Shallow NLG Exercise

Stephan Busemann

busemann@dfki.de

http://www.dfki.de/~busemann

Questions Answered by Slideset

- **How does shallow generation differ in principle from (standard) deep generation?**
- **Give advantages and disadvantages of shallow generation.**
- **How are sample corpora used to ensure the required coverage is available and the correct wordings are generated?**
- **Often an NLG problem is addressed by using as a resource clause-length pieces of prefabricated text with gaps that are filled with data during generation. Explain under which conditions this approach is justifyable. Or isn‘t it at all?**

TinyApp

- **Customer wants to report about time series data**

- **Data is a set of pairs**
  - Point of time: an integer
  - Value: an integer

- **User wants tendencies over time**
  - Values may
    - **go down, fall**
    - **stay at the same level**
    - **go up, rise**
    - **… and behave in ways that are hard to describe ;-)**
  - Within a given time interval

1 – 117
2 – 118
3 – 123
4 – 123
5 – 124
6 – 124
7 – 126
8 – 125
9 – 122
10 – 115
11 – 110
12 – 105
13 – …

Tasks

- **Input: a set of data pairs (like on the previous slide); a time interval**
- **Output samples (feel free to make your „system“ generate different wordings):**
  - „You have selected the time interval between 4 and 8. The data go slightly up during this time, rising from 123 to 128.“
  - „In the selected interval [1, 12], data slowly rise from 117 and reach 126 at 7. From 8 onwards they fall quite sharply to 105.“
- **Design input representations suitable for TG/2 (they will differ from the data themselves), and explain your choices.**
- **Sketch the algorithm needed to generate these input representations from the data.**
- **Write a TG/2-style grammar that produces short texts for three selected situations situations (e.g., data go up, go down, remain stable, behave in dizzy ways).**
- **Define access functions to the pieces of TG/2-style input your rules apply to.**
- **Use only few general test predicates in the rules with a well-defined semantics.**
- **In your texts, use hedges (*slightly*, *sharply*). When is this appropriate?**

- **In the lab session, show a „dry run“ through your „system“ for some input**
- **Completeness: Sit back and note what kinds of time series cannot be described by your „system“**