Foundations of Language Science and Technology

Technological Foundations II

Stephan Busemann

German Research Center for Artificial Intelligence
(DFKI GmbH)

Overview

- Language Technologies vs Human Language Processing
- Evaluation Techniques
- Exploring the LT World

Performance Revisited

Competence / Performance

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

Applied to Language

- People know a 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
  (possibly, this is due to a 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 an automated 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 ?? … ??

Human and Machine Performance: Out-of-Domain Talk

- Out of domain talk may lead to disrupture
- 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, or emergency measures are taken (*„Can you rephrase, please?“*)
- 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.;

- A tourist is satisfied with a trip that has been recommended by a computer agent in the course of a NL dialogue.
- An agent correctly translates a text from one language into another
- A user constructs an electric circuit based on NL advice provided by a computer assistant.
- A 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 to Havanna.*
- *…*

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

Situated Interaction: Evolving Functionality

Humans learn language in context –
they see, smell, feel, think and speak simultaneously.

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

- explore environment with laser scanner, various sensors
- represent perceived objects 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 probably I'm in the living room“)
- learn (generalize) from linguistic interaction („This is a cup!“), annotate ontology with linguistic terms
- understand and generate refererring 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

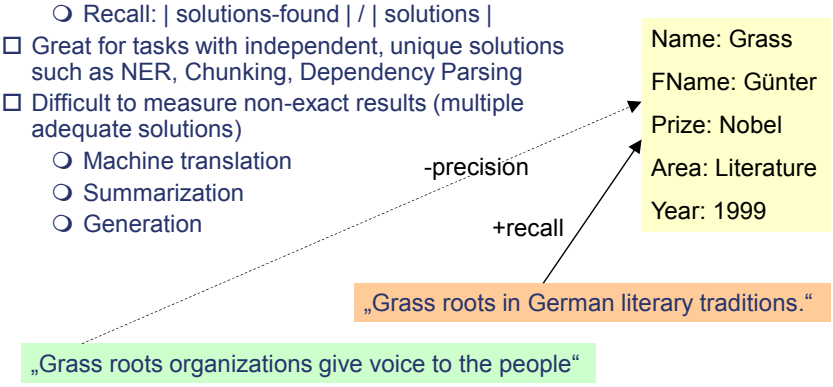
- Author of system sits back and checks what is plausible
- Self-evaluation
- No general validity of results

2. Group tests

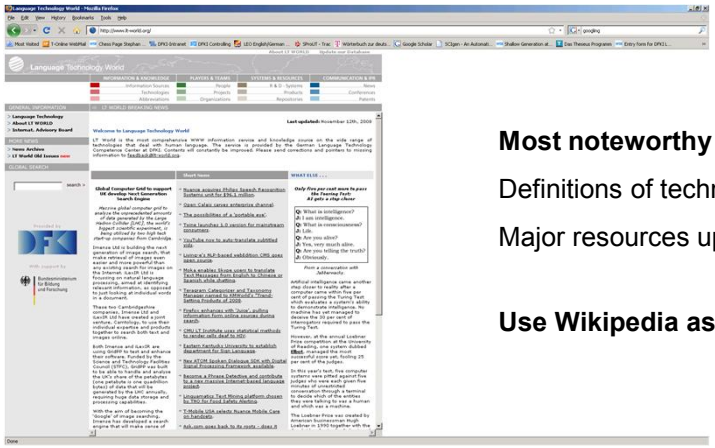
- A group of possible intended users (= hire a few motivated undergraduates) is testing the system and/or answering questionnaires
- Slow, costly, difficult to get reliable results
- General validity questionable

3. Comparison against „gold standard“

- Corpus of representative texts
- Annotated with the correct results („solutions“)
- Comparison with system results
- Measures
  - Precision: | solutions-found | / | items-found |
  - Recall: | solutions-found | / | 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



http://www.lt.world.org is a major Internet portal for language technologies. Top of results when googling „language technology“.



Most noteworthy today:

Definitions of technologies

Major resources up to 2012

Use Wikipedia as well!