

Terminology Applications and Human Language Technologies

Justus Roux
Stellenbosch University
Centre for Language and Speech Technology



Aim

To highlight a particular relationship existing between

- Terminology (and ontology) development and
- the development of Human Language Technologies (HLT)

Human Language Technologies?

- **Enabling technologies**

- Facilitate a process of Human-Machine interaction in a natural way – through language (text and speech)
- Render support to human operators in specific “language” tasks – e.g. machine aided translation
- Allow tasks to be performed by systems using speech - physically challenged persons

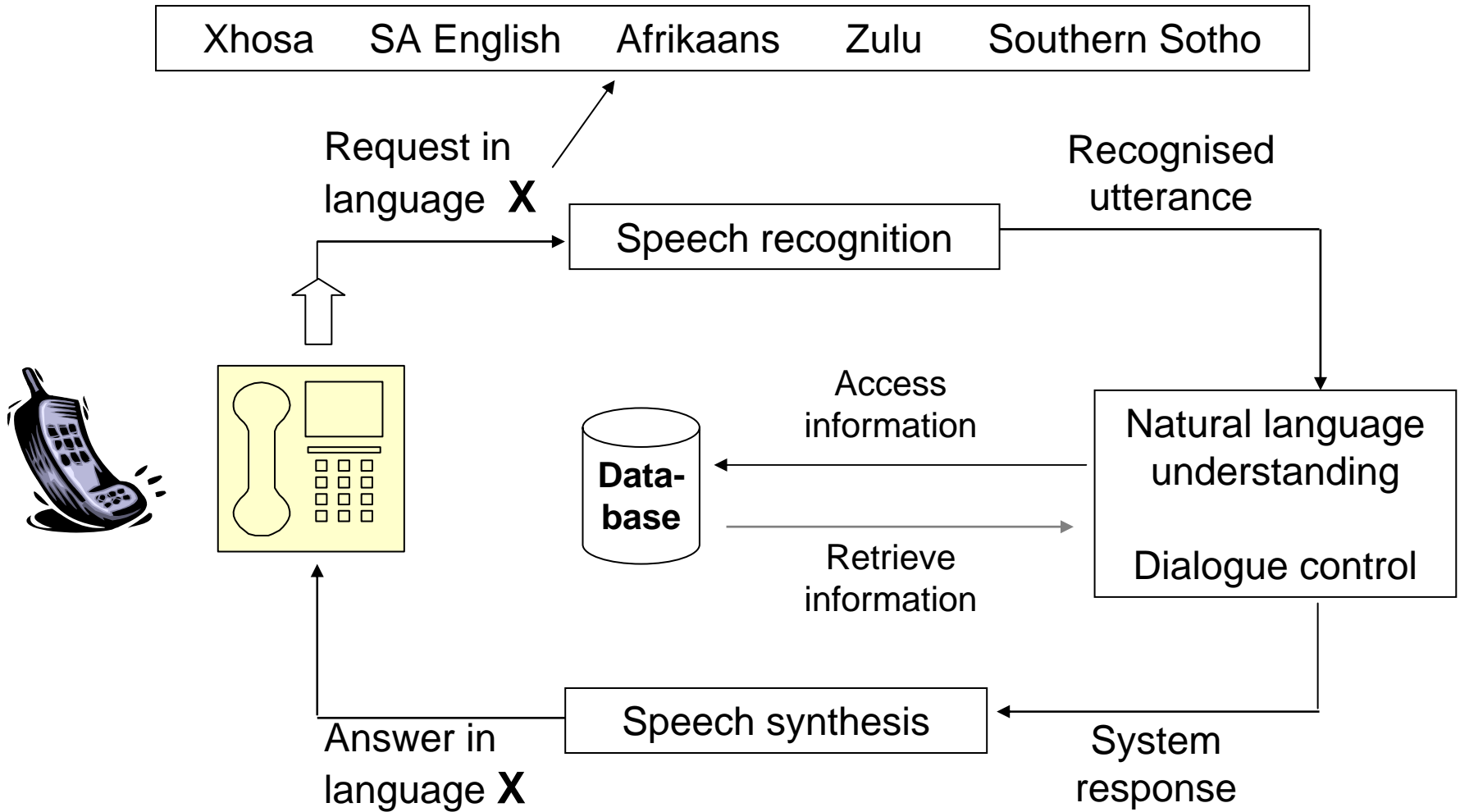
Typical/Potential HLT Applications

Telephone based information systems (automated)

- Tourism & Travel: Hotel booking systems (AST project); train, air, bus schedules
- Health services: First level medical help lines, Aids hotlines, TB hotlines
- Public services: Applications for pensions, travel documents, car registrations; telephone accounts, telephone number enquiries



Interactive voice driven information / transaction system



Demonstration

Text to speech system as used by AST project
(Implicit confirmation)

System:

OK, now you have to tell me when will you arrive
at **the Blue Bay Inn**?

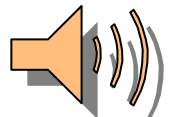
User:

On the first of June.

System:

So you will be arriving on **the first of June 2006**

[Pre-recorded →] [Synthesised →]



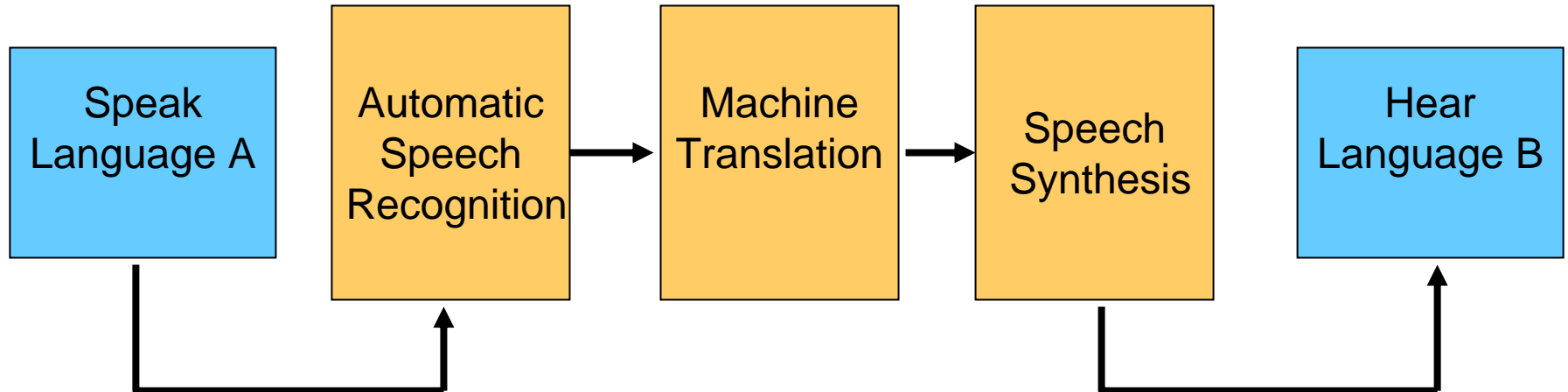
Multimedia information systems

- Education: Language learning, voice based training systems
- For the blind: Speaking books, newspapers – making Braille obsolete
- For the deaf: Screens on telephones converting speech into text

Translation systems / translation memory systems

- State services: Official documents, Hansard in national, provincial, local governments
- Education: Developing multilingual teaching material
- Business: Translation of technical manuals, instructions on the use of products, etc

Speech to Speech Translation Systems



Limited Domain Applications
English – Afrikaans: Naval terms (V1)
<English – Xhosa>

Intelligent Information Systems

Semantic Web

- Limitations of the Internet as information provider (e.g. access through keywords)
- Next generation of the Internet
- Intelligent searches using natural language and speech as input (HLT)
- Content queries understandable by machines
- **Need for descriptive terminology and ontologies**

The multi-faceted nature of Terminology

“service”¹

- Various disciplines involved in **on-line service offerings**:
 - Business Science (BS)
 - Computer Science (CS)
 - Information Science (IS)
- Manufacturing industries, service industries, governments move towards Internet use, i.e. **“e-service”**

1 All examples in next four slides: Baida, Z, Gordijn, J, Omelayenko, B. 2004. *A shared service terminology for online service provisioning*. ICEC`04 6th International Conference on Electronic Commerce. Jansen et al. (Eds)

Three definitions in BS community for term **e-services**

- (i) Traditional : “..deeds, processes, performances.., with outcomes or benefits ...”
- (ii) “where the Internet is used as a User Interface, a channel to interact with customers” [i.e. an Internet version of ‘traditional services’]
- (iii) “...the provision of service over electronic networks” [i.e. wider definition including Internet, ATMs, smart card networks, kiosks and “...all touch points with customers.”]

Computer Science community (Semantic Web)

Web-services: Software applications used on the Internet [No explicit reference to business processes or functionalities.]

E-services as synonym for Web-services

Business services used but not defined.

Information Science community

- **Web service** as used in Computer Science
- **Service** as used in Business Science
- **E-services** interpreted as Internet-based version of 'traditional' services (Business Science) or as Web-services (Computer Science)

Point to be made

- Intelligent HLT systems (Semantic Web) are based on the need for a **shared understanding of concepts and terminology across disciplines**
- Avoid confusions around terminology by “...**being aware** of the existence of multiple interpretations for the same terms”

(Baida, Z, et al. 2004:7)

Terminology within the shared domain of HLT

- **Different disciplines**
 - Linguistics (All sub-disciplines)
 - Computer science
 - Electronic engineering
- **Terminology**
 - “Speech”
 - “Grammar”
 - “Transcribe”
 - “Dialogue”
 - “Program”

Ontologies and the Semantic Web

- **Knowledge representation:**
 - Terminologies: Developed for human use
 - Ontologies: Developed for use with computers
- **Ontologies created from existing terminologies**
- **Machine readable taxonomies to facilitate a wide range of applications**
 - Information retrieval
 - Automatic summarisation etc

HLT in service of Terminology Development

- **Tools for**
 - Terminology extraction from large corpora
 - Terminology management

HLT in service of Translation and Editing

- **Machine (aided) translation**
- **Speech based editing: Listening to TTS**

HLT in service of Language Planning

- **HLT in the promotion of multilingualism**

The role of HLT in current NLS activities

