



Tools for Practical Dialogue Modelling

www.ling.uni-potsdam.de/~raquel/teaching/tools-DM.html

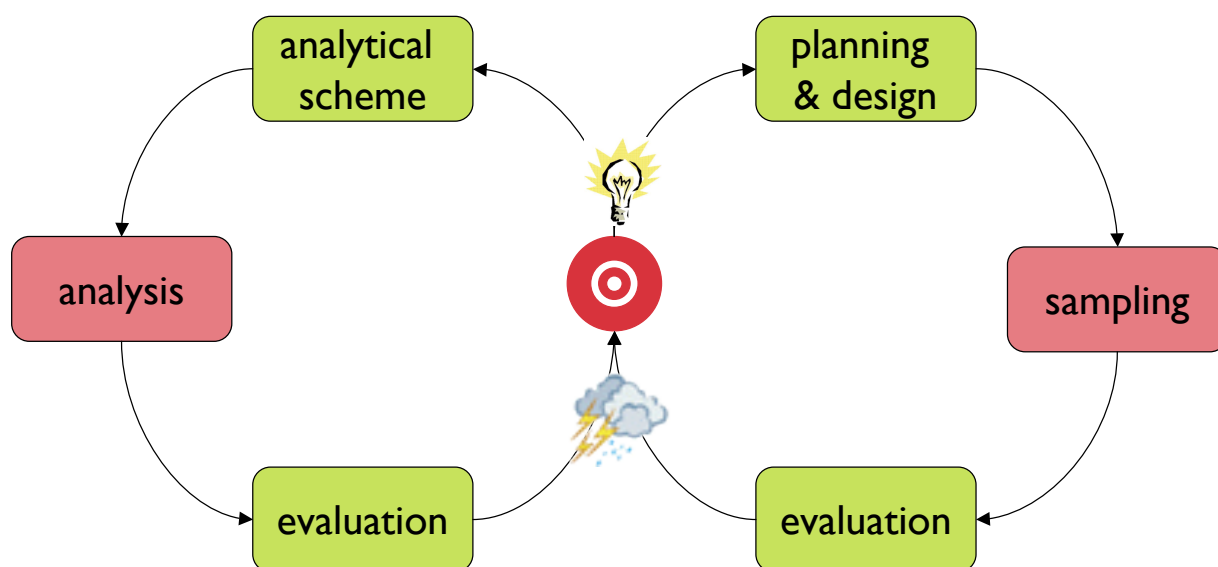
What's this course about?

- it focusses on the empirical and practical aspects of the study of dialogue
- it is a “hands on” seminar - you’ll be doing things
- it has 4 sessions - Fridays 10-18h (with breaks!)
 - Session 1: Annotation and Analysis of Data
 - Session 2: Talking Heads and TTS
 - Session 3: Finite-State Dialogue Modelling
 - Session 4: ISU Dialogue Modelling

Session I: Annotation and Analysis of Data

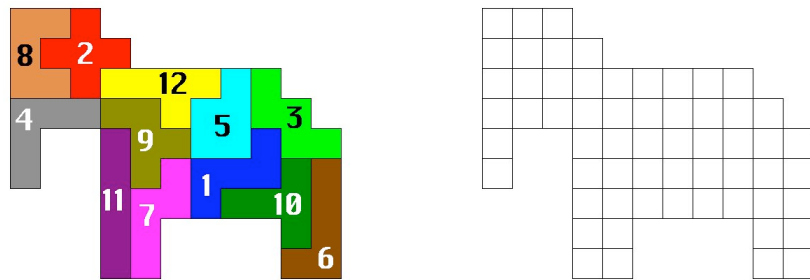
- experimental side of dialogue modelling
- empirical data is a must to get an understanding of the basic elements of dialogue
- dialogue corpora:
 - TRAINS
 - HCRC Maptask
 - Switchboard
 - etc.
- data collection

The experimental cycle



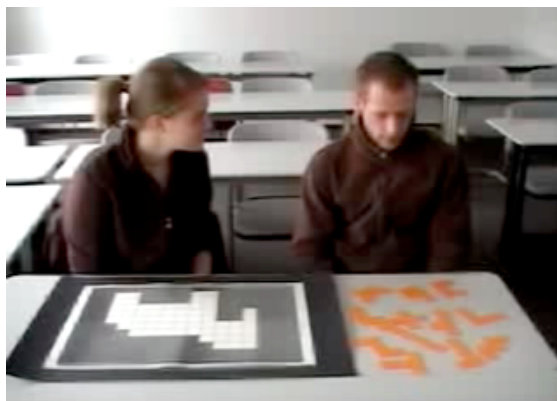
An example: experiment on turn-taking

- comparison of two turn-taking conditions:
 - free turn-taking (FTT)
 - push to talk (P2T)in instructional, task-oriented dialogue
- task: two participants (player and executor), who collaborate in reconstructing a puzzle



Going through the experimental cycle

- Pilot study:



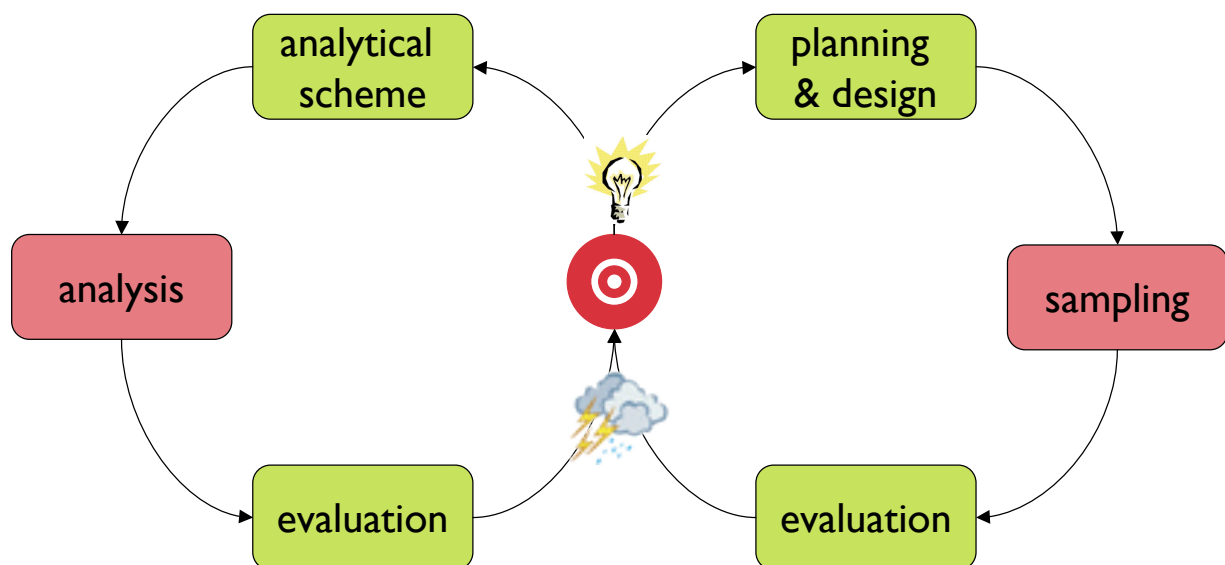
Going through the experimental cycle

- Pilot study:



- Final experimental setting:
 - player and executor in separate rooms
 - FTT v. P2T

The experimental cycle



Plan for today's session

1. use data collected in the turn-taking experiment
2. get a flavour of the transcription task (Praat)
3. decide on annotation schemes for
 - dysfluencies
 - dialogue acts
4. annotation (MMAX)
5. analyse correlations and differences (t-test) in data
 - speaker role
 - gender
 - turn-taking condition
6. evaluate annotation scheme (kappa)
7. present results

Guidelines to design the annotation scheme

- Dysfluencies:
switchboard corpus (Meteer & Taylor, 1995)
<http://www.ling.uni-potsdam.de/~raquel/teaching/DM-materials/DFL-book.pdf>
- Dialogue Acts:
DAMSL (Allen & Core, 1997)
<http://www.cs.rochester.edu/research/speech/damsl/RevisedManual/>

T-test

- used to determine whether the difference between the means of two samples of data is statistically significant
- the t value is evaluated with respect to
 - the set probability α that the null hypothesis is true (often $\alpha=.05$)
 - the degrees of freedom (df): estimate of the number of independent categories $(N_1+N_2) - 2$
- on-line t calculator:
http://www.physics.csbsju.edu/stats/t-test_bulk_form.html
- t table:
<http://www.ento.vt.edu/~sharov/PopEcol/tables/t.html>

T-test

- when reporting the results, you should always include the following:
 - whether the difference was significant,
 - the observed value of t,
 - alpha,
 - the degrees of freedom
- “The difference between the means is statistically significant (t=2.85, p<.05, df=4)”
- “The difference between the means is not statistically significant (t=2.09, p>.05, df=4)”

The *kappa* coefficient

- Kappa can be thought of as the chance-corrected proportional inter-annotator agreement, and possible values range from +1 (perfect agreement) via 0 (no agreement above that expected by chance) to -1 (complete disagreement).

$$\kappa = P(A) - P(E) / 1 - P(E)$$

- guideline for κ interpretation:

Kappa	Strength of agreement
0.00	Poor
0.01-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect