**CS6740/IS 6300, Lecture 23: Evaluation by/of textual inference (aka entailment)**

Goals:

- background and "appreciation" of the NLI task Chris Potts talked about last Thursday.
- Intro/review to various semantic phenomena
- A sub-problem: inferring downward-monotone operators

The natural-language inference (entailment) task may seem artificial. Why do so many NLP researchers spend time on it?

1. What are useful applications that can also serve to evaluate the goodness of the semantic component?

1. Information retrieval/Question answering: given a query, return an answer
2. Machine translation: given natural language text in the source language, return "semantically-equivalent" text in the target language
3. Summarization: given natural language text, return a "semantically-equivalent" but shorter version
4. Multi-document summarization: must also integrate/eliminate redundancy across docs

2. The claim: These can be represented by the single task, given a text T and a hypothesis H, is it the case that "typically, a human reading T would infer that H is most likely true"?[1]
So, NLI is not meant to be a "real" task in itself, but a sort of "condensation" or "crystallization" of a lot of different phenomena boiled down to a 2- or 3-way classification problem.

3. Some phenomena worth pointing out (from the FRACAS problems, but imagine you're building a QA system; there are definitely more real-world test suites.)

a) *Quantifiers*

"*all"* (fracas-003)
*All Italian men want to be a great tenor*
*There are Italian men who want to be a great tenor.*

"*no"* (fracas-006)
*No really great tenors are modest.*
*There are really great tenors who are modest.*

"*few*" (fracas-011)
*Few great tenors are poor.*
*There are great tenors who are poor.*

[1] Dagan, Glickman and Magnini (2006); see also the FRACAS consortium (1996)

"*at most*" (fracas-016)
*At most two tenors will contribute their fees to charity.*
*There are tenors who will contribute their fees to charity*. [answer unclear, <= 2]

b) *Monotonicity* (what is the effect of adding or removing "qualifiers"?)

(fracas-039) *Some delegates finished the survey.*
*Some delegates finished the survey on time*. [answer is "unknown"]

(fracas-038) *No delegate finished the report.*
*Some delegate finished the report on time.*

(fracas-056) *Many British delegates obtained interesting results from the survey.*
*Many delegates obtained interesting results from the survey.* [answer debated]

c) *Attitudes*
(fracas-334) *Smith knew that ITEL had won the contract in 1992*
*ITEL won the contract in 1992.*

(fracas-335) *Smith believed that ITEL had won the contract in 1992.*
*ITEL won the contract in 1992.* [answer is "unknown"]

(fracas-337) *ITEL tried to win the contract in 1992.*
*ITEL won the contract in 1992.* [answer is "unknown"]

4. More realistic examples from the RTE text (we'll display Manning;s (2006) defense of this dataset)

5. OK, how might we solve one subproblem in all the problems involved in inference: How might you learn what are downward-monotone operators? (This is a problem in lexical semantics)