PropS
Generic Proposition Extraction

Gabriel Satanovsky   Jessica Ficler   Ido Dagan   Yoav Goldberg

Online Demo available at:
http://u.cs.biu.ac.il/~stanovg/propextraction.html

PropS motivation

- Semantic application are primarily interested in the predicate-argument structure conveyed in texts

- Commonly extracted from dependency trees
  - Yet it is often a non-trivial and cumbersome process, due to syntactic over-specification, and the lack of abstraction & canonicalization

- Our goal:
  - Accurately get as much semantics as given by syntax
  - Stems from a technical standpoint
  - Yet raises some theoretic issues regarding the syntax – semantics interface
  - Over generalizing might result in losing important semantic nuances

PropS

- A simple, abstract and canonicalized sentence representation scheme

    - **Nodes** represent atomic elements of the proposition

    - Predicates, arguments or modifiers

   - **Edges** encode argument (solid) or modifier (dashed) relations

The janitor who likes plumbing didn't want to fix the pipe
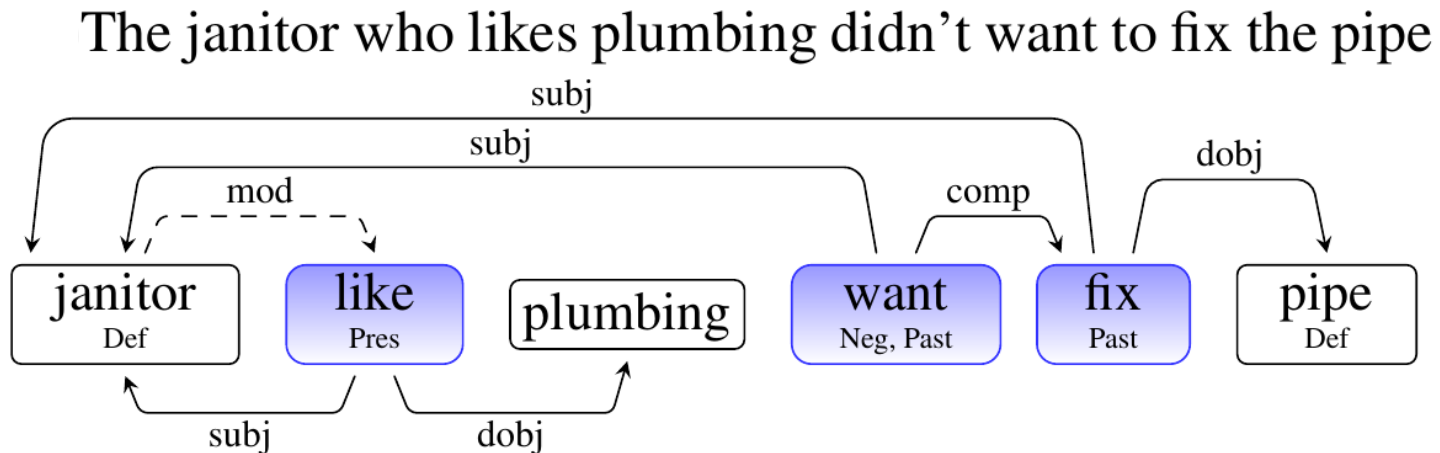
PropS Properties

- **Abstracts** away syntactic variations

  - Tense, passive vs. active voice, negation variants, etc.

- **Unifies** semantically similar constructions

  - Various types of predications

  - Different syntactic realizations of the same proposition

- **Differentiates** over semantically different propositions

  - E.g. restrictive vs. non-restrictive modification, different types of appositions
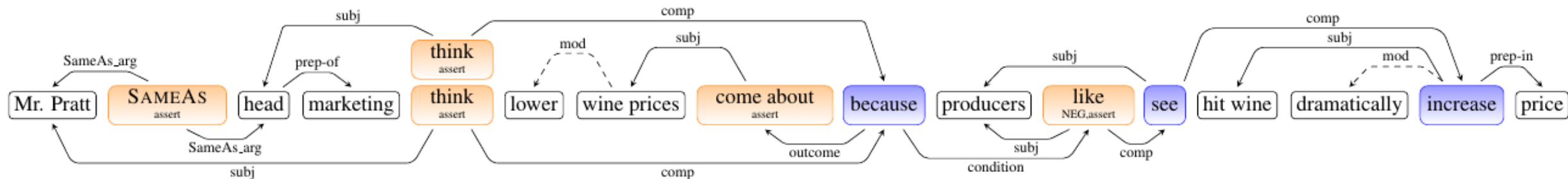
"Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don't like to see a hit wine dramatically increase in price."

Props (17 nodes and 19 edges)
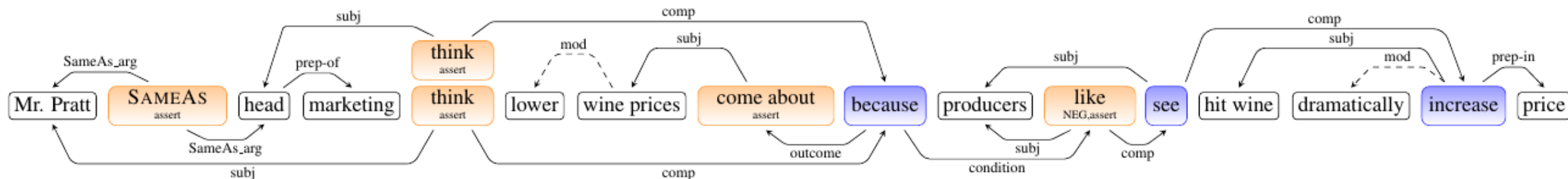
Dependency parsing (27 nodes and edges)
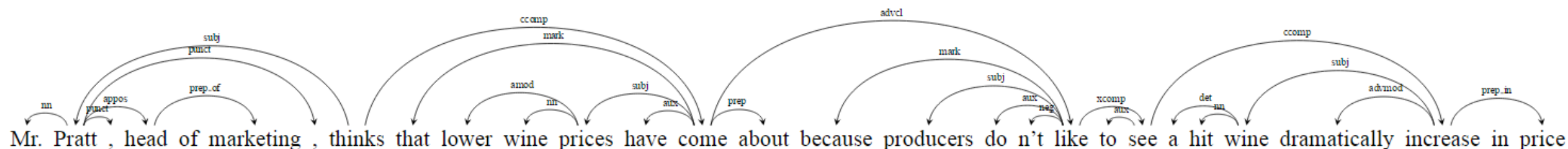
"Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don't like to see a hit wine dramatically increase in price."

- Extracted propositions:

  (1) lower wine prices have come about   [asserted]

  (2) hit wine dramatically increase in price

  (3) producers see (2)

  (4) producers don't like (3) [asserted]

  (5) Mr Pratt is the head of marketing [asserted]

  (6) (1) happens because of (4)

  (7) Mr Pratt thinks that (6) [asserted]

  (8) the head of marketing thinks that (6) [asserted]

PropS Methodology

- Corpus based analysis

  - Taking semantic applications perspective

  - Focusing on the most commonly occurring phenomena

  - Covering most of them

- Feasibility criterion

  - High accuracy "Gold Standard" would be feasibly derivable from available manual annotations

  - Reasonable accuracy for baseline parser from dependency parsing

PropS Handled Phenomena

- Certain syntactic details are abstracted into node features

  - Modality

  - Negation

  - Definiteness

  - Tense

  - Passive or active voice

- Restrictive vs. non restrictive modification

  - Implies different argument boundaries:

  - *[The boy who was born in Hawaii]* **went** *home*   [restrictive]

  - *[Barack Obama] who was born in Hawaii* **went** *home* [non-restrictive]

PropS Handled Phenomena (cont.)

- Distinguishing between **asserted** and **attributed** propositions
    - *John **passed** the test*
    - *the teacher **said** that John **passed** the test*

- Distinguishing the different types of appositives and copulas
    - *The company, Random House, didn't report its earnings* [**appositive**]
    - *Bill Clinton, a former U.S president , will join the board*   [**predicative**]

PropS Handled Phenomena (cont.)

- … and more:

  - Conditionals

  - Raising vs. control constructions

  - Non-lexical predications (expletives, possessives, etc.)

  - Temporal expressions

  - Adjectival modification

PropS Resources

- Human annotated gold-standard

  - 100 sentences from the PTB annotated with our gold structures

- High-accuracy conversion of the WSJ

  - Computed (rule-based) on top of integration of several manual annotations

  - PTB Constituency

  - Propbank

  - Vadas et al(2007)'s NP structure

- Baseline parser

  - Rule based converter over automatically generated dependency parse trees

PropS Conversion Accuracy

	Feature Computation			Modified LAS		
	P	R	F1	P	R	F1
WSJ	.95	.97	.96	.9	.92	.91
PROPS	.88(.88)	.89(.84)	.89(.86)	.83(.8)	.81(.81)	.82(.8)

Table 2: Conversion accuracy, WSJ is compared against gold standard, PROPS against the gold standard and WSJ (in parentheses).

http://u.cs.biu.ac.il/~stanovg/propextraction.html