Status: From queries, collect candidate lexical interpretations of at most one modifier relative to heads within noun phrases

2009 movies → movies of 2009, movies released in 2009, movies in 2009, movies from 2009, movies for 2009, ...
ayurvedic medicinal plants → medicinal plants in ayurveda, medicinal plants used in ayurveda, medicinal plants of ayurveda, medicinal plants used in ayurvedic medicines, medicinal plants used in ayurvedic products, ...
tsunami charities → charities for tsunami, charities for the tsunami, charities that are helping the tsunami, charities involved in tsunami, charities that have donated to the tsunami, ...

Diamonds in the Rough: Event Extraction from Imperfect Microblog Data

Ander Intxaurrondo, Eneko Agirre, Oier Lopez de Lacalle, Mihai Surdeanu

University of the Basque Country & University of Arizona

- Twitter to extract information about earthquakes.
- Distant supervision for event extraction.
- New evaluation metric.
- Two simple strategies that address inaccurate and ambiguous information.
- Public dataset containing a knowledge base of earthquakes and corresponding tweets.

eman ta zabal zazu
Universidad del País Vasco Euskal Herriko Unibertsitatea
THE UNIVERSITY OF ARIZONA
ixa

A Nonparanormal Approach to Predicting and Generating Popular Meme Descriptions

William Y. Wang (CMU)

Unsupervised Dependency Parsing: Let's Use Supervised Parsers

Phong Le & Willem Zuidema — ILLC, University of Amsterdam — {p.le,zuidema}@uva.nl

Supervised parsing develops fast with

- rich models
- word embedding integration
- efficient learning algorithms

Why is it difficult to reuse supervised parsers?

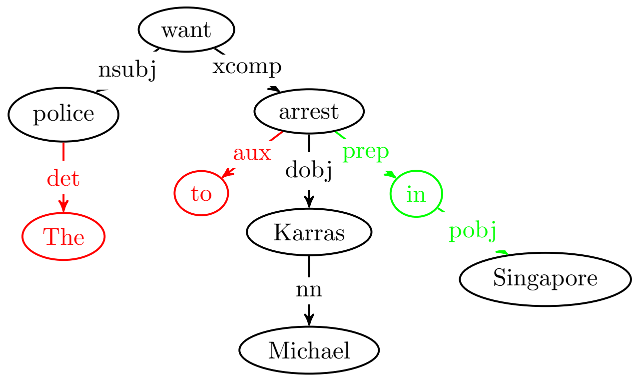
- supervised parsers are designed for being trained on manual annotated data
- using EM to train supervised parsing models is very expensive

Our proposal: *Iterated Reranking* (a variant of self-training)

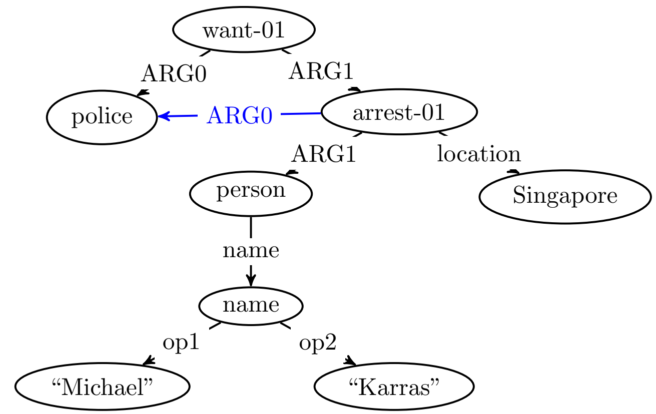
A Transition-based Algorithm for AMR Parsing

Chuan Wang, Nianwen Xue, Sameer Pradhan

There are many similarities between AMR and dependency structure.

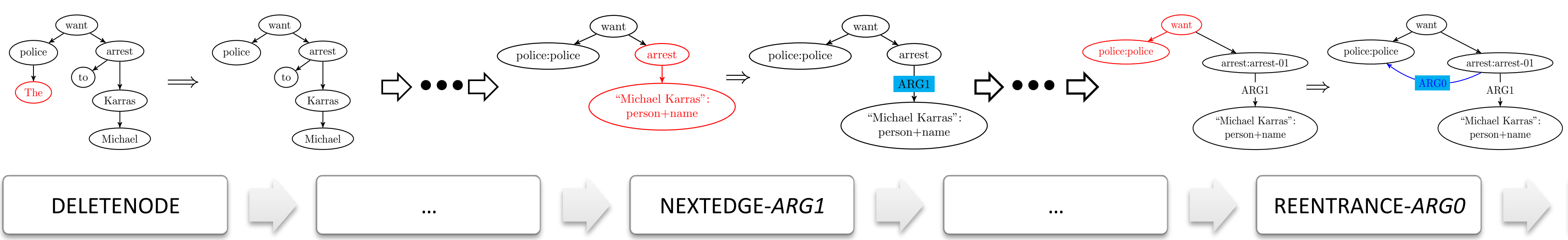


(a) Dependency tree



(b) AMR graph

AMR parsing as transitions from dependency tree to AMR Graph.



The Geometry of Statistical Machine Translation

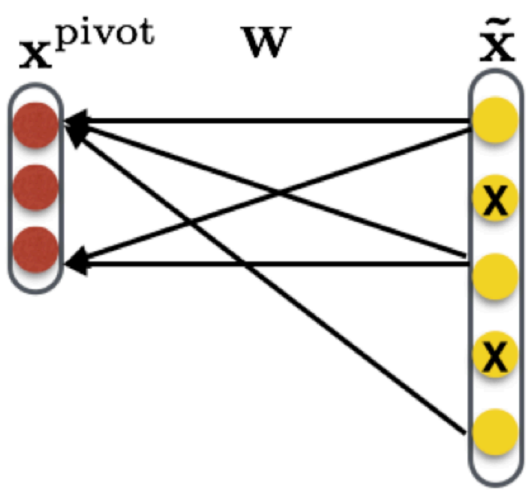
Aurelien Waite, William Byrne

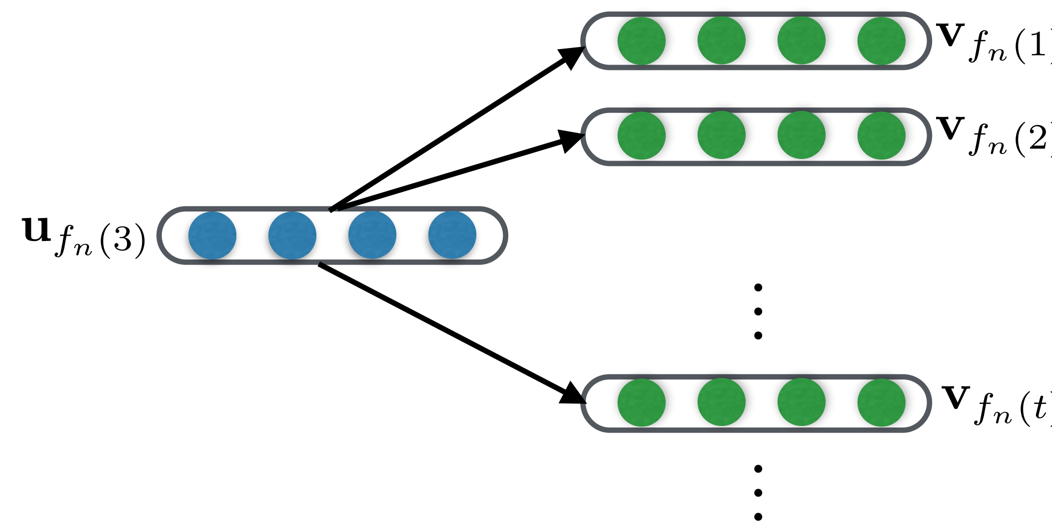
- Why do we struggle with feature-rich MT?
- Convex Geometry: A fresh perspective on the training of linear models (MERT, PRO, MIRA, etc.)
- Optimisers are constrained at low feature dimensions
- Polynomial-time multidimensional MERT

Unsupervised Multi-Domain Adaptation with Feature Embeddings

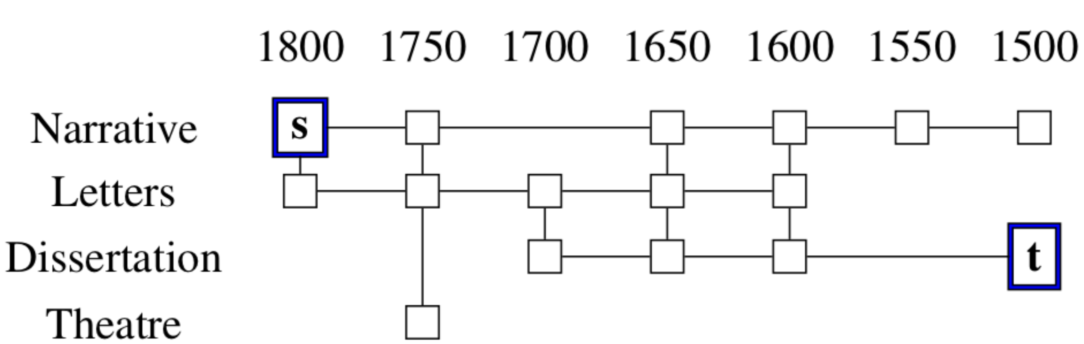
Yi Yang and Jacob Eisenstein (Georgia Tech)

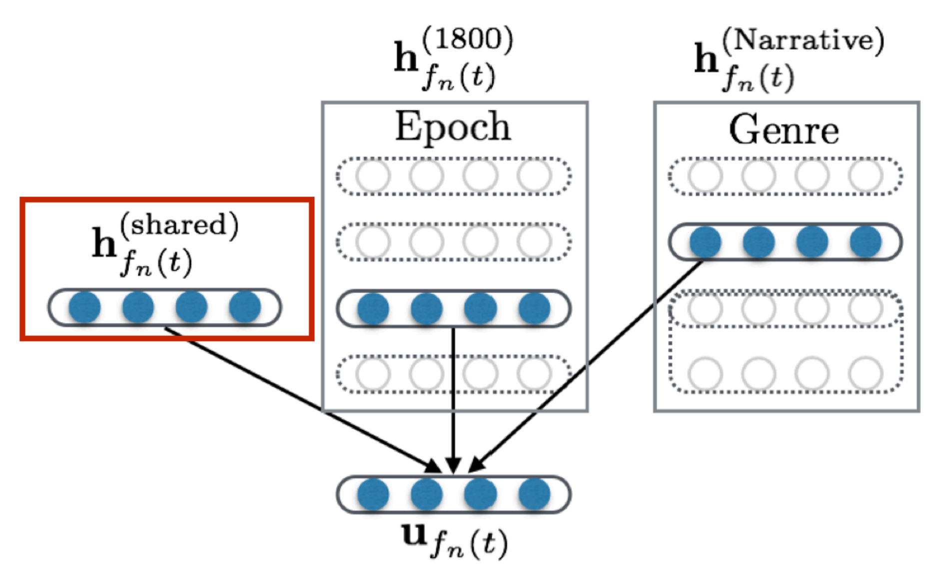
- Improve over pivot based approaches — leverage the full feature co-occurrence matrix





- Beyond single source and target domains setting — unsupervised multi-domain adaptation





We have **Big DATA** to train SMT systems.

- Thanks to Europarl, UN, Common Crawl, ...

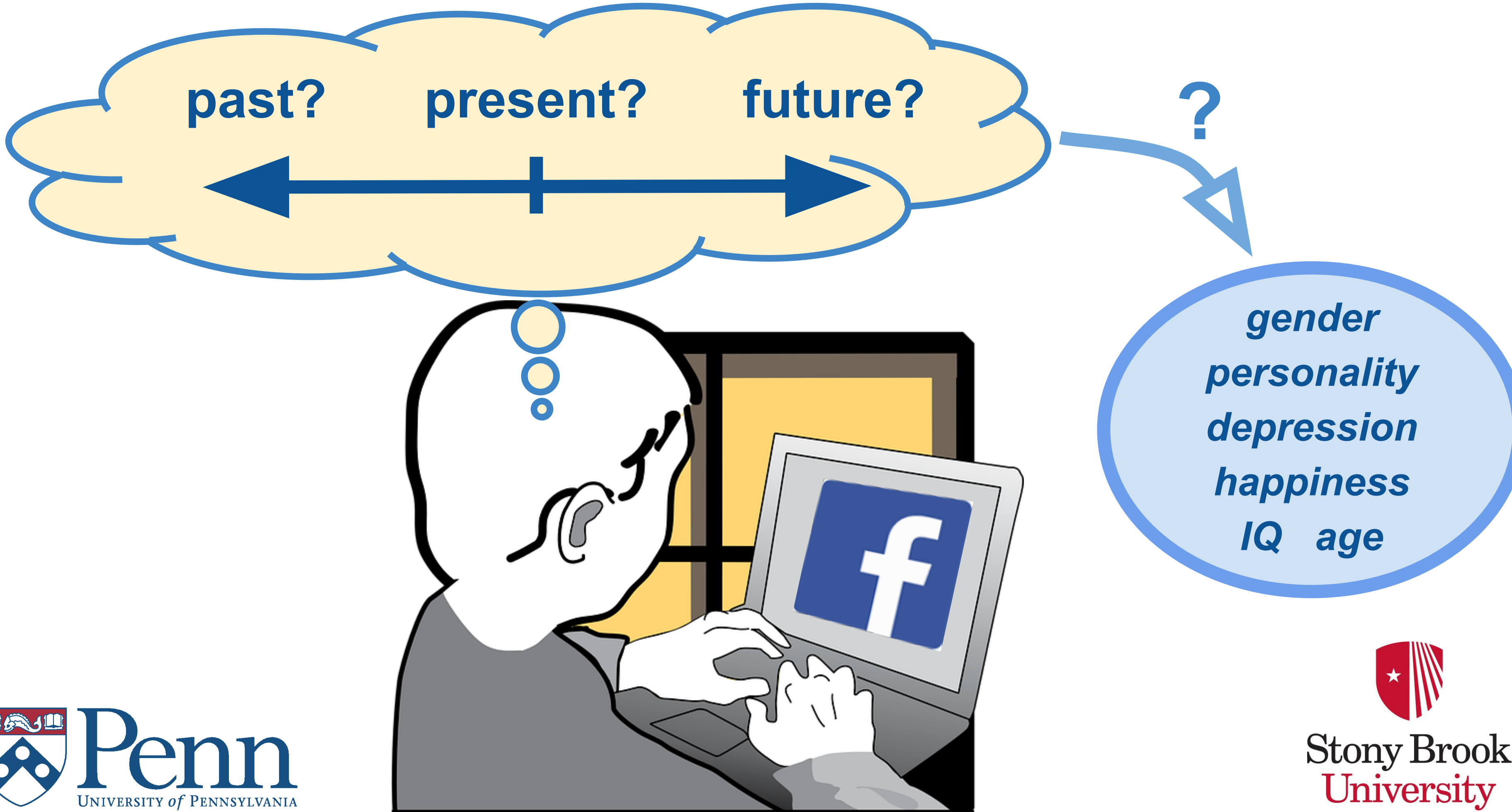
Wait ...

- Data come from **very different domains**.
- **How does this affect the word alignment accuracy**?

Bigger data $\neq$ producing better alignment quality

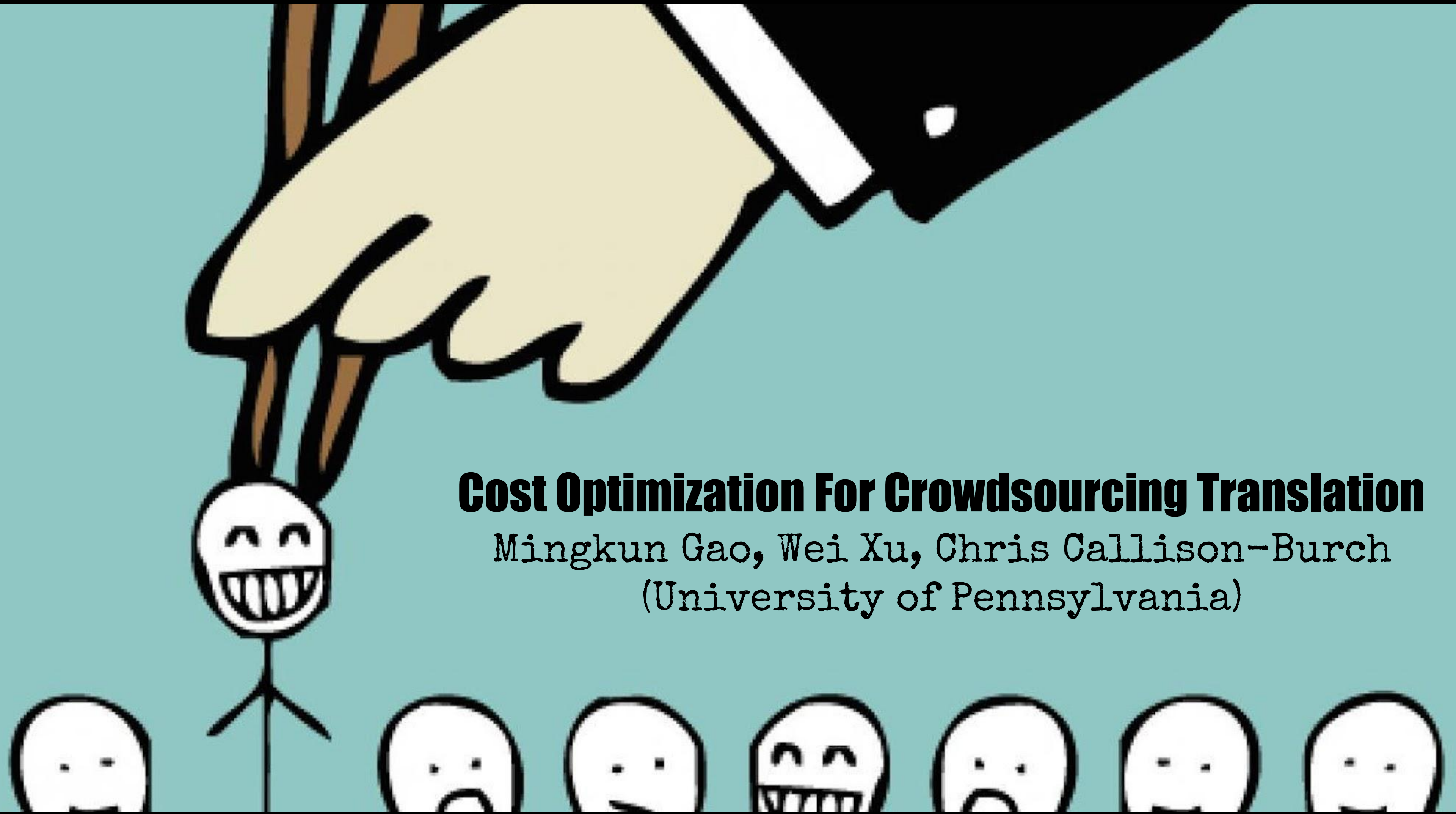
Extracting Human Temporal Orientation from Facebook Language

H. Andrew Schwartz, Greg Park, Maarten Sap, ..., Lyle H. Ungar



Penn
University of Pennsylvania

Stony Brook University



Cost Optimization For Crowdsourcing Translation

Mingkun Gao, Wei Xu, Chris Callison-Burch
(University of Pennsylvania)

An In-depth Analysis of the Effect of Text Normalization in Social Media

Tyler Baldwin and Yunyao Li

Past	Present	Future

TTS Normalization

Parser Normalization

NER Normalization

MT Normalization

… …

@someGuy idk y u think that tho

?

Multitask Learning for Adaptive Quality Estimation of Automatically Transcribed Utterances

José G. C. de Souza, Hamed Zamani, Matteo Negri, Marco Turchi, Daniele Falavigna

PROBLEM:

Predict ASR output quality without references nor access to confidence information…

???

"It's hard to wreck a nice beach"

"It's hard to recognize speech"

Input signal — ASR system — **Transcription** — Reference

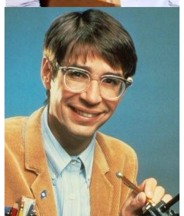
QUALITY = 0.03 (or "BAD")

…by learning from data coming from multiple, heterogeneous domains

Medical

"Hi, I'm Mister teeth"

IT



"Hi, I'm a brilliant, but socially inept computer hacker"

Legal

"Mother she wrote"

SOLUTION:

Come to poster session 1B

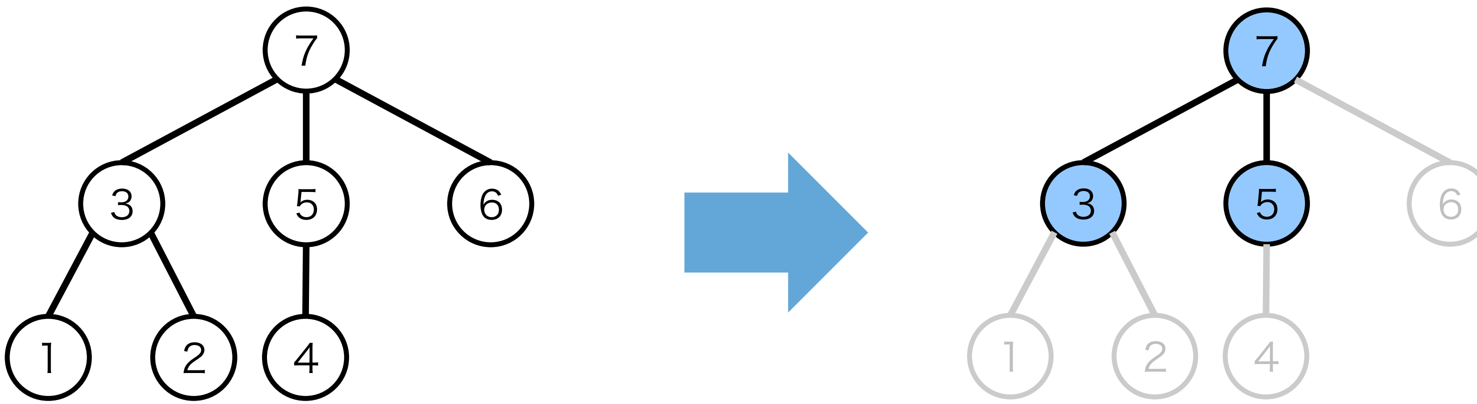
poster 460!

A Dynamic Programming Algorithm for Tree Trimming-based Text Summarization

Masaaki Nishino (NTT)

A tree trimming problem

Input **Output**



A dependency tree (sentence/document)

An optimal subtree within a length limit **L** **(summary)**

O(NL log N)

exact algorithms!

Innovative R&D by NTT

NTT

Sentiment after Translation: A Case-Study on Arabic Social Media Posts

Mohammad Salameh
University of Alberta

Saif M. Mohammad
National Research Council Canada

Svetlana Kiritchenko
National Research Council Canada

Dialectal Arabic:

مش غلط يكون اقلها مرتب (Positive)

Manual Translation:

it is not wrong to at least be neat (Neutral)

Automatic Translation:

it's not a mistake be less of their salary (Negative)

Questions:

- What is the accuracy of a state-of-the-art Arabic sentiment analysis (SA) system?
- What accuracy is expected when the posts are translated (manually or automatically) and run through an English SA system?
    - What is the impact of bad translation?
- How difficult is it for humans to determine sentiment of automatically translated text?

We don't seem to be doing well in the foreign beverage market. However, due to a mistranslation of our slogan we've become the leading international provider of embalming fluid

UNIVERSITY OF ALBERTA

NRC·CNRC

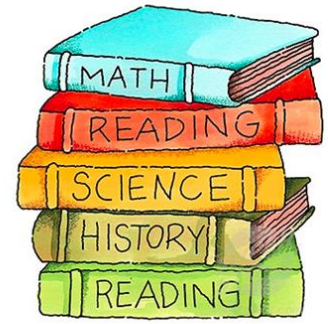
Corpus-based discovery of semantic intensity scales

They are synonyms !

Really?

Chomsky

word2vec + clustering



Derive clusters

good
great
excellent

Adjective scale

good < great < excellent

considerably
significantly
dramatically

Adverb scales

considerably < significantly < dramatically

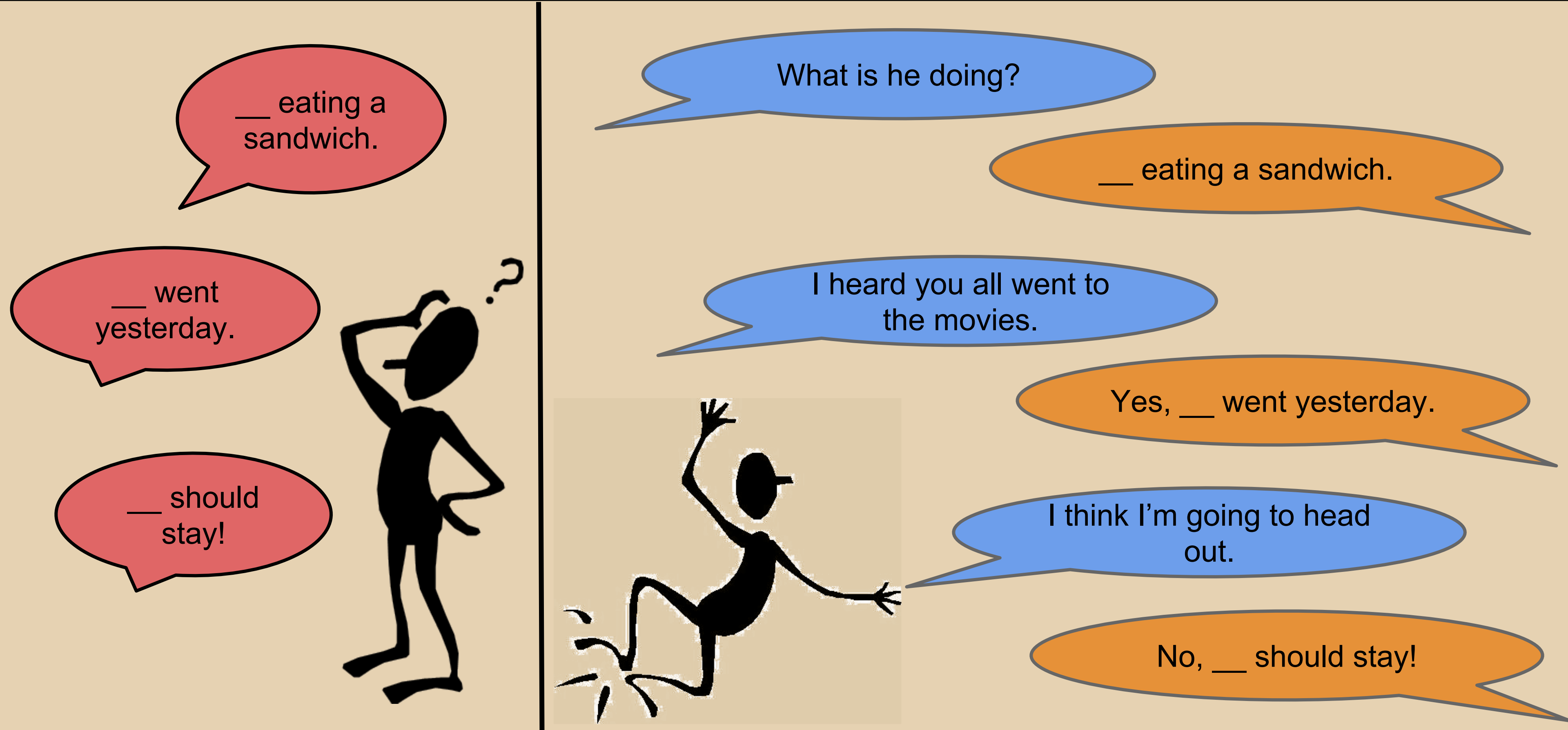
BONUS!

Applications
Sentiment Analysis

THE OHIO STATE UNIVERSITY

Dialogue focus tracking for zero pronoun resolution

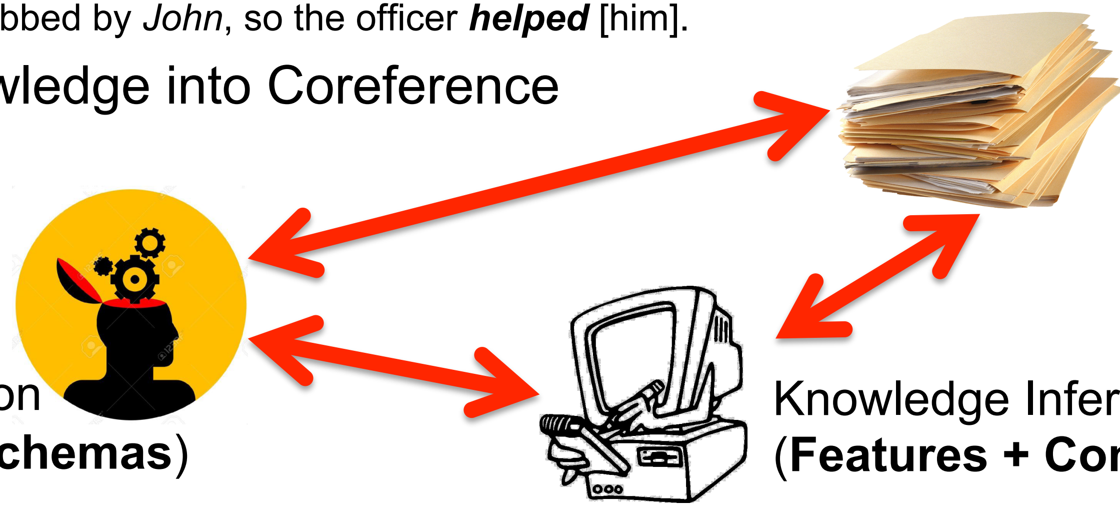
Sudha Rao, Allyson Ettinger, Hal Daumé III, Philip Resnik



Solving Hard Coreference Problems

Haoruo Peng, Daniel Khashabi and Dan Roth

- One fundamental difficulty of **coreference resolution** is to resolve instances that require background **knowledge**.
  - The *bird* perched on the *limb* and [it] ***bent***.
  - The *bird* perched on the *limb* and [it] ***sang***.
  - *Bill* was robbed by *John*, so the officer ***arrested*** [him].
  - *Bill* was robbed by *John*, so the officer ***helped*** [him].
- Inject Knowledge into Coreference



- We build a state-of-the-art coreference system that at the same time also handles hard instances at close to 90% precision.

Demo: http://cogcomp.cs.illinois.edu/page/demo_view/Coref

COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Reasoning about Quantities in Natural Language

Subhro Roy, Tim Vieira, Dan Roth

A bomb in a Hebrew University cafeteria killed **five Americans** and **four Israelis**.

A bombing at Hebrew University in Jerusalem killed **nine people**, including **five Americans**

Ryan has **72 marbles** and **17 blocks**. If he shares the marbles among **9 friends**, how many marbles does each friend get?

Each friend gets 72/9 = 8 marbles. The number of blocks is irrelevant.

Unsupervised Declarative Knowledge Induction for Constraint-Based Learning of Information Structure in Scientific Documents

Yufan Guo, Roi Reichart, Anna Korhonen

- Declarative Knowledge: $0.9 \leq p(\mathrm{CON} \mid \mathrm{suggest}) \leq 1$
- Automatic Induction: $? \leq p(? \mid ?) \leq ?$

UNIVERSITY OF CAMBRIDGE

TECHNION Israel Institute of Technology