Natural Language Processing 2004 Paper 9 Question 14

Note: This question is taken from lecture 7 of the course.

Part a

referring expressions fragments of language used to perform reference by a speaker (i.e., to refer to a real world entity)

cataphora a pronoun which doesn't corefer with a preceding noun phrase but which corefers with a subsequent noun phrase. E.g., at the start of a discourse:

> Although she couldn't see any dogs, Kim was sure she'd heard barking.

pleonastic pronoun A pronoun which is semantically empty and doesn't refer, e.g.,

> It is snowing

Part b The discourse model consists of a set of referring NPs arranged into equivalence classes, each class having a global salience value.

For each sentence (from the notes):

1. Divide by two the global salience factors for each existing equivalence class.
2. Identify referring NPs (i.e., exclude pleonastic *it* etc)
3. Calculate global salience factors for each NP (see below)
4. Update the discourse model with the referents and their global salience scores.
5. For each pronoun:
   (a) Collect potential referents (cut off is four sentences back).
   (b) Filter referents according to binding theory and agreement constraints.
   (c) Calculate the per pronoun adjustments for each referent (see below).
   (d) Select the referent with the highest salience value for its equivalence class plus its per-pronoun adjustment. In case of a tie, prefer the closest referent in the string.
   (e) Add the pronoun in to the equivalence class for that referent, and increment the salience factor by the non-duplicate salience factors pertaining to the pronoun.

Global salience factors mostly take account of grammatical function — they encode the syntactic hierarchy (i.e., subjects are more salient than objects which are more salient than PPs, etc). Recency weights mean that intrasentential binding is preferred.

The per-pronoun modifications are calculated each time a candidate pronoun is being evaluated. The modifications strongly disprefer cataphora and slightly prefer referents which have the same syntactic role.

Illustrating this on the sample text:

> Owners love the new hybrid cars.

Referring NPs: 'owners', 'the new hybrid cars'

Global salience factors: n and m such that $n > m$ (subject is higher than object) Discourse model:

- 'owners' : n
- 'the new hybrid cars' : m

No pronouns.

They all say that they have much better fuel economy than conventional vehicles.

Update equivalence class and include new NPs:

- 'owners' : $n/2$
- 'the new hybrid cars' : $m/2$
- 'much better fuel economy' : w
- 'conventional vehicles' : z

w and z are higher than $n/2$ and $m/2$ because of the recency factor.

First pronoun: 'They'. Exclude 'much better fuel economy' (agreement). Big negative for cataphora for 'conventional vehicles'. Hence prefer 'owners'.

Second pronoun: 'they'. As above. Binding theory does not exclude this referring to the same thing as the first 'they' (e.g., 'they all say they love their cars' is perfectly grammatical). Hence this is resolved incorrectly to 'owners'.

And it seems that the performance of hybrid cars matches all expectations.

Update equivalence class and include new NPs:

- 'owners', 'They', 'they' : $n/4 + n/2 + q/2$
- 'the new hybrid cars' : $m/4$
- 'much better fuel economy' : $w/2$
- 'conventional vehicles' : $z/2$
- 'the performance'
- 'hybrid cars'
- 'all expectations'

The 'it' in the sentence is non-referring, so no further pronoun resolution occurs.