# Our Quest for Interpretable Natural Language Processing

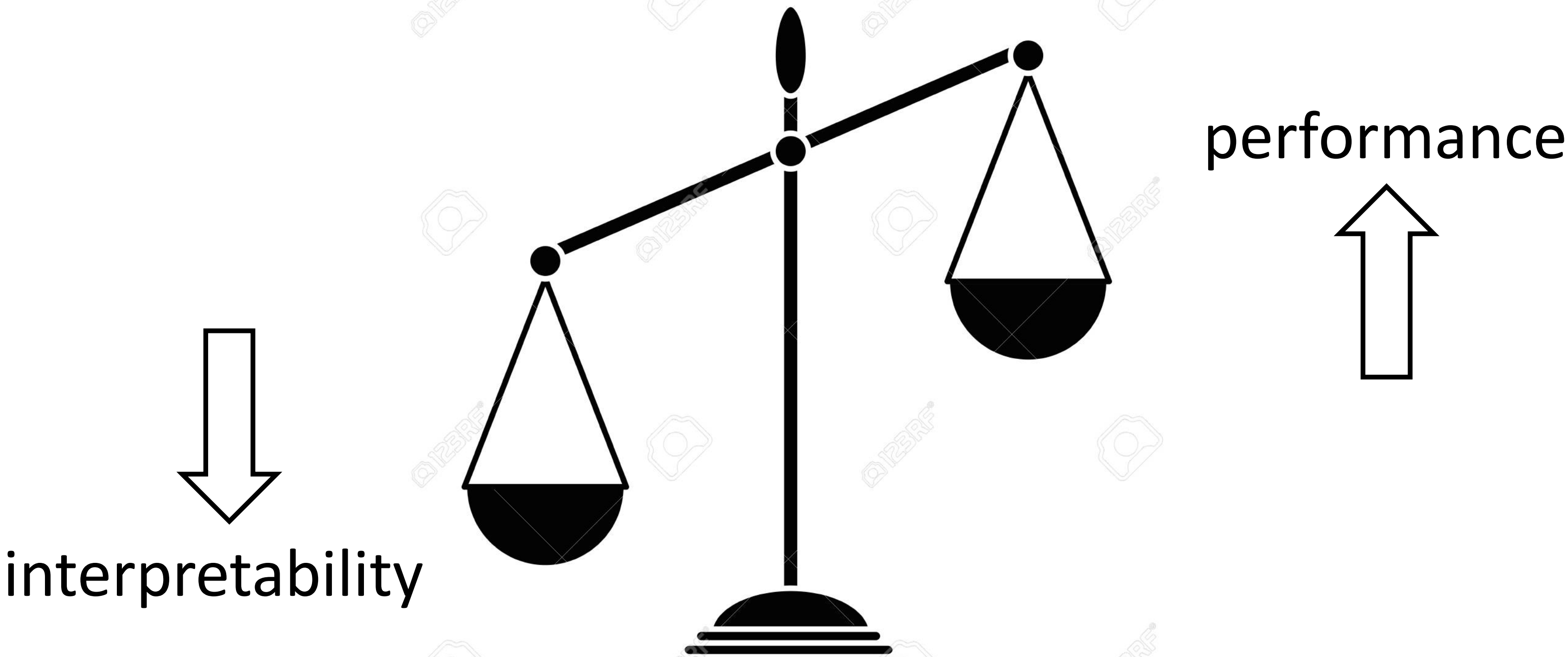
Mihai Surdeanu

April 2020

The University of Arizona

Computational Language Understanding

# Recent ML developments: a deal with the devil



# Interpretability

- Interpretability is an overloaded term in machine learning (ML)
- But we can classify it roughly in two classes

# 1. Post hoc interpretability (Ribeiro et al., 2016)

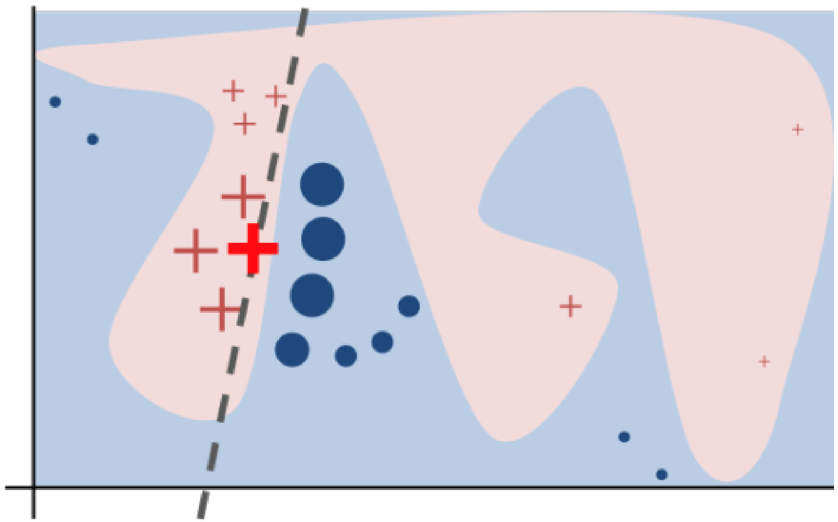
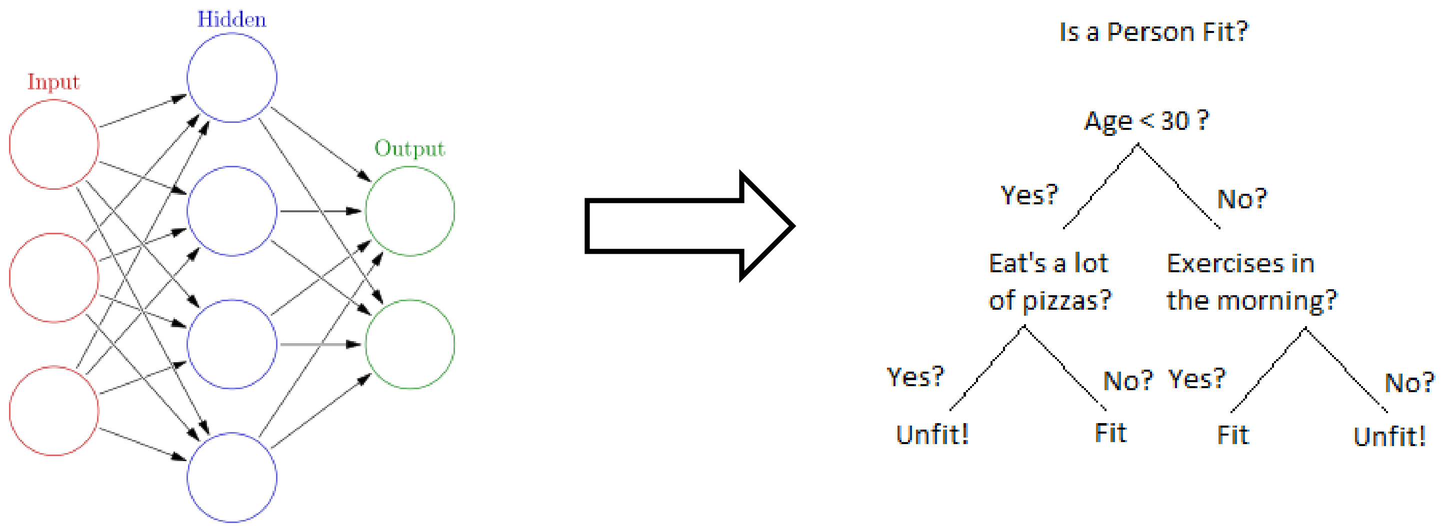


**Figure 3: Toy example to present intuition for LIME. The black-box model's complex decision function $f$ (unknown to LIME) is represented by the blue/pink background, which cannot be approximated well by a linear model. The bold red cross is the instance being explained. LIME samples instances, gets predictions using $f$, and weighs them by the proximity to the instance being explained (represented here by size). The dashed line is the learned explanation that is locally (but not globally) faithful.**

# 2. Global interpretability: Converting the statistical model into something interpretable (aka ”knowledge distillation”)




(Craven and Shavlik, 1996) (Hinton et al., 2015)

# Comparison

- Post hoc interpretability
  - May provide only an approximate explanation (Netflix lies to you ☺)
  - You can’t fix a problem with the original model when identified
- Global interpretability
  - May lose some performance in the conversion
  - Allows corrections to the model when problems are discovered

# Comparison

- Post hoc interpretability
  - May provide only an approximate explanation (Netflix lies to you ☺)
  - You can't fix a problem with the original model when identified
- Global interpretability
  - May lose some performance in the conversion
  - **Allows corrections to the model when problems discovered**

# Comparison

- Post hoc interpretability
  - May provide o on (Netflix lies to y
  - You can't fix a probl th the original model when identified

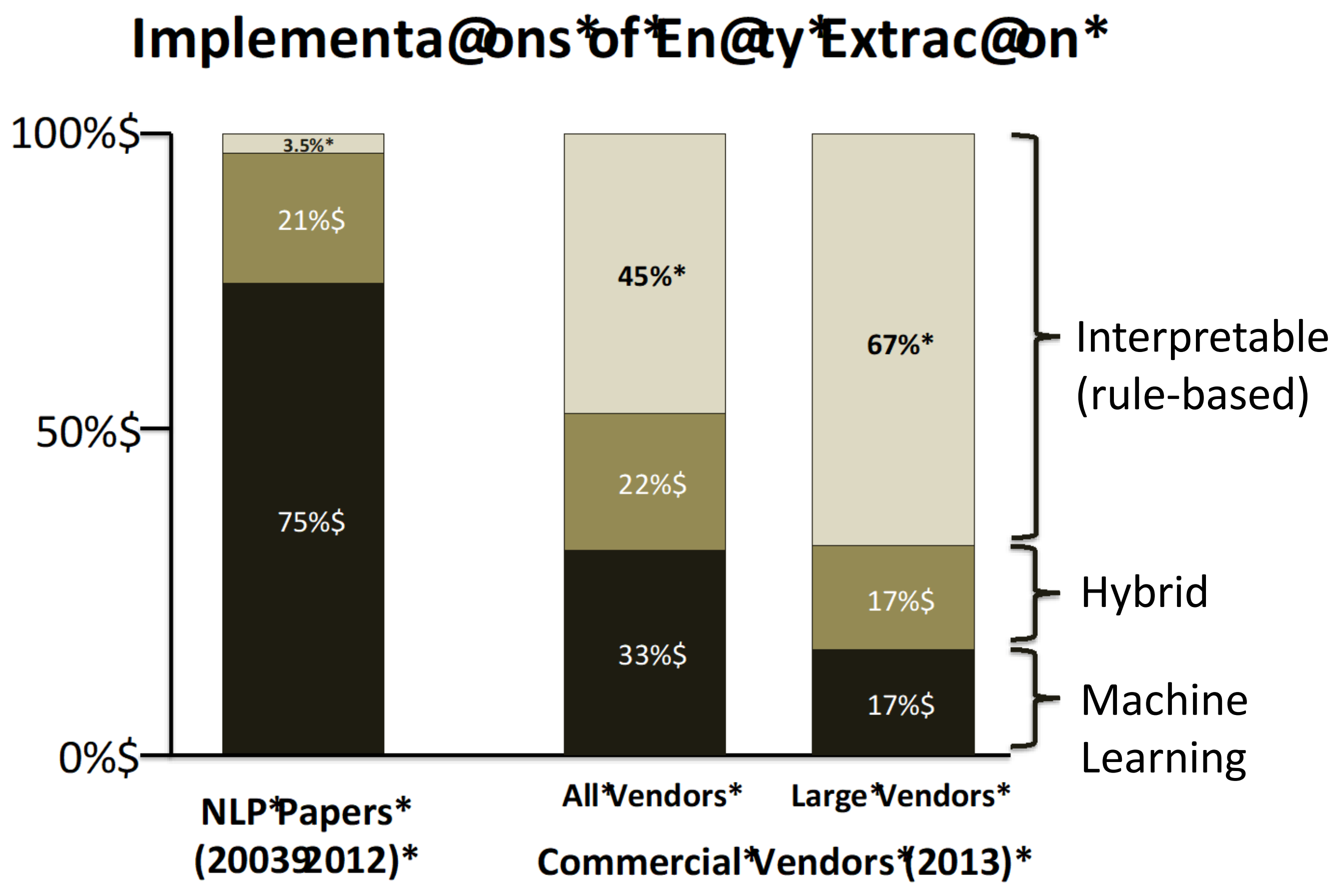
This is our focus!

- **Global interpretability**
  - May lose some performance in the conversion
  - **Allows corrections to the model when problems discovered**

# **Globally interpretable** models for natural language processing

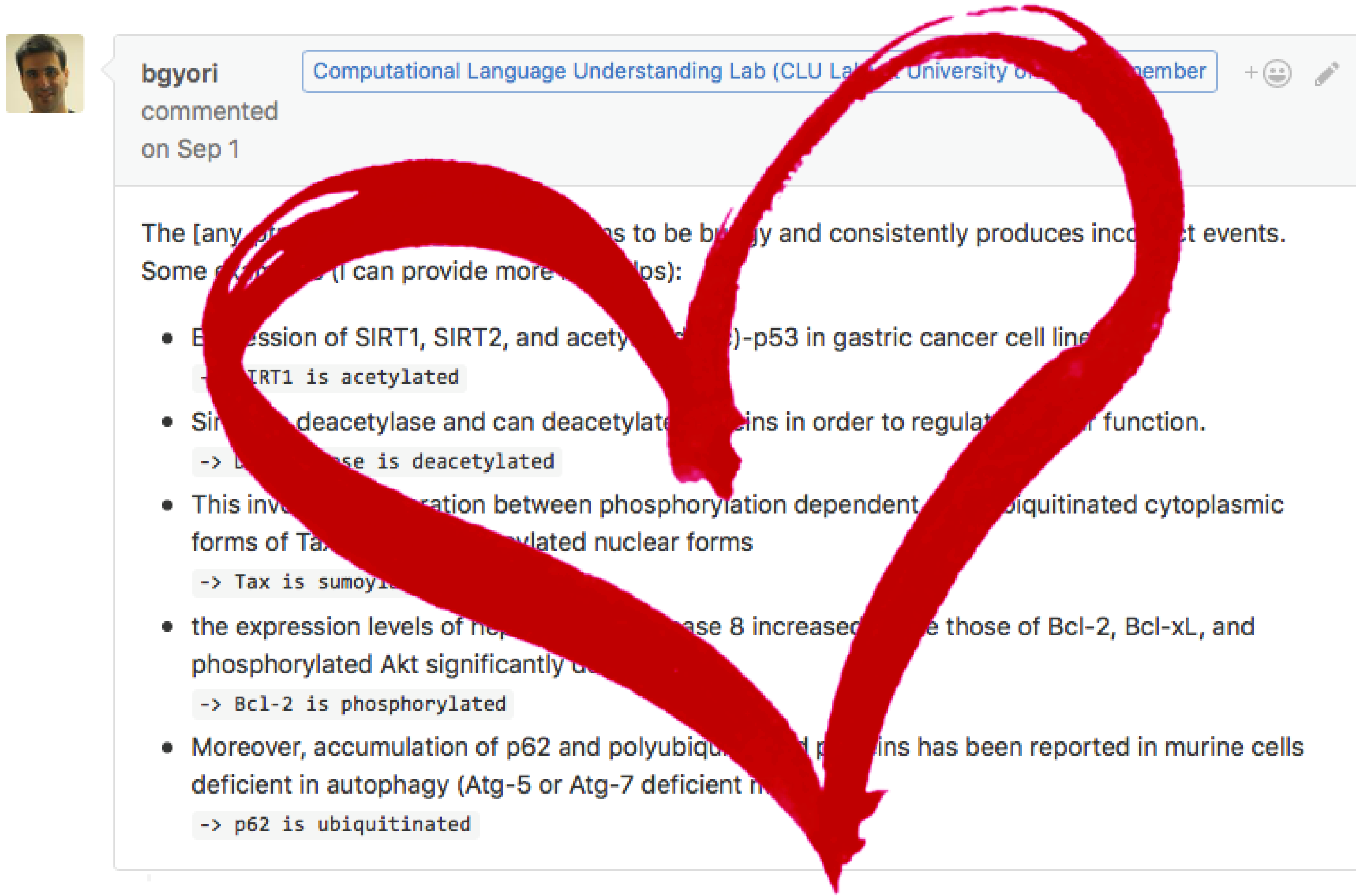
1. Humans can understand the model
2. Humans can change the model

# Academia vs. industry



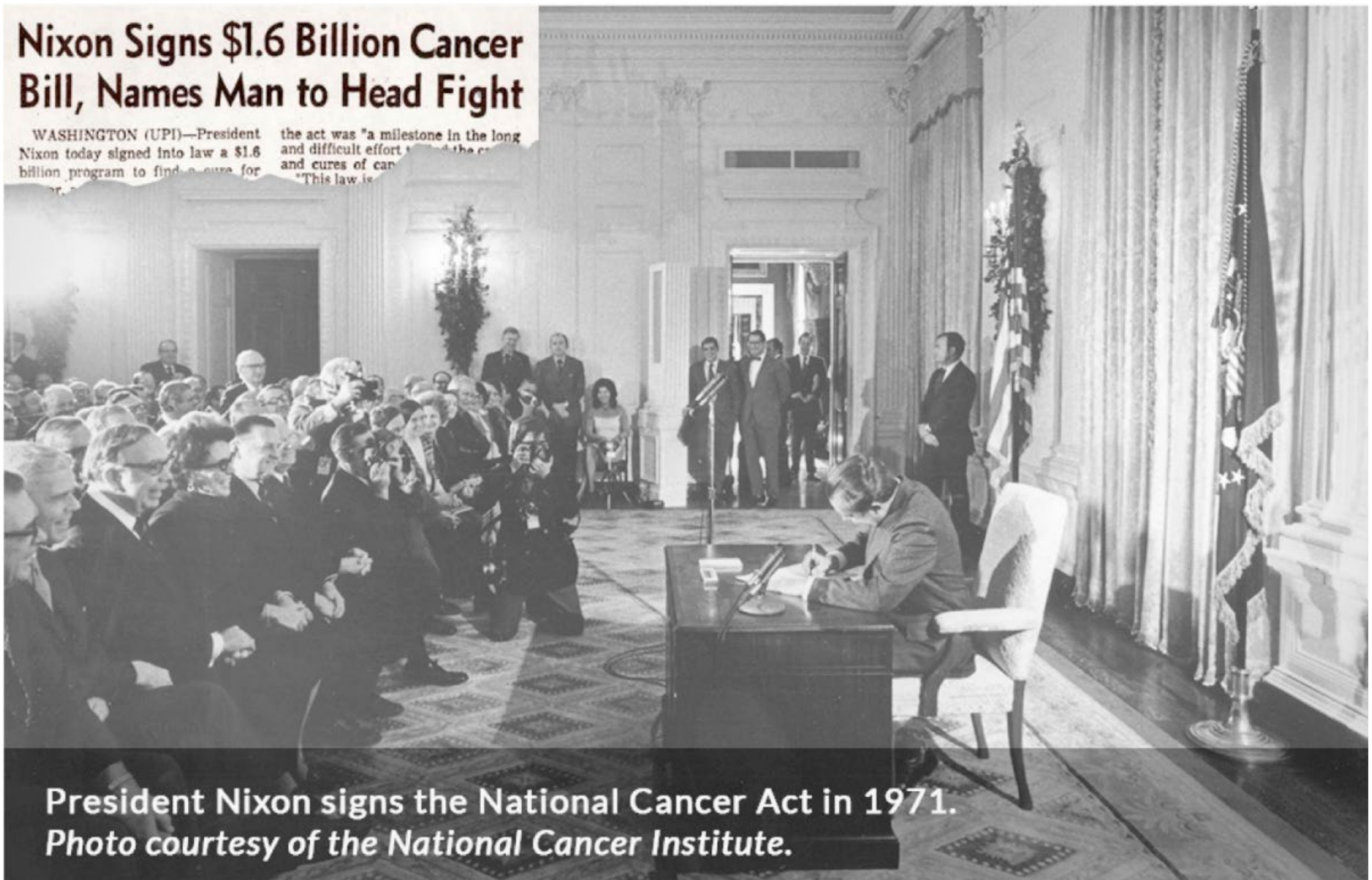
# Why do we care about global interpretability?

- Most of today's meaningful projects are inter-disciplinary, e.g., ML + medicine
- We can't expect an expert in another domain to understand (and fix!) our statistical classifiers
- But we need to iterate quickly…


bgyori
commented
on Sep 1
Computational Language Understanding Lab (CLU La
University o
ember
The [any
s to be b
y and consistently produces inco
t events.
Some
(I can provide more
ps):
ession of SIRT1, SIRT2, and acety
)-p53 in gastric cancer cell line
IRT1 is acetylated
deacetylase and can deacetylat
ins in order to regula
function.
-> 
se is deacetylated
This inv
ation between phosphorylation dependent
iquitinated cytoplasmic
forms of Ta
ylated nuclear forms
-> Tax is sumoyl
the expression levels of
ase 8 increase
e those of Bcl-2, Bcl-xL, and
phosphorylated Akt significantly
-> Bcl-2 is phosphorylated
Moreover, accumulation of p62 and polyubiqu
ins has been reported in murine cells
deficient in autophagy (Atg-5 or Atg-7 deficient
-> p62 is ubiquitinated

# USE CASE: MACHINE READING FOR CANCER RESEARCH

# Why cancer?

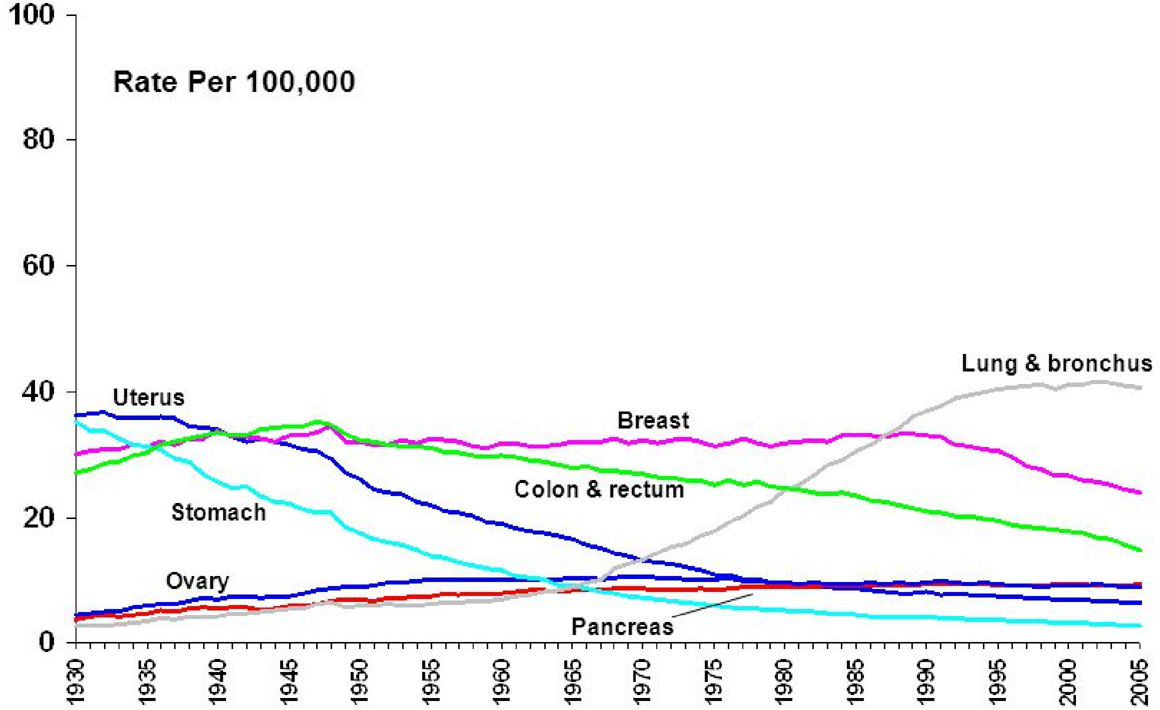


President Nixon signs the National Cancer Act in 1971.
*Photo courtesy of the National Cancer Institute.*

$200 billion have been invested in cancer research since (R. Barzilay, NAACL 2016)

# Why cancer?

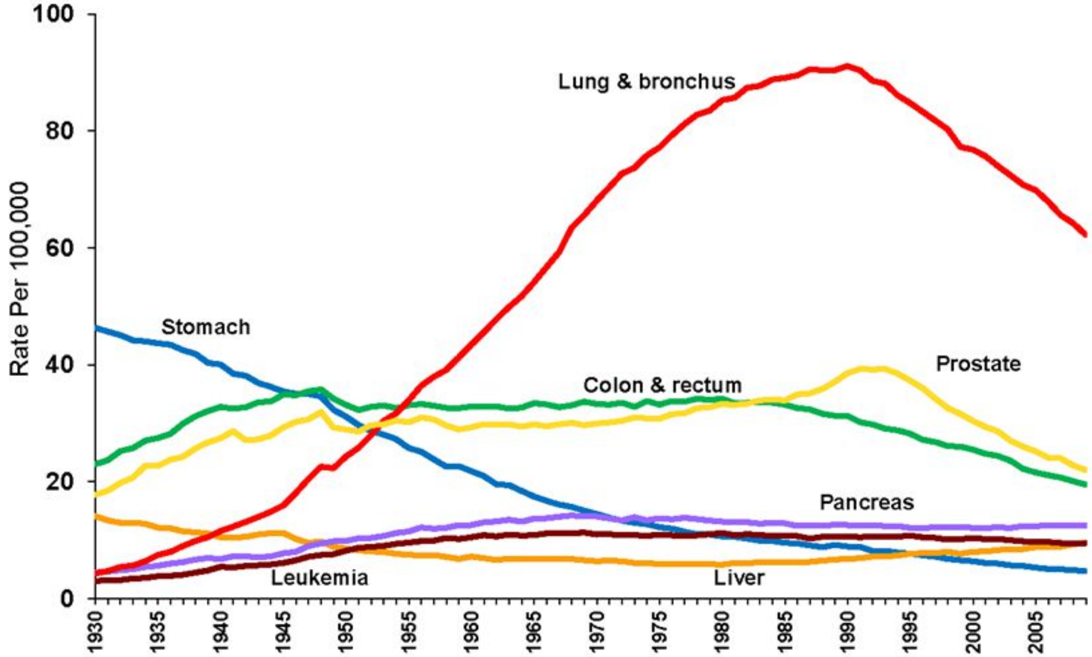
Cancer Death Rates* Among Women, US, 1930 – 2005




*Age-adjusted to the 2000 US standard population.
Source: US Mortality Data 1960-2005, US Mortality Volumes 1930-1959,
National Center for Health Statistics, Centers for Disease Control and Prevention, 2008.

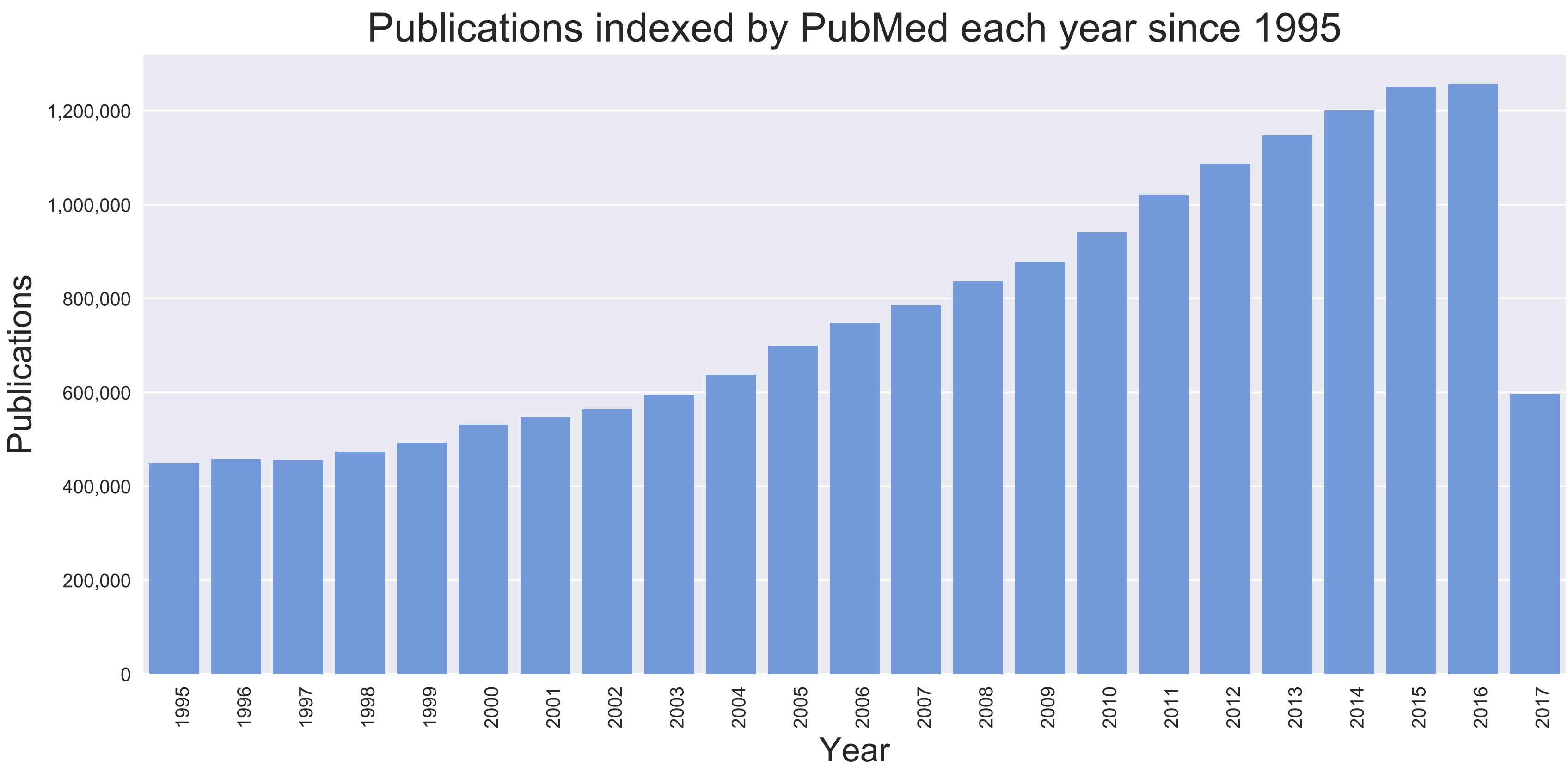
# Why cancer?

Cancer Death Rates* Among Men, US, 1930 – 2009

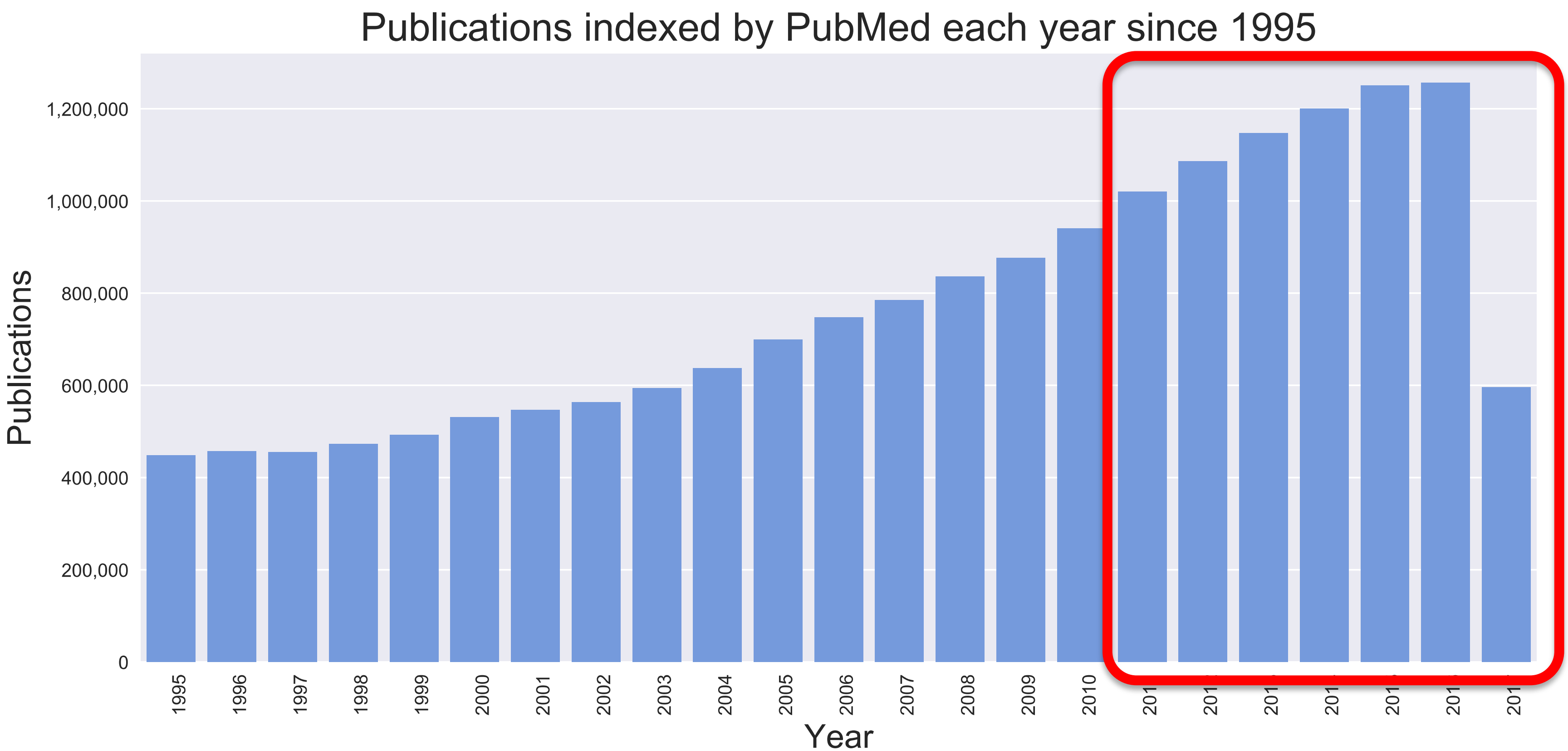



*Age-adjusted to the 2000 US standard population.
Source: US Mortality Data 1960-2009, US Mortality Volumes 1930-1959, National Center for Health Statistics, Centers for Disease Control and Prevention.

# Why the slow progress?



# Why the slow progress?




90% are never cited![1]

[1] http://www.smithsonianmag.com/smart-news/half-academic-studies-are-never-read-more-three-people-180950222/?no-ist

The problem of “undiscovered public knowledge”
(Swanson, 1986)

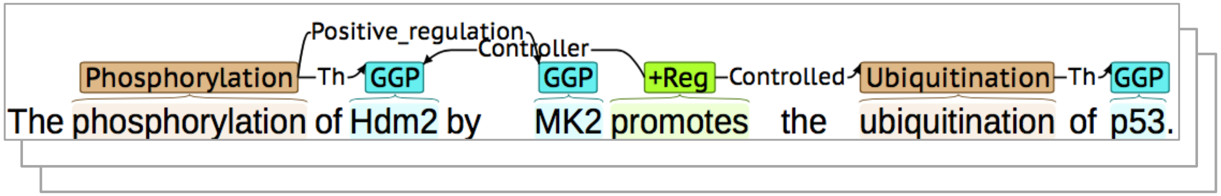
“A knotty puzzle may hold a scientist up for a century, when it may be that a colleague has the solution already and is not even aware of the puzzle that it might solve.”

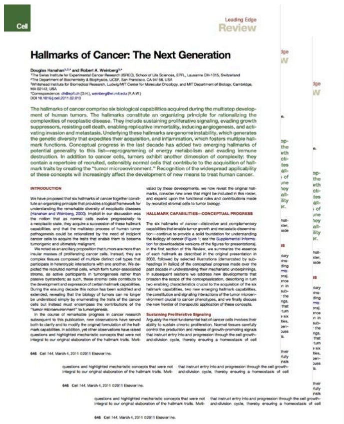
– Isaac Asimov, *The Robots of Dawn*

# We need machine reading

- If humans can’t process this much information, then machines must help!

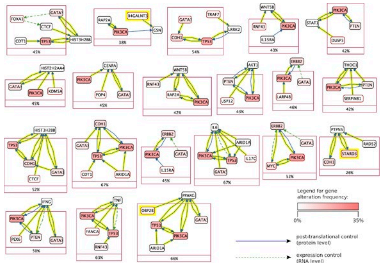
# Machine reading for biomedical literature




**Reading**

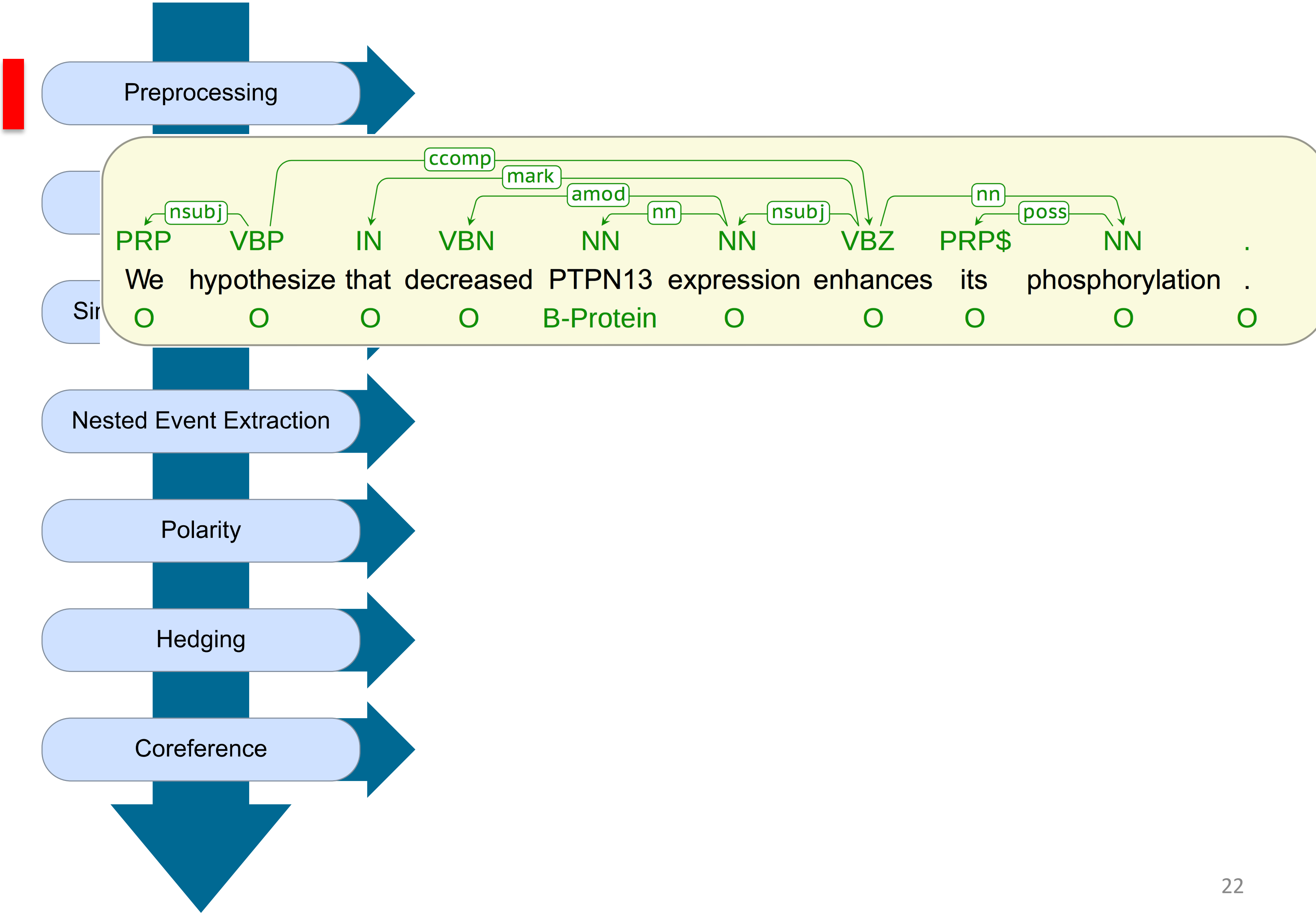
**Assembly**



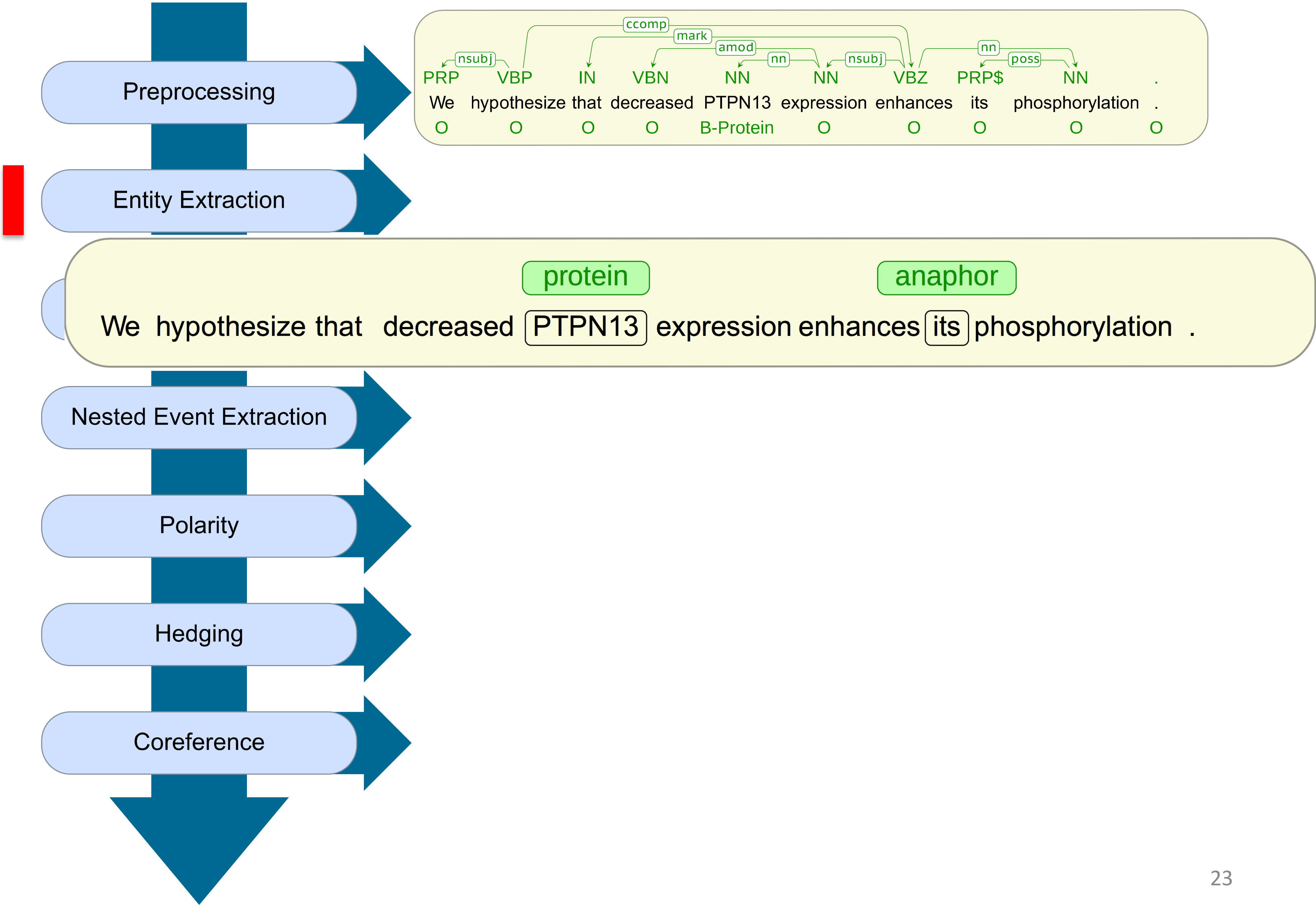
**Reasoning**

Predict, explain, test, curate, etc.

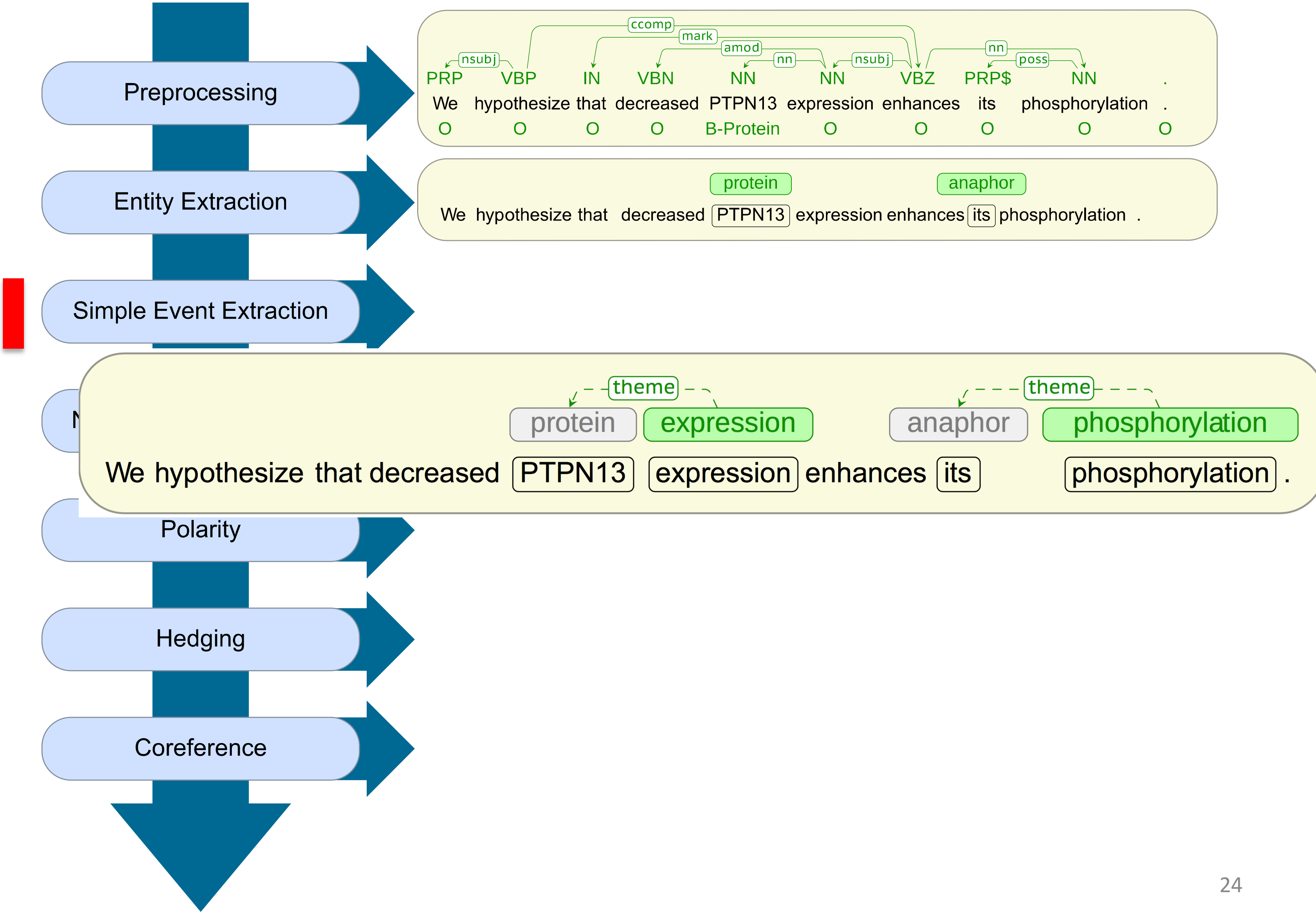
# REACH (**RE**ading and **A**ssembling **C**ontextual and **H**olistic mechanisms)



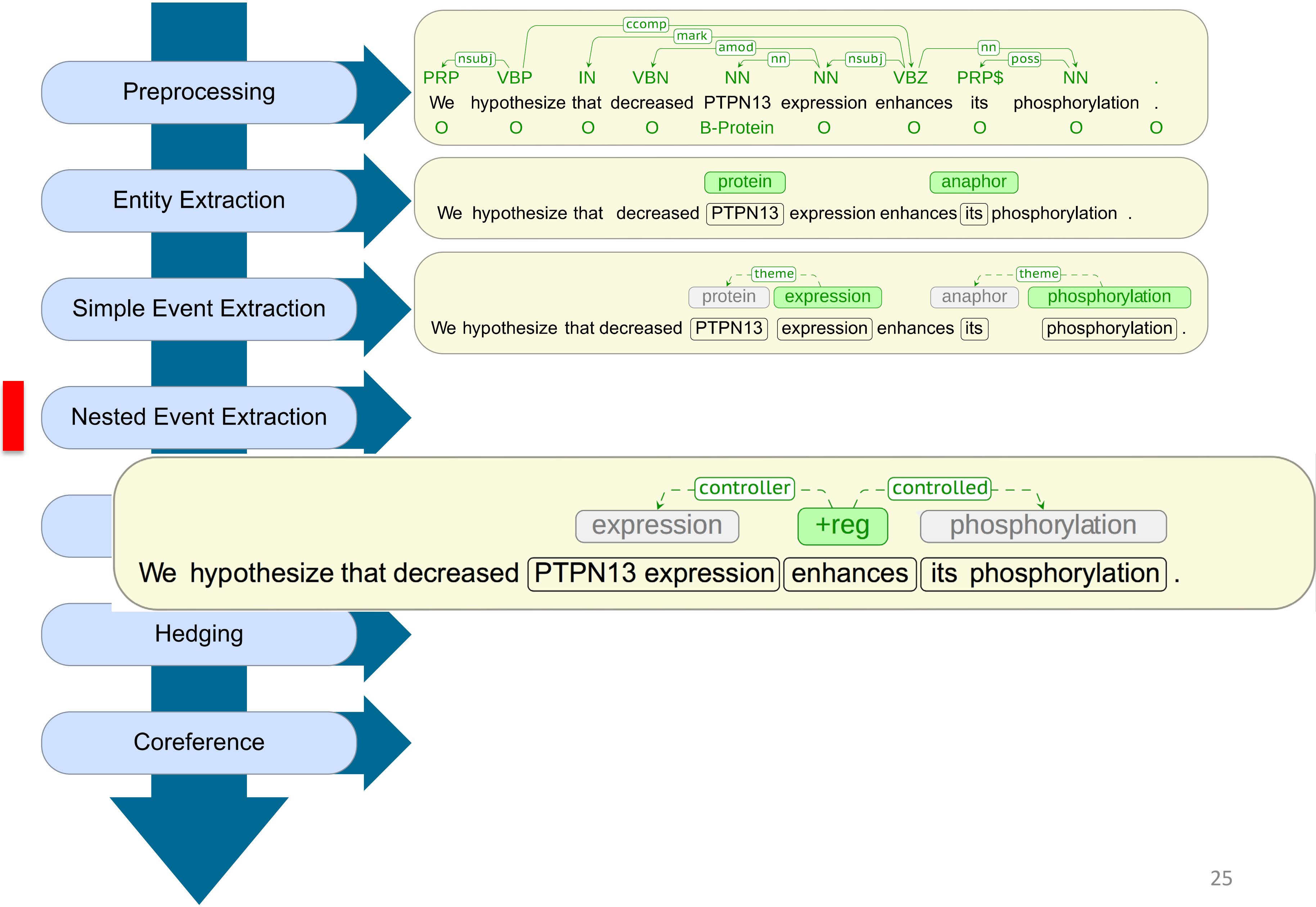
# REACH (**RE**ading and **A**ssembling **C**ontextual and **H**olistic mechanisms)



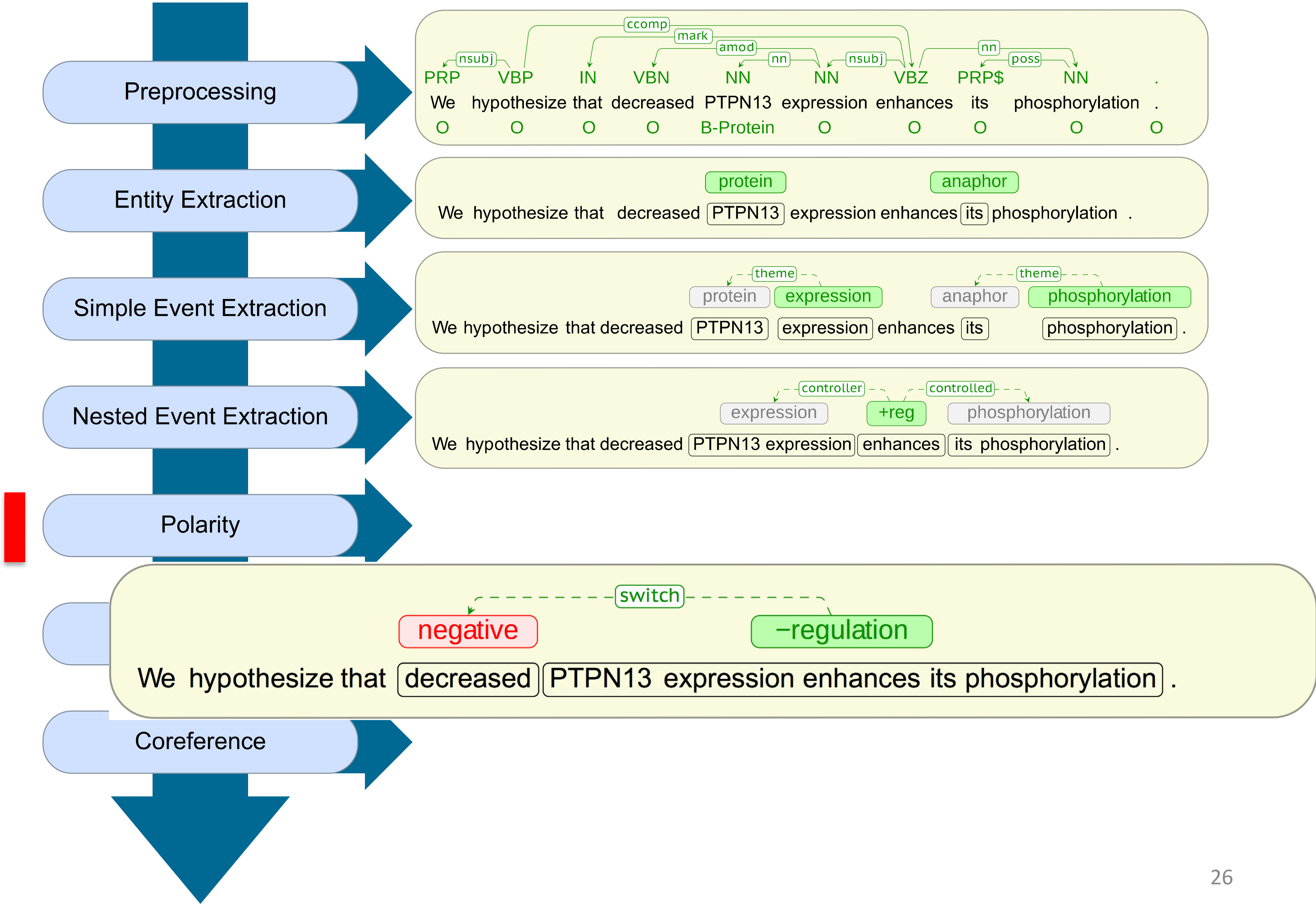
# REACH (**RE**ading and **A**ssembling **C**ontextual and **H**olistic mechanisms)



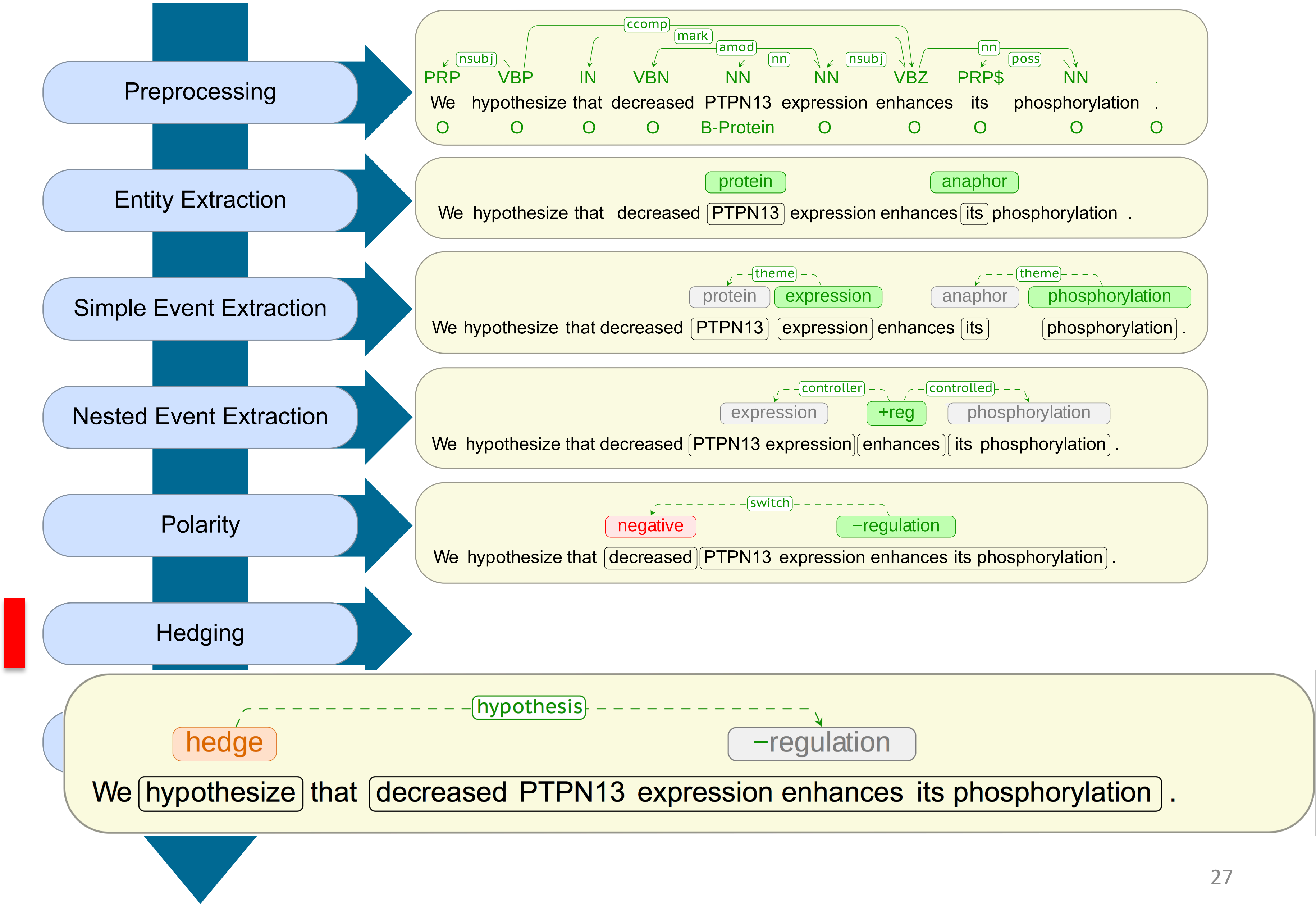
# REACH (**RE**ading and **A**ssembling **C**ontextual and **H**olistic mechanisms)



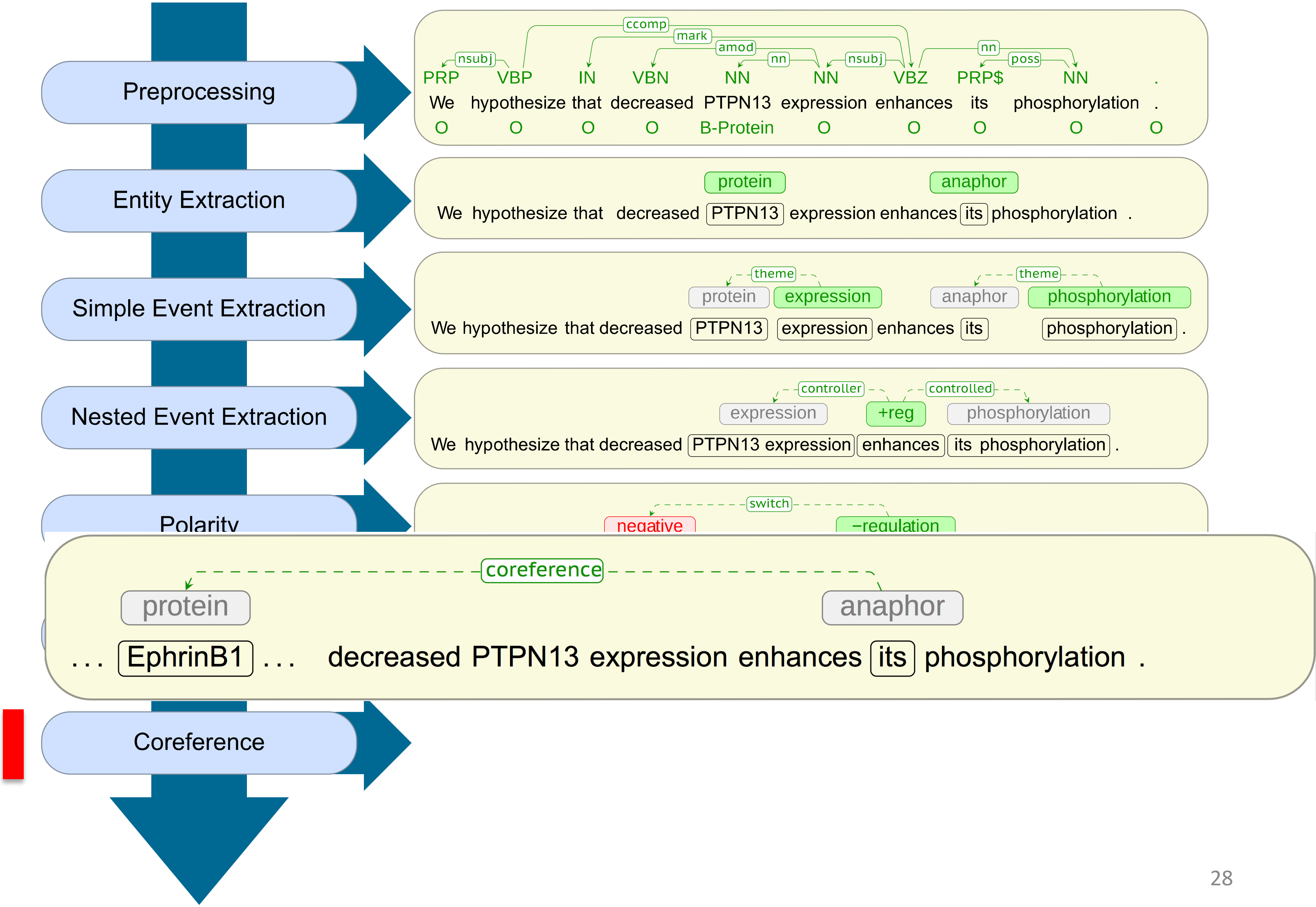
# REACH (**RE**ading and **A**ssembling **C**ontextual and **H**olistic mechanisms)



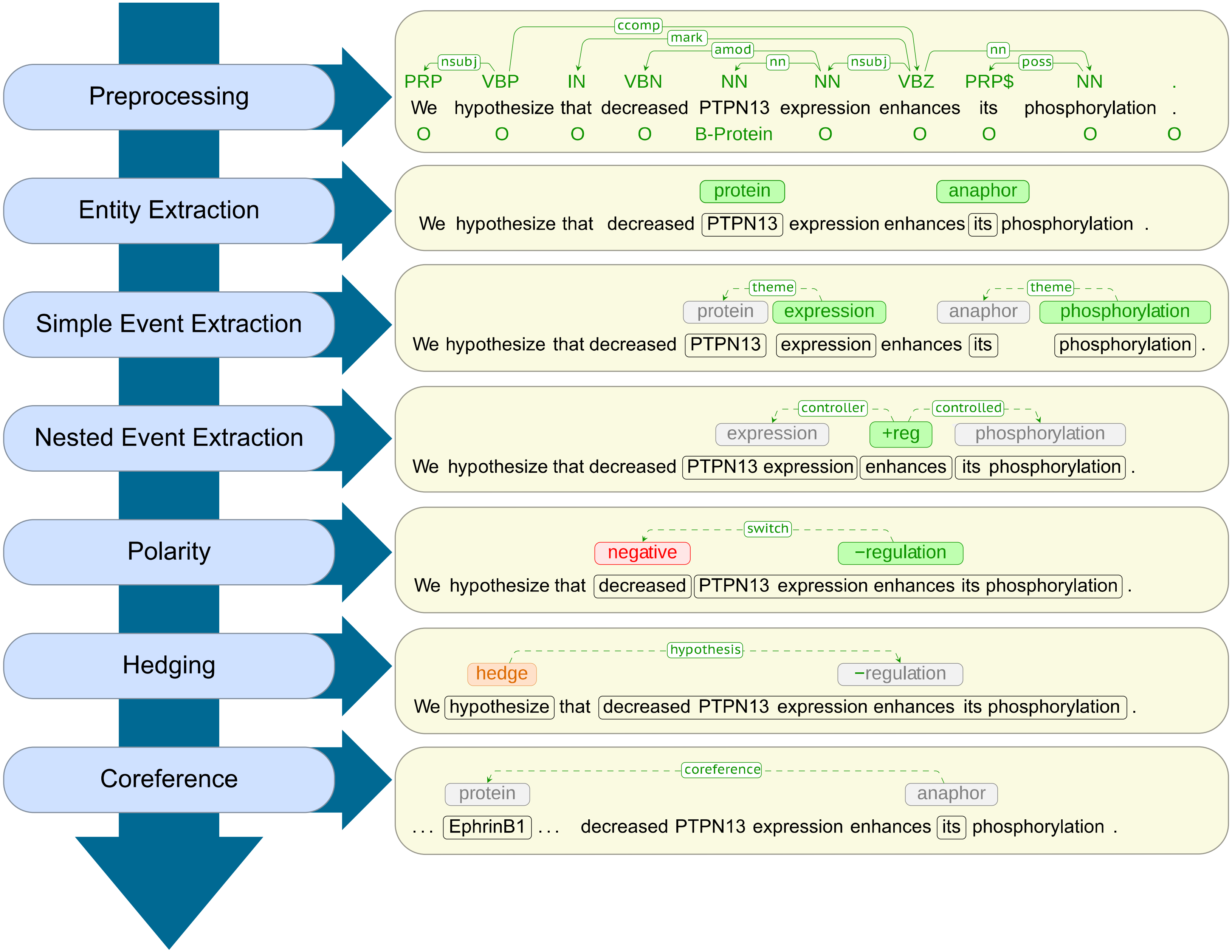
# REACH (**RE**ading and **A**ssembling **C**ontextual and **H**olistic mechanisms)



# REACH (**RE**ading and **A**ssembling **C**ontextual and **H**olistic mechanisms)



# REACH (**RE**ading and **A**ssembling **C**ontextual and **H**olistic mechanisms)



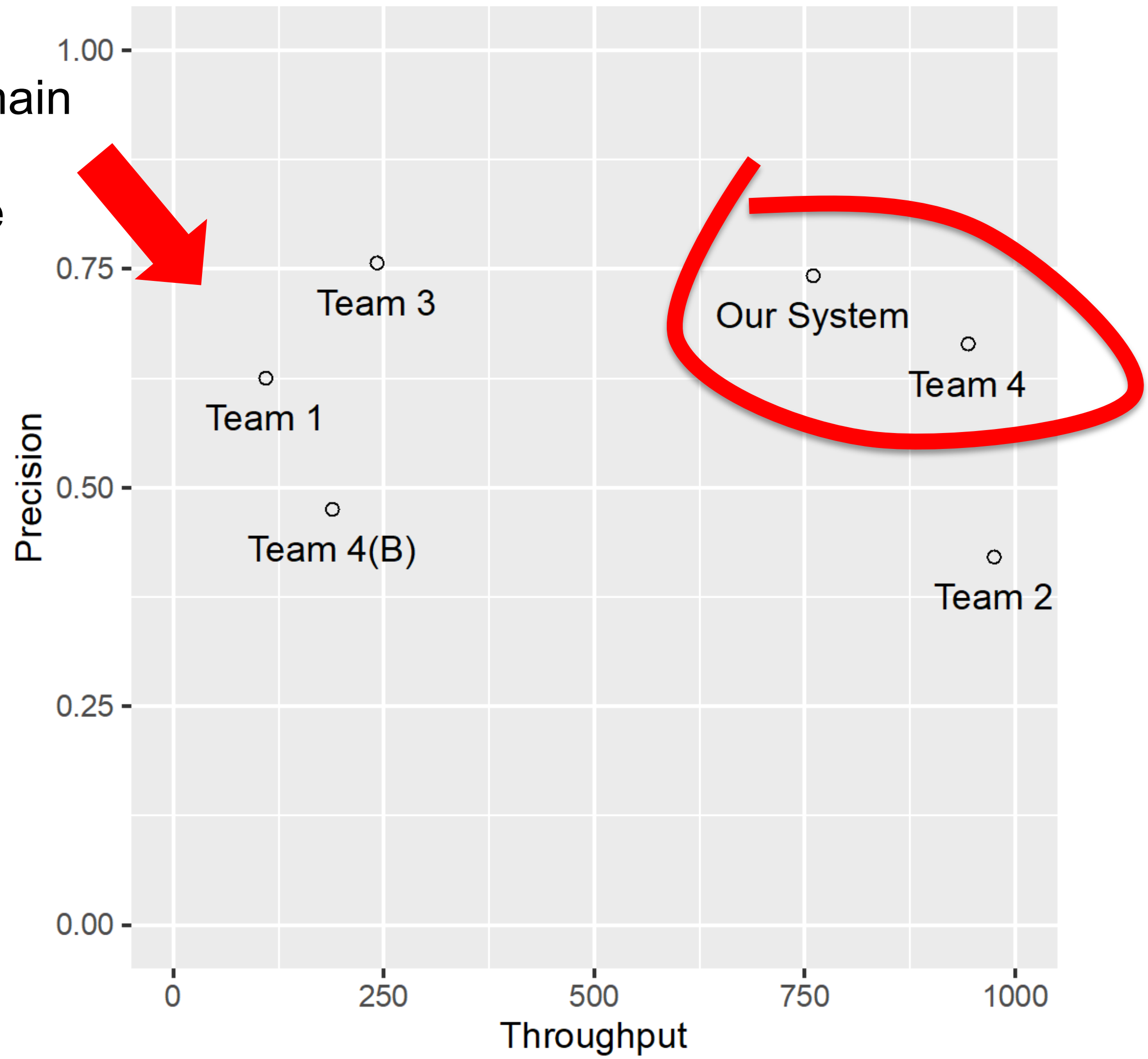
# Few grammar rules

| Type | Syntax | Surface | Total |
|---|---|---|---|
| Entities | 0 | 15 | 15 |
| Generic entities | 0 | 2 | 2 |
| Modifications | 0 | 6 | 6 |
| Mutants | 0 | 9 | 9 |
| Total entities | 0 | 32 | 32 |
| Simple event | [illegible] | [illegible] | 26 |
| Binding | [illegible] | [illegible] | 37 |
| Hydrolysis | [illegible] | [illegible] | 10 |
| Translocatio | [illegible] | [illegible] | 12 |
| Positive regu | [illegible] | [illegible] | 20 |
| Negative regu | [illegible] | [illegible] | 17 |
| Total events | 95 | 27 | 122 |
| Total | 95 | 59 | 154 |

These rules were manually written.
Later I will discuss our work on
using learn such rules using
representation learning.

# How well does machine reading work for a complete reading task?

Human domain experts are around here

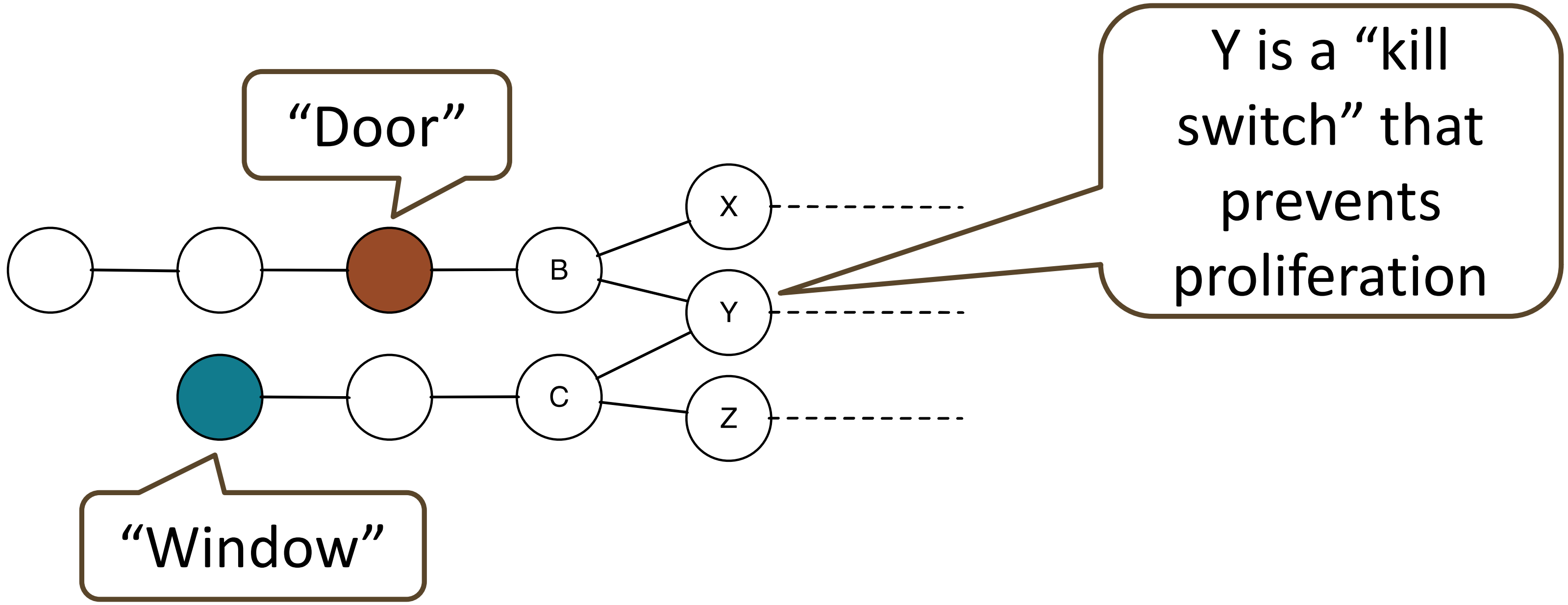


But is machine reading *actually* useful?

# Mutual exclusivity intuition

# The Mutex algorithm (by Emek Demir and colleagues)

Mutex insight: If a tumor "wants" to disable a mechanism, it will mutate something upstream, but it generally won't "pay" for two mutations that do the same thing. So mutually exclusive mutations plus a good model can tell us which mechanisms the tumor disables.






Slide courtesy of: Emek Demir, OHSU

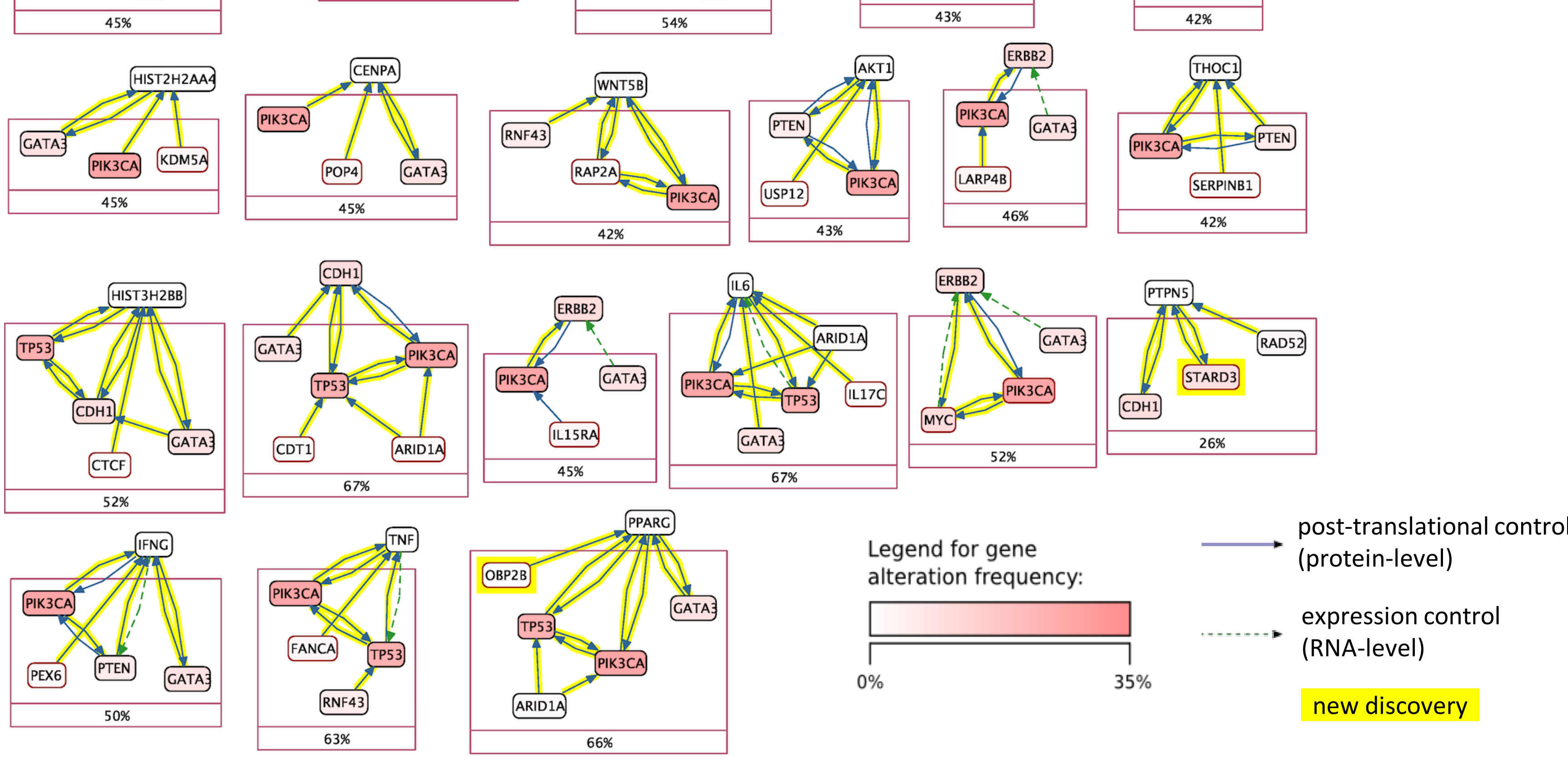
# How Mutex works

Patient data

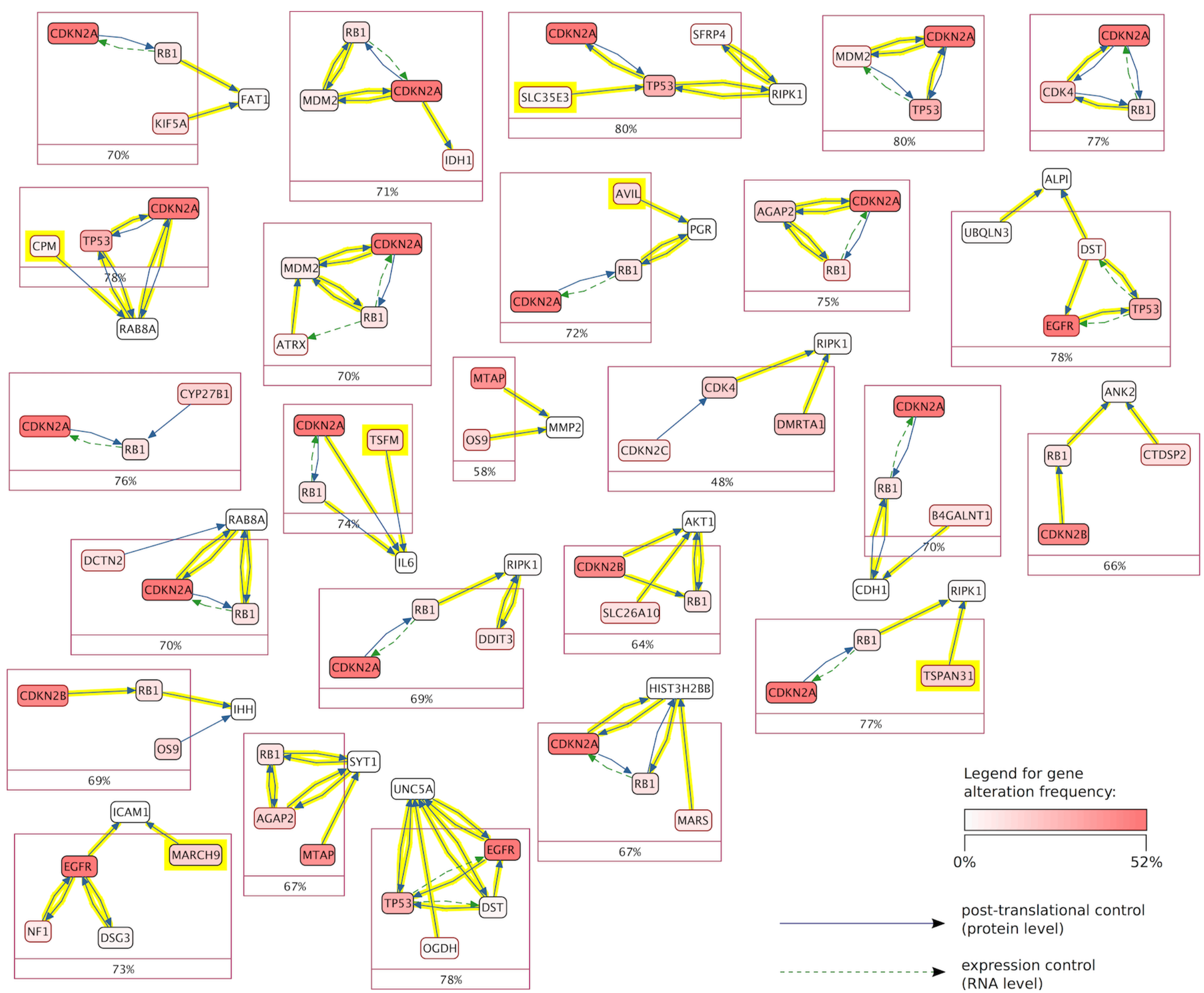
- Mutex does a graph search on the signaling network to find subgraphs of genes that
  - are altered in mutually exclusive manner, and
  - have a common downstream signaling target.

Machine reading contributes here!

# Breast (BRCA)


FOXA1
GATA3
CTCF
CDT1
TP53
HIST3H2BB
45%
RAP2A
B4GALNT3
PIK3CA
GSN
38%
GATA3
TRAF7
LRRK2
CDH1
TP53
54%
WNT5B
RNF43
PIK3CA
IL15RA
43%
PIK3CA
STAT1
PTEN
DUSP3
42%
HIST2H2AA4
GATA3
PIK3CA
KDM5A
45%
CENPA
PIK3CA
POP4
GATA3
45%
WNT5B
RNF43
RAP2A
PIK3CA
42%
AKT1
PTEN
USP12
PIK3CA
43%
ERBB2
PIK3CA
GATA3
LARP4B
46%
THOC1
PIK3CA
PTEN
SERPINB1
42%
HIST3H2BB
TP53
CDH1
GATA3
CTCF
52%
CDH1
GATA3
PIK3CA
TP53
CDT1
ARID1A
67%
ERBB2
PIK3CA
GATA3
IL15RA
45%
IL6
ARID1A
PIK3CA
TP53
IL17C
GATA3
67%
ERBB2
GATA3
PIK3CA
MYC
52%
PTPN5
RAD52
STARD3
CDH1
26%
IFNG
PIK3CA
PEX6
PTEN
GATA3
50%
TNF
PIK3CA
FANCA
TP53
RNF43
63%
PPARG
OBP2B
GATA3
TP53
PIK3CA
ARID1A
66%
Legend for gene alteration frequency:
0%
35%
post-translational control (protein-level)
expression control (RNA-level)
new discovery

# Brain (GBM)



CDKN2A
RB1
FAT1
KIF5A
70%
RB1
MDM2
CDKN2A
IDH1
71%
CDKN2A
SFRP4
TP53
SLC35E3
RIPK1
80%
MDM2
CDKN2A
TP53
80%
CDKN2A
CDK4
RB1
77%
CDKN2A
CPM
TP53
78%
RAB8A
CDKN2A
MDM2
RB1
ATRX
70%
AVIL
PGR
RB1
CDKN2A
72%
AGAP2
CDKN2A
RB1
75%
ALPI
UBQLN3
DST
TP53
EGFR
78%
CYP27B1
CDKN2A
RB1
76%
CDKN2A
TSFM
RB1
74%
IL6
MTAP
OS9
MMP2
58%
RIPK1
CDK4
DMRTA1
CDKN2C
48%
CDKN2A
RB1
B4GALNT1
70%
CDH1
ANK2
RB1
CTDSP2
CDKN2B
66%
RAB8A
DCTN2
CDKN2A
RB1
70%
RIPK1
RB1
DDIT3
CDKN2A
69%
AKT1
CDKN2B
RB1
SLC26A10
64%
RIPK1
RB1
TSPAN31
CDKN2A
77%
CDKN2B
RB1
IHH
OS9
69%
RB1
SYT1
AGAP2
MTAP
67%
HIST3H2BB
CDKN2A
RB1
MARS
67%
ICAM1
MARCH9
EGFR
NF1
DSG3
73%
UNC5A
EGFR
TP53
DST
OGDH
78%
Legend for gene alteration frequency:
0%
52%
post-translational control (protein level)
expression control (RNA level)

Machine reading suggests novel hypotheses that are missed by the authors of the individual publications.

# LEARNING INTERPRETABLE MODELS

# Motivation for rule learning

- It took us 1 – 2 person months to build the grammar of 150+ rules for the biomedical domain

- In many cases, one does not have this time
  - Scenario 1: But training data exists
  - Scenario 2: And training data does not exist either…

# Motivation

- It took us 1 – 2 person months to build the grammar of 154 rules for the biomedical domain

- In many cases, one does not have this time
  - **Scenario 1: But training data exists (2 papers)**
  - Scenario 2: And training data does not exist either…

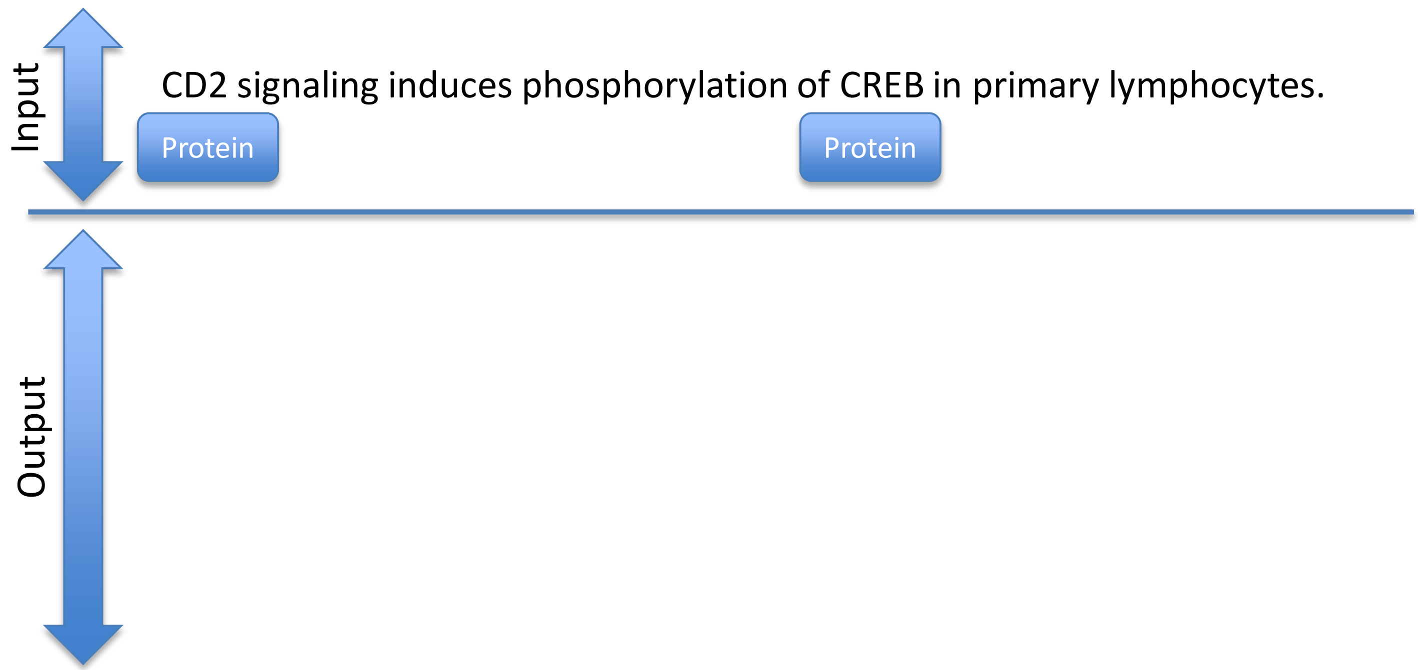
# SnapToGrid: From Statistical to Interpretable Models for Biomedical Information Extraction

**Marco A. Valenzuela-Escárcega, Gus Hahn-Powell, Dane Bell, Mihai Surdeanu**
University of Arizona
Tucson, AZ 85721, USA
{marcov, hahnpowell, dane, msurdeanu}@email.arizona.edu

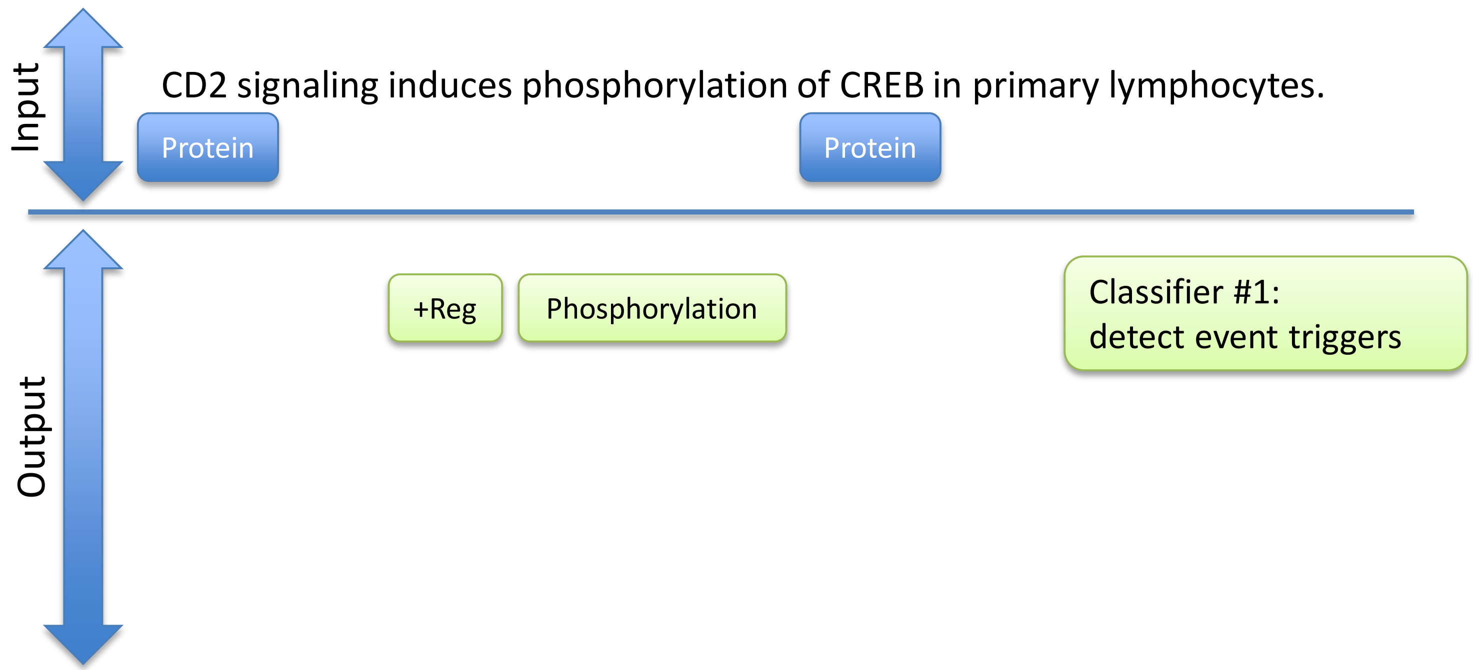
# A three-step process

1. Train a statistical classifier
   a) Aggressive feature selection using regularization
2. Convert model to rules
   a) Convert features to rules
   b) “Snap to grid”: throw away most statistics by discretizing feature weights
3. Model editing

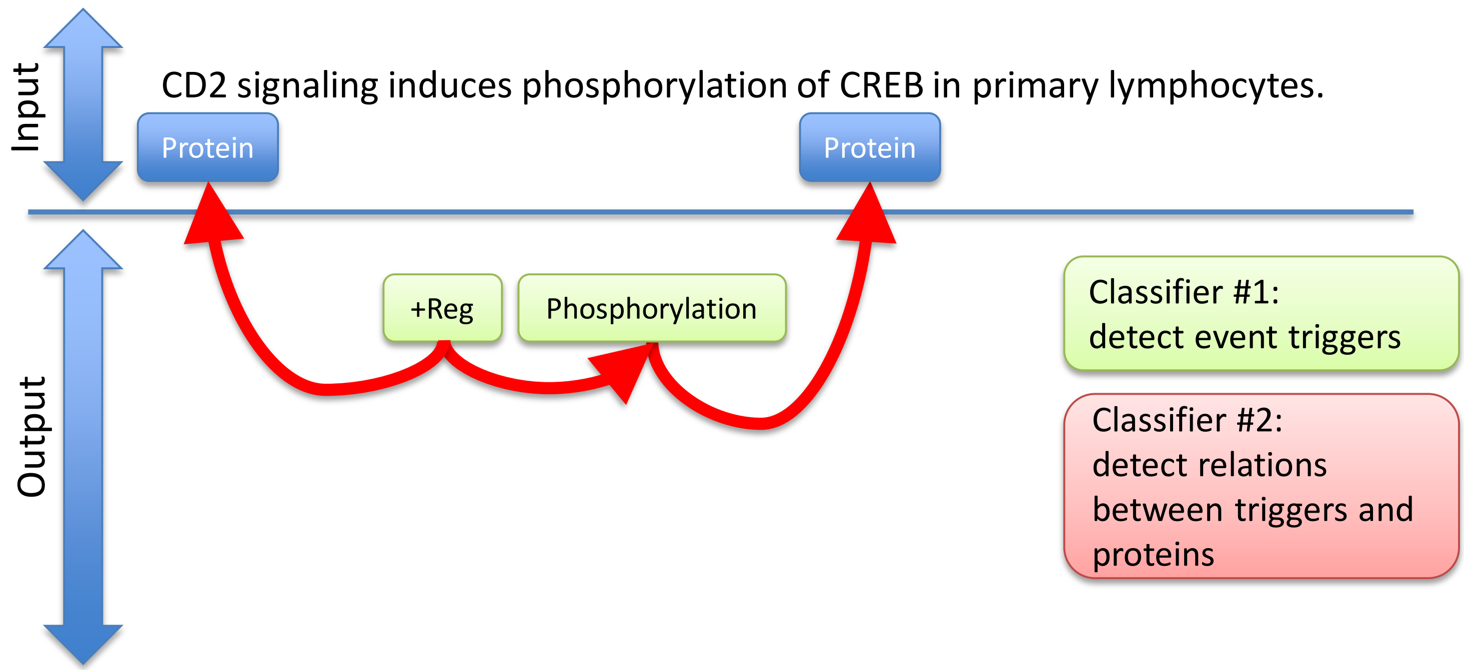
# BioNLP 2009 Shared Task



# BioNLP 2009 Shared Task



# BioNLP 2009 Shared Task



# Step 1: train statistical classifiers

- We use logistic regression for both classifiers
- Feature selection through regularization
  - L1 regularization: aggressive feature selection; slightly lower performance

# Machine learning features

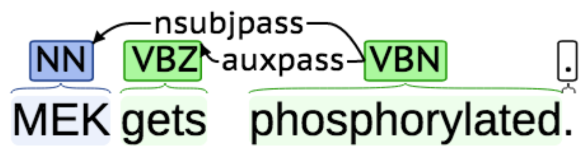
**Triggers**

- Token
  - Word, lemma, gazetteer
- Surface
  - Token features for a window around token of interest
  - Bigrams
- Syntax
  - Dependency paths up to depth 2
  - Token features for token at the end of each path
- Bag-of-words and entity count
  - For whole sentence
  - For window surrounding token of interest

**Relations**

- Path
  - Shortest dependency path between nodes of interest
  - With and without lexicalization
  - Path length
- Surface
  - Words surrounding and between nodes of interest
- Consistency
  - Soft constraints on edges between trigger and arguments (e.g., only regulations have causes)
- N-gram
  - Dependency, token, dependency
  - Token, dependency, token

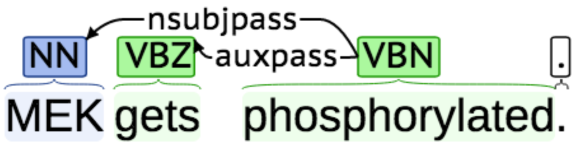
# Step 2: Convert model features to rules




"Passive subject of a phosphorylation trigger that is a protein"

- Features are just patterns
- We simply rewrote them using Odin syntax

# Step 2: Convert model features to rules

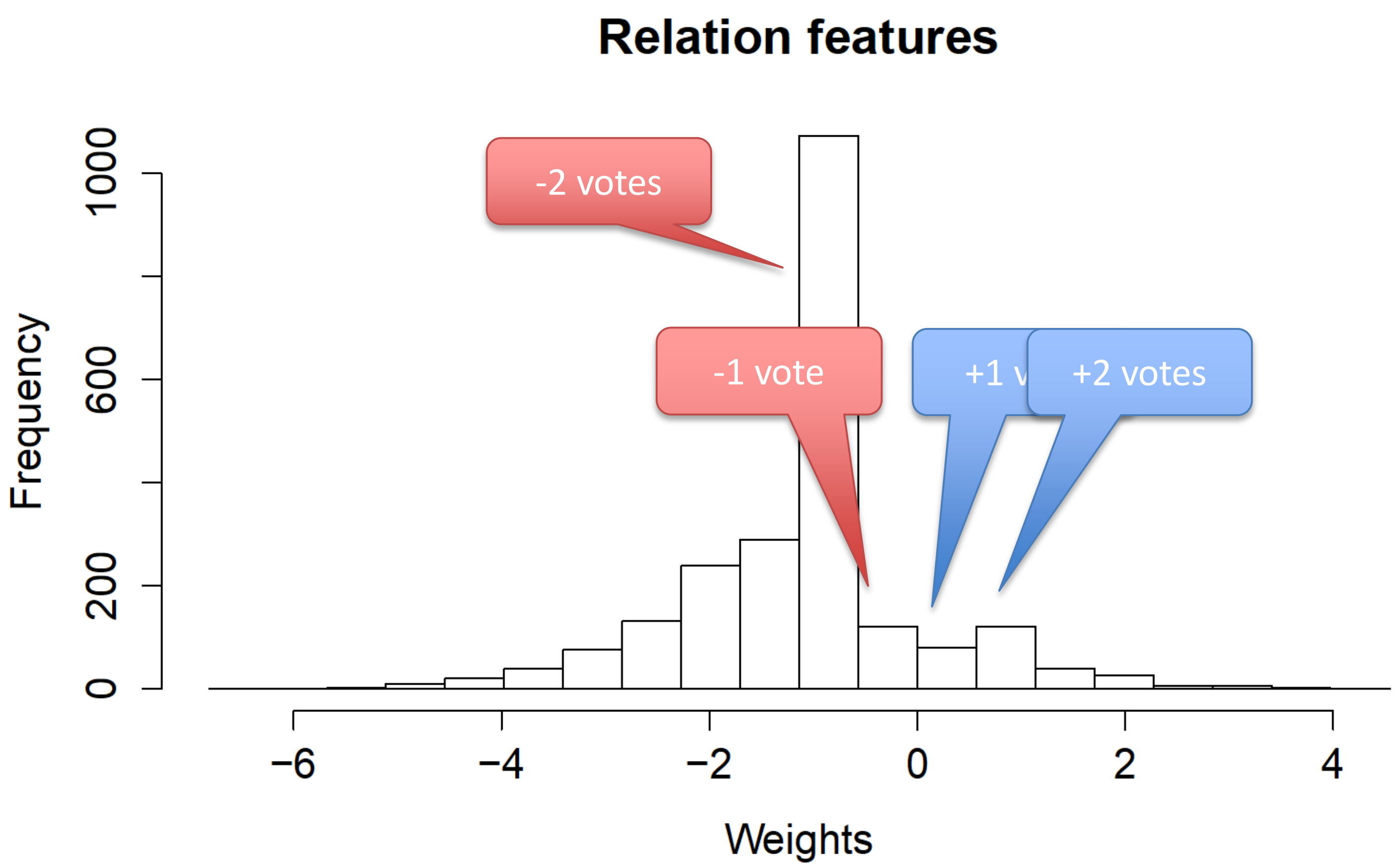



```
-    type : dependency
  label: Phosphorylation
  pattern: |
    trigger:Phosphorylation
    theme:Protein = >nsubjpass
```

# Step 2: Convert model features to rules

- We know have a decision list classifier. But:
- Feature weights are unbounded continuous values
  - Useful for resolving conflicts
  - But nearly impossible to understand/modify
- Rules still need to vote
  - We normalize and discretize weights ("votes") using Scott's rule (used for the generation of bins in histograms)

# Step 2: Convert model features to rules



# Step 2: Convert model features to rules

- Binding
  - PROTEIN recruits PROTEIN
- Localization
  - PROTEIN is recruited to the cytoplasm

```
# vote: +2
- name: Binding_1
  label: Binding
  type: token
  action: countMentions
  pattern: |
    [lemma=recruit & tag=/^(V|N|J)/]
```

```
# vote: +1
- name: Localization_1
  label: Localization
  type: token
  action: countMentions
  pattern: |
    [lemma=recruit & tag=/^(V|N|J)/]
```

# Step 3: Model editing

- Two linguists were given the task of improving the generated rules

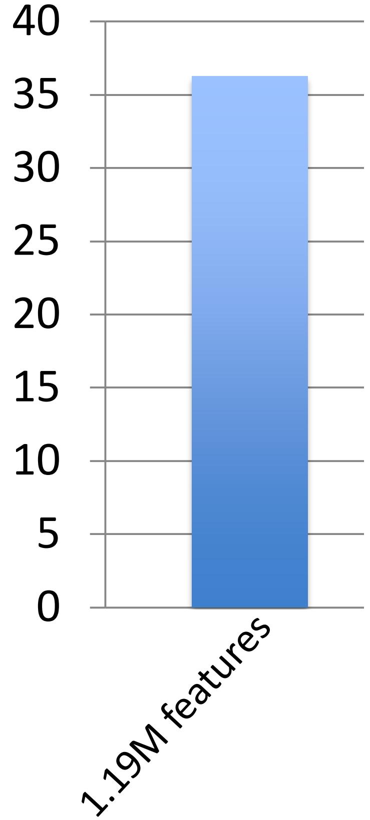
- Constraints:
  - Only have access to the model. No peeking at the training data
  - Approximately one hour to work on the task

# Expert recommendations

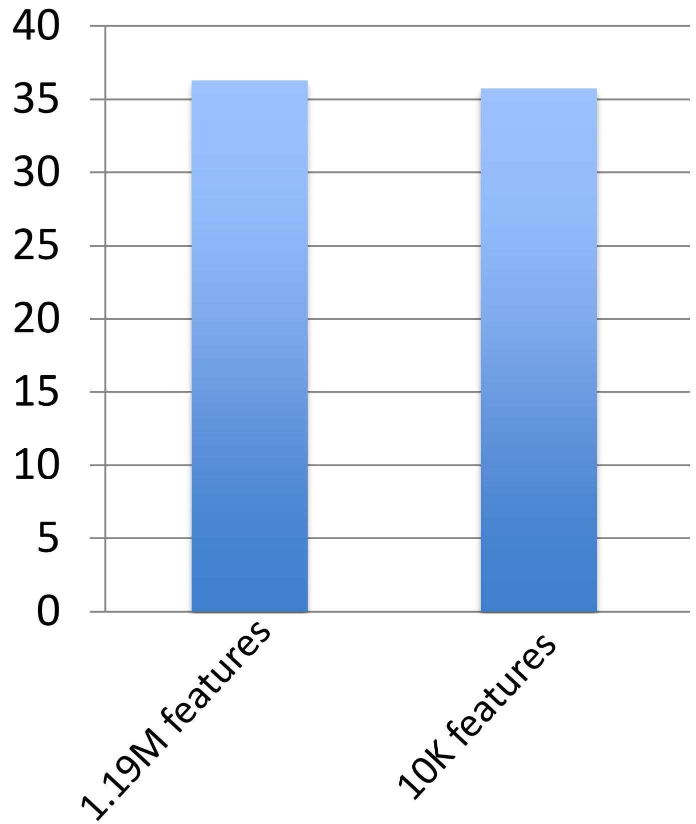
- Generalize syntactic patterns, e.g., participants in events may be heads or modifiers of noun phrases
  - E.g.: “K-Ras” or “the K-Ras protein”
- Eliminate trigger rules that were not sufficiently discriminative
- Make rules robust to common parsing mistakes

- (22 total specific recommendations)

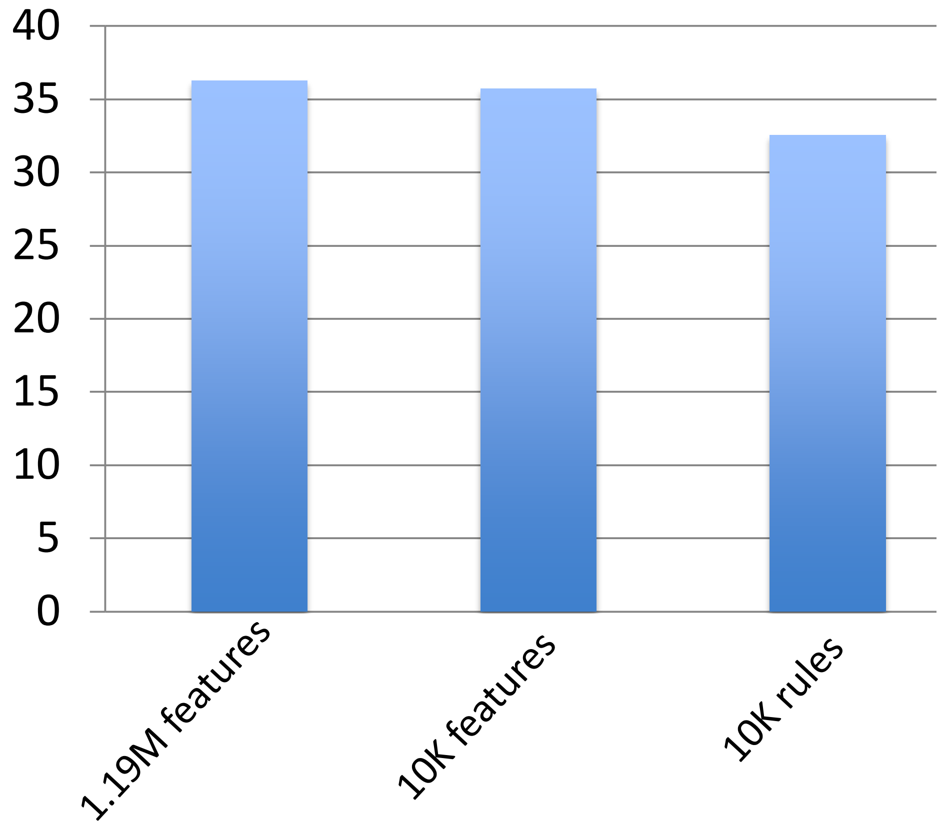
# Results on BioNLP 2009



# Results on BioNLP 2009

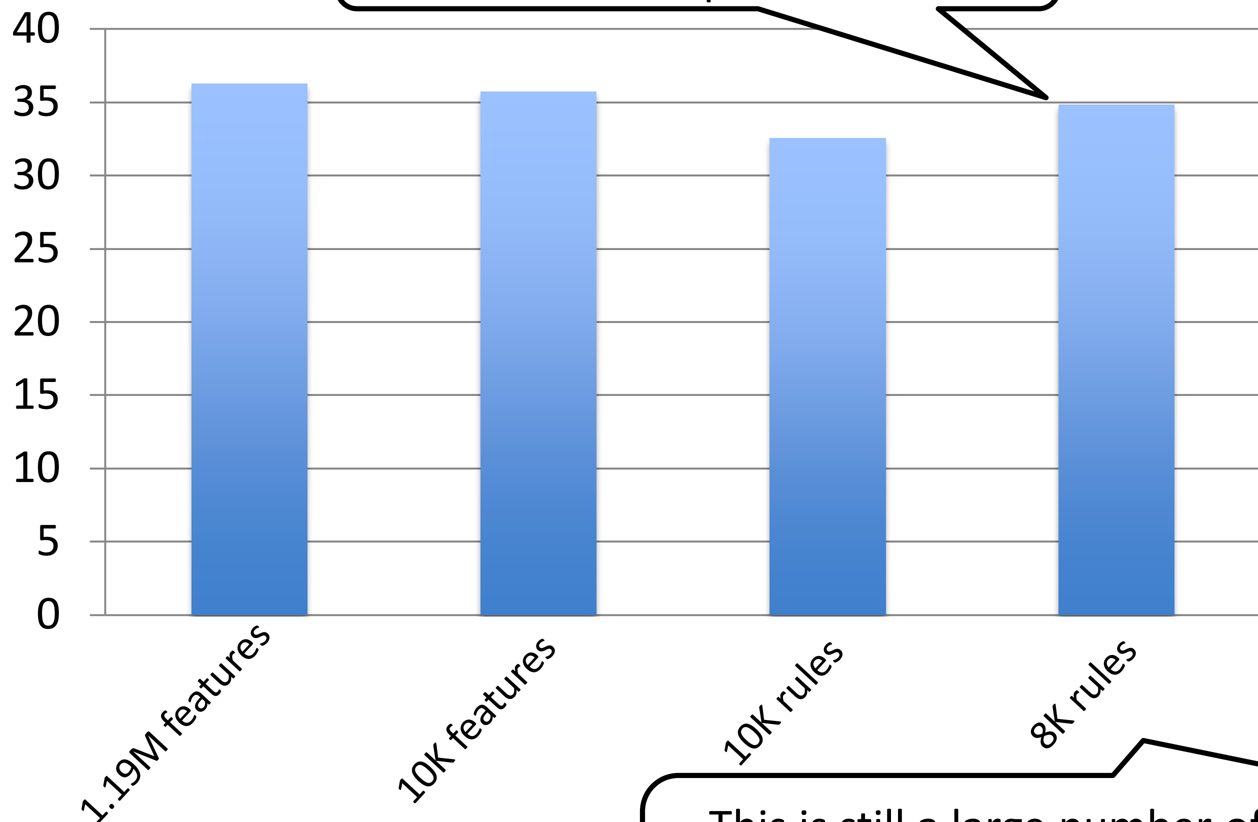


# Results on BioNLP 2009



# Results on BioNLP 2009

Nearly the same performance.
But with an interpretable model!




This is still a large number of rules.
We can reduce them by > 1 order of magnitude using automata theory.

# Summary

- Converting statistical models into a deterministic decision list classifier does impact performance negatively
- But keeping the human in the loop allows us to recover almost all the lost performance. And we end up with an interpretable model!

Fine. But how do you do this with neural networks that do not have explicit features?

# Exploring Interpretability in Event Extraction: Multitask Learning of a Neural Event Classifier and an Explanation Decoder

**Zheng Tang** [*], **Gus Hahn-Powell** [†], **Mihai Surdeanu** [*]
[*] Department of Computer Science
[†] Department of Linguistics
University of Arizona
{zhengtang, hahnpowell, msurdeanu}@email.arizona.edu

# Intuition: jointly training for prediction and interpretability!

1. Use neural classifiers instead of the LRs in the previous paper ☺
2. Decode rules from natural language texts (reusing ideas from machine translation)
   – Source language – original texts
   – "Target language" – grammar rule that matches
3. Train them jointly

# Encoder-decoder neural architectures for machine translation

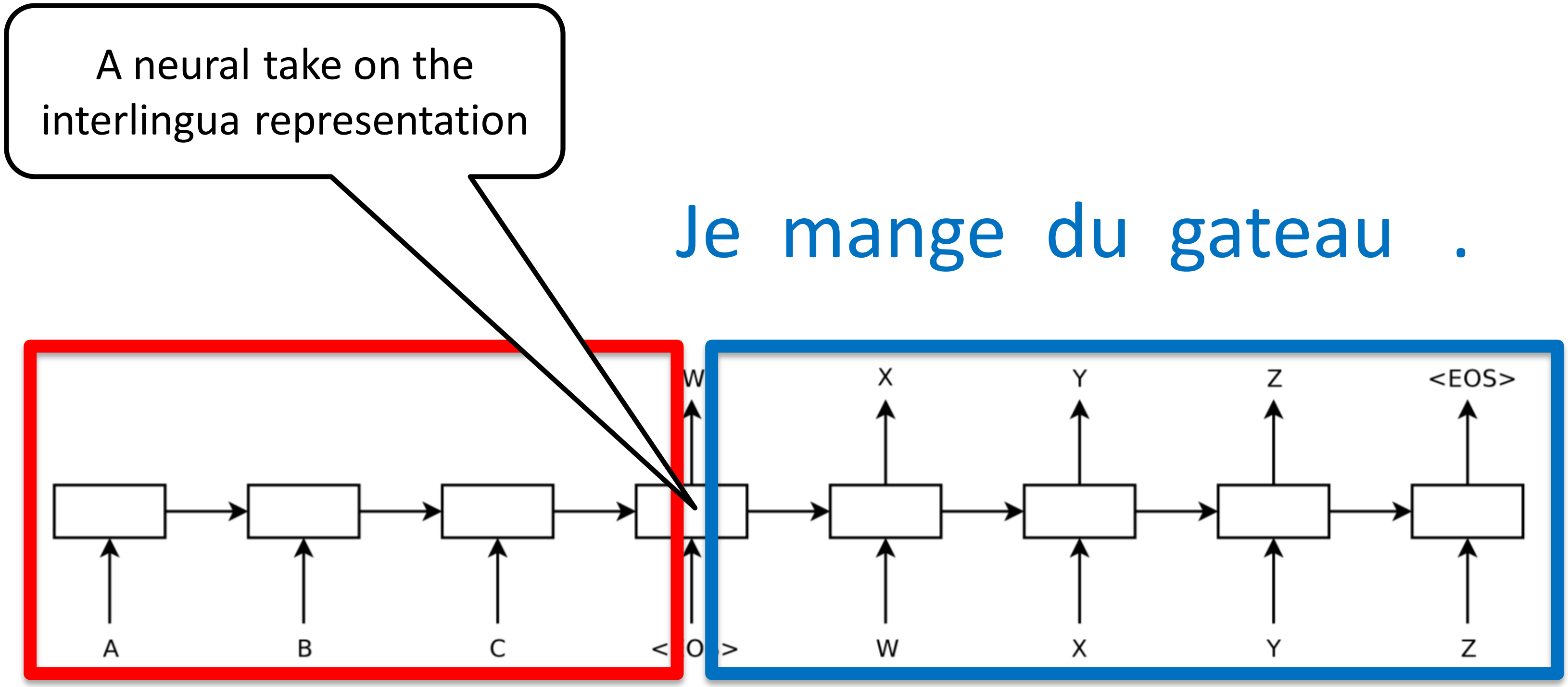



Figure 1: Our model reads an input sentence “ABC” and produces “WXYZ” as the output sentence. The model stops making predictions after outputting the end-of-sentence token. Note that the LSTM reads the input sentence in reverse, because doing so introduces many short term dependencies in the data that make the optimization problem much easier.

I eat cake .

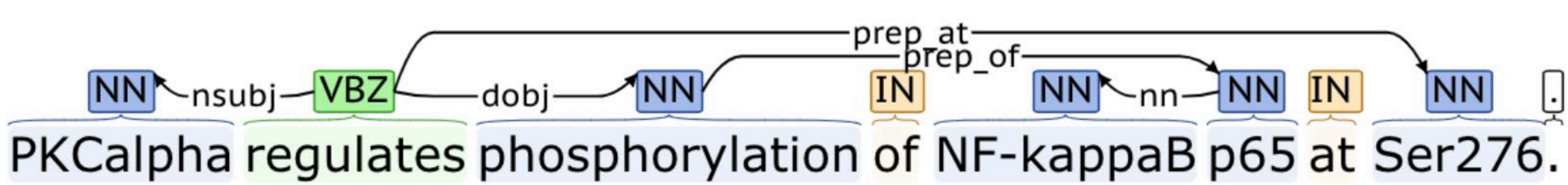
# Our “source” and “target languages”

(target)

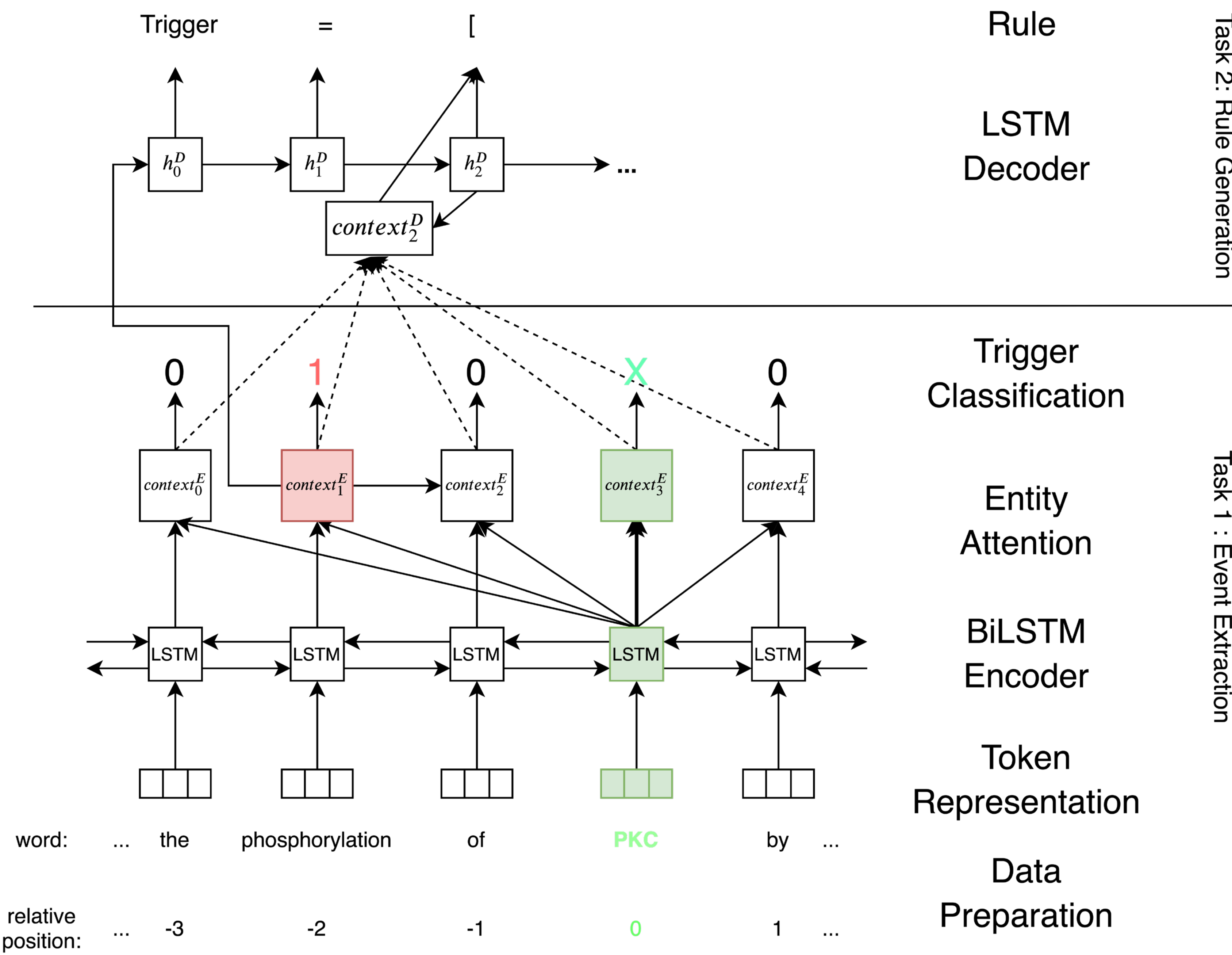
Phosphorylation_syntax_1a_noun:

trigger = [lemma=/phosphorylation/ !word=/(?i)^(de|auto)/]

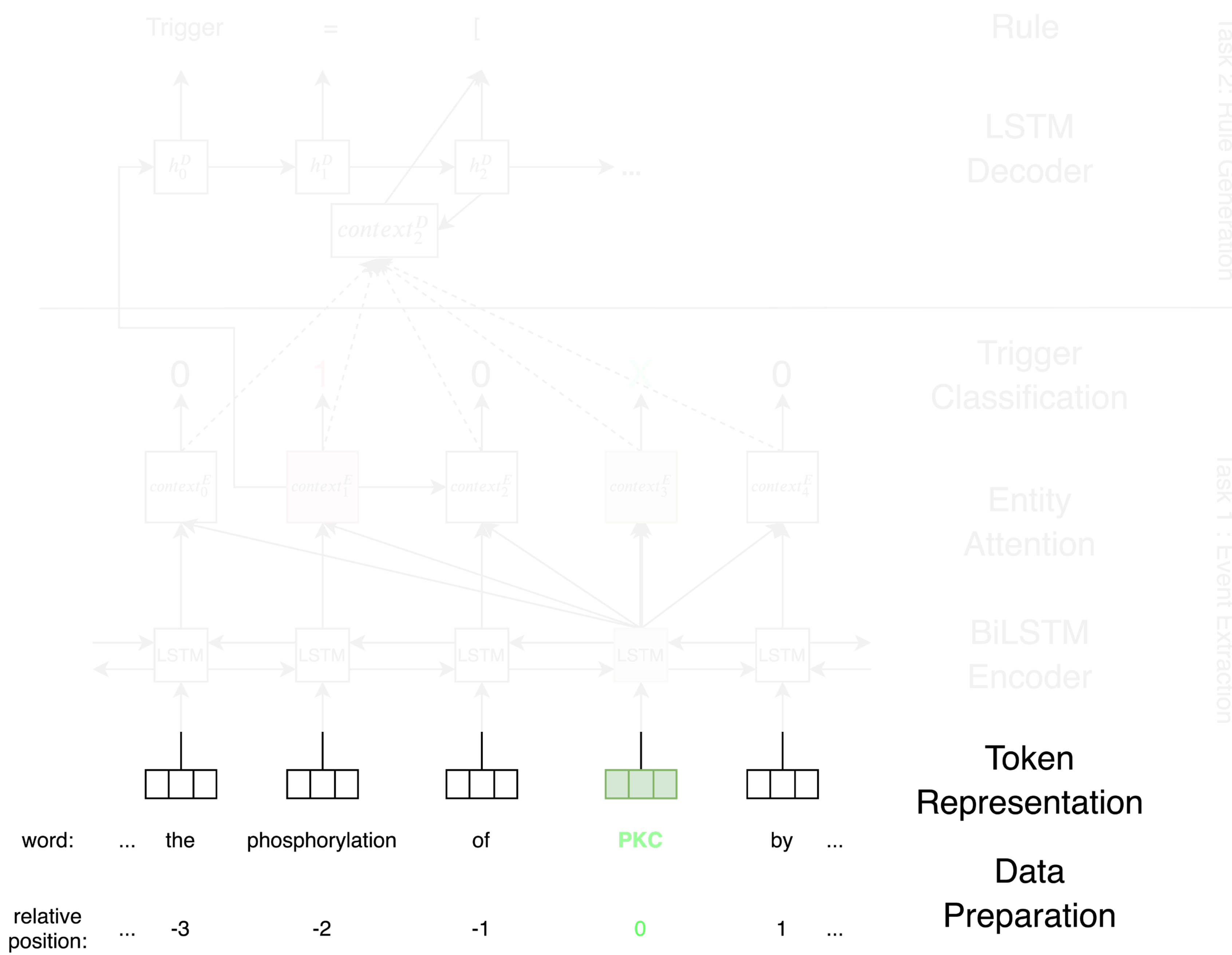
theme:BioChemicalEntity = prep_of appos? /nn|conj_(and|or|nor)|cc/{,2}

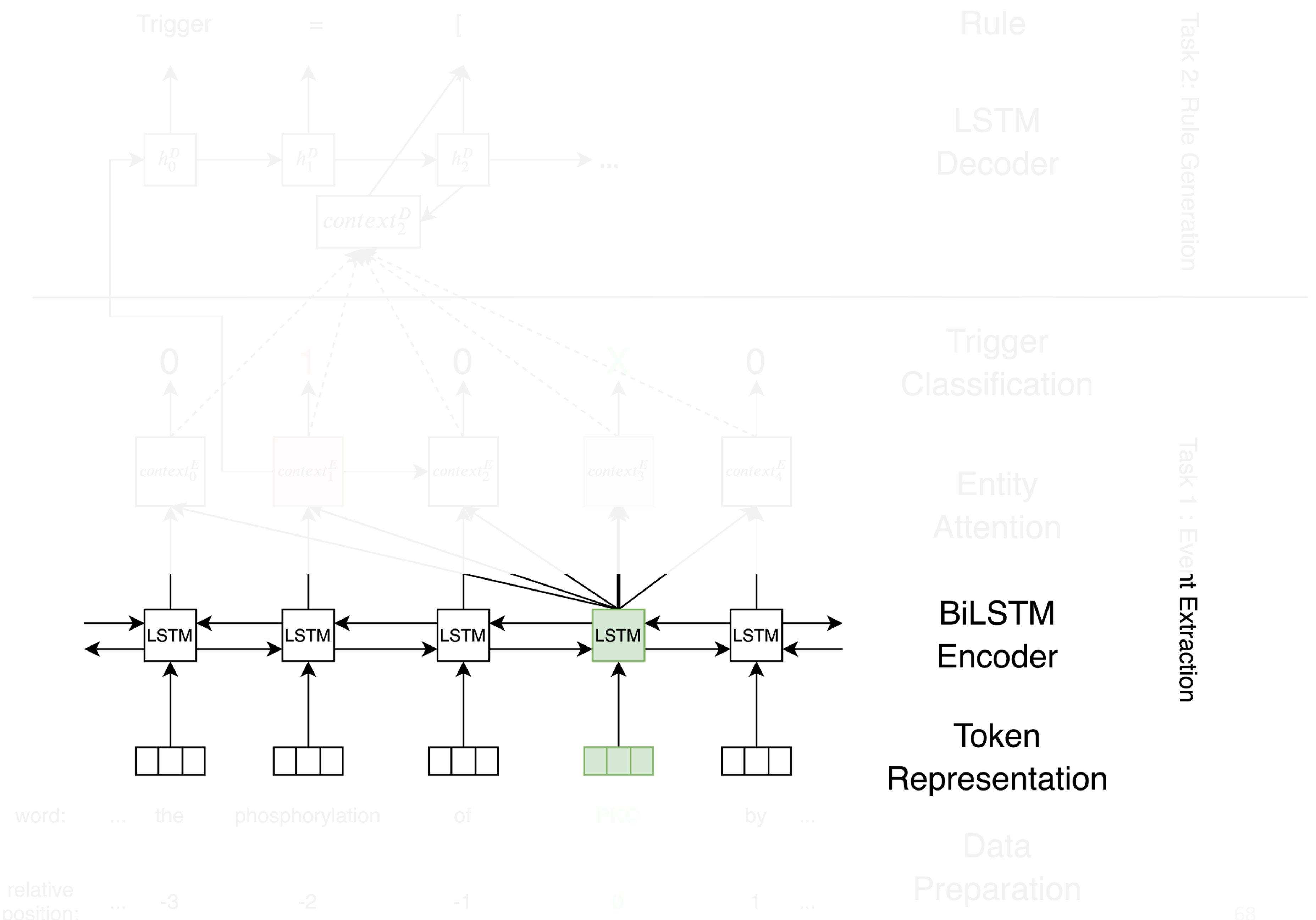



(source)

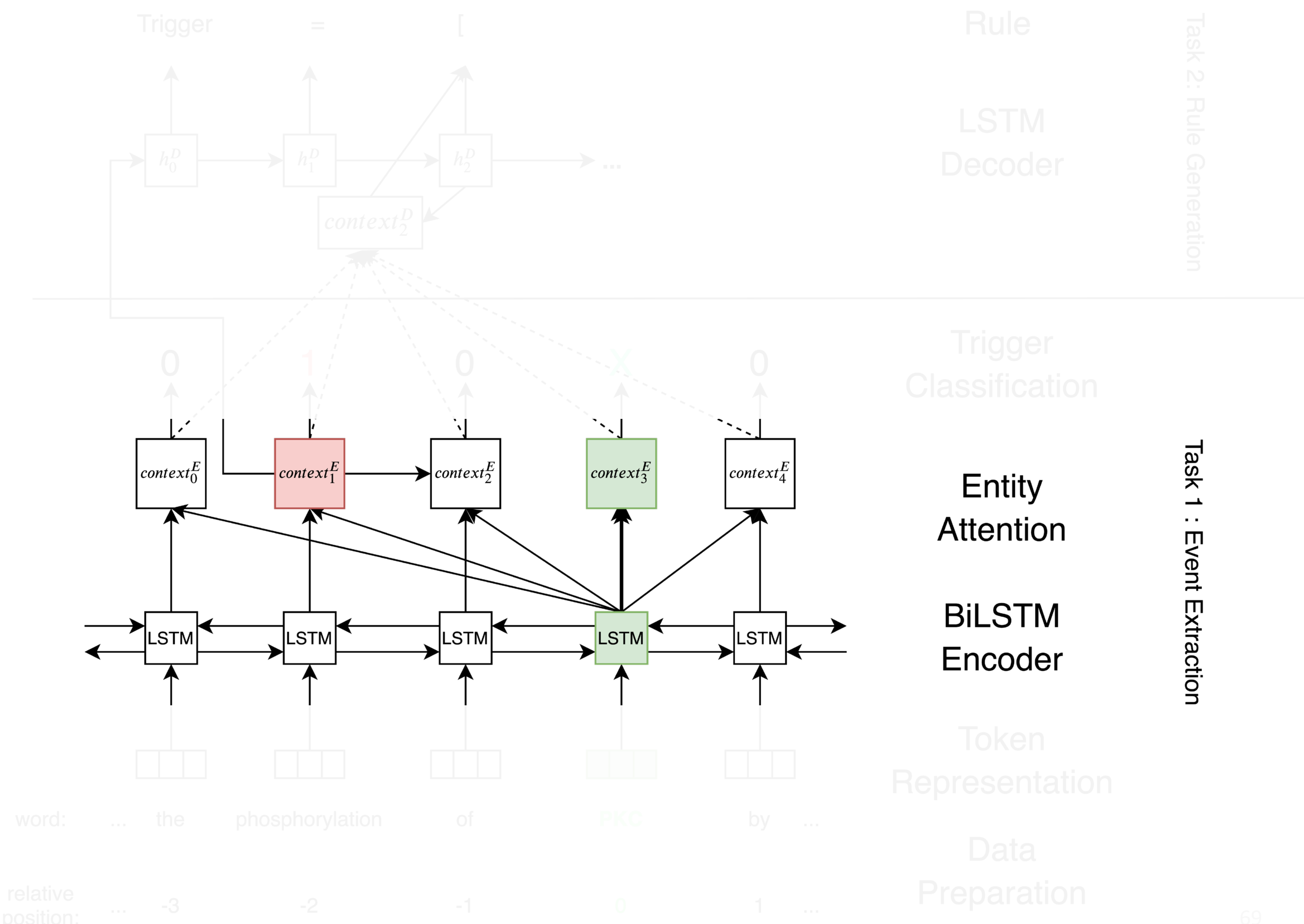


Trigger
=
[
Rule
LSTM Decoder
Task 2: Rule Generation
$h_0^D$
$h_1^D$
$h_2^D$
...
$context_2^D$
0
1
0
X
0
Trigger Classification
$context_0^E$
$context_1^E$
$context_2^E$
$context_3^E$
$context_4^E$
Entity Attention
Task 1 : Event Extraction
LSTM
LSTM
LSTM
LSTM
LSTM
BiLSTM Encoder
Token Representation
word: ... the phosphorylation of PKC by ...
relative position: ... -3 -2 -1 0 1 ...
Data Preparation

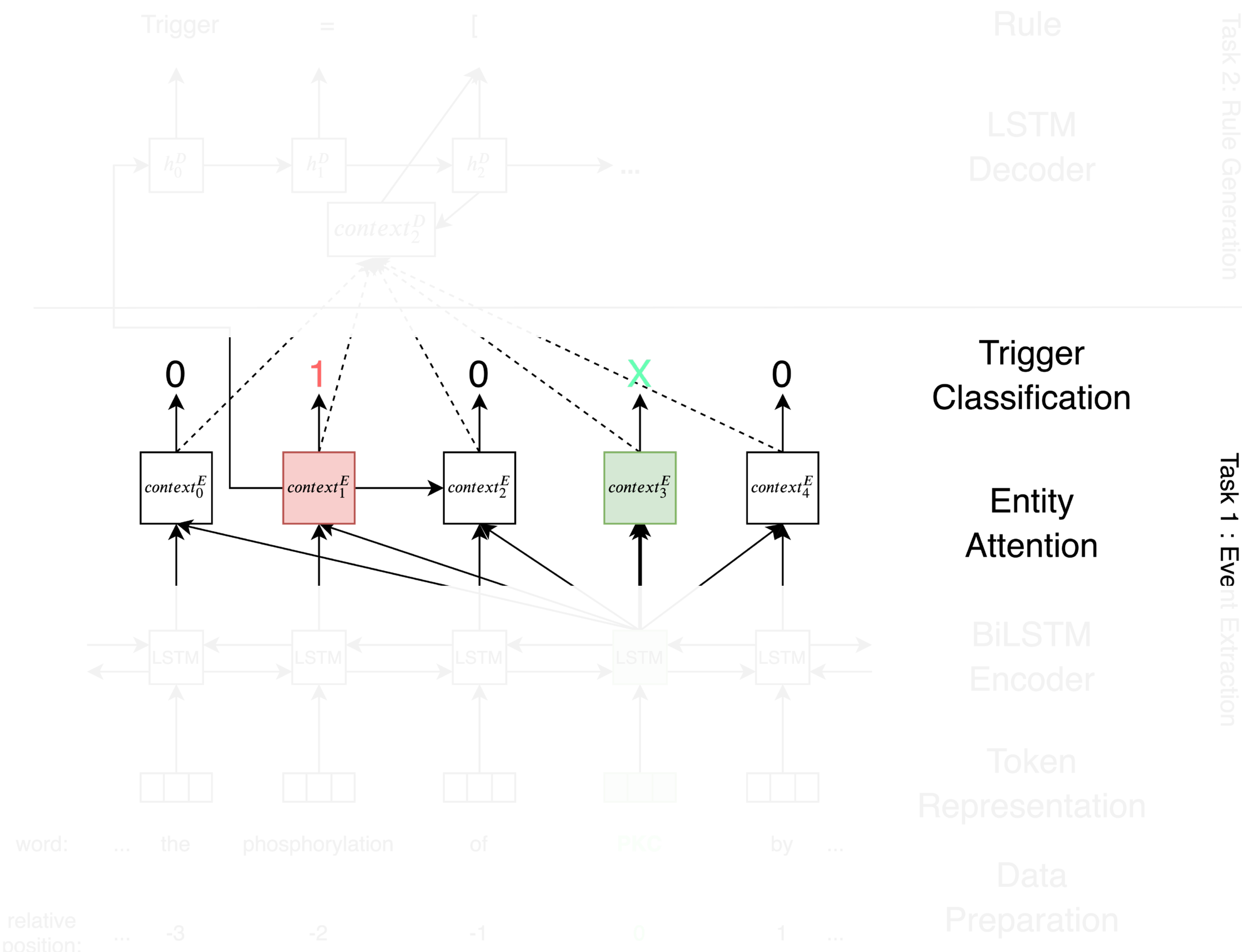

Trigger
=
[
Rule
LSTM Decoder
Task 2: Rule Generation
$h_0^D$
$h_1^D$
$h_2^D$
...
$context_2^D$
0
1
0
X
0
Trigger Classification
$context_0^E$
$context_1^E$
$context_2^E$
$context_3^E$
$context_4^E$
Entity Attention
LSTM
LSTM
LSTM
LSTM
LSTM
BiLSTM Encoder
Task 1 : Event Extraction
Token Representation
word: ... the phosphorylation of PKC by ...
Data Preparation
relative position: ... -3 -2 -1 0 1 ...

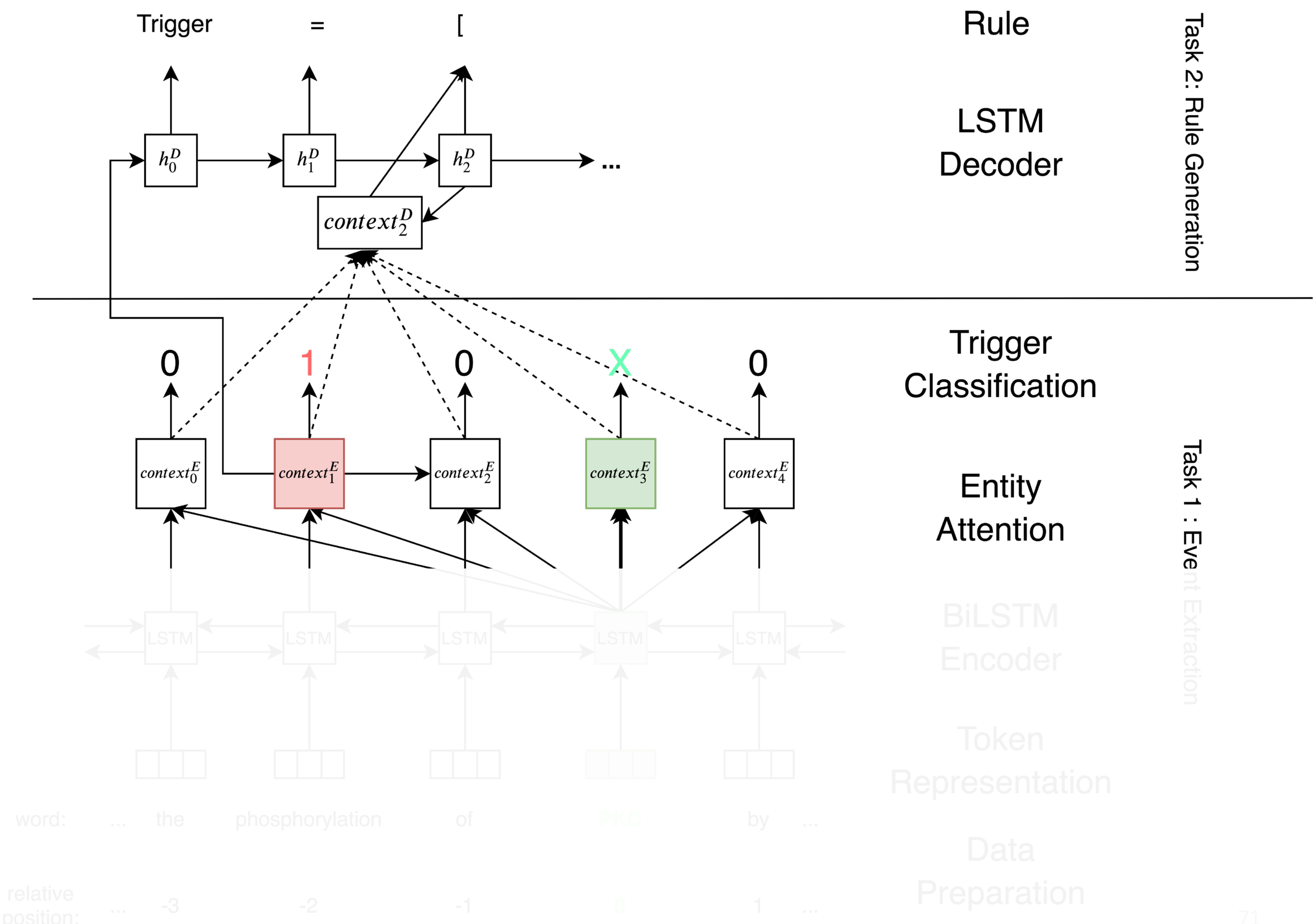


Trigger
=
[
Rule
LSTM
Decoder
Task 2: Rule Generation
Trigger
Classification
0
1
0
X
0
Entity
Attention
Task 1 : Event Extraction
LSTM
BiLSTM
Encoder
Token
Representation
word:
... the phosphorylation of by ...
Data
Preparation
relative
position:
... -3 -2 -1 0 1 ...

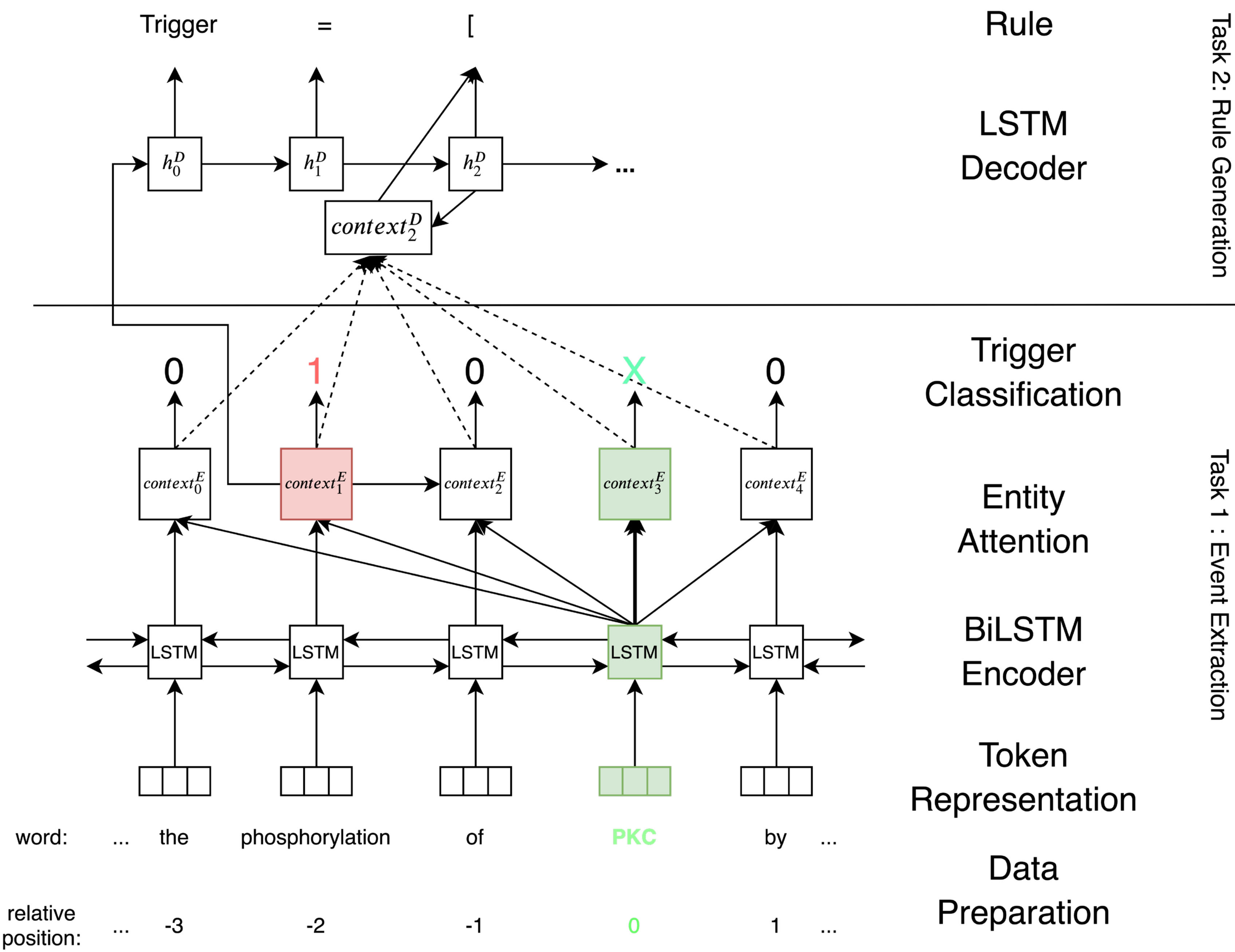


Trigger
=
[
Rule
LSTM Decoder
Task 2: Rule Generation
Trigger Classification
0
1
0
X
0
Entity Attention
BiLSTM Encoder
Task 1 : Event Extraction
LSTM
Token Representation
word: ... the phosphorylation of PKC by ...
Data Preparation
relative position: ... -3 -2 -1 0 1 ...


Trigger
=
[
Rule
LSTM Decoder
Task 2: Rule Generation
$h_0^D$
$h_1^D$
$h_2^D$
...
$context_2^D$
0
1
0
X
0
Trigger Classification
$context_0^E$
$context_1^E$
$context_2^E$
$context_3^E$
$context_4^E$
Entity Attention
Task 1 : Event Extraction
LSTM
LSTM
LSTM
LSTM
LSTM
BiLSTM Encoder
Token Representation
word:
...
the
phosphorylation
of
PKC
by
...
Data Preparation
relative position:
...
-3
-2
-1
0
1
...



Trigger
=
[
Rule
LSTM Decoder
Task 2: Rule Generation
$h_0^D$
$h_1^D$
$h_2^D$
...
$context_2^D$
Trigger Classification
0
1
0
X
0
$context_0^E$
$context_1^E$
$context_2^E$
$context_3^E$
$context_4^E$
Entity Attention
Task 1 : Event Extraction
LSTM
BiLSTM Encoder
Token Representation
word:
...
the
phosphorylation
of
PKC
by
...
Data Preparation
relative position:
-3
-2
-1
0
1



Trigger
=
[
Rule
$h_0^D$
$h_1^D$
$h_2^D$
...
LSTM
Decoder
$context_2^D$
Task 2: Rule Generation
0
1
0
X
0
Trigger
Classification
$context_0^E$
$context_1^E$
$context_2^E$
$context_3^E$
$context_4^E$
Entity
Attention
LSTM
LSTM
LSTM
LSTM
LSTM
BiLSTM
Encoder
Task 1 : Event Extraction
Token
Representation
word: ... the phosphorylation of PKC by ...
relative position: ... -3 -2 -1 0 1 ...
Data
Preparation

# Data

- For event classification: BioNLP 2013 (similar to the data used in the previous paper)
- For rule decoding: pairs of (sentence, rule) extracted by our machine reading system with manually written rules
  - Some come from the BioNLP dataset and are aligned with the gold annotations in BioNLP 2013
  - Approximately 15K pairs come from other publications ("silver" data)

# Results for event classification

We extract 3 types of events

| | Phosphorylation (P) | | | Localization (L) | | | Gene Expression (GE) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1 | Precision | Recall | F1 | Precision | Recall | F1 |
| Rule baseline | 92.68 | 48.12 | 63.35 | 66.13 | 44.44 | 53.16 | 51.08 | 69.79 | 58.98 |
| T1 | 87.78 | 49.38 | 63.20 | 100.00 | 4.04 | 7.77 | 89.32 | 64.30 | 74.77 |
| T1 + Silver | 62.75 | 82.50 | 71.28 | 54.55 | 34.34 | 42.15 | 68.43 | 74.31 | 71.25 |
| T1 + Silver + T2 | 84.38 | 68.75 | 75.77 | 76.60 | 39.39 | 52.03 | 69.92 | 71.24 | 70.58 |

- Rule baseline – system with manually written rules
- T1 – supervised neural event classifier
- T1 + Silver – semi-supervised neural event classifier
- T1 + Silver + T2 - semi-supervised neural event classifier trained jointly with the rule decoder

# Results for event classification

| | Phosphorylation (P) | | | Localization (L) | | | Gene Expression (GE) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1 | Precision | Recall | F1 | Precision | Recall | F1 |
| Rule baseline | 92.68 | 48.12 | 63.35 | 66.13 | 44.44 | 53.16 | 51.08 | 69.79 | 58.98 |
| T1 | 87.78 | 49.38 | 63.20 | 100.00 | 4.04 | 7.77 | 89.32 | 64.30 | 74.77 |
| T1 + Silver | 62.75 | 82.50 | 71.28 | 54.55 | 34.34 | 42.15 | 68.43 | 74.31 | 71.25 |
| T1 + Silver + T2 | 84.38 | 68.75 | 75.77 | 76.60 | 39.39 | 52.03 | 69.92 | 71.24 | 70.58 |

Semi-supervised learning helps!

# Results for event classification

| | Phosphorylation (P) | | | Localization (L) | | | Gene Expression (GE) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1 | Precision | Recall | F1 | Precision | Recall | F1 |
| Rule baseline | 92.68 | 48.12 | 63.35 | 66.13 | 44.44 | 53.16 | 51.08 | 69.79 | 58.98 |
| T1 | 87.78 | 49.38 | 63.20 | 100.00 | 4.04 | 7.77 | 89.32 | 64.30 | 74.77 |
| T1 + Silver | 62.75 | 82.50 | 71.28 | 54.55 | 34.34 | 42.15 | 68.43 | 74.31 | 71.25 |
| T1 + Silver + T2 | 84.38 | 68.75 | 75.77 | 76.60 | 39.39 | 52.03 | 69.92 | 71.24 | 70.58 |

Jointly training for prediction and interpretability helps prediction!

# Results for rule decoding

| | BLEU | Exact Matches | Non-exact Explainable Matches |
|---|---|---|---|
| P | 93.80 | 86.11 | 2/15 |
| L | 83.78 | 84.33 | 1/9 |
| GE | 78.99 | 76.45 | 10/43 |

# Rule error analysis

<table>
<tr><th>Hand-written Rule<br>Go to page 8</th><th>Decoded Rule</th></tr>
<tr><td>trigger = [ lemma = /phosphorylate/ & ! word = /(?i)^(de|auto)/ & tag = /^(V|JJ)/ & ! mention = ModificationTrigger ]<br><br>theme : BioChemicalEntity = > nsubjpass prep_of ? /conj_(and|or|nor)|nn|cc/ { , 2 }</td><td>trigger = [ lemma = /phosphorylate/ & ! word = /(?i)^(de|auto)/ & tag = /^(V|JJ)/ & ! = ModificationTrigger ]<br><br>theme : BioChemicalEntity = > nsubjpass prep_of ? /conj_(and|or|nor)|nn|cc/ { , 2 }</td></tr>
<tr><td>trigger = [ lemma = /detect|localiz|locat|releas|secret|translocat/ & ! word = /(?i)^de/ ]<br><br>theme : BioChemicalEntity = prep_of ? appos ? /conj_(and|or|nor)|cc|nn/ { , 2 }</td><td>trigger = [ lemma = /detect|localiz|locat|releas|secret|translocat/ & ! word = /(?i)^de/ & ! outgoing = /prep_(by|of)/ ]<br><br>theme : BioChemicalEntity = < /conj_(and|or|nor)/ ? /conj_(and|or|nor)|cc|nn|prep_of/ { , 2 }</td></tr>
<tr><td>trigger = [ lemma = / phosphorylation / & ! word = / ( ? i ) ^ ( de | auto ) / & ! outgoing = / prep_ ( by | of ) / ]<br><br>theme : BioChemicalEntity = < / conj_ ( and | or | nor ) / ? / conj_ ( and | or | nor ) | cc | nn | prep_of / { 2 } site : Site ? = nn | < dobj ? / prep_ ( at | on ) / num ?</td><td>trigger = [ lemma = / phosphorylate / & ! word = / ( ? i ) ^ ( de | auto ) / & tag = / ^ ( V | JJ ) / & ! mention = ModificationTrigger ]<br><br>cause : BioChemicalEntity ? = < xcomp ? ( nsubj | agent | < vmod ) / appos | nn | conj_ ( and | or | nor ) | cc / { 2 }<br><br>theme : BioChemicalEntity = ( dobj | xcomp ) / conj_ ( and | or | nor ) | dep | cc | nn | prep_of / { 2 } ( > > [ word = by ] ) { 2 } site : Site ? = dobj ? / prep_ ( at | on ) | nn | conj_ ( and | or | nor ) | cc / { 2 }</td></tr>
</table>

**Green** – missed text in the decoded rule
**Red** – hallucinated text

# Summary

- A neural approach that jointly trains for prediction and interpretability
- The joint training improves prediction!

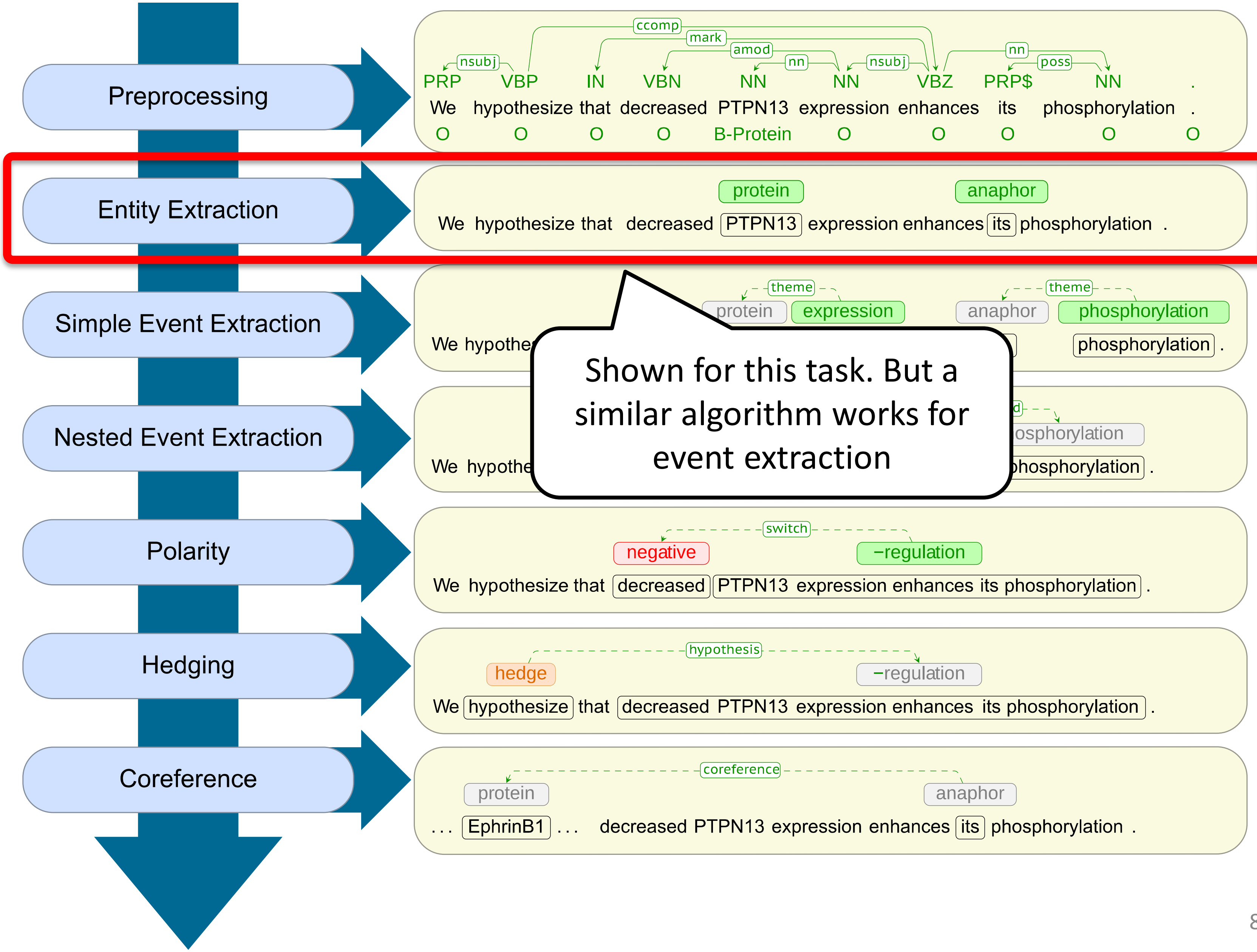
# Motivation

- It took us 1 – 2 person months to build the grammar of 154 rules for the biomedical domain
- In many cases, one does not have this time
  - Scenario 1: But training data exists
  - **Scenario 2: And training data does not exist either…**

# Lightly-supervised Representation Learning with Global Interpretability

**Andrew Zupon, Maria Alexeeva, Marco A. Valenzuela-Escárcega,**
**Ajay Nagesh,** and **Mihai Surdeanu**
University of Arizona, Tucson, AZ, USA
{zupon, alexeeva, marcov, ajaynagesh, msurdeanu}@email.arizona.edu

# Architecture and walkthrough example



# Traditional rule bootstrapping algorithm for entity classification

Problem: brittle statistics, estimated only with respect to known information!

city of San Francisco ✓

**En** **ns**

Seeds → LOC of ___

LOC or of ___

# Learning to read

We want to combine:

- The advantages of representation learning, aka "word embeddings"
  - Neural network language models handle unsupervised data well
- The interpretability of our current approach
  - Produce patterns in the end
- Keep the human in the loop, but minimally
  - Just a few examples

# Starting point: distributional hypothesis

- Distributional hypothesis:
  - By looking at a word's context, one can infer its meaning (Harris, 1954)
  - You shall know a word by the company it keeps (Firth, 1957)

# Example

- tasty *X*
- *X* with butter
- *X* and coffee
- greasy *X*

# Example

- tasty *X*
- *X* with butter
- *X* and coffee
- greasy *X*

# Starting point: word2vec, skip-gram

You shall know a word by the company it keeps.

- Firth, 1957

You shall know the company by the word it keeps.

- Word2vec, skip-gram

# Starting point: word2vec, skip-gram

The city of Tucson is the place to vacation !

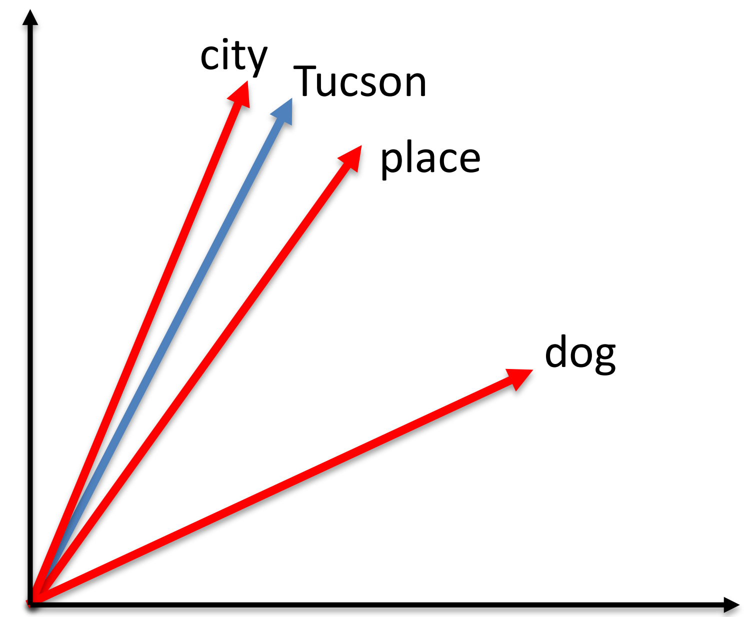
P(city|Tucson) ++

P(place|Tucson) ++

...

P(dog|Tucson) - -

...



# Two important changes

- We will learn embeddings for *both named entities and patterns*
  - An entity's context is defined by the patterns that match it
- Added supervision in the objective function, to incorporate human-provided information
  - We have "seed" names in each category

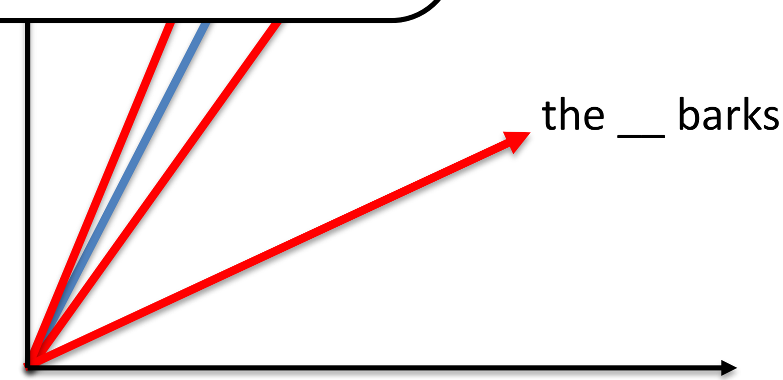
# Context as patterns

The city of Tucson is the place to vacation !

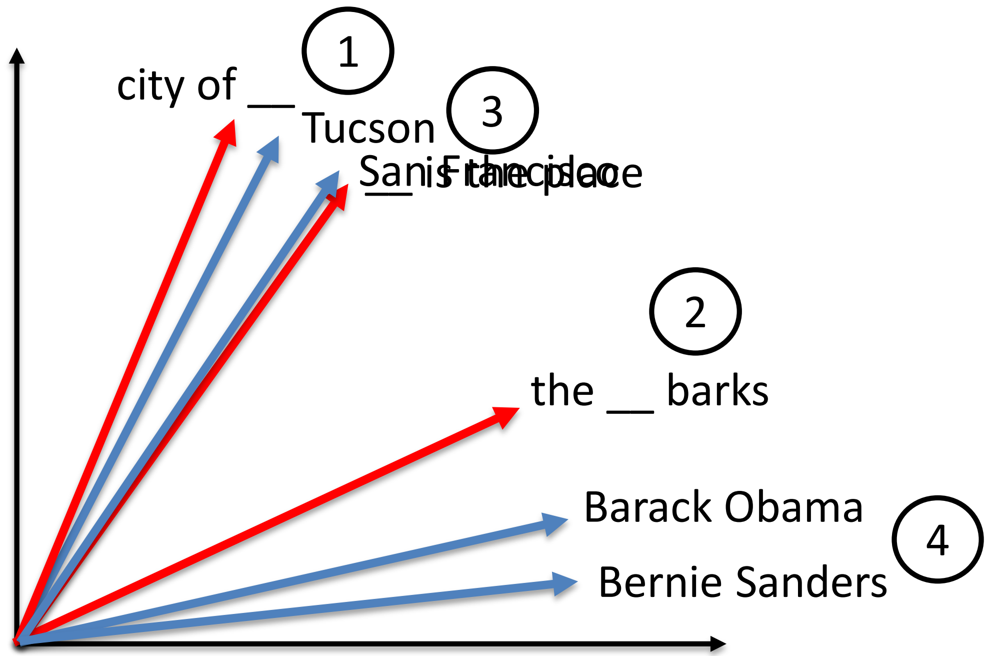
We can export the patterns closest to a category to be used in our grammars!

P(city of __|Tucs
P(__ is the place
…
P(the __ barks|Tucson) - -
…

Just n-grams…



# Objective function




1 2 Unsupervised, directly “inherited” from skip-gram

3 4 Light supervision from a few seed examples, iteratively expanded

# EmBoot

SGD on the previous objective function

**<u>Entities</u>**

Seeds → LOC: San Francisco

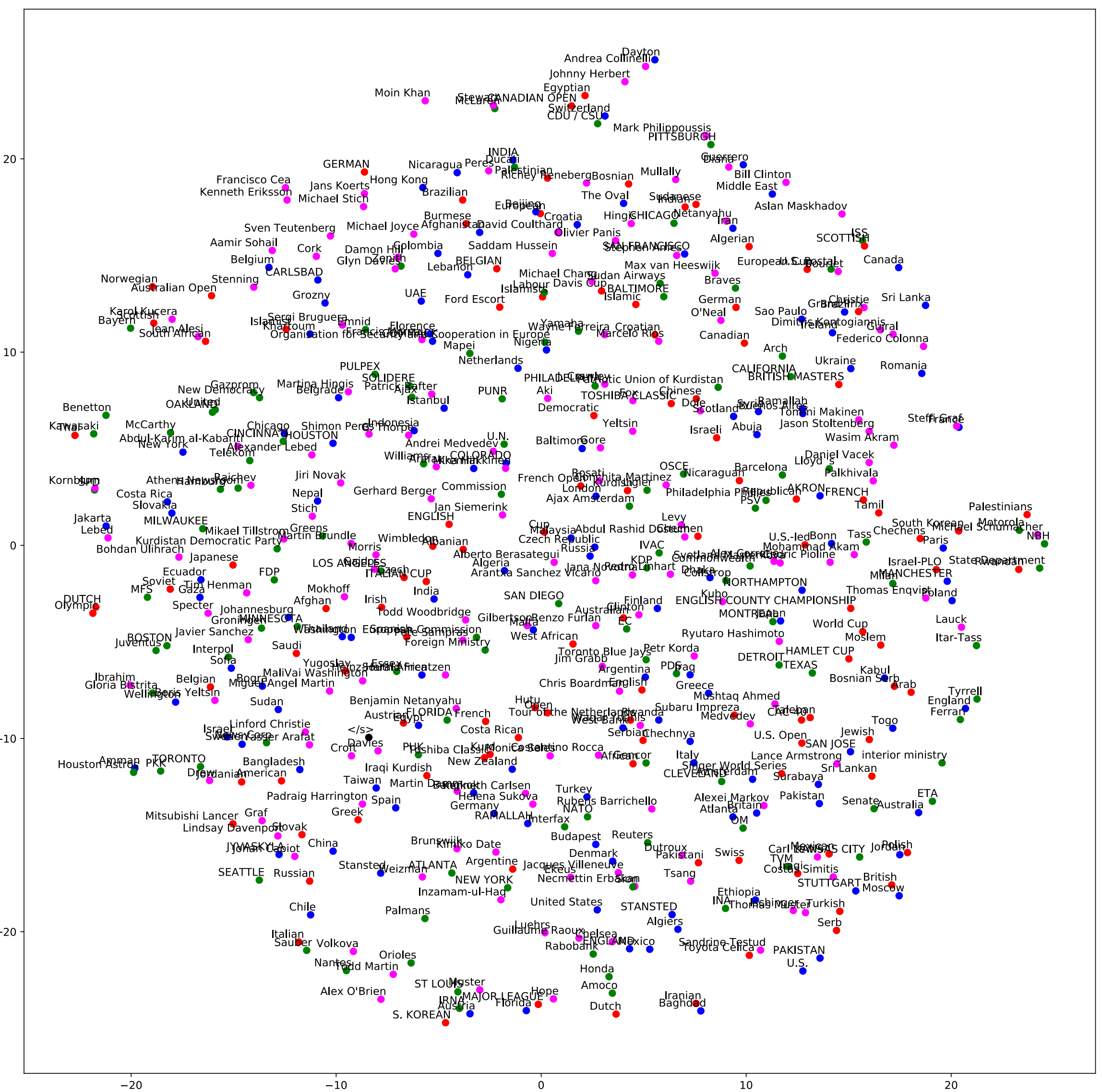
LOC: Los Angeles

**embedding vectors for entities and patterns**

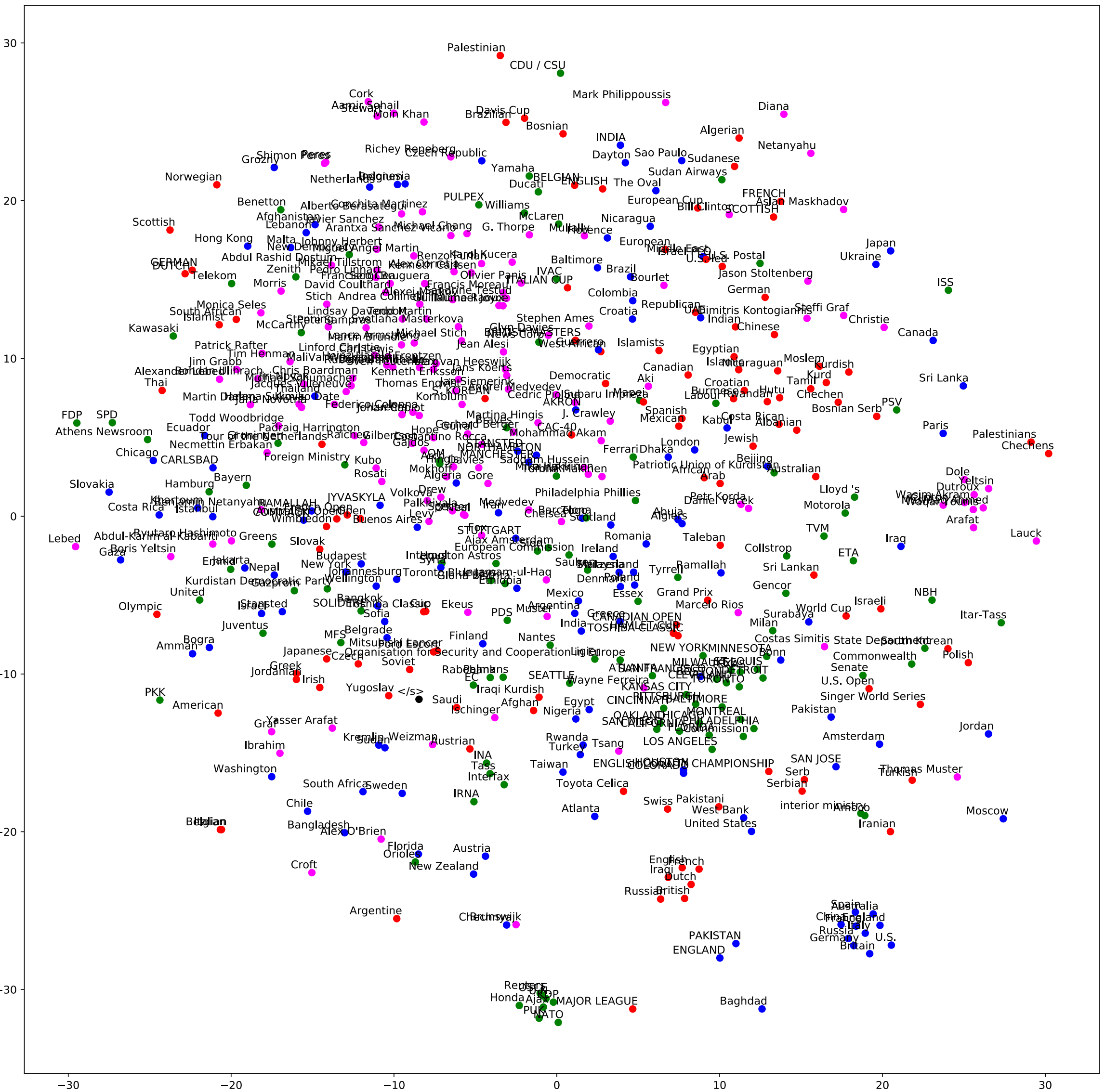
Promote entities closest to the seeds

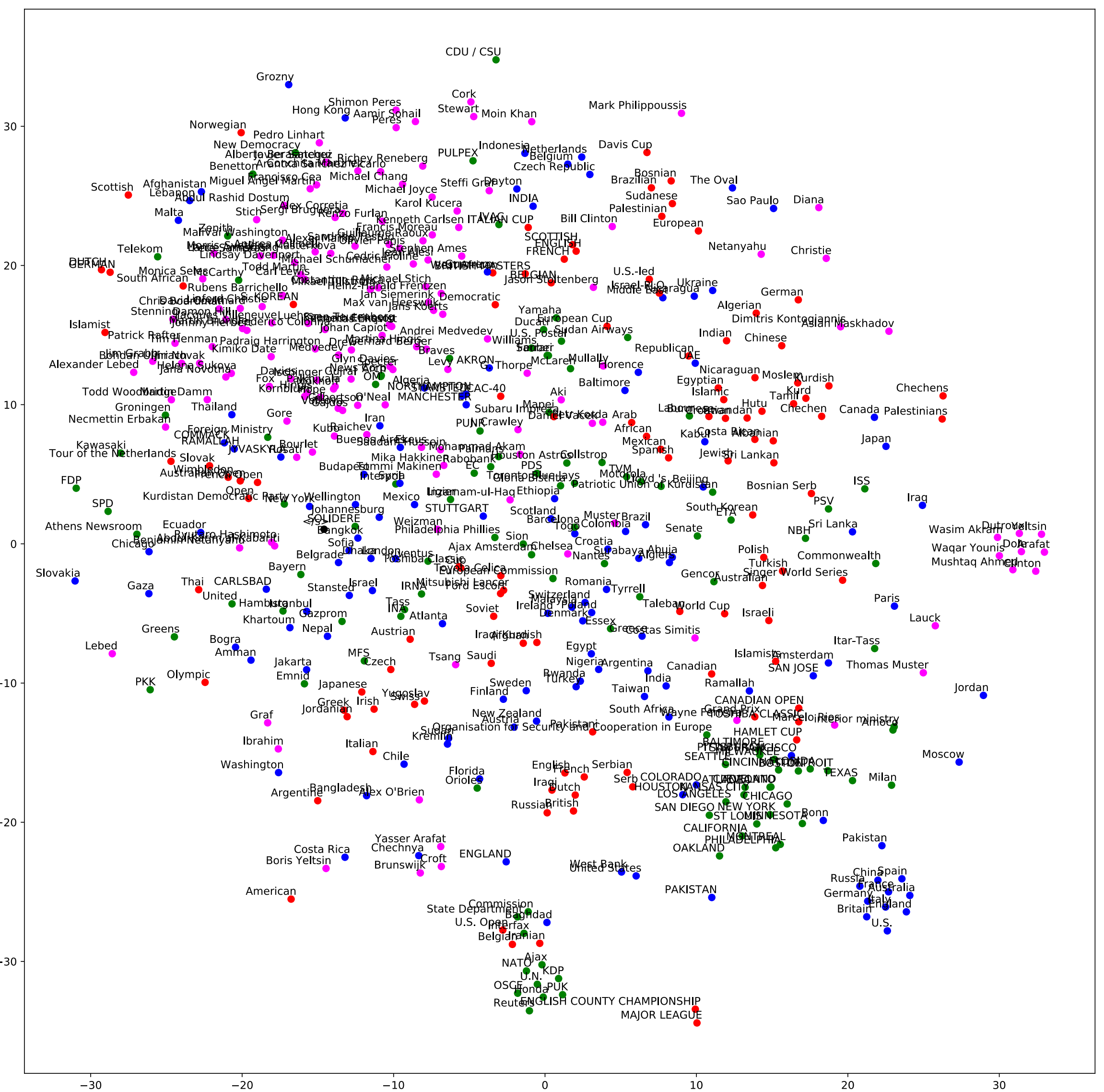
# Visualization of the training procedure

- On the CoNLL-2003 dataset
  - PER = **purple**
  - LOC = **blue**
  - ORG = **green**
  - MISC = **red**
- Human contribution: seed set with 10 entities in each category
  - PER: Clinton, Dole, …
  - LOC: U.S., Germany, …
  - ORG: Reuters, U.N., …
  - MISC: Russian, German, …

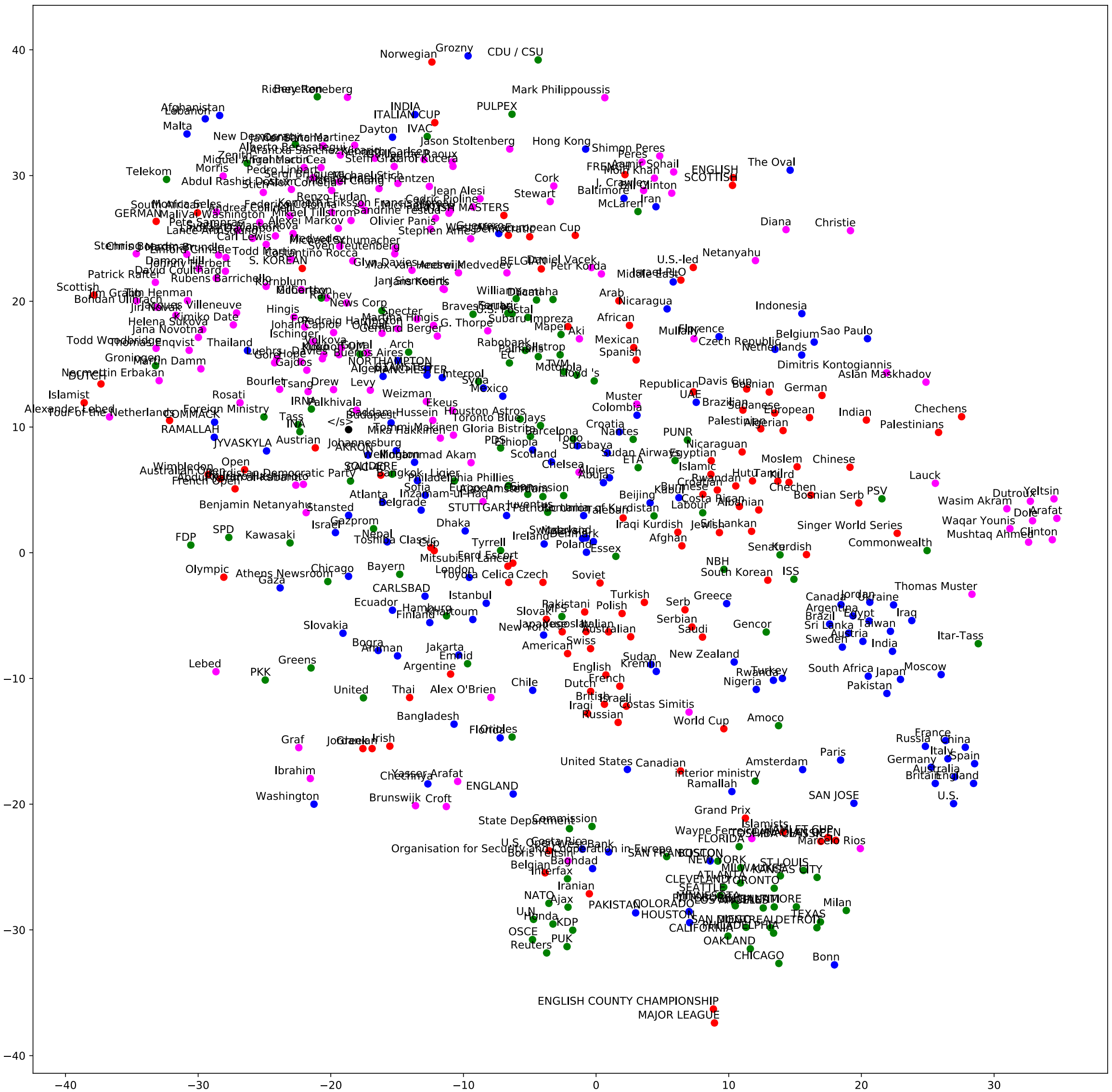


Dayton
Andrea Collinelli
Johnny Herbert
Egyptian
Moin Khan
CANADIAN OPEN
Switzerland
CDU / CSU
Mark Philippoussis
PITTSBURGH
INDIA
GERMAN
Nicaragua
Palestinian
Bosnian
Mullally
Bill Clinton
Middle East
Francisco Cea
Kenneth Eriksson
Jans Koerts
Hong Kong
Michael Stich
Brazilian
The Oval
Sudanese
Aslan Maskhadov
Croatia
CHICAGO
Netanyahu
Iran
Sven Teutenberg
Michael Joyce
David Coulthard
Olivier Panis
Algerian
SCOTTISH
Aamir Sohail
Cork
Damon Hill
Colombia
Saddam Hussein
SAN FRANCISCO
Canada
Belgium
CARLSBAD
Glyn Davies
BELGIUM
Lebanon
Max van Heeswijk
Norwegian
Stenning
Australian Open
Michael Chang
Sudan Airways
Braves
Grozny
UAE
Ford Escort
BALTIMORE
Islamic
German
Sri Lanka
Karol Kucera
Sergi Bruguera
O'Neal
Sao Paulo
Bayern
Khartoum
Florence
Yamaha
Croatian
Ireland
South African
Organisation for Security and Cooperation in Europe
Nigeria
Canadian
Federico Colonna
Mapei
Arch
Netherlands
Ukraine
Romania
PULPEX
CALIFORNIA
Gazprom
Martina Hingis
SOLIDERE
BRITISH MASTERS
New Democracy
Belgrade
Patrick Rafter
Aki
Chinese
Benetton
OAKLAND
Istanbul
PUNR
Democratic
Dole
Scotland
Ramallah
McCarthy
Indonesia
Yeltsin
Tommi Makinen
Steffi Graf
Chicago
Shimon Peres
Abuja
Jason Stoltenberg
Abdul-Karim al-Kabariti
CINCINNATI
HOUSTON
U.N.
Israeli
Wasim Akram
New York
Alexander Lebed
Andrei Medvedev
Gore
Telekom
COLORADO
Daniel Vacek
OSCE
Lloyd 's
Jiri Novak
Nicaraguan
Barcelona
Palkhivala
Korn
SPD
Athens
Raichev
Gerhard Berger
Commission
London
Costa Rica
Nepal
Ajax Amsterdam
AKRON
FRENCH
Slovakia
Stich
Jan Siemerink
Tamil
Palestinians
Jakarta
MILWAUKEE
ENGLISH
Levy
South Korean
Motorola
Lebed
Mikael Tillstrom
Greens
CUP
Abdul Rashid Dostum
Chechens
Paris
NBH
Martin Brundle
Wimbledon
Albanian
Czech Republic
Bonn
Tass
Kurdistan Democratic Party
Russia
IVAC
Mohammad Akam
Bohdan Ulihrach
Alberto Berasategui
Japanese
LOS ANGELES
Algeria
KDP
Israel-PLO
State Department
Ecuador
FDP
ITALIAN CUP
Arantxa Sanchez Vicario
Commonwealth
MANCHESTER
Soviet
Tim Henman
Milan
MFS
Gaza
Mokhoff
NORTHAMPTON
DUTCH
India
Thomas Enqvist
Olympic
Afghan
Irish
SAN DIEGO
Finland
ENGLISH COUNTY CHAMPIONSHIP
Poland
Specter
Johannesburg
Todd Woodbridge
Australian
Clinton
Groningen
MINNESOTA
Gilbertson
Renzo Furlan
EC
MONTREAL
Lauck
World Cup
BOSTON
Javier Sanchez
Foreign Ministry
West African
Ryutaro Hashimoto
Juventus
Moslem
Itar-Tass
Interpol
Saudi
Toronto Blue Jays
Petr Korda
HAMLET CUP
Sofia
Jim Grabb
DETROIT
TEXAS
Argentina
Ibrahim
Belgian
MaliVai Washington
PDS
Kabul
Gloria Bistrita
Miguel Angel Martin
Chris Boardman
English
Greece
Bosnian Serb
Arab
Tyrrell
Wellington
Boris Yeltsin
Mushtaq Ahmed
England
Sudan
Benjamin Netanyahu
Hutu
Ferrari
Linford Christie
FLORIDA
French
Tour of the Netherlands
Subaru Impreza
Medvedev
Togo
Israel
Austria
Egypt
West Bank
Rwanda
</s>
Costa Rican
Serbian
Chechnya
Jewish
U.S. Open
Croft
Davies
SAN JOSE
Toshiba Classic
Costantino Rocca
Houston Astros
Amman
TORONTO
PKK
Bangladesh
New Zealand
Italy
Lance Armstrong
interior ministry
Jordanian
American
Iraqi Kurdish
Singer World Series
CLEVELAND
Sri Lankan
Taiwan
Martin Dahlin
Magnus Carlsen
Surabaya
Turkey
Padraig Harrington
Spain
Helena Sukova
Alexei Markov
Pakistan
ETA
Germany
Rubens Barrichello
Senate
Australia
Mitsubishi Lancer
Graf
Greek
RAMALLAH
NATO
Atlanta
Britain
Lindsay Davenport
Slovak
Interfax
Budapest
Reuters
China
Brunswijk
Kimiko Date
Dutroux
Denmark
Swiss
Polish
JYVASKYLA
Johan Cabiot
Stansted
Weizman
ATLANTA
Argentine
Jacques Villeneuve
Pakistani
Mexican
Jordan
SEATTLE
Russian
NEW YORK
Necmettin Erbakan
Tsang
Simitis
STUTTGART
British
Moscow
Inzamam-ul-Haq
Ethiopia
Chile
United States
STANSTED
INA
Turkish
Palmans
Algiers
Serb
Italian
Sauber
Volkova
Guillaume Raoux
Chelsea
Mexico
Sandrine Testud
Toyota Celica
Rabobank
PAKISTAN
Orioles
ENGLAND
Nantes
Todd Martin
Honda
U.S.
ST LOUIS
Amoco
Alex O'Brien
IRNA
MAJOR LEAGUE
Hope
Iranian
Florida
Baghdad
Austria
Dutch
S. KOREAN
20
10
0
−10
−20
−20
−10
0
10
20

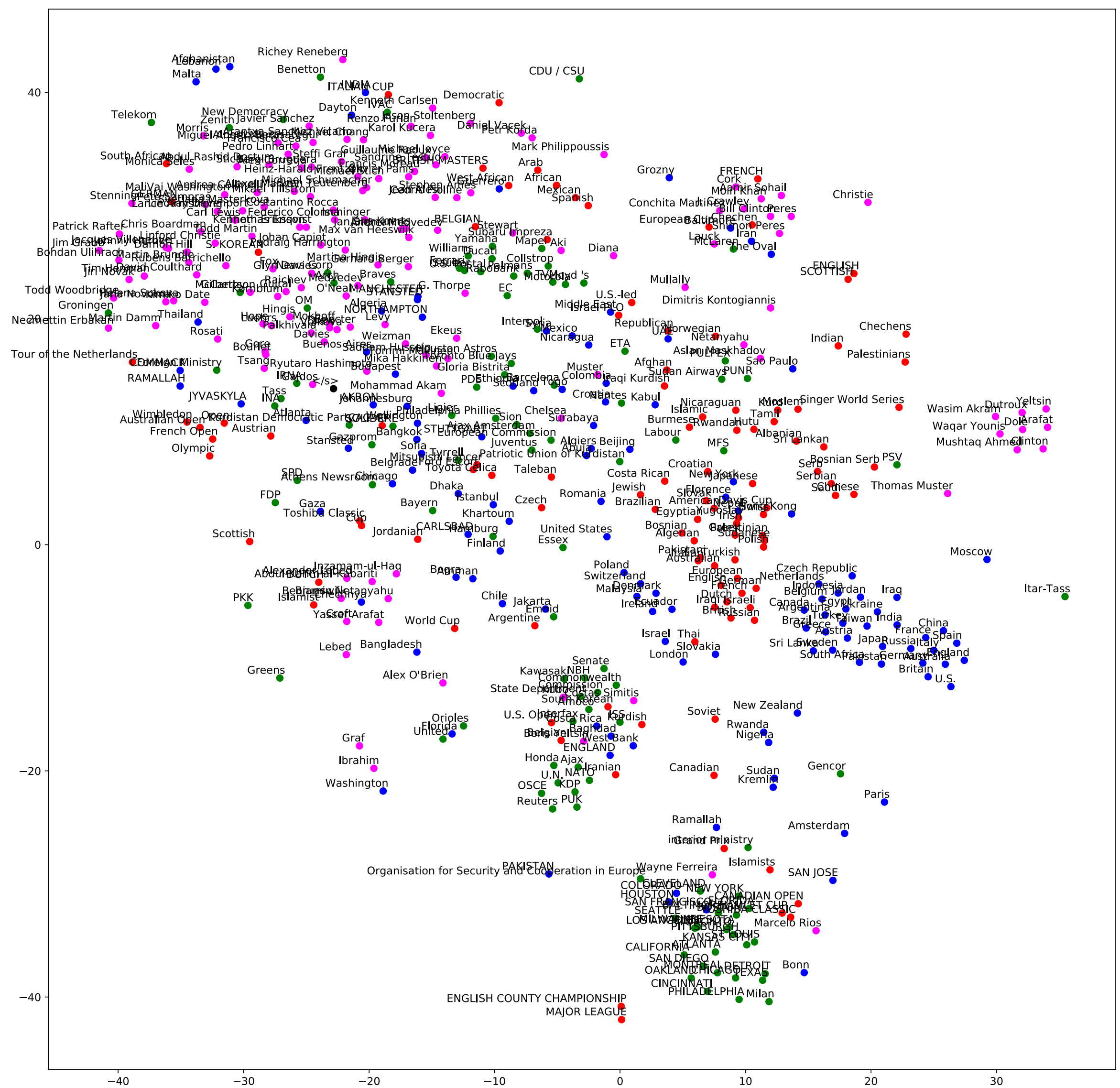

Palestinian
CDU / CSU
Mark Philippoussis
Diana
Algerian
Netanyahu
Bosnian
INDIA
Norwegian
Scottish
Hong Kong
Japan
Ukraine
ISS
Canada
Sri Lanka
PSV
Paris
Palestinians
Chechens
FDP
SPD
Athens Newsroom
Chicago
CARLSBAD
Slovakia
Hamburg
Costa Rica
Lebed
Olympic
Amman
PKK
American
Washington
South Africa
Chile
Croft
Argentine
New Zealand
Austria
Atlanta
Moscow
Jordan
Amsterdam
Baghdad
PAKISTAN
ENGLAND
MAJOR LEAGUE
Lauck
Arafat
Iraq
ETA
NBH
Itar-Tass
Polish
Iranian
Turkish
U.S.
Britain
Germany
Russia
Saudi
</s>
−30
−20
−10
0
10
20
30

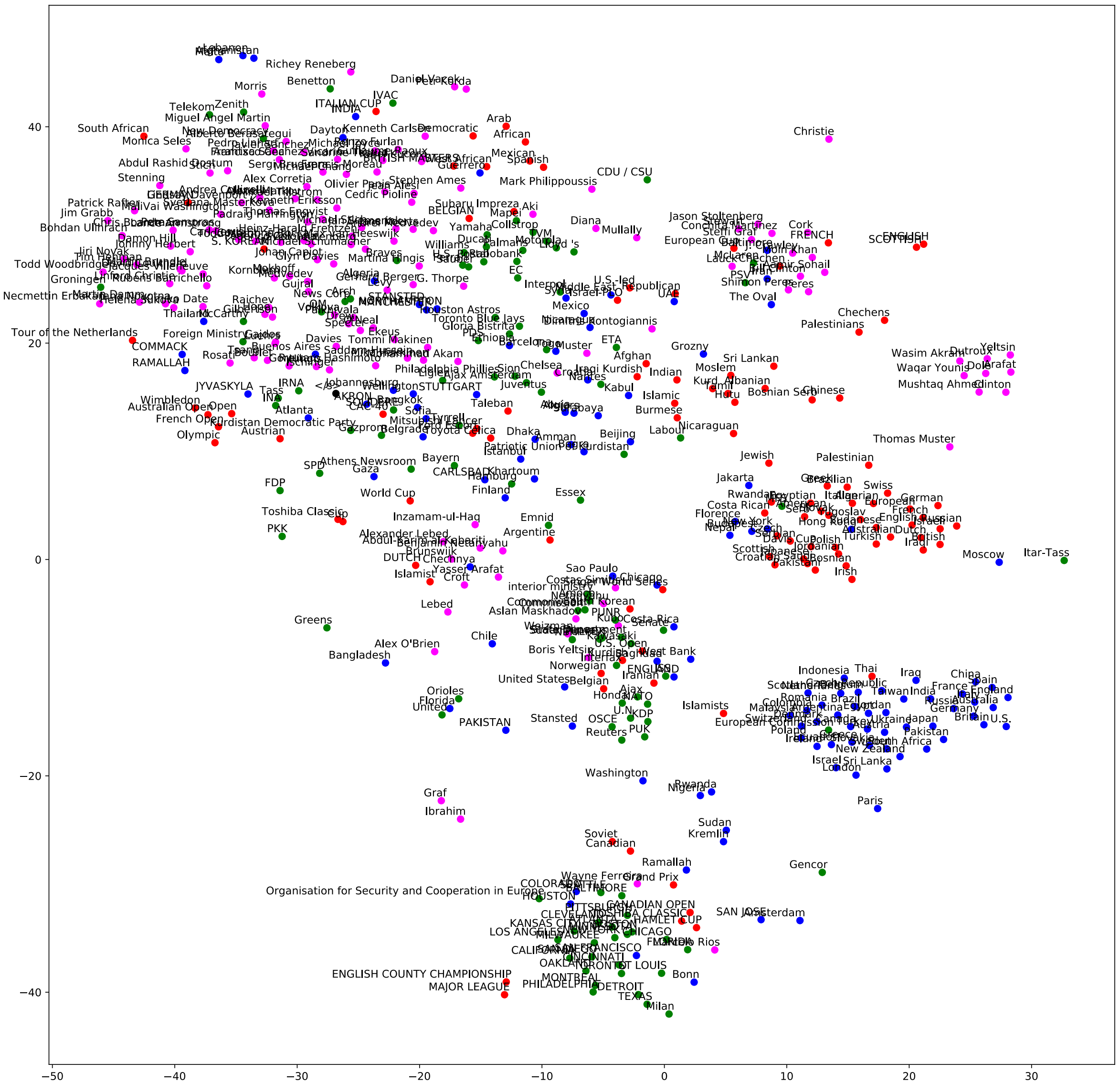


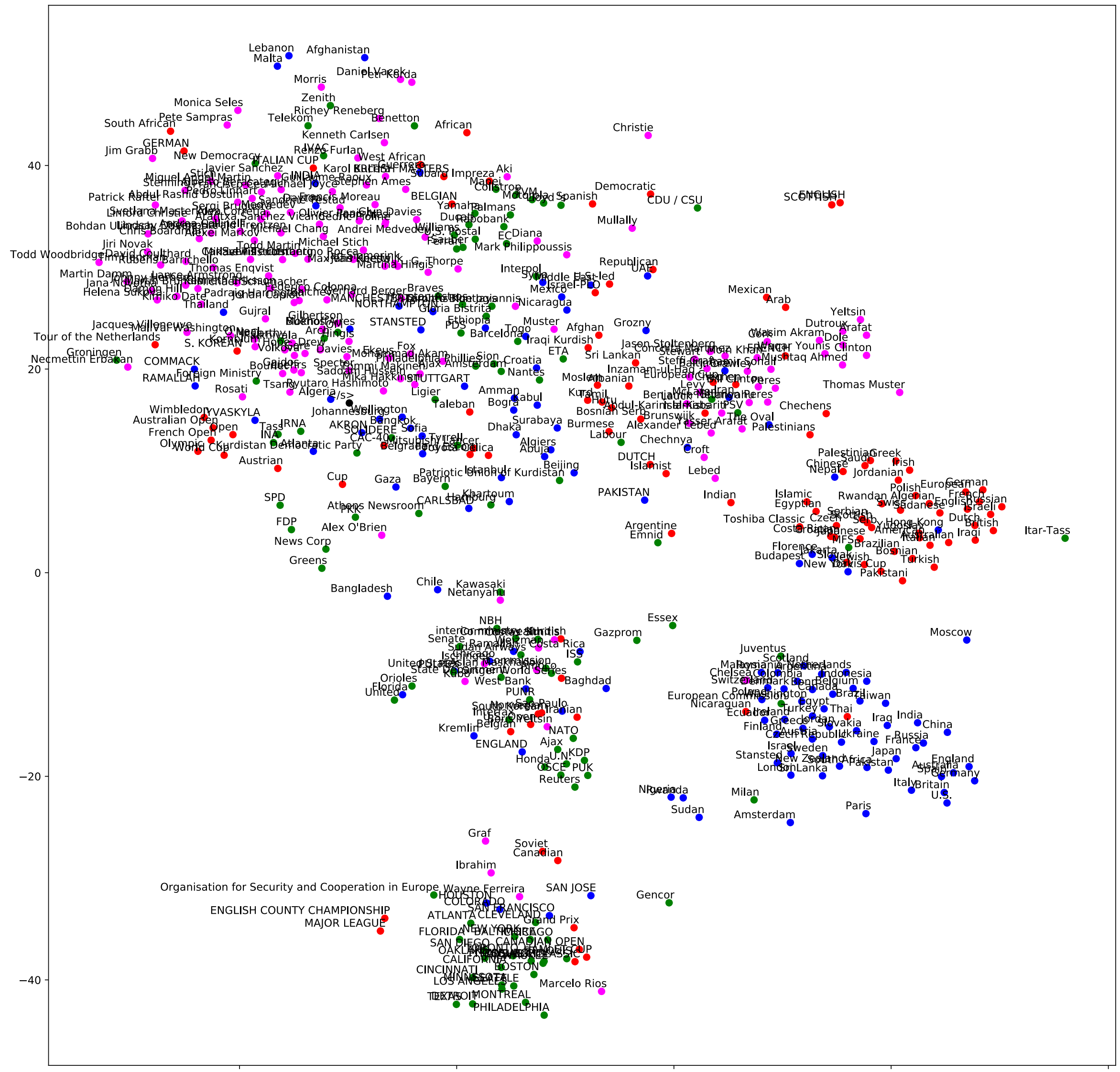
CDU / CSU
Grozny
Shimon Peres
Cork
Mark Philippoussis
Hong Kong
Aamir Sohail
Stewart
Moin Khan
Norwegian
Pedro Linhart
New Democracy
Indonesia
Netherlands
Davis Cup
PULPEX
Belgium
Czech Republic
Bosnian
Brazilian
The Oval
Scottish
Afghanistan
Michael Chang
Michael Joyce
Steffi Graf
Dayton
INDIA
Sudanese
Palestinian
Sao Paulo
Diana
Lebanon
Abdul Rashid Dostum
Karol Kucera
Malta
Kenneth Carlsen
ITALIAN CUP
Bill Clinton
European
Zenith
Francis Moreau
SCOTTISH
ENGLISH
FRENCH
Netanyahu
Christie
Telekom
DUTCH
GERMAN
Stephen Ames
Monica Seles
McCarthy
South African
Todd Martin
Carl Lewis
Michael Stich
BELGIAN
Jason Stoltenberg
U.S.-led
Israel-PLO
Ukraine
Middle East
Rubens Barrichello
KOREAN
Jan Siemerink
Democratic
German
Max van Heeswijk
Yamaha
Algerian
Dimitris Kontogiannis
Aslan Maskhadov
Islamist
Ducati
European Cup
Sudan Airways
Patrick Rafter
Johan Capiot
Andrei Medvedev
U.S. Postal
Indian
Chinese
Tim Henman
Padraig Harrington
Kimiko Date
Medvedev
Braves
Williams
Republican
Jim Grabb
Jana Novotna
Mullally
UAE
Alexander Lebed
Helena Sukova
AKRON
Thorpe
McLaren
Florence
Nicaraguan
Moslem
Kurdish
Egyptian
Baltimore
Islamic
Kurd
Todd Woodbridge
Martin Damm
O'Neal
MANCHESTER
CAC-40
Aki
Tamil
Chechens
Groningen
Thailand
Hutu
Chechen
Canada
Palestinians
Necmettin Erbakan
Gore
Iran
PUNK
Crawley
Arab
African
Costa Rican
Albanian
Japan
Foreign Ministry
Raichev
Mexican
Kawasaki
COMMACK
RAMALLAH
Bourlet
Spanish
Jewish
Sri Lankan
Tour of the Netherlands
Slovak
Mika Hakkinen
Collstrop
Budapest
Tommi Makinen
PDS
TVM
Bosnian Serb
Motorola
FDP
Wimbledon
Australian Open
Patriotic Union of Kurdistan
ISS
Kurdistan Democratic Party
Wellington
Mexico
Ethiopia
Iraq
SPD
Johannesburg
STUTTGART
Scotland
PSV
Open
Barcelona
Brazil
South Korean
Athens Newsroom
Ecuador
SOLIDERE
Weizman
Philadelphia Phillies
Togo
Colombia
ETA
Senate
Sri Lanka
Wasim Akram
Dutroux
Yeltsin
Chicago
Ryutaro Hashimoto
Bangkok
Sion
Croatia
NBH
Dole
Arafat
Sofia
Ajax Amsterdam
Chelsea
Polish
Waqar Younis
Belgrade
London
Juventus
Nantes
Surabaya
Abuja
Algeria
Commonwealth
Slovakia
Bayern
Turkish
Mushtaq Ahmed
Clinton
European Commission
Genco
Australian
Singer World Series
Gaza
Thai
CARLSBAD
Israel
IRNA
Romania
Stansted
United
Hamburg
Istanbul
Switzerland
Tyrrell
Paris
Tass
Soviet
Ireland
Poland
Taleban
World Cup
Israeli
Gazprom
Atlanta
Denmark
Khartoum
Essex
Lauck
Greens
Nepal
Austrian
Greece
Costas Simitis
Bogra
Iraqi
Kurdish
Itar-Tass
Lebed
Egypt
Islamists
Amsterdam
Amman
Tsang
Saudi
Nigeria
Argentina
Canadian
Thomas Muster
Jakarta
MFS
Czech
SAN JOSE
Olympic
Emnid
Rwanda
India
Ramallah
PKK
Japanese
Sweden
Turkey
Jordan
Finland
Taiwan
Yugoslav
Swiss
CANADIAN OPEN
Greek
Irish
Grand Prix
Graf
Jordanian
New Zealand
South Africa
Interior ministry
Austria
Ibrahim
Pakistani
Kremlin
Organisation for Security and Cooperation in Europe
HAMLET CUP
Italian
BALTIMORE
SAN FRANCISCO
Chile
SEATTLE
CINCINNATI
DETROIT
Moscow
Washington
English
French
Serbian
TEXAS
Milan
Florida
Oriole
COLORADO
CLEVELAND
Serb
HOUSTON
KANSAS CITY
Argentine
Bangladesh
Iraqi
Alex O'Brien
Dutch
LOS ANGELES
CHICAGO
SAN DIEGO
NEW YORK
Russian
British
ST LOUIS
MINNESOTA
Bonn
CALIFORNIA
MONTREAL
PHILADELPHIA
Yasser Arafat
Pakistan
Costa Rica
Chechnya
OAKLAND
Boris Yeltsin
Croft
ENGLAND
Brunswik
West Bank
United States
China
Spain
Russia
France
Australia
PAKISTAN
American
Germany
Italy
Finland
Commission
Britain
State Department
Baghdad
U.S.
U.S. Open
Belgian
Iranian
Ajax
NATO
KDP
U.N.
OSCE
PUK
Honda
Reuters
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE
30
20
10
0
−10
−20
−30
−30
−20
−10
0
10
20
30



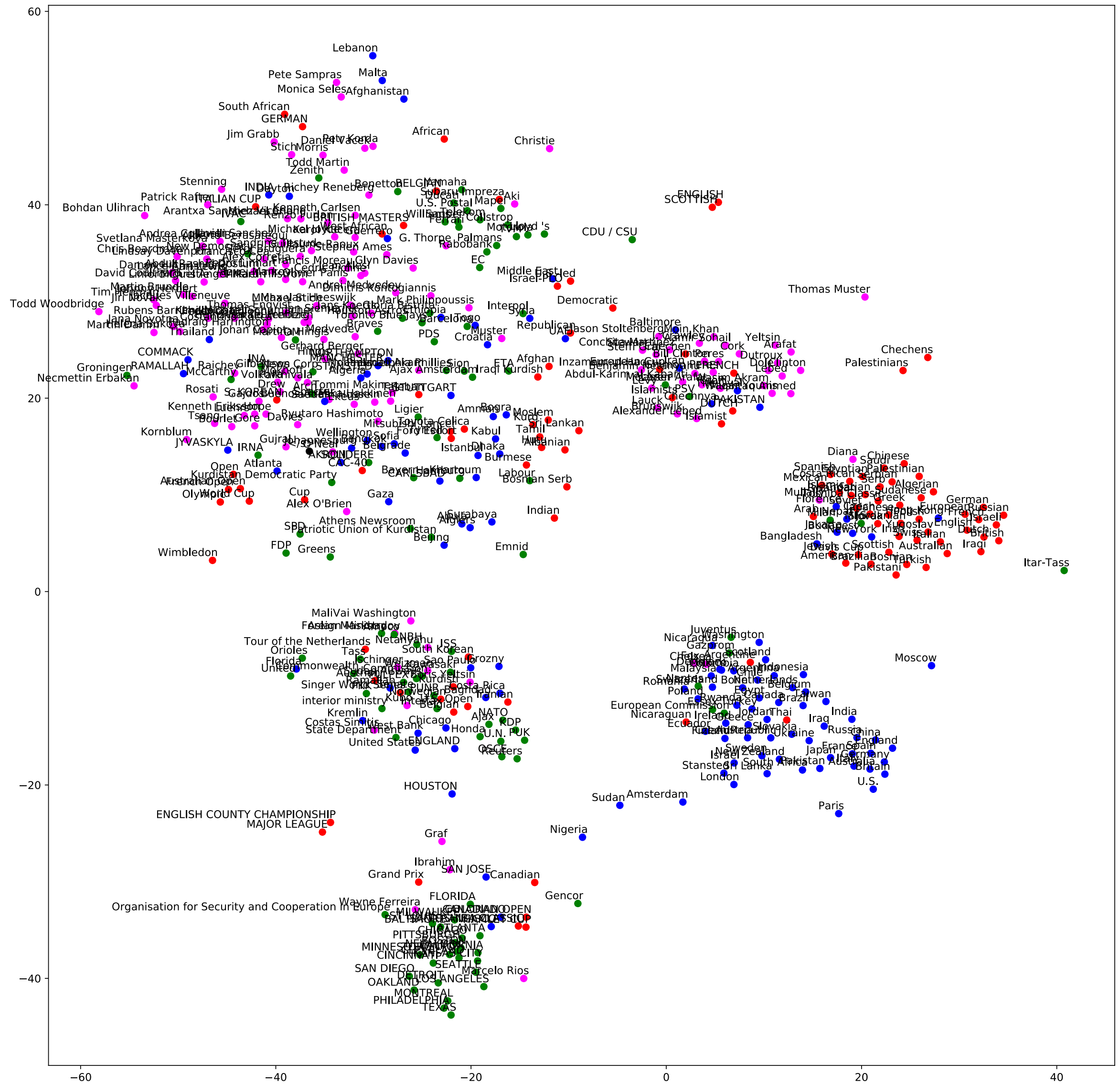
40
30
20
10
0
−10
−20
−30
−40
Norwegian
Grozny
CDU / CSU
Mark Philippoussis
Afghanistan
Malta
PULPEX
INDIA
Dayton
Jason Stoltenberg
Hong Kong
Telekom
The Oval
Diana
Christie
Netanyahu
Indonesia
Sao Paulo
Dimitris Kontogiannis
Aslan Maskhadov
Islamist
Chechens
Palestinians
Lauck
Yeltsin
Wasim Akram
Waqar Younis
Mushtaq Ahmed
Singer World Series
Commonwealth
FDP
SPD
Kawasaki
Olympic
Thomas Muster
Itar-Tass
Slovakia
Lebed
PKK
Greens
United
Thai
Alex O'Brien
Chile
World Cup
Amoco
Moscow
Pakistan
Graf
Ibrahim
Washington
ENGLAND
United States
Paris
SAN JOSE
Grand Prix
Marcelo Rios
Organisation for Security and Cooperation in Europe
Iranian
PAKISTAN
Milan
OSCE
Reuters
CHICAGO
Bonn
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE

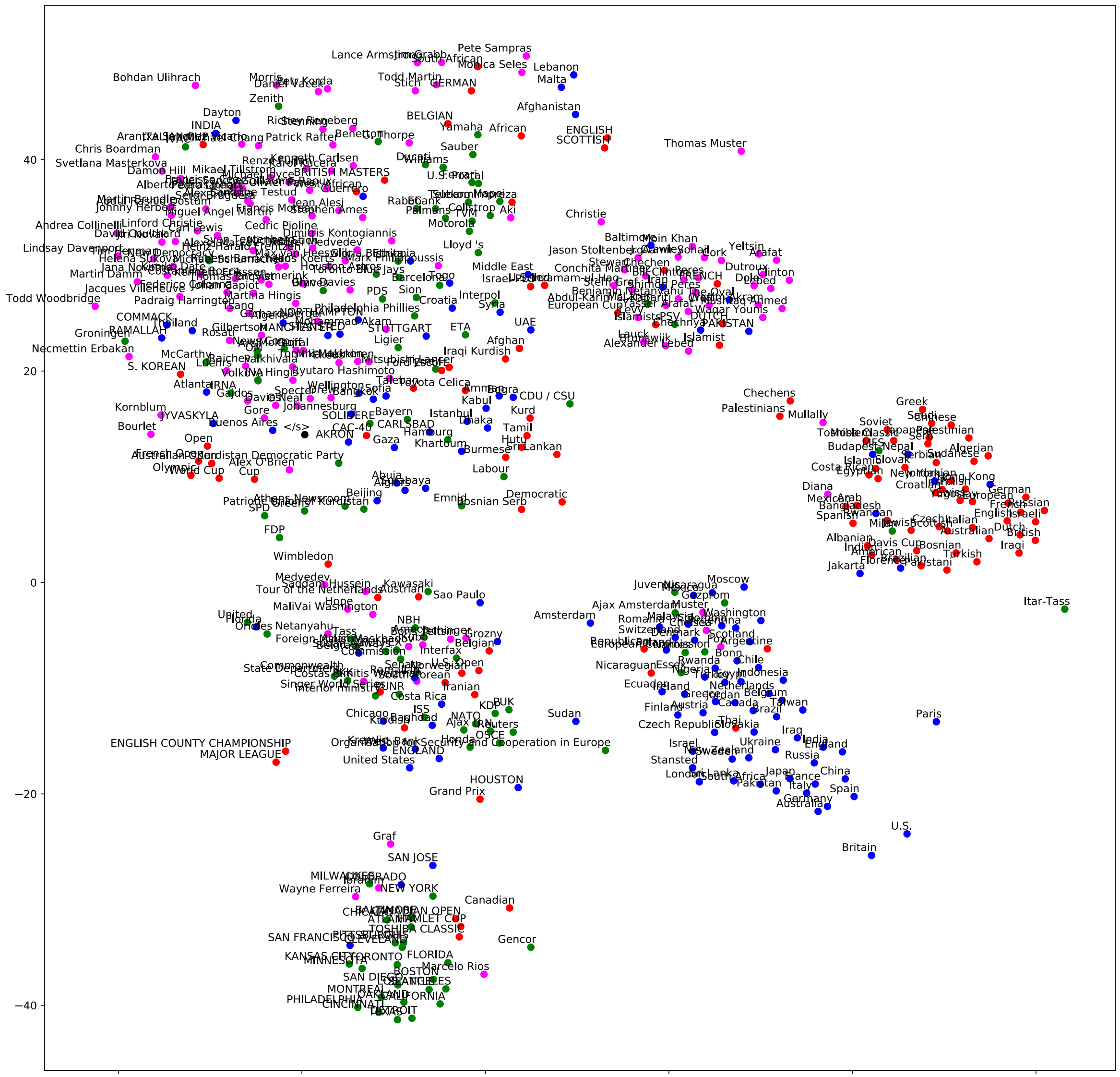

40
20
0
−20
−40
−40
−30
−20
−10
0
10
20
30
Afghanistan
Malta
Richey Reneberg
Benetton
CDU / CSU
Democratic
Telekom
Dayton
Karol Kucera
Mark Philippoussis
Arab
African
Mexican
Spanish
Grozny
FRENCH
Cork
Christie
Conchita Martinez
Stewart
Diana
Mullally
ENGLISH
SCOTTISH
Dimitris Kontogiannis
Chechens
Indian
Palestinians
Netanyahu
Sao Paulo
Singer World Series
Wasim Akram
Yeltsin
Waqar Younis
Mushtaq Ahmed
Arafat
Clinton
Thomas Muster
Bosnian Serb
PSV
Serbian
Moscow
Itar-Tass
Czech Republic
Netherlands
Iraq
China
France
Spain
U.S.
Britain
Patrick Rafter
Chris Boardman
Bondan Ulimrac
Todd Woodbridge
Groningen
Thailand
Rosati
Tour of the Netherlands
RAMALLAH
JYVASKYLA
French Open
Olympic
Scottish
Atlanta
Austrian
SPD
FDP
Gaza
Toshiba Classic
Bayern
Istanbul
Khartoum
Hamburg
Finland
Czech
Romania
United States
Essex
Poland
Switzerland
Denmark
Malaysia
Moslem
Albanian
Sri Lankan
Croatian
Brazilian
Jewish
PKK
World Cup
Bangladesh
Lebed
Greens
Alex O'Brien
Chile
Jakarta
Argentine
Israel
Thai
London
Slovakia
Senate
Soviet
New Zealand
Rwanda
Nigeria
Sudan
Kremlin
Gencor
Paris
Canadian
Graf
Ibrahim
Washington
Honda
OSCE
Reuters
PUK
KDP
Ramallah
Amsterdam
Islamists
SAN JOSE
Organisation for Security and Cooperation in Europe
Wayne Ferreira
Marcelo Rios
CALIFORNIA
ATLANTA
SAN DIEGO
Bonn
CINCINNATI
PHILADELPHIA
Milan
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE


40
20
0
−20
−40
−50
−40
−30
−20
−10
0
10
20
30
Richey Reneberg
Christie
South African
Mark Philippoussis
CDU / CSU
Tour of the Netherlands
Thomas Muster
Moscow
Itar-Tass
Greens
Bangladesh
PAKISTAN
Washington
Paris
Gencor
Graf
Ibrahim
Organisation for Security and Cooperation in Europe
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE
SAN JOSE
Amsterdam
Milan
Bonn

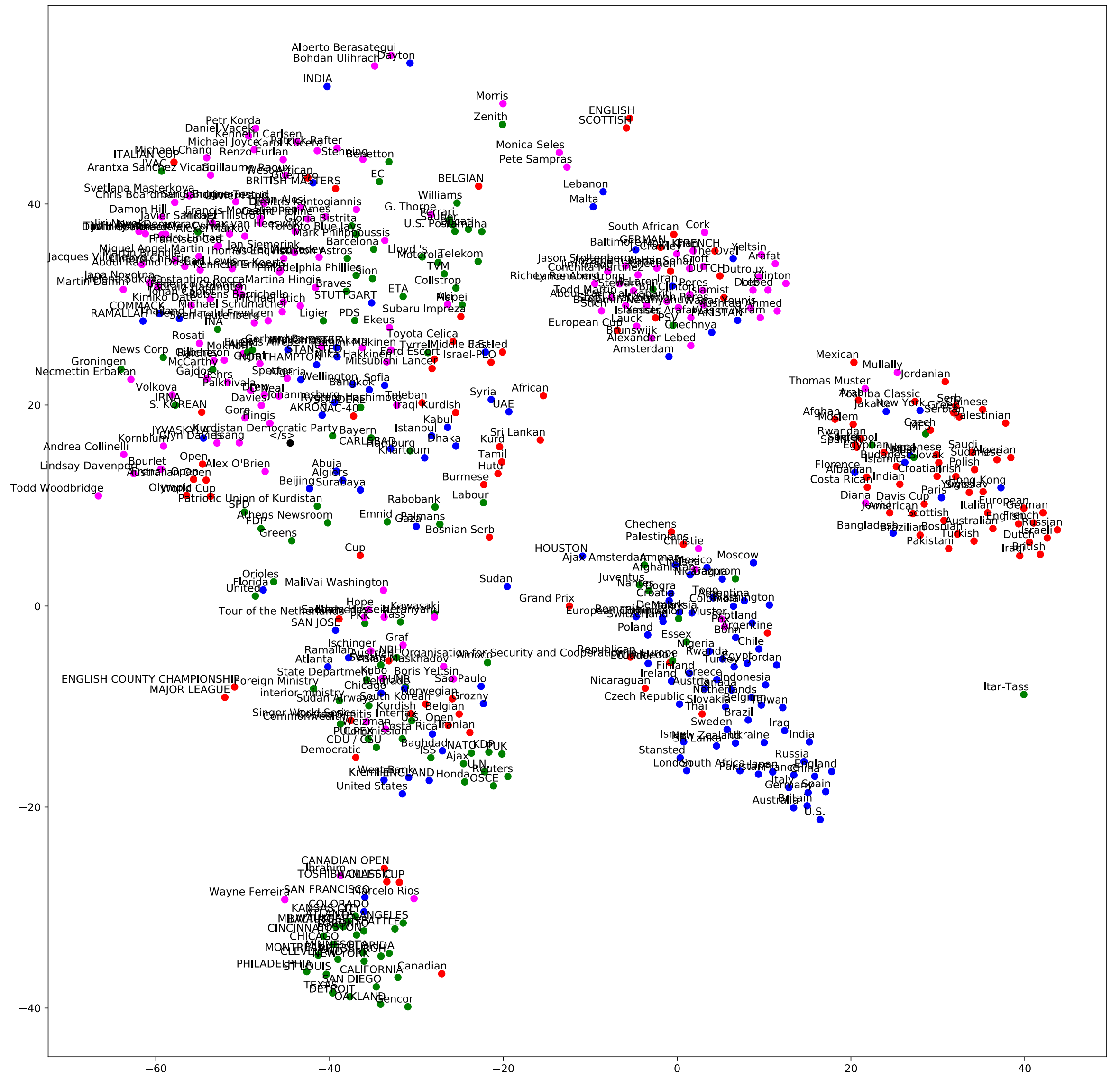


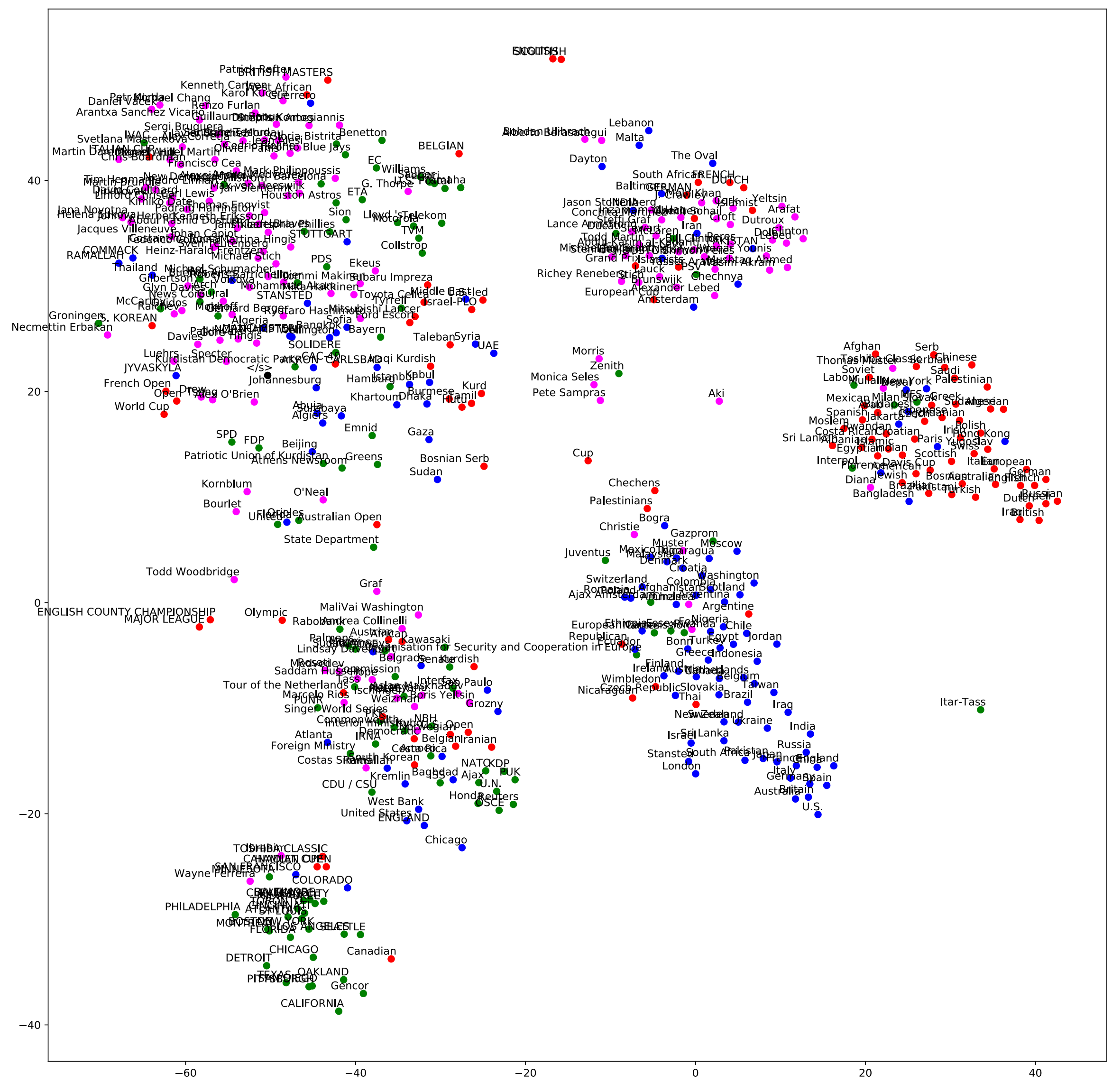
Lebanon
Malta
Afghanistan
Morris
Monica Seles
Pete Sampras
South African
Telekom
Benetton
African
Christie
Kenneth Carlsen
GERMAN
Jim Grabb
New Democracy
ITALIAN CUP
West African
Democratic
CDU / CSU
Spanish
Mullally
Todd Woodbridge
Jiri Novak
Mark Philippousis
Republican
Martin Damm
Thomas Enqvist
Interpol
Mexican
Arab
Yeltsin
Thailand
Gujral
Jacques Villeneuve
Tour of the Netherlands
Groningen
Necmettin Erbakan
COMMACK
RAMALLAH
Foreign Ministry
Rosati
Thomas Muster
Grozny
Afghan
Sri Lankan
Iraqi Kurdish
Croatia
Nantes
Amman
Kabul
Taleban
Chechens
Palestinians
Wimbledon
Australian Open
French Open
Austrian
Surabaya
Burmese
Labour
Chechnya
DUTCH
Islamist
Lebed
Greek
Irish
Jordanian
Cup
Gaza
Bayern
SPD
FDP
News Corp
Greens
Alex O'Brien
PAKISTAN
Indian
Islamic
Egyptian
Argentine
Emnid
Toshiba Classic
Itar-Tass
Pakistani
Turkish
Bangladesh
Chile
Kawasaki
Netanyahu
Essex
Gazprom
Moscow
NBH
Senate
Baghdad
West Bank
Orioles
Florida
Juventus
Scotland
European Commission
Nicaraguan
India
China
Kremlin
ENGLAND
NATO
Ajax
KDP
PUK
Reuters
Stansted
London
Japan
Italy
Britain
U.S.
Nigeria
Rwanda
Sudan
Milan
Amsterdam
Paris
Graf
Soviet
Canadian
Ibrahim
Organisation for Security and Cooperation in Europe
SAN JOSE
Gencor
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE
ATLANTA
CLEVELAND
Grand Prix
CALIFORNIA
BOSTON
CINCINNATI
Marcelo Rios
MONTREAL
PHILADELPHIA
40
20
0
−20
−40
−40
−20
0
20
40


Lebanon
Malta
Pete Sampras
Monica Seles
Afghanistan
South African
GERMAN
Jim Grabb
African
Christie
Todd Martin
Zenith
Stenning
Patrick Rafter
Bohdan Ulihrach
ENGLISH
SCOTTISH
CDU / CSU
Thomas Muster
Todd Woodbridge
Middle East
Democratic
Baltimore
Yeltsin
Arafat
Chechens
Palestinians
COMMACK
Groningen
RAMALLAH
Necmettin Erbakan
Rosati
Kenneth Eriksson
Kornblum
JYVASKYLA
IRNA
Atlanta
Open
Kurdistan Democratic Party
World Cup
Cup
Alex O'Brien
Gaza
Indian
Athens Newsroom
Patriotic Union of Kurdistan
Beijing
SPD
FDP
Greens
Wimbledon
Emnid
Diana
Chinese
Saudi
Spanish
Palestinian
Mexican
Algerian
German
Russian
Israeli
Bangladesh
English
Iraqi
Australian
Scottish
Bosnian
Turkish
Pakistani
American
Itar-Tass
MaliVai Washington
Tour of the Netherlands
Orioles
South Korean
Grozny
Costa Rica
interior ministry
Kremlin
Costas Simitis
State Department
United States
ENGLAND
NATO
Reuters
Chicago
Honda
Juventus
Washington
Nicaragua
Indonesia
Romania
Poland
Taiwan
European Commission
Nicaraguan
Ecuador
India
Iraq
Russia
China
England
Japan
France
Spain
Germany
Sweden
Britain
Stansted
London
U.S.
Moscow
Sudan
Amsterdam
Paris
HOUSTON
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE
Graf
Nigeria
Ibrahim
Grand Prix
SAN JOSE
Canadian
Gencor
Organisation for Security and Cooperation in Europe
Wayne Ferreira
FLORIDA
PITTSBURGH
MINNESOTA
CINCINNATI
SEATTLE
SAN DIEGO
DETROIT
Marcelo Rios
OAKLAND
LOS ANGELES
MONTREAL
PHILADELPHIA
TEXAS
60
40
20
0
−20
−40
−60
−40
−20
0
20
40

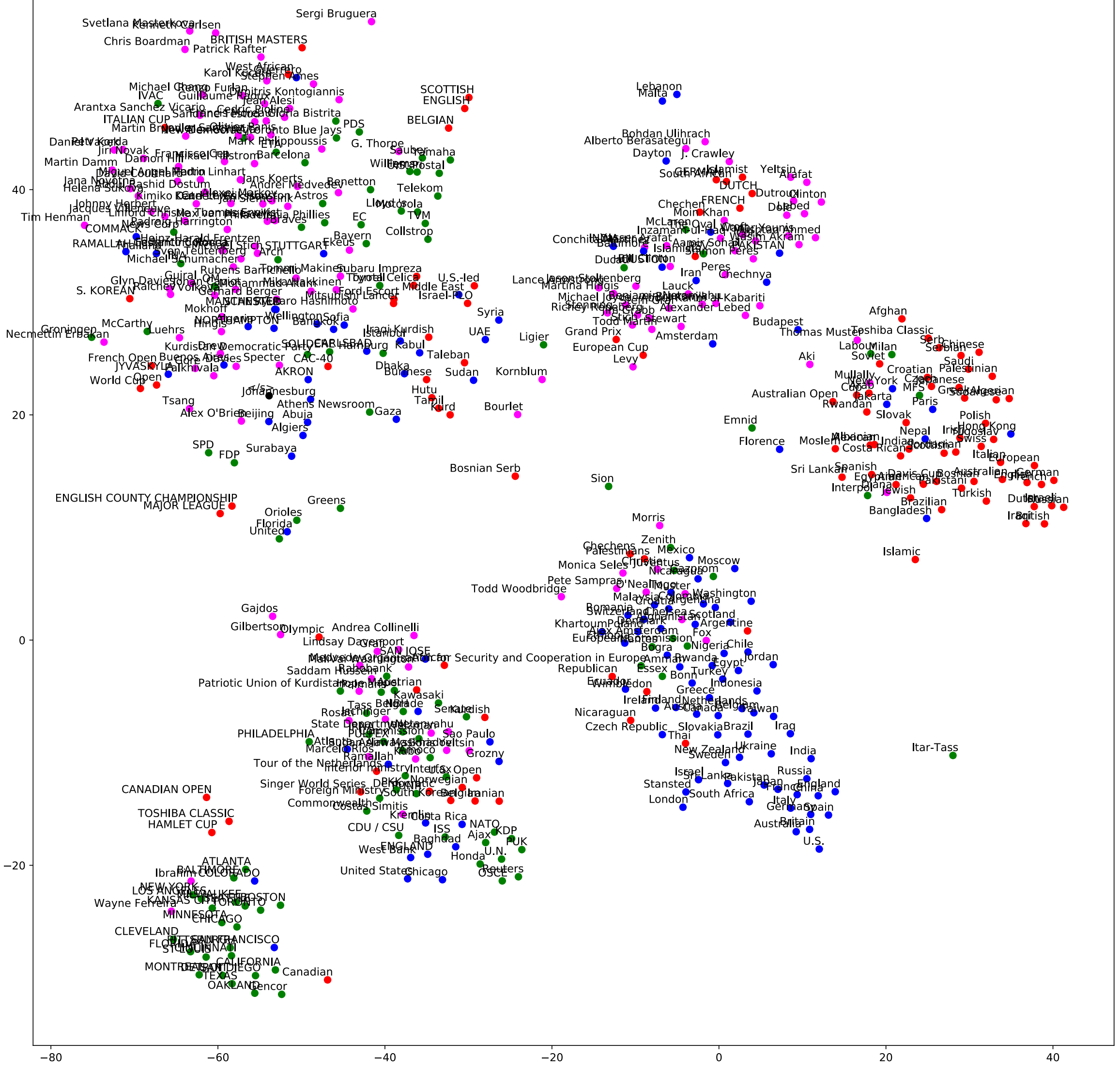


Lance Armstrong
Pete Sampras
Lebanon
Bohdan Ulihrach
Malta
Afghanistan
Thomas Muster
ENGLISH
SCOTTISH
BELGIAN
Sauber
Chris Boardman
Svetlana Masterkova
Christie
Andrea Collinelli
Lindsay Davenport
Todd Woodbridge
Middle East
Yeltsin
Arafat
Kornblum
Bourlet
Chechens
Palestinians
Greek
Chinese
Serbian
Algerian
German
Russian
Israeli
British
Iraqi
Turkish
Wimbledon
Amsterdam
Moscow
Washington
Itar-Tass
Sudan
Paris
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE
United States
HOUSTON
Grand Prix
Graf
SAN JOSE
Wayne Ferreira
NEW YORK
Canadian
Gencor
Marcelo Rios
U.S.
Britain
China
Spain
Russia
England
India
Iraq
Japan
France
Italy
Germany
Australia
40
20
0
−20
−40
−60
−40
−20
0
20
40

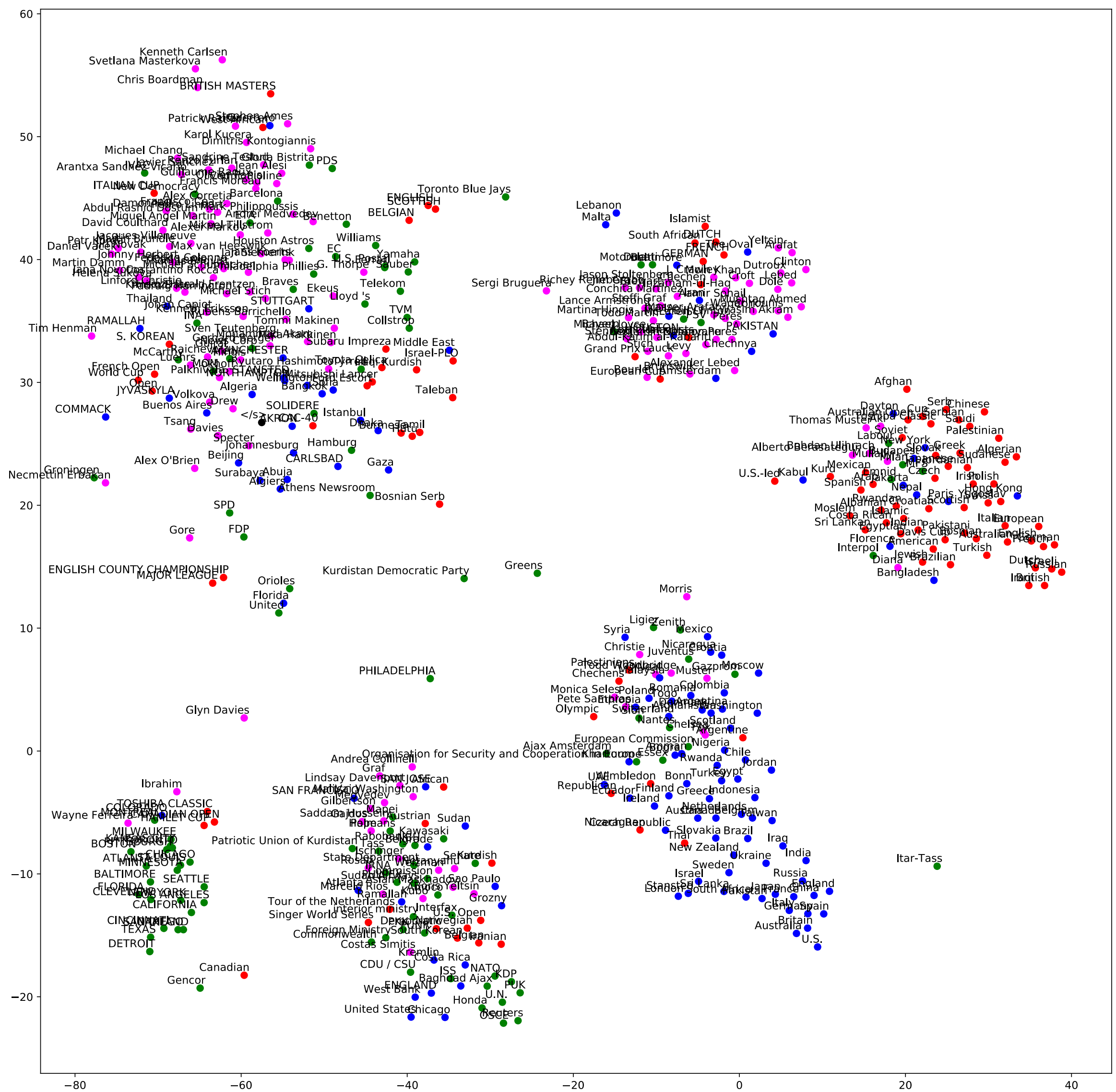

Alberto Berasategui
Bohdan Ulihrach
Dayton
INDIA
Morris
Zenith
ENGLISH
SCOTTISH
Monica Seles
Pete Sampras
Lebanon
Malta
BELGIAN
South African
Cork
Yeltsin
Arafat
Clinton
European Cup
Brunswijk
Alexander Lebed
Amsterdam
Chechnya
Mexican
Mullally
Jordanian
Thomas Muster
Toshiba Classic
Jakarta
New York
Serb
Serbian
Palestinian
Afghan
Moslem
Czech
Rwandan
Florence
Albanian
Costa Rican
Indian
Croatian
Irish
Polish
Saudi
Nigerian
Hong Kong
Diana
Davis Cup
Paris
European
German
Italian
Scottish
Bangladesh
Bosnian
Australian
Russian
Israeli
Turkish
Dutch
Pakistani
British
Iran
Petr Korda
Daniel Vacek
Kenneth Carlsen
Rafter
Renzo Furlan
Benetton
EC
ITALIAN CUP
BRITISH MASTERS
Svetlana Masterkova
Damon Hill
Williams
G. Thorpe
Barcelona
Lloyd 's
Telekom
Motorola
Philadelphia Phillies
Braves
STUTTGART
ETA
Subaru Impreza
Ligier
PDS
Ekeus
Martina Hingis
Michael Schumacher
COMMACK
RAMALLAH
Jacques Villeneuve
Jana Novotna
Kimiko Date
Rosati
News Corp
Groningen
Necmettin Erbakan
Volkova
IRNA
S. KOREAN
Toyota Celica
Israel-PLO
Wellington
Sofia
Bangkok
Teleban
Iraqi Kurdish
AKRONAC-40
Gore
Davies
Syria
UAE
African
Kabul
Istanbul
Dhaka
Khartoum
Sri Lankan
Kurd
Tamil
Hutu
Burmese
Labour
Kurdistan Democratic Party
</s>
Bayern
Andrea Collinelli
Lindsay Davenport
Todd Woodbridge
Open
Alex O'Brien
Abuja
Algiers
Surabaya
Beijing
Patriotic Union of Kurdistan
SPD
Athens Newsroom
FDP
Greens
Rabobank
Emnid
Gaza
Palmans
Bosnian Serb
Cup
HOUSTON
Chechens
Palestinians
Moscow
Sudan
Grand Prix
Orioles
Florida
United
MaliVai Washington
Tour of the Netherlands
SAN JOSE
Poland
Essex
Nigeria
Chile
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE
Foreign Ministry
State Department
Boris Yeltsin
Sao Paulo
Grozny
Nicaraguan
Czech Republic
Thai
Brazil
Iraq
India
Russia
Sweden
Stansted
England
Spain
Australia
U.S.
Itar-Tass
Democratic
Baghdad
NATO
Ajax
Reuters
OSCE
Honda
United States
CANADIAN OPEN
Wayne Ferreira
Marcelo Rios
SAN FRANCISCO
COLORADO
CINCINNATI
CHICAGO
PHILADELPHIA
CALIFORNIA
Canadian
SAN DIEGO
TEXAS
DETROIT
OAKLAND
Cencor
40
20
0
−20
−40
−60
−40
−20
0
20
40



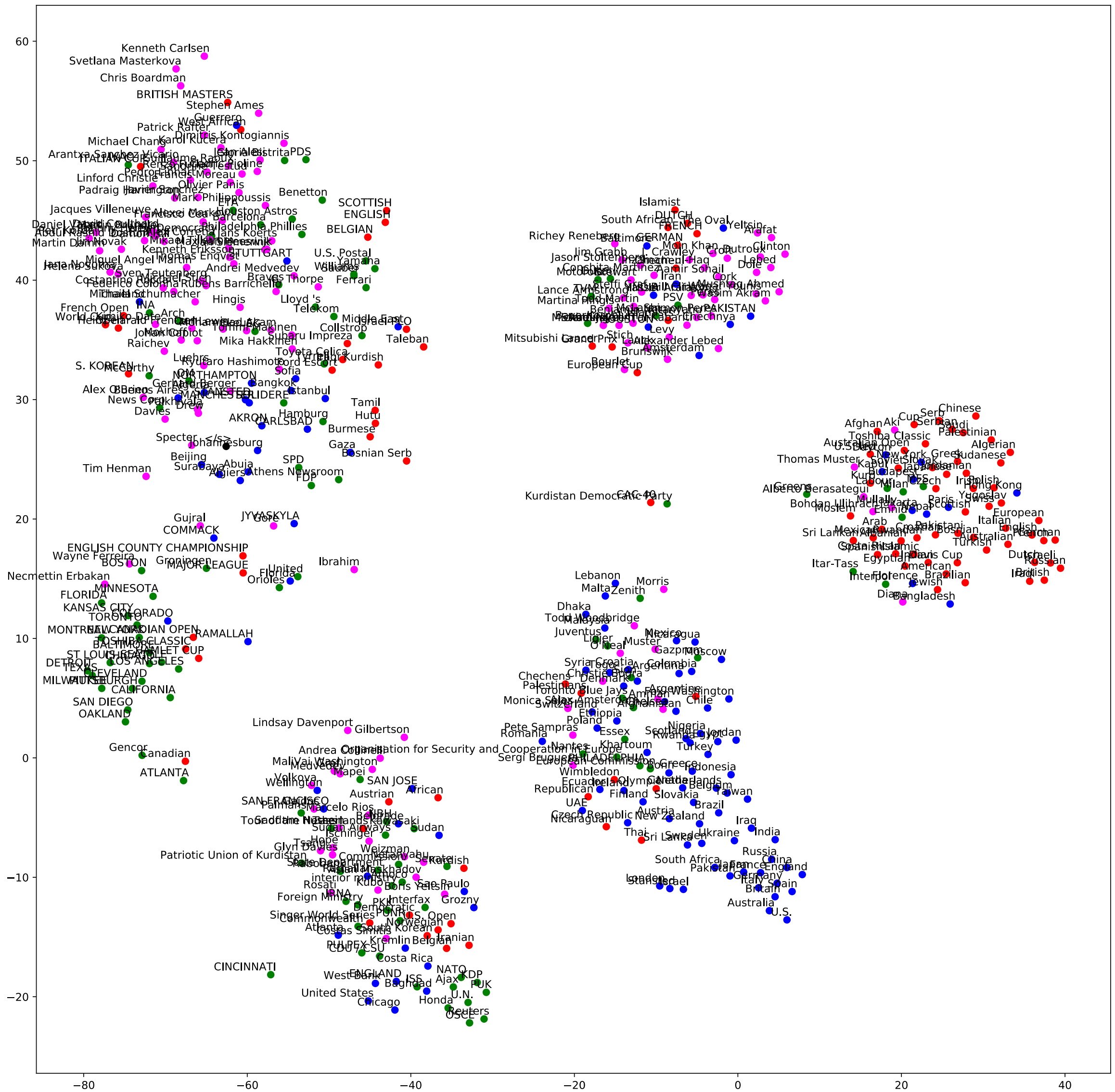
40
20
0
−20
−40
−60
−40
−20
0
20
40
ENGLISH
SCOTTISH
BRITISH MASTERS
BELGIAN
Lebanon
Malta
Dayton
The Oval
South African
FRENCH
DUTCH
Yeltsin
Arafat
Iran
Chechnya
Alexander Lebed
Amsterdam
European Cup
Richey Reneberg
Arantxa Sanchez Vicario
Kenneth Carlsen
Renzo Furlan
Sergi Bruguera
Svetlana Masterkova
Francisco Cea
Benetton
Jana Novotna
Jacques Villeneuve
Heinz-Harald Frentzen
Michael Stich
RAMALLAH
Thailand
Michael Schumacher
Glyn Davies
Groningen
S. KOREAN
Necmettin Erbakan
Davies
Luehrs
Specter
JYVASKYLA
Kurdistan Democratic Party
Johannesburg
French Open
World Cup
Bayern
Syria
UAE
SOLIDERE
Iraqi Kurdish
Kabul
Khartoum
Dhaka
Burmese
Kurd
Tamil
Abuja
Algiers
Emnid
SPD
FDP
Beijing
Patriotic Union of Kurdistan
Athens Newsroom
Greens
Gaza
Bosnian Serb
Sudan
Kornblum
O'Neal
Bourlet
Australian Open
State Department
Todd Woodbridge
Graf
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE
Olympic
MaliVai Washington
Andrea Collinelli
Kawasaki
Kurdish
Belgrade
Senate
Saddam Hussein
Commission
Tour of the Netherlands
Marcelo Rios
Sao Paulo
Boris Yeltsin
Grozny
PUNR
Singer
World Series
Weizman
Commonwealth
Atlanta
IRNA
Democratic
Belgian
Iranian
Foreign Ministry
Costa Rica
Costas Simitis
South Korean
NATO
KDP
PUK
Kremlin
Baghdad
Ajax
U.N.
CDU / CSU
Honda
Reuters
West Bank
United States
ENGLAND
Chicago
Morris
Zenith
Monica Seles
Pete Sampras
Aki
Cup
Chechens
Palestinians
Bogra
Christie
Gazprom
Moscow
Juventus
Muster
Nicaragua
Denmark
Croatia
Switzerland
Washington
Colombia
Scotland
Argentina
Argentine
Nigeria
Chile
European
Republican
Egypt
Jordan
Turkey
Greece
Indonesia
Finland
Ireland
Netherlands
Belgium
Wimbledon
Czech Republic
Slovakia
Nicaraguan
Brazil
Taiwan
Thai
Iraq
New Zealand
Ukraine
India
Israel
Sri Lanka
Russia
Stansted
South Africa
Pakistan
Japan
England
London
Italy
Germany
Spain
Britain
Australia
U.S.
Itar-Tass
Afghan
Serb
Chinese
Soviet
Saudi
Palestinian
Labour
Mexican
Greek
Spanish
Moslem
Polish
Hong Kong
Costa Rican
Croatian
Sri Lankan
Egyptian
Yugoslav
Islamic
Scottish
Interpol
Swiss
Italian
European
Diana
Jewish
American
German
Bosnian
Australian
English
French
Bangladesh
Turkish
Russian
Dutch
Iraqi
British
TOSHIBA CLASSIC
CANADIAN OPEN
Wayne Ferreira
SAN FRANCISCO
COLORADO
PHILADELPHIA
ATLANTA
ST LOUIS
NEW YORK
LOS ANGELES
SEATTLE
FLORIDA
CHICAGO
Canadian
DETROIT
OAKLAND
TEXAS
PITTSBURGH
Gencor
CALIFORNIA



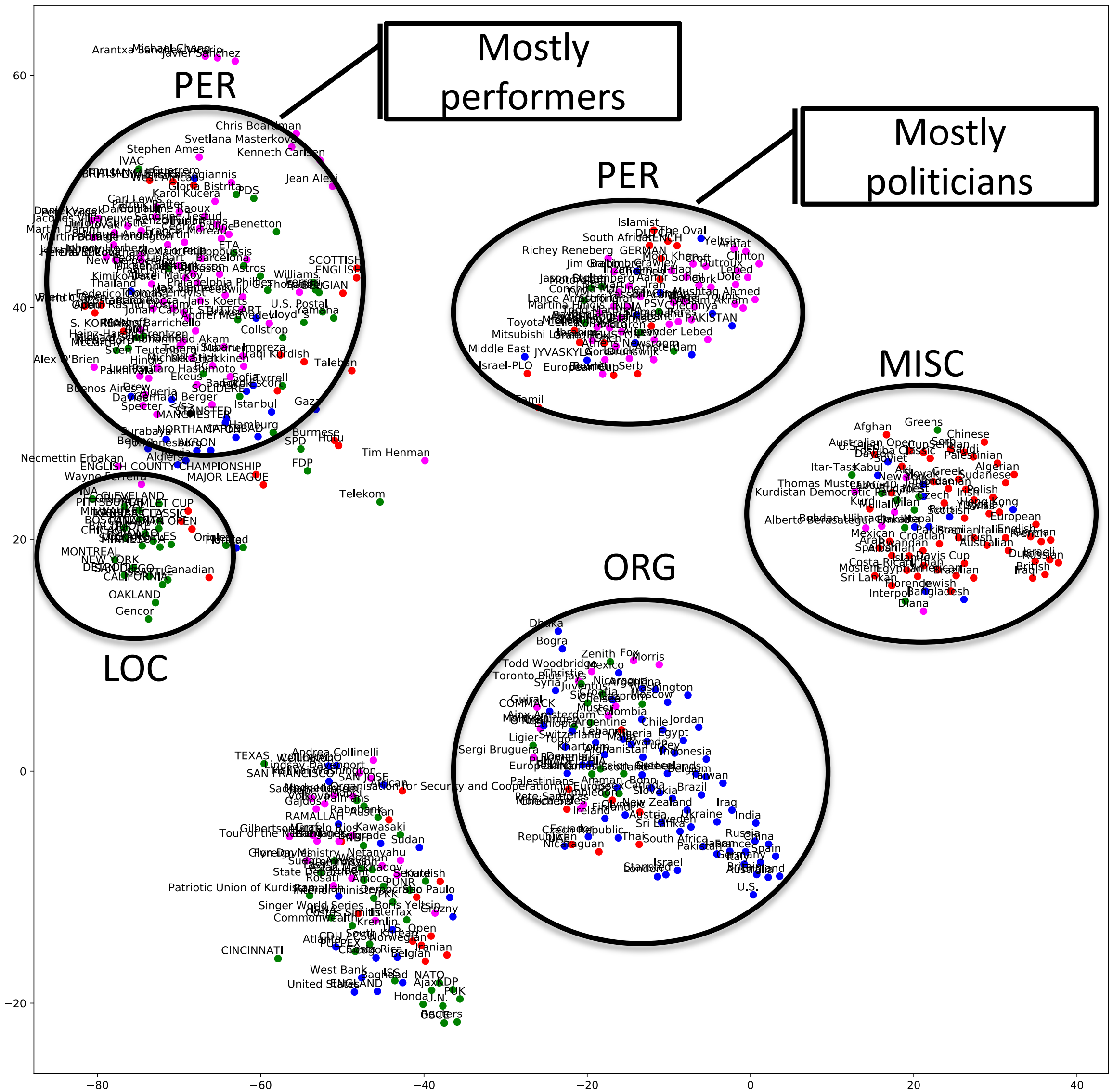
40
20
0
−20
−80
−60
−40
−20
0
20
40
Sergi Bruguera
BRITISH MASTERS
SCOTTISH
ENGLISH
BELGIAN
Lebanon
Malta
Tim Henman
S. KOREAN
Groningen
Necmettin Erbakan
World Cup
Tsang
SPD
FDP
Surabaya
Bosnian Serb
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE
Orioles
Florida
Greens
Sion
Morris
Islamic
Todd Woodbridge
Gajdos
Gilbertson
Olympic
Andrea Collinelli
Patriotic Union of Kurdistan
PHILADELPHIA
Tour of the Netherlands
CANADIAN OPEN
TOSHIBA CLASSIC
HAMLET CUP
Itar-Tass
Czech Republic
Nicaraguan
United States
West Bank
Kornblum
Bourlet
Ligier
Amsterdam
Budapest
Emnid
Florence
Afghan
Canadian
OAKLAND



60
50
40
30
20
10
0
−10
−20
−80
−60
−40
−20
0
20
40
Kenneth Carlsen
Svetlana Masterkova
Chris Boardman
BRITISH MASTERS
Karol Kucera
Dimitris Kontogiannis
Michael Chang
PDS
Toronto Blue Jays
BELGIAN
Barcelona
Benetton
David Coulthard
Williams
Houston Astros
EC
Yamaha
Telekom
Sergi Bruguera
Thailand
STUTTGART
TVM
Tommi Makinen
Tim Henman
RAMALLAH
S. KOREAN
Subaru Impreza
Middle East
Israel-PLO
French Open
World Cup
Taleban
Algeria
COMMACK
SOLIDERE
Buenos Aires
Istanbul
Tamil
Tsang
Davies
Specter
Johannesburg
Hamburg
Alex O'Brien
Beijing
CARLSBAD
Gaza
Groningen
Necmettin Erbakan
Surabaya
Abuja
Algiers
Athens Newsroom
Bosnian Serb
SPD
Gore
FDP
ENGLISH COUNTY CHAMPIONSHIP
MAJOR LEAGUE
Orioles
Florida
United
Kurdistan Democratic Party
Greens
PHILADELPHIA
Glyn Davies
Lebanon
Malta
Islamist
South Africa
DUTCH
GERMAN
FRENCH
Yeltsin
Arafat
Clinton
Dole
Lebed
Richey Reneberg
Lance Armstrong
PAKISTAN
Chechnya
Grand Prix
Alexander Lebed
European Union
Amsterdam
Afghan
Chinese
Thomas Muster
Palestinian
Alberto Berasategui
Algerian
Sudanese
U.S.-led
Kabul
Kurd
Mexican
Spanish
Nepal
Irish
Polish
Hong Kong
Moslem
Costa Rican
Sri Lankan
Indian
Pakistani
European
Florence
American
Interpol
Turkish
Jewish
Brazilian
Dutch
Israeli
Bangladesh
Iranian
British
Russian
Morris
Syria
Ligier
Zenith
Mexico
Christie
Juventus
Chechens
Moscow
Colombia
Monica Seles
Pete Sampras
Olympic
Switzerland
Scotland
Argentine
Nigeria
European Commission
Ajax Amsterdam
Organisation for Security and Cooperation in Europe
Chile
Rwanda
Jordan
Wimbledon
Turkey
Egypt
Republican
Ecuador
Finland
Greece
Indonesia
Ireland
Netherlands
Taiwan
Nicaragua
Czech Republic
Slovakia
Brazil
Iraq
New Zealand
Ukraine
India
Sweden
Israel
Russia
Itar-Tass
England
Italy
Spain
Britain
Australia
U.S.
Ibrahim
TOSHIBA CLASSIC
Wayne Ferreira
MILWAUKEE
BOSTON
CHICAGO
ATLANTA
MINNESOTA
BALTIMORE
SEATTLE
FLORIDA
CLEVELAND
NEW YORK
LOS ANGELES
CALIFORNIA
CINCINNATI
TEXAS
DETROIT
Canadian
Gencor
Andrea Collinelli
Graf
Lindsay Davenport
SAN FRANCISCO
Gilbertson
Saddam Hussein
Sudan
Patriotic Union of Kurdistan
Kawasaki
State Department
Kurdish
Sao Paulo
Grozny
Tour of the Netherlands
Singer World Series
Interior ministry
Interfax
U.S. Open
Foreign Ministry
Commonwealth
Costas Simitis
Kremlin
Costa Rica
CDU / CSU
NATO
KDP
PUK
ENGLAND
West Bank
Baghdad
Ajax
Honda
U.N.
United States
Chicago

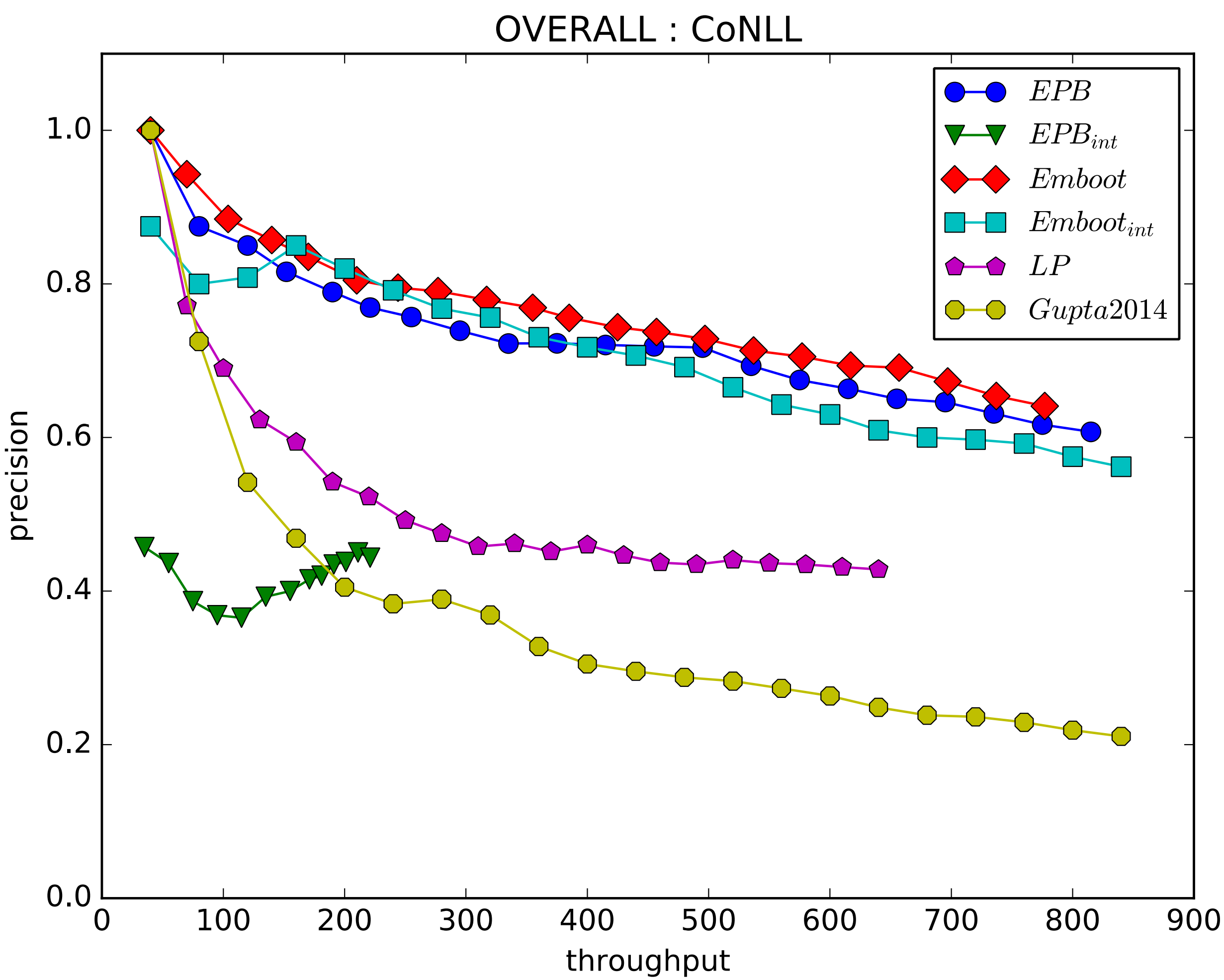


60
50
40
30
20
10
0
−10
−20
−80
−60
−40
−20
0
20
40
Kenneth Carlsen
Svetlana Masterkova
Chris Boardman
BRITISH MASTERS
Stephen Ames
Dimitris Kontogiannis
Michael Chang
Linford Christie
Olivier Panis
Benetton
SCOTTISH
ENGLISH
BELGIAN
Jacques Villeneuve
Martin Damm
Kenneth Eriksson
STUTTGART
U.S. Postal
Andrei Medvedev
Ferrari
Michael Schumacher
Hingis
Lloyd 's
Telekom
French Open
Middle East
Collstrop
Taleban
Raichev
Mika Hakkinen
Subaru Impreza
Toyota Celica
S. KOREAN
NORTHAMPTON
Gerhard Berger
Bangkok
Istanbul
Tamil
AKRON
CARLSBAD
Hamburg
Burmese
Johannesburg
Gaza
Bosnian Serb
Beijing
SPD
Abuja
Tim Henman
FDP
Gujral
COMMACK
JYVASKYLA
ENGLISH COUNTY CHAMPIONSHIP
Wayne Ferreira
BOSTON
Groningen
MAJOR LEAGUE
Orioles
Ibrahim
Necmettin Erbakan
MINNESOTA
FLORIDA
KANSAS CITY
COLORADO
TORONTO
MONTREAL
CANADIAN OPEN
RAMALLAH
ST LOUIS
DETROIT
TEXAS
CLEVELAND
LOS ANGELES
MILWAUKEE
CALIFORNIA
SAN DIEGO
OAKLAND
Gencor
Canadian
ATLANTA
Lindsay Davenport
Gilbertson
Organisation for Security and Cooperation in Europe
MaliVai Washington
Volkova
Wellington
SAN JOSE
Austrian
African
SAN FRANCISCO
Sudan
Patriotic Union of Kurdistan
Weizman
Kurdish
Rosati
Foreign Ministry
San Paulo
Boris Yeltsin
Interfax
Grozny
Democratic
Singer World Series
Commonwealth
Norwegian
South Korean
Iranian
Kremlin
Belgian
Costa Rica
CINCINNATI
NATO
ENGLAND
West Bank
ISS
Baghdad
Ajax
KDP
PUK
United States
U.N.
Chicago
Honda
Reuters
OSCE
Islamist
South Africa
The Oval
Yeltsin
Arafat
Richey Reneberg
GERMAN
FRENCH
Clinton
Jim Grabb
Moin Khan
Crawley
Jason Stoltenberg
Dole
Lebed
Conchita Martinez
Iran
York
Lance Armstrong
Martina Hingis
PSV
Wasim Akram
Ahmed
Mohammad
Chechnya
PAKISTAN
Levy
Stich
Mitsubishi Grand Prix
Alexander Lebed
Brunswik
Amsterdam
European Cup
Chinese
Serb
Afghan
Palestinian
Algerian
Thomas Muster
Sudanese
Serbian
Kurdish
Milan
Czech
Polish
Hong Kong
Yugoslav
Bohdan Ulihrach
European
Arab
Italian
Pakistani
Sri Lankan
English
Turkish
Australian
Davis Cup
Dutch
Israeli
Itar-Tass
Egyptian
American
Jewish
Brazilian
Florence
Bangladesh
British
Russian
Diana
Iraqi
Kurdistan Democratic Party
CAC-40
Lebanon
Malta
Zenith
Morris
Dhaka
Todd Woodbridge
Juventus
Mexico
Nicaragua
Muster
Gazprom
Moscow
Syria
Croatia
Colombia
Chechens
Palestinians
Blue Jays
Toronto
Monica Seles
Amsterdam
Argentina
Washington
Switzerland
Chile
Ethiopia
Poland
Pete Sampras
Romania
Essex
Nigeria
Scotland
Jordan
Khartoum
Turkey
Nantes
Sergi Bruguera
European Commission
Greece
Wimbledon
Indonesia
Ecuador
Republican
Finland
Slovakia
Belgium
Taiwan
UAE
Brazil
Austria
Czech Republic
New Zealand
Nicaraguan
Iraq
Thai
Sri Lanka
Sweden
Ukraine
India
Russia
South Africa
China
England
Pakistan
France
Germany
Italy
Spain
Britain
London
Israel
Australia
U.S.

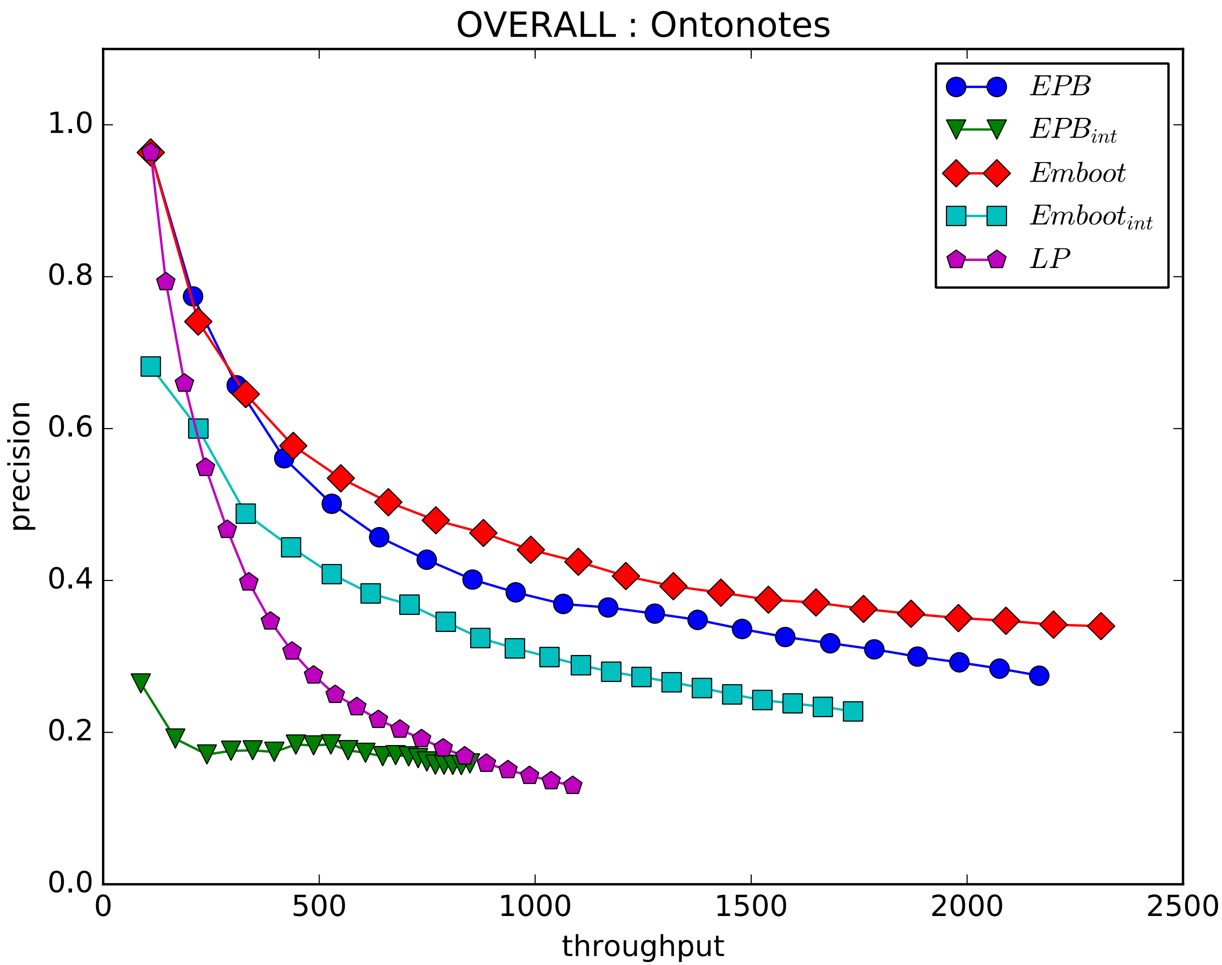


PER
Mostly performers
PER
Mostly politicians
MISC
ORG
LOC

# Results, CoNLL dataset (4 classes)



# Results, OntoNotes (11 classes)



# Summary

- A lightly-supervised approach that jointly learns representations for entities and patterns that extract them
- State-of-the-art results for semi-supervised learning
- The rules can be edited by domain experts, and this leads to further improvements in performance (not shown)

# Take-home message

- For large, inter-disciplinary projects we need to move beyond "black-box" methods to approaches that produce globally interpretable models
- We can produce such interpretable models using deep learning (best of both worlds?)

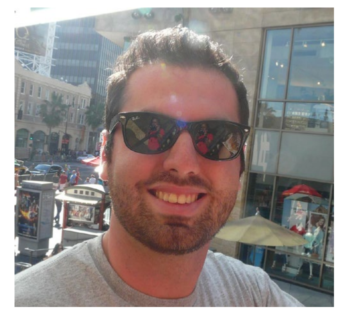
# Bibliography

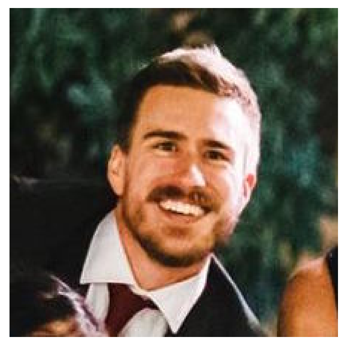
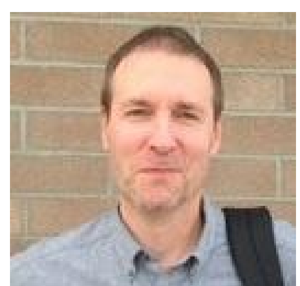

1. M.A. Valenzuela-Escarcega, G. Hahn-Powell, T. Hicks, and M. Surdeanu. *A Domain-independent Rule-based Framework for Event Extraction*. ACL-IJCNLP, 2015.
2. M. A. Valenzuela-Escarcega, G. Hahn-Powell, and M. Surdeanu. *Description of the Odin Event Extraction Framework and Rule Language*. arXiv:1509.07513, 2015.
3. D. Bell, G. Hahn-Powell, M. A. Valenzuela-Escarcega, and M. Surdeanu. *An Investigation of Coreference Phenomena in the Biomedical Domain*. LREC, 2016.
4. M. A. Valenzuela-Escarcega, G. Hahn-Powell, and M. Surdeanu. *Odin's Runes: A Rule Language for Information Extraction*. LREC, 2016.
5. G. Hahn-Powell, D. Bell, M. A. Valenzuela-Escarcega, and M. Surdeanu. *This before That: Causal Precedence in the Biomedical Domain*. BioNLP, 2016.
6. Large-scale Automated Machine Reading Discovers New Cancer Driving Mechanisms. Valenzuela-Escarcega, M. A.; Babur, O.; Hahn-Powell, G.; Bell, D.; Hicks, T.; Noriega-Atala, E.; Wang, X.; Surdeanu, M.; Demir, E.; and Morrison, C. T. Database: The Journal of Biological Databases and Curation. 2018.
7. Lightly Supervised Representation Learning with Global Interpretability. Zupon, A.; Alexeeva, M.; Valenzuela-Escarcega, M. A.; Nagesh, A.; and Surdeanu, M. In Proceedings of the 3rd Workshop on Structured Prediction for Natural Language Processing, 2019.
8. Z. Tang, M.A. Valenzuela-Escarcega, and M. Surdeanu. Exploring Interpretability in Event Extraction: Multitask Learning of a Neural Event Classifier and an Explanation Decoder. In preparation.

# Many thanks to my collaborators!








# THANK YOU!
## QUESTIONS?

# Conflict of interest disclosure

- M. Surdeanu discloses a financial interest in Lum.ai. This interest has been disclosed to the University of Arizona Institutional Review Committee and is being managed in accordance with its conflict of interest policies.