

## Language Technology 2 Summer Semester 2009

### Exam preparation questions

For the **MR** and **NLI** blocks, the questions below are intended to **help you revise** for the exam. This does not necessarily mean that these very questions will actually appear in the exam. For info about the **MT** part of the exam, see below.

If you have questions about the **NLI** block, please send email to Magdalena Wolska to arrange a meeting to discuss the questions.

#### Multilingual Resources

- What does the term "hybrid NLP" refer to?
- What is the Matrix Grammar?
- Describe the role of semantics in the HoG integration architecture.
- What are the advantages and/or disadvantages of integrating shallow and deep NLP components using semantics as an integrating strategy and interface. Please, be as specific as you can.
- Describe briefly the ParGram Architecture.

#### Natural Language Interaction

- What is a conversation turn? Explain the turn taking rules.
- Explain (and illustrate) what collaboration in dialogue means. What are collaborative responses?
- List and explain the Gricean maxims of cooperation.
- What are adjacency pairs in dialogue, what characteristics do they have, and how is the concept of adjacency pair useful in dialogue modelling?
- What is the general global structure of a conversation? Give some typical examples of the respective parts.
- Explain the notion of a speech act.
- Explain the notion of common ground. When is a piece of information considered common ground between dialogue participants?
- What is the process of grounding?
- What levels of interpretation does Clark (1996) distinguish? Explain what they mean and illustrate with examples.
- What are the various reasons for grounding problems (why can an agent have difficulties grounding an utterance)? Explain and illustrate.
- What (three) kinds of verification strategies can a grounding model of a dialogue system implement? Discuss advantages and disadvantages.
- Explain the difference between finite state machine-based, frame-based and information-state update based dialogue models. What are their advantages and disadvantages? Give examples of applications in which you would use them.
- Explain the concepts behind information state update based dialogue modelling. What kinds of rules have to be defined in such a model?
- What is the difference between fixed and mixed initiative? What advantages and disadvantages has fixed system initiative in comparison to mixed initiative model?
- Explain the concepts of technical evaluation, usability evaluation, and customer evaluation. Give some (at least 3) questions that users may be asked to assess the usability of a system. Discuss the PARADISE and SASSI frameworks?

- What are basic motivations to carry out WoZ experiments? What are the challenges in conducting a WoZ experiment? What requirements must a WoZ experiment meet in order to serve the purpose of a simulated human-machine interaction? Give examples for methods or tools that are employed to meet the requirements.
- Reinforcement Learning (RL) is based on the concept of Markov Decision Process (MDP). Give a short and informal description of the components of an MDP and the relevant notions (i.a. the Markov Property and the three kinds of rewards). Explain how MDP is used in RL in general. RL for natural languages interaction uses recorded dialogue data for training. There are two kinds of information derived from the empirical data which is fed to the learning process. What are they?
- Explain the difference between model-based and simulation-based reinforcement learning.
- Explain the concepts of exploration vs. exploitation in reinforcement learning.

## Machine Translation

*Below is an outline exam for the MT block. The actual exam will be a subset of these with the details of the questions filled out as appropriate.*

- In a rule-based machine-translation system, what role would be played by
  - a morphographemic transducer
  - a morphological analyzer/generator
  - syntactic parser/generator
  - the transfer component
- In a statistical machine-translation system, what role is played by
  - the noisy channel model
  - the translation model
  - the language model
  - the distortion model
  - the notion of fertility
  - the Bleu score
- Bernard Vauquois used a triangular diagram to discuss various strategies used in machine translation. The following question relate to such a triangle.
  - Draw the diagram
  - Why is the diagram narrower at the top?
  - What name is given to the top point in the triangle.
  - What would the diagram look like for a syntax based transfer (or some other) system.
- Why would a language with one of the following properties be easier/harder for a rule-based/statistical system to translate into/out of English?
  - Rich morphology
  - Genders
  - Noun-noun compounds
  - VSO, SOV, free, word order.
  - Separable verbs or verbal particles (like German "ankommen" or "to look up")
- Which of the following properties of a machine-translation system are claimed as advantages by the advocates of statistical/rule-based machine-translation system?
  - (Non)determinism
  - Rapid development
  - Broad domain coverage
  - Robustness
  - Ability to handle long-distance dependencies.

- speed
- better resolution of ambiguities

- Explain briefly:

- Triangulation
- Interlingua
- Vauquois' triangle
- fertility
- the Blue score
- Robnustness

- Which of the following sentences would you expect to be handled better a statistical/rule-based MT system and explain briefly why.

- ...