day 1: screen for Prages
- AI check-in. Fri. CMS report. 2-3 pgs. whatever first [handout]
- discussion phenomena hinting @ structure.
- (handout)

9/9/14
Laba 5

(remarks: discussion re: AI projects not scanned)

quick "regrep" synchrim to check feasibility: mosh data? (msinfo)

quick: "visualization" by clustering based on feature-vector choices.

question assumptions from paper, extend concepts from paper.

data cleaning = data carpentry is important. this is why AI helps you do it.

sanity-checking code on synthesized data.
Lecture 6: discourse phenomena, hinting at structure (Write this last)
No class next Tuesday

Last lecture (Tuesday) we talked about on-line conversations, an obvious form of language-based social interaction.

For the next couple of lectures, we'll be briefly introducing some of the classic work on understanding the structure of conversations and other discourse.

Do all ax's on handout to speed things along.

"coherent structured group of sentences [Jurafsky; Martin Text]
- monologues; conversations.

The hard work bit is th
The executive summary is that there is a tremendous amount going on behind the scenes when you speak or listen 'naturally'.

[Note: some will be mostly not currently fully implemented]

For reasons that will rapidly become obvious.

But I still want to present this material b/c I think it's fundamental to understanding discourse, and I think this stuff may represent big opportunities for future systems.

(2) "rules" of conversation [not our focus]

I just want to quickly mention some important related work, that's interesting but not what I want to focus on.

You may have intuitions about how conversations are speed to go.

Like, if someone asks you a question, you're not speed to leave it hanging, but instead at least acknowledge it.

But there are also subtler 'rules' that seem to influence how we
example: Grice's [1975, 1978] maxims (theory of conversational implication) developed as part of his

'maxim of quantity': <see handout #1>

I put in there a recent Google-plus post regarding the accepted papers @ NIPS (a machine learning conference) impressed that Michael Jordan has five papers.

Then there's a comment by the original poster as follow-up. Why is the OP making this comment? What do they have to apologize for? (after all, mesh was true) ... the start post = exactly five: that's the inference we draw.

'maxim of relevance': be relevant.

[Rogers; Norton 2011]: 'Artful dodgers' political scientists exploring the strategy: "Don't answer the question you were asked. Answer the question you wish you were asked." (McNamara) Robert

- people who answer
- experiment: same q, two groups of respondents answered the q, but not very fluently other gave an off-topic answer that was fluent I eval'd more highly

That's all quite cool, but let's now turn to something else.

(W) structure of conversation (what I want to focus on)

(A) excursion into pronounal anaphora: pronouns referring to an antecedent entity - demonstrate a way to infer hidden structure in language

so let's talk about how people determine what a pronoun refers to - since that's usually part of figuring out what someone is talking about.

* Note: pronoun behavior is very dependent. Our goal is to use it to develop intuition about structure, but other longs can indicate structure in different ways.
Starting w/ a very simple example

2(a) Jill blames herself.

(clearly, 'herself' is Jill)

2(b)

/* himself
syntactically unacceptable
(assuming Jill is F. No x if Jill is M)

so, gender and other features have to match.
Fine, but that's not structural.

2(c) Jill thinks Bob blames herself.

* This is not a legal sentence.

What's wrong? In (a) 'herself' was allowed to refer to 'Jill'.

But what here, for some reason it seems like the only possible referent for 'herself' is Bob, which doesn't match gender-wise.

* This is not a legal sentence.

So, maybe there's a locality constraint.

locality constraint??

Three Free counterexamples (??):

2(d) Jill thinks Bob blames her.

is ok, not local

non-reflexive pronoun
maybe these have different roles.

indeed, w/ (a) you couldn't have

'Still, Bill blames her'.

2(e) Bob confronted Bill (all by) himself.

perhaps Bill was all by himself?

But hopefully the point then it's 'more literally', Bob stands.

seen the local matches in feature.
Example #3 on handout.

- "It" is at first blush the wine, despite semantic constraints.
  (Wine could perhaps be brown.
   Maybe "brown" is one of those wine words, like "oaky".)

And the fact that there's a semantically valid possible referent
much more local

=> more evidence for hidden structure that can over-ride strong
"world-knowledge" constraints.
Another convo, artificial, but painted in order to make a pt, w' as few changes as possible.

The example on your handout are a little continued. We wanted to have as little as possible change than different variants, whereas in real life there are often many redundant signals.

So, just bear with me.

Again, using pronouns as a probe...

2(a) "it" : the book & vs.

2(b) which differs only by an 'anyway'

"it" more likely to be theory.

So, the single word 'anyway' is changing how we understand the structure of this discourse.

Whereas now look @: 2(a)

2(a) mixes 2(b) with an extra line

"they're"... "they're" = quarks

It seems weird; Can try to then claim LS, "it" is "theory", but it would have or alternatively in LS say, 'eh by the way back to what I was saying', q' or re-analyze as "it" was the theory that "is" the problem? Or it just change of topic that is the actual "inaccessibility"?

Incoherent diary if short, structure yields.

---

1. Albert
2. The book
3. It actually
4. Theory, quarks
5. Anyway, it
6. They've

Signal: only into new topic?

Signal of return: why does speak better?