NORSK REGNESENTRAL
NORWEGIAN COMPUTING CENTER

Dialogue systems (3)

Pierre Lison
plison@nr.no

IN4080, Autumn 2025, video lecture

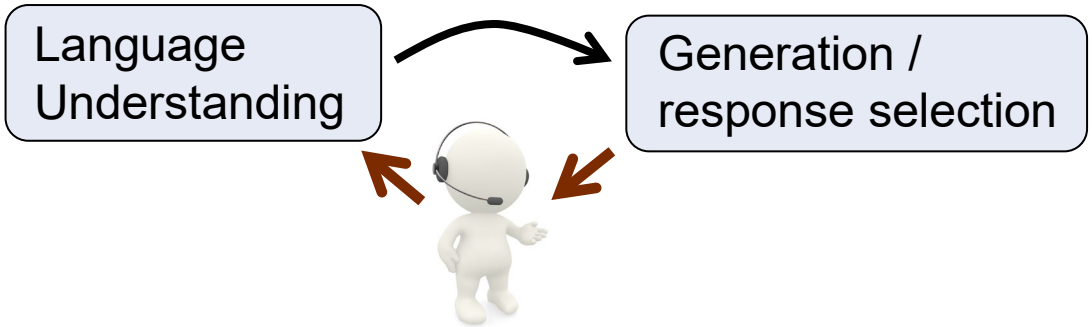
Week 43

- The second lecture on dialogue systems will be done **remotely**, through a series of shorter videos
- Don’t hesitate to reach out to me (plison@nr.no) or on our Discourse channel if you have any questions!

Plan:

- Video 1: generative dialogue models
- **Video 2: dialogue management**
- Video 3: system development and evaluation

Limitation of chatbot models

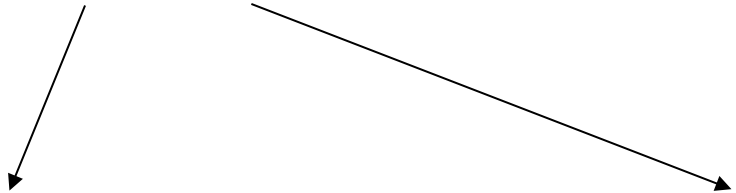


- Chatbots are typically built for *short interactions*
  - Often an initial question/request + perhaps some follow-up response(s)
- Can we build systems able to handle *longer* and more *complex* interactions?

Many *dialogue turns*, lists of tasks to complete, etc.

Rich external context, multi-party dialogues, etc.

Dialogue management tasks



Keeping track of the dialogue state over time (**dialogue state tracking**)

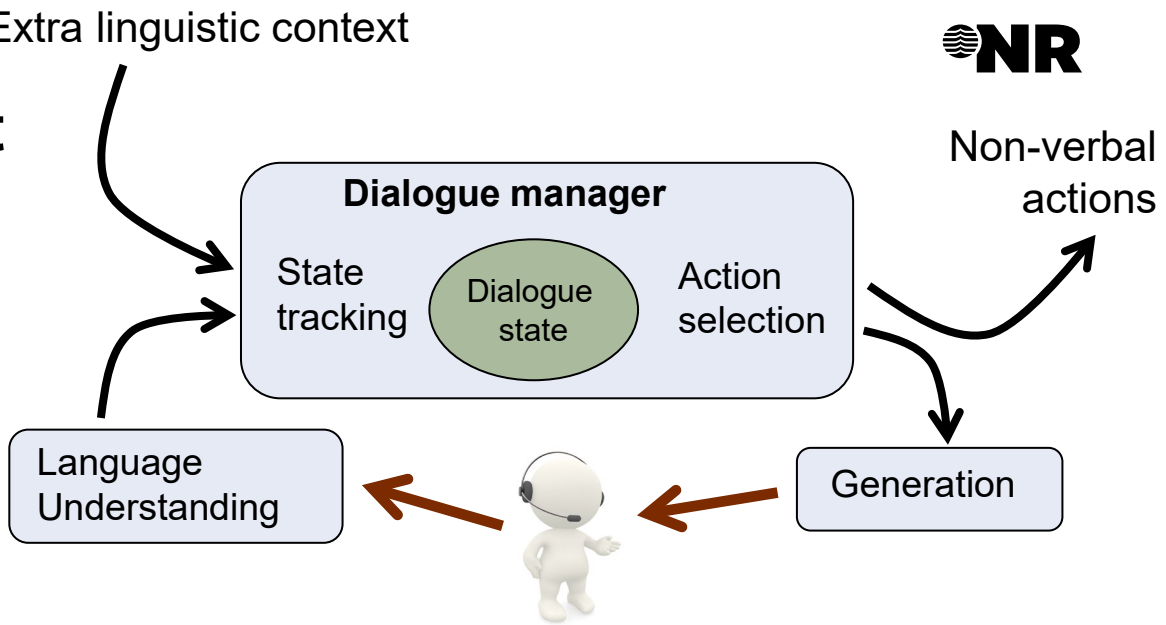
Often done by defining the *variables/slots* that are relevant for the task and *updating* those based on incoming *inputs/observations*

Selecting what to say/do next based on this state (**dialogue policy**)

Decision-making under uncertainty: find the best way to fulfill the system goals, taking uncertainties into account (misunderstandings, unclear intents, ambiguous context, etc.)

Dialogue management

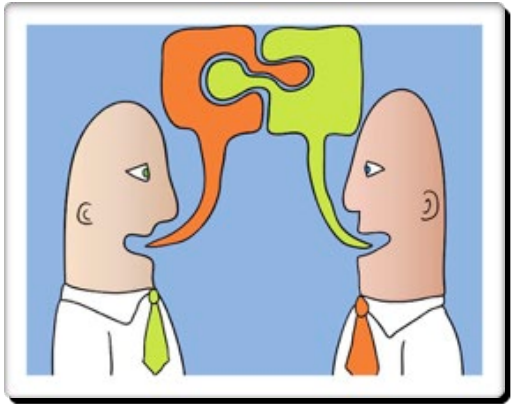
- Dialogue management can be implemented via a *separate module* that mediates understanding and generation



- … or, as is increasingly the case nowadays, integrated into generative models through *prompting* and *LLM chaining*
  - One can use a first LLM for state tracking, and insert its output in the prompt of another LLM employed to generate the response

Dialogue management

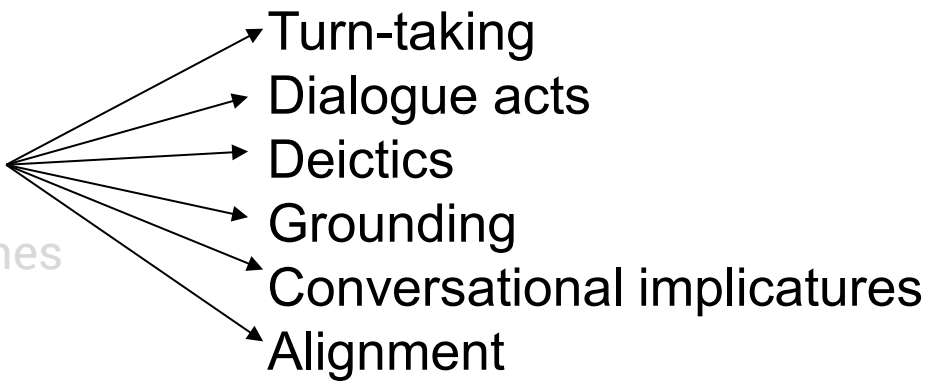
- The role of the dialogue manager is to control the *dialogue flow*
  - *What* to say/do and *when* to say/do it
- Conversational skills to emulate:
  - Interpret utterances *contextually*
  - Manage *turn-taking*
  - Fulfill *conversational obligations* & *social conventions*
  - Manage *uncertainties*



➔ **What are those skills, and how are humans using them?**

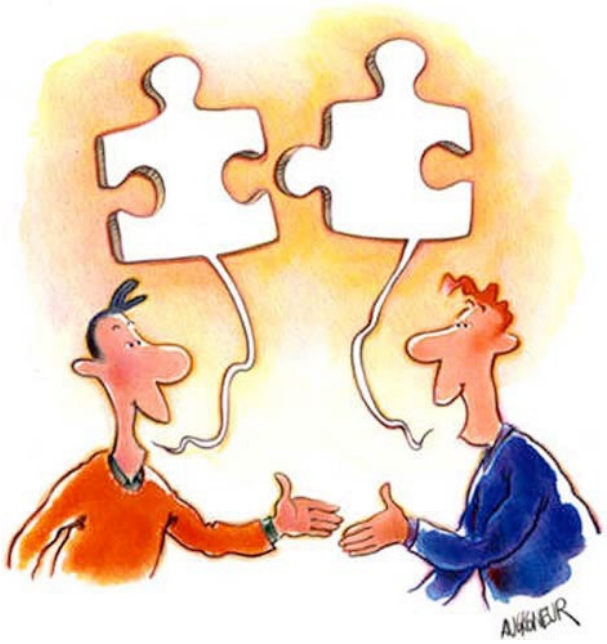
Outline

- Intro to dialogue management
- **Properties of human dialogues**
- Dialogue management approaches



What is dialogue?

- Spoken (“verbal”) + possibly non-verbal interaction between two or more participants
- Dialogue is a *joint*, *social activity*, serving one or several purposes for the participants
- What does it mean to view dialogue as a **joint activity**?



Turn-taking

- Dialogue participants take *turns*
  - **Turn** = continuous contribution from one speaker
  - Turn-taking is a *resource allocation problem*
- Surprisingly fluid in normal conversations:
  - Minimise both gaps (no speaker) and overlaps (more than one speaker)
  - Interval between speakers is around 250 ms

[Duncan (1972): «Some Signals and Rules for Taking Speaking Turns in Conversations», in *Journal of Personality and Social Psychology*]

Turn-taking

- How are turns taken or released?
- Markers for turn boundaries:
  - Complete syntactic/semantic unit?
  - Dialogue structure (greetings → greetings, question → answer)
  - Intonation (falling intonation → speaker is finished)
  - Non-verbal cues (eye gaze, gestures)
  - Silence & hesitation markers (unfilled pauses ≠ filled pauses)
  - Social conventions





Example of turn-taking

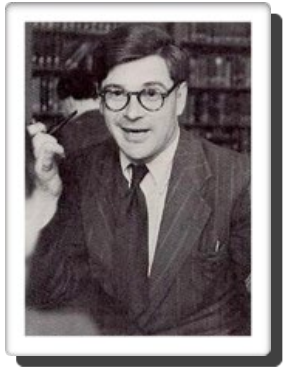
Speaker 1:	han vil bo i skogen ?
Speaker 2:	# altså hvis jeg hadde kommet og sagt " skal vi flytte i skogen ? " så h adde han sagt ja
Speaker 1:	mm
Speaker 2:	men jeg vil ikke bo i skogen
Speaker 1:	nei det skjønner jeg
Speaker 2:	så vi må jo finne et sted som er mellomting og det jeg vil ikke bo utp å landet # i hvilken som helst (uforståelig) ...
Speaker 1:	* men det kommer jo an på hvor i skogen da

[«*Norske talespråkskorpus - Oslo delen*» (NoTa), collected and annotated by the Tekstlaboratoriet]

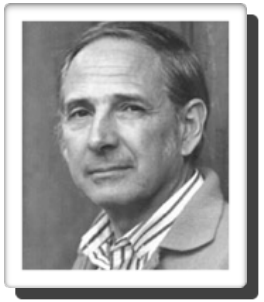
Dialogue acts

Each utterance is an *action* performed by the speaker

- The speaker has a specific **goal**
  (which might be only to establish or maintain *rapport* with the listeners)
- The utterance produces specific **effects** upon the listeners, or the world at large
- «*Language as action*» perspective



J.L. Austin (1911-1960)
philosopher of language



J. Searle (1932, -)
philosopher of language

[J. L. Austin (1955), *How to do things with words*.]

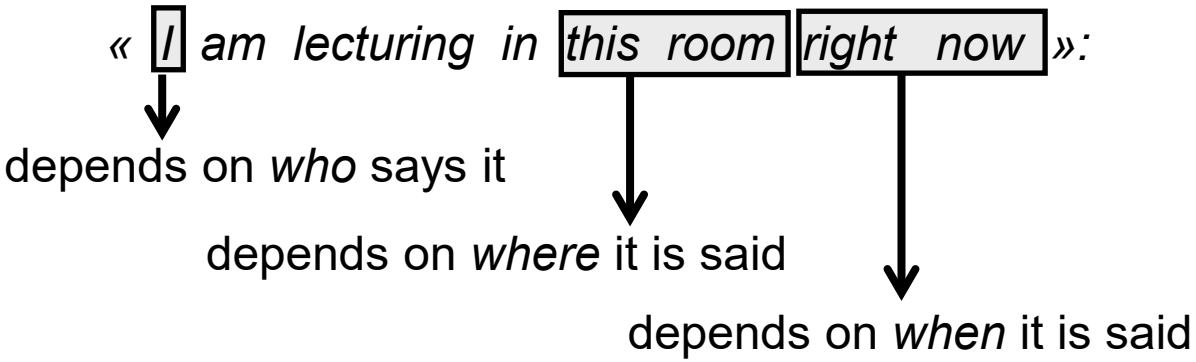
Dialogue acts



- The mother reaction has a specific **purpose**
  - Communicating her suprise/anger, and stop Calvin
- Her question will trigger some **effects**:
  - A psychological reaction from Calvin (e.g. surprise)
  - Possibly a real-world effect as well (Calvin stopping his action)

Deixis

- Dialogue often *referential* to a **spatio-temporal context**
- Such references are called deictics
  - Related concepts: *indexicals*, *anaphora*
- The meaning of a deictic depends on the context in which it is uttered (including the speaker perspective)



Deictics can be:

Pronouns

Adverbs of time and place

Demonstratives

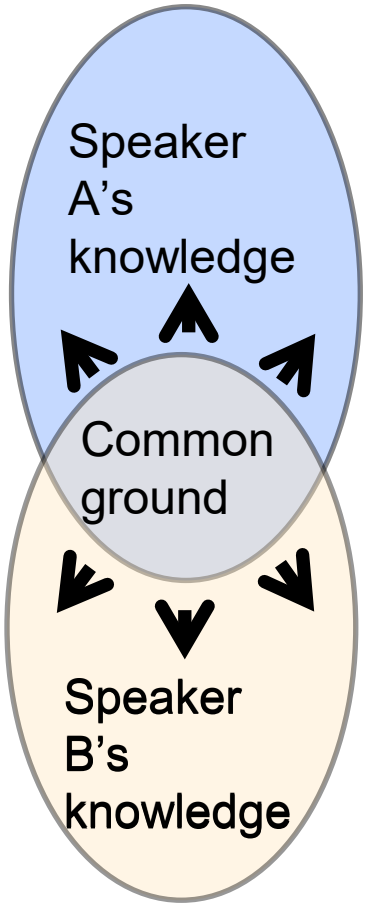
Tense markers

Non-verbal signs

…

Grounding

- Dialogue is a *joint, collaborative process* between the participants
  - Need to ensure mutual understanding
- Gradual expansion and refinement of **common ground**
  - Common ground = shared knowledge



[H. H. Clark and E. F. Schaefer (1989), «Contributing to discourse», in *Cognitive Science*]

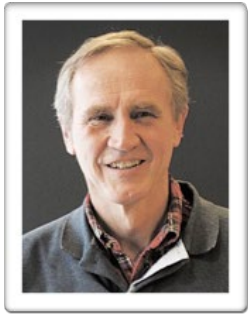
Grounding

- Grounding is the process of gradually augmenting the common ground during the interaction
  - Variety of signals and strategies
- Multiple levels:
  - Contact (attention to interlocutor)
  - Perception (detection of utterance)
  - Understanding (comprehension of utterance)
  - Attitudinal reactions

[Jens Allwood (1992), «On discourse cohesion», in *Gothenburg papers in Theoretical Linguistics*.]



Herbert H. Clark
psycholinguist



Jens Allwood
linguist

Grounding acts

- Backchannels: «*uh-uh*», «*mm*», «*yeah*»
- Explicit feedback: «*ja det skjønner jeg*»
- Implicit feedback: A: «*I want to fly to Rome*» → B: «*there are two flights to Rome on Wednesday: ...* »
- Clarification strategies: «*Did you mean to Rome or to Goa?*», «*could you confirm that ...*»
- Repair strategies: «*OK, you're not going to Goa. Where do you want to go then?*»

Examples of grounding

Speaker 1:	vi vasker den hver dag vi # vi har mopp
Speaker 2:	**mm** ## **ja** det er fort og faren til M27 legger nytt teppe han # det er gjort på to timer ## så det er fort gjort
Speaker 1:	**ja** ## da er ikke noe sak
Speaker 2:	vi har skifta teppe tre ganger allerede han gjør det gratis
Speaker 1:	**hæ ?**
Speaker 2:	vi har skifta teppe tre ganger og # han han ...
Speaker 1:	* **jeg skjønner ikke** hvorfor dere har teppe
Speaker 2:	jeg syns det var rart jeg **òg** # men e # (sibilant)



[«*Norske talespråkskorpus - Oslo delen*» (NoTa),
collected and annotated by the Tekstlaboratoriet]

Examples of grounding

Speaker 1:	e # **nei** det er ikke mange
Speaker 2:	**ja * nei**
Speaker 1:	men heldigvis så var ikke Petter Rudi tatt ut denne gangen da
Speaker 2:	**ja** # jeg skjønner ikke hva han skal på landslaget å gjøre
Speaker 1:	*** nei han har ingen ting på landslaget**
Speaker 2:	**nei** # definitivt
Speaker 1:	å gjøre # han er ubrukelig
Speaker 2:	* moldensere
Speaker 1:	**hm?**
Speaker 2:	ja disse moldenserne
Speaker 1:	**en gang til?**
Speaker 2:	disse moldenserne
Speaker 1:	* **å ja** (fremre klikkelyd) # unnskyld # jeg hørte ikke hva du sa

implicit feedback
(repetition of *landslaget*)

clarification requests

[«*Norske talespråkskorpus - Oslo delen*» (NoTa), collected and annotated by the Tekstlaboratoriet]

Conversational implicatures

- Very often, part of the meaning of utterance is not explicitly stated, but only implied

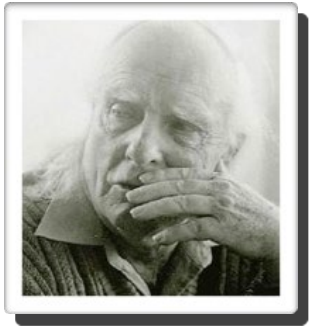
A: «Is William working today?»

B: «He has a cold»

- How can we retrieve this «suggested» meaning, and go beyond literal interpretations?
  - Need to make some *assumptions* about the speaker to help us infer the hidden part

Conversational implicatures

- Same idea again: dialogue as a **collaborative process**
- Grice's *Cooperative Principle*:
  - Maxim of Quality: «be truthful»
  - Maxim of Quantity: «be exactly as informative as required»
  - Maxim of Relation: «be relevant»
  - Maxim of Manner: «be clear»



Paul Grice (1913-1988) philosopher of language

1. We assume our conversational partners adhere to those maxims
2. If an utterance initially seems to violate a maxim, we (unconsciously) infer *additional hypotheses* to make sense of our partner's contribution

[Paul Grice (1975), *Logic and Conversation*.]

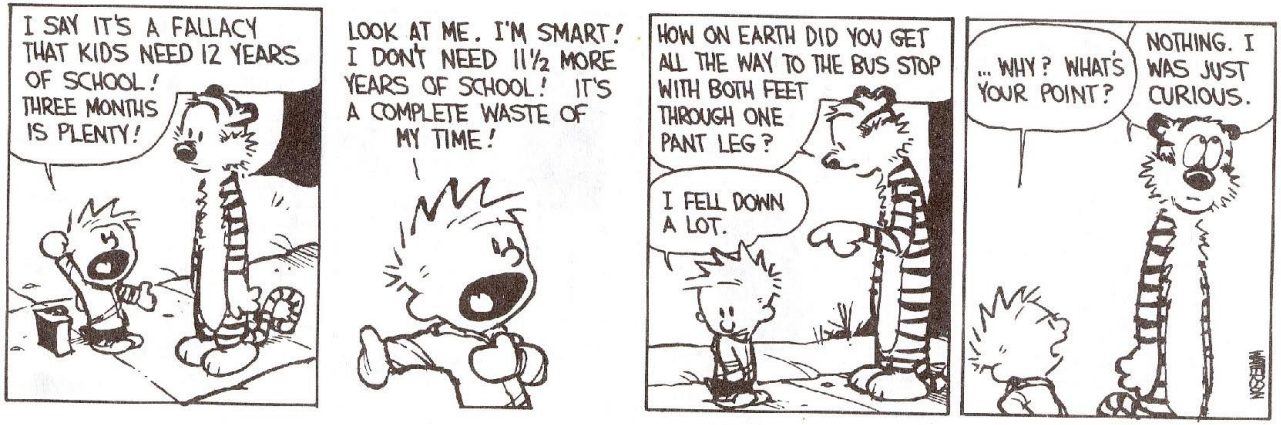
Conversational implicatures

A: «Is William working today?»

B: «He has a cold»

- At first glance, B seems to violate the maxim of relevance - he does not directly answer A’s question
- But looking at the utterance more closely, we can read it as implying that (due to his cold) he is probably at home, and thus not working today
- This is because we assume that B is cooperative and wouldn’t have uttered «he has a cold» if it didn’t help answering A’s question

Conversational implicatures



Hobbes' question is *suggesting* something about Calvin's need for schooling, without stating it explicitly

We can understand it because we assume that Hobbes' contribution is cooperative and thus relevant to the discussion

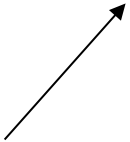
Alignment

- Participants in a dialogue continuously **align** their mental representations
  - Notion of *common ground* discussed earlier
- But dialogue participants also align at other levels, by unconsciously *imitating* each other
- As the interaction unfolds, the participants automatically align their *wording*, *pronunciation*, *speech rate*, and *gestures*

[Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*]

Outline

- Intro to dialogue management
- Properties of human dialogues
- **Dialogue management approaches**

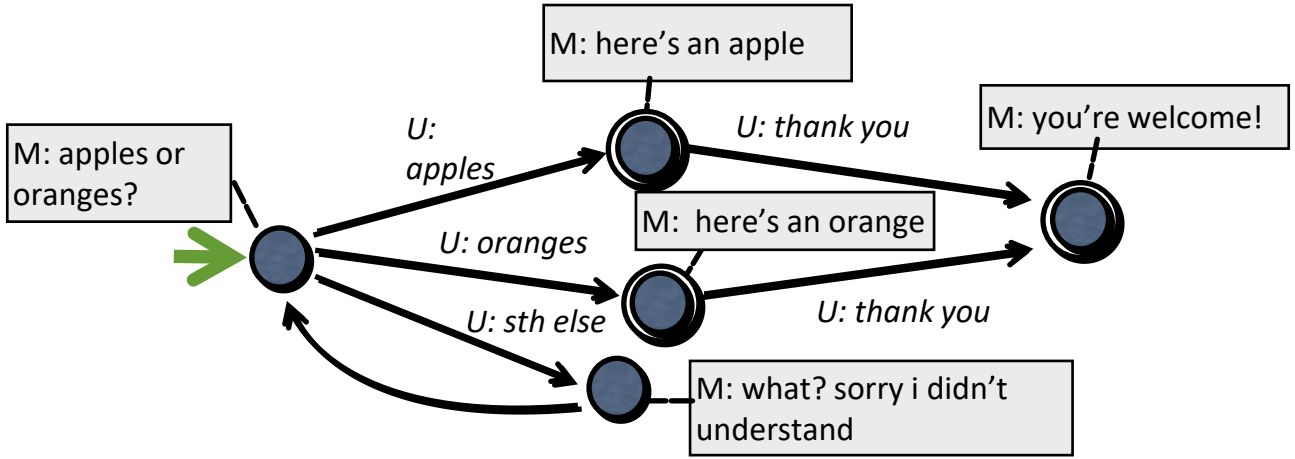


Flowcharts
Data-driven techniques
(based on either supervised or reinforcement learning)

Flowcharts

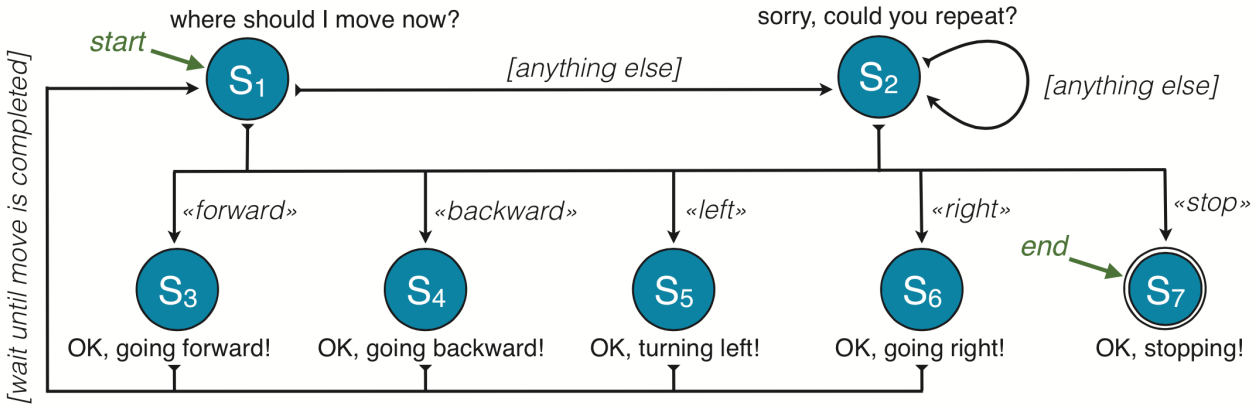
= encode dialogue strategies as **finite-state automata**

- the nodes represent *machine actions*
- and the edges possible (mutually exclusive) *user responses*



Flowcharts

- Transitions can relate to other signals than user inputs (for instance, external events)
- And can also express complex conditions (pattern matching on the user input, confidence thresholds, etc.)



Interaction style

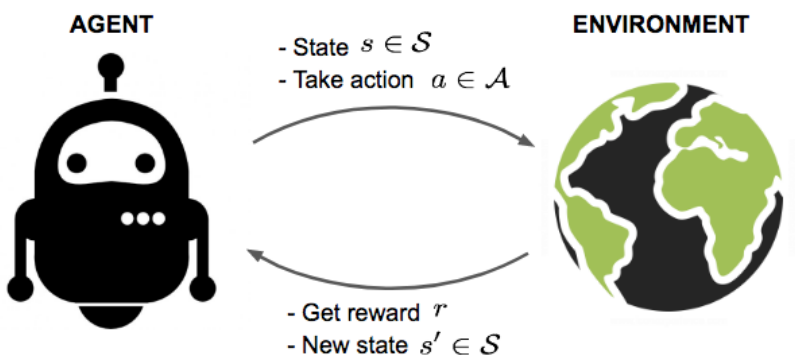
- Rigid, repetitive structure of the interaction
- Irritating confirmations & acknowledgements
  - No good management of uncertainties
- No user or context adaptivity



“Saturday night live” sketch comedy, 2005

Data-driven techniques

- Can we optimize dialogue strategies from data? **Yes!**
- Several approaches:
  - **Supervised learning**: imitate «good dialogues» from experts
  - Learn from **human preferences** (using e.g. Direct Preference Optimization)
  - Learn dialogue strategies through **Reinforcement Learning**



Reward model

- Interactive tasks must often encode trade-offs between *multiple competing objectives*
  - A flight booking system must book the right ticket
  - But it must do so with the fewest number of requests
- Typically encoded via **rewards** (utilities) associated to particular state/action pairs

State	Action	Reward
User wants to book ticket *x*	Booking x	+10
User wants to book ticket *x*	Booking y ≠ x	−30
User wants to book ticket *x*	Clarification request	−1

Policy optimisation

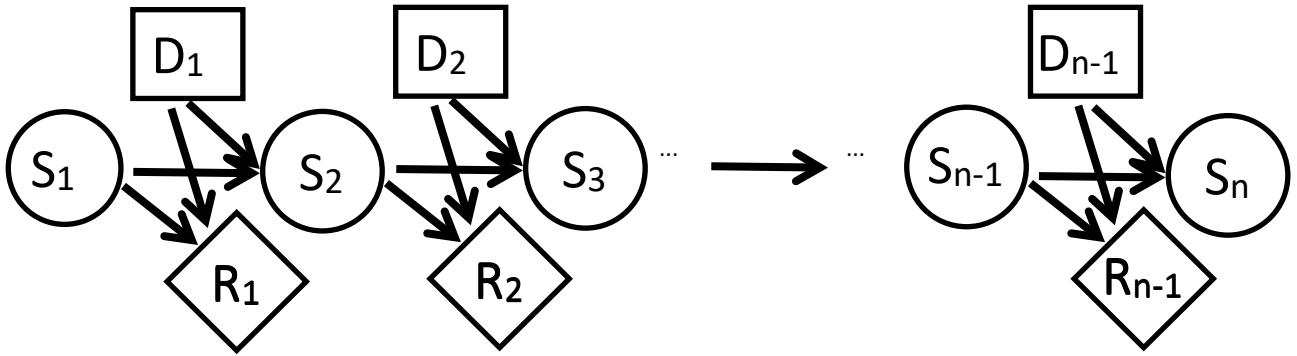
- Given a reward model, we can **optimize** a model for selecting the *best action* to execute at any given time
- This works by letting the system *interact* many times with its «environment» and gradually improve its action selection based on the rewards it receives
  - Here, the «environment» = the user
  - This user may be a real human, but in most cases, it is easier to develop a *simulated user*

The action may be a high-level command («ConfirmPurchase») or, for generative models, the next token to produce

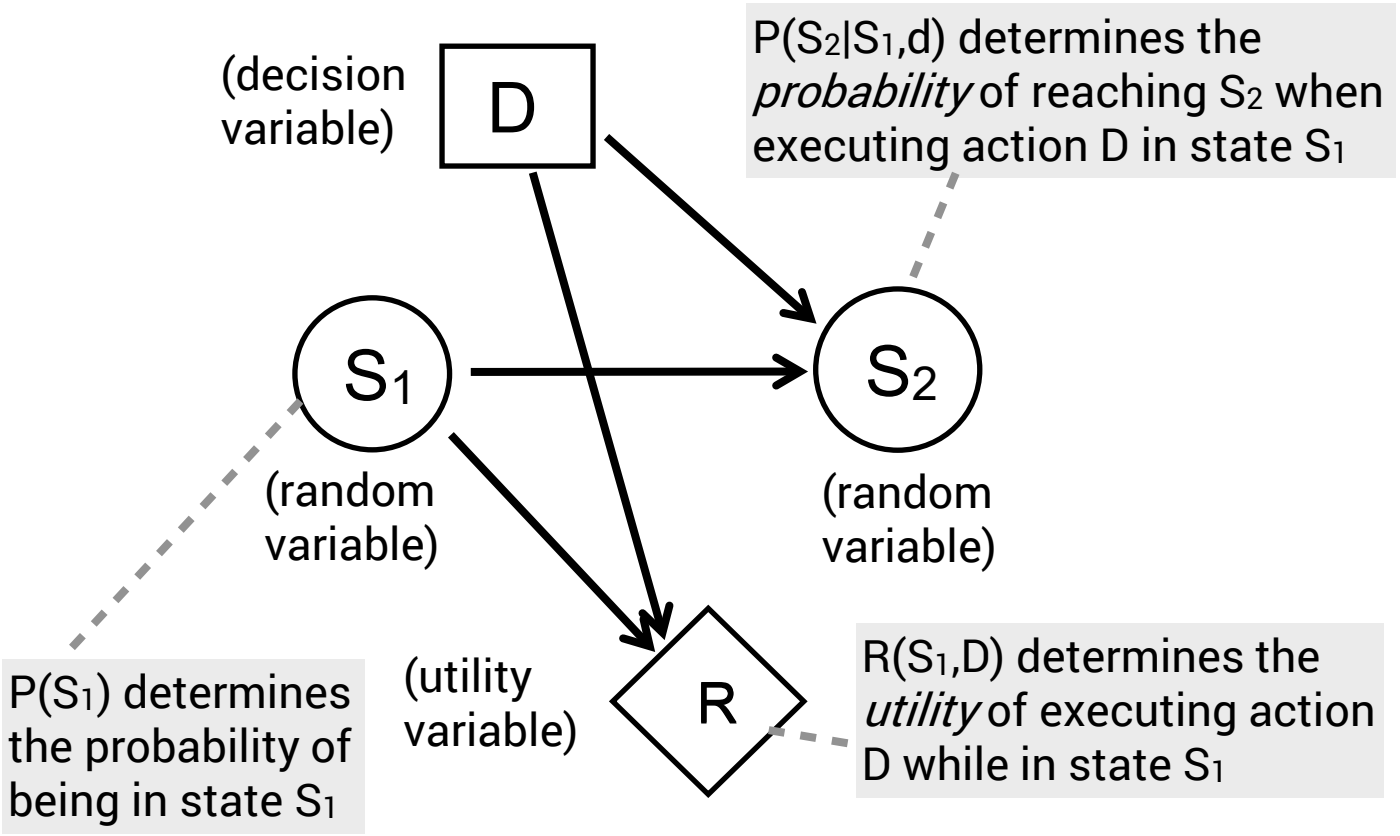
The technical name for this action selection process is a **policy** (mapping from states to actions)

Markov Decision Processes

- We can define these ideas more precisely using a formalism called **Markov Decision Processes** (MDPs)
- Markov Decision Processes are an extension of Markov Chains where the agent *selects an action* at each state
  - This action will then modify the state space
  - And will yield a particular reward for the agent



Graphical notation



Markov Decision Processes

A MDP is as a tuple **<S,A,T,R>**, where:

- **S** is the *state space* (possible states in the domain)
- **A** is the *action space* (possible actions for the agent)
- **T** is the *transition function*, defined as $T(s, a, s') = P(s'|s, a)$. It is the probability of arriving to state **s'** after executing action **a** in state **s**.
- **R** is the *reward function*, defined as $R : S \times A \rightarrow R$. It is a real number encoding the utility for the agent to perform action **a** while in state **s**.

Expected cumulative reward

- In an MDP, the agent seeks to maximise its *expected cumulative reward* **Q(s,a)**

The rewards accumulate over time

The agent must try to predict future inputs/rewards

- How much worth is a reward expected at time (t+i) compared to one received right now?
  - We use a *discount factor* γ to capture this balance
  - Related to *delayed gratification* in psychology

Bellman equation

The *Bellman equation* tells us that we can write the expected cumulative reward Q in a recursive fashion:

$$Q(s,a) = R(s,a) + \gamma \sum_{s'} P(s'|s,a) \max_{a'} Q(s',a')$$

Notice that we are estimating the Q-values based on... our estimation of the Q-values (can be used to iteratively refine these estimates until convergence)

[R. Bellman (1957): «*Dynamic Programming*»]

MDP policy

- Given an MDP, a (dialogue) policy tells us which action to execute in each state
- A dialogue policy is a *mapping* π: S → A from states to actions
- An *optimal* dialogue policy π* is a policy that always outputs the action yielding the maximum expected cumulative reward:

$$\pi^*(s) = \underset{a}{\operatorname{argmax}}\, Q(s, a)$$

Types of RL techniques

- *Model-based* vs. *model-free* methods
  - *model-free*: Directly estimate the policy (or Q-values) from data
  - *model-based*: Estimate the MDP model from interaction data, and then derive the policy from it

- As the state-action space is high-dimensional, the policy or Q-value model is typically expressed as a deep neural network
- Many techniques exist, but a popular choice in NLP is *Proximal Policy Optimization* (PPO)

Schulman, et al (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

R. Sutton & A. Barto (2018): «*Reinforcement Learning: An Introduction*» (complete book available online!)

Challenges



- **Real** or **simulated** interactions?
  - Simulated users often necessary due to slow convergence speed
  - However, this may lead to poor behaviours when applied to *real users*
- **Reward design**
  - What constitutes a “good” dialogue?
  - Possible to learn a reward model from *human preferences* … but we might as well then directly learn from those?

Next video

- In the next video, we will talk about how to
  - Practically design and develop conversational systems
  - Evaluate their performance of those systems



- Don’t hesitate to get in touch if you have any questions!