Natural Language Processing: Assignment III

Reykjavik University - School of Computer Science

October 2015

1 Natural Discourse - 40%

Collect a sample of naturally occurring spontaneous face-to-face discourse. "Naturally-occurring" can be broadly construed to include talk shows, conversation among friends, storytelling, classroom interactions and tutoring sessions. "Spontaneous" means it should *not* be a recital of written text like a movie, TV series or a news cast. "face-to-face" means you need to have visual access to the bodies of the speakers and hearers. "collect" means here that you can either do your own video (and audio) recording (e.g. on your cell phone), or you can find an existing recording. Remember that if you do your own recording, you need the consent of everyone involved (no spying!). You should collect a minimum of 5 minutes, and then transcribe at least 2 continuous minutes. "Transcribe" means you should make a record on paper/computer of what you saw/heard - a good enough record so that when the transcript is read the reader knows what went on.

The point is to push you to think about what discourse is and what makes it hard to model discourse in a computational system. Discuss in about one page the following questions:

1. What makes a sufficient record of the spoken discourse? Did you include everything that matters for understanding? If not, why not?

2. Think about how a computer could replace a participant in the discourse (possibly you as the observer). Supposing that it had perfect speech recognition and word tagging, what could become the most challenging issues for the computer when processing your discourse?

2   Discourse Function vs. Device - 20%

Decide whether the following phenomena are examples of a *Discourse Function* or a *Discourse Device*.

1. Nodding the head

2. The words "and now"

3. Strongly agreeing

4. Changing the body posture

5. Referring to a discourse entity

6. Confirming you heard correctly

7. Touching a sheet of paper

8. Saying "mhm"

9. Making a "thumbs up" gesture

10. Contrasting two options

3   Discourse Model - 40%

Consider the following discourse:

The raven had returned to its lookout in the church tower. I could feel its piercing eyes on my back as I slipped between the tombs towards the gate. Where had it been? What had it seen? A petty thief could expect no mercy from this winged angel of darkness. It did not have to act, just watch in silence. Implant a seed of doubt that my secret was safe. It proved my undoing of course.

3.1   Discourse Entities and Co-Reference

Manually build a *discourse model* for this discourse that contains its *Discourse Entities* and all *Co-References*. Create a table with two columns: First column should contain a unique name for the discourse entity, and the second column should contain (in the same row) all the referring expressions that co-refer to that entity. For example:

Entity	Referring Expressions
RAVEN1	The raven, its

3.2 Where Simple Computers Fail

Explain **four** different referring expressions in the text that a computer using a *Recency List* and *Morphosyntactic Feature Agreement* would not associate with the correct discourse entity. What would the computer need to know to draw the right conclusion in each of these cases?

4 Due Date and What to Return

Return this assignment by Wednesday the 28th of October at 23:59. There is no source code or program to return, only the written answers to the above problems. You do not have to return your video material, just your transcript.