Generation of Referring Expressions: Failures of Audience Design

For CS50AD, NLG Course

Kees van Deemter
Computing Science
University of Aberdeen
Shared Knowledge explored: the egocentricity debate

Reading material:

“Imagine the hypothetical question posed by speaker A to listener B: ‘Did you eat of the tree from which I had forbidden you to eat?’

... The Speaker seems to be sensitive to the listener’s knowledge about possible tree referents.”

H&K 1996
“Imagine the hypothetical question posed by speaker A to listener B: ‘Did you eat of the tree from which I had forbidden you to eat?’

... The Speaker seems to be sensitive to the listener’s knowledge about possible tree referents.” 

H&K 1996
The standard theory (the “Initial Design model”)

- H. Clark and others: Speakers and Hearers orient themselves towards their “common ground”.

- Prototypical case: S and H are co-present as observers of some physical scene (and of each other). E.g., Clark & Marshall 1981.
Clark’s model

But what if S and H see different things?

For example, S might have “privileged information” that H does not know about.
In this situation, would it be OK for S to refer to “the ball”? – Probably not!

What’s the best course of action for S?
- S can decide to use only those facts p such that S knows that H knows that p.
- S can guide H’s attention towards the ball, and then proceed as normal, e.g. “Look behind you; the ball …”

Keysar et al. have asked what Speakers actually do. They have also asked what Hearers do. Let’s start there.
Keysar et al. about hearer’s behaviour

- Children **can** reason about other minds from at least age 6
- Yet even adults often **do** not use this ability (experiments by Keysar et al.)
- The simplest experiment:
  - hearer sees three candles of different sizes
  - speaker does not see the largest one
  - hearer knows this
  - speaker says “**Pick up the large candle**”

Caveat: these pictures are mine
This candle not seen by speaker
This candle not seen by speaker

“Pick up the large candle”
Keysar et al. on Hearers

**Results** *(Keysar, Barr, Balin and Brauner 2000)*

- hearers **often** grab the largest of the three, even though they know the speaker cannot see it
- hearers’ interpretation of these utterances is delayed

**Similar in other experiments** *(Keysar, Lin, and Barr 2003)*
Contrast two models:

1. **Initial Design (ID) model**: Speakers use “common ground” when planning utterances. (“Audience design” in the style of Clark.)

2. **Monitoring and Adjustment (M&A) model**: Speakers use their own information when planning utterance. They sometimes use common ground to correct errors in their utterance plans.

Second model seems attractive: S and H often share all crucial information. In these situations it’s not necessary to reason about H’s knowledge.
ID and M&A

M&A:

production → adjustment

ID:

reasoning → production

Tacit assumptions:
• Reasoning is always done
• Adjustment is optional
Choosing between these two models

- Crucial question: does initial plan distinguish between privileged and common knowledge?
  - Initial Design (ID) model: yes
  - Monitoring & Adjustment (M&A) model: no

- H&K: give speakers a reference task. One group is put under pressure to start talking as soon as possible [speeded]. The other group can speak at leisure [unspeeded]. Predictions according to H&K:
  - ID: all should avoid using privileged information (Audience design!)
  - M&A: speeded speakers might not bother; they might use words that rely on privileged information
Experiment

- S and H observe two halves of a screen.
- S’s half contains 2 objects: a target object (which can move) and a context object (which doesn’t move).
- H’s half does not contain a target object (yet). It may contain a context object [shared condition]. Or it may not [privileged condition]. If it does, it’s the same context object as S’s.
- S knows whether he/she is in shared or privileged condition.
- One object moves from S’s half to H’s half.
- Objects can be different sizes and shapes.
Experiment

S describes the moving object to H, e.g.
- “a small ball”, “the small ball”
- “a ball about the size of a pingpong ball”, …

S could use degree adjectives like ‘large’. But these only have meaning to H if a context object is available to H.

H&K: do “speeded” speakers use fewer degree adjectives in privileged than in shared condition?
- ID: yes (unless speeding messes up S’s reasoning!)
- M&A: no …
If speeded speakers use the same number of degree adjectives in privileged as in shared then this may have other explanations:

- maybe they say `the large ball` because the target is (somehow) intrinsically large

To rule out this explanation, every target occurs twice:

- 1x in combi with larger/darker/.. context object
- 1x in combi with smaller/lighter/.. context object

H&K’s analysis focusses on the second time each object is referred to.
Results consistent with M&A model

How many degree adjectives were used? (Calculated as a fraction of the number of words in the NP.)

- Unspeeded: 29% (shared), 9% (privileged)
  The difference was significant.
- Speeded: 19% (shared), 18% (privileged)
  The difference was not significant

So: **speeded speakers did not distinguish between shared and privileged info!**
How long does it take before S starts speaking?
- Un speeding speakers should take more time
- Post hoc (?) predictions:
  - M&A: the difference in onset times should be greater for the privileged condition. For the difference lies in whether Adjustment is done or not, and there is no Adjustment work to be done in shared.
  - ID: the difference in onset times should be the same for the privileged and the shared condition (because the two conditions require an equal amount of reasoning and other work)

Again, M&A confirmed, and ID’s rejected.
My understanding of the argument

M&A:

production ← adjustment

production ← no adjustment

(unspeeded)

(speeded)

In privileged condition, there is much adjustment to do, so M&A predicts a big difference in latency times between speeded and unspeeded (bigger than in shared condition, where there is little or no adjustment to do.)

ID:

reasoning ← production

reasoning ← production

(unspeeded)

(speeded)

No such difference in the effects of speeding is predicted between • shared and privileged: the same tasks need to be performed in both conditions.
H&K conclude

- Common Ground is not routinely reasoned about, but only consulted as a last resort, when time permits
- A current fashion in psycholinguistics (and much of computational linguistics):
  - Diminish importance of “strategic processes”
Assessment of Horton & Keysar 1996

- Their finding that speeded S don’t distinguish between privileged and shared is surprising.
- We have definitely learned: “When speeded, S think less carefully about H”.
- But should we believe M&A?
  - Experiments say little about the **time** when epistemic reasoning happens
  - Maybe ID is correct for **some** situations. *(e.g., If speeded then M&A else ID )*
  - Predictions for latency times seem doubtful: What if reasoning **takes more time** in the *privileged* condition? Then both models predict speeding to make a large difference in latency times
Lessons for theoretical & computational linguists

- Both use idealised notions of the knowledge relevant for speaking and hearing.
- It’s interesting to know what happens when S have to start (or finish!) quickly, or when the situation is not fault critical.
- We’d like to understand how deep speaker’s/hearer’s reasoning tends to be. (Cf. Conway paradox.)
- Formal models of shared information (as used in NLG) need to become more sophisticated.
- Egocentricity might also help to explain other phenomena such as presupposition accommodation.
Wider lesson

- Computational Linguistics can learn from psycholinguistics.
  - Current standards of data gathering/experimentation in CL are primitive
- Psycholinguistics can learn from CL.
  - Current standards for specifying models are primitive.
Wider lesson

- Computational Linguistics can learn from psycholinguistics.
  - Current standards of data gathering/experimentation in CL are primitive

- Psycholinguistics can learn from CL.
  - Current standards for specifying models are primitive. (E.g., M&A model!)