

A Syntactic Rule-base for Textual Entailment

Amnon Lotan, Tel Aviv University
Ido Dagan, Bar Ilan University

Asher Stern, Bar Ilan University
Jonathan Berant, Stanford University

ISCOL Sep 7 2014

Textual Entailment

Given a *text* **T** and a *hypothesis* **H**, Recognize whether **T** entails **H**

Example

Text: The boy was located by the police.
Hypothesis: The police found the child.

Sequence of Transformations (A Proof)

Text: The boy was located by the police
 passive to active
 The police located the boy
 X locate Y → X find Y
 The police found the boy
 boy → child
 The police found the child → **Hypothesis**

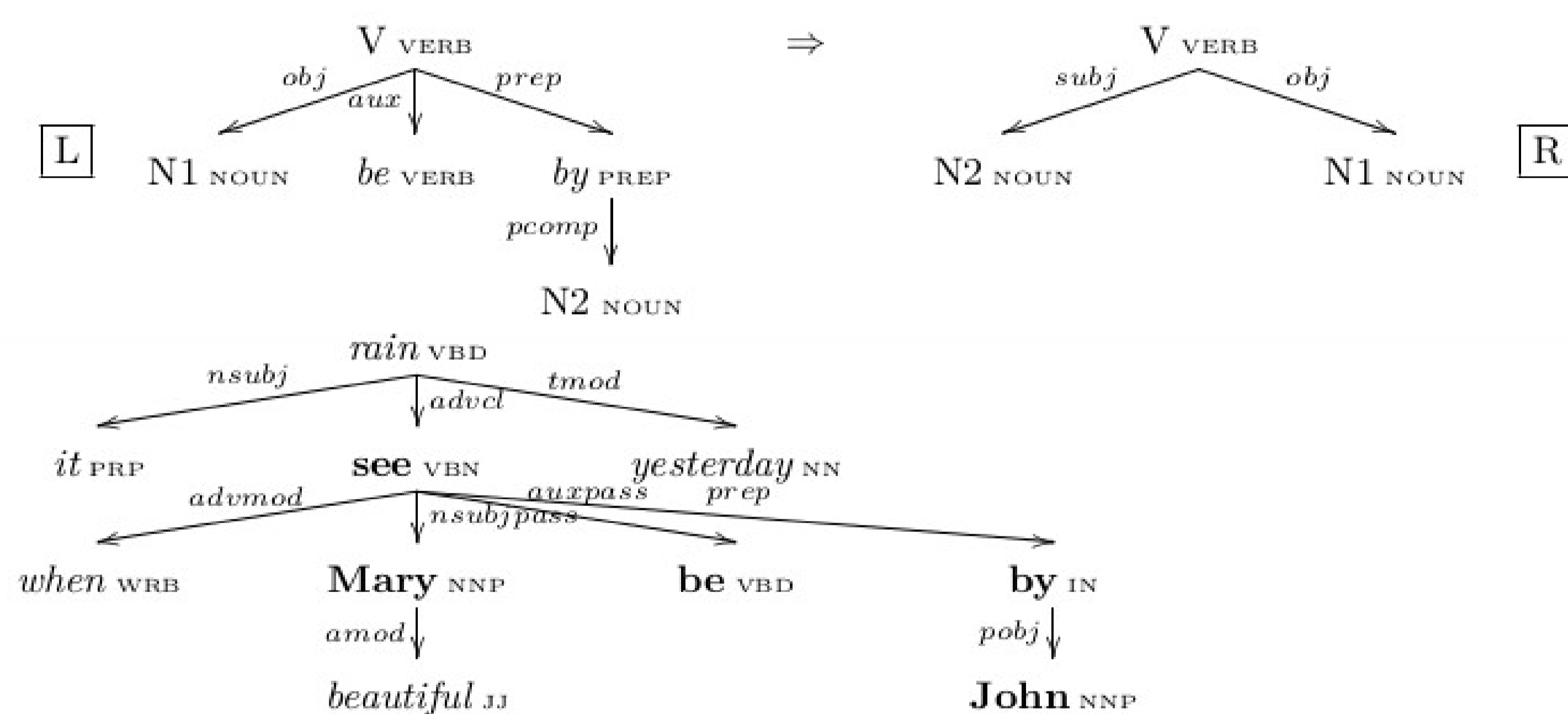
Previous Work

- Formalism, first draft of rules (Bar Haim et al. 2007)
- Adjectival and adverbial structures (Amoia and Gardent 2008)

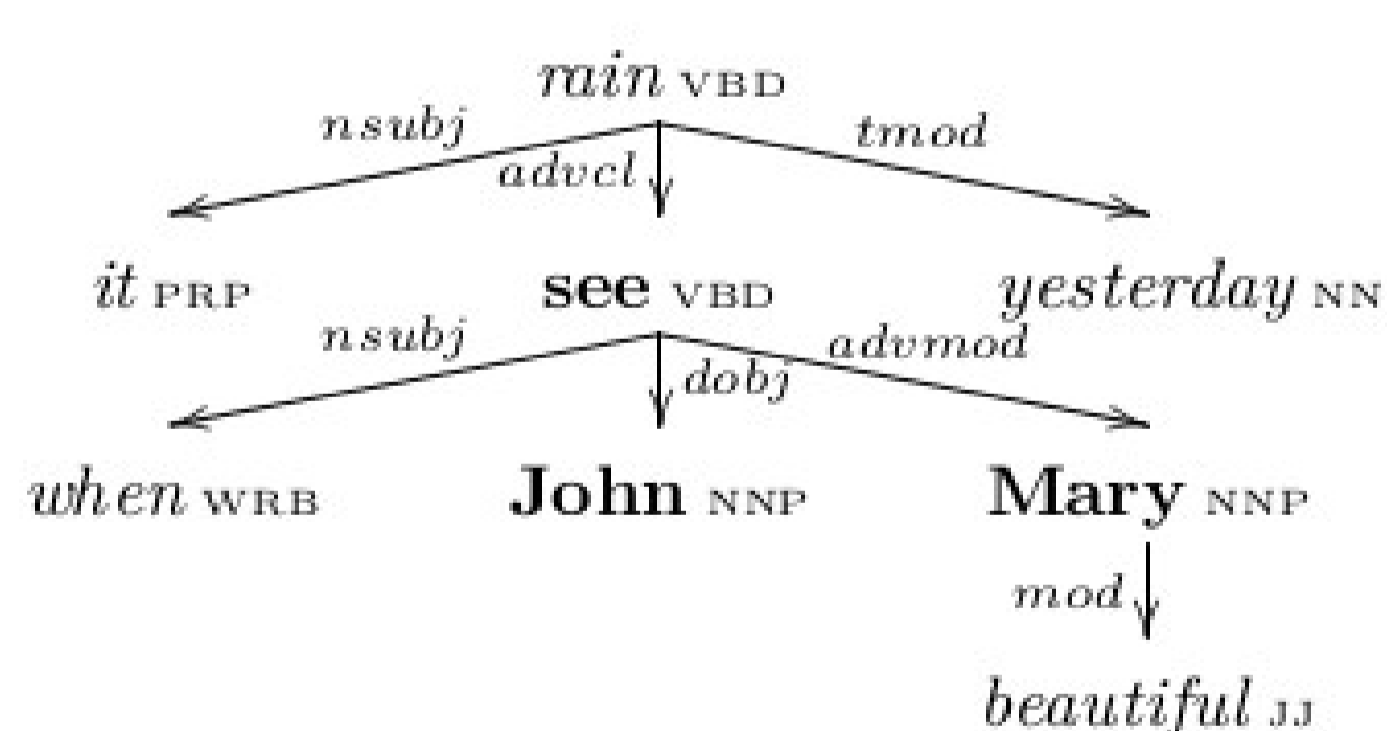
This is a red table ↔ *This table is red*
- Hearst Patterns (Hearst 1992, Pantel et al. 2004)

She likes apples and other fruit → *Apples are fruit*

Passive to Active

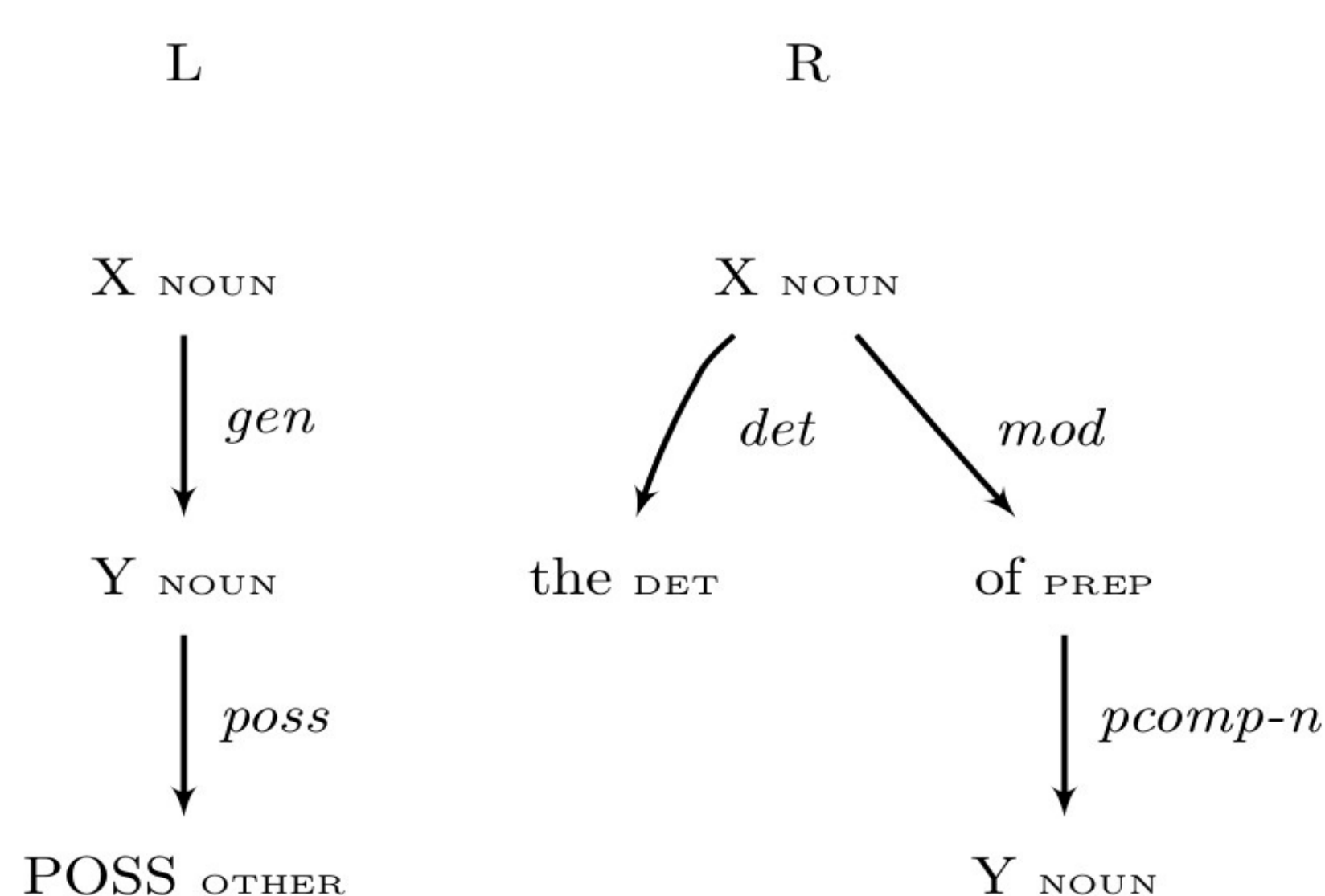


Source: it rained when beautiful Mary was seen by John yesterday



Derived: it rained when John saw beautiful Mary yesterday

“an X’s Y” ↔ “the Y of X”



High Potential in RTE Datasets

- Manually solved 50 TH pairs in RTE5 + 50 in RTE6
- Used 1.16 syntactic operations per proof
- 78% of pairs need syntactic inference

Features

- Generic, Syntax-based
- Hand coded, syntactically motivated rules
- Over **twice as large** as previous rule-bases
- Open source**, supports Java, XML, CoNLL
- Handy **GUI** editor and rule parser

Usage in Gold Standard and BiuTee*

Syntactic Category	Manual Analysis	BiuTee Usage
Apposition	26.3%	6.7%
Relative Clause	19%	3.7%
Determiners	19%	26.3%
Possessives	14%	16.9%
Active/Passive	12.3%	35.9%
Coordination	5.3%	10.2%
Adjectival	3.5%	0%
Case Correction	0%	0.2%
IS-A implications	0%	0.1%

*The Bar Ilan Textual Entailment Engine is a transformation based system that plugs in diverse knowledge resources, like this one (Stern and Dagan 2011)

High Quality :)

Precision 94% - based on system rule applications

Recall 70% - based on manually constructed gold standard

Ablation Tests

Configuration	Recall	Precision	F1	Accuracy
Baseline RTE6	43.07%	55.37%	48.45%	N/A
Rule-base Test RTE6	43.17%	57.71%	49.4%	N/A
Baseline RTE5	70%	61.04%	65.21%	0.6333
Rule-base Test RTE5	73.66%	63.32%	68.1%	0.6483

Error Analysis – based on BiuTee System Usage

Applied / Why not applied	% of manually identified syntax-based rule applications
Applied by the system	26.3%
Unavailable rule, within the formalism	30%
System found alternative proof	15.8%
Parser error	10.5%
System lacking prerequisite rules	7%
Resource lacking prerequisite rules	5.3%
Unavailable rule, out of the formalism	3.5%