

# Natural Language Processing: Assignment III

Reykjavik University - School of Computer Science

November 2009

## 1 Exercise I - 40%

Collect a sample of naturally occurring spontaneous spoken discourse. "Naturally-occurring" can be broadly construed to include radio or talk shows, storytelling, classroom interactions, task-oriented conversations and classroom lectures. "Spontaneous" means it should not be a recital of written text like a book reading or a news cast. "collect" means here that you can either do your own video or audio recording, or you can find an existing recording. You should collect a minimum of 10 minutes, and then transcribe at least 2 continuous minutes (usually the middle of the discourse is the most natural). "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. You do not need to return your actual collected data, just the transcript and the discussion.

The point is to push you to think about what discourse is and what makes it hard to model discourse in a computational system. You may want to have an interactive NLP system in mind when you choose your sample. 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 in processing your discourse?
3. Are some of these challenges specific to the kind of discourse you chose?

## 2 Exercise II - 20%

In this exercise reflect on the main purpose of language in different situations. For each of the example texts, explain whether you think the primary purpose of its language is **transactional** or **interactional**. Do you believe both purposes can be served at the same time? Support your argument.

1. *In a TV commercial*: "We both know picking the right plan can be difficult!"
2. *Heard in a bakery*: **A** "The total is 1300 Kr." **B** "You accept VISA?"  
**A** "No." **B** "What!?"
3. *Two friends*: **A** "Wazzup?" **B** "Howzit goin?"
4. *On a web page*: "Welcome to my humble home, feel free to browse through my stuff here and don't hesitate to send me comments."

## 3 Exercise III - 16%

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

1. Starting a story
2. Emphasizing a point
3. Making the voice louder and stronger
4. Saying "can you close the window"
5. Being angry
6. Making a fist with the hand
7. Shaking someone's hand
8. Greeting someone

## 4 Exercise IV - 24%

What **Discourse Entities** (DE) get created in the **Discourse Model** when this text is analyzed? Don't just consider noun phrases here, but also consider that DEs can be objects, abstract concepts or events. For each **Coreference** you resolved, briefly state how you could tell what was being referred to.

*In a small house by the sea lived an old sailor named Joe. When Joe was a lad, he travelled the seven seas and saw all the wonders of the world. The sailor had friends in every harbour and they greeted him like a long lost relative whenever he came ashore. They always had a great time together at the pub. This was something the old man relived in his mind every day as he walked along the shore. It was a memory he cherished more than anything.*

## 5 Due Date

Return this assignment by Monday the 16th of November at 23:59.