



Texts are Knowledge

Christopher Manning
AKBC 2013 Workshop

Gabor Angeli, Jonathan Berant, Bill MacCartney Mihai Surdeanu, Arun Chaganty,
Angel Chang, Kevin Aju Scaria, Mengqiu Wang, Peter Clark, Justin Lewis, Brittany
Harding, Reschke, Julie Tibshirani, Jean Wu, Osbert Bastani, Keith Siillats



Jeremy Zawodny [ex-Yahoo!/Craigslist] sez ...

Back in the late 90s when I was building things that passed for *knowledge management tools* at Marathon Oil, there was all this talk about *knowledge workers*. These were people who'd have vast quantities of knowledge at their fingertips. All they needed was ways to organize, classify, search, and collaborate.

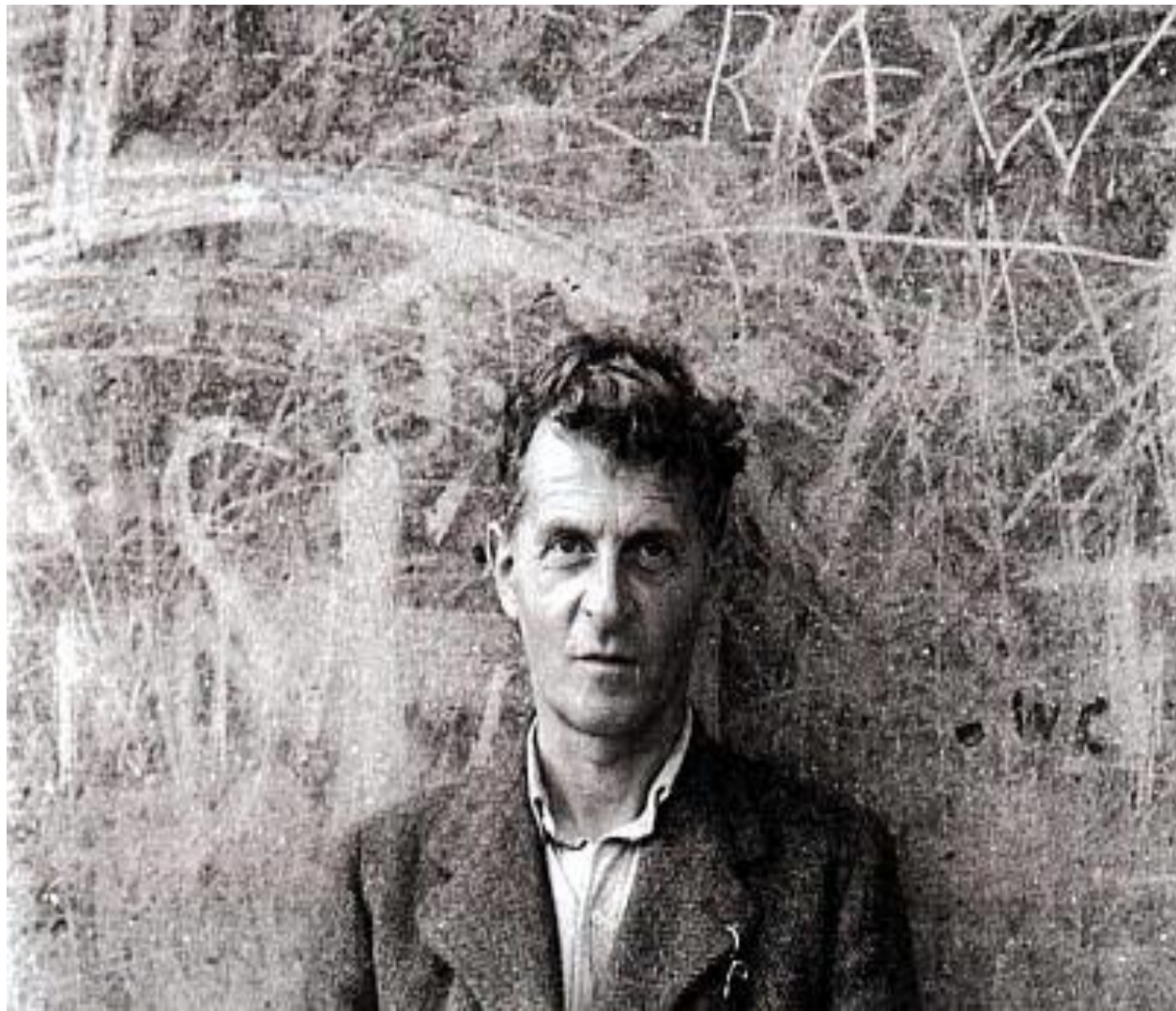
I think we've made it. But the information isn't organized like I had envisioned a few years ago. It's just this big ugly mess known as **The Web**. Lots of pockets of information from mailing lists, weblogs, communities, and company web sites are loosely tied together by hyperlinks. **There's no grand schema or centralized database. There's little structure or quality control. No global vocabulary.**

But even with all that going against it, it's all indexed and easily searchable thanks largely to Google and the companies that preceded it (Altavista, Yahoo, etc.). Most of the time it actually works. Amazing!



From language to knowledge bases – Still the goal?

- For humans, going from the largely unstructured language on the web to information is effortlessly easy
 - But for computers, it's rather **difficult!**
 - This has suggested to many that if we're going to produce the next generation of intelligent agents, which can make decisions on our behalf
 - Answering our routine email
 - Booking our next trip to Fiji
- then we still first need to construct knowledge bases
- To go *from languages to information*





“You say: the point isn’t the word, but its meaning, and you think of the meaning as a thing of the same kind as the word, though also different from the word. Here the word, there the meaning. The money, and the cow that you can buy with it. (But contrast: money, and its use.)”



1. Inference directly in text: Natural Logic (van Benthem 2008, MacCartney & Manning 2009)

Q *Beyoncé Knowles's husband is X.*

A *Beyoncé's marriage to rapper Jay-Z and portrayal of Etta James in Cadillac Records (2008) influenced her third album I Am... Sasha Fierce (2008).*





Inference directly in text: Natural Logic (van Benthem 2008, MacCartney & Manning 2009)

- P *James Dean refused to move without blue jeans.*
H *James Byron Dean didn't dance without pants*

The example is contrived, but it compactly exhibits containment, exclusion, and implicativity





Step 1: Alignment

P	James	refused			move	without	blue	jeans
H	Dean	to						
	James		did	n't	dance	without		pants
	Byron							
	Dean							
edit index	1	2	3	4	5	6	7	8
edit type	SUB	DEL	INS	INS	SUB	MAT	DEL	SUB

- Alignment as sequence of *atomic phrase edits*
- Ordering of edits defines path through intermediate forms
 - Need not correspond to sentence order
- Decomposes problem into atomic inference problems



Step 2: Lexical entailment classification

- Goal: predict entailment relation for each edit, based solely on lexical features, independent of context
- Done using lexical resources & machine learning
- Feature representation:
 - WordNet features: synonymy ($=$), hyponymy (\sqsubset/\sqsupset), antonymy ($|$)
 - Other relatedness features: Jiang-Conrath (WN-based), NomBank
 - Fallback: string similarity (based on Levenshtein edit distance)
 - Also lexical category, quantifier category, implication signature
- Decision tree classifier



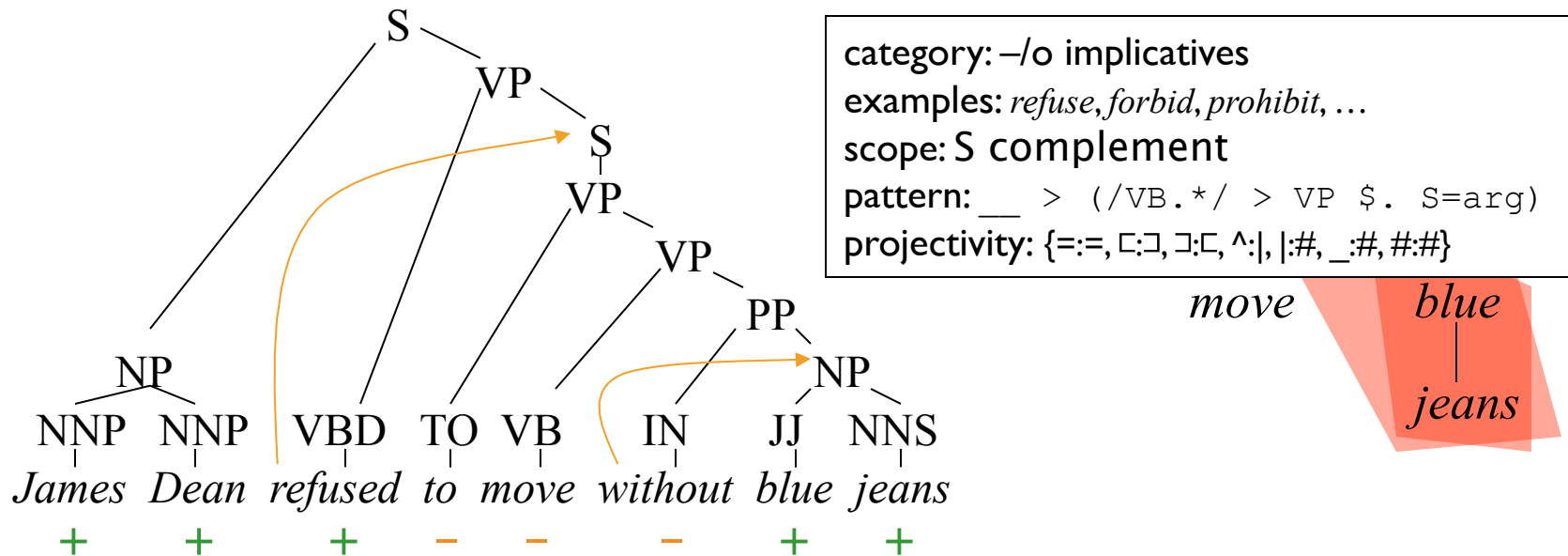
Step 2: Lexical entailment classification

P	<i>James Dean</i>	<i>refused to</i>			<i>move</i>	<i>without</i>	<i>blue</i>	<i>jeans</i>
H	<i>James Byron Dean</i>		<i>did</i>	<i>n't</i>	<i>dance</i>	<i>without</i>		<i>pants</i>
edit index	1	2	3	4	5	6	7	8
edit type	SUB	DEL	INS	INS	SUB	MAT	DEL	SUB
lex feats	strsim= 0.67	implic: -/o	cat:aux	cat:neg	hypo			hyper
lex entrel	=		=	^	⊃	=	⊆	⊆



Step 3: Sentence semantic analysis

- Identify items w/ special projectivity & determine scope

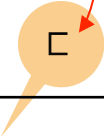


Specify scope in trees



Step 4: Entailment projection

P	<i>James Dean</i>	<i>refused to</i>			<i>move</i>	<i>without</i>	<i>blue</i>	<i>jeans</i>
H	<i>James Byron Dean</i>		<i>did</i>	<i>n't</i>	<i>dance</i>	<i>without</i>		<i>pants</i>
edit index	1	2	3	4	5	6	7	8
edit type	SUB	DEL	INS	INS	SUB	MAT	DEL	SUB
lex feats	strsim= 0.67	implic: -/o	cat:aux	cat:neg	hypo			hyper
lex entrel	=		=	^	⊃	=	⊃	⊃
projectivity	↑	↑	↑	↑	↓	↓	↑	↑
atomic entrel	=		=	^	⊃	=	⊃	⊃



inversion



Step 5: Entailment composition

P	James Dean	refused to			move	without	blue	jeans
H	James Byron Dean		did	n't	dance	without		pants
edit index	1	2	3	4	5	6	7	8
edit type	SUB	DEL	INS	INS	SUB	MAT	DEL	SUB
lex feats	strsim= 0.67	implic: -/o	cat:aux	For example:				hyper
lex entrel	=		=	<div style="border: 1px solid black; padding: 5px; margin: 5px;"> <p>fish human → fish ⊆ nonhuman</p> <p style="text-align: center;">human ^ nonhuman ↓</p> </div>				⊆
projectivity	↑	↑	↑					↑
atomic entrel	=		=	^	⊆	=	⊆	⊆
composition	=			⊆	⊆	⊆	⊆	⊆

final answer



The problem

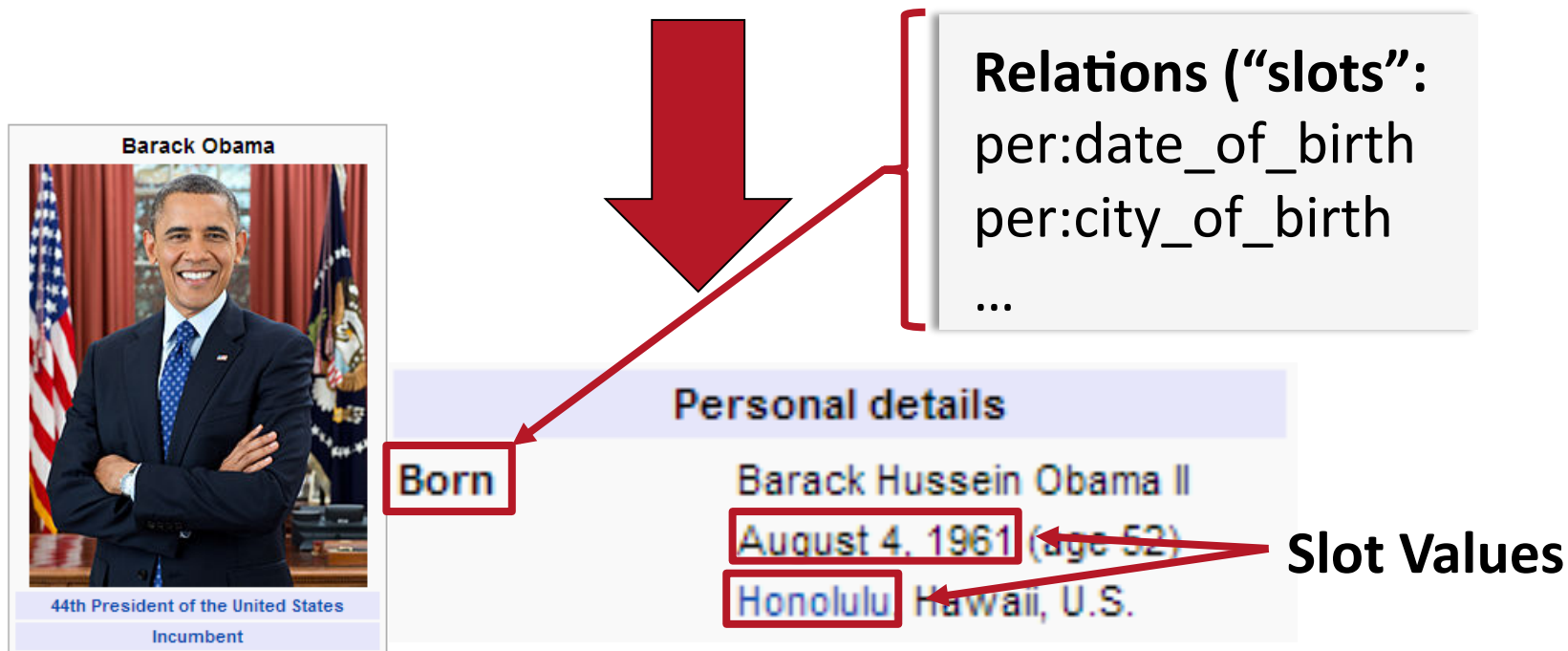
Much of what we could and couldn't do depended on knowledge of word and multi-word relationships ...

which we mainly got from knowledge-bases (WordNet, Freebase, ...)



2. Knowledge Base Population (NIST TAC task) [Angeli et al. 2013]

“Obama was born on August 4, 1961, at Kapi’olani Maternity & Gynecological Hospital (now Kapi’olani Medical Center for Women and Children) in Honolulu, Hawaii.”





The opportunity

- There is now a lot of content in various forms that naturally aligns with human language material
- In the old days, we were either doing unsupervised clustering or doing supervised machine learning over painstakingly hand-annotated natural language examples
- Now, we can hope to learn based on found, naturally generated content



Distantly Supervised Learning



44th President of the United States
Incumbent
Assumed office
January 20, 2009
Vice President Joe Biden
Preceded by George W. Bush
United States Senator
from Illinois
In office
January 3, 2005 – November 16, 2008
Preceded by Peter Fitzgerald
Succeeded by Roland Burris

Lots of text »

Barack Obama is the 44th and current President of the United States.

« Millions of infoboxes in Wikipedia

Over 22 million entities in Freebase, each with multiple relations »



Person /people/person	X
Place of birth /people/person/place_of_birth	
Honolulu	

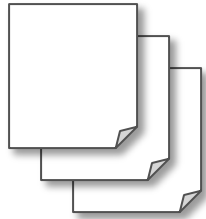


Traditional Supervised Learning





Not so fast! Only distant supervision



Barack Obama is the 44th and current President of the United States.

United States President Barack Obama meets with Chinese Vice President Xi Jinping today.

Obama was born in the United States just as he has always said.

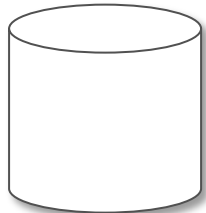
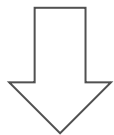
Obama ran for the United States Senate in 2004.

EmployedBy

?
EmployedBy

BornIn

--



Same Query

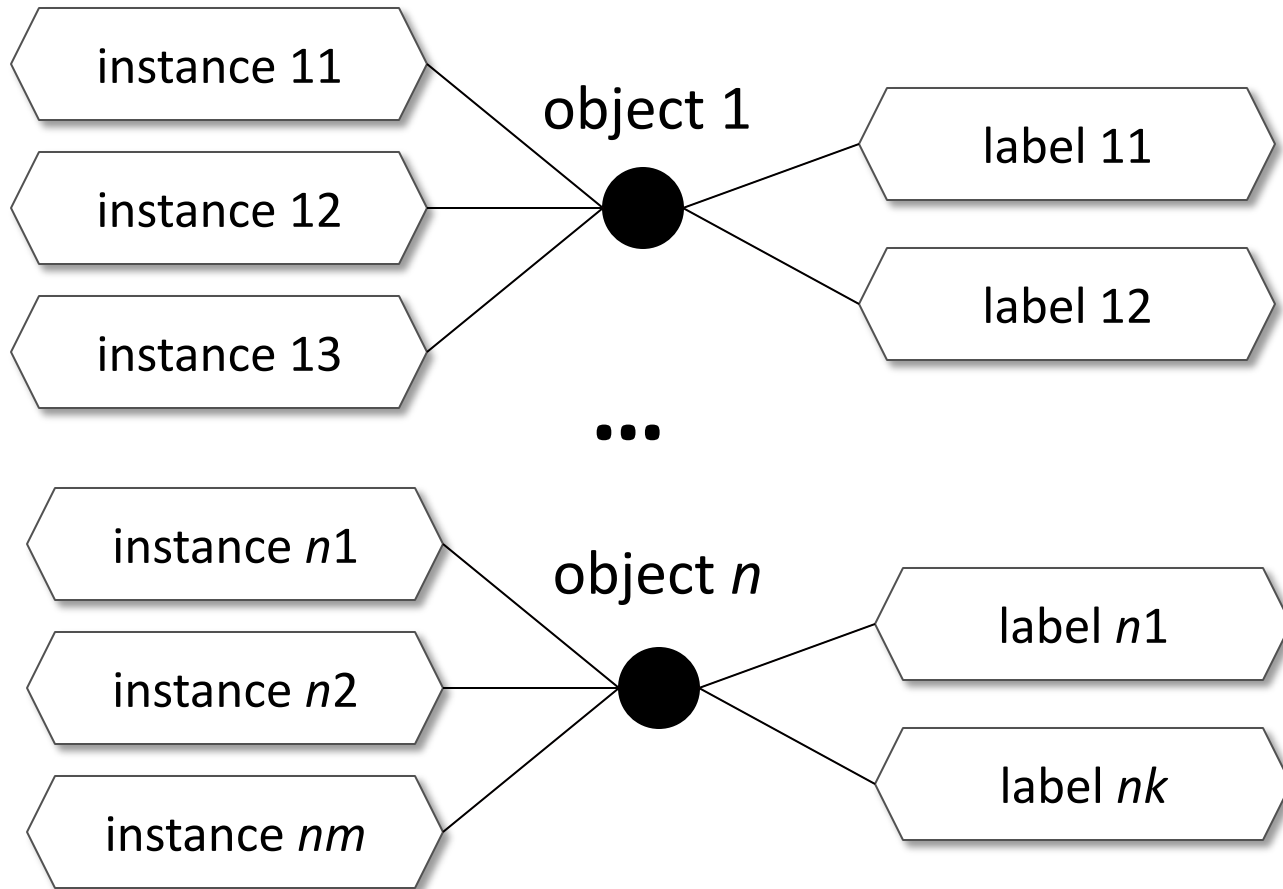
EmployedBy(Barack Obama, United States)

BornIn (Barack Obama, United States)

...



Multi-instance Multi-label (MIML) Learning



(Surdeanu et. al 2012)



MIML for Relation Extraction

Barack Obama is the 44th and current President of the United States.

United States President Barack Obama meets with Chinese Vice President Xi Jinping today.

Obama was born in the United States just as he has always said.

Obama ran for the United States Senate in 2004.

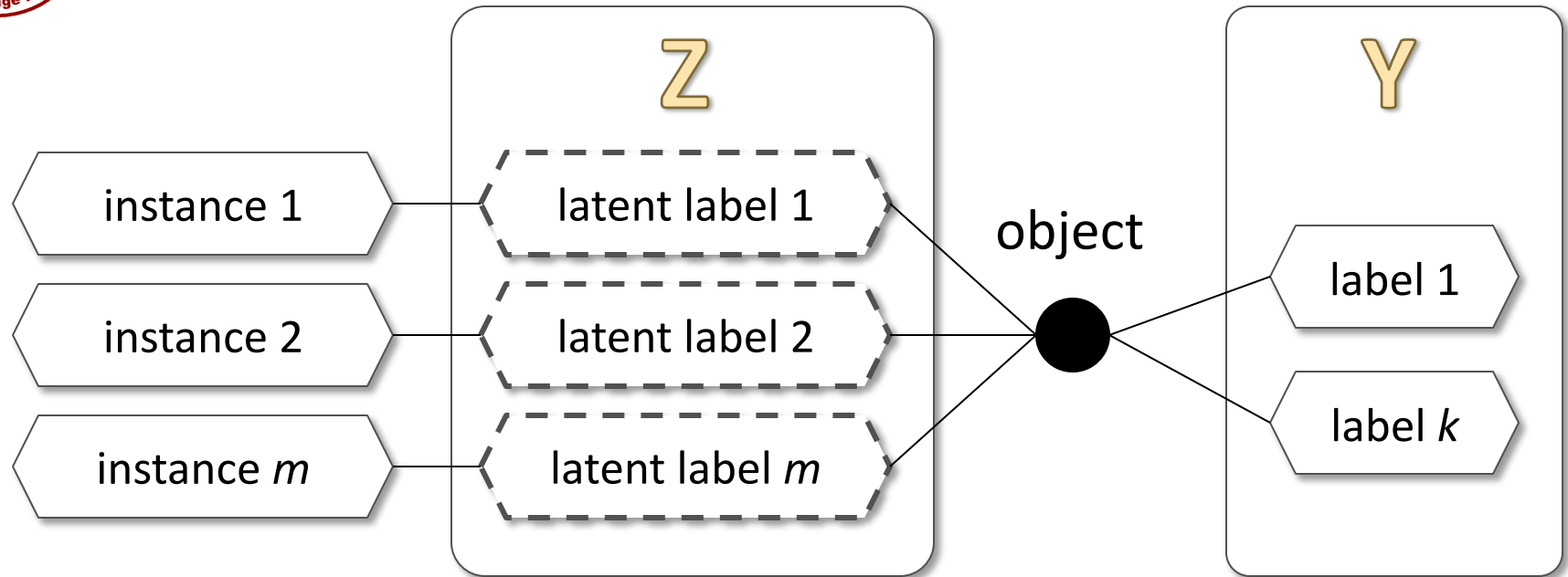
(Barack Obama,
United States)

EmployedBy

BornIn



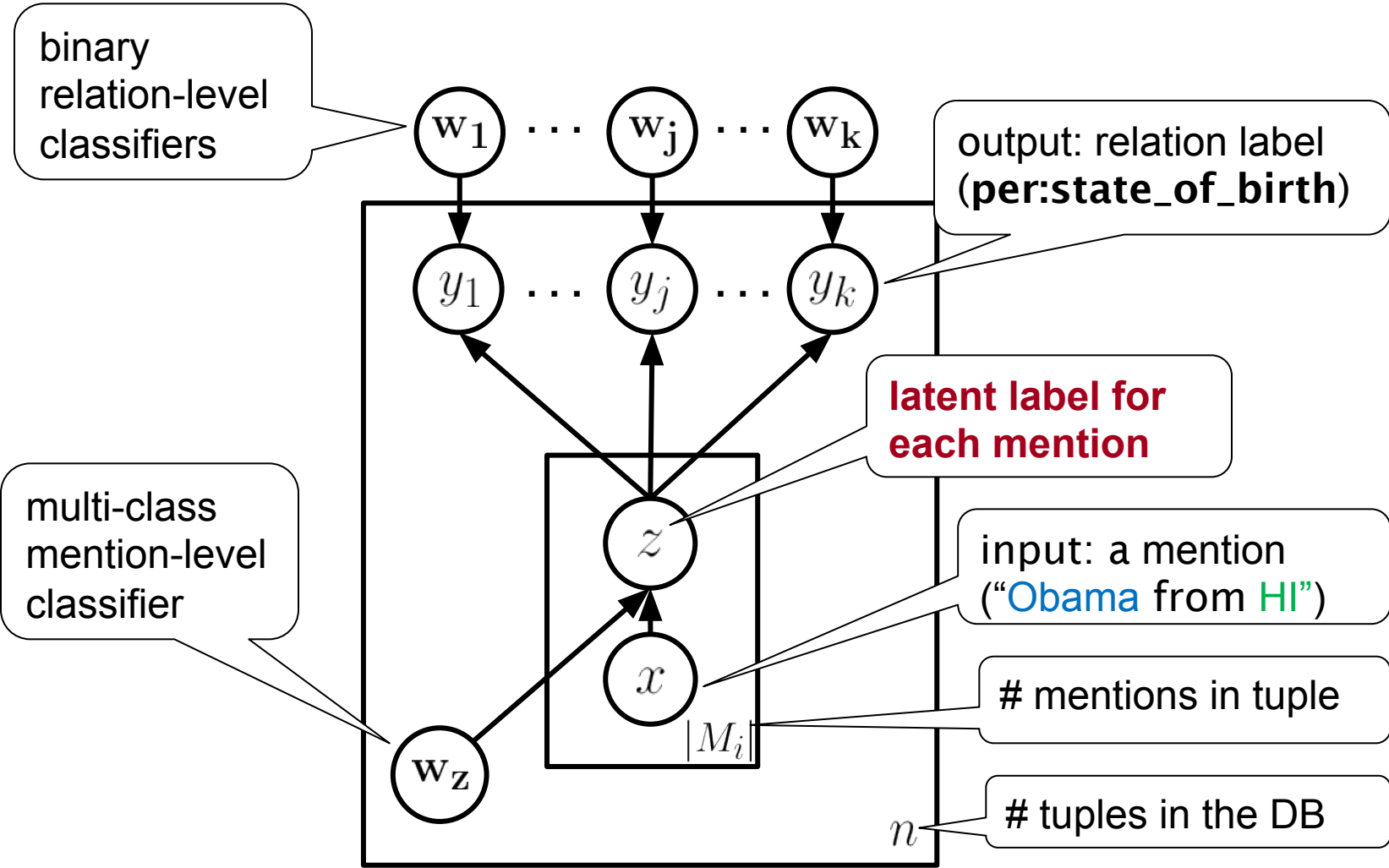
Model Intuition



- Jointly trained using discriminative EM:
 - E-step: assign latent labels using current θ_z and θ_y .
 - M-step: estimate θ_z and θ_y using the current latent labels.

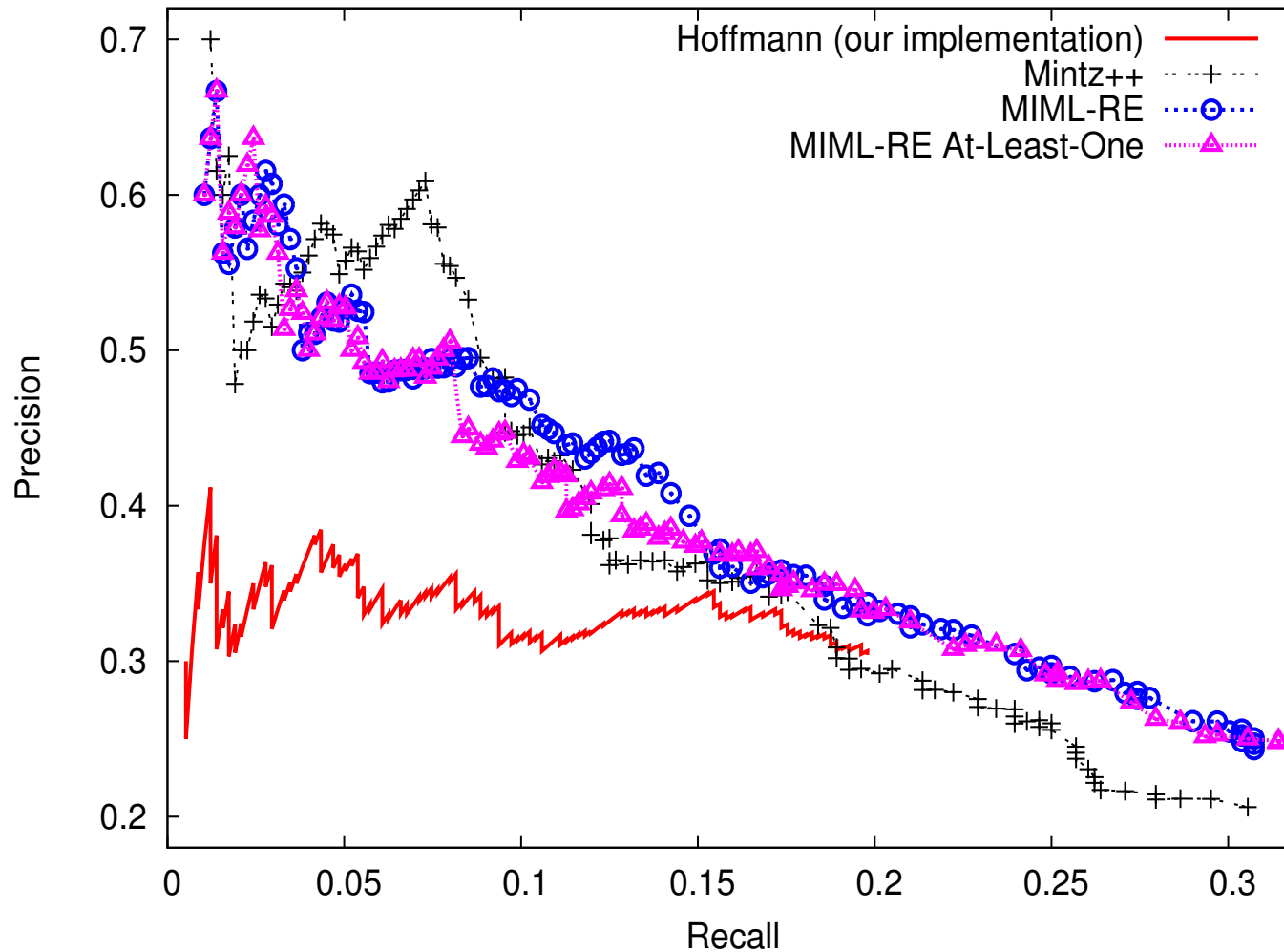


Plate Diagram



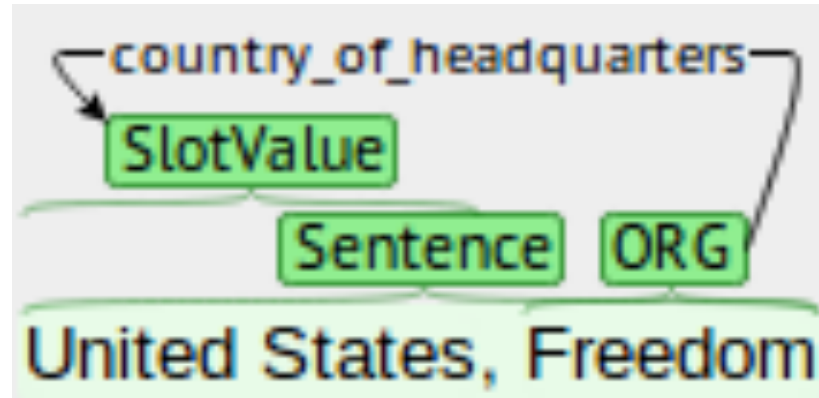


Post-hoc evaluation on TAC KBP 2011





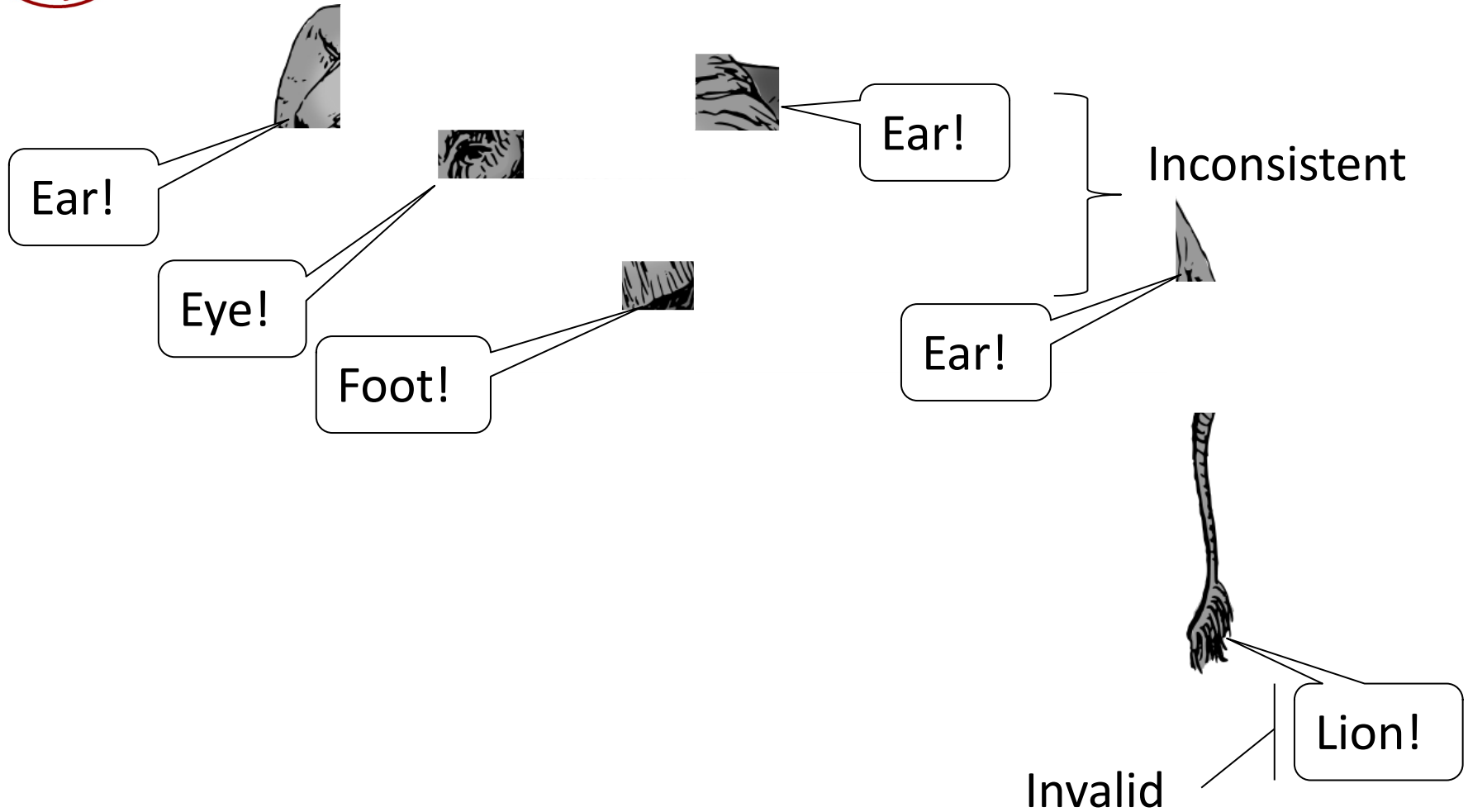
Error Analysis



* Actual output from Stanford's 2013 submission



Consistency: Relation Extractor's View





Consistency as a CSP

Nodes: Slot fills proposed by relation extractor

- ✓ Obama born in Hawaii
- ✓ Obama born in Kenya
- ✓ Obama age -12
- ✓ Obama origin American



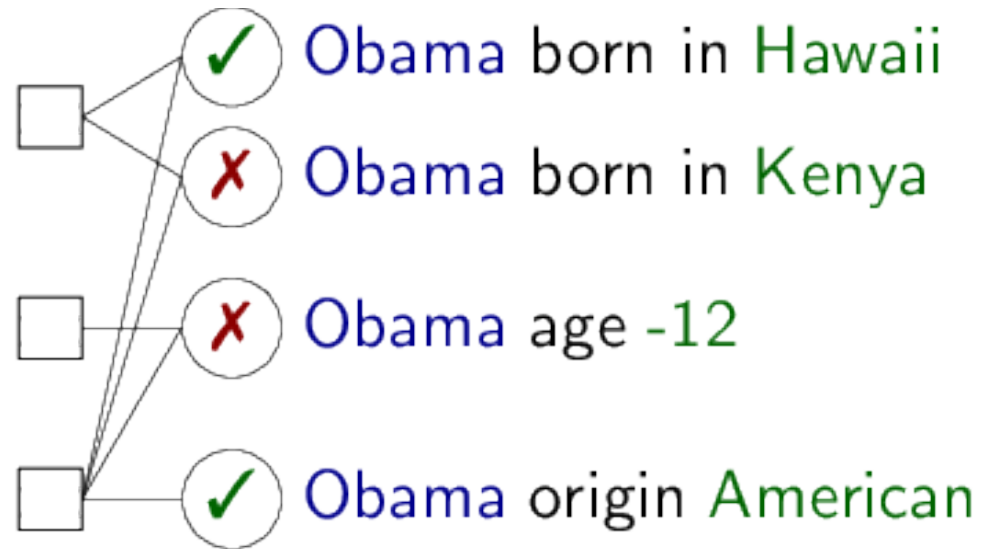
Consistency as a CSP

End Result: More robust to classification noise

Hawaii \cap Kenya = \emptyset

-12 invalid

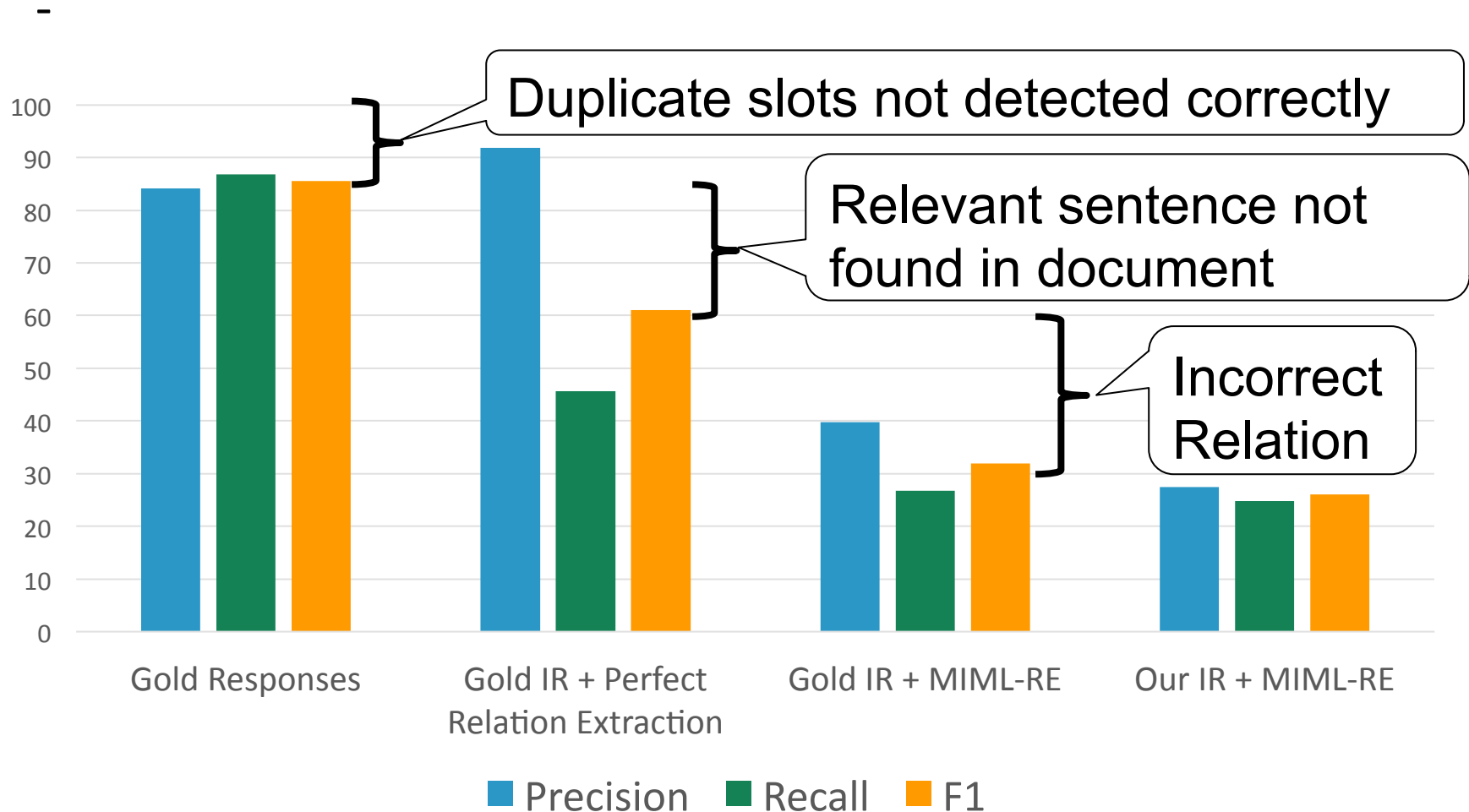
birthplace
matches origin





Error Analysis: Quantitative

Our IR + MIML-RE: our KBP system as run in the evaluation





3. What our goal should be: Not just knowledge bases but inference

Populating knowledge-bases is a method

If it has a use, it's to enable easier inferences!



Test-case: modeling processes

(Berant et al. 2013)

- Processes are all around us.
- Example of a biological process:

Also, slippage can occur during DNA replication, such that the template shifts with respect to the new complementary strand, and a part of the template strand is either skipped by the replication machinery or used twice as a template. As a result, a segment of DNA is deleted or duplicated.



Answering non-factoid questions

Also, slippage can occur during DNA replication, such that the template shifts with respect to the new complementary strand, and a part of the template strand is either skipped by the replication machinery or used twice as a template. As a result, a segment of DNA is deleted or duplicated.

In DNA replication, what causes a segment of DNA to be deleted?

- A. A part of the template strand being skipped by the replication machinery.*
- B. A part of the template strand being used twice as a template.*



Answering non-factoid questions

Also, slippage can occur during DNA replication, such that the template shifts with respect to the new complementary strand, and a part of the template strand is either skipped by the replication machinery or used twice as a template. As a result, a segment of DNA is deleted or duplicated.

In DNA replication, what causes a segment of DNA to be deleted?

- A. *A part of the template strand being skipped by the replication machinery.*
- B. *A part of the template strand being used twice as a template.*

Entities

Events



Answering non-factoid questions

Also, slippage can occur during DNA replication, such that the template shifts with respect to the new complementary strand, and a part of the template strand is either skipped by the replication machinery or used twice as a template. As a result, a segment of DNA is deleted or duplicated.

In DNA replication, what causes a segment of DNA to be deleted?

- A. A part of the template strand being skipped by the replication machinery.
- B. A part of the template strand being used twice as a template.

Entities

Events



Answering non-factoid questions

Also, slippage can occur during DNA replication, such that the template shifts with respect to the new complementary strand, and a part of the template strand is either skipped by the replication machinery or used twice as a template. As a result, a segment of DNA is deleted or duplicated.

In DNA replication, what causes a segment of DNA to be deleted?

- A. *A part of the template strand being skipped by the replication machinery.*
- B. *A part of the template strand being used twice as a template.*

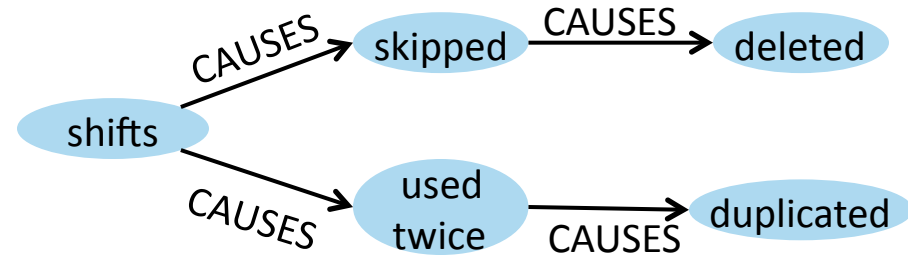
Entities

Events



Process structures

Also, slippage can occur during DNA replication, such that the template **shifts** with respect to the new complementary strand, and a part of the template strand is either **skipped** by the replication machinery or **used** twice as a template. As a result, a segment of DNA is **deleted** or **duplicated**.





Example: Biology AP Exam

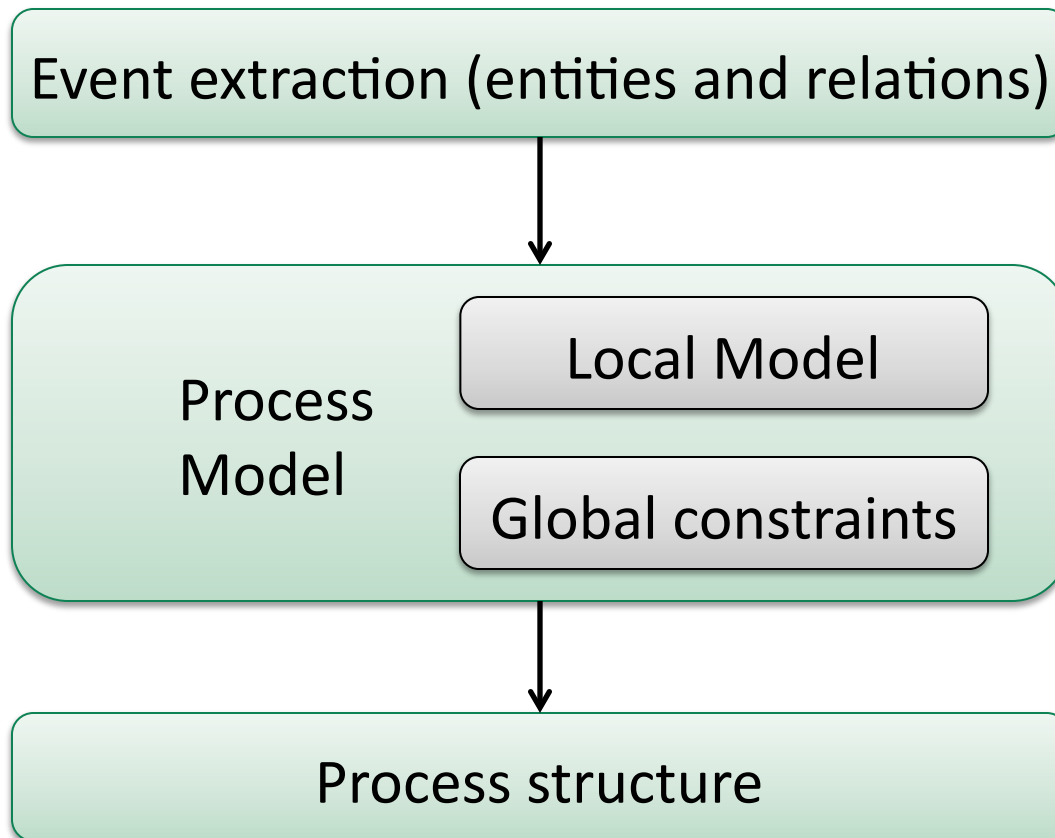
In the development of a seedling, which of the following will be the last to occur?

- A. Initiation of the breakdown of the food reserve
- B. Initiation of cell division in the root meristem
- C. Emergence of the root
- D. Emergence and greening of the first true foliage leaves
- E. Imbibition of water by the seed

Actual order: E A C B D



Setup





Local event relation classifier

- Input: a pair of events i and j
- Output: a relation
 - 11 relations including
 - Temporal ordering
 - Causal relations – cause/enable
 - Coreference
- Features:
 - Some adapted from previous work
 - A few new ones – to combat sparseness
 - Especially exploiting function words that suggest relations (e.g., *hence*)



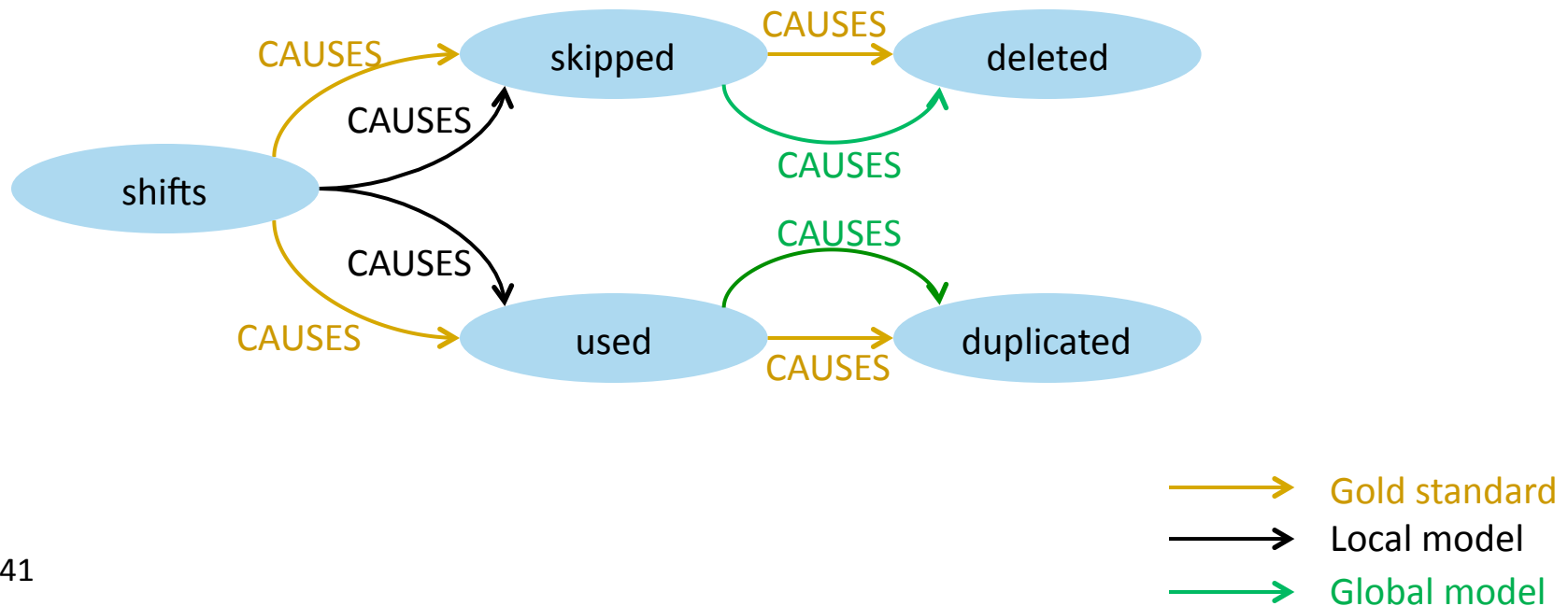
Global constraints

- Ensure consistency using hard constraints.
- Favor structures using soft constraints.
- Mainly in three flavors
 - Connectivity constraint
 - The graph of events must be connected
 - Chain constraints
 - 90% of events have degree ≤ 2
 - Event triple constraints
 - E.g.: Event co-reference is transitive
- Enforced by formulating the problem as an ILP.



Connectivity constraint

- ILP formulation based on graph flow for dependency parsing by Martins et. al. (2009).





Dataset

- 148 paragraphs describing processes.
 - From the textbook “Biology” (Campbell and Reece, 2005).
- Annotated by trained biologists
- 70-30% train-test split.
- Baselines
 - All-Prev: For every pair of adjacent triggers in text, predict PREV relation.
 - Local_{base}: MaxEnt classifier with features from previous work.
 - Chain: Join adjacent events with highest probability relation



Event-event relation results

	P	R	F1
All-Prev	34.1	32.0	33.0
Local _{base}	52.1	43.9	47.6
Local	54.7	48.3†	51.3
Chain	56.1	52.6†‡	54.3†
Global	56.2	54.0†‡	55.0†‡

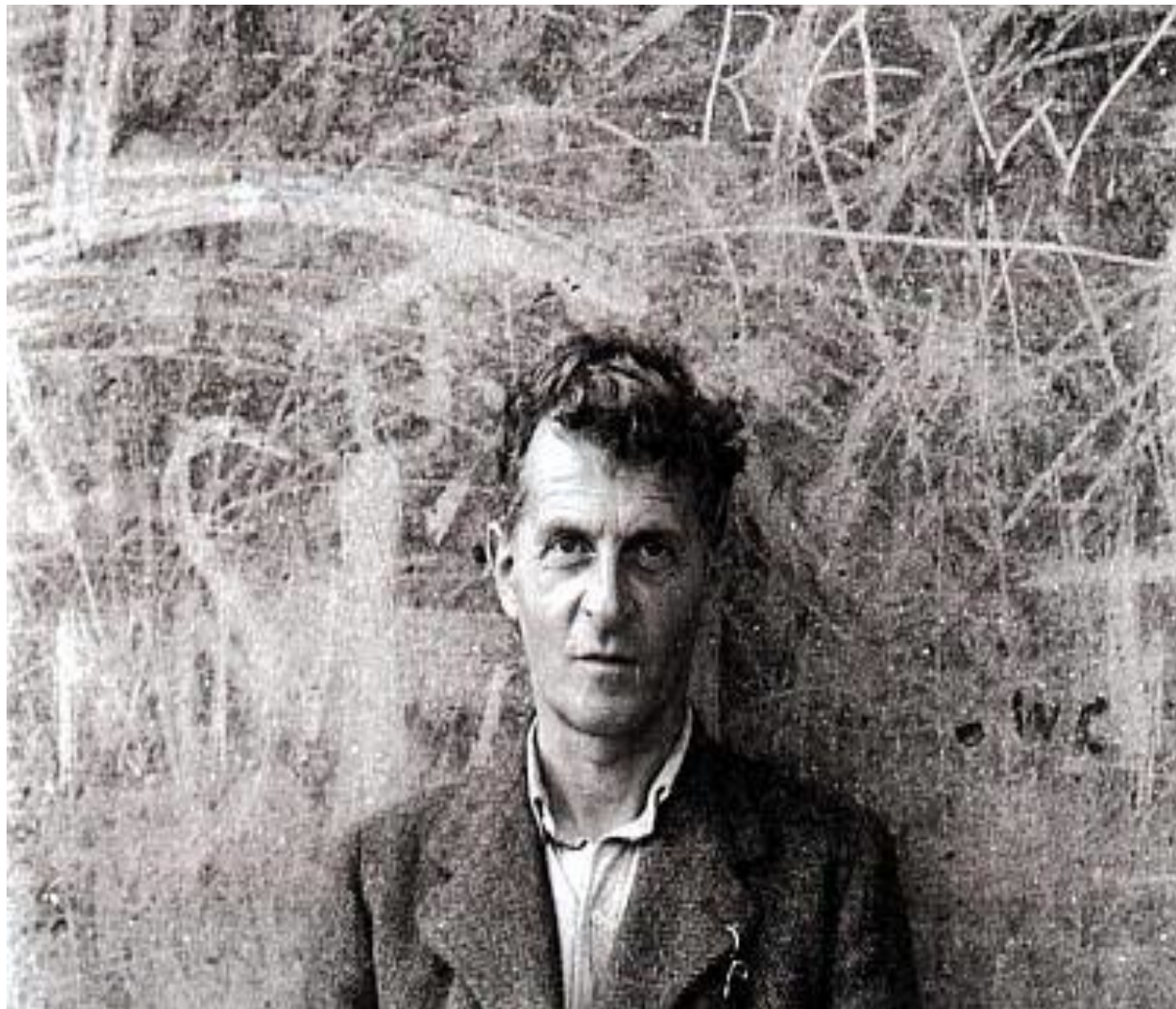
NOTE:

† and ‡ denote statistical significance against *Local_{base}* and *Local* baselines.



Towards understanding processes

- We built a system that can recover process descriptions in text
- We see this as an important step towards applications that require deeper reasoning, such as building biological process models from text or answering non-factoid questions about biology
- Our current performance is still limited by the small amounts of data over which the model was built
- But we want to go on to show that these process descriptions can be used for answering questions through reasoning



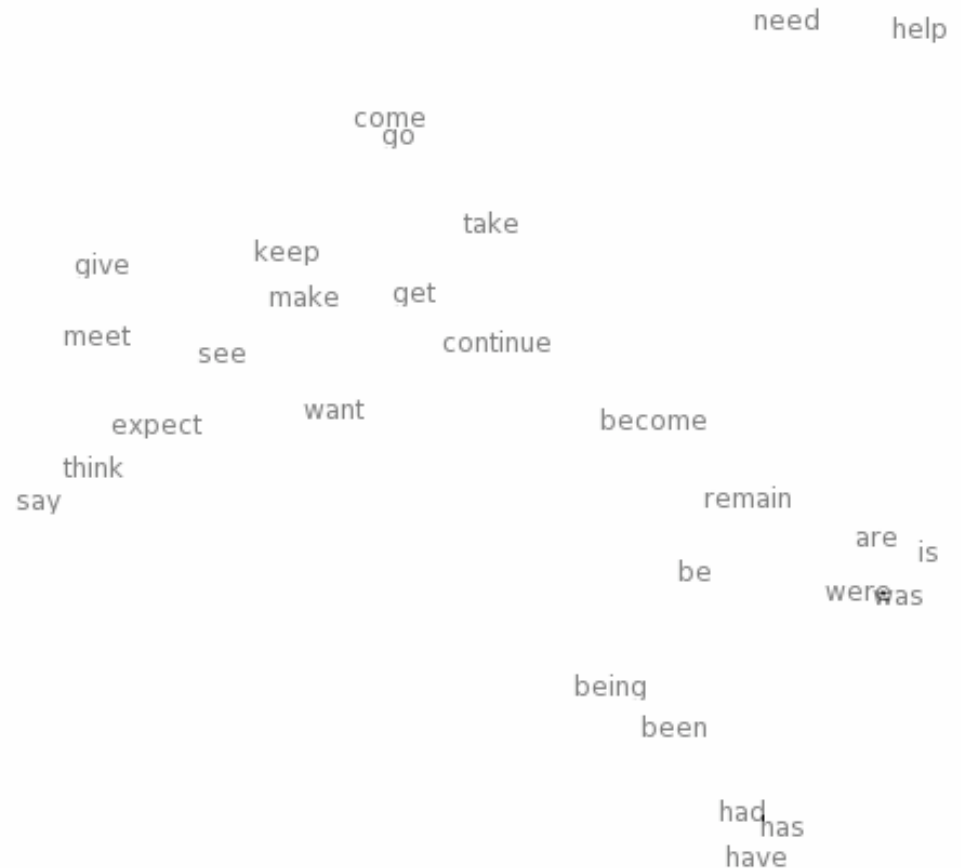


“When I talk about language (words, sentences, etc.) I must speak the language of every day. Is this language somehow too coarse and material for what we want to say? Then how is another one to be constructed?—And how strange that we should be able to do anything at all with the one we have!”



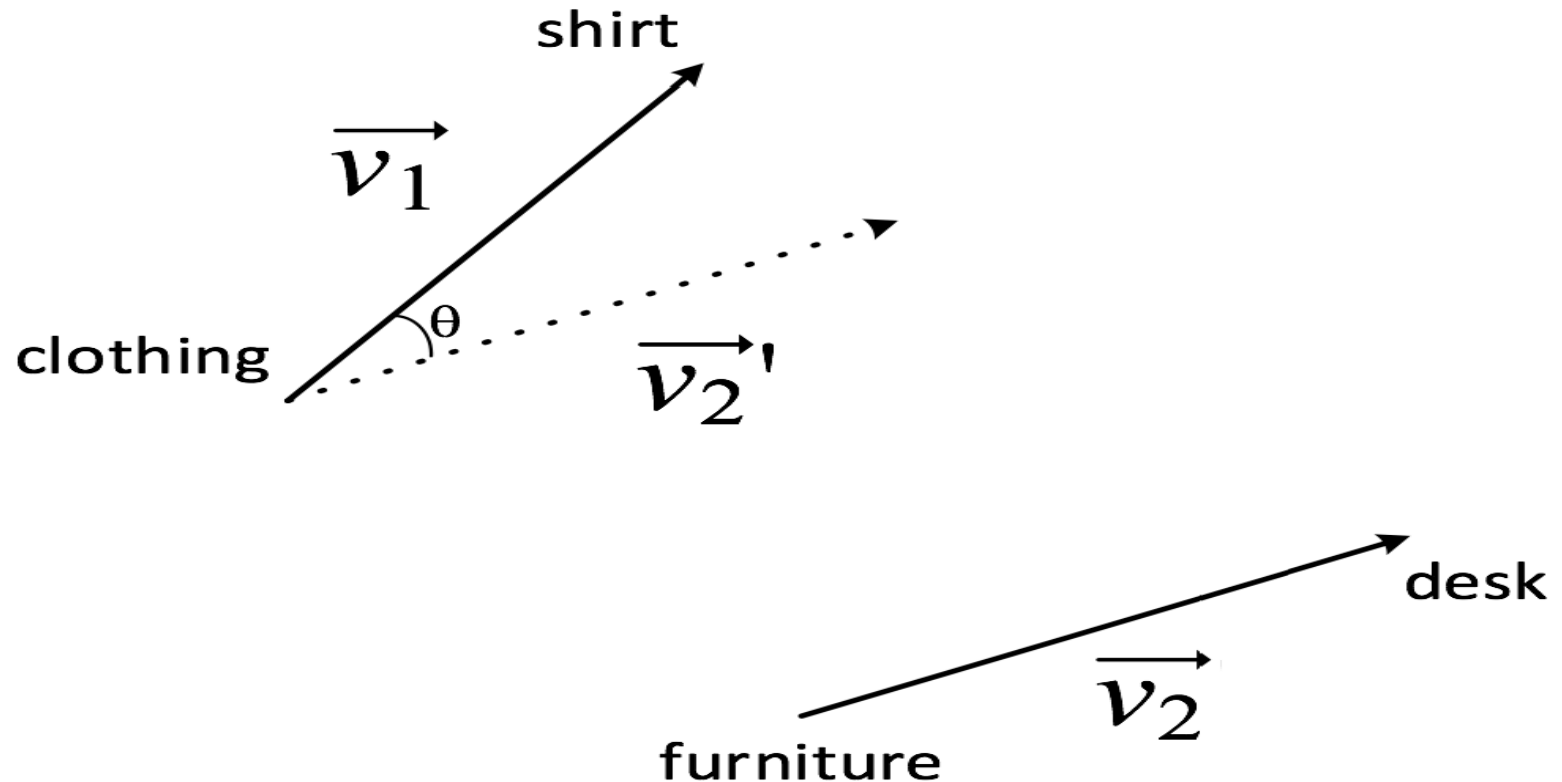
Word distributions can be used to build vector space models

linguistics =
$$\begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \\ 0.487 \end{pmatrix}$$





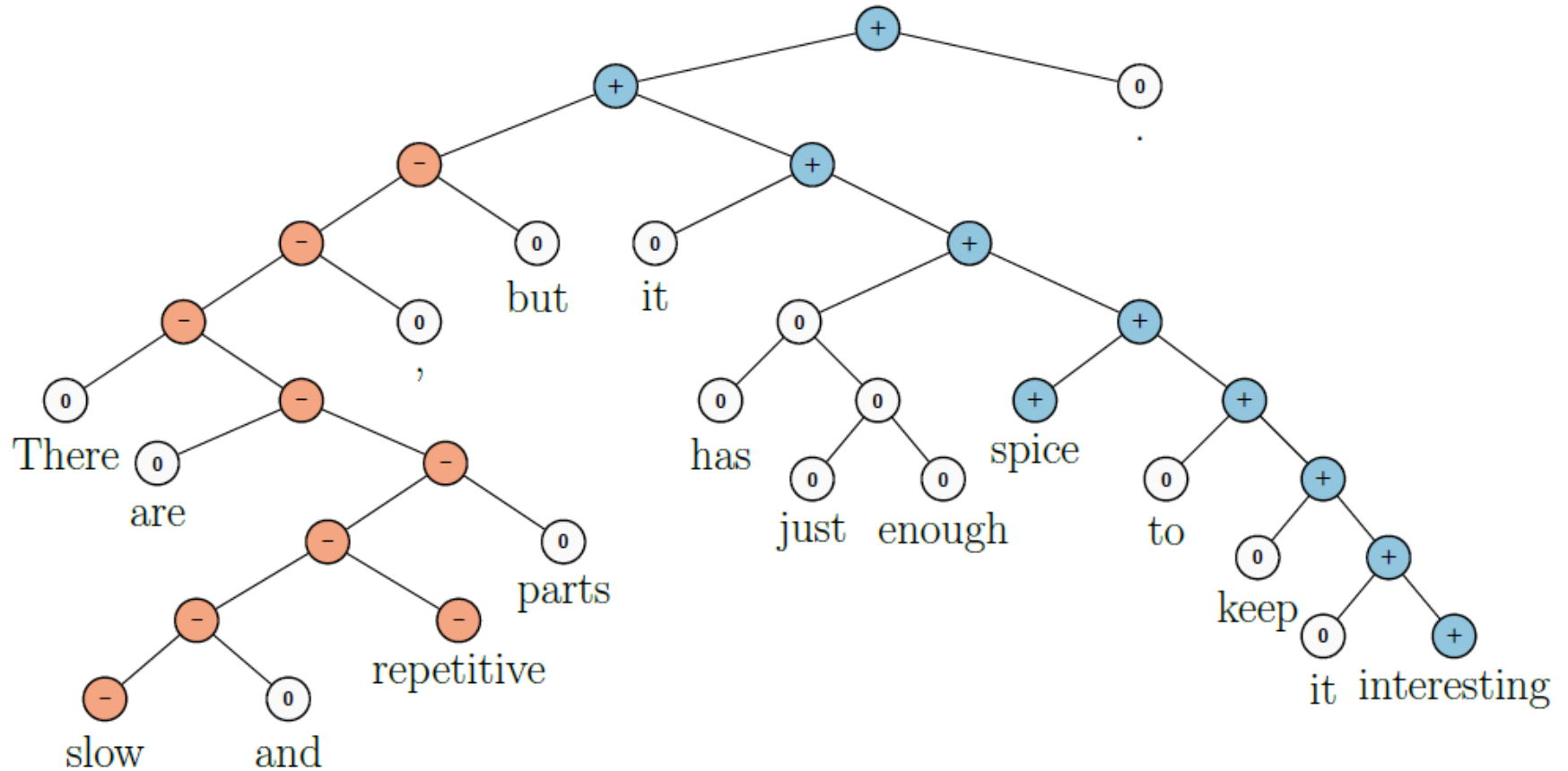
Neural/deep learning models: Mikolov, Yih & Zweig (NAACL 2013)





Socher et al. (2013)

Recursive Neural Tensor Networks



Texts are Knowledge

