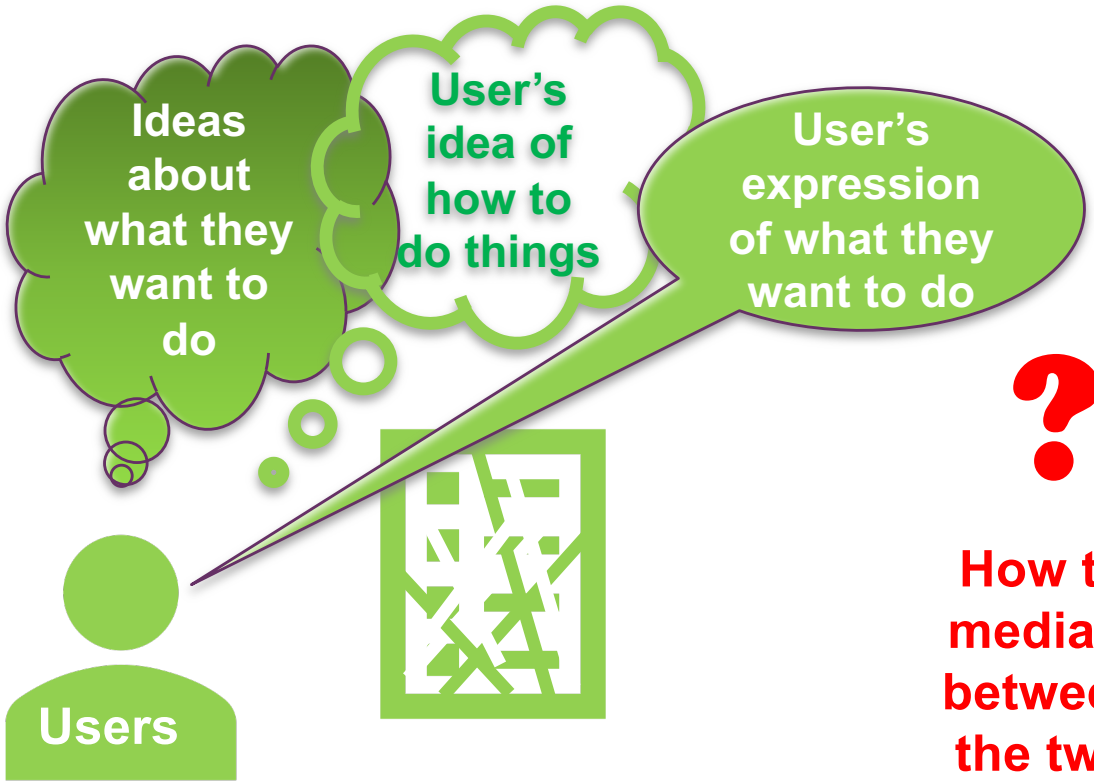
Dialog Challenges

CS 136a

February 28, 2020

Professor Meteer

The dialog problem: Where to start?

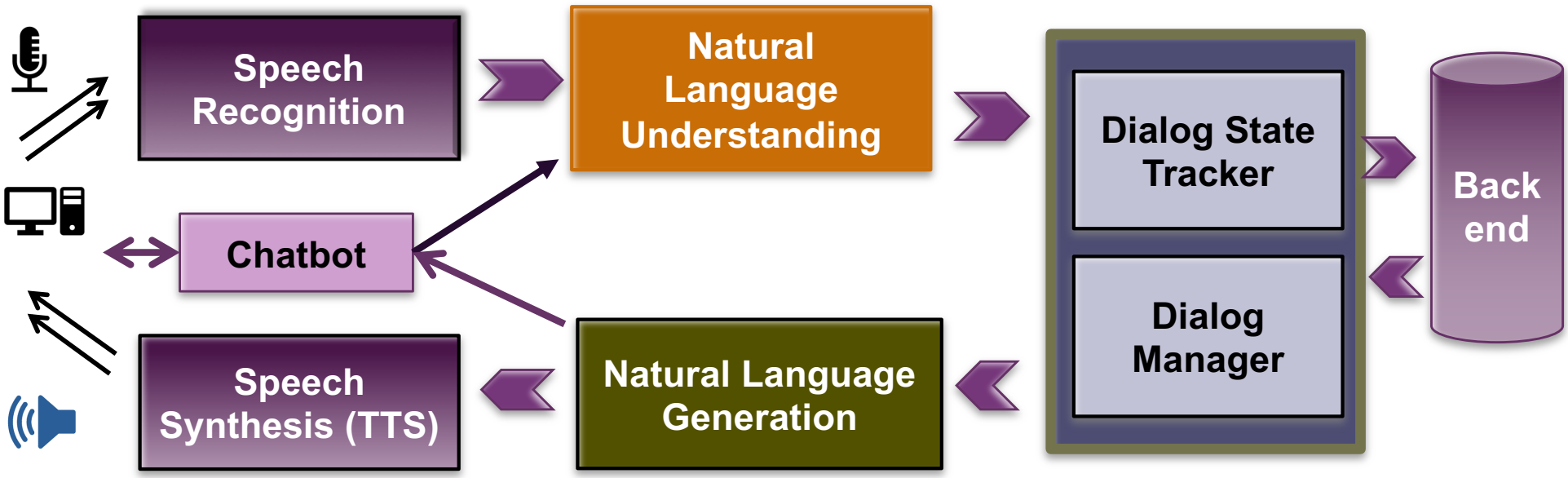


How to mediate between the two

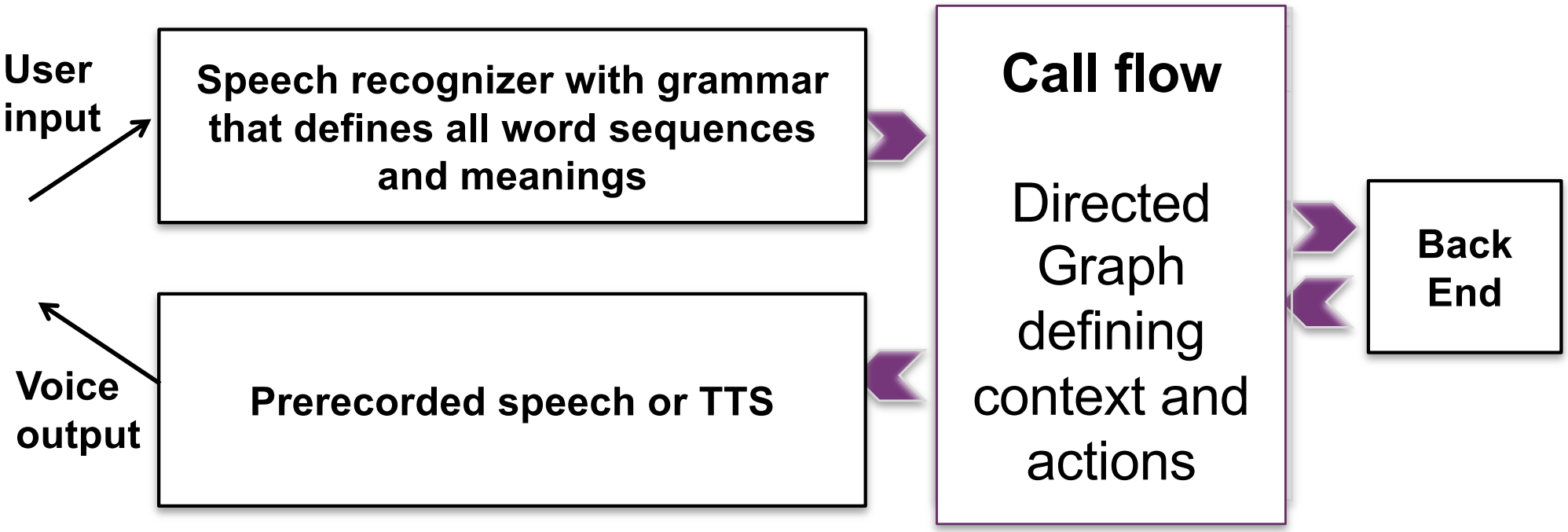
Instructions for how to do things

Things that can be done

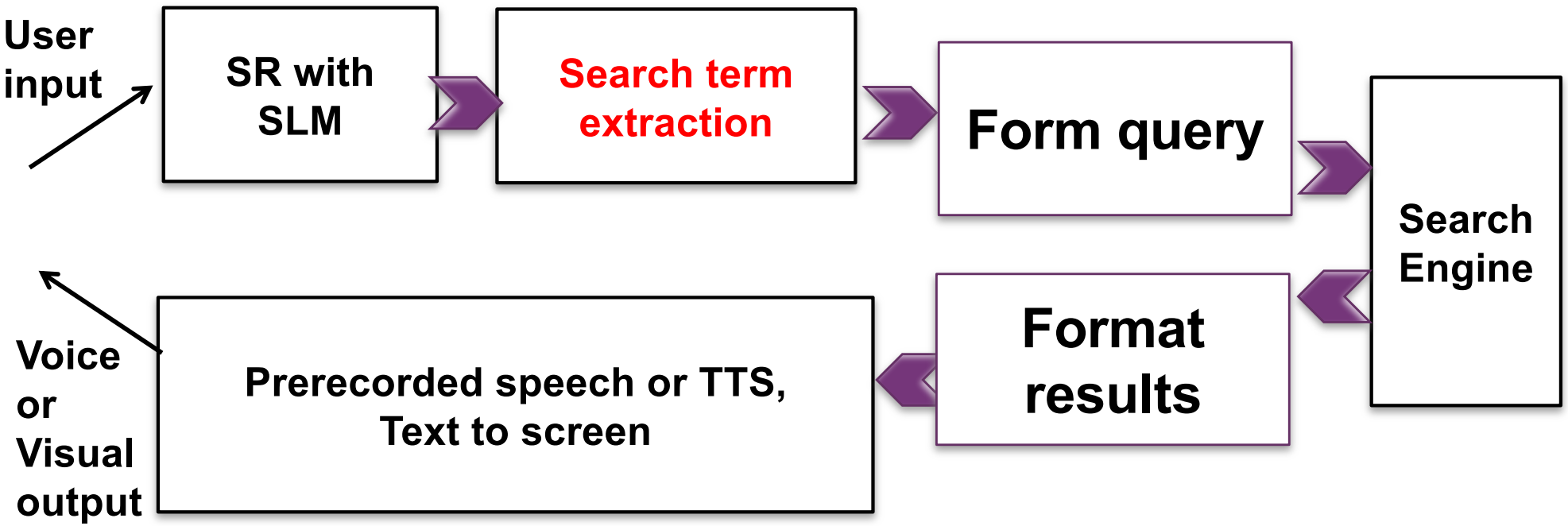
Dialogue System Architecture



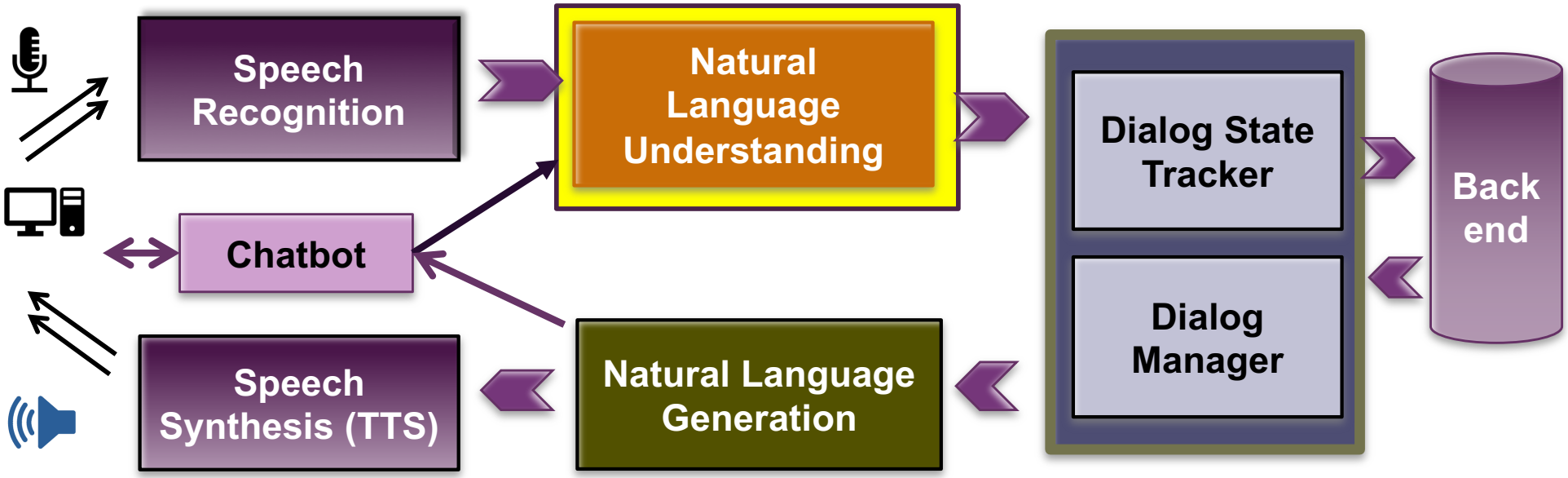
Standard speech “IVR”—automated voice phone interaction



+ Voice Search



Dialogue System Architecture



- Text in, “meaning” out
- Each utterance is handled separately
- Some tools keep “context” to help interpretation

Natural Language Understanding

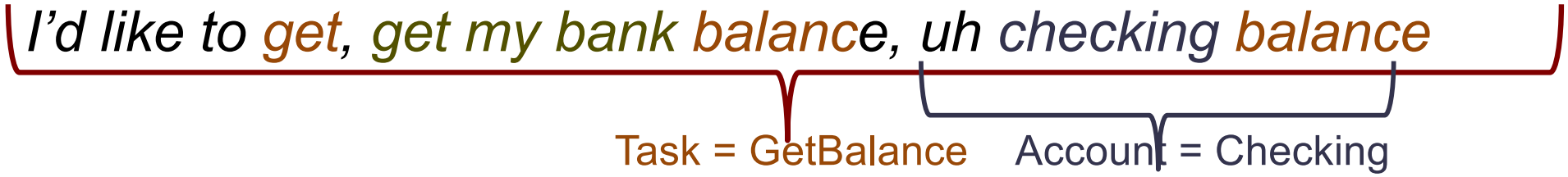
- For speech dialogue systems, most common is "Frame and slot semantics"

Show me morning flights from Boston to SF on Tuesday.

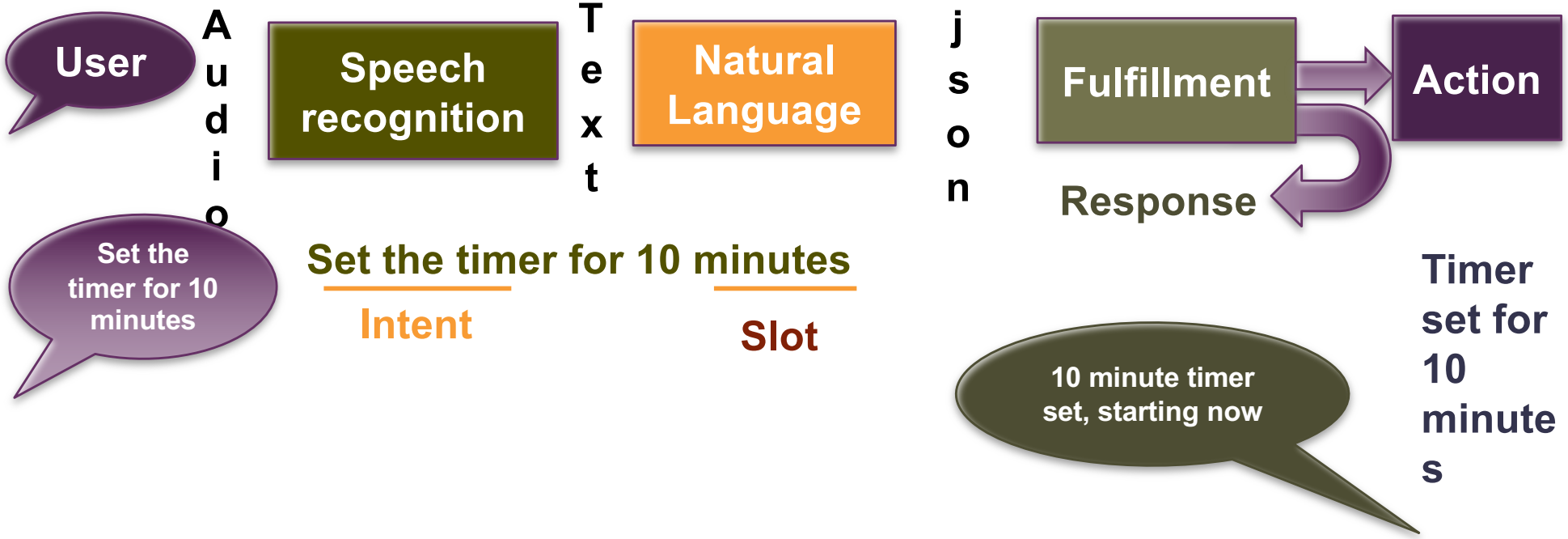
```
SHOW:
  FLIGHTS:
      ORIGIN:
            CITY: Boston
            DATE: Tuesday
            TIME: morning
      DEST:
            CITY: San Francisco
```

Other types of NLU

- Topic identification
  - The first NLU for call routing
    - Briefly tell me the reason for your call today
- Partial parsing
  - Just pick out the phrases that match a grammar
  - "Information extraction"

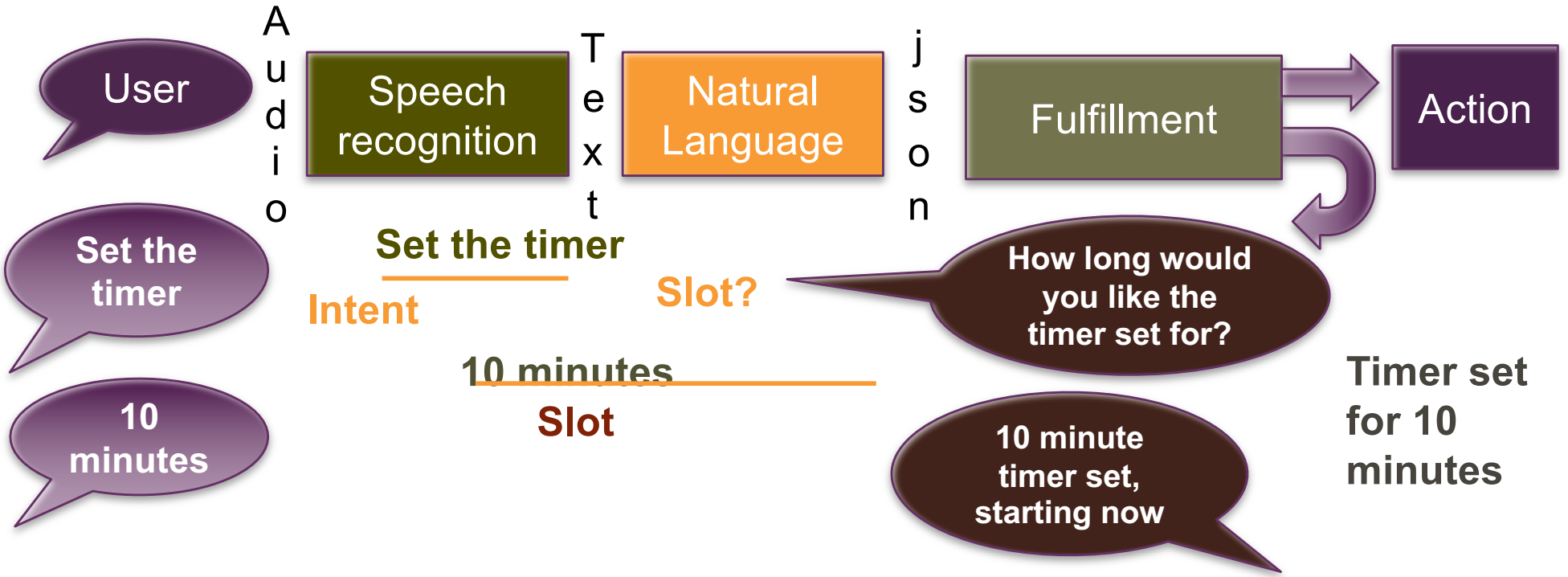


+ Single turn, Single action



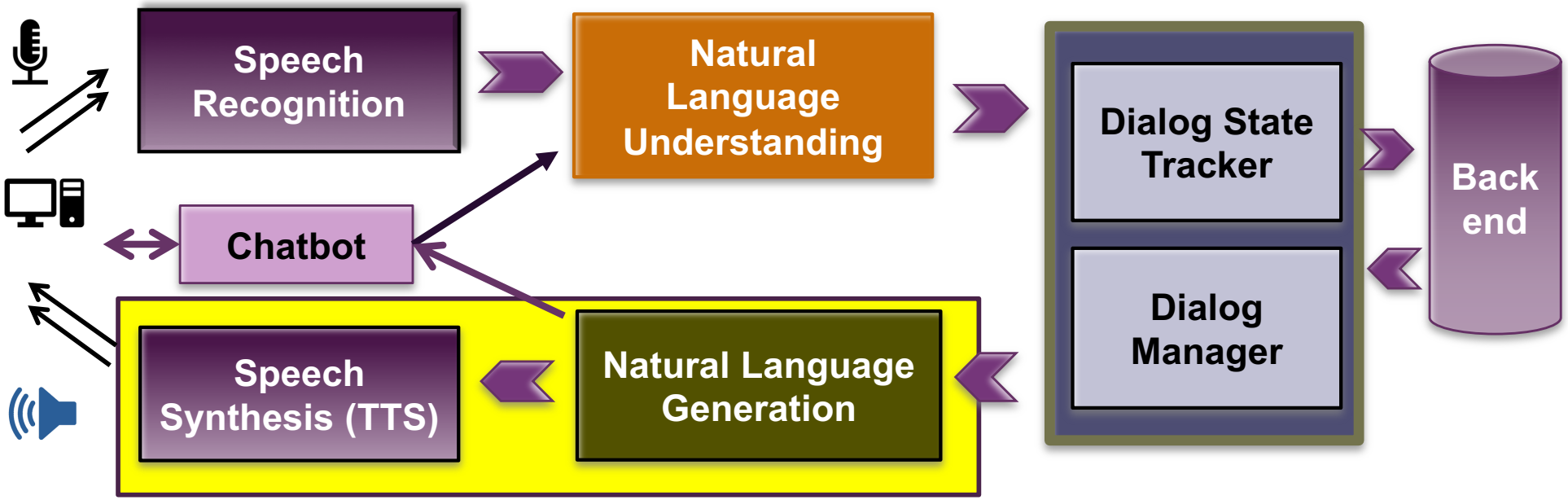
- Voice command containing
  - Intent: What the user wants done
  - Slots: Any other information required

+ Multiple turns, Single action



- **Voice command containing**
  - **Intent: What the user wants done**
  - **Slots: Any other information required**

+ Dialogue System Architecture

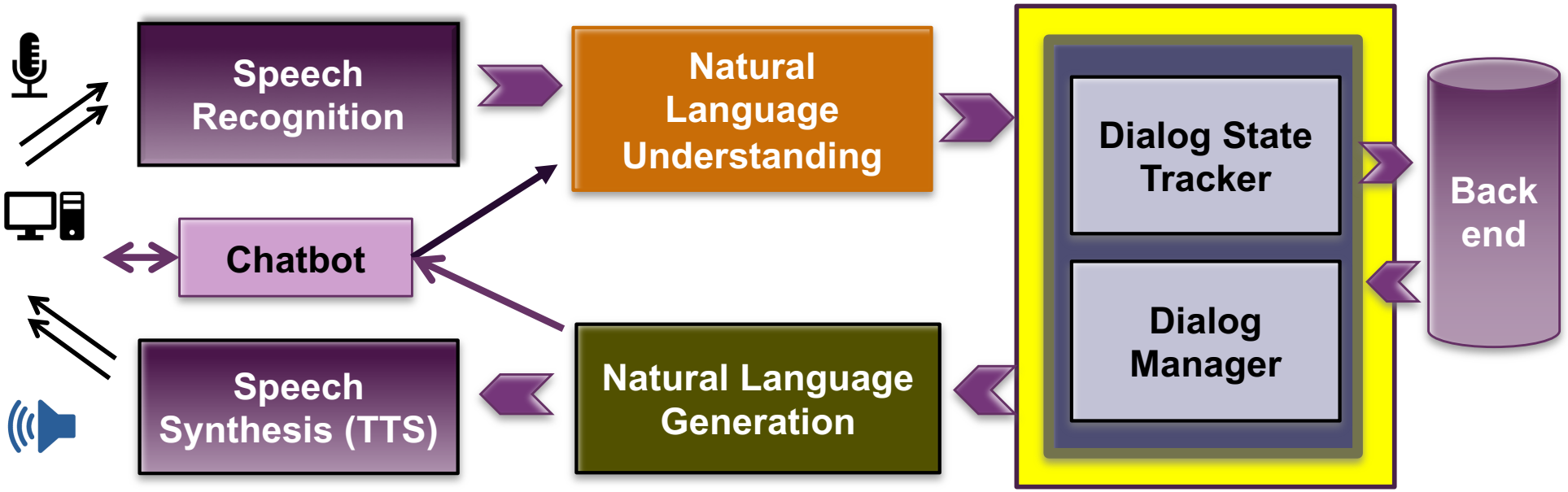


- Controls what goes out to the user
  - Text that will be synthesized (audio)
  - Visuals: Search results
  - Links
- Output may be to some other device (e.g. texting or emailing a link or set of instructions)

Generation and TTS

- Generation component
  - Chooses concepts to express to user
  - Plans out how to express these concepts in words
  - Assigns any necessary prosody to the words
- TTS component
  - Renders words into audio
    - Need to know more than just the word, e.g. read
- In practice both often based on canned sentences

Dialogue System Architecture



- Dialog state
  - Context from
    - Previous utterances
    - User profile
    - Back end
- Dialog manager
  - Takes the utterance & context and decides next actions
    - Interactions with the back end
    - Output to the user

Characteristics of an "Advanced Dialog System"

- Mixed initiative
  - User can change the topic or revisit previous dialog elements
  - System can take control to get more clarifying information
  - User can say anything at any point and get some intelligent response
- Multimodal
  - Input and output can be audio/visual/tactile
  - Issue: processing the timing between speaking & gesturing
- "Natural" interaction, variability of expression
- "Self aware"
  - System identifies the confidence of an interpretation
  - Knows when to ask for clarification and when to move forward
  - Knows when it's being asked something beyond the applications capability
- Helpful
  - Offers additional information or incorporates multiple steps
  - Recognizes

Critical Element: Context

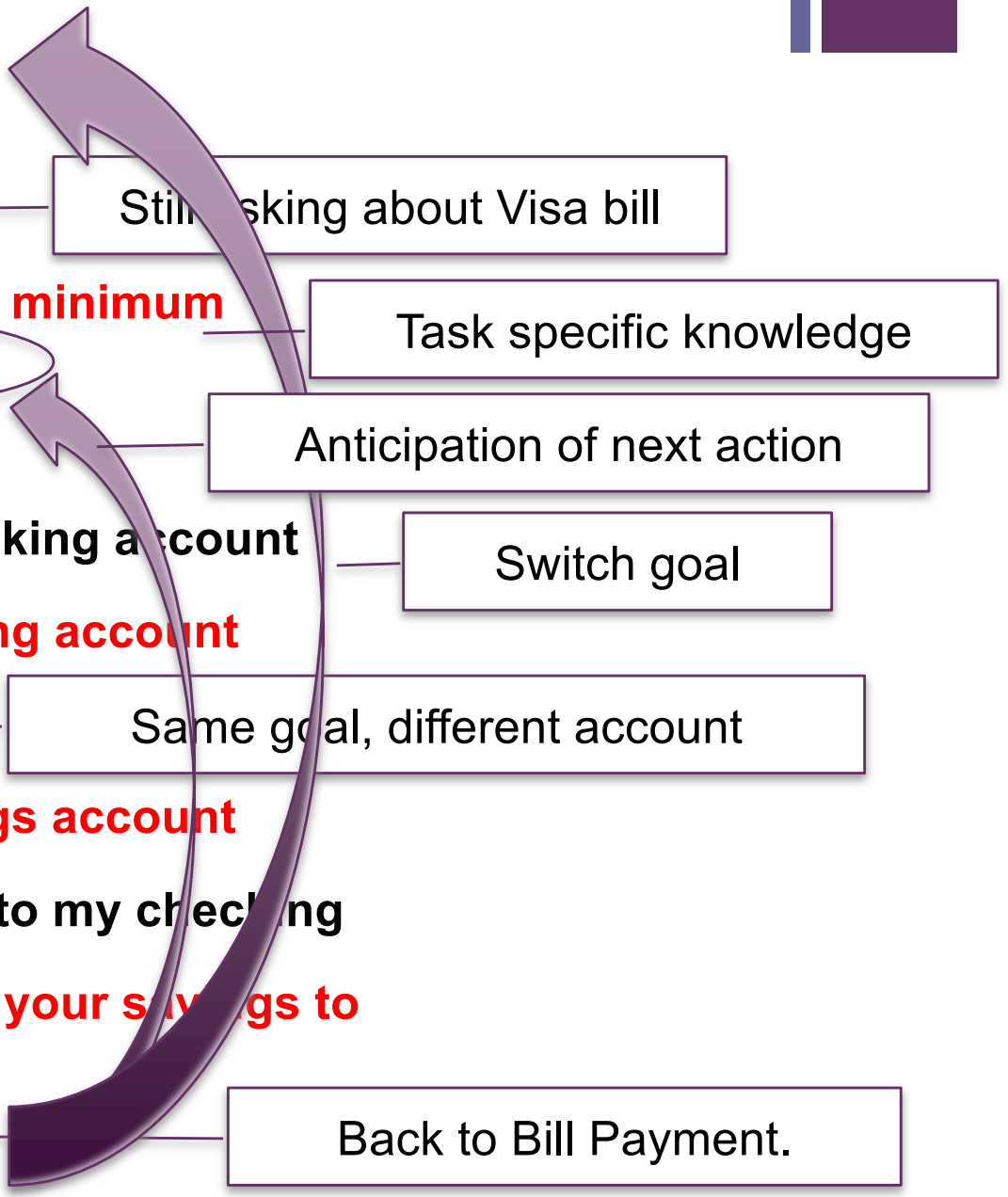
- Context aware
  - “Remembers” what the user has already said and uses that information
  - Recognizes the users’ goals based on what has happened so far
  - Tracks the users “focus” so speakers can refer back to objects already mentioned

What does context give us?

- Narrow the search
  - Restaurants near here.
  - Which ones are open?
- Pointing with words
  - The second one
  - The Indian one
- Tracking objects mentioned already
  - Reuse for new tasks
  - Going back to an unfinished task
- Track user goals to understand tasks in context
- Ability to “recover” when user is misunderstood or user changes his/her mind
  - Guide “clarification” dialogs

Example

- When is my Visa bill due
- Monday, April 15th
- How much do I owe?
- Your balance is $583.24 and the minimum payment is twenty five dollars.
- Would you like to pay your bill?
- How much do I have in my checking account
- You have 164.25 in your checking account
- How about my savings
- You have 2019.97 in your savings account
- Transfer $500 from my savings to my checking
- $500 has been transferred from your savings to your checking account
- OK. Now pay the minimum

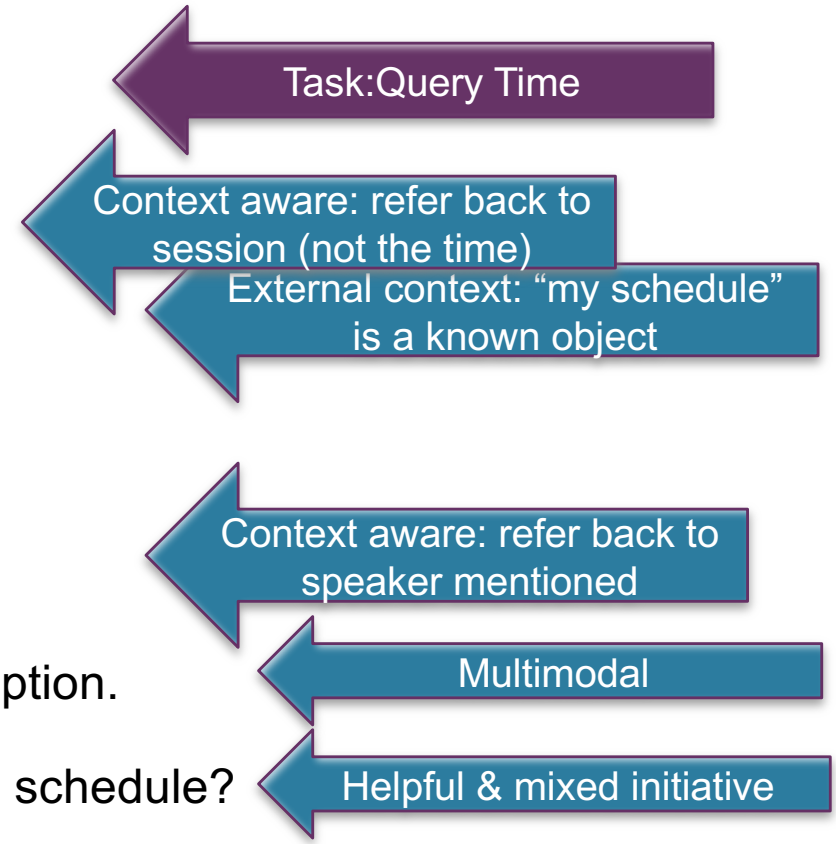


What's in the "Discourse Manager" box?

- Context created by the discourse so far
  - Allows "abbreviated" sentences
    - How much do I owe **[on my visa bill]**
- Expectation of what might happen next
  - Task representation with sequences of events (Agenda)
  - Proactively offer of next steps
- Knowledge of the tasks the system is capable of
  - Connection of tasks to backend systems
  - Representation of what can't be done
- Knowledge of the input and output modalities and how they relate
  - Need to know what's in the list to act on "next one"

Examples of where we want to go

- U: When am I speaking?
- S: 2 pm on Tuesday
- U: Put **that** on my schedule
- U: Is anyone from Sensory speaking?
- S: Yes. Jeff Rogers and **Todd Mozer**
- U: What is **Todd** speaking on?
- S: **He'**s on the panel. Here's the description.
- S: Would you like that on your personal schedule?
- U: Thanks. **How about** Jeff?



Continued Examples

- U: **How about** Jeff?
- S: Jeff Rogers is speaking on Truly Handsfree **at 11 am on Tuesday**.
- U: (notices time) Do I have a session **scheduled** now?
- S: No, until 11 am.
- U: **OK**. What's in the **other** track?
- S: Here are the talks in the Business track at 11: <shows results>
- U: I'd like to hear **<Pat>.** Schedule me for **that one**.
- S: ?? **No "Pat" in that session**
  - Ask to repeat
  - Ask if user means Matt?
  - Ignore misreco and execute command
- S: It's 11 o'clock. Time for your next session, which is in Ballroom B

Context Aware

Helpful: Provide additional info

Context: Return to previous task

Discourse context "Other"

Self aware: Misrecognition

Make choices

Mixed Initiative, helpful

Integrated components

- Not a strictly sequential process
  - Components constrain and feed each other