Introduction to Computational Linguistics

Computer-assisted language learning

Jan-Philipp Söhn

jp.soehn@uni-tuebingen.de

October 31st, 2007

Today's Topic

- Computer-assisted language learning (CALL)
- Goals:
  - provide comprehensible foreign-language material in spoken and written form
  - help students understand the material
  - provide exercises and tests
- What we ignore in this class:
  - Diagnosis and Assistance of verbally handicapped people
  - Development of advanced writing skills

Motivation

- Language teaching is much work!
- 60 to 100 hours for an adult to function minimally in a foreign language
- CALL is a good alternative, when teachers are
  - too far away
  - too expensive
  - inconvenient due to scheduling difficulties

Application Sectors

- Academic (Schools, Universities)
- Industrial ("Business English")
- Self-study ("Edutainment")

Second Language Learning (1)

Coverterm for learning a foreign language either in class ("learning") or in a natural environment ("acquisition").

- Krashen (1989): Adults comprehend more initally
  - Use of non-linguistic cues
  - Inference based on world-knowledge
- De Keyser (2000):
  - Adults need to learn grammar rules explicitly
  - Children perform better in abstracting away from the input and inferencing the rules (good adult learners also have good inferencing abilities)

Second Language Learning (2)

What is important:

- holding the attention of learners
- encouraging repetition
- aiming for varied practical excercise

Less agreement is on the

- value of formal grammatical tutoring
- value of correcting errors
- time at which to encourage speaking

History of CALL (1)

- Computer-aided language learning began in the 60ies
- 70ies/80ies: applications on PCs which where didactically quite backward
- 90ies to present: development of multi-media results in more and more CALL applications

History of CALL (2)

"Although CALL employs the computer to assist in language teaching and in language self-study, it primarily uses non-language technology."
The basic technologies used:

- database technology
- digital audio/video
- hypertext
- network communication

See also:
http://www.eurocall-languages.org/resources/history_of_call.pdf

NLP and CALL

- "Linguistics has not been able to encode the complexity of natural language..."
- However, there is need for NLP (syntax, morphology, intelligent tutoring schemes)
- The following techniques are useful for CALL
  - concordancing
  - text alignment
  - speech recognition and synthesis
  - morphological processing (lemmatization)
  - syntactic processing (parsing)
- CALL systems that use NLP techniques are named iCALL systems

Corpora

- 'unmediated' language resource
- useful for teachers and very advanced learners
- supplemented with concordancing and lemmatization
- bilingual corpora
  - convenient translation
  - comparison
  - pragmatic equivalence as opposed to formal linguistic equivalence illustrated in bilingual dictionaries
  - text alignment

Morphology

- For example, GLOSSER (see below) shows:
  - lemmatization of inflected form
  - information on grammatical significance
  - dictionary entry
  - examples of the word from corpora
- enables the learner to read more
- makes texts more accessible
- not included: multi-word lexemes; disambiguation

Syntax

- applications doing language generation walk learners through the rules for sentence construction
- parsers can give an idea of syntactic structure

Error Recognition and Diagnosis (1)

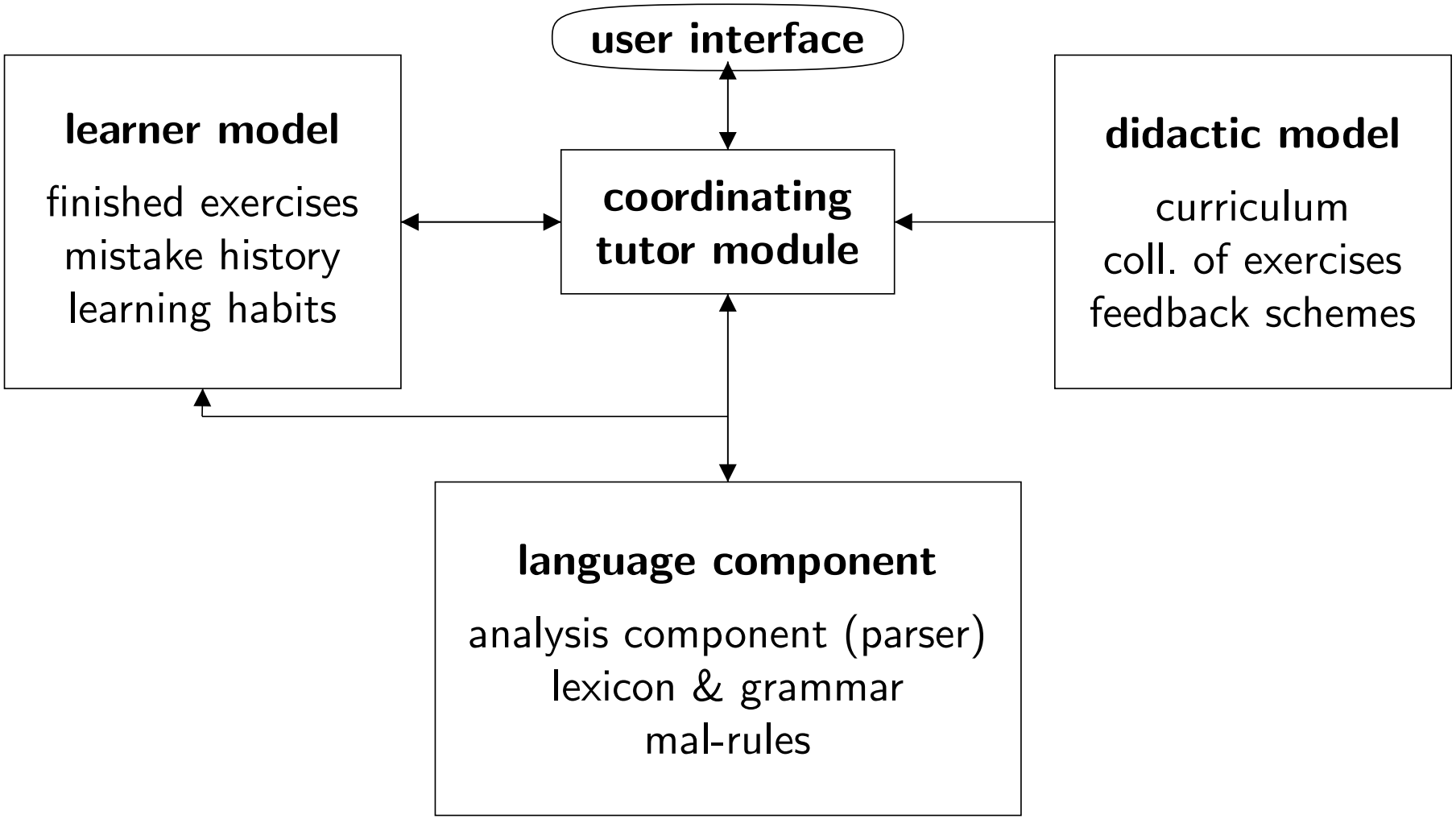
- Allen (1996/97): learners should not be corrected on every minor error
- normal grammar checkers fail to spot problems
- problems for current technology:
  - parsing return multiple analysis (ambiguity)
  - when an error is found, it is beyond the bounds of standard NLP to classify it

Error Recognition and Diagnosis (2)

Still, there is hope...

- The input that is required by the learner is short and thus not ambiguous (multiple-choice, single-word input,...)
- Errors are predictable
- mal-rules which cover errors

Architecture of an iCall System

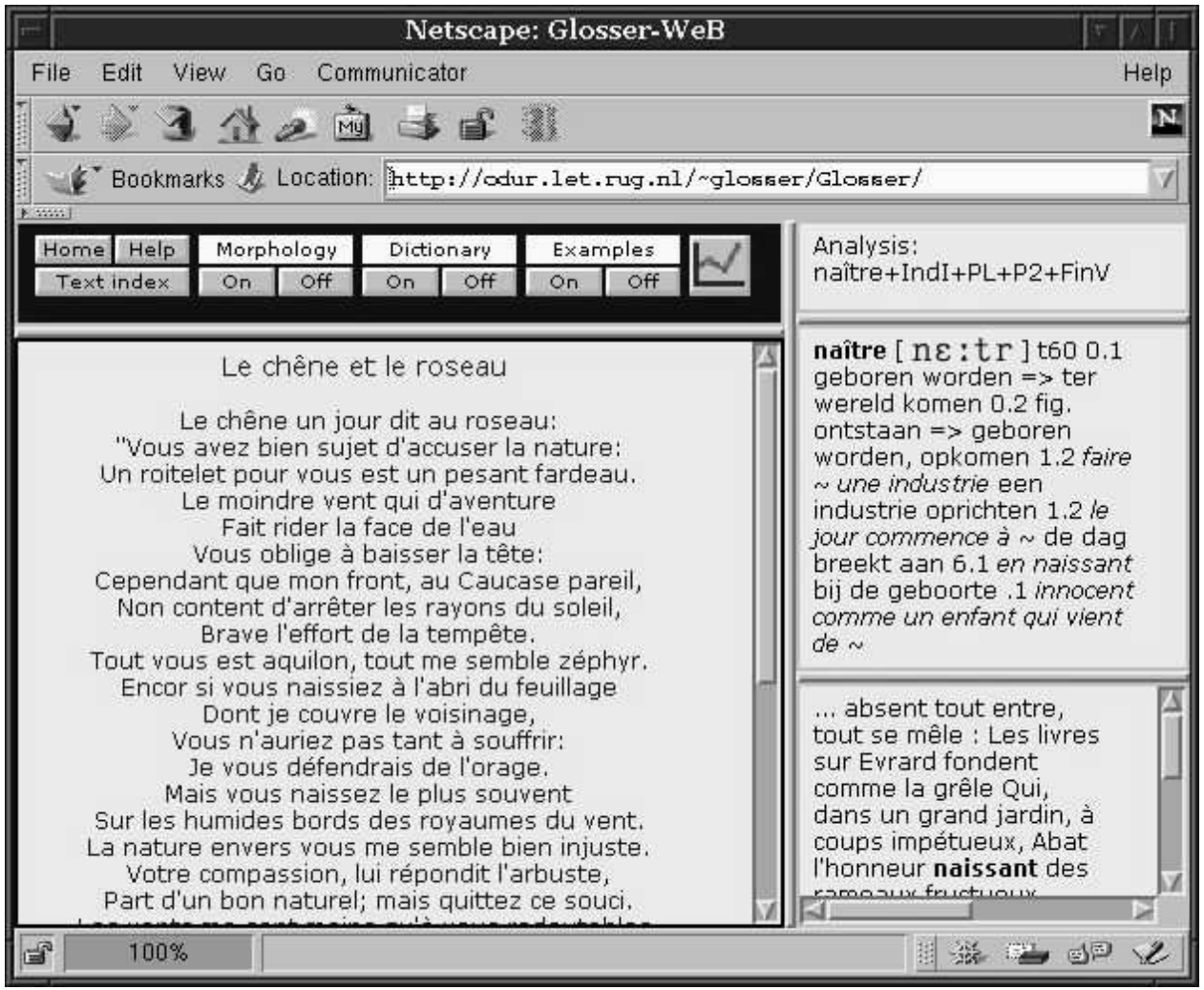


Some Applications

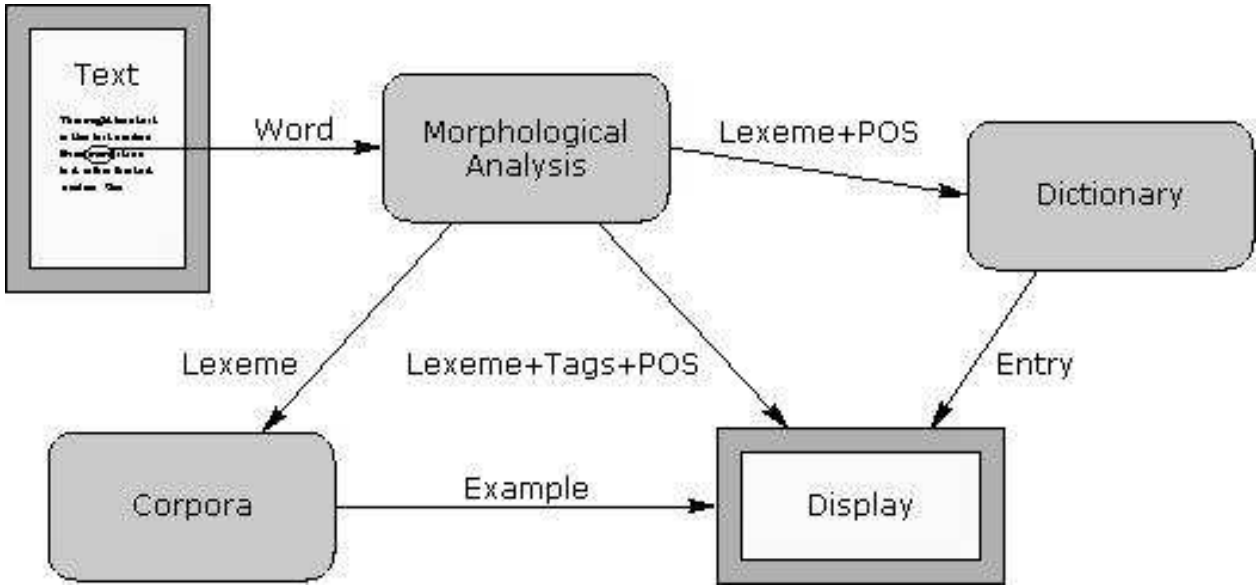
Difference Tools vs. Tutorial Systems

- Glosser-Rug (tool)
  http://www.let.rug.nl/∼nerbonne/papers/gloss-web/
- ALICE-chan (text-based tutor)
- Herr Kommissar (dialogue-based tutor)

Some Applications: Glosser



Some Applications: Glosser



Some Applications: ALICE-chan

- Modular system (extensible, adaptable to other languages)
- For 1st and 2nd year Japanese learners
- NLP hidden from users
- Graphical user interface (GUI)

Some Applications: ALICE-chan (cont'd)

NLP in ALICE-chan:

- Character-independent processing
- NLP lexicon
- Morphological analysis and Segmentation
- Analysis of syntactic structure (parsing, mapping, disambiguation)
- Grammar: LFG with Lisp functions; coverage: 1st–2nd year
- Error detection (mal-rules for parsing errorful structures)

Some Applications: Herr Kommissar

Herr Kommissar (Mr. Inspector)

- role playing detective game
- dialogue-based tutorial system for intermediate German
- extralinguistic, goal-oriented setting

NLP in Herr Kommissar:

- Lexical identification (NLP lexicon)
- Syntactic analysis (agreement, declension types, tense,. . .)
- Semantic interpretation (incl. selectional restrictions) which is tested against a knowledge representation schema
- Response Formulation