

# Language Technology II

## Language-Based Interaction

Dialogue design,  
usability, evaluation

Manfred Pinkal

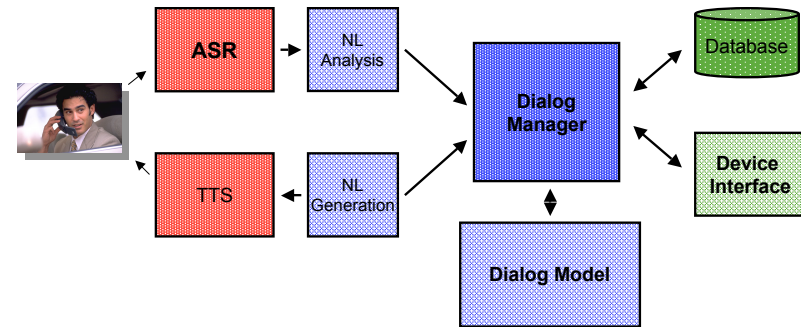
Ivana Kruijff-Korbayová

Course website:

[www.coli.uni-saarland.de/courses/late2](http://www.coli.uni-saarland.de/courses/late2)



## Basic Architecture of a Dialog System (3)



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

2

## Evaluation of ASR Systems

- WER
- Speed (real-time performance)
- Size of lexicon
- Perplexity



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

3

## Word-error Rate

- Standard performance measure is "Word Error Rate" (WER):
  - The **Minimum Edit Distance** between best hypothesis and correct string:  
Insertions+Substitutions+Deletions
  - divided by total number of words in correct string.



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

4

## A Problem for Measurement

- WER is an objective quantitative measure, but: does it really measure the relevant properties of the system?
- Example: All words vs. content word vs. concepts
  - *Ja, das wäre eine gute Idee. Das könnten wir dann machen.*
  - *Ja, dann wäre eine gute Idee. Das könnten wir*



machen.  
7/4/05

LaTelI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

5

## Another Problem

- Confidence values:
  - a measure (usually a probability value between 0 and 1) for the degree to which the system believes into its best hypothesis (sentence level/word level)
- Problem: How can the assignment of confidence values be assessed at all?



7/4/05

LaTelI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

6

## Evaluation of TTS

- Intuitive evaluation by users with respect to
  - intelligibility
  - pleasantness
  - naturalness
- No objective (though quantitative) criteria, but extremely important for user satisfaction



7/4/05

LaTelI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

7

## Measuring of Dialogue Model Quality

- "Query density": Number of concepts per user query/turn
- "Concept efficiency": Number of user turns per concept understood by the system



7/4/05

LaTelI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

8

## Different levels of evaluation

- Technical evaluation
- Usability evaluation
- Customer evaluation
  
- According to: L. Dybkjaer/ N.Bernsen/ W.Minker, "Overview of evaluation and usability", in: W. Minker et al., Spoken multimodal human-computer dialogue in



## Different levels of evaluation

- Technical evaluation
  - Typically component evaluation (ASR, TTS, Grammar, but e.g.: System robustness)
  - To some extent quantitative and objective
  - Possibly glass-box evaluation
- Usability evaluation
- Customer evaluation



## Different levels of evaluation

- Technical evaluation
- Usability evaluation
- Customer evaluation



## Different levels of evaluation

- Technical evaluation
- Usability evaluation
  - Evaluation of user satisfaction
  - Typically end-to-end evaluation
  - Mostly subjective and qualitative measures
- Customer evaluation



## A user satisfaction questionnaire

- Was the system easy to understand?
- Did the system understand what you said?
- Was it easy to find the information you wanted?
- Was the pace of interaction with the system appropriate?
- Did you know what you could say at each



point in the dialogue?

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

13

## A user satisfaction questionnaire

- How often was the system sluggish and slow to reply to you?
- Did the system work the way you expected it to?
- From your current experience with using the system, do you think you would use the system regularly?



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

14

## PARADISE

- An attempt to provide an objective, quantitative, operational basis for qualitative user assessments
- M. Walker/ C. Kamm/ D. Litman: "Towards developing general models of usability with PARADISE", in: Natural Language Engineering 6, 2000



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

15

## PARADISE

- An attempt to provide an objective, quantitative, operational basis for qualitative user assessments, in terms of
- a weighted linear combination of a task-based success-measure and dialogue costs.
- M. Walker/ C. Kamm/ D. Litman: "Towards developing general models of usability with PARADISE", in: Natural Language Engineering 6,

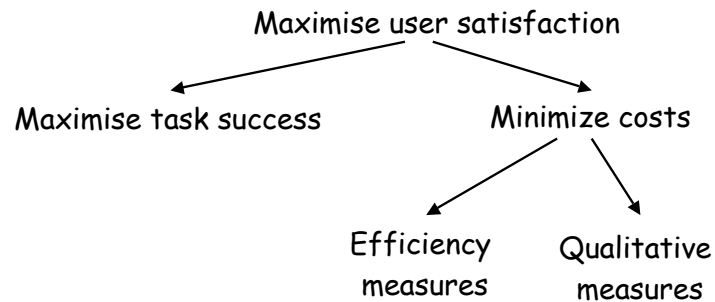


7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

16

## PARADISE Structure



7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

17

## Efficiency and Quality Measures

- Efficiency measures
  - Elapsed time
  - System turns
  - User turns
- Quality measures
  - # of timeout prompts
  - # of rejects
  - # of helps
  - # of cancels
  - # of barge-ins



7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

18

## User Satisfaction

- Measured by adding the scores assigned to 8 questions by the subjects.
  - Systems tested:
    - e-mail message access
    - train information
  - Training set of 500 dialogues with efficiency, quality, task success features
- + satisfaction score



7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

19

## Problems

- Is the user satisfaction score a good measure for user satisfaction?
- Criterion for the feature selection is the easy availability of features through log-files. Is it really the interesting features that are selected?
- Does the methodology extend to more complex dialogue applications in real-world environments?



7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

20

One application: Re-enforcement learning

## Practical Usability Testing

- Implementing a dialogue system and testing it afterwards is usually very inefficient.
- First empirical access: Data collection via human-human dialogue
- Limited usefulness because:
  - Human dialogue behavior differs from dialogue system behavior

 7/4/05 LaTeI: Language-based Interaction 21  
~~- Human attitude towards humans differs from their attitude towards machines.~~

## Wizard-of-Oz Experiments

- Experimental WoZ systems allow to test a dialogue design (to some extent) before it has been implemented.
- The WoZ is not just a person in a box: WoZ setups should
  - constrain the options of the wizards in an appropriate way
  - support real-time reaction of the wizard
  - constrain the access of the wizard to information, to bring him on a par with the artificial dialogue manager

 7/4/05 LaTeI: Language-based Interaction 22  
~~- simulate short-comings of S&L Front-ends, if required~~  
Manfred Pinkal & Ivana Kruijff-Korbyová

## Wizard-of-Oz Experiments

- Ideally, a WoZ environment or tool is set up in a modular way, allowing to replace functions contributed by humans subsequently in the course of system implementation.
- Gradual transition between WoZ and fully artificial system.

## Challenges

- Flexible dialogue modeling in the ISU framework.
- Adaptivity as a feature of advanced dialogue systems
- The difficulty is to find the balance between required degree of freedom and necessary restriction of choices.

## A shift of motivation for WoZ experiments

- Originally
  - Collection of realistic data for speech and language processing (sound files and transliterations, lexical, grammatical variation)
  - Studying user behavior and satisfaction in spite of bottle-necks with speech and language front-ends
- Currently
  - shift of focus to the exploration of wizard's/system's behavior



7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

25

## An example

- TALK project
- MP3 Player
- Multi-modal dialogue, language German
- In-car/in-home scenario
- Saarland University, DFKI, CLT



7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

26

## Goals of WOZ MP3 Experiment

- Gather pilot data on human multi-modal turn planning
- Collect wizard dialogue strategies
- Collect wizard media allocation decisions
- Collect wizard speech data
- Collect user data (speech signals and spontaneous speech)



7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

27

## Tasks for the Subjects

- MP3 domain
  - "in-car" with primary task Lane Change Task (LCT)
  - "in-home" domain without LCT
- Tasks for the subject:
  - Play a song from the album "New Adventures in Hi-Fi" by REM
  - Find a song with "believe" in the title and play it.
- Task for the wizard:
  - Help the user reach their goals



7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

28

## A Walk Through One Exchange

- Wizard: "Ich zeige Ihnen die Liste an."  
*I am displaying the list.*
- User: "Ok. Zeige mir bitte das Lied aus dem ausgewählten Album und spiel das vor."  
*Ok. Please show me that song ("Believe") from the selected album and play it.*



7/4/05

LaTelI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

29

## User View

- Primary task: driving
- Secondary task on second screen: MP3 player



7/4/05

LaTelI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

30

## Video Recording



7/4/05

LaTelI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

31

## DFKI/USAAR WOZ system

- System features:
  - 14 (via OAA) communicating components distributed over
  - 5 machines (3 windows, 2 linux)
  - Plus LCT on a separate machine
- People involved to run an experiment: 5
  - 1 experiment leader
  - 1 wizard
  - 1 subject

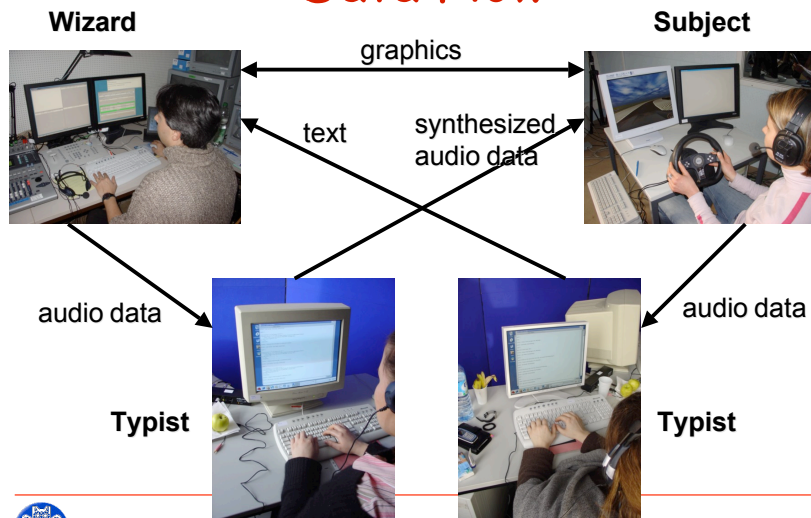


7/4/05  
2 typists

LaTelI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

32

## Data Flow

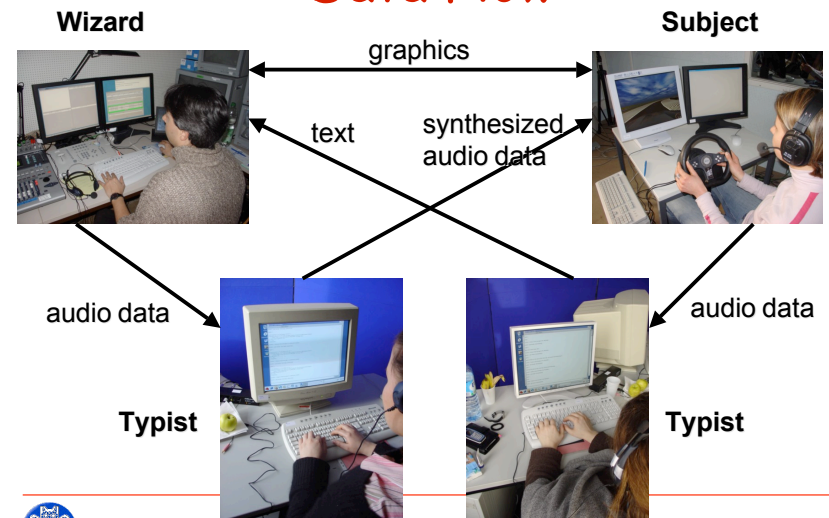


7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

33

## Data Flow



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

34

## A Walk Through the Final Turns

- Wizard:
  - Database search
  - Select "album presentation" (vs. songs or artists)
  - Select "list presentation" (vs. tables or textual summary)
  - "Ich zeige Ihnen die Liste an."  
*I am displaying the list.*
  - Audio is sent to typist
  - Text is sent to speech synthesis
- User: "Ok. Zeige mir bitte das Lied aus dem ausgewählten Album und spiel das vor."  
*Ok. Please show me that song ("Believe") from the selected album and play it.*



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

35

## Example(1) Wizard

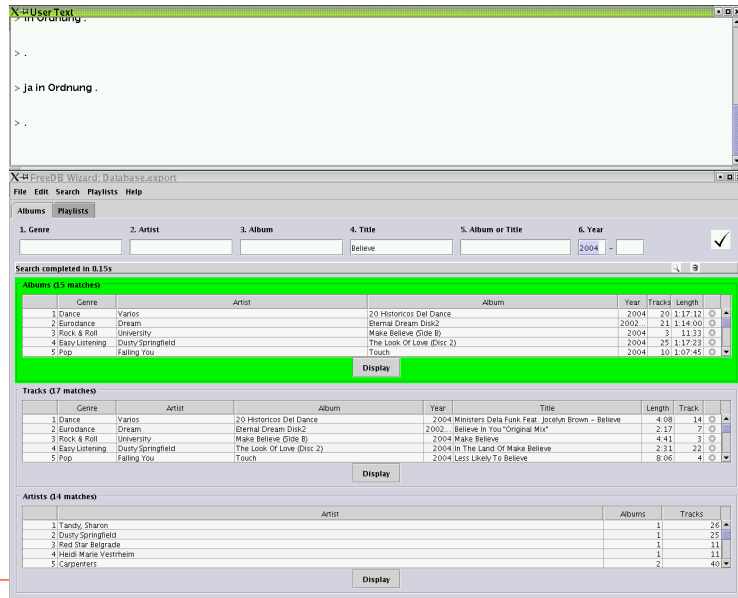
*says: "Ich zeige Ihnen die Liste an."  
(I am displaying the list.) and clicks on the list presentation*



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

36



7/4/05 La Fei: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová 37

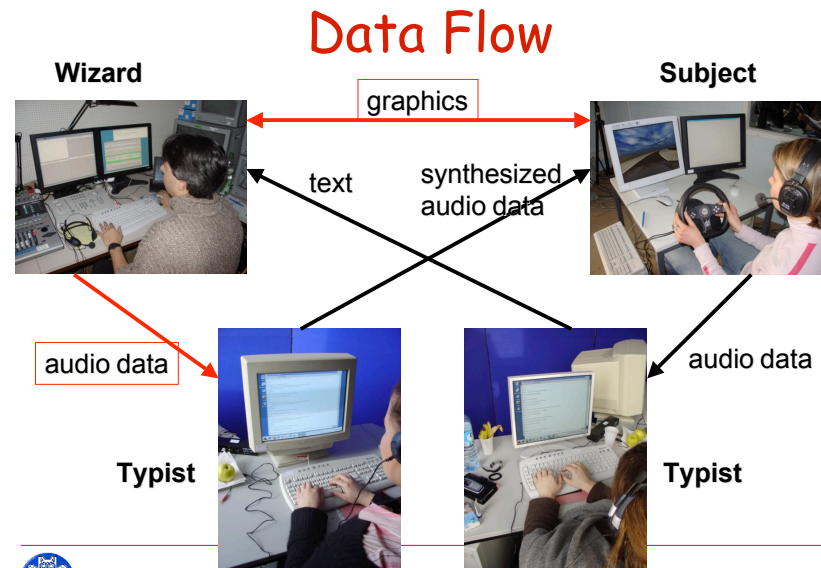


Manfred Pinkal & Ivana Kruijff-Korbayová 38

## Options presenter with User-Tab



7/4/05 Manfred Pinkal & Ivana Kruijff-Korbayová 39



7/4/05 La Fei: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová 40

## Example(2) WizardTypist

types the wizard's spoken text  
I am displaying the list.

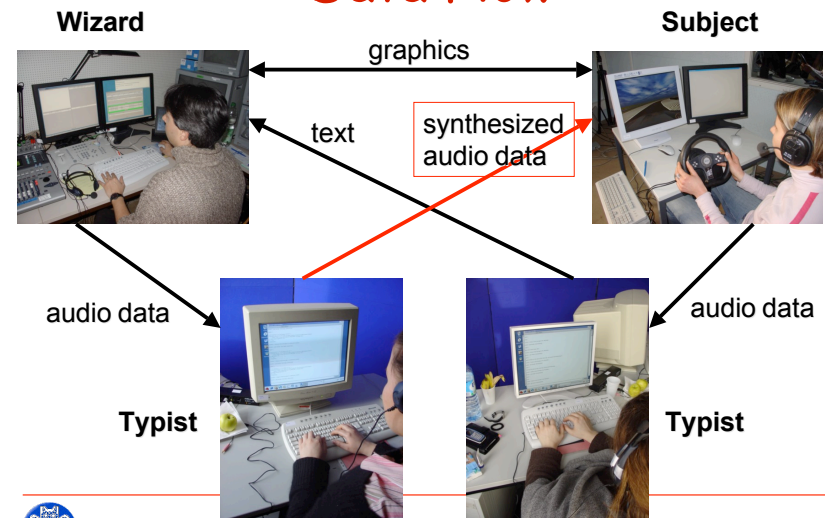


7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

41

## Data Flow



7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

42

## Example(3) User



Listens to wizard text synthesized by Mary and receives the selected list presentation



7/4/05

LaTeII: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

43

## Example(4) User

- Selects one album and says: "Ok. Zeige mir bitte das Lied aus dem aus gewählten Album und spiel das vor."

Ok. Please show me that song ("Believe") from the selected album and play it.



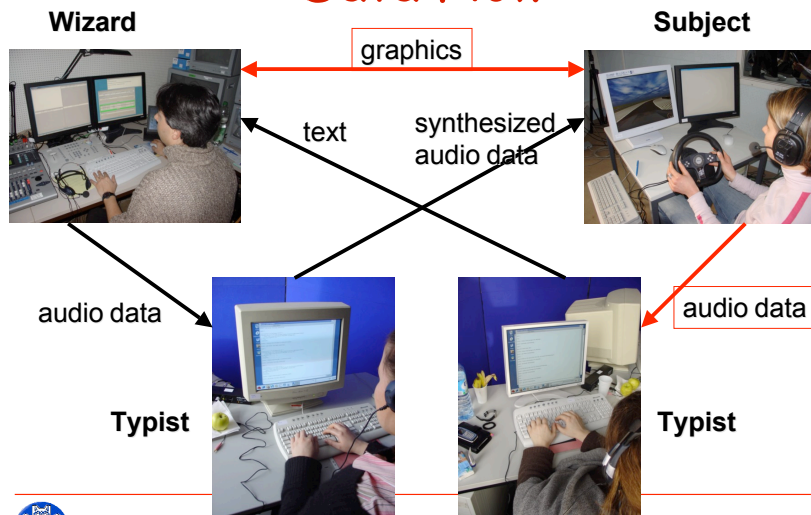
7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

45



## Data Flow



7/4/05

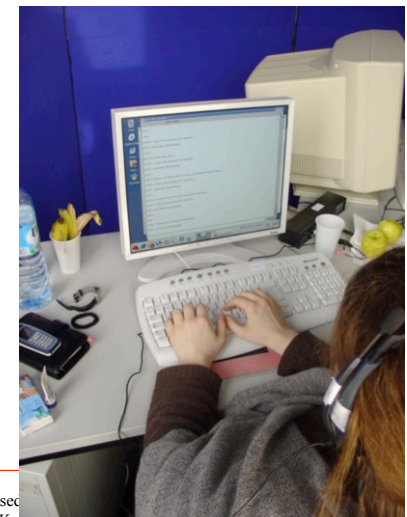
LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

47

## Example(5) User Typist

- Types the user's spoken text

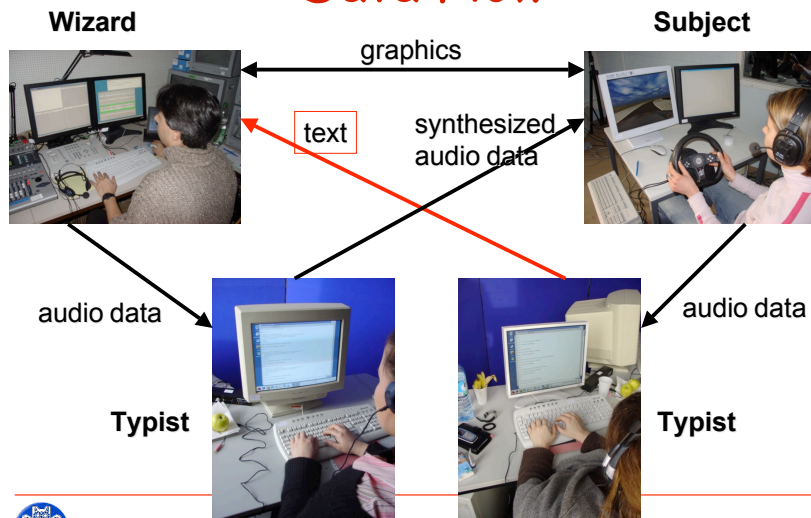
Ok. Please show me that song ("Believe") from the selected album and play it.



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

## Data Flow



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

49

## Example(6) Wizard



Gets a correspondingly updated TextBox Window



7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

50

## Usability testing of System Prototypes

- Debugging (in the system and in the dialogue model)
  - First tests of full prototype system always reveal drastic mistakes
- Polish the dialogue design - Needed:
  - Experimental test runs with
    - developers
    - inexperienced subjects
    - dialogue designers



Dialogue design is more art than engineering.

7/4/05

LaTeI: Language-based Interaction  
Manfred Pinkal & Ivana Kruijff-Korbayová

51