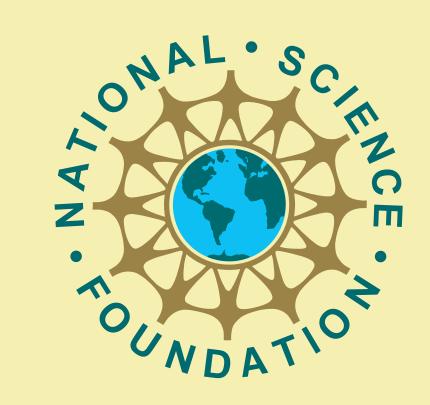


Automated Event Coding Using Machine Learning Techniques



Adam Stepinski (adamstep@rice.edu), Rice University, Houston, TX

What is event data?

- -information about political events
- -coded from lead sentences in newswire articles
- -multiple events in one sentence
- -consists of date and three parts:

Part	Description	Examples	
Actor	valid political entity initiates political action	France PLO Russian officials Iraqi spokesman	
Target	valid political entity target of action		
Event	action performed by actor classified into categories	Comment Force Reduce relationship Propose	

Problems with current methods

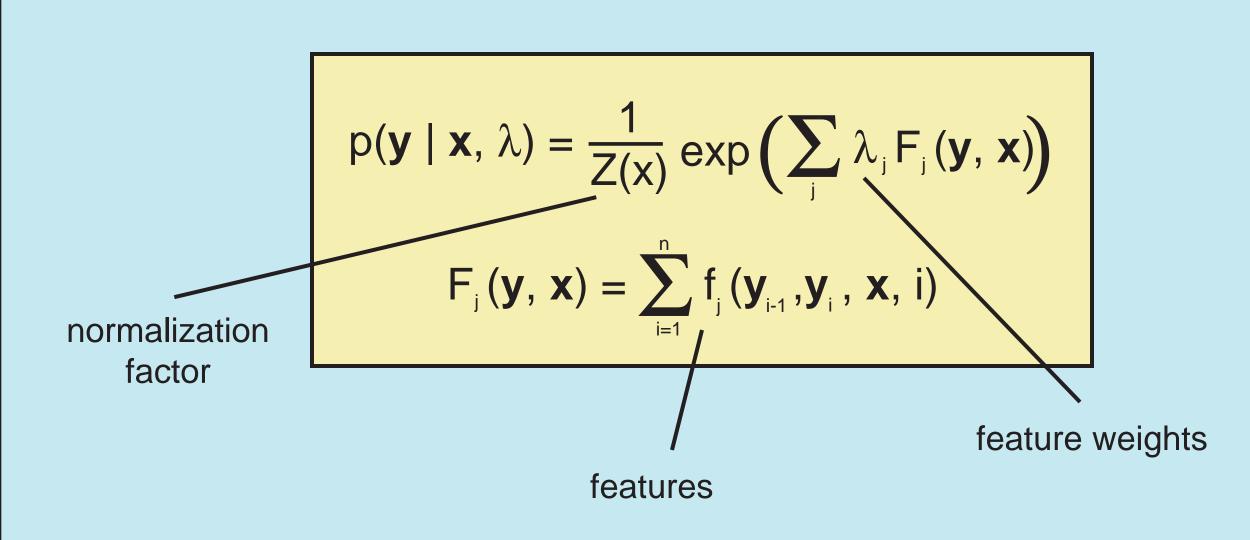
- -human coders (train humans to code events) -slow
 - -inconsistent across coders
 - -long training period
- -TABARI (automated coding using dictionaries)
 - -small number of events extracted
 - -phrase matching produces poor results
 - -poor extrapolation beyond dictionary

A smarter automated coder

- -analyze sentence before attempting event coding
 - -create grammatical parse tree of sentence -use grammar rules to turn sentence into
 - fragments -use parse tree structure to resolve
 - pronouns
- -apply conditional random fields to code events
 - -identify actors and targets
 - -identify event category
 - -use grammatical information in features

Conditional random fields

- -probabilistic framework for labeling sequential observations.
- -probability of label sequence y given observation sequence x:



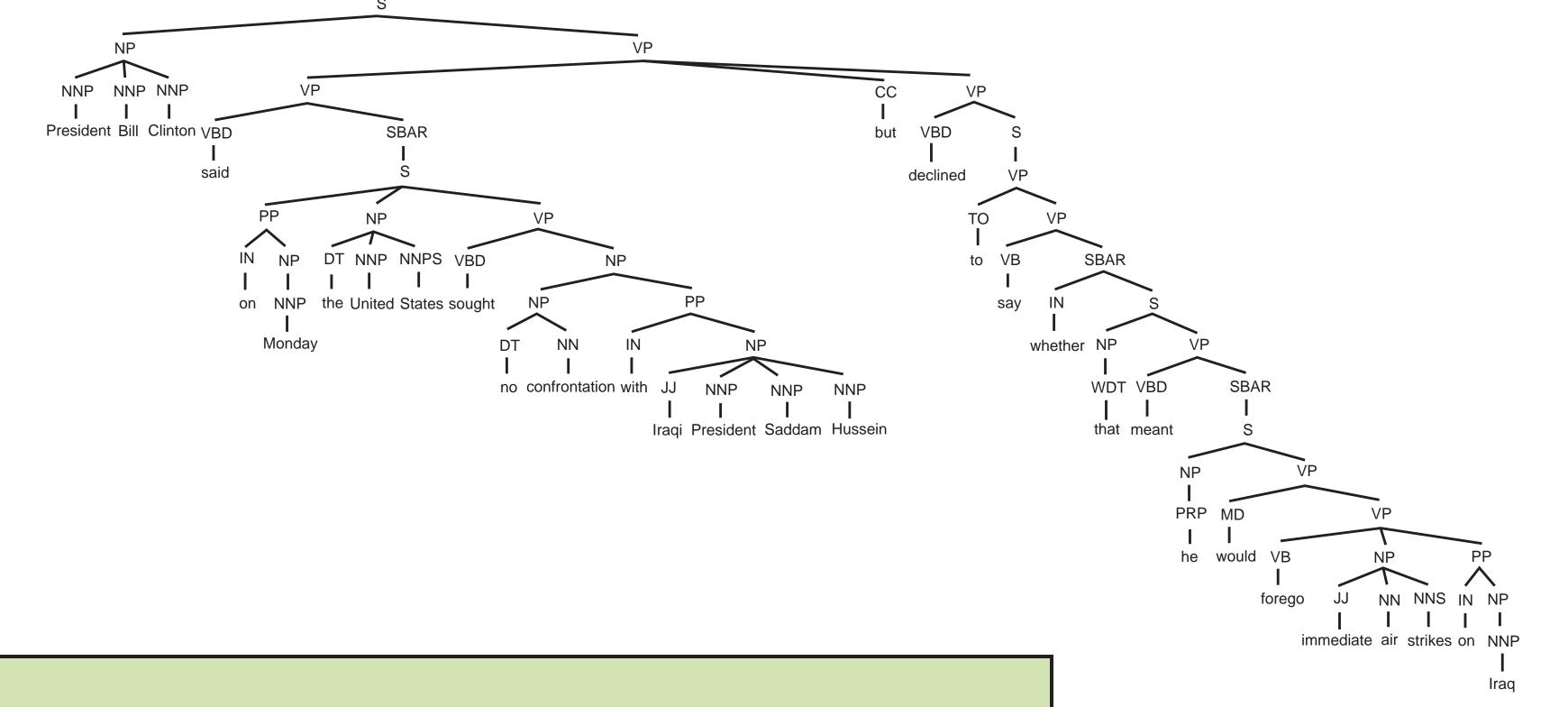
-features return 1 if present, 0 if not -adjusting feature weights trains algorithm -dynamic programming calculates sequence

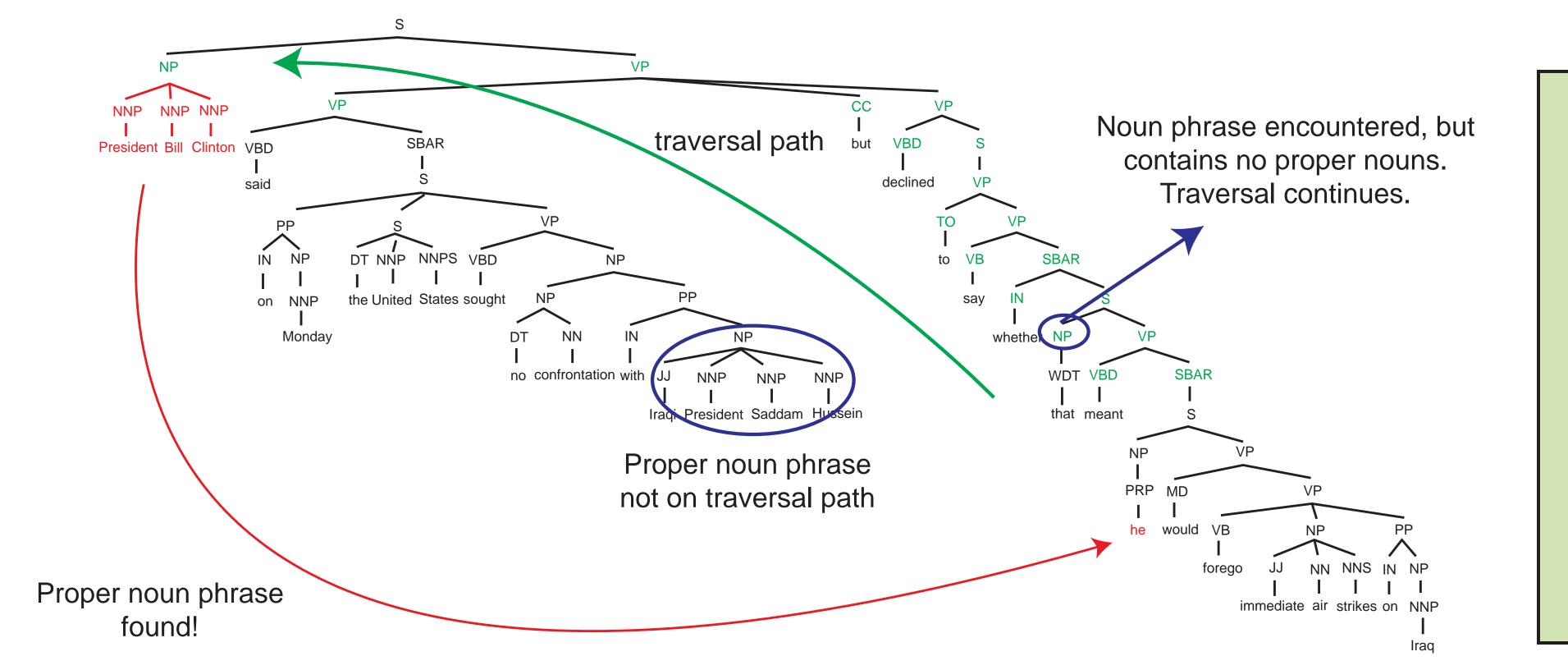
"President Bill Clinton said on Monday the United States sought no confrontation with Iraqi President Saddam Hussein but declined to say whether that meant he would forego immediate air strikes on Iraq."

1. Original sentence

2. Sentence parsing

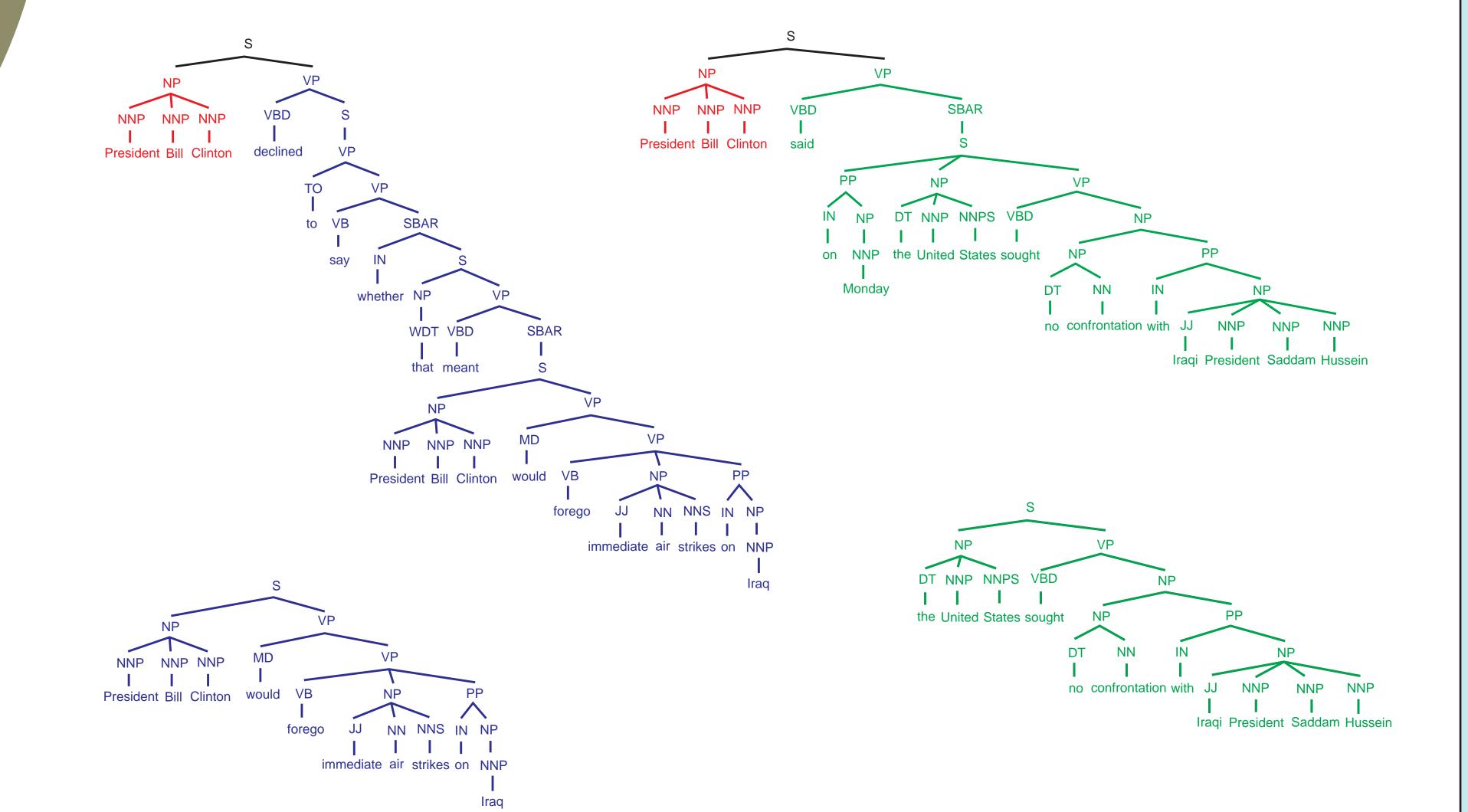
-parse tree created using Stanford Factored Parser





3. Pronoun resolution

- -pronouns detected
- -parse tree traversed starting at pronoun
 - -traversal moves up and to the left
- -first noun phrase on traversal path replaces pronoun
 - -noun phrase must contain proper nouns or traversal continues -noun phrase must agree in number with pronoun



4. Sentence fragmentation

- -New sentences created from compound structures -verb phrases linked by correlative conjunctions -main noun phrase used in both sentences
- -Fragments created from independent and subordinate clauses -identified by co-coordinating and subordinate conjunctions

Event data Fragment

President Bill Clinton said on Monday the United States sought no confrontation with Iraqi President Saddam Hussein	USA	Comment	USA
the United States sought no confrontation with Iraqi President Saddam Hussein	USA	Deny	IRQ
President Bill Clinton declined to say whether that meant President Bill Clinton would forego immediate air strikes on Iraq	USA	Comment	USA
President Bill Clinton would forego immediate air strikes on Iraq	not part of event phrase (did not actually happen)		

5. Sentence fragment labeling

-Condtional random fields used to label all fragments

Actor & target labels

-countries, capitals, nationalities -words in proper noun phrases

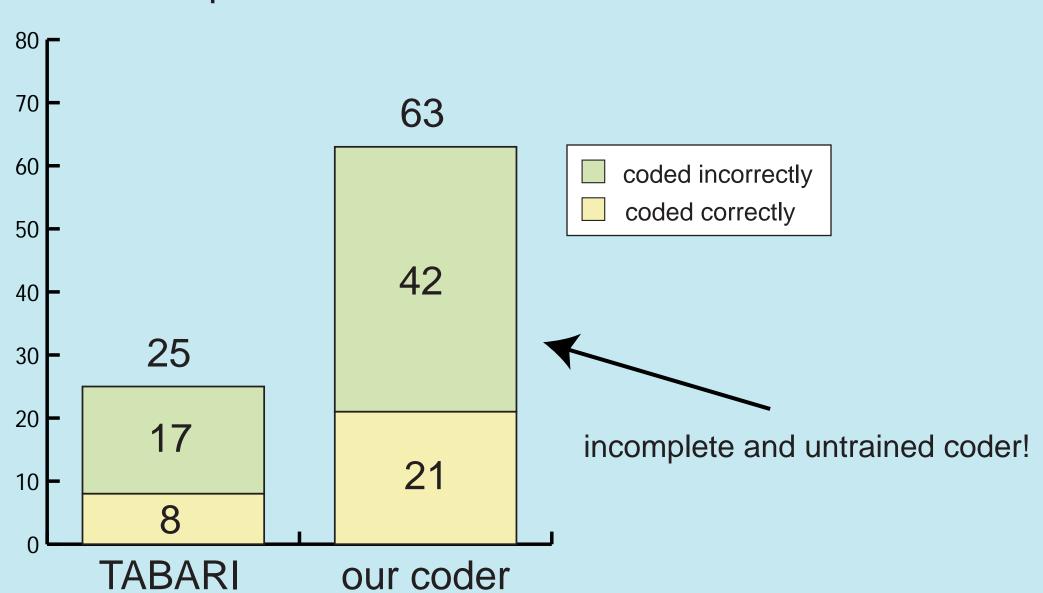
- -actors occur before main verb -actors at higher levels of tree
- -targets occur after main verb -targets at lower levels of tree

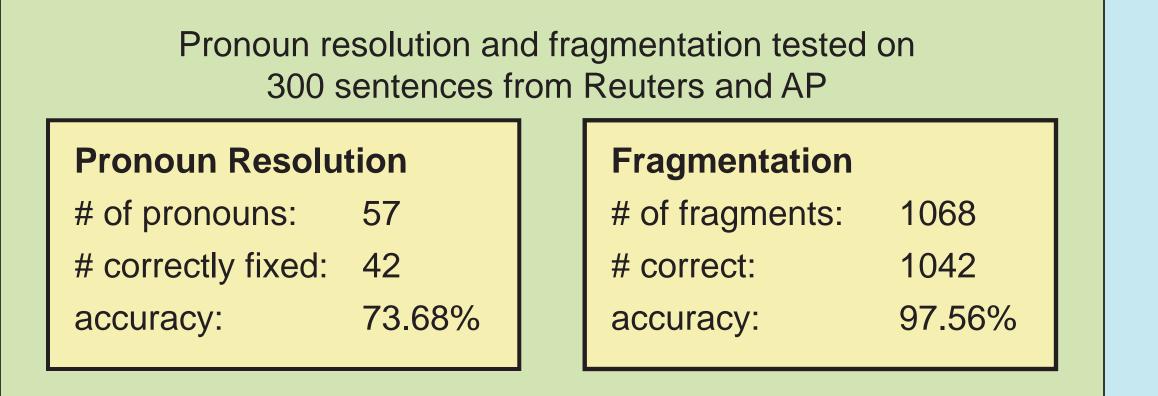
Event category labels

- -specific event keywords -words in main verb phrase
- -specific parts of speech
- -not modified by negative words -part of event phrase

Preliminary results

TABARI comparison tested on 189 Reuters sentences





Results show promise

-pronoun resolution

-good for "he" and "she" pronouns -not so good for "it" pronoun

-sentence fragmentation

-very low rate of malformed fragments

-performance improvement over TABARI

- -more events coded
- -better actor and target detection -better event detection

Future work

-improve pronoun resolution

-Resolution of Anaphora Procedure

-improve actor and target features -disregard unofficial entities

-add new entities automatically

-create smarter event features -utilize grammatical rules

-add new keywords automatically -train conditional random fields

-gather sentence/event data pairs -determine useful features

References

Wallach, Hannah M. "Conditional Random Fields: An Introduction." Feb 24,

Schrodt, Philip A. "Automated Coding of International Event Data Using Spase Parsing Techniques." International Studies Association,

Klein, Dan and Manning, Christopher D. "Fast Exact Inference with a Factored Model for Natural Language Parsing." Advances in Neural Information Processing Systems, 15 Dec 2002.

Gucker, Philip. "Essential English Grammar." Dover Publications, 1966.