

Textual Entailment Engine for Semantic Information Access

Ido Dagan
Bar-Ilan University, Israel

Semantic Information Access

Question answering

- *Which companies insure pets?*

■ Information extraction

- Telecom investments: *company, bank, amount, date*

■ Multi-document summarization

- *Pakistani terrorists attacked Mumbai*
Mumbai terror act traces to Pakistani militants

■ Even keyword search

- *Water pollution [in China]*

Need a Semantic Engine

- It's desired to encompass semantic processing under a unified engine
 - Cf. morphology, syntax
- But it's not clear what semantics processing is!
- Engine desiderata
 - 1) *Generic module for applications*
 - Common underlying task, unified interface (API)
 - 2) *Encompasses all phenomena (& knowledge)*
 - 3) *Unified knowledge representations*

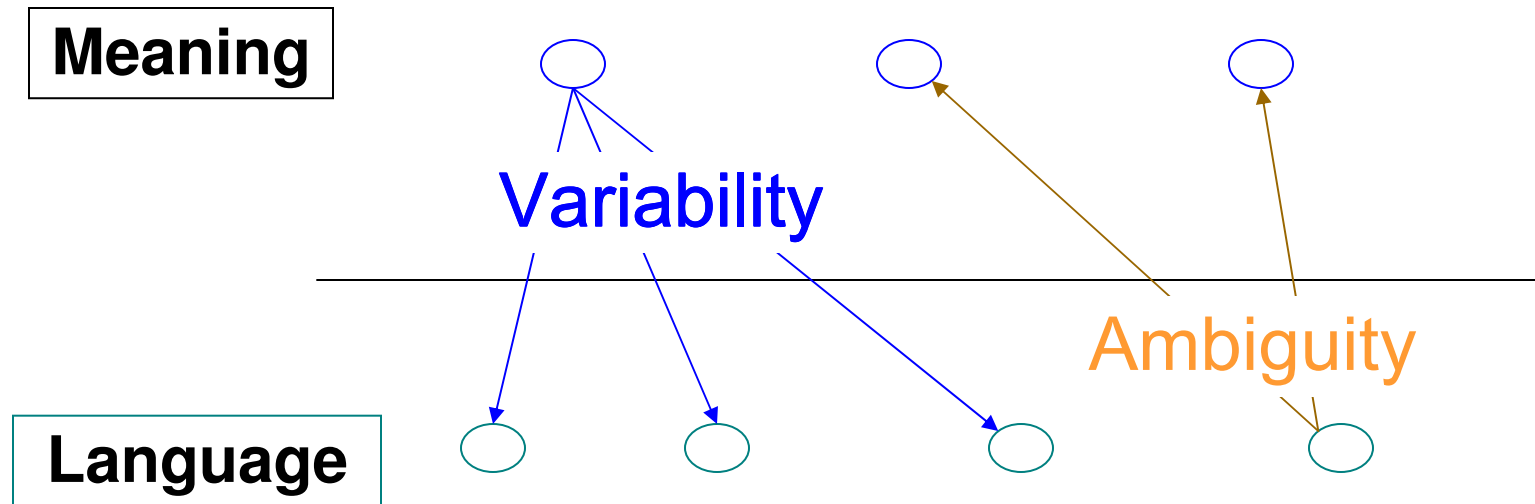
Talk Outline

- 1) *Textual Entailment* - and why it may be a good framework for semantic inference
- 2) How we approach it at Bar-Ilan
 - See poster
- 3) A proposal for a concrete inference engine – API
 - How it can be used by applications

Textual Entailment: Task and Framework

Natural Language and Meaning

Why is it difficult?



Variability of Semantic Expression

The Dow Jones Industrial Average closed up 255

Dow ends up

Dow climbs 255



Dow gains 255 points

Stock market hits a
record high

Model variability as relations between texts:

- Equivalence: $text1 \Leftrightarrow text2$ (paraphrasing)
- **Entailment:** $text1 \Rightarrow text2$ – the general case

Typical Application Inference

Question

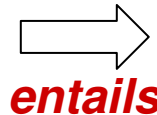
Who acquired Overture?

Expected answer template

>> X acquired Overture

Yahoo's buyout
of Overture

text



Yahoo acquired Overture

hypothesized answer

Application inferences can be reduced to entailment

- IE: X acquire Y
- Summarization (multi-document): identify redundant sentences
- MT: paraphrasing, evaluation
- Educational applications: student answer vs. reference

Applied Textual Entailment

- Directional relation between two text fragments: *Text (t)* and *Hypothesis (h)*:

t **entails** *h* ($t \Rightarrow h$) if **humans** reading *t* will infer that *h* is **most likely** true

- Operational (applied) definition:
 - Human gold standard
 - Entailment judgment matches applications judgments
 - Assuming common background knowledge
 - Language & world knowledge

Evaluation: PASCAL RTE Challenges

	TEXT	HYPOTHESIS	TASK	ENTAILMENT
1	<i>Regan attended a ceremony in Washington to commemorate the landings in Normandy.</i>	<i>Washington is located in Normandy.</i>	IE	False
2	<i>Google files for its long awaited IPO.</i>	<i>Google goes public.</i>	IR	True
3	<i>...: a shootout at the Guadalajara airport in May, 1993, that killed Cardinal Juan Jesus Posadas Ocampo and six others.</i>	<i>Cardinal Juan Jesus Posadas Ocampo died in 1993.</i>	QA	True
4	<i>The SPD got just 21.5% of the vote in the European Parliament elections, while the conservative opposition parties polled 44.5%.</i>	<i>The SPD is defeated by the opposition parties.</i>	IE	True

RTE Participation and Impact

- Very successful challenges, world wide:
 - Over 40 groups so far
 - Hundreds of downloads
- RTE-4 (2008) – under NIST
 - New TAC conference (with QA and summarization)
- High interest in research community
 - Papers, sessions and areas, PhD's, funding proposals
 - Special issue at of NLE Journal, ACL-07 tutorial, AVE@CLEF, ...

Initial use of RTE systems in applications

- QA

- Harabagiu & Hickl, ACL-06
- Answer Validation Exercise at CLEF

- Relation extraction

- Romano et al., EACL-06

- Educational applications

- Nielsen et al., ACL-08 workshop

- Ongoing projects

Bar-Ilan's Approach for Textual Entailment Modeling

Teaser: see Roy's Bar-Haim poster this afternoon

Our Inference Formalism

Bar-Haim et al., AAAI-07 & RTE3

- A **proof system** over parse trees:
 - Represents diverse kinds of semantic knowledge uniformly as entailment (inference) rules
 - Allows unified inference mechanism
- Analogy to logic proof systems:

Propositions	Parse Trees
Inference Rules	Tree transformations
Proof	A sequence of trees, generated by rule application

- Given *Text* (T) and *Hypothesis* (H), try to *generate* H from T
- Formalizes common transformation-based approaches

A Sample Proof

Text: McCain congratulated the elected president, Barack Obama.

Consequent

⇒ Barack Obama is the elected president.

⇒ Barack Obama won the elections

⇒ The Democratic nominee won the elections.

Entailment Rule

Apposition
(syntactic)

X is the elected president
→ *X won the elections*
(lexical-syntactic)

Barack Obama
→ *the Democratic nominee*
(lexical)

Hypothesis: The Democratic nominee won the elections.

Diverse Knowledge Types as Entailment Rules

- Syntactic transformations:
 - *Passive, appositives, relative clause, conjunctions, determiners*
- Lexical-syntactic predicate rules:
 - Learned with unsupervised algorithms (DIRT, TEASE)
 - Derived automatically by integrating WordNet and Nomlex
 - $X \text{ bought } Y \rightarrow Y \text{ was sold to } X$
 - $X \text{ is a maker of } Y \rightarrow X \text{ produces } Y$
- Lexical rules:
 - From WordNet and Wikipedia
 - $steal \rightarrow take, Janis Joplin \rightarrow singer, Amazon \rightarrow South America$
- Polarity rules:
 - Manually-composed, utilizing VerbNet and PARC's polarity lexicon
 - *Verbal negation, modal verbs, conditionals, verb polarity*

Extracting Lexical Rules from Wikipedia

E.T. the Extra-Terrestrial

From Wikipedia, the free encyclopedia

(Redirected from E.T. (film))

E.T. the Extra-Terrestrial is a 1982 science fiction film co-produced and directed by Steven Spielberg, written by Melissa Mathison and starring Henry Thomas, Robert MacNaughton, Drew Barrymore, Dee Wallace and Peter Coyote. It tells the story of Elliott (played by Thomas), a lonely boy who befriends a friendly alien dubbed "E.T.", who is stranded on Earth. Elliott and his Extraterrestrial life

•*Be-complement*

Nominal complements of 'be'

•*Redirect*

various terms to canonical title

•*Parenthesis*

used for disambiguation

•*Link*

Maps to a title of another article

Towards a Concrete Semantic Engine

Entailment-based Inference Engine

- Base a generic ***inference*** engine on the ***entailment*** task
 - Reduce application inferences to entailment API
- *Different* than traditional approach of semantic ***analysis/annotation***
 - E.g. semantic “parser”
 - Assuming that applications would make their own inferences, facilitated by these annotations
- Provide the eventual inference by the generic engine
 - *Avoid forcing semantic annotations on the application (vs. morphology & parsing!)*
- Need to specify API, and usage by applications (reductions)

Basic Function - Recognize Entailments

- RTE challenge mode
- Given (t, h) pair:
 - Is h 's truth *entailed* by t ? (*contradicted/unknown*)
- Validation mode for applications
- Target meaning considered as h :
 - QA – candidate answer entailed by passage
 - IE – candidate extraction entailed by passage
 - *Summarization* – avoid entailments within summary
 - *MT* paraphrasing – sentence/paraphrase equivalence
 - *Tutoring* – student answer entails reference

Typed Variables in Hypothesis

- Address “slot filling” (QA, IE)
- Allow **variables** in h
 - h : $X \text{ elect } Y$
 - t : a candidate text for an election event
- Entailment engines can be naturally extended to *instantiate* variables
- ***Textual typing*** for variables
 - Types are *unrestricted textual expressions*, for names *and* terms
 - vs. pre-defined NER types
 - Election example: $\langle X: \text{executive board}, Y: \text{company official} \rangle$
 - QA: “which” questions (*river, president, treatment, ...*)

Meaning Disambiguation via Contextual Preferences

- Specify/disambiguate the meaning of the target h by *textual* context information
- Context types:
 - Textual types for variables (as above)
 - Disambiguating context terms
 - IE: $X \text{ attack } Y$ - specify $\{\text{military, war, terrorism}\}$ for template
 - Session/user context in QA&IR, doc context in summarization
- Avoid symbolic annotations to specify meaning (e.g. senses) – stick to textual representations
 - *Contextual Preferences* – Szpektor et al., ACL-08

Operation modes

- **Recognition:** recognize entailment given t/h pair
 - *Validation in applications*
- **Search:** given h and corpus/doc, find all entailing texts
 - *QA, IR, IE against corpus/doc*
- **Generation:** given text, generate all entailed statements
 - *Paraphrase generation for MT*
- **Partial entailment:** identify entailments of parts of h
 - *Summarization, partial match*

API Summary – Supply application needs

1. Recognize entailment for a (t,h) pair
2. Variables with textual types in hypothesis
3. Textual contextual preferences for disambiguation
4. Support all possible operation modes

Enabling New Capabilities, e.g. - Integrating IE and Search (w. HRL)

Structured Search

Results Summary

- [-] Hypertension causes
 - [-] Lifestyle
 - [-] Food
 - Sodium**
 - Coffee drinking
 - Smoking
 - Alcohol drinking
 - The North American diet
 - Fat
 - Imported licorice
 - [+] Activities
 - [+] Medicines with side effects
 - [+] Genetics
 - [+] Medical problems with side effects
 - [+] Others

Results for causes of hypertension: Sodium

Diet and Hypertension

People trying to control **hypertension** often are advised to decrease **sodium**, increase potassium, watch their calories, and maintain a reasonable weight. ...
www.ext.colostate.edu/pubs/foodnut/09318.html - 57k - [Cached](#) - [Similar pages](#) - [Note this](#)

Sodium Reduction and Hypertension

Obesity, for example, may be an even bigger contributor to **hypertension** than **sodium** intake. Scientists have long stressed the high correlation between ...
www.aces.edu/dept/extcomm/newspaper/feb13a01.html - 10k - [Cached](#) - [Similar pages](#) - [Note this](#)

Is Sodium Restriction Important to Hypertension?

Debate: the argument for and against implementing **sodium** restrictions to patients with **hypertension**.
www.medscape.com/viewarticle/480719 - [Similar pages](#) - [Note this](#)

Reduce Salt and Sodium in Your Diet

... Pressure, Hormone, Replacement, Therapy, Isolated systolic, **hypertension** (ISH) ... A key to healthy eating is choosing foods lower in salt and **sodium**. ...
www.nhlbi.nih.gov/hbp/prevent/sodium/sodium.htm - 17k - [Cached](#) - [Similar pages](#) - [Note this](#)

Research Challenges

■ **Knowledge acquisition**

- Learning (corpora and web), extraction from resources, manual

■ Inference

- Principled models (inspired by logical & probabilistic reasoning)
- Representation, efficient search, scoring, approximate match

■ Model specific semantic phenomena

- Derivations, compounds relations, typing, modals, polarity...

■ **Using** generic entailment technology for applications

Takeouts

- It's time to think of generic semantic engines
 - Would provide needed inference for current applications
 - Enable new applications
- My proposal:
 - Base engine on *entailment* rather than *annotation*
 - Much room for formalization
 - Initial API
 - Highlights challenges for semantic research
- How long did it take for useful parsers?

Thank You!