



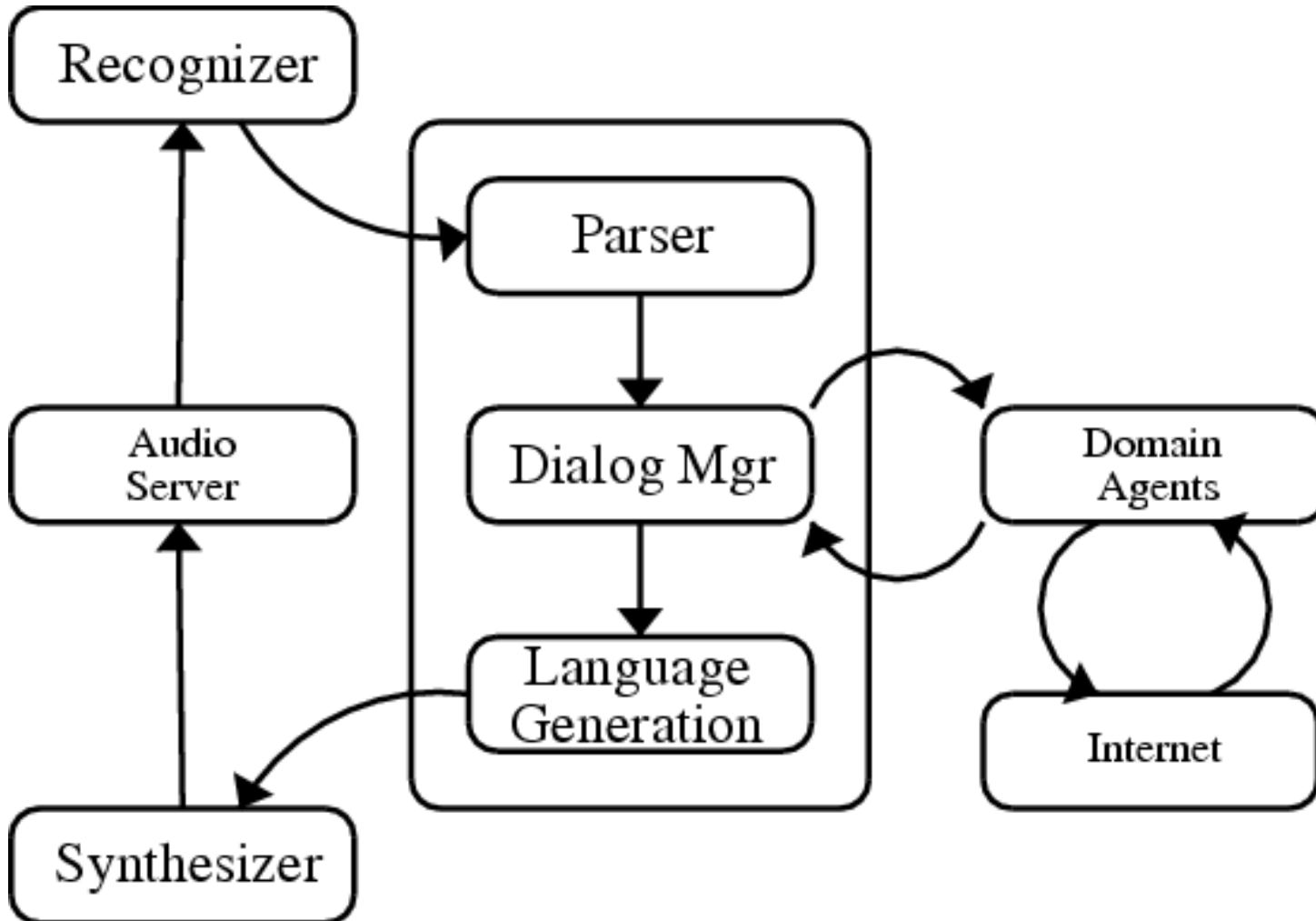
Speech Processing 15-492/18-492

Spoken Dialog Systems
SDS components

Spoken Dialog Systems

- ◆ *More than just ASR and TTS*
 - *Recognition*
 - *Parsing*
 - *Manipulation of utterances*
 - *Generation of new information*
 - *Text generation*
 - *Synthesis*

SDS Architecture



SDS Internals

- ◆ *Parser*
 - *From words to structure*
- ◆ *Dialog Manager*
 - *State of dialog (who is talking)*
 - *Direction of dialog (what next)*
 - *References, user profile etc*
 - *Interaction of database/internet*
- ◆ *Language Generation*
 - *From structure to words*

Parsing

◆ *Parsing of SPEECH not TEXT*

- *Eh, I wanna go, wanna go to Boston tomorrow*
- *If its not too much trouble I'd be very grateful if one might be able to aid me in arranging my travel arrangements to Boston, Logan airport, at sometime tomorrow morning, thank you.*
- *Boston, tomorrow*

Parsing: Output structure

- ◆ *“I wanna go to Boston, tomorrow”*
 - *Destination: BOS*
 - *Departure: 20081028, AM*
 - *Airline: unspecified*
 - *Special: unspecified*
- ◆ *Convert speech to structure*
 - *Sufficient for further processing/query*

Phoenix Parser

[Place]

(carnegie mellon university)

(downtown)

(robinson towne center)

(the airport)

(south hills junction)

(mount oliver)

(the south side)

(oakland)

(bloomfield)

(polish hill)

(the strip district)

(the north side)

;

[NextBus]

(*WHEN_IS *the next *BUS)

(*WHEN_IS *the BUS after that *BUS)

WHEN_IS

(when is)

(when's)

BUS

(bus)

(one)

;

Phoenix Parser

- ◆ *Parse what is important*
- ◆ *Ignore other parts*
- ◆ *Map know parts to usually information*

Parsing vs Language Model

◆ *Language Model*

- *Model what actually gets says*

◆ *Parsing*

- *Extract the information you want*

◆ *Models *can* be shared*

- *Only accept things in the grammar*
- *Can be over limiting*

Dialog Manager

◆ *Maintain state*

- *Where are we in the dialog*
- *Whose turn is it*
 - ⊗ *Waiting for speaker*
 - ⊗ *Waiting for database query (stall user)*
- *Deal with barge-in*

Language Generation

- ◆ *Query for flights to Boston*
- ◆ *Template fill answer(s)*
 - *The next flight to DEST leaves at DEPART_TIME arriving at ARRIVE_TIME.*
- ◆ *Templates may be much more complex*

Language Generation

- ◆ *Choose which template to use*
 - *Based on state, answer type*
 - *Natural variation*
 - *Statistical variation*
- ◆ *Include <ssml> tags to help synthesis*
 - *Can <emph>emphasize</emph> parts*
 - *Can identify dates, numbers etc.*
- ◆ *Humans like variation in the output*
 - *It is rare for a human to repeat things exactly*

Language Generation

- ◆ *Frames structures to (marked up) text*
 - *START: Pittsburgh*
 - *END: Boston*
 - *DATE: 20081028*
 - *TIME: 07:45*
 - *FLIGHT: US075*
- ◆ *Can generation*
 - *I have US 075 leaving at 07:45 tomorrow*
 - *US Airways has a flight departing tomorrow at 07:45*

Standardized things

◆ *Help*

- *User should be able to get help at any time*
- *Explain where they are and what they are expected to say (with explicit examples)*

◆ *Errors*

- *“I didn’t understand” ...*

◆ *Confirmation*

- *Did you say “Boston”?*

Confirmation

◆ *Explicit confirmation*

- *Where are you traveling to ?*

Boston

- *Boston, did I get that right?*

Yes

Confirmation

◆ *Implicit confirmation*

- *Where are you traveling to?*

Boston

- *Boston, where ...*
<can barge in>

Confirmation

- ◆ *Explicit confirmation*
 - *Safe but slow*
- ◆ *Implicit confirmation*
 - *Natural, but requires good support for barge-in*

Grounding

- ◆ *Showing evidence the system understands*

- *Where are you traveling to?*

Boston.

Right. Where

Boston, right. Where

Designing Prompts

- ◆ *Constrain your questions:*
 - *How may I help you?*
 - ⊗ *Long story reply*
 - *What bus number would like schedules for?*
 - ⊗ *Expect bus number replies*

