

Foundations of Language Science and Technology

## **Technological Foundations II**

Nov. 11, 2009

Stephan Busemann

German Research Center for Artificial Intelligence  
(DFKI GmbH)

### Overview

- Language Technologies vs Human Language Processing
- Evaluation Techniques
- Exploring the LT World (<http://www.lt-world.org>)
- Exercise

**Competence / Performance**

- Competence: skills and abilities needed to solve a problem. Can not be observed directly.
- Performance: behaviour in solving a problem. Can be observed.

**Applied to Language**

- People know the grammar of English. This is their **competence**.
- People produce utterances. This is their **performance**.
- Different people show different performance.
- Their utterances may be deviant or ungrammatical (performance-competence mismatch).

**LT systems**

- No distinction between competence and performance.
- However, a system's performance usually differs in specific ways from human performance when given the same task.

## Human and Machine Performance: Out-of-domain talk

Assume a two-party dialogue application. USR is a human customer in a travel agency. SYS is a consultation system for travel recommendations.

After some talk...

**USR** *I'd like one of the smaller hotels, with a pool. I'm a nonswimmer.*

**SYS** *You may wish to stay at the BelAir. They have both an indoor and a large outdoor pool.*

**USR** *Are these pools deep?*

**SYS** *?? ... ??*

- Out of domain talk may lead to disupture
- System doesn't know the concept of a pool's depth. It doesn't have data about pool depth either. It can't reason about this situation. At most: *„I don't know what you mean by a pool being deep.“*
- A human agent should be able to explain, infer and cooperate: *„I don't know how deep they are. But the hotel has wading pools, too. So you'll most certainly find a safe area in the water.“*

## Human and Machine Performance: Avoiding Errors

Humans try to anticipate and avoid errors by quickly choosing a „safer solution“.

Ex.: style used in foreign language text production

- Speaker should like to say: „*improve the public image of LT*“
- Speaker preverbal message: „*improve the public picture of LT*“
- Speaker realizes that „*picture*“ is the wrong word
- Speaker doesn't use metaphor at all, replans and utters: „*create positive connotations for LT in public*“

LT systems don't usually have a dedicated mechanism for error anticipation and avoidance.

- LT errors from basic methods or component technologies show in the output
- No feedback architecture allowing inter-component interaction

## What is Language Understanding?

**Understanding „understanding“:** Verifiable Scenario in which an actor demonstrates an intellectual effort that involves reasonable action (verbal or nonverbal) as a consequence of a linguistic stimulus

There are different ways to define „language understanding“, e.g.;

- Tourist satisfied with a trip recommended by a computer agent in the course of a NL dialogue
- Agent translates a text from one language into another
- User constructing an electric circuit based on NL advice provided by a computer
- Robot seeks, finds and fetches a book after being told to bring it

What language understanding is not:

- Successful runs of a parser that maps text input onto a logical form output (no reasonable action)
- Phone routing systems (predefined interpretations of digits)
- Airport flight information (predefined utterances)

## Modeling Language Understanding is Always Partial

### Linguistic coverage

- I'd like to fly to Cuba.
- Are there still flights to Cuba?
- Can you please book me a flight to Cuba!
- ...

### Conceptual (out of domain) coverage

- With BA, food is better.
- I have fear of flying.
- Why not by car?
- ...

### Social coverage (adolescence)

- No social learning
- No social experience
- No social integration

*In constructing a model we necessarily exclude anything that is not modeled*

*Constructing models is not the right way to making computer performance more similar to human performance*

*Adding models of thought, behavior, social roles etc. will improve performance, but still remain deficient*

## Social interaction: evolving functionality

Humans learn language in context – they see, smell, touch and think at the same time. Humans do this over many years.

A human-like – more holistic – view on a computer acquiring human language is based on situated interaction:

- explore environment with laser scanner, various sensors
- represent objects perceived in a knowledge space (ontology)
- spatial recognition (shape, size, color – must be a cup)
- understand the concept of space and reason about it (I see a sofa, so maybe I'm in the living room)
- learn (generalize) from linguistic interaction („This is a cup!“), annotate ontology with linguistic terms
- understand and generate referring expressions („the large blue cup“)

The talking robots group at DFKI is building cognitive robots --  
<http://talkingrobots.dfki.de>

## Evaluation Techniques I

How can we assess whether our technology lives up to expectations?  
How can we compare a technology with other technologies that do the same thing?

**Glassbox evaluation** (competence predicted by theory) vs.  
**Blackbox evaluation** (performance of implemented system)

### 1. Introspection

- Author of system sits back and checks what is plausible
- Self-evaluation
- No general validity of results

### 2. Group tests

- A group of possible intended users (= hire a few motivated undergraduates) is testing the system
- Slow, costly, difficult to get reliable results
- General validity questionable

## Evaluation Techniques II

### 3. Comparison against „gold standard“

- Corpus of representative texts
- Annotated with the correct results („solutions“)
- Comparison with system results
- Measures
  - Precision: | found-solutions | / | items-found |
  - Recall: | found-solutions | / | solutions |
- Great for tasks with independent, unique solutions such as NER, Chunking, Dependency parsing
- Difficult to measure non-exact results (multiple adequate solutions)
  - Machine translation
  - Summarization
  - Generation

Name: Grass  
FName: Günter  
Prize: Nobel  
Area: Literature  
Year: 1999

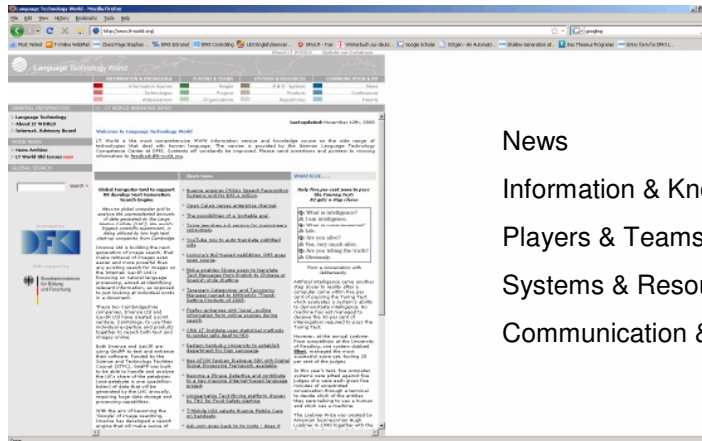
„Grass roots in German literary traditions.“

„Grass roots organizations give voice to the people“

-precision

+recall

<http://www.lt.world.org> is the major Internet portal for language technologies.  
 First result when googling „language technology“. Maintained at DFKI.



News  
 Information & Knowledge  
 Players & Teams  
 Systems & Resources  
 Communication & IPR

**Improving on the LT World: <http://beta.lt-world.org>**

Choose a language technology in one of the subsections of the Technologies area.

1. Consider the information associated with it.
  - Is it still current?
  - Can you find newer relevant information on Google?
  - Do available link resources maintain relevant newer information?
2. Do you have other recommendations regarding missing / outdated technologies?

Write up your findings (one page only) and present them on Friday (5-10 mins).

If your results are used for an update of LT World, you will be duly acknowledged.