

Recent Trends in the Development of Speech Technologies and Dialog Systems

David Griol Barres

Applied Artificial Intelligence Group

Universidad Carlos III de Madrid

dgriol@inf.uc3m.es



Outline

1. Introduction
2. Spoken Dialog Systems
3. Recent Trends: Adaptation
 - a) Statistical dialog management
 - b) User Modeling
 - c) Emotion Recognition
 - d) Context-aware system
 - e) Multimodal Interaction
4. Application domains
5. Conclusions



Outline

1. **Introduction**
2. Spoken Dialog Systems
3. Recent Trends: Adaptation
 - a) Statistical dialog management
 - b) User Modeling
 - c) Emotion Recognition
 - d) Context-aware system
 - e) Multimodal Interaction
4. Application domains
5. Conclusions



Introduction

- New devices to access information in the digital domain



- Necessary to provide straightforward ways of retrieving it.



- Objective: effective, easy, safe and transparent interaction between the user and the system.



- Simulation of spoken human-to-human communication.

Dialog Systems

Outline

1. Introduction
- 2. Spoken Dialog Systems**
3. Recent Trends: Adaptation
 - a) Statistical dialog management
 - b) User Modeling
 - c) Emotion Recognition
 - d) Context-aware system
 - e) Multimodal Interaction
4. Application domains
5. Conclusions



Spoken Dialog Systems

- **Dialog system:** man-machine interface that is able to recognize and understand a spoken input and to produce an oral output as an answer.
- Different modules usually take part to carry out this final goal.
- Spoken conversational agents are generally based on:
 - Communication channel: telephone line, microphone.
 - The task is restricted to a certain knowledge domain.
 - Mixed initiative strategy.
 - Natural language inputs are accepted/generated.
- **Applications:** The complexity of the conversation can vary from a simple option choosing in a menu, to a conversational companion.



Spoken Dialog Systems

- Recently, dialog systems have been augmented in their input and output with other modalities different from voice (e.g. gestures, hand-writing or facial expressions). These are **called multimodal dialog systems**.
- Humans communicate using natural language. Interfaces based on keyboard and mouse force users to adapt to the machine, which makes the access to new technologies difficult for some population groups like the elderly or disabled.
- Oral communication is more straightforward and natural, but today SDSs still cannot dialog like humans:
 - Misunderstanding, world knowledge, flexibility, adaptability, learning from experience, etc.



Spoken Dialog Systems

Applications through examples ...

- One of the most wide-spread applications of dialog systems is information retrieval.
 - Is there anything more natural than just ask for the information you want?
- These systems usually have the added functionality of booking.
 - E.g. ask about train timetables and also book the tickets.
- Some examples:
 - Railway, flights, buses and other transport information
 - ARISE - Automatic Railway Information Systems for Europe (several languages)
 - Let's Go: A spoken dialog system for non-native and elderly. For the domain of bus information, around Pittsburgh.
 - Mercury flight reservation system Weather forecast
 - Jupiter from MIT
 - Tourist and city practical information
 - AUGUST - Swedish spoken dialog system about the capital.
 - Voyager - Tourist and travel information for the Greater Boston area



Spoken Dialog Systems

Applications through examples ...

- Some examples:
 - Other...
 - Adapt - Multimodal spoken dialog system for browsing apartments on the Stockholm real estate market
 - Amities project - Automated Multilingual Interaction with Information and Services
 - DARPA Communicator: Intelligent conversational interfaces to distributed information
 - DisCoH Spoken Dialog System for Conference Help
 - SmartKom German multimodal dialog system. Several application domains including cinema booking
 - CTT-Bank Speech controlled telephone banking system.



Spoken Dialog Systems

Applications through examples ...

- Some examples:
 - Spoken interaction can be the only way to access information in some cases, e.g., the screen is too small to display information (e.g. handheld devices) or the eyes of the user are busy in other tasks (e.g. driving)
 - Athosmail - a multilingual adaptive spoken dialog system for e mail message reading on a mobile phone
 - DICO - a multimodal dialog system that lets the driver control devices and access Internet services using natural speech.
 - MUST - Multimodal, multilingual information services for small mobile Terminals
 - ISL-VODIS and VODIS: Voice Operated Driving Information System
 - VICO: Virtual Intelligence CO-driver enabling natural interaction between humans and digital devices and services in the car.

Spoken Dialog Systems

Applications through examples ...

- Some examples:
 - Spoken interaction is also useful for remote control of devices and robots:
 - ARGUS - An ergonomic dialog system to control electrical devices in houses using gesture recognition
 - D'Homme project - dialogs in the home machine environment.
 - Cogniron - the cognitive robot companion
 - Clarissa Voice-operated procedure browser for the International Space Station (ISS)
 - The most demanding application for fully natural and understandable dialogs, are virtual agents and companions:
 - AVATALK - natural, interactive dialogs with responsive virtual humans
 - Collagen - Conversational Assistants. Building collaborative agents



Spoken Dialog Systems

Applications through examples ...

- Example of multimodal dialog systems
 - Smartkom
 - Welcome to the SmartKom information system. I am Smartakus. How may I help you?
 - What is playing at the cinema tonight?
 - These are the movies playing tonight. The cinemas have been marked on the map

Spoken Dialog Systems

... Apps. through examples

The screenshot displays the 'COMPANIONS DIALOGUE SYSTEM MONITOR' interface. On the left is a 3D avatar of a woman with blonde hair wearing a white blazer. The central area shows a dialogue log with alternating system and user messages, including emotion fusion labels like 'Negative-Active'. On the right is the 'ENGLISH COMPANION PROTOTYPE' architecture diagram, which consists of a grid of colored boxes representing different system components: ATT (blue), ASR (blue), EV (blue), DAT (purple), SA (purple), EM (purple), NLU (orange), DM (orange), NLG (orange), ECA (yellow), KB (green), and IM (green).

COMPANIONS DIALOGUE SYSTEM MONITOR

ENGLISH COMPANION PROTOTYPE

Dialogue Log:

- SYSTEM (emo) Hello john, how did your workday go today?
EMOTION FUSION: Negative-Active
- USER I worst day never I arrived late because of the traffic
- EMOTION FUSION: Negative-Active
- SYSTEM (emo) you have my sympathy!
SYSTEM (emo) what happened next?
- EMOTION FUSION: Negative-Active
- USER I I miss the meeting on project responsibility I got more work dumped on me just what I don't need as I'm already likely to miss the deadline from a report
- EMOTION FUSION: Negative-Active
- SYSTEM (emo) I'm sorry to hear that!
SYSTEM (emo) who was at the meeting?
- EMOTION FUSION: Negative-Active
- USER I only my team they gave me an set of PowerPoint slides from the meeting I spent the rest of the day leading the slides but my computer crashed several times because of the updates
- EMOTION FUSION: Negative-Active
- SYSTEM (emo) oh dear!
- SYSTEM (emo) it's understandable how you must be feeling
- SYSTEM (emo) it's right to be unhappy about such a bad situation
- SYSTEM (emo) try not to worry about missing the deadline because of the technical problems
- SYSTEM (emo) it's sad that your motivation could be effected by them
- SYSTEM (emo) I can sympathise with your situation but unfortunately the technical problems are unavoidable

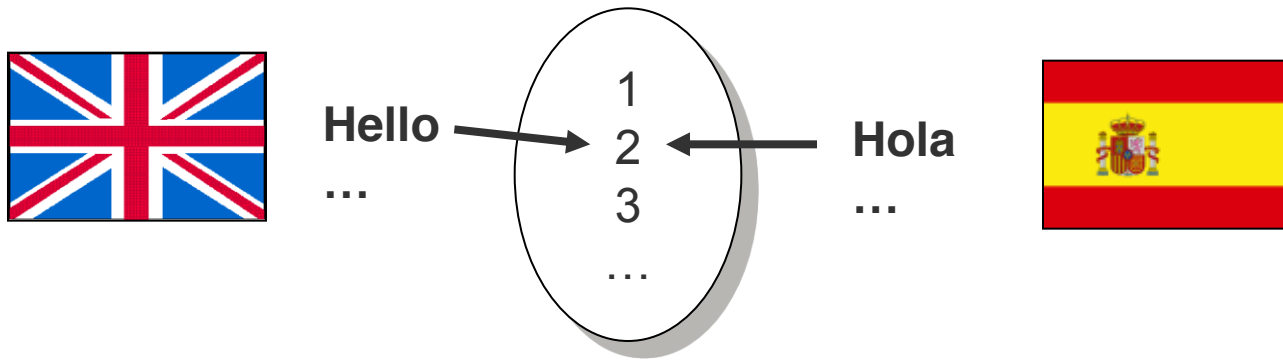
System Architecture Diagram:

- ATT (blue)
- ASR (blue)
- EV (blue)
- DAT (purple)
- SA (purple)
- EM (purple)
- NLU (orange)
- DM (orange)
- NLG (orange)
- ECA (yellow)
- KB (green)
- IM (green)

Spoken Dialog Systems

... Brief historical review ...

- In the XVII century, some philosophers like Descartes or Leibniz began with the construction of universal languages in form of numerical codes.



- First serious attempts to build talking systems where initiated in the XVIII and XIX centuries. First automata able to imitate human behaviour were constructed.

Spoken Dialog Systems

... Brief historical review ...

- Baron Von Kempelen – 1770
 - Talking heads
 - First automata to produce whole words and short phrases.
- Josef Faber – 1857
 - Imitation of human speech production
 - Euphonia. 16 keys to form any word



Spoken Dialog Systems

... Brief historical review ...

- At the end of XIX century, scientists (e.g. Von Helmholtz) concluded that speech could be produced electrically instead of mechanically.

This was the theory, what about the practice??

- JQ Stewart in the beginning of XX century built a machine that could generate vocalic sounds electrically.

Spoken Dialog Systems

... Brief historical review ...

- **30s**

- First electric machines, for example:

- VOCODER, speech analyzer and synthesizer developed in Bell Laboratories that could be operated by a keyboard.
- First translating machines are patented – France (automatic, simple) and Russia (advanced, human intervention).

- **40s**

- First computers were developed during WW2

- 1938 – 1941: Konrad Zuse, first programmable computer.
- 1943: Colossus, Bletchley Park.
- 1943 – 1945: Univ. Pennsylvania, ENIAC.

- The potential of these machines begins to be considered for other purposes:

- Alan Turing
- Warren Weaver



Spoken Dialog Systems

... Brief historical review ...

- **50s**

- 50 - Alan Turing – Intelligent Machines
- 51 – Cooper, Liberman and Borst – Spectrogram
- 52 – Machine translation conferences
- 57 – Chomsky – Syntactic Structures
- 58 – Peterson, Wang and Siversten – Diphones

- **60s**

- First conversational agents, e.g. Weizenbaum ELIZA
→ Not computational linguistics

- **70s**

- Computational linguistic
- Based in theoretical work about semantic and language carried out by Chomsky, Montague, Wood
- Rule based speech synthesis



Spoken Dialog Systems

... Brief historical review ...

- **80s**
 - Telephony services
 - Dictation systems
- **90s**
 - Corpora development: Wordnet
 - Isolated word recog. Telephony dialog systems
 - Dedicated companies.
- **2000s**
 - Statistical systems – Rules still in practical use
 - Standards for easy development.
- **2010s**
 - Usability, user centred systems.

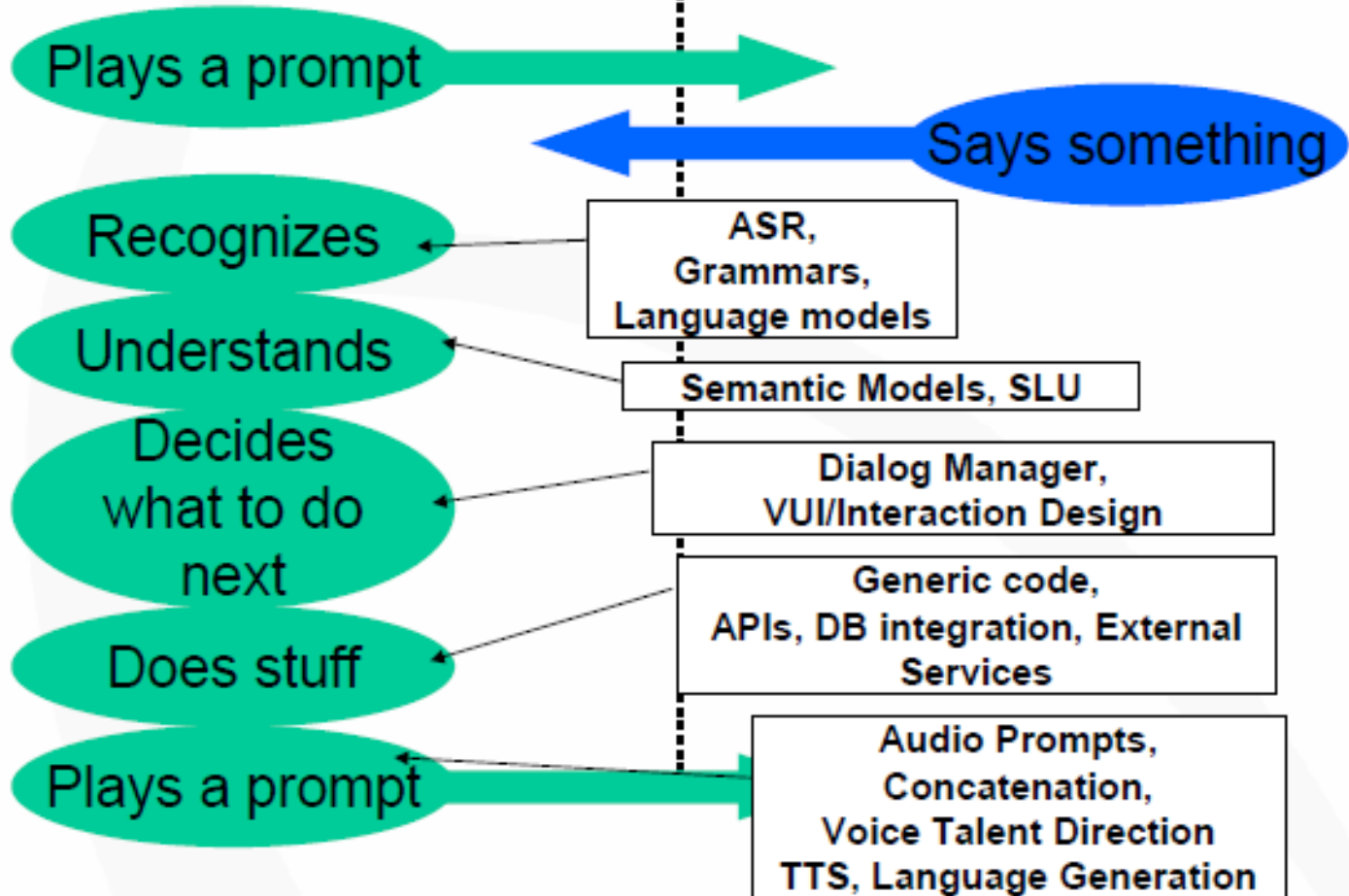


Spoken Dialog Systems

Architecture

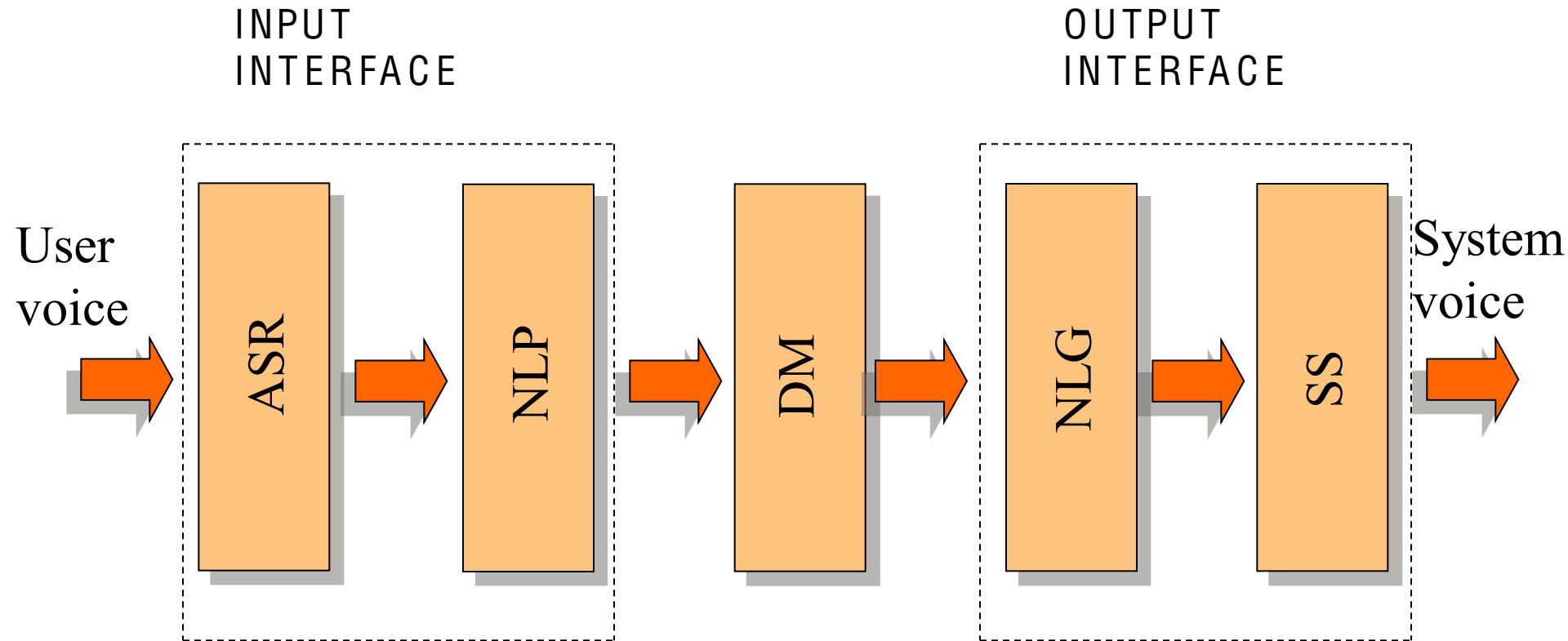
SYSTEM

USER



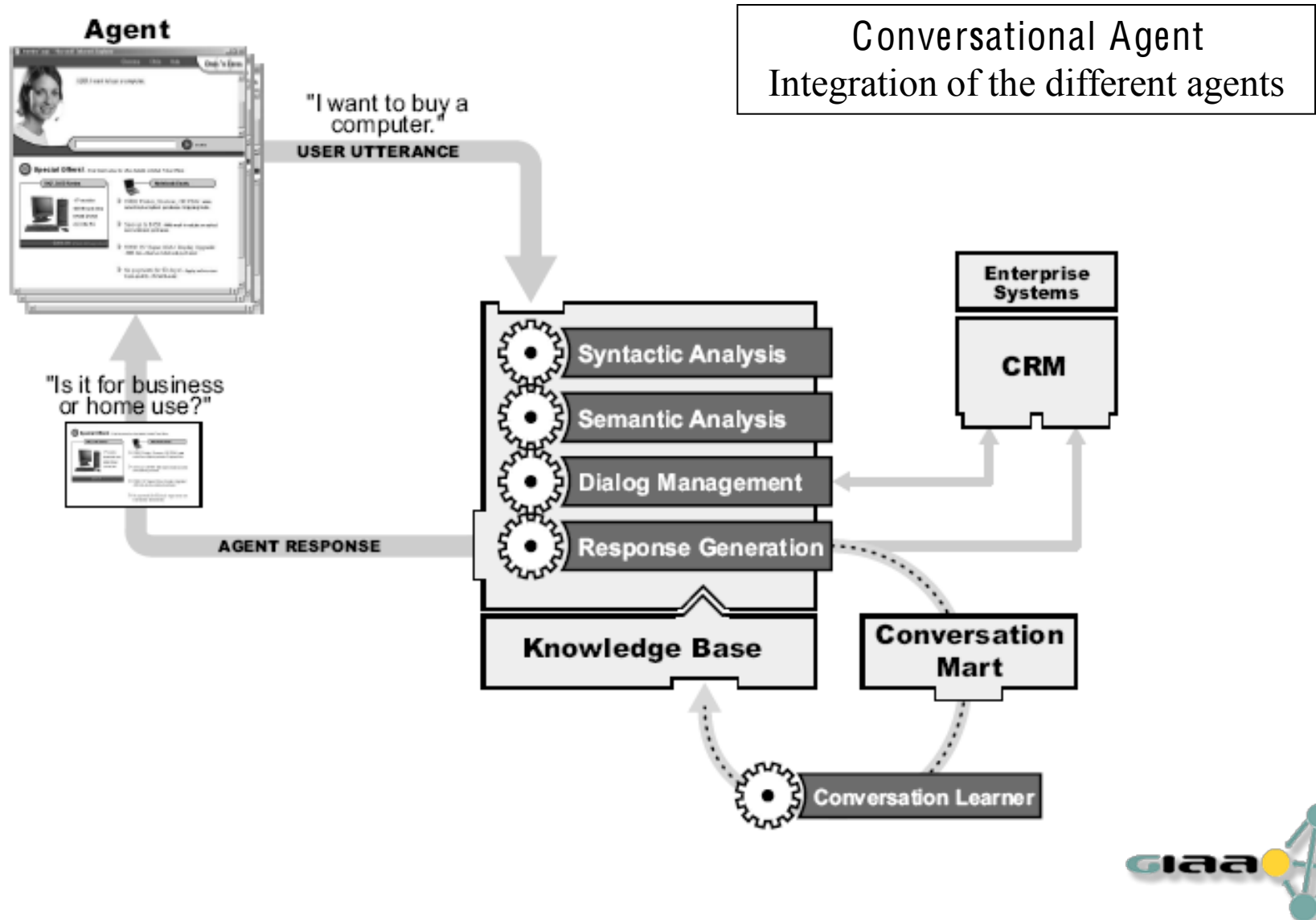
Spoken Dialog Systems

Architecture



Spoken Dialog Systems

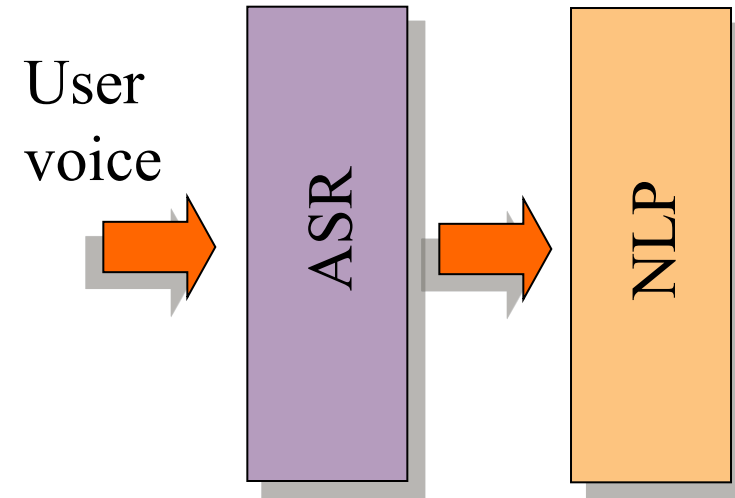
Architecture



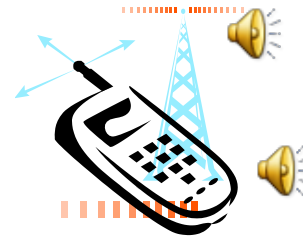
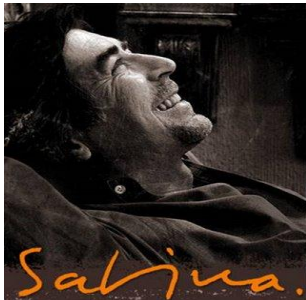
Spoken Dialog Systems

Automatic Speech Recognition

- Some ASR complexities:
 - Linguistic variability
 - Inter-speaker variability
 - Intra-speaker variability
 - Channel variability



- Maybe you could add your own to the list??



Spoken Dialog Systems

Automatic Speech Recognition

- **Statistical Approach: Markov Hidden Models**
 - Objective: To find the sequence of words W given a sequence of acoustic data A :

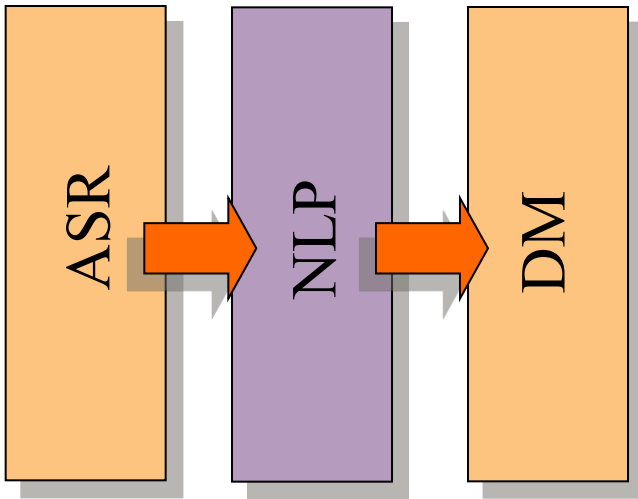
$$P(W|A) = \frac{P(A|W)P(W)}{P(A)}$$

The diagram shows the equation $P(W|A) = \frac{P(A|W)P(W)}{P(A)}$. An orange arrow points from the term $P(A|W)$ in the numerator to a box labeled "Acoustic Model". Another orange arrow points from the term $P(W)$ in the numerator to a box labeled "Language Model".

- Training data!
- ASR systems: HTK, Sphinx, Loquendo...

Spoken Dialog Systems

Natural Language Processing



Levels of NLP:

- Morphological
- Lexical
- Syntactical
- Semantic
- Discourse
- Pragmatical

In dialog systems:

Obtains the meaning of the recognized user utterance

In a wider sense:

Understanding human natural language

It has many applications, among them:

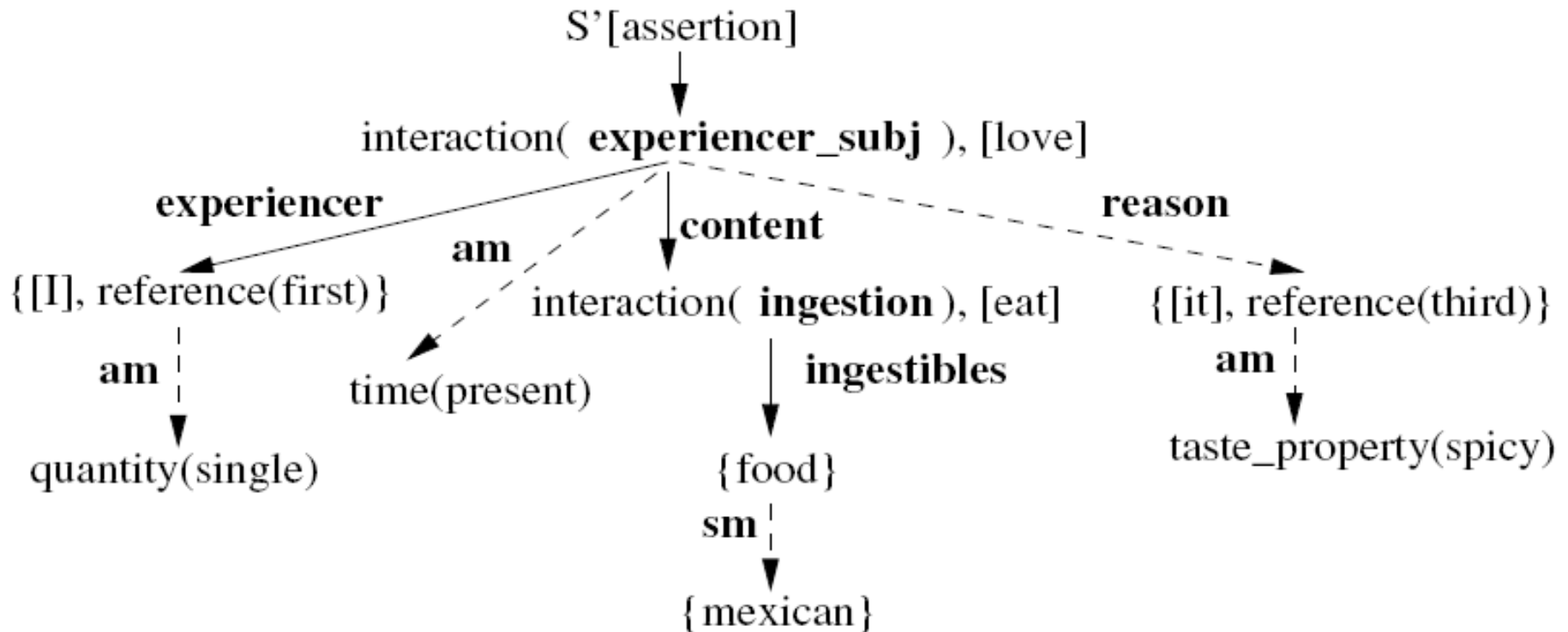
- Machine translation
- Question answering, text summarization, text simplification
- Information classification, filtering and recovering, e.g. spam filters, orthographic correctors
- Dialog systems 😊

Spoken Dialog Systems

Natural Language Processing

- The goal of semantic analysis is to represent the meaning of the sentences in a logical language

I love to eat Mexican food, because it is spicy.



Spoken Dialog Systems

Natural Language Processing

- Incremental Interpretation
 - The discourse is a sequence of sentences that can be structured in a way or another to produce different meanings.
 - Rhetoric
- Reference resolution
 - Anaphora
 - Hyponyms
 - Pronouns
 - NNs (p.e. What was previously commented).
 - Other expressions (p.e. he was feeding the pigs).
- World knowledge needed
 - Suppositions (e.g. I came in my bike. The wheel makes a strange noise).
 - Implications (e.g. Some kids don't like vegetables).

Spoken Dialog Systems

Natural Language Processing

NLP - Ambiguity

- **Complexity of natural language is in its ambiguity. Natural language is ambiguous at various levels:**
 - Lexical
 - Same word might have different grammatical categories.
 - Syntactic
 - Phrases with more than one associated syntactic structure.
 - Semantic
 - Same word different meanings.
 - Same syntactic structure, different meanings.
 - Discourse
 - Referential
 - Pragmatic
 - Irony, sarcasm, double sense...



Spoken Dialog Systems

Natural Language Processing

- Oral natural language has peculiarities that make it even more difficult to process:
 - **LEXICAL**
 - Several users talking at same time → Separation
 - Homophonic words
 - Colloquial form of words
 - Fragments, hesitations, repetitions
 - Quasilexical (e.g. aha)
 - **MORPHOLOGICAL**
 - Hesitations
 - Word classes that only appear in oral language, like for example discourse markers (e.g. well)
 - Interjections
 - Multi-word expressions (e.g. I see)



Spoken Dialog Systems

Natural Language Processing

- Syntactic
 - Incomplete structures
 - Are you ok?
 - Yes, but I still feel a bit...
 - Worried?
 - Yes
 - Reparations
 - On Monday, on Tuesday, I went to the beach
 - Repetitions
 - My sister, my sister is very friendly
 - Syntactic combinations
 - The capital of Spain is ... how was it?... Madrid



Spoken Dialog Systems

Natural Language Processing

- **Possible implementations**
 - **Rule-based methodologies.**
 - Extract the semantic information from a syntax-semantic analysis of the sentences:
 - Specific grammars defined for the task.
 - Detection of single/sequences of key words with semantic meaning.
 - **Statistical methodologies.**
 - Definition of linguistic units with semantic content and construction of a statistical model from a corpus of labeled samples.
 - Training the model: Capture correspondences between input text and its corresponding semantic interpretation.
 - Training data!!



Spoken Dialog Systems

Natural Language Processing

EDECÁN project

Semantics of the task:

- USERS: Semantic interpretation using **frames** (concepts and attributes).
 - Task-dependent concepts: *Availability, Booking, Booked and Cancellation.*
 - Task-independent concepts: *Acceptance, Rejection and Not-Understood.*
 - Attribute: *Sport, Hour, Date, Court-Type, Court-Number and Order-Number.*

Yes, I want to book a tennis court for tomorrow afternoon.

(Affirmation)

(Booking)

Sport: tennis

Date: tomorrow

Hour: afternoon

Spoken Dialog Systems

Natural Language Processing

EDECÁN project

Semantics of the task:

- SYSTEM: Semantic representation using frames.
 - Concepts: *Availability, Booking, Booked, Cancellation, Sport, Date, Hour, Court-Type, Confirmation-Availability, Confirmation-Booking, Confirmation-Booked, Confirmation-Cancellation, Confirmation-Sport, Confirmation-Date, Confirmation-Hour, Confirmation-CourtType, Rule-Info, Booking-Choice.*
 - Atributtes: *Sport, Hour, Date, Court-Type, Court-Number, Order-Number y Availability-Number.*

Do you want to book the squash court number 1 for February 2nd?

(Confirmation-Booking)

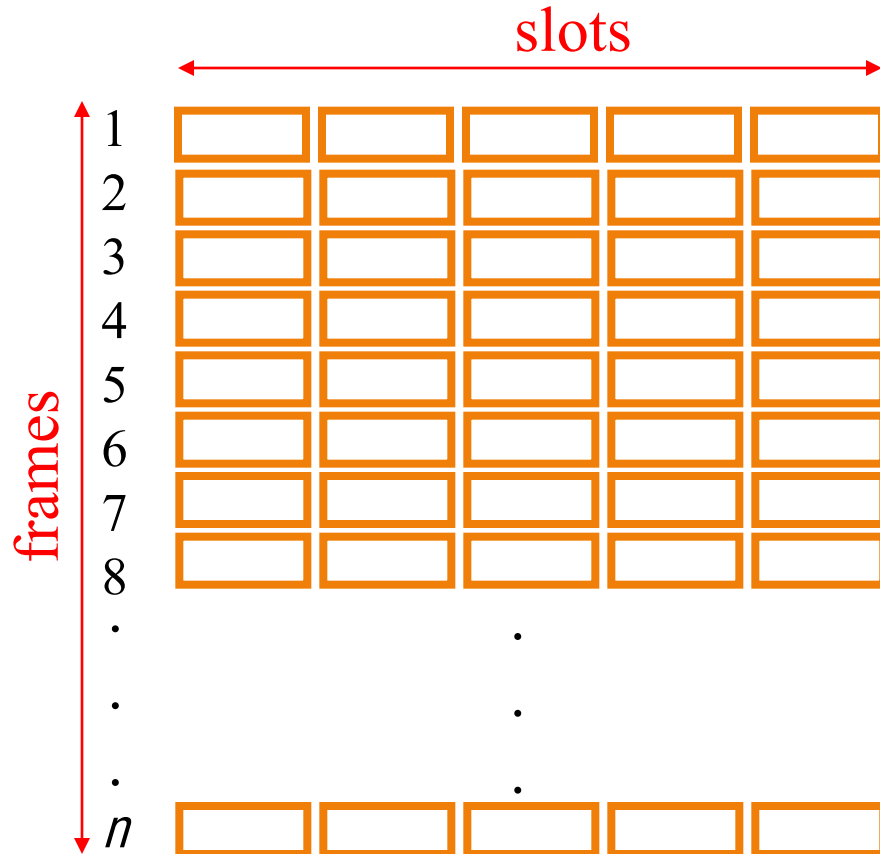
Sport: squash

Date: 02-02-2012

Court-Number:1

Spoken Dialog Systems

Natural Language Processing



Process of unification of frames: to combine several frames to generate a new frame

Spoken Dialog Systems

(Viajero system)

S: Welcome to the system...

U: I want to go to Madrid

S: Tell me the origin city

U: From Valencia

Unification
frames 1 and 2

S: Do you want a direct bus?

U: Yes

Unification
frames 3 and 4

S: There are four buses from Valencia to Madrid at 9, 10, 11 and 12.

Tell me the departure hour

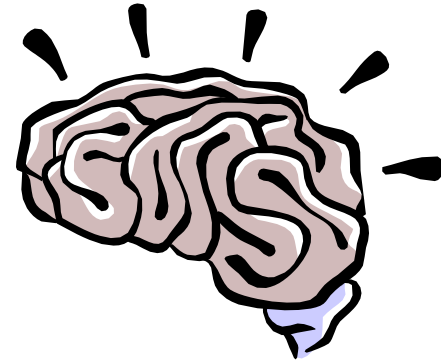
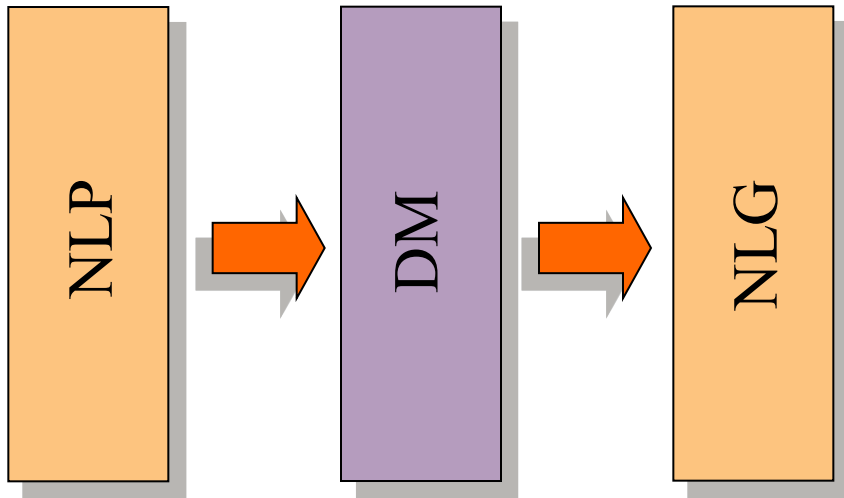
U: At 12

Unification
frames 5 and 6

#	FR	STAT	ORG	DEST	TYPE	HOUR
1		<input checked="" type="checkbox"/>	<input type="checkbox"/>	Madrid	<input type="checkbox"/>	<input type="checkbox"/>
2		<input checked="" type="checkbox"/>	Valencia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3		<input checked="" type="checkbox"/>	Valencia	Madrid	<input type="checkbox"/>	<input type="checkbox"/>
4		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Direct	<input type="checkbox"/>
5		<input checked="" type="checkbox"/>	Valencia	Madrid	Direct	<input type="checkbox"/>
6		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	12:00
7		<input type="checkbox"/>	Valencia	Madrid	Direct	12:00

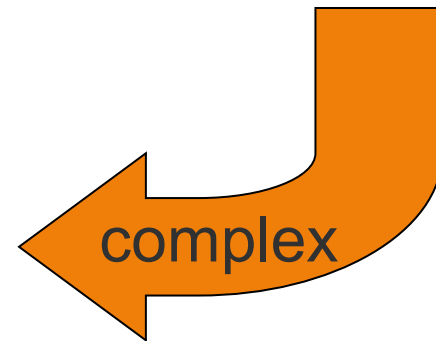
Spoken Dialog Systems

Dialog management



Intelligence – Decision - Coordination

- Deals with different sources of information
 - NLP results, database query results
 - Application domain – World knowledge
 - Knowledge about users and their intentions
 - Dialog history
- Depends on:
 - Task
 - Flexibility
 - Initiative



Spoken Dialog Systems

Dialog management

- There is not an agreement about what the task of a dialog manager are. Here there is a short list:
 - Decide the next system turn and which information to provide.
 - What information to provide? How?
 - Where to get the information? Manage multiple sources
 - How is the decision taken? From last user turn? From whole dialog history?
 - Manage interaction
 - Who has the initiative in the dialog?
 - Manage errors in recognition or understanding modules
 - Is the input to the manager correct?
 - If not, how to correct it?
 - Manage confirmations
 - Which pieces of information are necessary to confirm?
 - How?
 - Keep track of system history
 - Which information is necessary to store?
 - How?



Spoken Dialog Systems

Dialog management

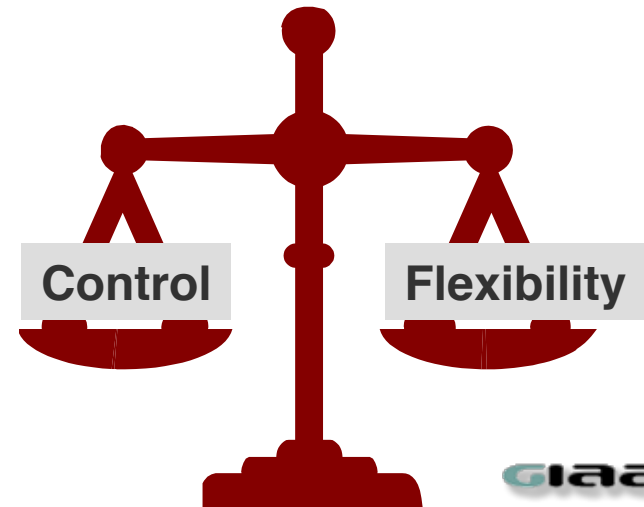
- **Concepts involved:**
 - **Initiative** - The interlocutor who controls the dialog at a certain point has the *initiative* at this point and decides on what to talk about next.
 - **Confirmation** - System way of checking whether the last piece of information gathered from the user is correct or not.
 - **Prediction** - Expectation on what the user will say next. → User modelling.
 - **Dialog history** - Log of information that has already been exchanged throughout the dialog. It can be saved in form of dialog acts list.



Spoken Dialog Systems

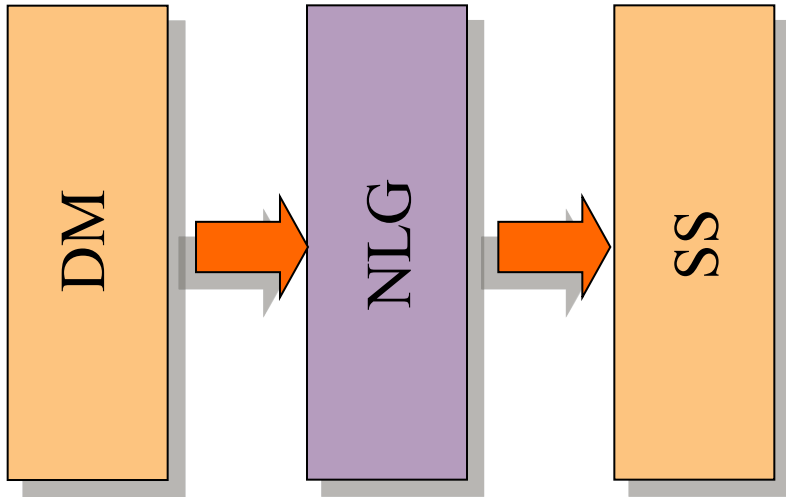
Dialog management

- Initiative can be:
 - System directed
 - User directed (pioneer: How may I help you? AT&T)
 - Mixed
 - Static – Always the same initiative type
 - Dynamic – Initiative changes during the interaction
- Confirmation strategies:
 - Implicit
 - Explicit



Spoken Dialog Systems

Natural language generation



Generates natural language from a machine representation of the content to be conveyed

Id-message: 000
Relation: SHOWING
Arguments:

Film: VÉRTIGO
Room: 2
Session: 18:30

5 basic steps:

- Content organization → Discourse planning
- Content distribution in sentences
- Lexicalization
- Referential expressions generation
- Linguistic realization

Spoken Dialog Systems

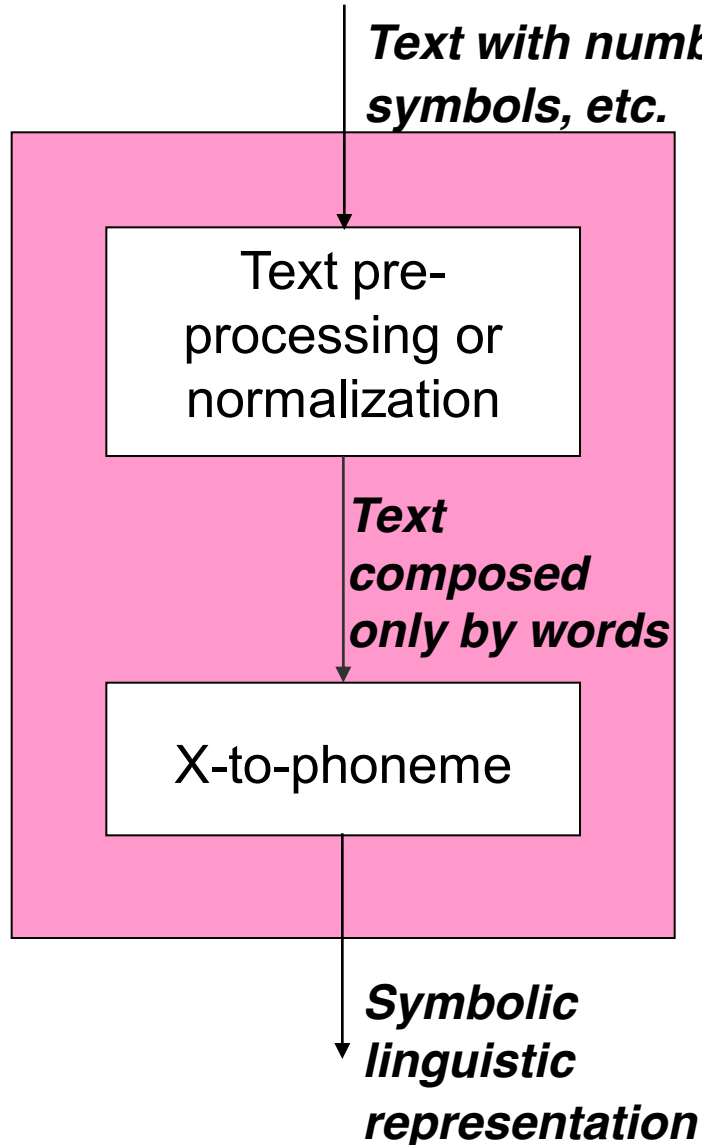
Natural language generation

- Discourse planning:
 - Structure and organize messages to be generated.
 - It can have several levels of complexity:
 - Organization based on planning – Artificial intelligence, preconditions and effects.
 - Organization based on templates – Small number of choices that can be generated instantiating templates.
 - Conjunction
 - Ellipsis
 - Building of word sets
 - Complex sentences



Spoken Dialog Systems

Natural language generation

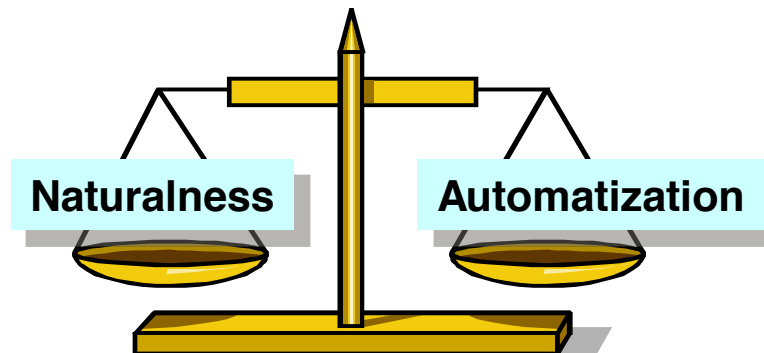


- Transforms the input text into a symbolic representation.
 1. Transforms numbers, abbreviations and symbols into words.
 2. Assigns phonetic transcriptions to each word and distinguishes prosodic units like phrases and sentences.
 - Grapheme-to-phoneme – GTP
 - Text-to-phoneme - TTP

Spoken Dialog Systems

Natural language generation

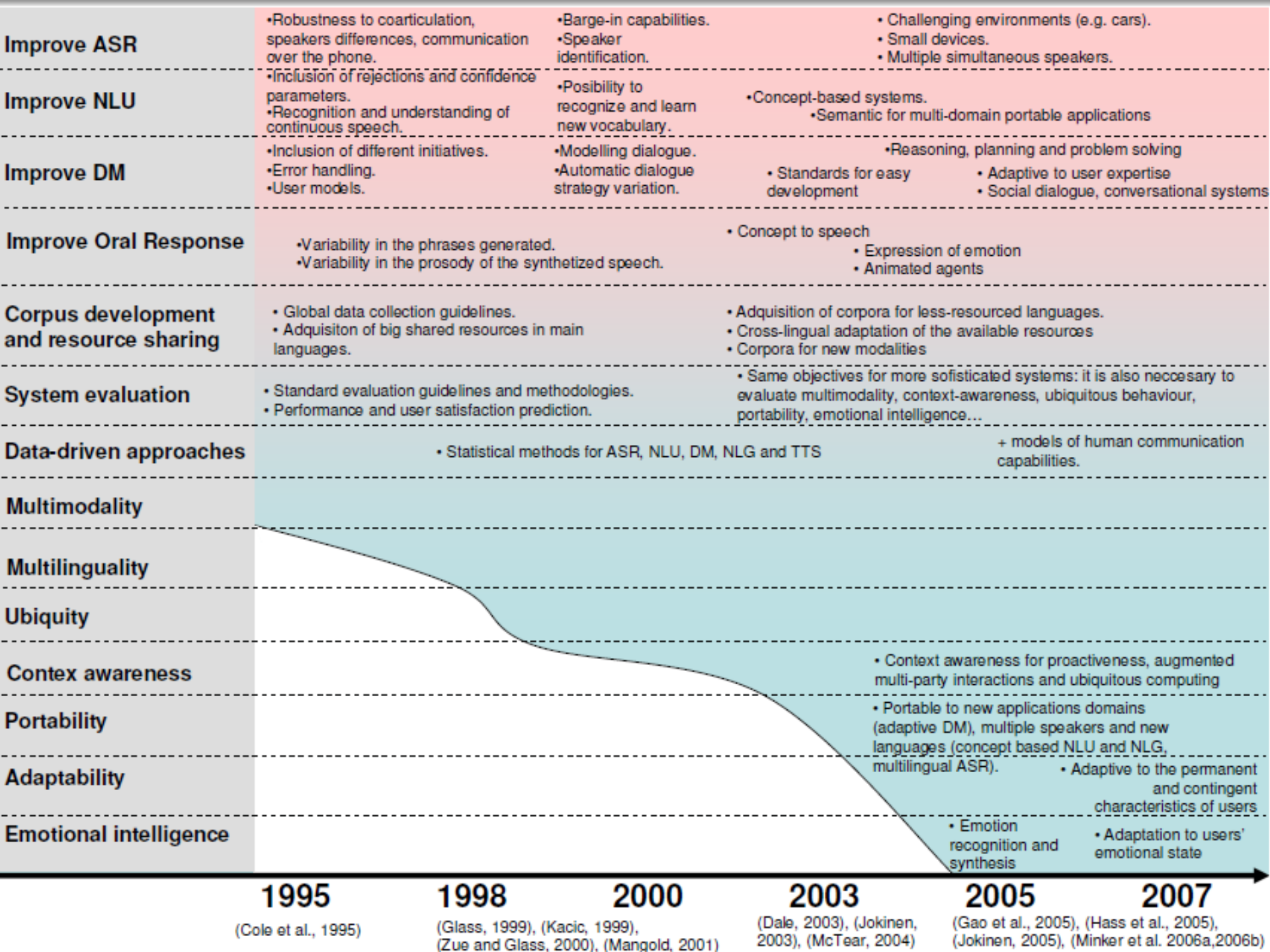
- Different techniques of Speech Synthesis:
 - Based in human speech production
 - Parametric – Model several physiological parameters about vocal tract to produce sounds.
 - Based in formants – Formants model vibration of vocal chords.
 - Without physiological base
 - Concatenative synthesis → Join pre-recorded units of human voice → Unit size??



Outline

1. Introduction
2. Spoken Dialog Systems
- 3. Recent Trends: Adaptation**
 - a) Statistical dialog management
 - b) User Modeling
 - c) Emotion Recognition
 - d) Context-aware system
 - e) Multimodal Interaction
4. Application domains
5. Conclusions





Recent Trends

Recent Trends

- Proactivity
- AmI environments
- Automatic resolution of problems
- Optimal dialog strategies
- Context-aware systems
- Emotional intelligence
- Portability
- Multimodal, multilanguage, multidomain systems
- Recommending systems
- Comparative evaluation
- Time and effort minimization

ADAPTATION

Outline

1. Introduction
2. Spoken Dialog Systems
3. **Recent Trends: Adaptation**
 - a) **Statistical dialog management**
 - b) User Modeling
 - c) Emotion Recognition
 - d) Context-aware system
 - e) Multimodal Interaction
4. Application domains
5. Conclusions



Recent Trends

Dialog Management

- Design of dialog system:
 - Specification of the system actions in response to user utterances and environmental states
 - Dialog strategy that defines the conversational behavior of the system.



- Great effort is employed to empirically design dialog strategies for commercial systems (not a trivial task)

Recent Trends

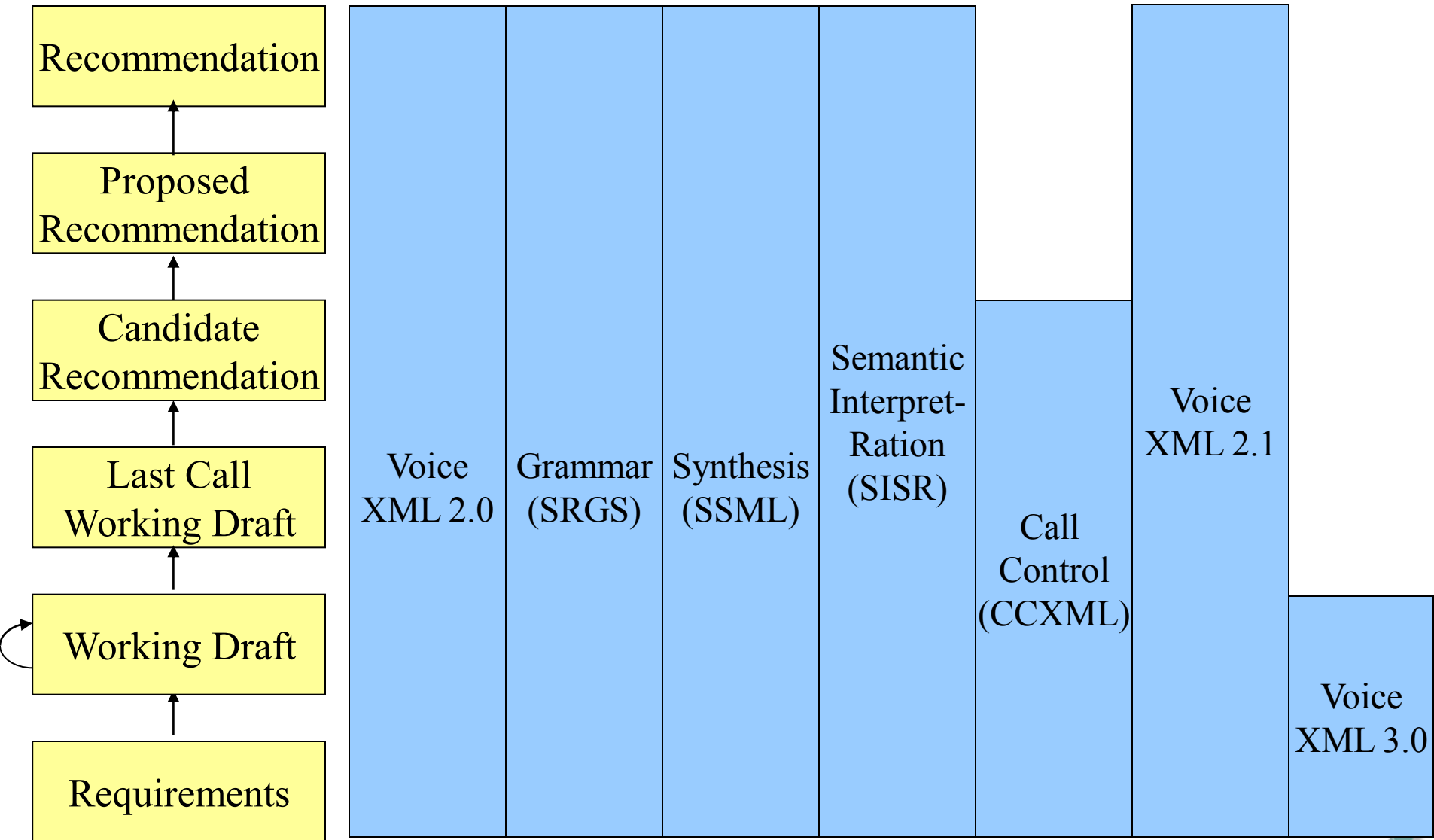
Dialog Management

- **VoiceXML (VXML)**

- W3C's standard XML format for specifying interactive voice dialogs between a human and a computer.
- It allows voice applications to be developed and deployed in an analogous way to HTML for visual applications.
- Just as HTML documents are interpreted by a visual web browser, VoiceXML documents are interpreted by a voice browser.
- Evolution:
 - **v1.0 (2000)**
 - **v2.0 (2004)**
 - **v2.1 (2007)** <http://www.w3.org/TR/voicexml20/>
 - **V3.0 (2010)** <http://www.w3.org/TR/voicexml21/>



Recent Trends



Recent Trends

Dialog Management

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/vxml
    http://www.w3.org/TR/voicexml20/vxml.xsd">
<form id="weather_info">
  <block>Welcome to the weather information service.</block>
  <field name="state">
    <prompt>What state?</prompt>
    <grammar src="state.grxml" type="application/srgs+xml"/>
    <catch event="help">
      Please speak the state for which you want the weather.
    </catch>
  </field>
  <field name="city">
    <prompt>What city?</prompt>
    <grammar src="city.grxml" type="application/srgs+xml"/>
    <catch event="help">
      Please speak the city for which you want the weather.
    </catch>
  </field>
  <block>
    <submit next="/servlet/weather" namelist="city state"/>
  </block>
</form>
</vxml>
```

C (computer): Welcome to the weather information service. What state?
H (human): Help
C: Please speak the state for which you want the weather.
H: Georgia
C: What city?
H: Tblisi
C: I did not understand what you said. What city?
H: Macon
C: The conditions in Macon Georgia are sunny and clear at 11 AM ...



Recent Trends

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml xmlns="http://www.w3.org/2001/vxml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/vxml
    http://www.w3.org/TR/voicexml20/vxml.xsd"
  version="2.0">
  <var name="bye" expr="'Ciao'"/>
  <link next="operator_xfer.vxml">
    <grammar type="application/srgs+xml" root="root" version="1.0">
      <rule id="root" scope="public">operator</rule>
    </grammar>
  </link>
</vxml>
```

ROOT
app-

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml xmlns="http://www.w3.org/2001/vxml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/vxml
    http://www.w3.org/TR/voicexml20/vxml.xsd"
  version="2.0" application="app-root.vxml">
<form id="say_goodbye">
  <field name="answer">
    <grammar type="application/srgs+xml" src="/grammars/boolean.grxml"/>
    <prompt>Shall we say <value expr="application.bye"/>?</prompt>
    <filled>
      <if cond="answer">
        <exit/>
      </if>
    </filled>
  </field>
</form>
```

LEAF
leaf.vxml

A
P
P
L
I
C
A
T
I
O
N



Recent Trends

Dialog Management

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/vxml
  http://www.w3.org/TR/voicexml20/vxml.xsd">
<menu>
  <prompt>
    Welcome home. Say one of: <enumerate/>
  </prompt>
  <choice next="http://www.sports.example.com/vxml/start.vxml">
    Sports
  </choice>
  <choice next="http://www.weather.example.com/intro.vxml">
    Weather
  </choice>
  <choice next="http://www.stargazer.example.com/voice/astronews.vxml">
    Stargazer astrophysics news
  </choice>
  <noinput>Please say one of <enumerate/></noinput>
</menu>
</vxml>
```

C: Welcome home. Say one of: sports; weather; Stargazer astrophysics news.

H: Astrology.

C: I did not understand what you said. (a platform-specific default message.)

C: Welcome home. Say one of: sports; weather; Stargazer astrophysics news.

H: sports.

C: (proceeds to <http://www.sports.example.com/vxml/start.vxml>)



Recent Trends

Dialog Management

```
<grammar mode="voice" xml:lang="en-US" version="1.0" root="command">
  <!-- Command is an action on an object -->
  <!-- e.g. "open a window" -->
  <rule id="command" scope="public">
    <ruleref uri="#action"/> <ruleref uri="#object"/>
  </rule>

  <rule id="action">
    <one-of>
      <item> open </item>
      <item> close </item>
      <item> delete </item>
      <item> move </item>
    </one-of>
  </rule>

  <rule id="object">
    <item repeat="0-1">
      <one-of> <item> the </item> <item> a </item> </one-of>
    </item>
    <one-of>
      <item> window </item>
      <item> file </item>
      <item> menu </item>
    </one-of>
  </rule>
</grammar>
```

GRAMMAR
SRGS - XML



Recent Trends

Dialog Management

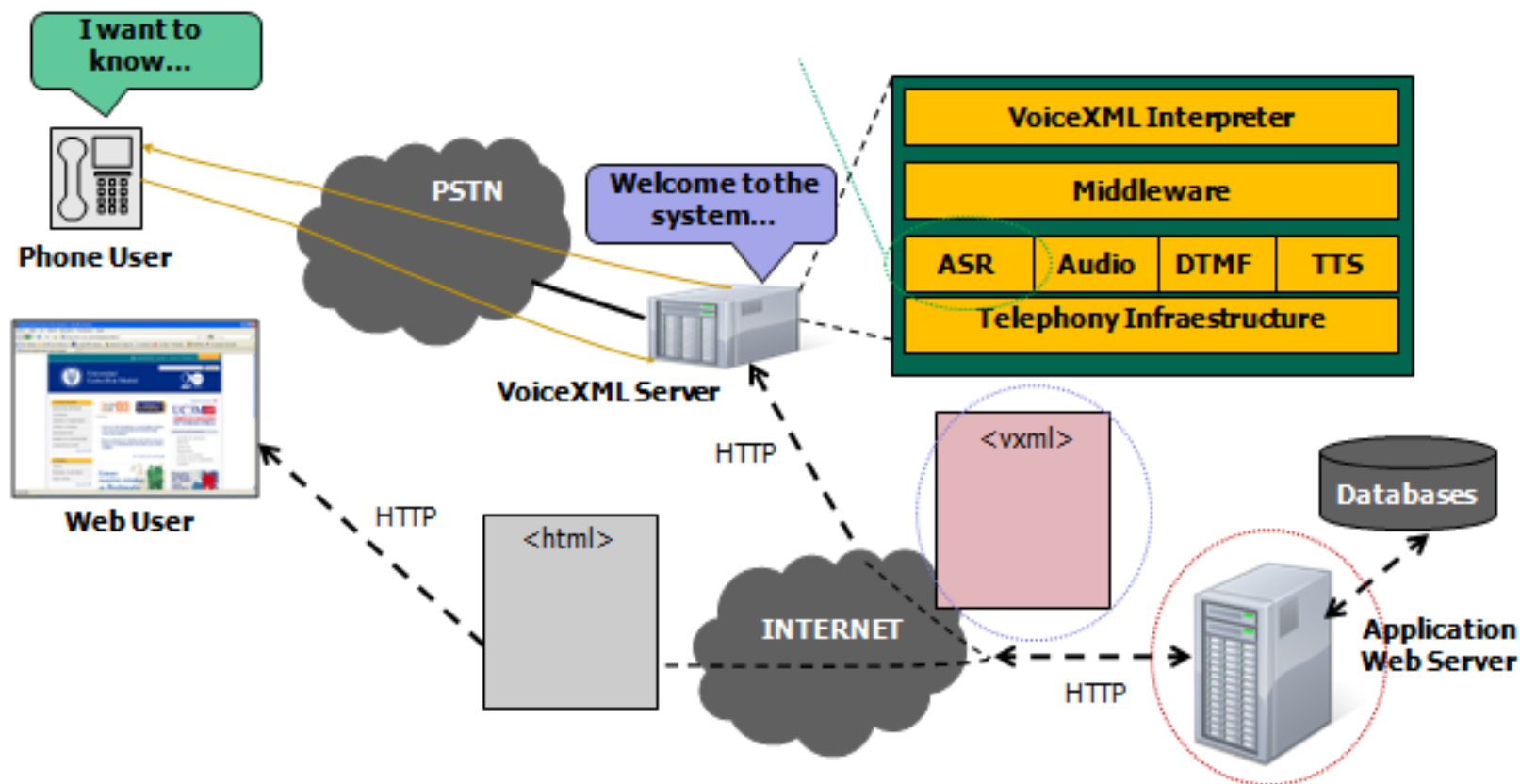
```
<grammar mode="voice" type="application/srgs">  
#ABNF 1.0;  
language en-US;  
mode voice;  
root $command;  
  public $command = $action $object;  
  $action = open | close | delete | move;  
  $object = [the | a] (window | file | menu);  
</grammar>
```

GRAMMAR
SRGS -



Recent Trends

Dialog Management



VoiceXML architecture for the design of dialog systems

Recent Trends

Dialog Management

- **VOXEO COMMUNITY** (<http://community.voxeo.com>)
- **HeyAnita RapidApp** (<http://freespeech.heyanita.com>)
- **Tellme Debug Log** (<http://studio.tellme.com>)
- **Voice Genie Call Log Explorer** (<http://developer.voicegenie.com>)
- **BeVocal Vocal Scripter** (<http://cafe.bevocal.com>)
- **Apple** (<http://www.apple.com/macos/speech>)
- **AT&T** (<http://www.att.com/aspg>)
- **BBN** (<http://www.bbn.com/departments/dept-slp.html>)
- **IBM** (<http://ibm.com/software/speech>)
- **Lernout & Hauspie** (<http://www.lhs.com>)
- **Microsoft** (<http://www.microsoft/speech>)
- **Nuance** (<http://www.nuance.com>)
- **Philips** (<http://www.speech.be.philips.com>)
- **Speechworks** (<http://www.speechworks.com>)



Recent Trends

Dialog Management

- Once the dialog strategy has been designed, the implementation of the system can be carried out using the standard VoiceXML... but how to design this strategy?
 - Application of statistical approaches to dialog management



- OUR SOLUTION: Combine the flexibility of statistical dialog management with the facilities that VoiceXML offers,
- Introduce statistical methodologies for the development of commercial (and not strictly academic) conversational agents.

Recent Trends

Statistical Dialog Management

- The use of statistical models that are automatically learnt from data has provided very interesting results:
 - Speech recognition.
 - Language understanding.
 - **Dialog management.**
- The problem of coverage:
 - All the possible situations that can occur during a dialog must be modeled using only the training corpus → **Classification problem.**
 - Considerable effort is needed to acquire and label a dialog corpus → **Automatic dialog generation.**



Recent Trends

Statistical Dialog Management

- POMDP representation:

Formally, a POMDP is defined as a tuple $\{S, A, T, R, O, Z, \lambda, b_0\}$ where:

- S is a set of the agent states.
- A is a set of actions that the agent may take.
- T defines the transition probability $P(s'|s, a)$.
- R defines the immediate reward obtained from taking a particular action in a particular state $r(s, a)$.
- O is a set of possible observations that the agent can receive.
- Z defines the probability of a particular observation given the state and machine action $P(o'|s', a)$.
- λ is a geometric discount factor $0 \leq \lambda \leq 1$.
- b_0 is an initial belief state $b_0(s)$.



Recent Trends

Statistical Dialog Management

- POMDP representation:

- Belief space:

$$b'(s') = P(s' | o', a_m, b) = k \cdot P(o' | s', a_m) \sum_{s \in S} P(s' | a_m, s) b(s)$$

- Reward:

$$R = \sum_{t=0}^{\infty} \lambda^t R(b_t, a_{m,t}) = \sum_{t=0}^{\infty} \lambda^t \sum_{s \in S} b_t(s) r(s, a_{m,t})$$

Each action $a_{m,t}$ is determined by the policy $\pi(b_t)$

- Using Point-based Value Iteration or Q-learning, in combination with a user simulator:

$$\bar{a}_p = \operatorname{argmax}_{\bar{a}} \bar{Q}(a, p)$$

Recent Trends

Statistical Dialog Management

- Most slot-filling dialog systems: the dialog manager takes its decisions based only on the information provided by the user in the previous turns and its own model (**DIHANA task**).
- Formal description:
 - A_i : the system answer at time i .
 - U_i : semantic representation of the user turn at time i .

Dialog $\rightarrow (A_1, U_1), \dots, (A_i, U_i), \dots, (A_n, U_n)$



S_i : State of the dialog sequence at time i



Recent Trends

Statistical Dialog Management

- Formal description:
 - At time i , the objective of the dialog manager is to find the best system answer A_i
 - The selection is made by maximizing:

$$\hat{A}_i = \operatorname{argmax}_{A_i \in \mathcal{A}} P(A_i | S_1, \dots, S_{i-1})$$



All the possible system answers.

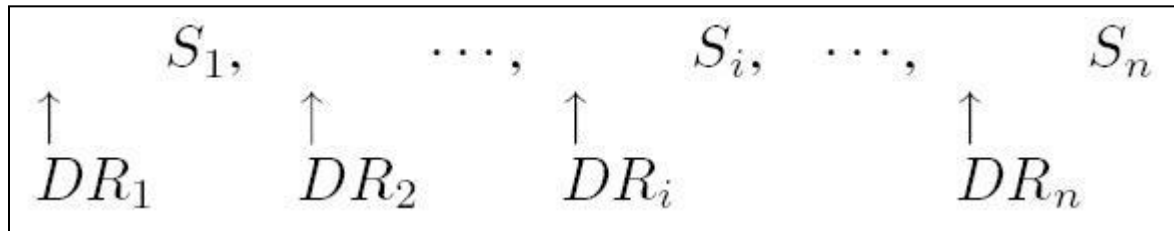
- We establish a partition in the space of sequences of states:
 - DR_i : Dialog register at time i (concepts and attributes).



Recent Trends

Statistical Dialog Management

- Formal description:
 - For a sequence of states of a dialog, there is a corresponding sequence of DR :



- Two different sequences of states are considered to be equivalent if they lead to the same DR_i
 - Great reduction in the number of different histories in the dialogs.
 - A loss in the chronological information.

Recent Trends

Statistical Dialog Management

- Formal description:
 - The selection of the best A_i is given by:

$$\hat{A}_i = \operatorname{argmax}_{A_i \in \mathcal{A}} P(A_i | DR_{i-1}, S_{i-1})$$

- Each user turn:
 - supplies the system with information about the task;
 - provides other kinds of information, such as task-independent information (*Affirmation*, *Negation*, and *Not-Understood* dialog acts).



Recent Trends

Statistical Dialog Management

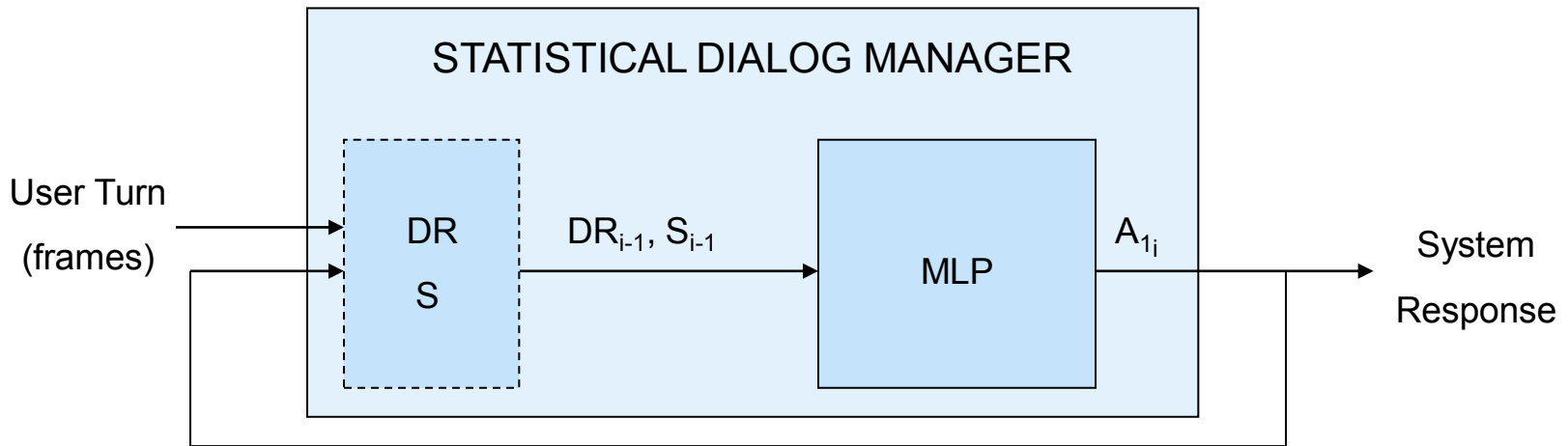
DIHANA project

- Dialog Register representation:
 - The DR is a sequence of 15 fields:
 - Five concepts: *Hour*, *Price*, *Train-Type*, *Trip-Time*, and *Services*.
 - Ten attributes: *Origin*, *Destination*, *Departure-Date*, *Arrival-Date*, *Departure-Hour*, *Arrival-Hour*, *Class*, *Train-Type*, *Order-Number*, and *Services*.
 - We have assumed that the exact values of the attributes are not significant to determine the next system answer:
 - **0**: The concept is not activated, or the value of the attribute is not given.
 - **1**: The concept or attribute is activated with a confidence score that is higher than a given threshold.
 - **2**: The concept or attribute is activated with a confidence score that is lower than the given threshold.
 - $DR = 15$ length string of elements from $\{0,1,2\}$.



Recent Trends

Statistical Dialog Management



Recent Trends

Statistical Dialog Management

- Example of a dialog:

System₁: Welcome to the railway information system. How can I help you?

A₁: Opening DR: 00000-1000001000

User₁: I want to go to Barcelona. U₁=()

DR: 00000-1100001000 + Opening + U₁ → A₂ = (Confirmation:Departure-Hour:Nil)

System₂: Do you want to know the timetables?

User₂: Yes, for the Euromed train. U₂=(Affirmation)

DR: 10000-1100001100 + Confirmation:Departure-Hour + U₂ →

A₃: (Question:Departure-Date:Nil)

System₃: Tell me the departure date.

Recent Trends

- Example of a dialog:

User₃: Tomorrow $U_3=()$

DR: 10000-1120001100 + Question:Departure-Date + $U_3 \rightarrow$

A_4 = Confirmation:Departure-Date

System₄: Do you want to leave tomorrow?

User₄: Yes $U_4=(\text{Affirmation})$

DR: 10000-1110001100 + Confirmation:Departure-Date + $U_4 \rightarrow$

A_5 = Answer:Departure-Hour:Number-Trains, Train-Type, Departure-Hour

New-Query

System₅: There are several Euromed trains. The first one leaves at 08:54 and the last one at 23:45. Anything else?

User₅: No, thank you $U_5=(\text{Negation})$

DR: 10000-1110001100 + Answer-New-Query:Departure-Nil + $U_5 \rightarrow$

A_6 = Closing

System₆: Thanks for using this service. Have a good trip.

Recent Trends

Statistical Dialog Management

- More complex dialog systems: the dialog manager selects the system turn taking into account:
 - the information provided by the user.
 - the information generated by the module that controls the application.
- Application manager (AM):
 - Performs the queries to the database.
 - Verifies if the user query follows the regulations defined for the task.

The result of queries to the *AM* has to be considered in order to generate the following system turn.



Recent Trends

Statistical Dialog Management

We have established two phases for the selection of the next system turn in this type of tasks:

- First phase: Select the best request to be made to the AM.

$$\hat{A}_{1_i} = \operatorname{argmax}_{A_{1_i} \in \mathcal{A}_1} P(A_i | DR_{i-1}, S_{i-1})$$

- Second phase: Generate the final system turn.

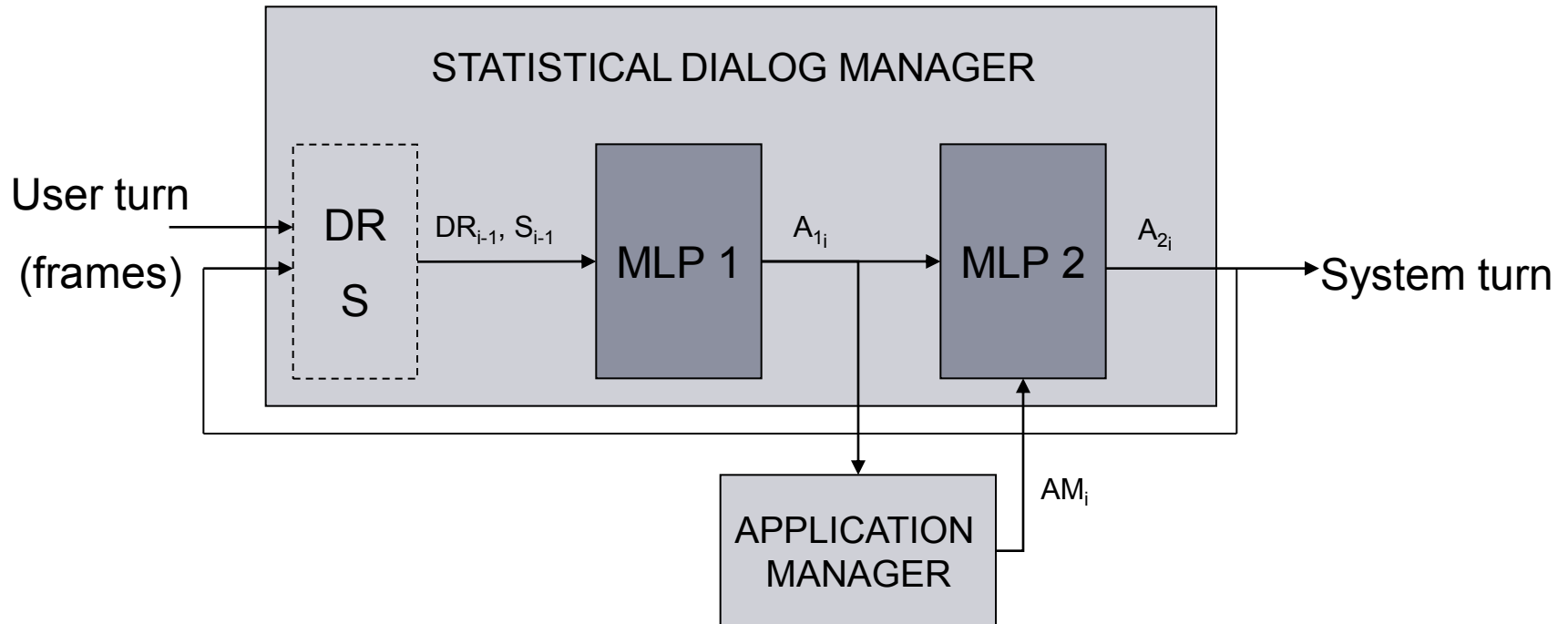
$$\hat{A}_{2_i} = \operatorname{argmax}_{A_{2_i} \in \mathcal{A}_2} P(A_i | AM_i, A_{1_i})$$

We propose the use of a multilayer perceptron (MLP) to obtain the system answer.



Recent Trends

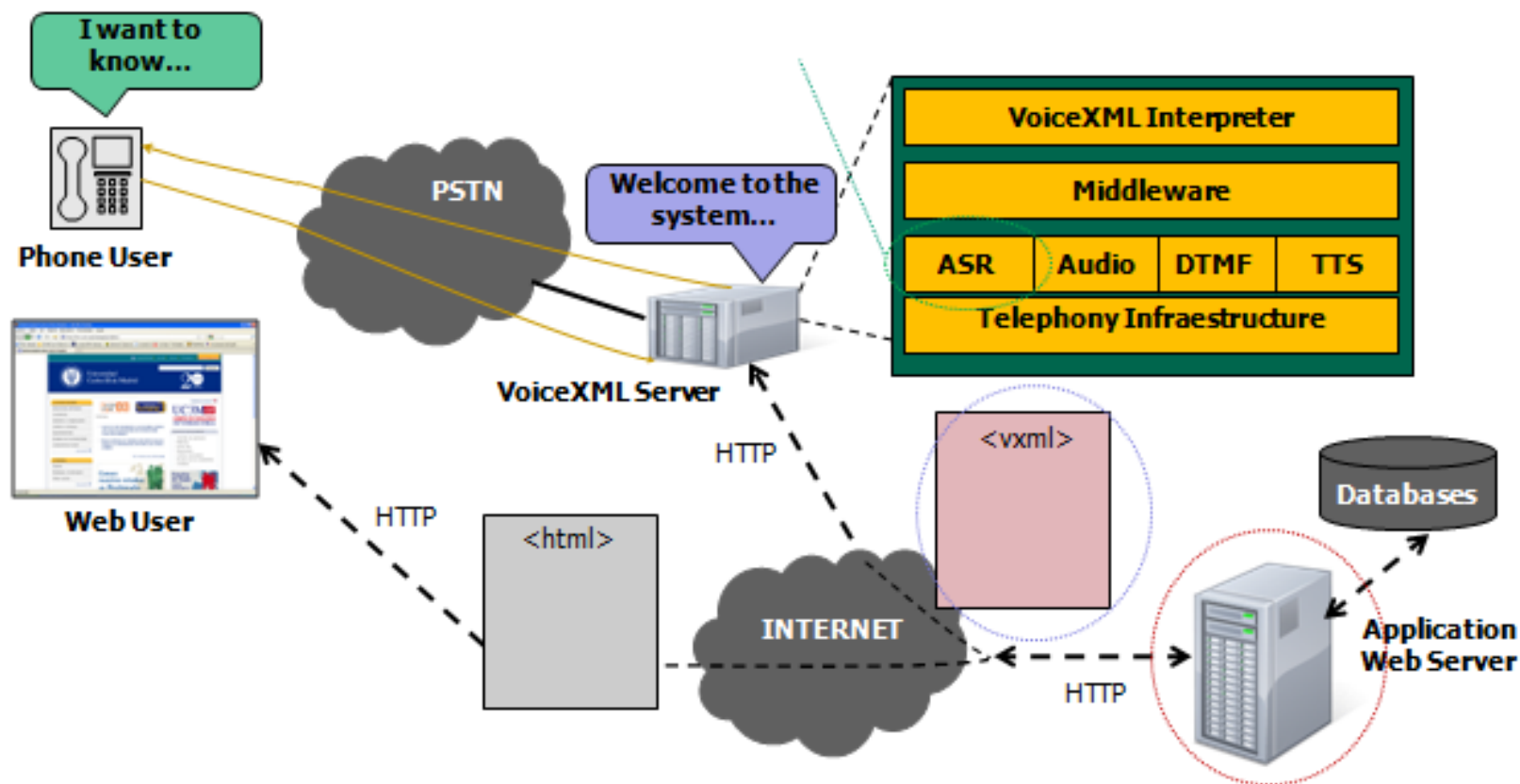
Statistical Dialog Management



- First phase (MLP 1): Select the best request to be made to the *AM*.
- Second phase (MLP 2): Generate the final system turn.

Recent Trends

Statistical Dialog Management



VoiceXML architecture for the design of dialog systems

Recent Trends

Statistical Dialog Management

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml xmlns="http://www.w3.org/2001/vxml"
  xmlns:xsi="http://www.w3.org/2001/
  XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/vxml
  http://www.w3.org/TR/voicexml20/vxml.xsd"
  version="2.0" application="app-trains.vxml">
<form id="origin_form">
  <field name="origin">
    <grammar type="application/srgs+xml"
      src="/grammars/origin.grxml"/>
    <prompt>Tell me the origin city.</prompt>
    <filled>
      <return namelist="origin"/>
    </filled>
  </field>
</form>
</vxml>
```

```
#JSGF V1.0;
grammar origin;
public <origin> = [<desire>]
[<travel> <city> {this.destination=$city}]
[<proceed> <city> {this.origin=$city}];
<desire> = I want [to know] | I would like
[to know] | I would like | I want | I need
| I have to;
<travel> = go to | travel to | to go to
| to travel to;
<city> = Murcia | Vigo | Sevilla | Huelva |
Cuenca | Lugo | Granada | Salamanca |
Valencia | Alicante | Albacete | Barcelona
| Madrid;
<proceed> = from | going from | go from;
```

VoiceXML documents to implement the system prompts



Outline

1. Introduction
2. Spoken Dialog Systems
3. **Recent Trends: Adaptation**
 - a) Statistical dialog management
 - b) **User Modeling**
 - c) Emotion Recognition
 - d) Context-aware system
 - e) Multimodal Interaction
4. Application domains
5. Conclusions



Recent Trends

To learn a Statistical Dialog Model



Training Dialogs



Dialog System



Dialog Manager



Statistical Dialog Model ???

User Modeling and Simulation

1) Wizard of Oz technique

- Pros: Acquisition under control conditions, not system fully developed.
- Cons: Difficult to train the wizard(s) and to recruit people.

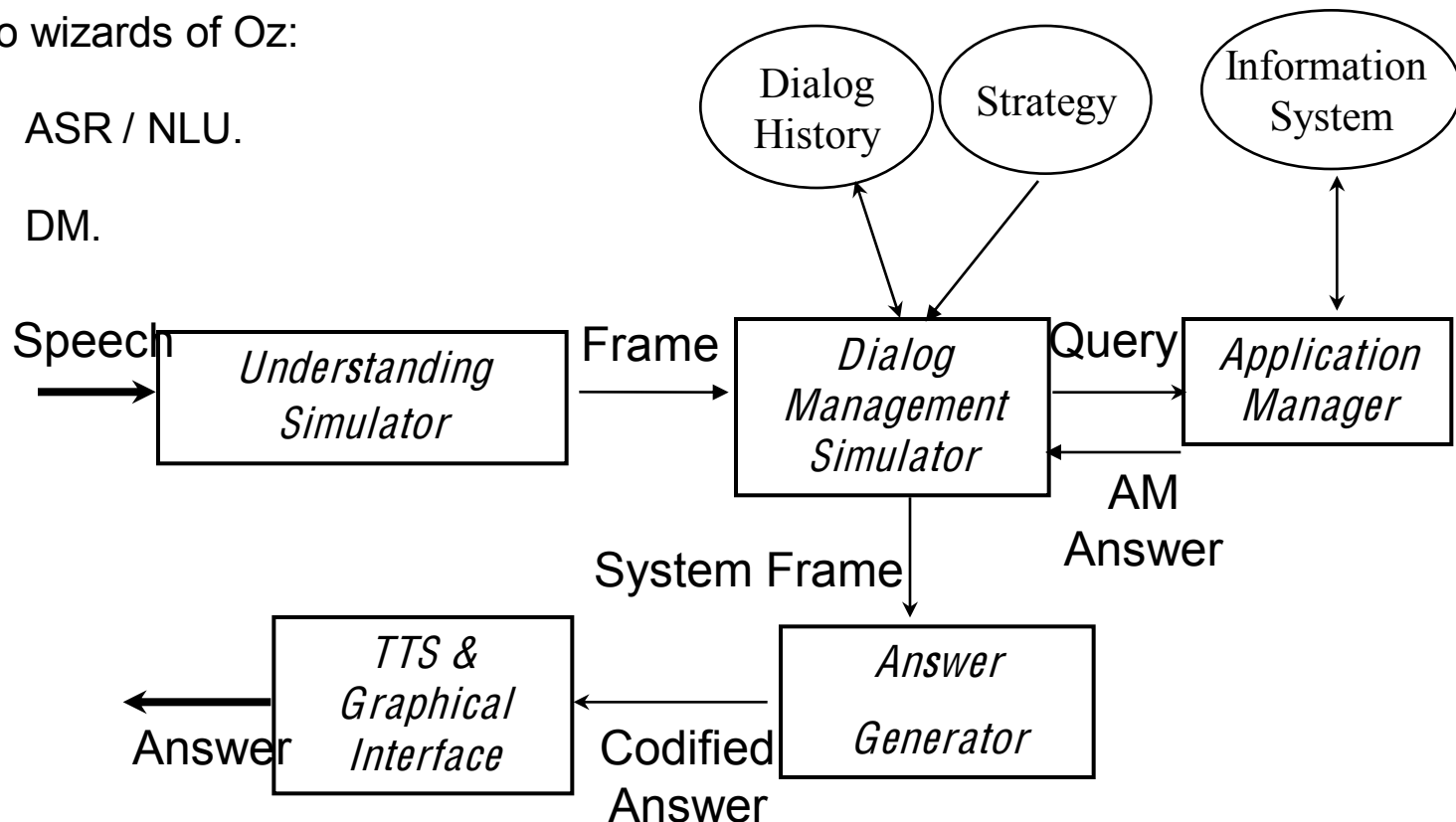
- DIHANA project:
- We acquired a corpus of 900 dialogs.
- We use a complete dialog system and the WOz simulates the role of the DM.
- If we want to develop a complete statistical DM:
 - To learn when a confirmation should be done → confidence measures.

Recent Trends

User Modeling and Simulation

EDECAN project:

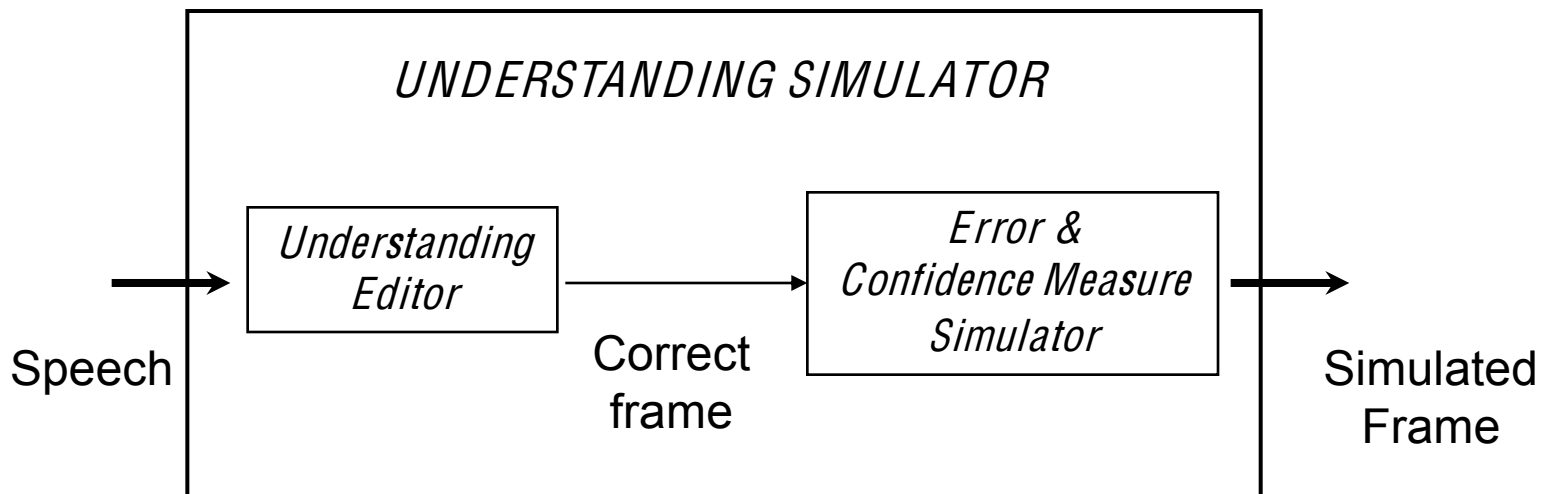
- Two wizards of Oz:
 - ASR / NLU.
 - DM.



Recent Trends

User Modeling and Simulation

1) First wizard:



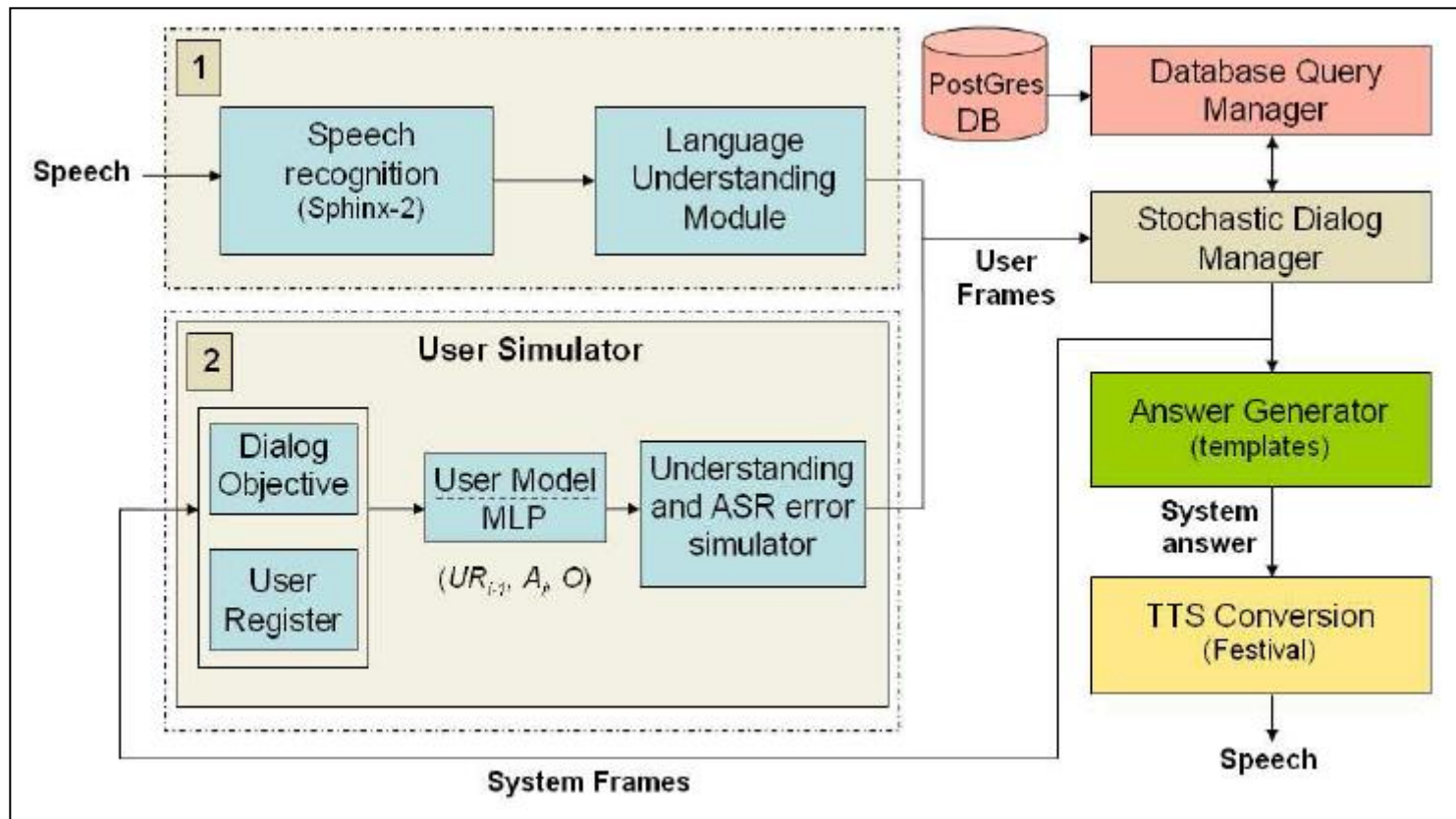
2) Second wizard:

- Only supervises the output generated by an initial dialog model.

Recent Trends

User Modeling and Simulation

2) Statistical user modeling



Recent Trends

User Modeling and Simulation

- Formal description:
 - If the most probable user answer U_i is selected at each time i , the selection is made using the following maximization:

$$\hat{U}_i = \operatorname{argmax}_{U_i \in \mathcal{U}} P(U_i | S_1, \dots, S_{i-1}, A_i, \mathcal{O})$$



All the possible user answers.

- We establish a partition in the space of sequences of states:

- UR_i : User register at time i (concepts and attributes).

Two different sequences of states are considered to be equivalent if they lead to the same UR_i



Recent Trends

User Modeling and Simulation

- Formal description:
 - The selection of the best U_i is given by:

$$\hat{U}_i = \operatorname{argmax}_{U_i \in \mathcal{U}} P(U_i | UR_{i-1}, A_i, \mathcal{O})$$

- We propose using a MLP to make the assignation of a user turn:
 - the input layer receives the current situation of the dialog.
 - the output layer can be viewed as the a posteriori probability of selecting the different user answers.



Recent Trends

User Modeling and Simulation

- User Register representation:
 - The UR is a sequence of 15 fields:
 - Five concepts: *Hour*, *Price*, *Train-Type*, *Trip-Time*, and *Services*.
 - Ten attributes: *Origin*, *Destination*, *Departure-Date*, *Arrival-Date*, *Departure-Hour*, *Arrival-Hour*, *Class*, *Train-Type*, *Order-Number*, and *Services*.
 - We have assumed that the exact values of the attributes are not significant to determine the next user answer:
 - **0**: The concept is not activated, or the value of the attribute is not given.
 - **1**: The concept or attribute is activated.
 - The error simulator perform error generation and the addition of confidence measures.



Recent Trends

User Modeling and Simulation

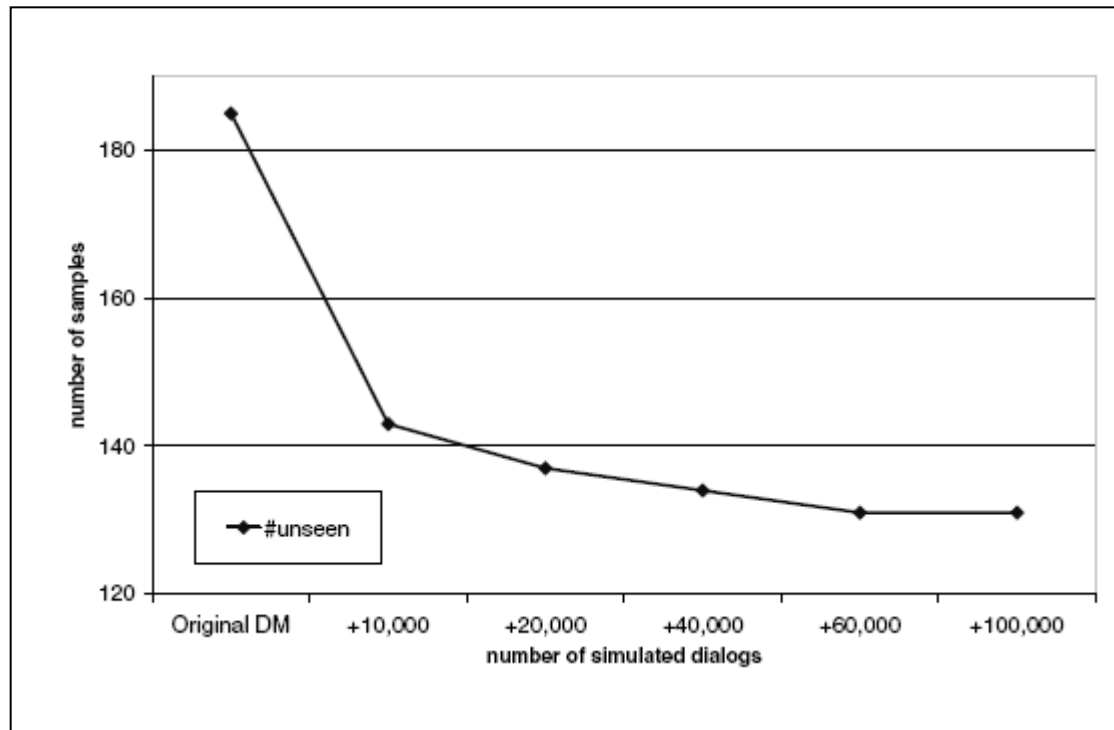
- First, we evaluated the behavior of the original DM that was learned using the training corpus (obtained by WOz).
- Then, we evaluated its evolution when the successful simulated dialogs were incorporated to the training corpus.
- We defined four measures to evaluate the evolution of the dialog manager:
 - #unseen: the number of unseen situations.
 - #error: the number of answers provided by the DM that would cause the failure of the dialog).
 - %strategy: the percentage of answers provided by the DM that exactly follow the strategy defined for the WOz to acquire the training corpus.
 - %coherent: percentage of answers provided by the DM that are coherent with the current state of the dialog.



Recent Trends

User Modeling and Simulation

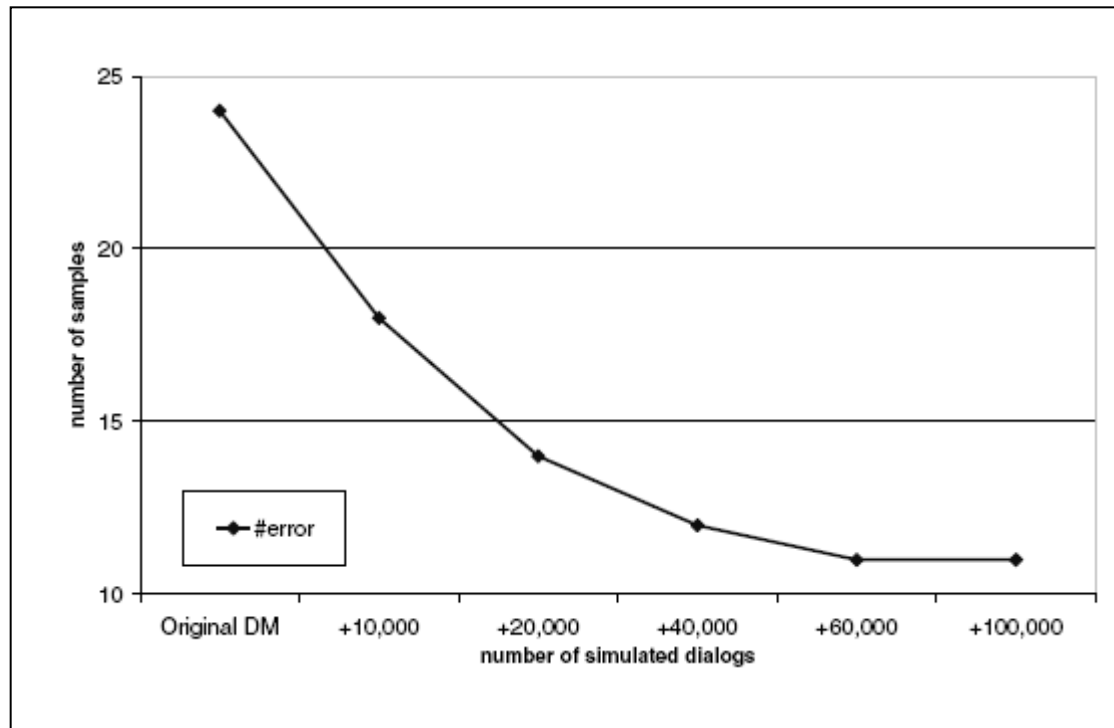
#unseen



Recent Trends

User Modeling and Simulation

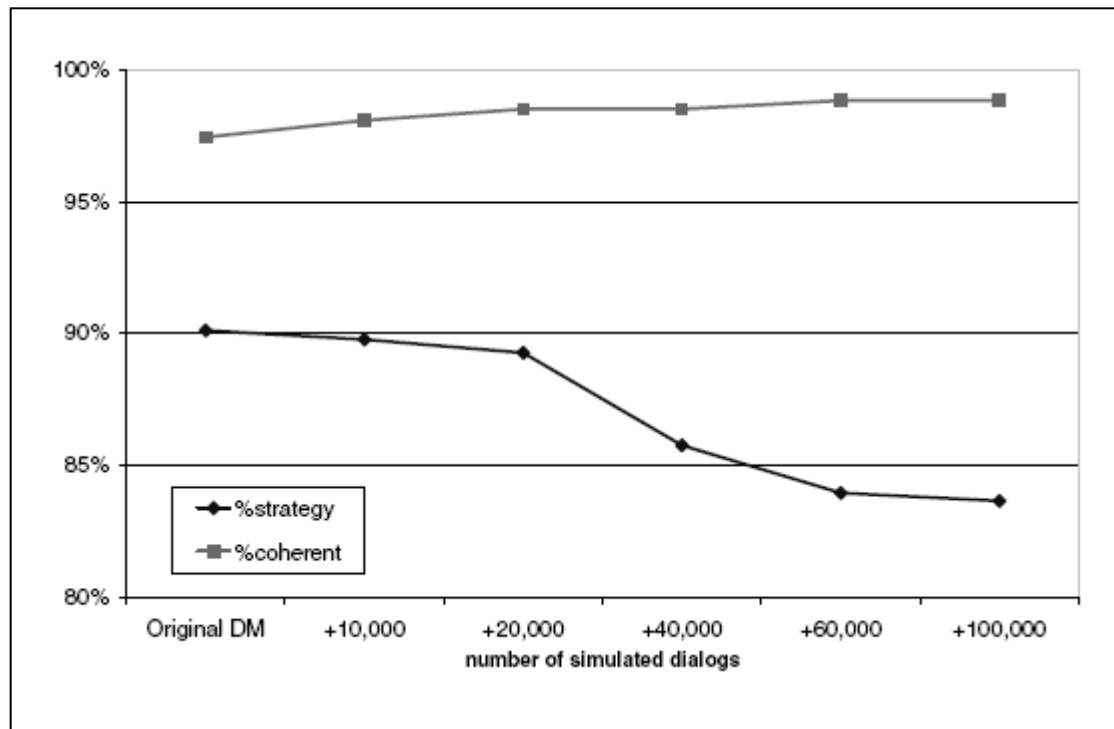
#error



Recent Trends

User Modeling and Simulation

%strategy and %coherent



Recent Trends

User Modeling and Simulation

3) Automatic Dialog Generation:

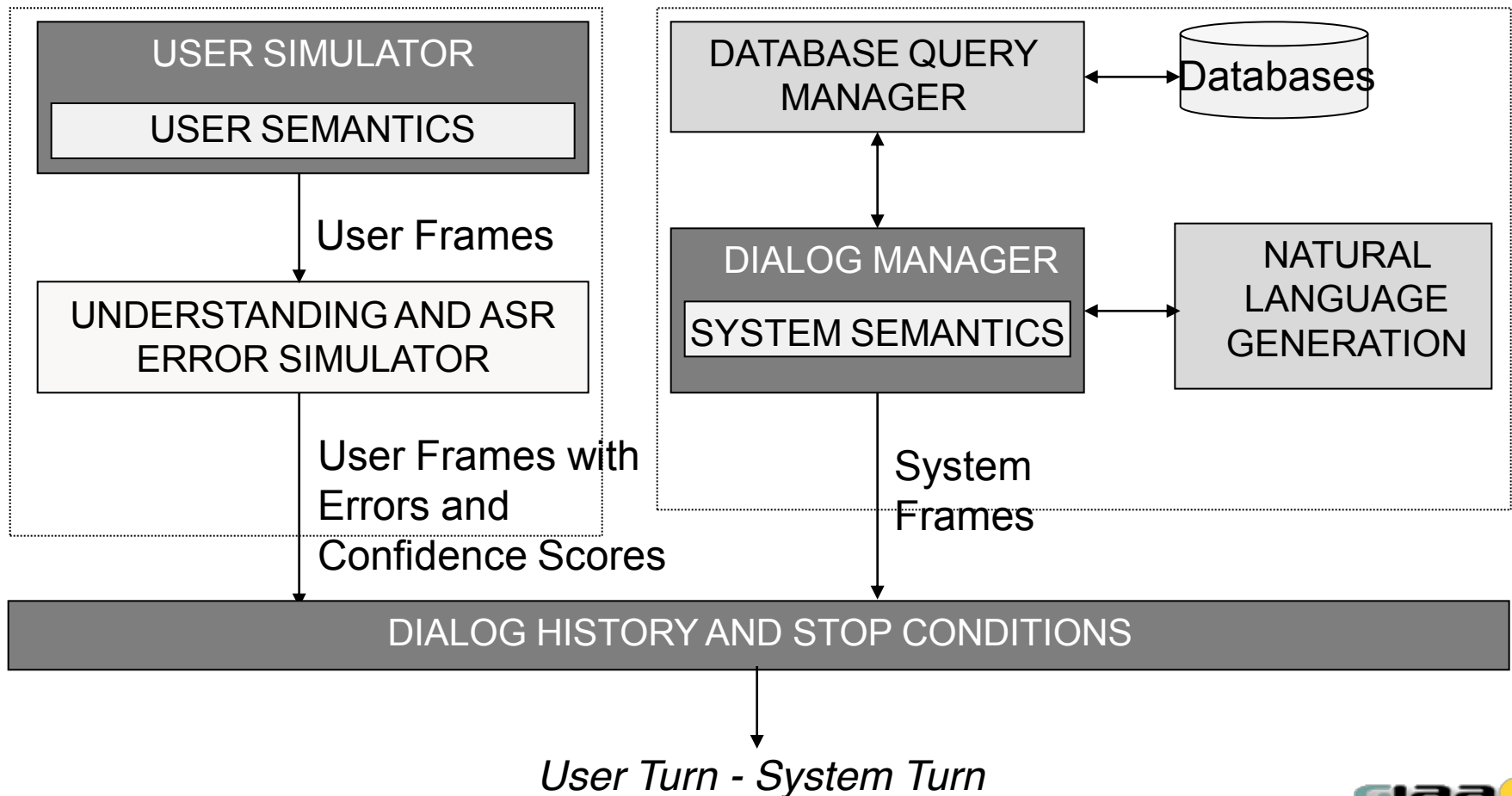
- Interaction of a user simulator and a dialog manager simulator.
- Initial model: Random selection of one of the possible answers defined for the semantics of the task (user and system dialog acts).
- Learning: the probabilities of the answers selected by the dialog manager during that dialog are incremented before beginning a new simulation.
- An error simulator module has been designed:
 - Performs error generation.
 - Addition of confidence measures.
- Evaluation: A set of stop conditions is defined to automatically evaluate if the dialog is successful or not.



Recent Trends

User Modeling and Simulation

- Automatic Acquisition of a Dialog Corpus:



Recent Trends

- Example of a dialog:

S1: (*Opening*)

Welcome to the sport service. How can I help you?

U1: (*Booking*) [0.9]

Sport: Tennis [0.9]

Date: 03-15-2011 [0.1]

Hour: 08.00-09.00 [0.9]

S2: (*Confirmation-Date*)

Do you want to play on the 15th March?

U2: (*Negation*) [0.9]

Date: 03-14-2011 [0.9]

S3: (*Confirmation-Booking*) Sport: Tennis Date: 2011-03-14 Hour: 08.00-09.00

{*One court available*} Do you want to book tennis court number 2?

U3: Yes.

S4: (*Booking*) Sport: Tennis Date: 2011-03-14 Hour: 08.00-09.00 (*New-Query*)

Tennis court number 2 has been booked. Anything else?

U4: (*Negation*) [0.9]

S5: (*Closing*)

Thank you for using the sport service. Goodbye.

Objective:

Booking

Sport: tennis

Date: tomorrow

Outline

1. Introduction
2. Spoken Dialog Systems
3. **Recent Trends: Adaptation**
 - a) Statistical dialog management
 - b) User Modeling
 - c) **Emotion Recognition**
 - d) Context-aware system
 - e) Multimodal Interaction
4. Application domains
5. Conclusions



Recent Trends

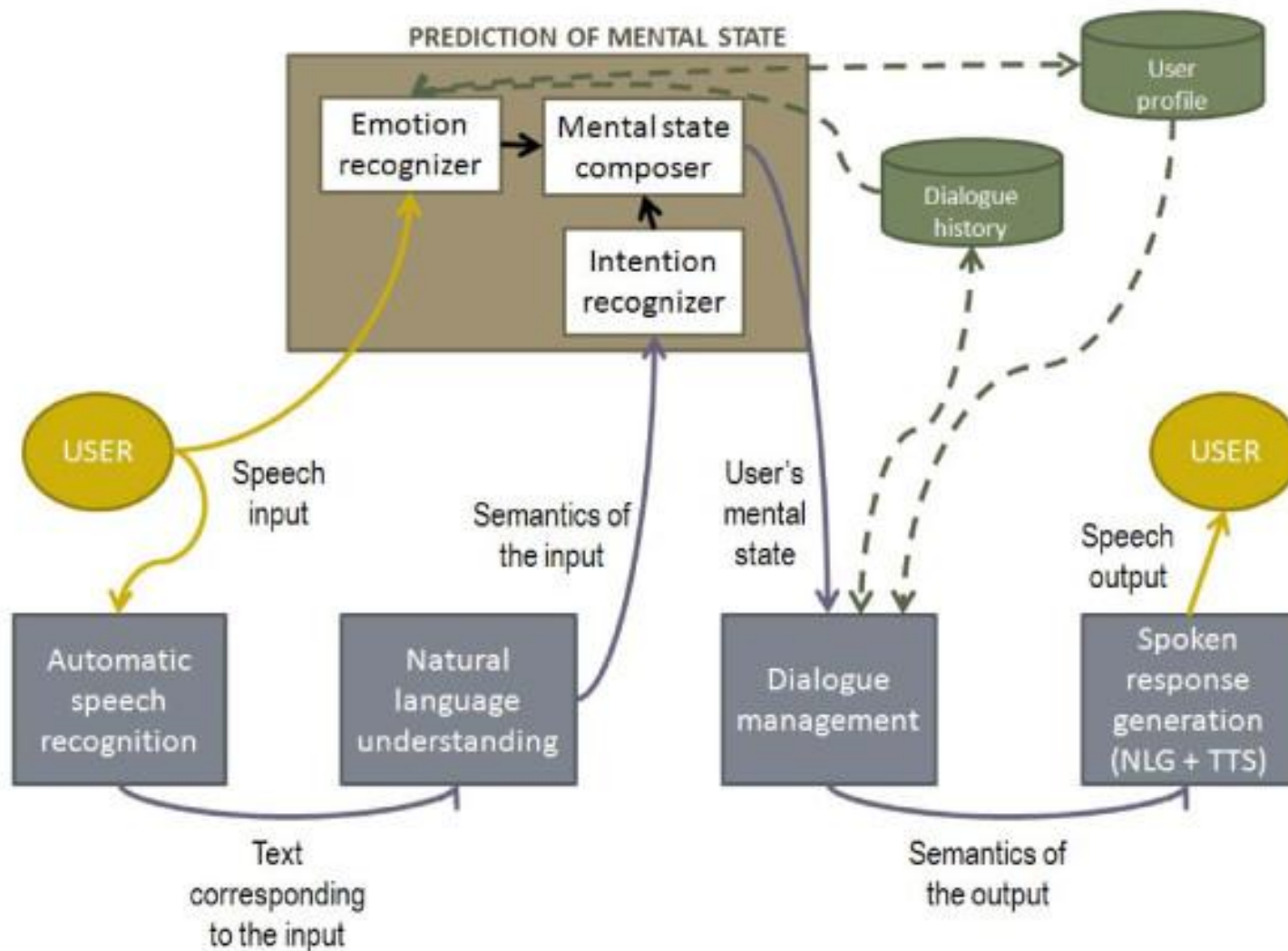
Emotion recognition

- **Affective Computing:**
 - System that are able to:
 - Detect users' emotional state.
 - Adapt the operation according to them.
 - Provide emotional responses:
 - Facilitate positive users' assessment.
 - Improve human-computer interaction.



Recent Trends

Emotion recognition



Recent Trends

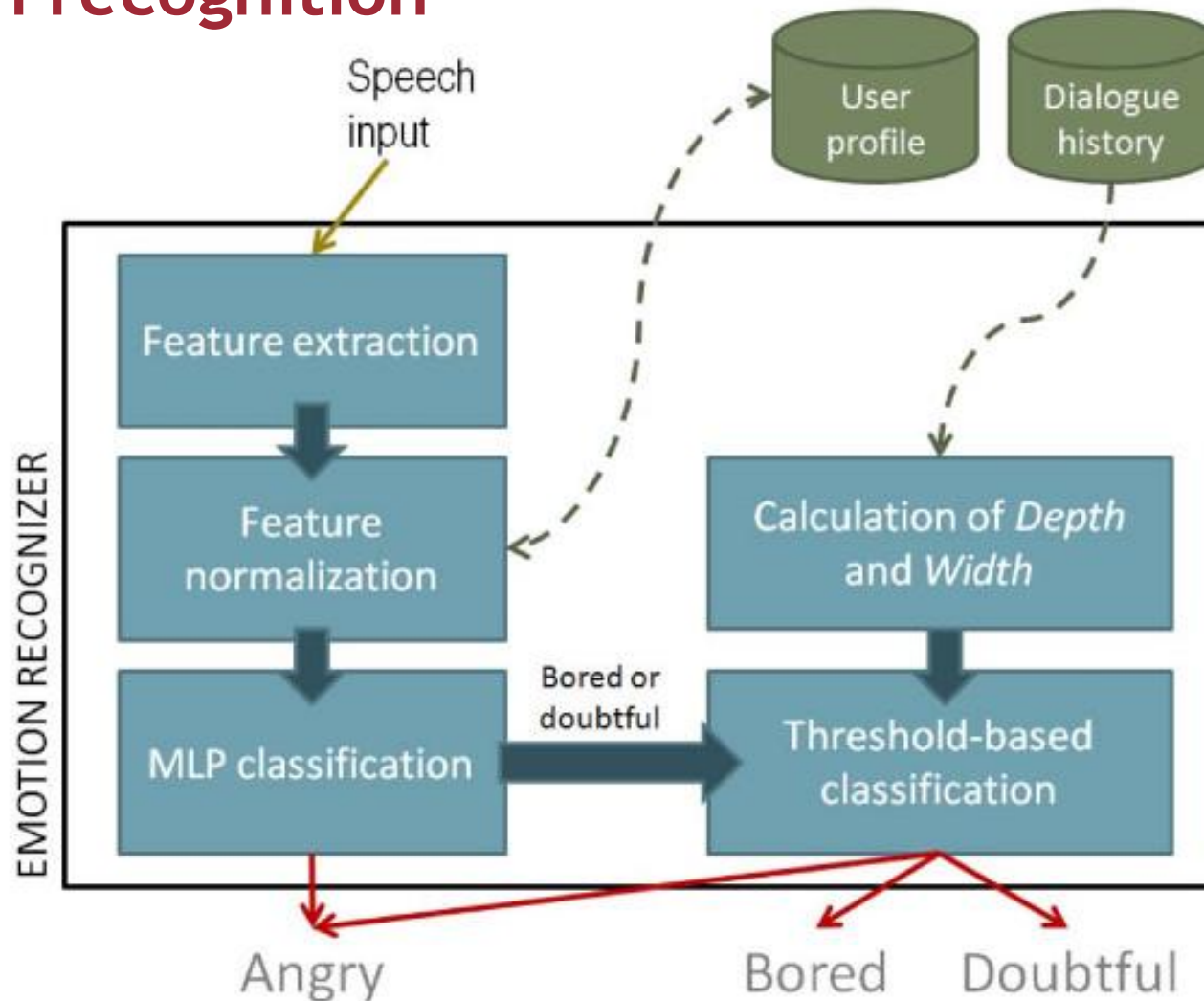
Emotion recognition

- **The emotion recognizer**
 - Based solely in acoustic and dialog information:
 - firstly takes **acoustic information** into account to distinguish between the emotions which are acoustically more different,
 - secondly **dialog information** to disambiguate between those that are more similar.
 - Recognizing negative emotions that might discourage users from employing the system again or even lead them to abort an ongoing dialog: *anger*, *boredom* and *doubtfulness*.
 - Our emotion recognizer employs acoustic information to distinguish *anger* from *doubtfulness* or *boredom* and dialog information to discriminate between *doubtfulness* and *boredom*.



Recent Trends

Emotion recognition



Recent Trends

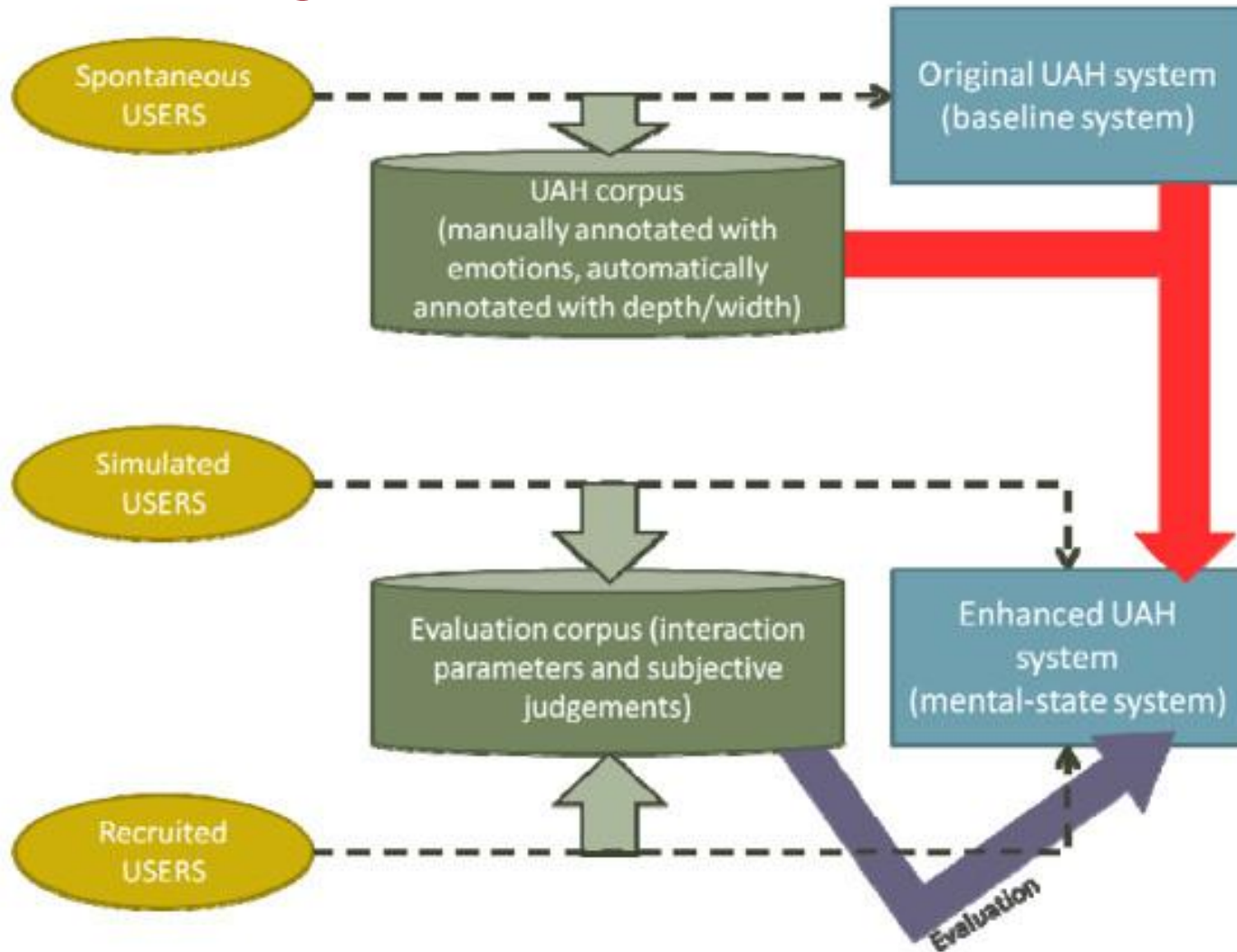
Emotion recognition

Features employed for emotion detection from the acoustic signal

Groups	Features	Physiological changes related to emotion
Pitch	Minimum value, maximum value, mean, median, standard deviation, value in the first voiced segment, value in the last voiced segment, correlation coefficient, slope, and error of the linear regression	Tension of the vocal folds and the sub glottal air pressure
First two formant frequencies and their bandwidths	Minimum value, maximum value, range, mean, median, standard deviation and value in the first and last voiced segments	Vocal tract resonances
Energy	Minimum value, maximum value, mean, median, standard deviation, value in the first voiced segment, value in the last voiced segment, correlation, slope, and error of the energy linear regression	Vocal effort, arousal of emotions
Rhythm	Speech rate, duration of voiced segments, duration of unvoiced segments, duration of longest voiced segment and number of unvoiced segments	Duration and stress conditions
References	Hansen [59], Ververidis and Kotropoulos [60], Morrison et al. [61] and Batliner et al. [62]	

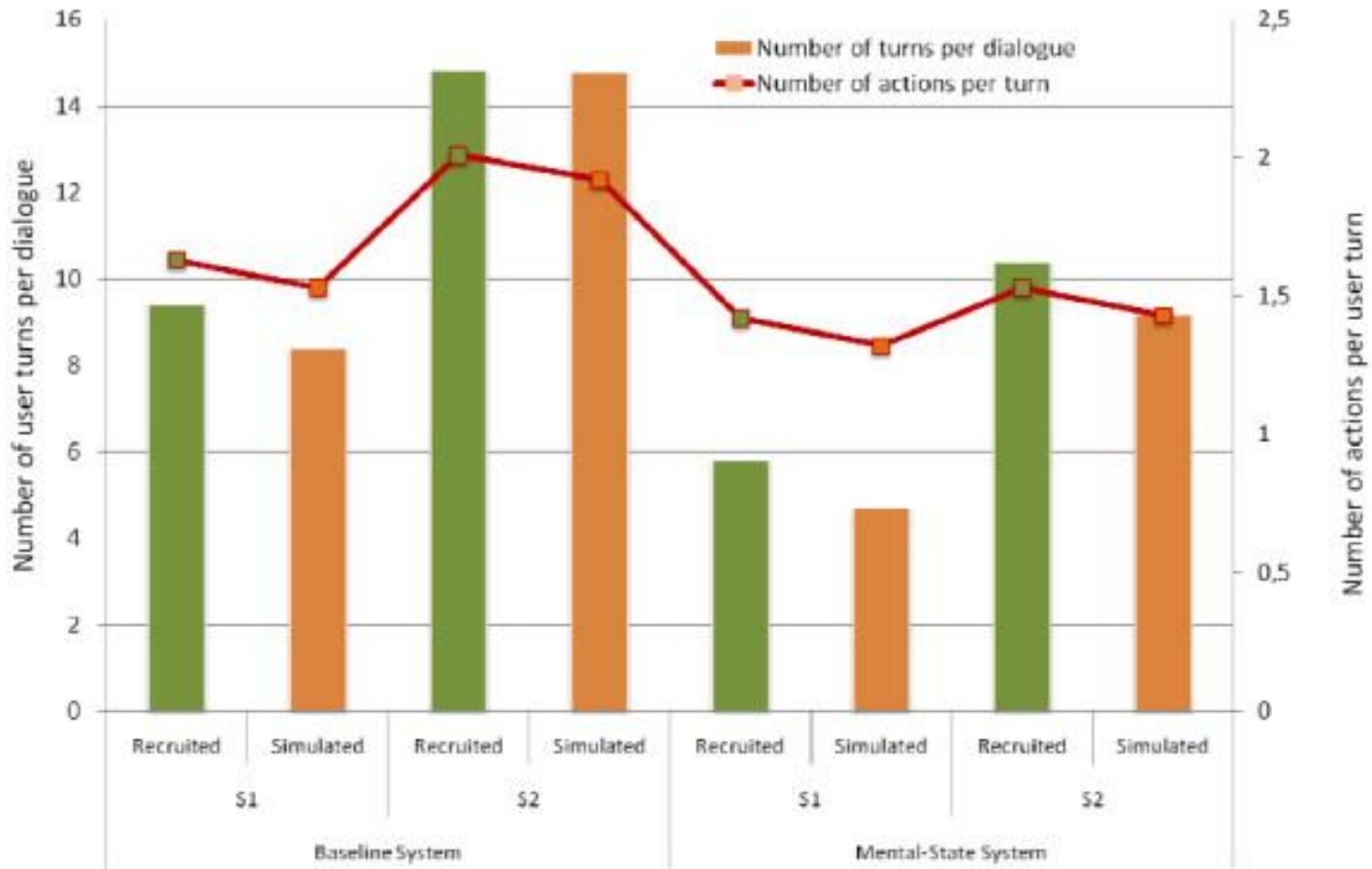
Recent Trends

Emotion recognition



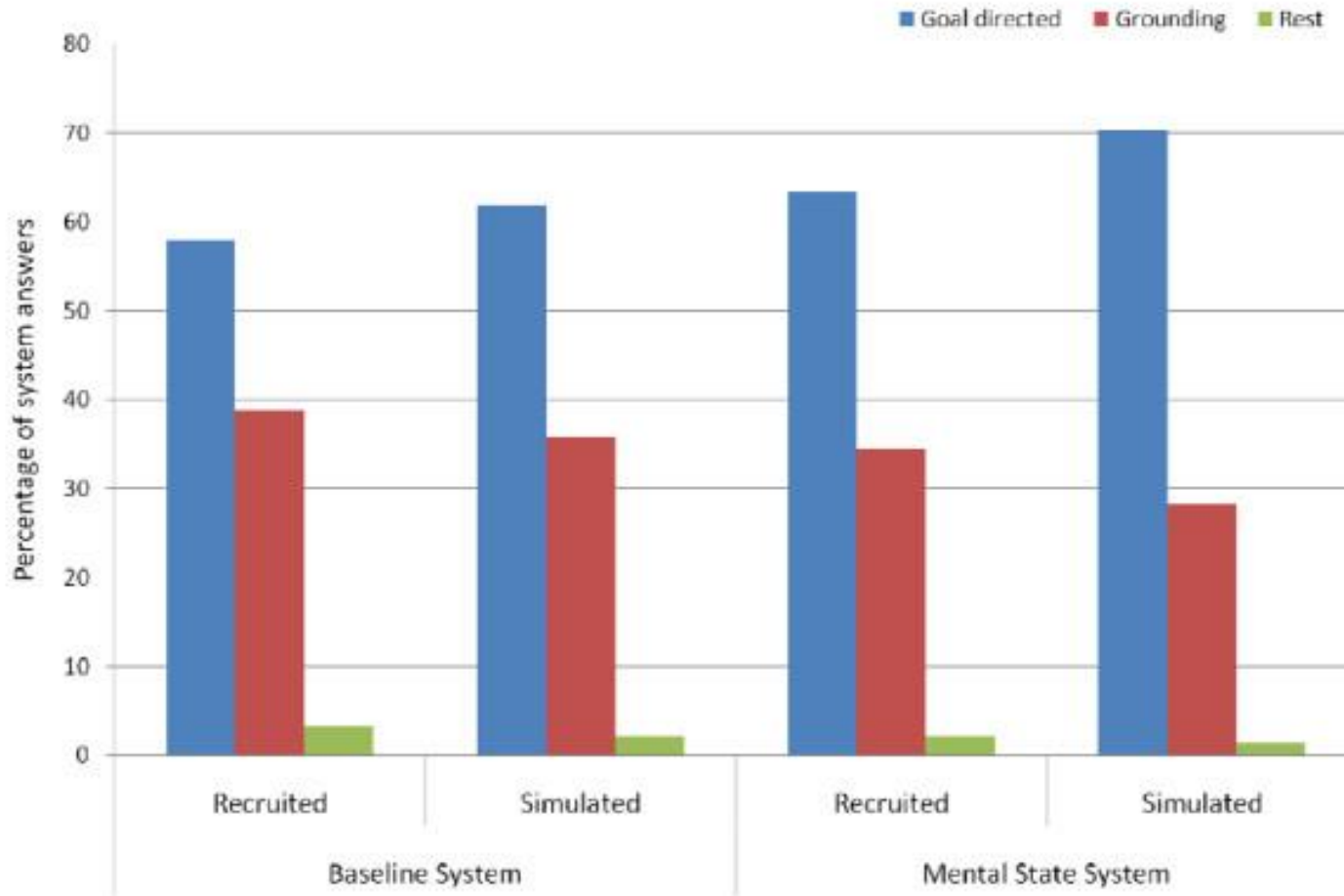
Recent Trends

Emotion recognition



Recent Trends

Emotion recognition



Outline

1. Introduction
2. Spoken Dialog Systems
3. **Recent Trends: Adaptation**
 - a) Statistical dialog management
 - b) User Modeling
 - c) Emotion Recognition
 - d) **Context-aware system**
 - e) Multimodal Interaction
4. Application domains
5. Conclusions



Recent Trends

Context-aware systems

- Users can access the huge amount of information and web services:
 - from almost everywhere;
 - through different communication modalities.
- Effectively bringing information and services to people can only be done:
 - supporting pervasive computing;
 - contextual information must be conveniently gathered and processed.



Recent Trends

Context-aware systems

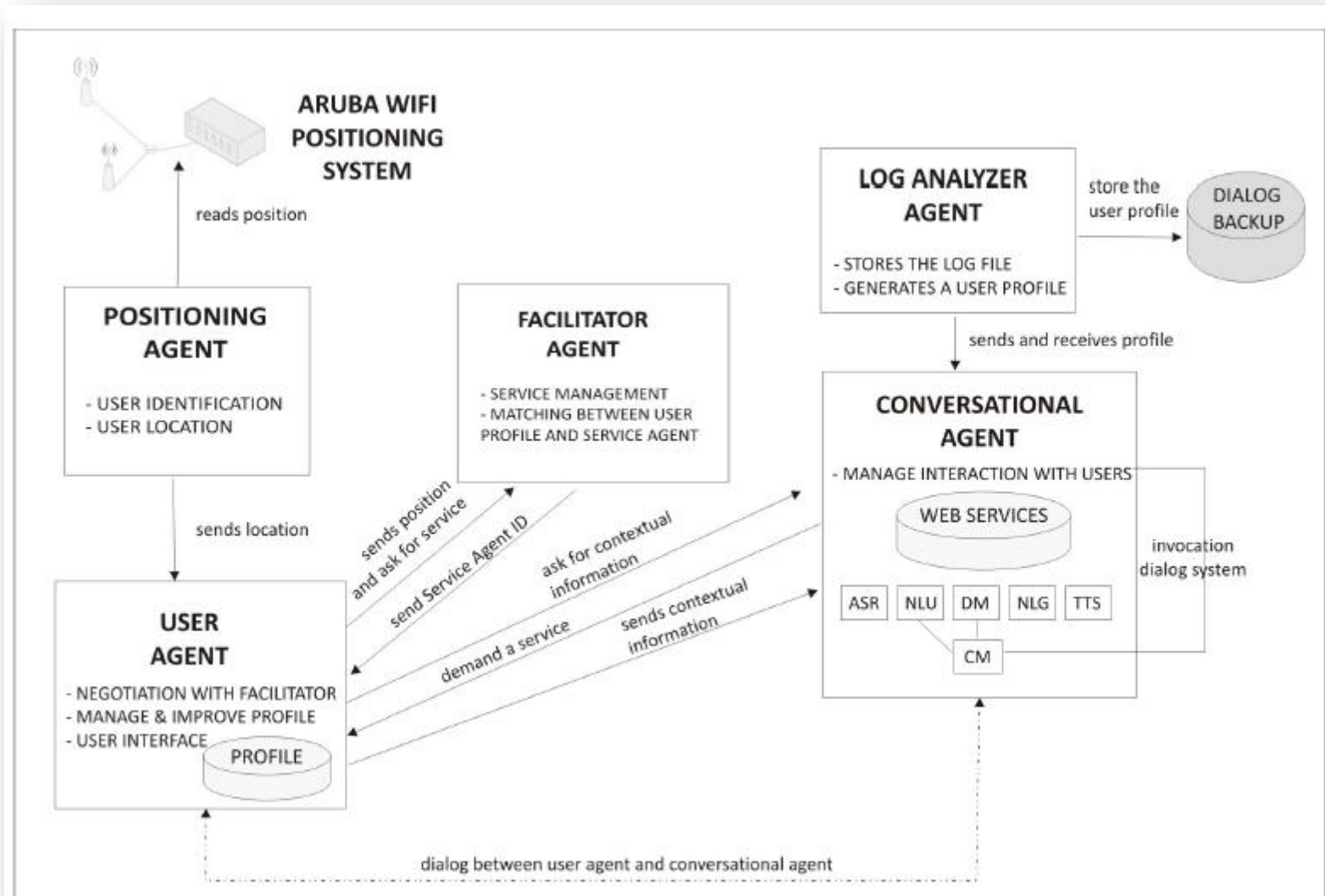
- Ambient Intelligence (Aml) may support this desirable feature of personalization in web services:
 - People surrounded by a set of interconnected computing and sensing devices invisible to them:
 - Recognize the surrounding environment;
 - Facilitate the provision of personalized services through intelligent and intuitive interfaces.



- Services personalized and dynamically adapted to the interaction context.
- Users naturally perceive the service interface, by means they interact, as intelligent.

Recent Trends

Context-aware systems



Proposed multi-agent architecture



Recent Trends

Context-aware systems

- **Context Manager:**
 - Create, update and load the context information by means of specific profiles associated to each user.
 - Communicate this information to the DS modules during the interaction.
- **Context Profile:**
 - **General user information:** user's name, gender, age, current language, skill level when interacting with dialog systems, pathologies or speech disorders...
 - **General statistics:** number of dialogs and dialog turns, their durations, the date of the last interaction with the system, etc.
 - **Usage statistics and user privileges:** specific domain knowledge.



Recent Trends

Context-aware systems

Process followed to adapt the service

1. A user's device is discovered by the WiFi sensor,
2. Zone name identification with his location and a set of categories of services are provided as icons on his GUI;
3. The user decides to use the spoken communication interface
4. The ACE sends an XML package to the DS Context Manager: identification and current location.
5. The DS Context Manager selects the user profile and communicates this information to the different modules of the DS.



Recent Trends

Context-aware systems

Process followed to adapt the service

6. Each module uses this information to load its specific information and models
7. The user starts the interaction with the DS.
8. The *Context Manager sends the value of the* new features only to the modules that requires such information;
9. At the end of the interaction, the user profile is updated using the information of the last dialog session.



Recent Trends

Context-aware systems

- Task: Design an adaptive system that provides railway information in natural language.
- Users dialog acts:
 - Task-dependent: *Hour, Price, Train-Type, Trip-Time, and Services.*
 - Task-independent: *Affirmation, Negation, Not-Understood.*
- *Attributes needed*:
 - *Origin, Destination, Departure-Date, Arrival-Date, Ticket-Class, Departure-Hour, Arrival-Hour, Train-Type, Order-Number, and Services.*
- *System responses*:
 - *Opening, Closing, Not-Understood, Waiting, New-Query, Acceptance, Rejection, Question, Confirmation, and Answer.*

Recent Trends

Dialog without including context information

S₁: Welcome to the railway information system. How can I help you?

U₁: I want to know timetables to Torrelodones.

S₂: From which station do you want to leave?

U₂: Atocha Station.

S₃: Please, tell me date.

U₃: Today.

S₄: There are five trains leaving from Atocha Station to Torrelodones at 10:00am, 11:30am, 15:00pm and 19:00pm. Anything else?

U₄: Tell me the price of the first one.

S₅: Please, tell me the ticket class.

U₅: Business class.

S₆: The price in business class is 2.5 euros. Anything else?

U₆: No, thank you.

S₇: Thank you for using the railway information system. Have a good trip!

*No prompts
personalization.*

*There is not information
about the current location
of the user.*

No date stored.

*No information about the
user preferences:
timetables.*

*No information about the
user preferences: ticket
class.*



Recent Trends

Dialog including context information

S₁: Good morning José, I consult you trains information leaving from Atocha Station. Do you want timetables information?

Prompt personalization including the user's name. The user location is known and his preference about consulting timetables.

U₁: Yes, I want to know timetables to Torrelodones.

S₂: The following two trains leaves today at 10:00am and 10:27am from Atocha Station to Torrelodones. Anything else?

The system uses the current hour and the date to personalize the timetables service. There is not any information about other possible queries performed by the user in previous dialogs.

U₂: Please, tell me the price of the first one.

S₃: The price in business class is 2.5 euros. Anything else?

The system uses the ticket-class preferences included in the user profile.

U₃: No, thank you.

S₄: Thank you José. Have a good trip!

Prompt personalization including the user's name.



Recent Trends

User name: *José García*

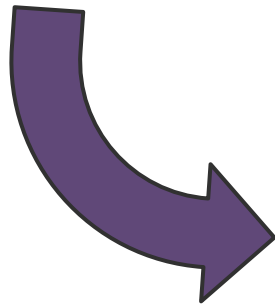
Location: *Atocha Station*

Date and Time: *2009-05-01, 9:00AM*

Device: *PDAQ 00-18-41-32-0B-59*

Objective: *To know timetables and prices to
Torrelodones station*

**Dialog
Scenario**



**XML
Package**

```
<xsd:complexType name="Identification">
  <xsd:user>
    <xsd:name="José García" />
    <xsd:role="Passenger" />
    <xsd:location="Atocha Station" />
    <xsd:date="2009-05-01" />
    <xsd:time="9:00am" />
    <xsd:MAC_address="00-18-41-32-0B-59" />
  </xsd:user>
</xsd:complexType>
```

Recent Trends

Context-aware systems

User Profile

Name: José García

Gender: Male | Age: 29 | Language: Spanish | Skill level: High |
Pathologies: None ...

Sessions: 12 | Average_Turns: 5.7 | Average_Duration: 2.2 |
Last_Session: 2009-04-29 ...

Preferences: Timetables, Talgo Train,
Business Class, Atocha Station, 10:00am...

Current_Location: Atocha Station | Platform zone



Recent Trends

Context-aware systems

- Acquire a total of 300 successful dialogs including or not context information
 - User simulation technique.
 - Two types of scenarios: one objective (S1) or two (S2).
 - Compare both corpora using seven statistical measures.

	Without Context Information		Using Context Information	
	Type S1	Type S2	Type S1	Type S2
Percentage of successful dialogues	1.6%	0.1%	11.4%	2.5
Average number of turns per dialogue	9.4	12.6	4.9	6.2
Percentage of different dialogues	92.9%	98.3%	71.9%	83.7%
Number of repetitions of the most seen dialogue	5	3	12	7
Number of turns of the most seen dialogue	7	9	5	7
Number of turns of the shortest dialogue	5	7	5	7
Number of turns of the longest dialogue	25	27	17	19

- Easiness during the simulation.
- Reduction in the number of turns.

Recent Trends

Context-aware systems

- **High-level features (dialog and turn lengths)**

	Initial Convers. Agent	Final Convers. Agent
Average number of user turns per dialog	4.99	3.75
Number of turns of the most seen dialog	2	2
Number of turns of the shortest dialog	2	2
Number of turns of the longest dialog	14	12

- Reduction in the average number of turns required to fulfill the complete set of objectives defined in the scenarios.

Recent Trends

Context-aware systems

- **Dialog style**

	Initial Convers. Agent	Final Convers. Agent
Request to the system	31.74%	35.43%
Provide information	20.72%	24.98%
Confirmation	10.81%	7.34%
Yes/No answers	31.47%	28.77%
Other answers	3.26%	3.48%

	Initial Convers. Agent	Final Convers. Agent
Confirmation of concepts and attributes	13.51%	10.23%
Questions to require information	18.44%	19.57%
Answers generated after a database query	68.05%	70.20%

- Users need to employ less confirmation turns in the final agent.
- Reduction in the number of system confirmations for data items.
- Both results show that the final conversational agent carries out a better selection of the system responses.

Outline

1. Introduction
2. Spoken Dialog Systems
3. **Recent Trends: Adaptation**
 - a) Statistical dialog management
 - b) User Modeling
 - c) Emotion Recognition
 - d) Context-aware system
 - e) **Multimodal Interaction**
4. Application domains
5. Conclusions



Recent Trends

Multimodal Interaction

- The great interest and extension which have currently reached social networks makes possible:
 - quickly introduction of a great number of applications;
 - new types of communication.
- Possibility of interacting in virtual worlds:
 - perfect scenario to develop multimodal systems;
 - take into account new information sources;
 - adapt systems to the specific users preferences.
- Integration of conversational agents in virtual worlds:
 - bot that provides academic information in Spanish in Second Life.



Recent Trends

Multimodal Interaction

- Second Life (SL) is a three-dimensional virtual world accessible on Internet.
- Second Life Viewer allows its users:
 - interact with each other by means of avatars with capacity of movement;
 - advanced level of social network.
- different types of communication among residents
 - gestures (animations),
 - text messages (chat),
 - voice.



Recent Trends

Multimodal Interaction

- SL is currently used with success as an educational platform in many institutions:
 - it is one of the most popular virtual social worlds (millions of users);
 - it uses very advanced technologies for the development of realistic simulations
 - many possibilities for customization;
 - encourages innovation.



Recent Trends

Multimodal Interaction

- We have developed a conversational metabot that facilitates academic information (subjects, lecturers, doctoral studies, and deadlines) based on the features provided by a previously developed conversational agent.



Recent Trends

Multimodal Interaction

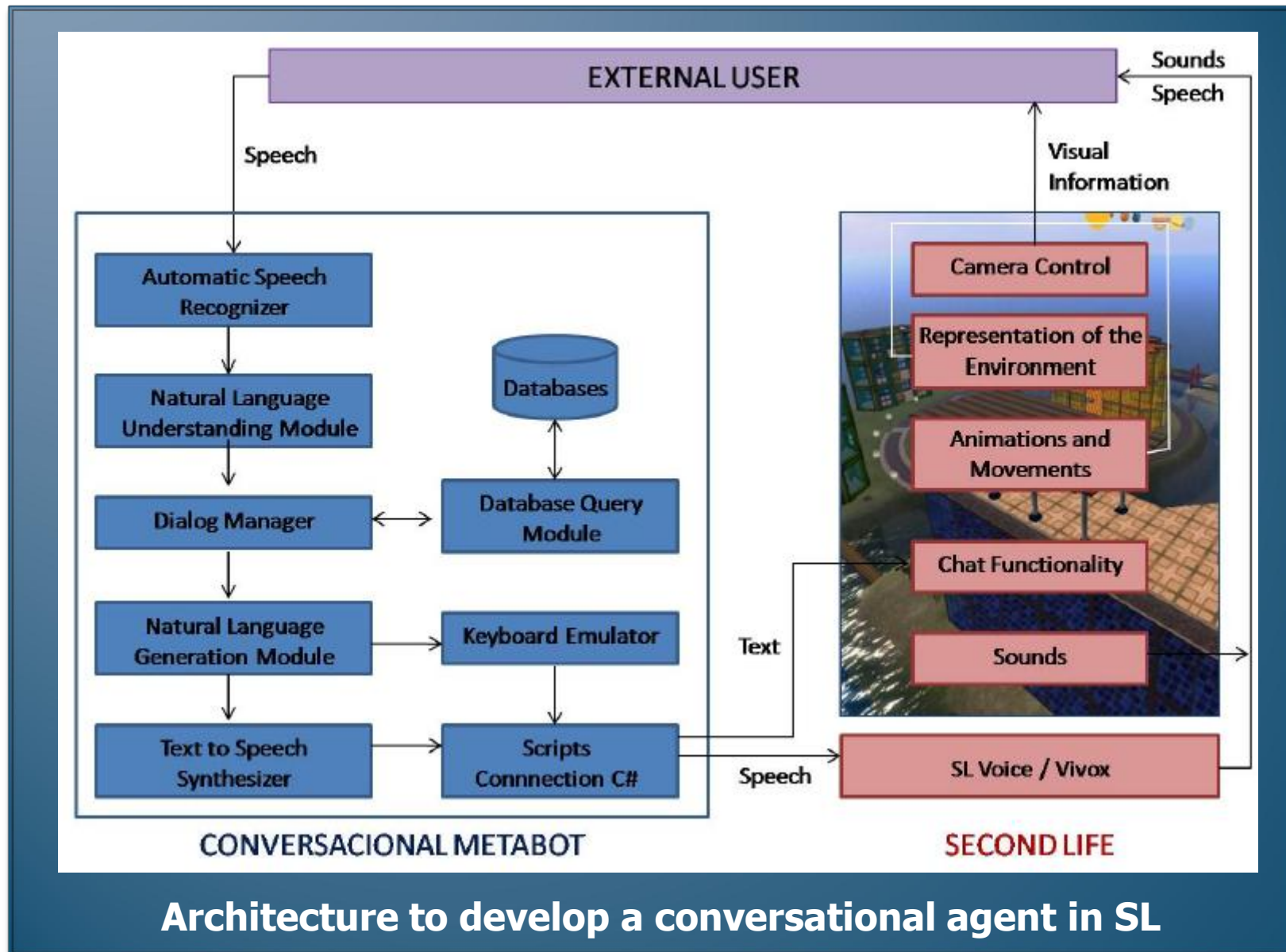
- **Conversational agent:**

Category	Information provided by the user (including examples)		Information provided by the system
Subject	<i>Name</i>	Compilers	Degree, lecturers, responsible lecturer, semester, credits, web page
	<i>Degree</i> , in case that there are several subjects with the same name	Computer science	
	<i>Group name</i> and optionally <i>type</i> , in case he asks for information about a specific group	A Theory A	Timetable, lecturer
Lecturers	Any combination of <i>name</i> and <i>surnames</i>	Zoraida Zoraida Callejas Ms. Callejas	Office location, contact information (phone, fax, email), groups and subjects, doctoral courses
	Optionally <i>semester</i> , in case he asks for the tutoring hours	First semester Second semester	Tutoring timetable
Doctoral studies	Name of a doctoral program	Software development	Department, responsible
	Name of a course if he asks for information about a specific course	Object-oriented programming	Type, credits
Registration	Name of the deadline	Provisional registration confirmation	Initial time, final time, description

- Placed outside the virtual world;
- Voice and web servers
- Technologies: C#.NET, SpeechLib, SLVoice, Vivox, Microsoft SAPI, Loquendo TTS.



Recent Trends



Architecture to develop a conversational agent in SL



Recent Trends

Multimodal Interaction

- Preliminary evaluation by means of the acquisition of 50 dialogs

Average number of turns per dialog	4.99
Percentage of confirmations from the metabot	13.51%
Questions from the metabot to request information	18.44%
Prompts generated by the metabot after a database query	68.05%

- Similar behaviour to the conversational agent using only voice
- Required:
 - to evaluate new features to be included in the conversational metabot to improve the communication process.
 - to carry out a detailed analysis of the integration of different modalities.



Outline

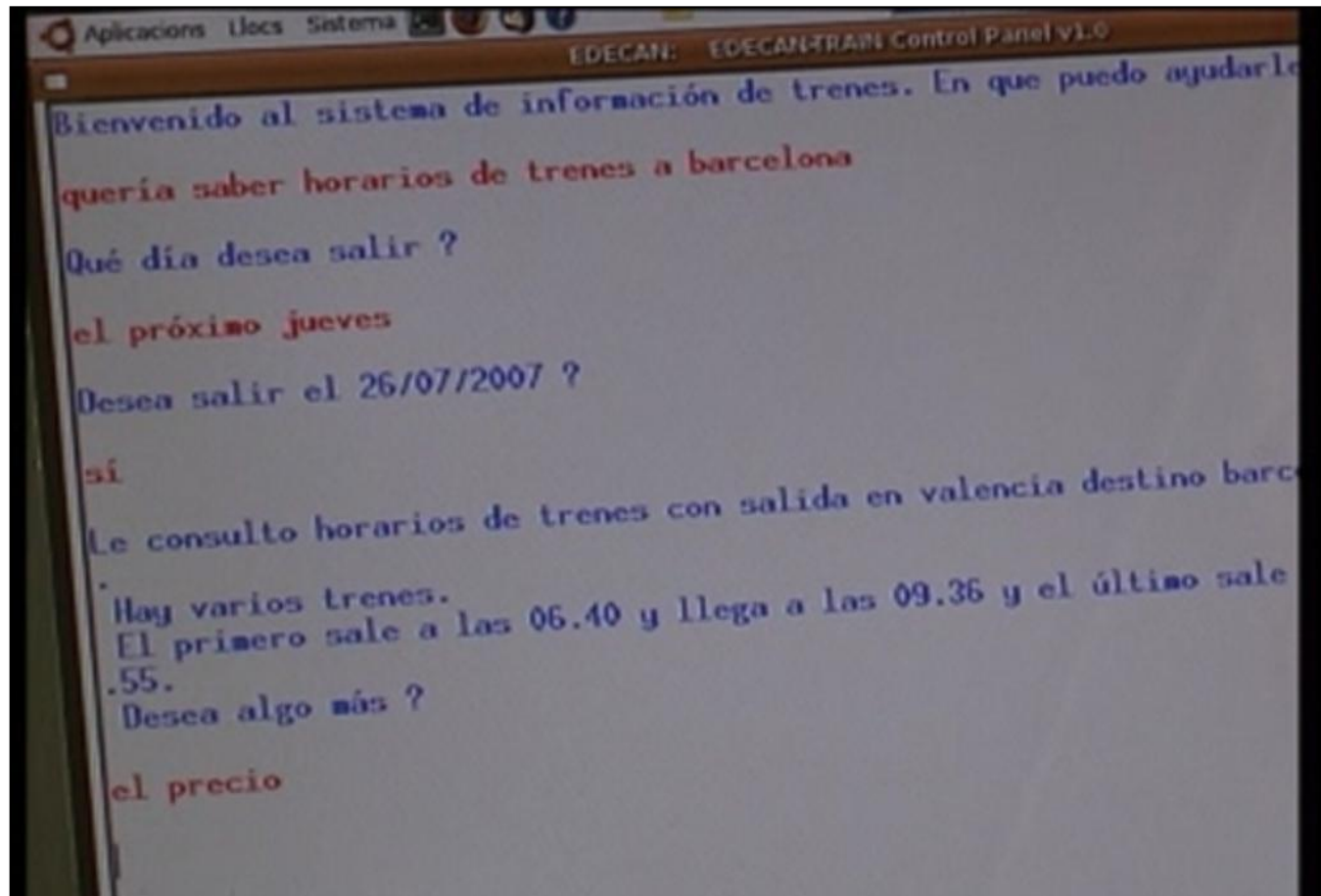
1. Introduction
2. Spoken Dialog Systems
3. Recent Trends: Adaptation
 - a) Statistical dialog management
 - b) User Modeling
 - c) Emotion Recognition
 - d) Context-aware system
 - e) Multimodal Interaction
4. **Application domains**
5. Conclusions



Application Domains

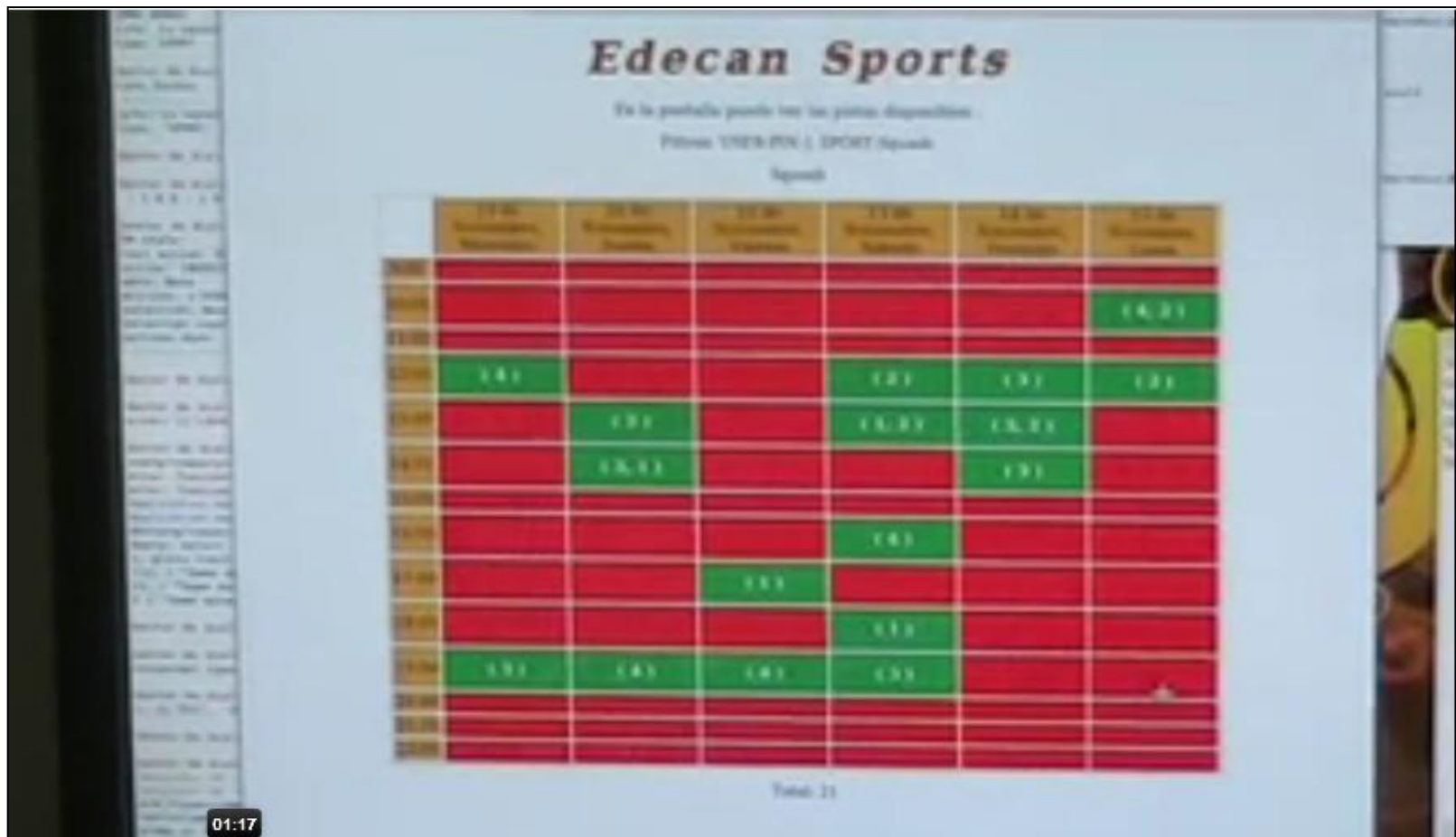
- **Provide Information:**

- Conversational agent providing information in natural language about train services, schedules, and fares in Spanish.



Application Domains

- **Carry out a specific task:**
 - Development of a conversational agent for the booking of sport facilities.



The screenshot displays the Edecan Sports website interface. At the top, the title "Edecan Sports" is visible, followed by the tagline "En la práctica puede ser la mejor disposición." and the text "Pólizas: USER PIN + SPORT Squads". Below this, a calendar grid is shown for the month of August. The calendar has columns for different days of the week and rows for dates. The cells in the calendar are colored red, indicating that the facilities are booked, and green, indicating they are available. The green cells contain the number "100", which likely represents the number of available spots or a booking count. The calendar is titled "Ago" and "Total: 21".

	100	100	100	100	100	100
1						100
2						
3	100			100	100	100
4		100		100	100	
5		100			100	
6				100		
7			100			
8				100		
9	100	100	100	100		
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						



Application Domains

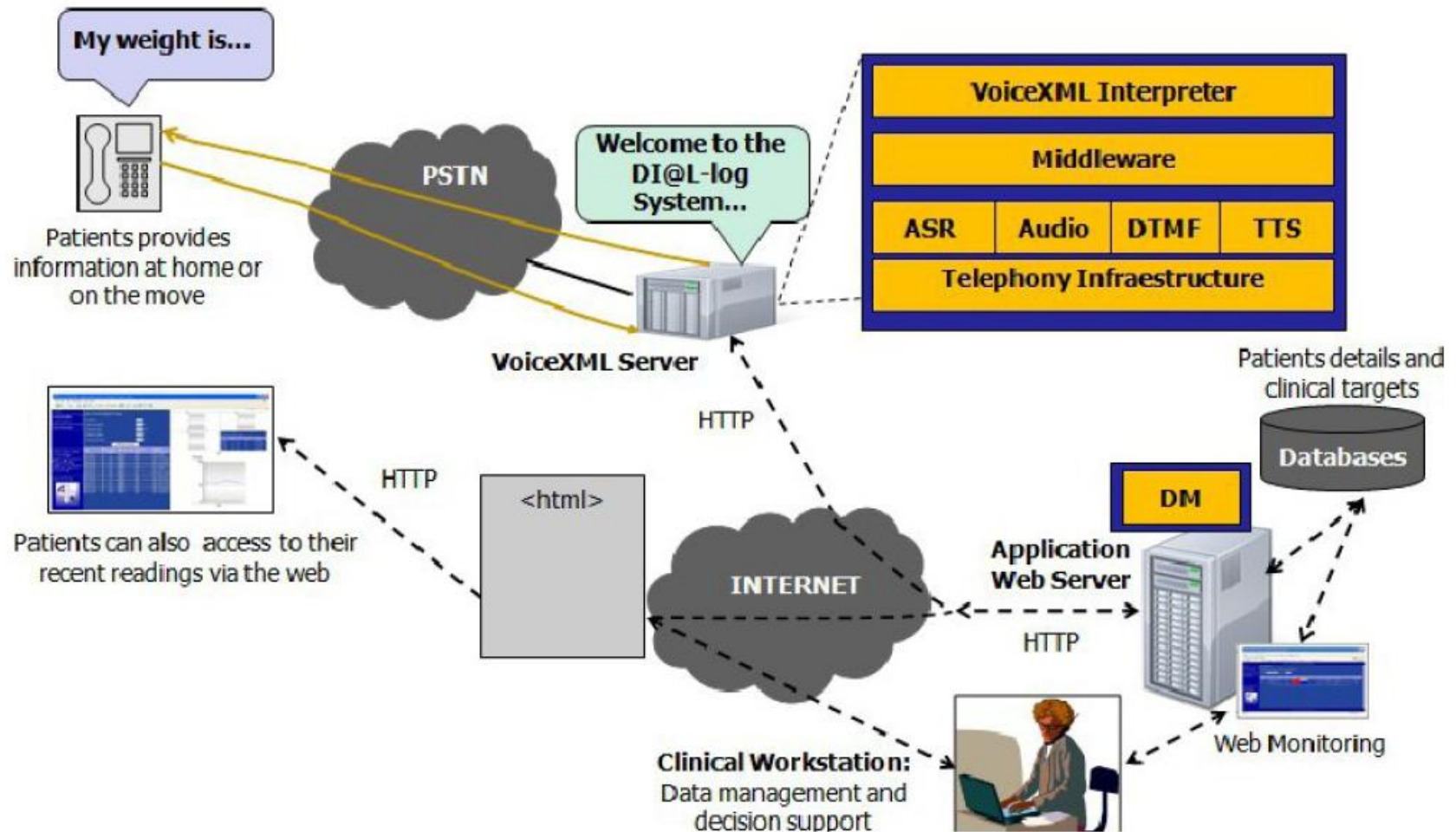
- **E-Health:**

- DI@L-log is a conversational agent which acts as a voice logbook to collect home monitored data from patients suffering from Type-2 diabetes.
- The data collected by the system are the patient's weight, blood pressure (systolic and diastolic values) and sugar levels.
- The system validates and analyses the data:
 - providing some immediate feedback to the patients regarding their current progress;
 - communicating the results to doctors at the hospital who are able to review the patient's progress graphically and deal with any alerts generated by the system concerning abnormal developments.



Application Domains

- E-Health:**



Application Domains

- **E-Health:**

- Taking into account the requirements of the task, an initial strategy was designed for the DM:

S1: Welcome to the DI@L-log System. Please say or enter your four digit pin.
U1: 1356.
S2: Are you ready to give your measurements?
U2: Yes.
S3: Let's begin with your weight. How many stones?
U3: 12.
S4: I have understood 12 stones. Is it correct?
U4: Yes.
S5: What is your blood sugar?
U5: 11.
S6: I have understood 11. Is it correct?
U6: Yes.
S7: And finally, for your blood pressure you will be asked for the top number and then the lower number. What is the top number of your blood pressure?
U7: 160.
S8: I have understood 160. Is it correct?
U8: Yes.
S9: And what is the lower number?
U9: 86.
S10: I have understood 86. Is it correct?
U10: Yes.
S11: Your readings have been successfully updated. According to them, your weight is higher than last week, be very careful. Thanks for contacting DI@L-log. Goodbye.

Identification and waiting phase

The system analyzes which data is required for the current user and requires it.

Every item is confirmed after the user has provided its value.

The user can only provide one item at a time.

System analysis and diagnosis

Application Domains

- Carry out a specific task:
 - Learning foreign languages

Voice Pronunciation

HOME DICTIONARY PRONUNCIATION LIBRARY VIDEOCLUB GOOGLE

Max Puntuation: 100 pts
Puntuation: 0 pts

Pictures

Picture:

Description:
This is a PLACE or a BUILDING

EXIT SOLUTION NEXT

Legal Information | Privacy Policy | Copyright 2010 | Contact Us

W3C CSS

Voice Applications

HOME DICTIONARY PRONUNCIATION LIBRARY VIDEOCLUB GOOGLE

Voice Dictionary

With the application Voice Dictionary you can search the meaning of any word you want using your voice. The application will say the meaning of this word and show it on the screen to make easy you can search any other word or use the hipertext you find.

Voice Dictionary searches the definitions and also the history and all the information about this word. All of functionalities of this application can be used with your own voice or with your keyboard and your mouse.

If you want to open Voice Dictionary you must to say: "DICTIONARY".

Voice Pronunciation

With Voice Pronunciation you can play a little game to improve pronunciation. The application shows you a word and you must to say what it is with your voice. If it's correct, the application shows you a picture of the word. If you want to open Voice Pronunciation you must to say: "PRONUNCIATION".

Voice Library

This application helps you to find books in internet with your voice. You can obtain by means of the voice the summary, the price and the opinion of other lectors with Voice Library, and you will receive the results by voice.

Voice Library offers the possibility to accede to the page where the application found the results and you can buy the book. The application also offers the possibility to link to other books that will be related.

If you want to open Voice Library you must to say: "LIBRARY".

Voice Videoclub

Voice Videoclub is the application used to search films in the internet by means of your own voice. You tell the application the name of the film and Voice Videoclub show in your screen the results, also by means of voice.

The application search the summary of the film, if it is for sale Voice Videoclub search the price, and if the film is uploaded in YouTube the application search the trailer.

If you want to open Voice Videoclub you must to say: "VIDEOCLUB".

Voice Google

This application use all the power of Google adding the functionality of voice. Voice Google allow you to use Google with your own voice and you will obtain the same results with the possibility that you will can accede to the results with your voice.

If you want to open Voice Google you must to say: "GOOGLE".

Legal Information | Privacy Policy | Copyright 2010 | Contact Us

W3C CSS

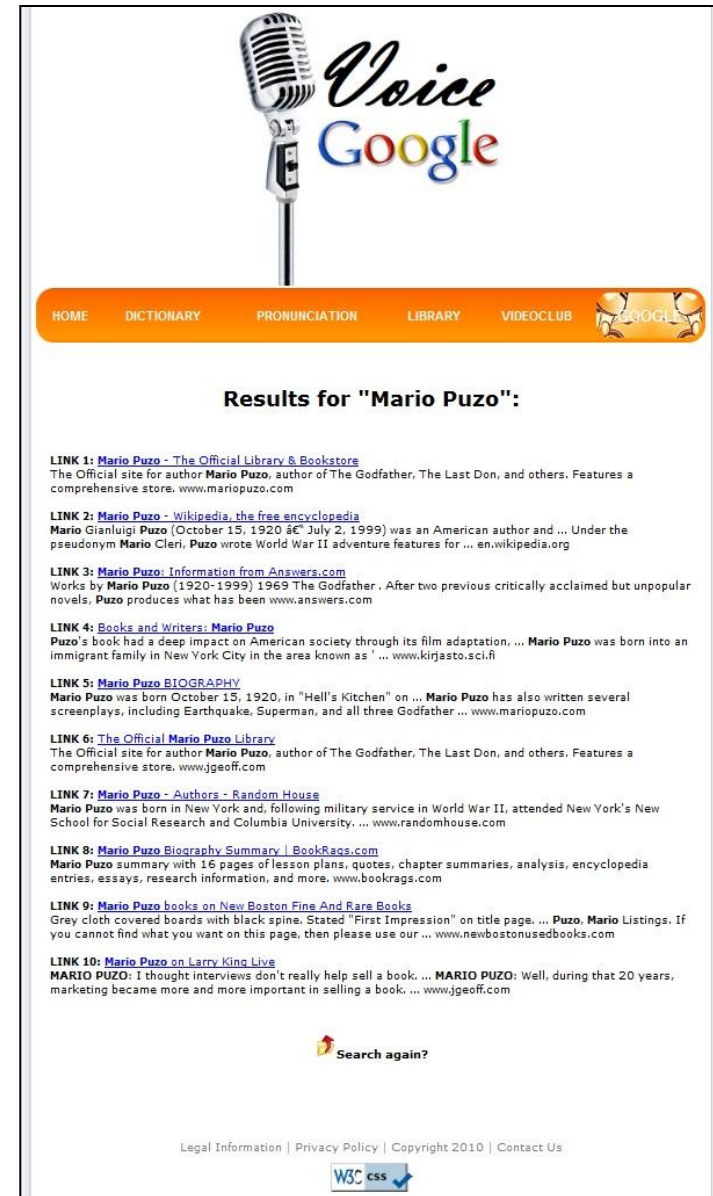


Application Domains

- Access for disabled people:
 - Voice Google



The screenshot shows the Voice Google search page. At the top, there is a microphone icon and the text "Voice Google". Below this is a navigation bar with links for HOME, DICTIONARY, PRONUNCIATION, LIBRARY, and VIDEOCLUB. The main content area features a search bar with the text "Enter query words:" and a "Send" button. There is also a link for "Instructions" and a "W3C CSS" logo at the bottom.



The screenshot shows the search results for "Mario Puzo". At the top, there is a microphone icon and the text "Voice Google". Below this is a navigation bar with links for HOME, DICTIONARY, PRONUNCIATION, LIBRARY, and VIDEOCLUB. The main content area features the text "Results for 'Mario Puzo':" followed by a list of search results. Each result is a link to a different website or resource related to Mario Puzo. At the bottom, there is a "Search again?" button and a "W3C CSS" logo.

Results for "Mario Puzo":

LINK 1: [Mario Puzo - The Official Library & Bookstore](#)
The Official site for author **Mario Puzo**, author of The Godfather, The Last Don, and others. Features a comprehensive store. [www.mariopuzo.com](#)

LINK 2: [Mario Puzo - Wikipedia, the free encyclopedia](#)
Mario Gianluigi Puzo (October 15, 1920 â€” July 2, 1999) was an American author and ... Under the pseudonym **Mario Cleri**, **Puzo** wrote World War II adventure features for ... [en.wikipedia.org](#)

LINK 3: [Mario Puzo: Information from Answers.com](#)
Works by **Mario Puzo** (1920-1999) 1969 The Godfather. After two previous critically acclaimed but unpopular novels, **Puzo** produces what has been [www.answers.com](#)

LINK 4: [Books and Writers: Mario Puzo](#)
Puzo's book had a deep impact on American society through its film adaptation. ... **Mario Puzo** was born into an immigrant family in New York City in the area known as ' ... [www.kirjasto.sci.fi](#)

LINK 5: [Mario Puzo BIOGRAPHY](#)
Mario Puzo was born October 15, 1920, in "Hell's Kitchen" on ... **Mario Puzo** has also written several screenplays, including Earthquake, Superman, and all three Godfather ... [www.mariopuzo.com](#)

LINK 6: [The Official Mario Puzo Library](#)
The Official site for author **Mario Puzo**, author of The Godfather, The Last Don, and others. Features a comprehensive store. [www.jgeoff.com](#)

LINK 7: [Mario Puzo - Authors - Random House](#)
Mario Puzo was born in New York and, following military service in World War II, attended New York's New School for Social Research and Columbia University. ... [www.randomhouse.com](#)

LINK 8: [Mario Puzo Biography Summary | BookRags.com](#)
Mario Puzo summary with 16 pages of lesson plans, quotes, chapter summaries, analysis, encyclopedia entries, essays, research information, and more. [www.bookrags.com](#)

LINK 9: [Mario Puzo books on New Boston Fine And Rare Books](#)
Grey cloth covered boards with black spine. Stated "First Impression" on title page. ... **Puzo, Mario** Listings. If you cannot find what you want on this page, then please use our ... [www.newbostonusedbooks.com](#)

LINK 10: [Mario Puzo on Larry King Live](#)
MARIO PUZO: I thought interviews don't really help sell a book. ... **MARIO PUZO:** Well, during that 20 years, marketing became more and more important in selling a book. ... [www.jgeoff.com](#)

[Search again?](#)

Legal Information | Privacy Policy | Copyright 2010 | Contact Us

W3C CSS

Application Domains

- **Interaction in virtual worlds:**
 - Design of a chatbot to provide academic information



Application Domains

Salto al contenido | Accesibilidad | Mapa del sitio | EMGIASA | MAVA | UPA | IMEPE

Puedes ver Alcorcón en nuestra cuenta de Flickr

Ayuntamiento de Alcorcón
Alcorcón, Comunidad de Madrid, España

Introduce el texto a buscar...

in nova alcorcón
ciudad digital

ADMINISTRACIÓN ELECTRÓNICA
Gestiones on-line

Personas mayores | Familias | Mujer | Menores | Personas con discapacidad | Jóvenes | Inmigración

CONÉCTATE CON TU ALCALDE

Portada

- Ayuntamiento
- Alcorcón
- Sala de prensa
- Eventos
- Servicios y Trámites
- Subvenciones
- Perfil del contratante
- Galería de Fotos
- Áreas temáticas
 - Bienestar Social
 - Circulación y Transporte
 - Cultura y Ocio
 - Deportes

Noticias

Seguridad

El Ayuntamiento de Alcorcón organiza la III Semana de la Prevención de Incendios, una iniciativa pionera en la Comunidad de Madrid

Lo más consultado

- Listado de Noticias
- Listado de Eventos
- Galería de Fotos
- index
- Página de Servicios y Trámites
- Listado de Documentos
- Licitaciones
- Resultados de búsqueda
- Portada de Ayuntamiento
- Portada de empleo

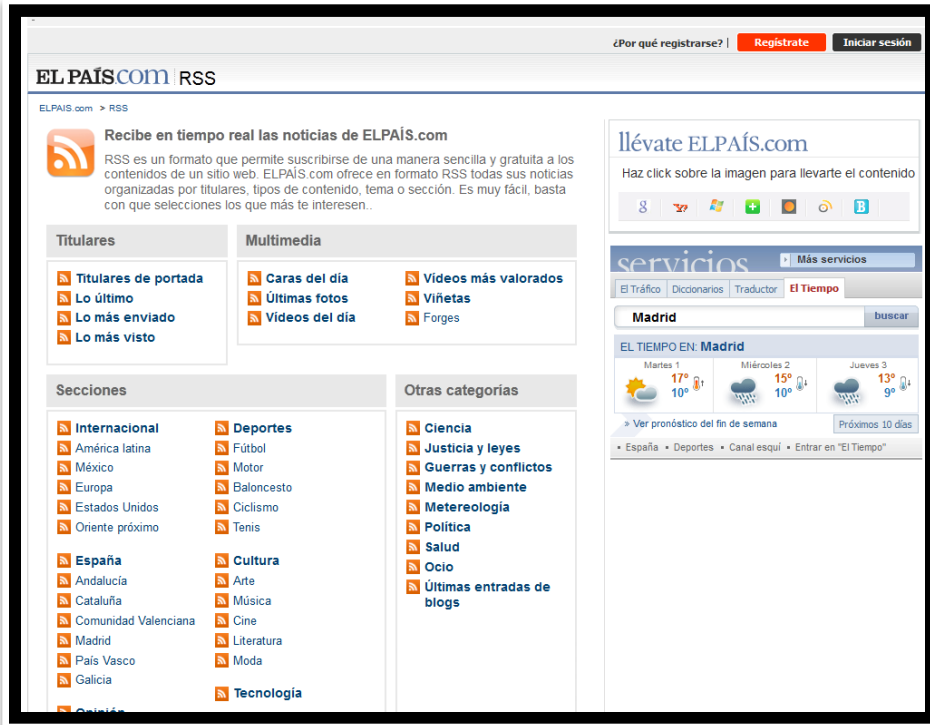
Eventos hoy | Esta semana | Este mes

El Tiempo en Alcorcón Hoy 18°
Max Min 18° 12°

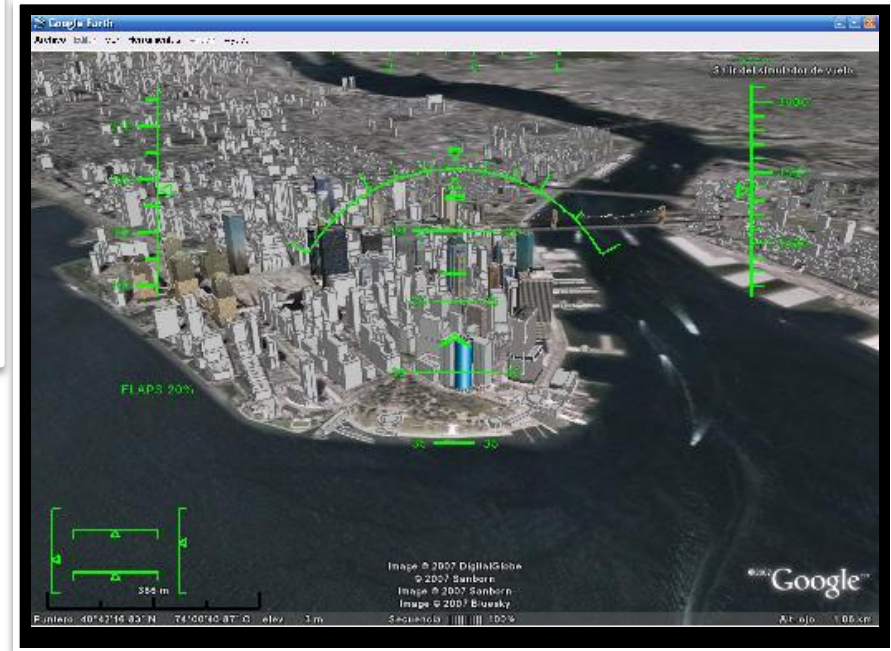
Voice Portals



Application Domains



RSS news speech-based service




Flight Simulation Assistant



Application Domains

20 aniversario UC3M

Carlos III University



Subjects Information

In this section you can consult information of the different subjects taught in the different degrees from the Carlos III University. This information is grouped in different parts such as subject teachers, course objectives, evaluation method, the subject programme, bibliography and practical sessions. The information can be very useful for both students and faculty, but we must remember that it is a guide.

AUDIO OPTIONS:

- CALCULUS
- ARTIFICIAL INTELLIGENCE
- PHYSICS
- COMPUTER TECHNOLOGY
- INFORMATION SKILLS

-RETURN (main page)

•Select one of the following subjects:

Calculus


Send

- HOME
- BACHELORS DEGREE INFORMATION
- SUBJECTS INFORMATION
- BACHELOR DEGREES' SCHEDULES
- WHICH DEGREE SHOULD I CHOOSE?
- PREPARATION TEST
- APPLICATION QUESTIONNAIRE
- USERS STATISTICS

Students Guidance (recommending systems)

20 aniversario UC3M

Carlos III University



Application Questionnaire

By the completion of the following test you will help us to evaluate our website and improve the application. Your opinion is very important for us and we will consider it as much as possible.

- 1.-State on a scale from 1 to 5 your previous knowledge about new technologies for information access.
 1 2 3 4 5
- 2.-State on a scale from 1 to 5 your previous experience using Opera Voice or similar systems.
 1 2 3 4 5
- 3.-How many times have you used "Opera voice" before?
1 : I have not used it before
2 : Rarely
3 : Sometimes
4 : Few times
5 : Many times
 1 2 3 4 5
- 4.-How well did the system understand you?
1 : Extremely bad
2 : Bad

- HOME
- BACHELORS DEGREE INFORMATION
- SUBJECTS INFORMATION
- BACHELOR DEGREES' SCHEDULES
- WHICH DEGREE SHOULD I CHOOSE?
- PREPARATION TEST
- APPLICATION QUESTIONNAIRE
- USERS STATISTICS

On-line surveys

Outline

1. Introduction
2. Spoken Dialog Systems
3. Recent Trends: Adaptation
 - a) Statistical dialog management
 - b) User Modeling
 - c) Emotion Recognition
 - d) Context-aware system
 - e) Multimodal Interaction
4. Application domains
- 5. Conclusions**



Conclusions

- **Main research lines in the Applied Artificial Intelligence Group:**
 - Development of dialog systems for specific tasks;
 - Integration of dialog systems in multiagent systems;
 - Context-aware systems;
 - Multimodal applications (speech+images+video+text, interaction in virtual worlds);
 - Affective computing and HCI.
 - Statistical dialog methodologies for dialog management;
 - User modeling and simulation;
 - Design of user-centered applications and recommending systems.

