Learning Dialogue Strategies with a Simulated User

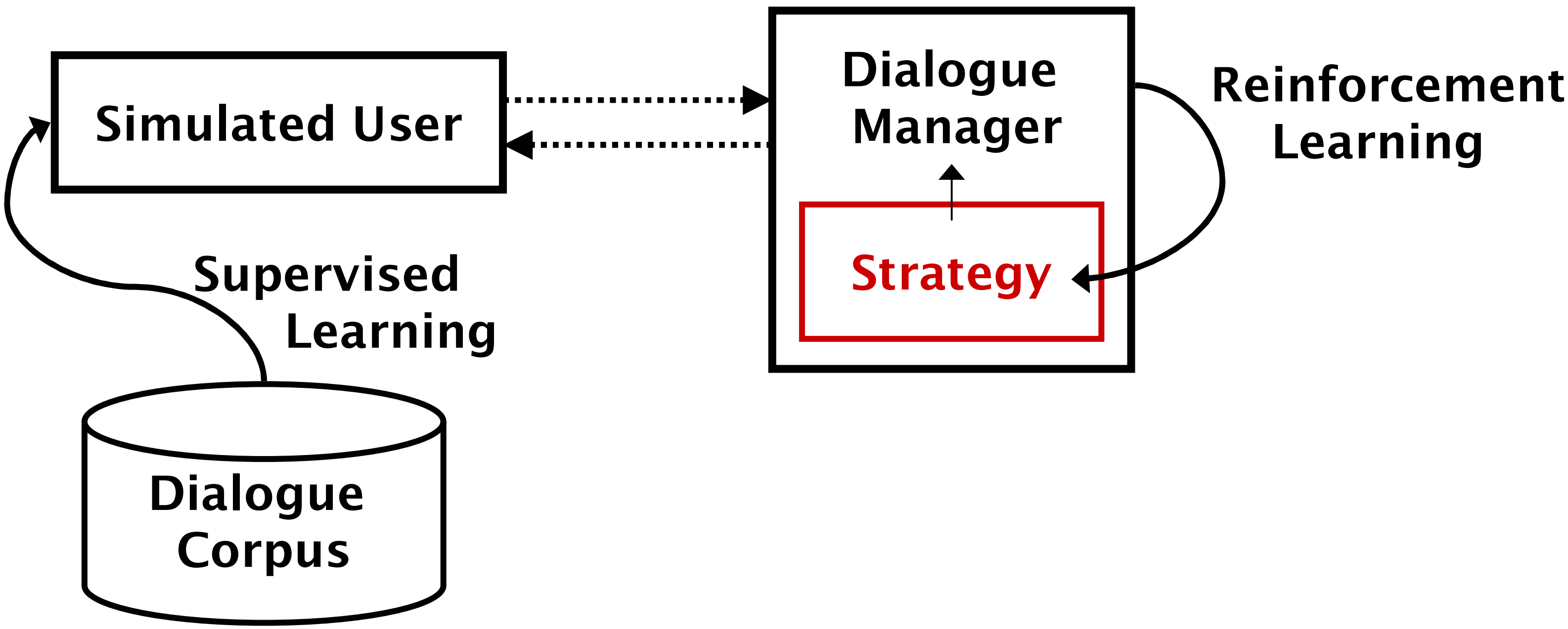
Jost Schatzmann and Steve Young

Cambridge University Engineering Department
Trumpington Street, Cambridge CB21PZ, UK
{js532, sjy}@eng.cam.ac.uk

Dialog on Dialogs Meeting
Carnegie Mellon University, 19 August 2005

User Simulation-Based Learning

- **Learn dialogue strategies through trial-and-error interaction with with a simulated user**



Agenda

- **Work on Evaluation: Experiments and Results**
- **Agenda-based User Modelling**

Research Questions

- How good are the currently available simulation techniques? Can they...
  - produce human-like behaviour?
  - cover the variety of real user behaviour?

SIGdial paper

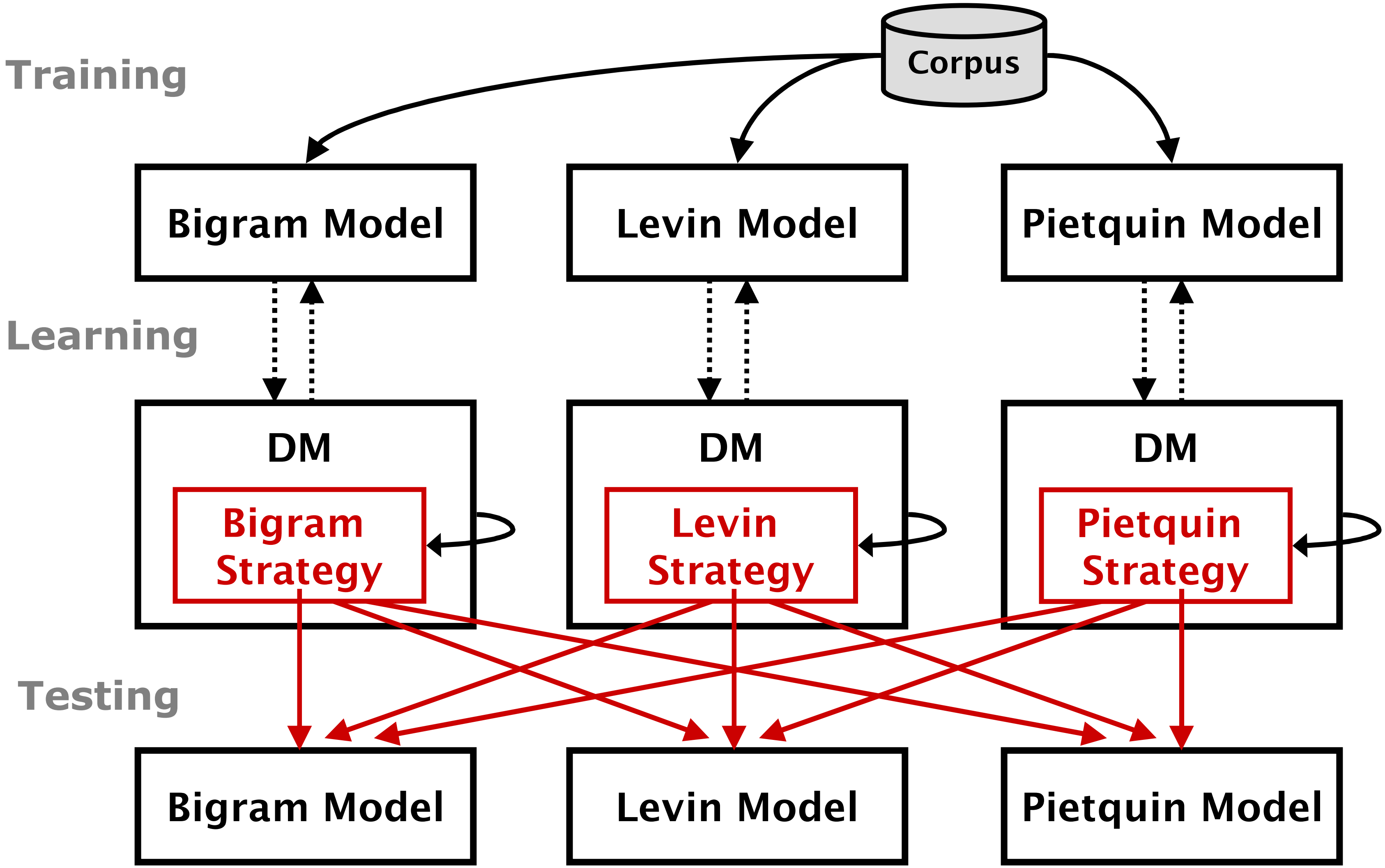
- What is the effect of the user model on the learned strategy?
  - Influence on strategy performance?
  - Influence on strategy characteristics?
  - Are the strategies merely fitted to a particular UM?
  - Can we find UM-independent forms of strategy evaluation?

ASRU paper

User Modelling Techniques

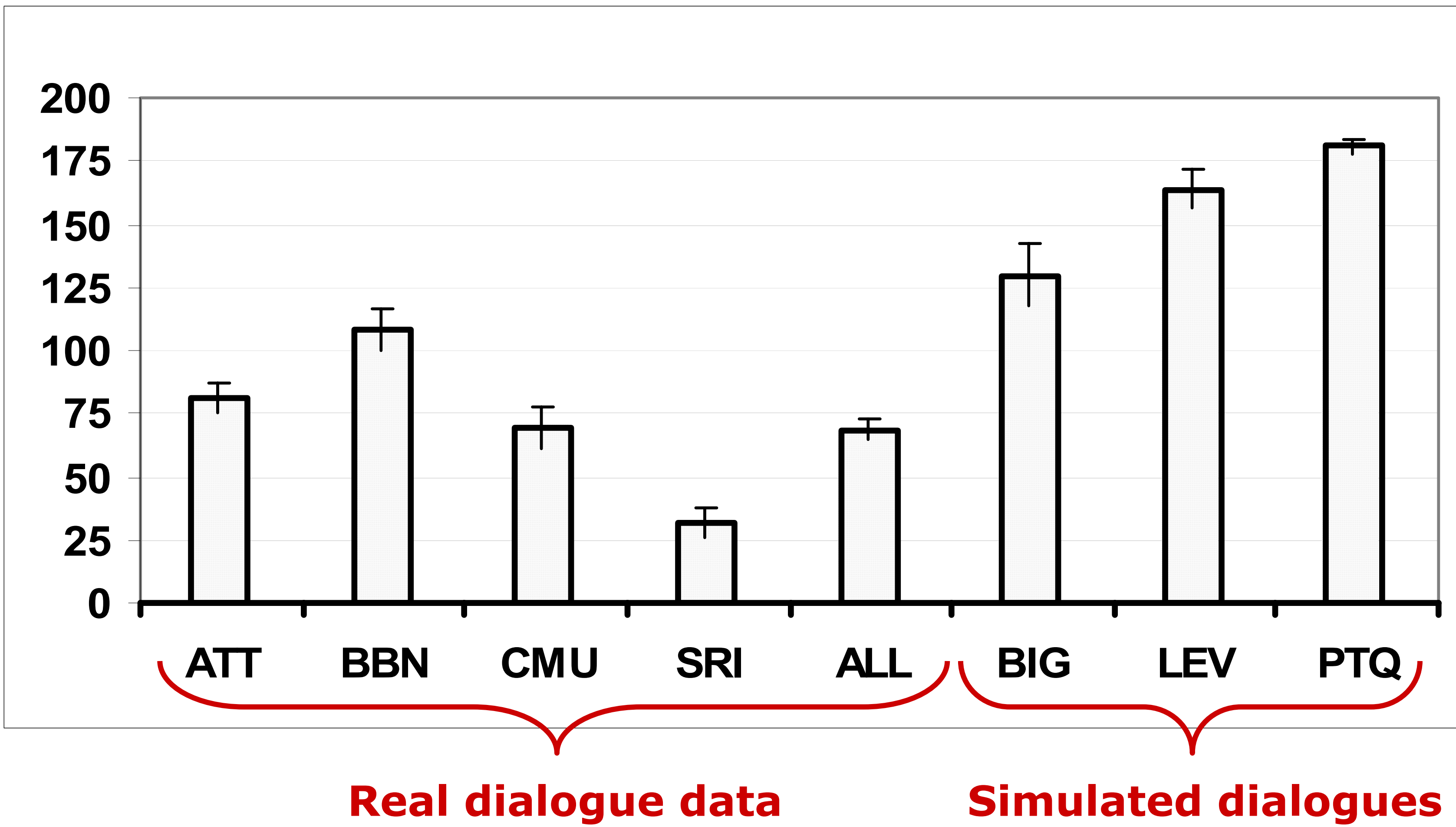
- State of the art in intention-level modelling:
  - Bigram model: $p(a_u|a_s)$
  - Levin model: p(yes_answer|expl_conf)
  - Pietquin model: p(yes_answer|expl_conf, goal)
- UMs typically not trained on real data
- Standard evaluation practice is to test learned strategy on the user model used for learning

Experiments



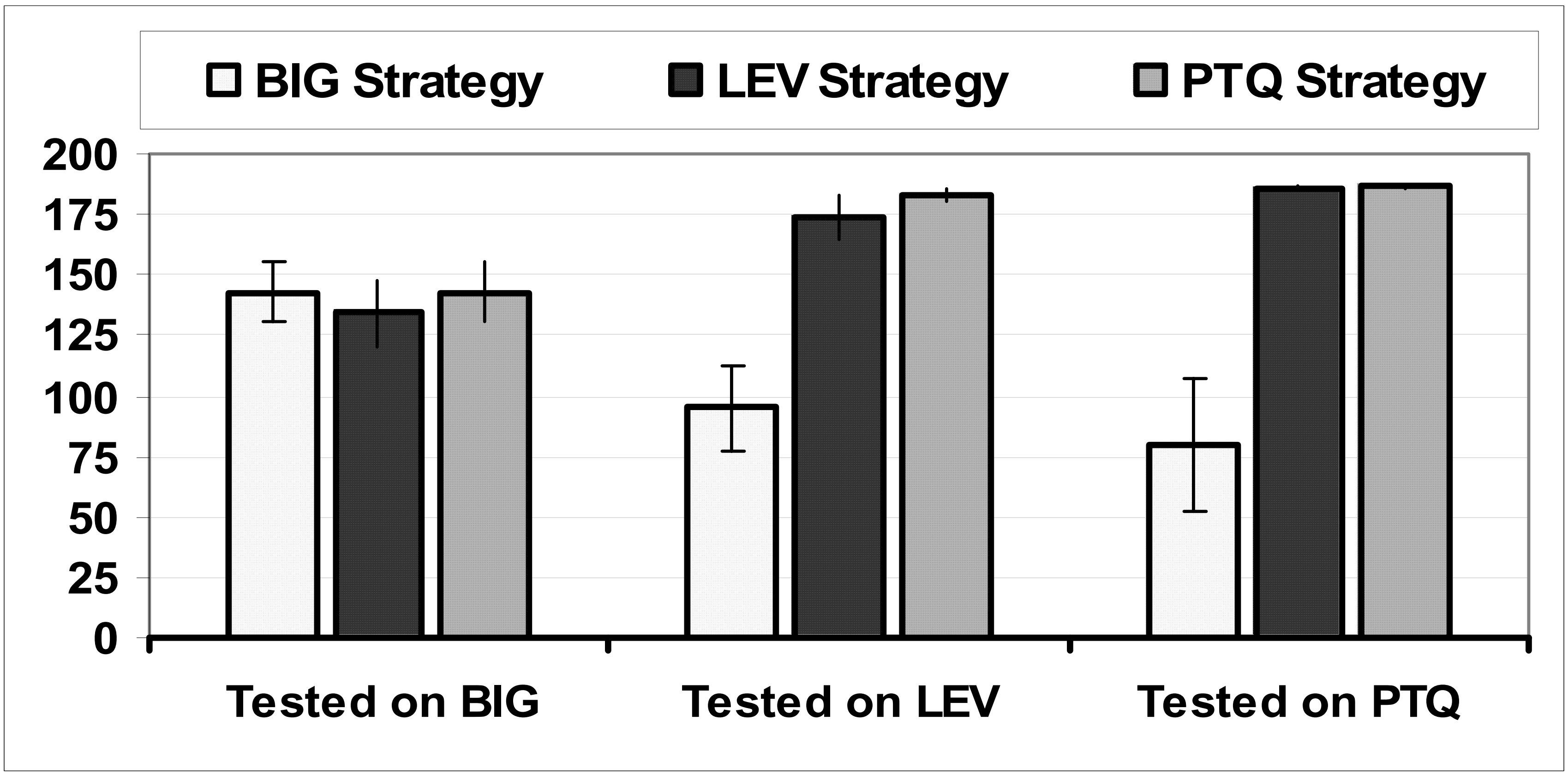
Comparative Evaluation

- **Performance of the learned strategy depends on the quality of the UM**



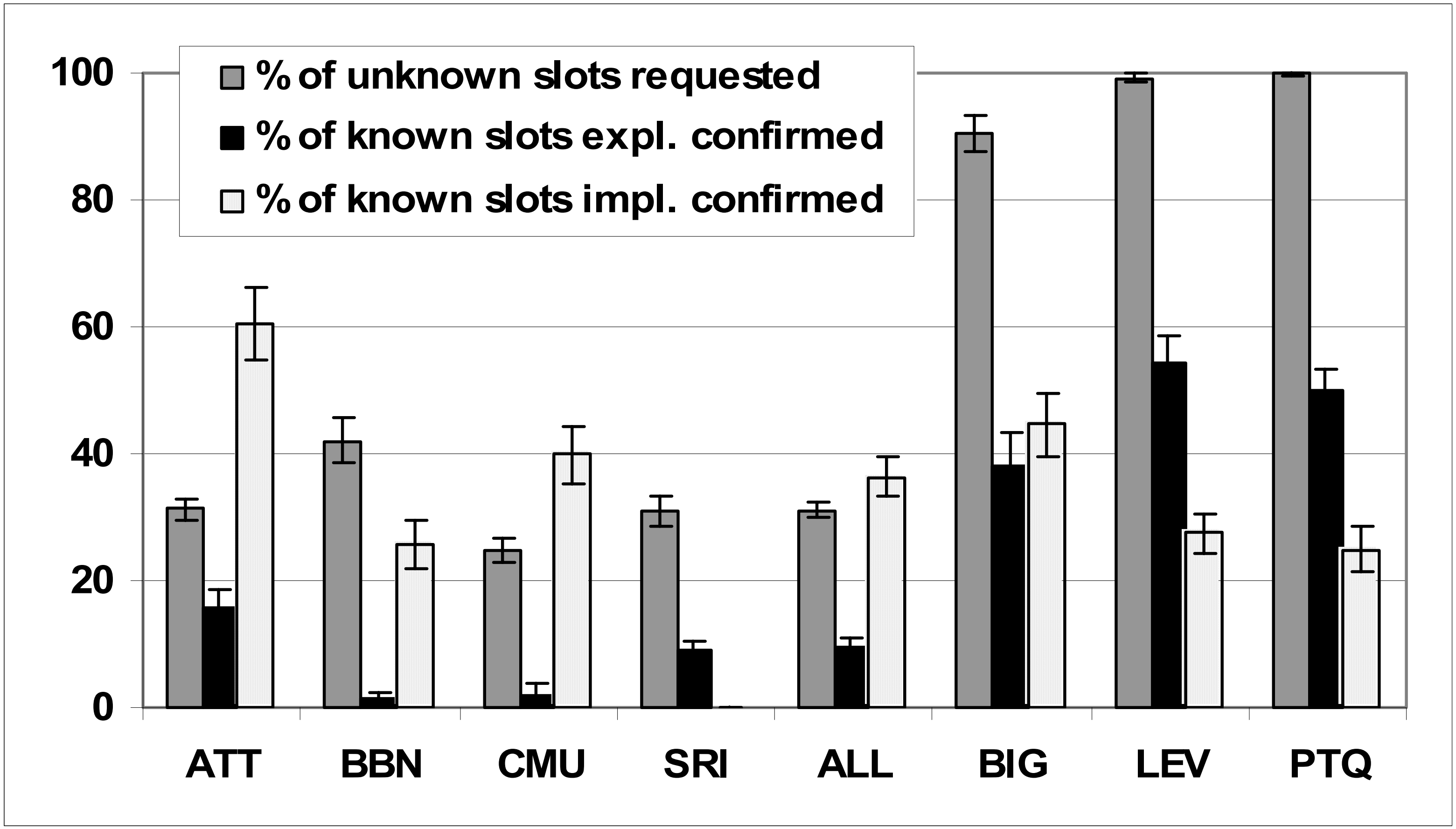
Cross-model Evaluation

- **Strategies learned with a poor UM can fail when tested on a better UM**



Strategy Characteristics

- **Learned strategies exploit weaknesses in UMs**



UM-independent Evaluation

- **Techniques for evaluating new strategies on real dialogue data would be helpful**

Agenda

- **Work on Evaluation: Experiments and Results**
- **Agenda-based User Modelling**

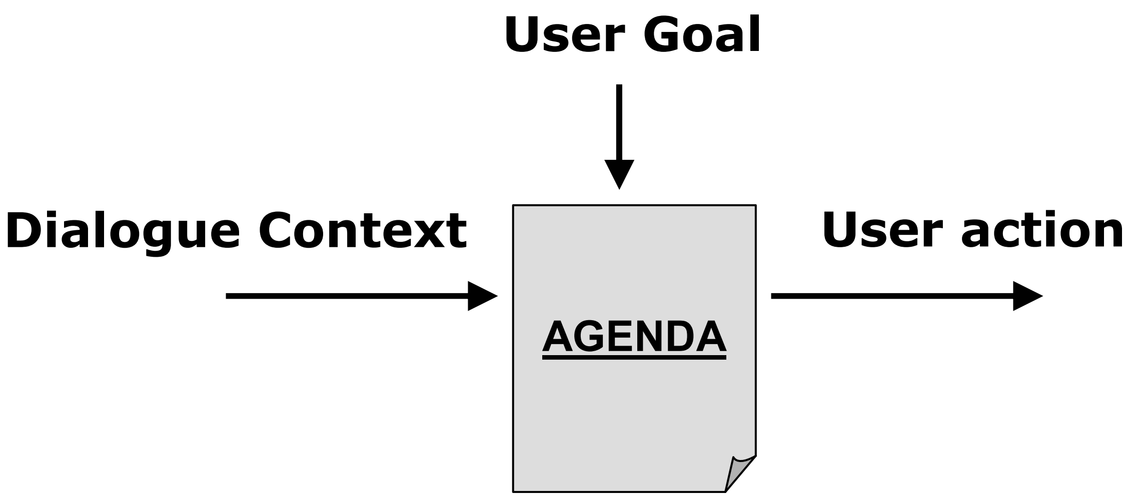
Motivation

- **Currently have drastically different levels of sophistication for DM and UM**
- **Fail to model context which extends beyond the previous dialogue turn**

User:	**I want to go from Boston to London.**
System:	**Going from Austin to London. And when do you want to fly?**
User:	**No, from Boston to London.**
System:	**From Boston to London, is that correct?**
User:	**Yes. And I‘m flying on March 15th.**

Agenda-based User Model

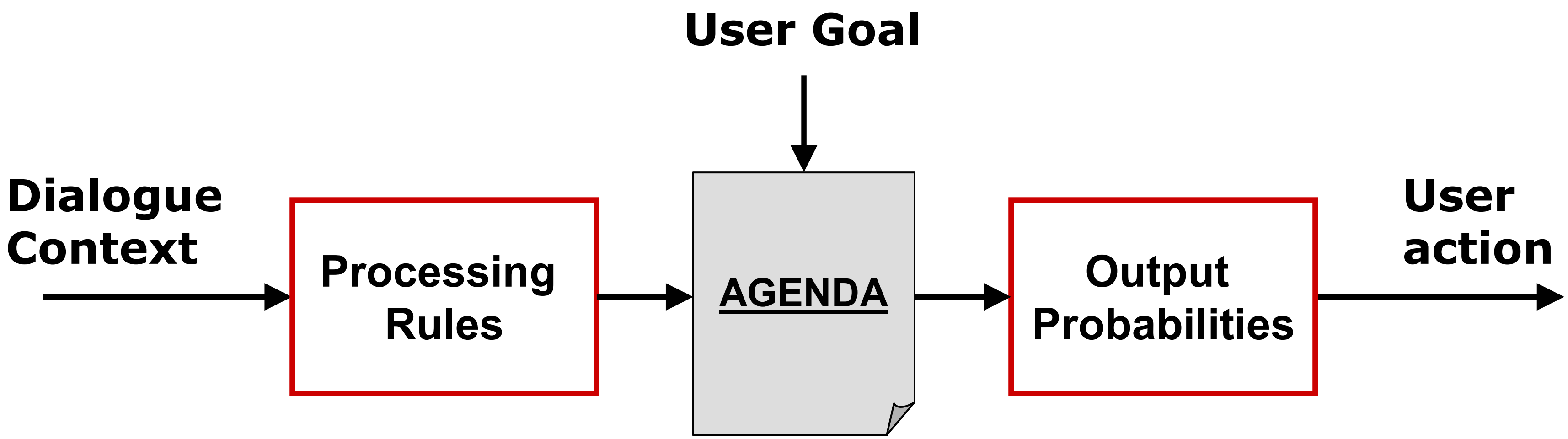
- **<u>Idea</u>: MDP User Model with agenda-based state representation**



  - **Combines user state and user goal representation**
  - **Naturally encodes dialogue history**
  - **Allows delayed user responses (priority of actions)**

Agenda-based User Model

- Assume cooperative user behaviour to label dialogues
- Learn output probabilities to model user behaviour



- Potential scope for modelling uncertainty about true state of user agenda (‘Hidden Agendas‘)

Summary

- **Current lack of solid user models and reliable evaluation standards is a major roadblock to simulation-based strategy learning**

- **Work on agenda-based user models may help to enhance our model of the user state and improve simulation quality**

Thank you!

js532@cam.ac.uk
http://mi.eng.cam.ac.uk/~js532/

J. Schatzmann, K. Georgila, and S. Young. *"Quantitative Evaluation of User Simulation Techniques for Spoken Dialogue Systems"*. 6th SIGdial Workshop on Discourse and Dialogue, Lisbon, September 2-3, 2005 (to appear)

J. Schatzmann, M. N. Stuttle, K. Weilhammer and S. Young. *"Effects of the User Model on Simulation-based Learning of Dialogue Strategies"*. IEEE Automatic Speech Recognition and Understanding Workshop, Cancun, Mexico, November 27 - December 1, 2005 (submitted)

Backup slides

- **The rest of this slide deck is only a backup for further questions.**

Strategy Confidence Scores (1/3)

- **Need to deviate from known strategies to explore new and potentially better ones**

System:	**Where are you flying from, where are you flying to, on what date are you flying, when is your preferred time, do you have a preferred airline and would you like a window-seat?**
Sim. User:	**Flying from Boston to London on March 15 at 9am with Delta Airlines. Window seat please.**
Real User:	**?????**

Strategy Confidence Scores (2/3)

- **<u>Idea:</u> System designer needs a confidence measure indicating how reliable the learned strategy is**

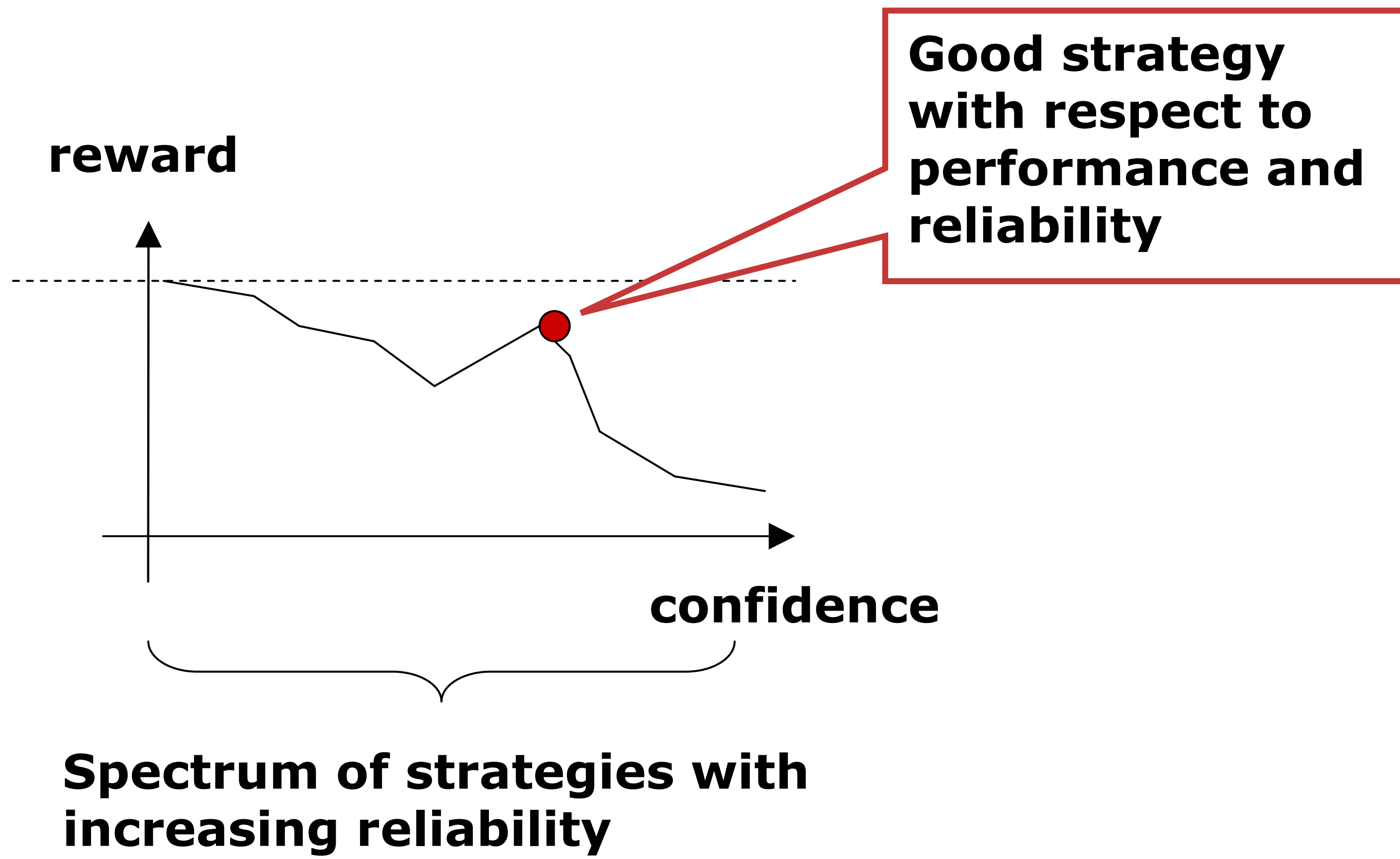
- **Define strategy confidence as function of the likelyhood of the user response in the given context**

$$conf(\pi) = \frac{1}{N} \sum_{i=0}^{N} \frac{1}{N_i} \sum_{t=0}^{N_i} conf(a_{u,t,i}, a_{s,t,i}, s_{t,i})$$

$$conf(a_u, a_s, s) = p(a_u \mid a_s) p(a_s \mid s)$$

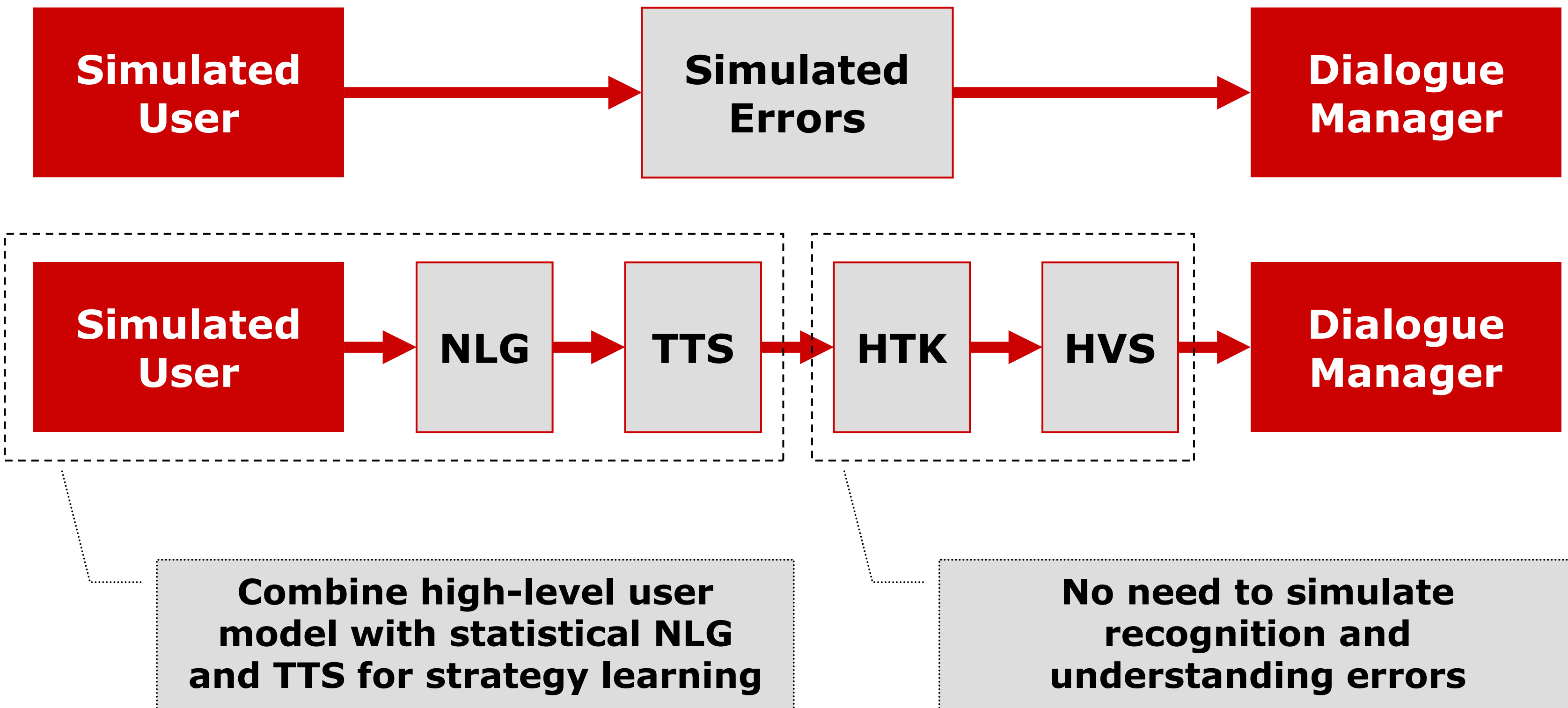
Strategy Confidence Scores (3/3)

- **Reliability score can integrated into the learning process by weighting the reward**



Error Generation (1/1)

- **<u>Idea:</u> Produce acoustic-level output and optimize strategy for system-specific error conditions**



User Studies (1/1)

- **Evaluate performance of new user models using real users**
  - **Test simulation quality using listening tests**
  - **Test strategy performance using questionnaires**
  - **Test usefulness of reliability scores**

Summary

- Work on Evaluation (January to July 2005)
  - Experiments and Results

- Project Proposals (Summer 2005 to Summer 2007)
  - Introduction of strategy confidence scores
  - Agenda-based User Models
  - Strategy learning under system-specific error conditions
  - User studies

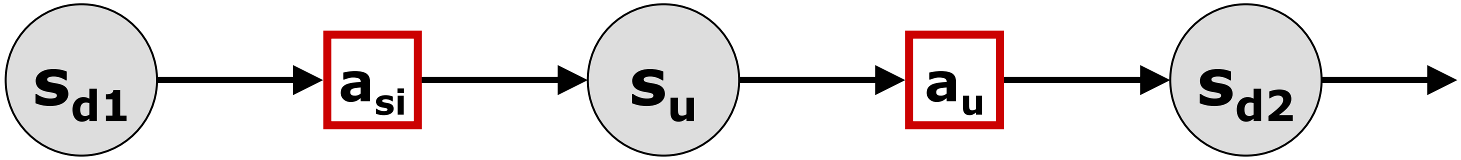
Experiments

- **Implemented a handcrafted DM and trained three different UMs**
  - **Bigram model:** $p(a_u|a_s)$
  - **Levin model:** **p(yes_answer|expl_conf)**
  - **Pietquin model:** **p(yes_answer|expl_conf, goal)**

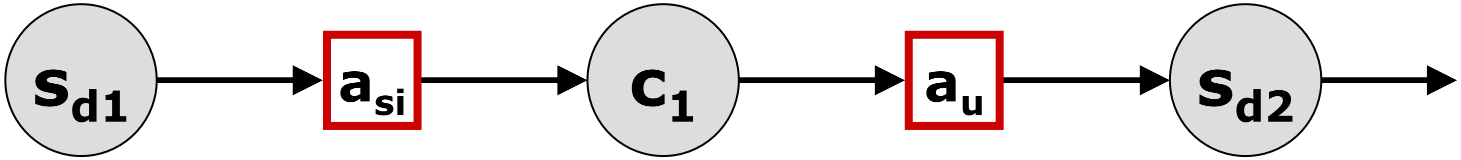
- **Implemented Q-Learning DM, learned strategies with each UM and compared performance and characteristics**
- **Cross-model evaluation of strategies**
- **Investigated user-model independent techniques for testing learned strategies**

Phase II: New User Models (2/3)

- **Idea 2**: Use clustering to construct networks of user behaviour
- Motivation: Networks are well-suited for encoding dialogue context, but their manual construction is expensive
- Represent each dialogue as follows:

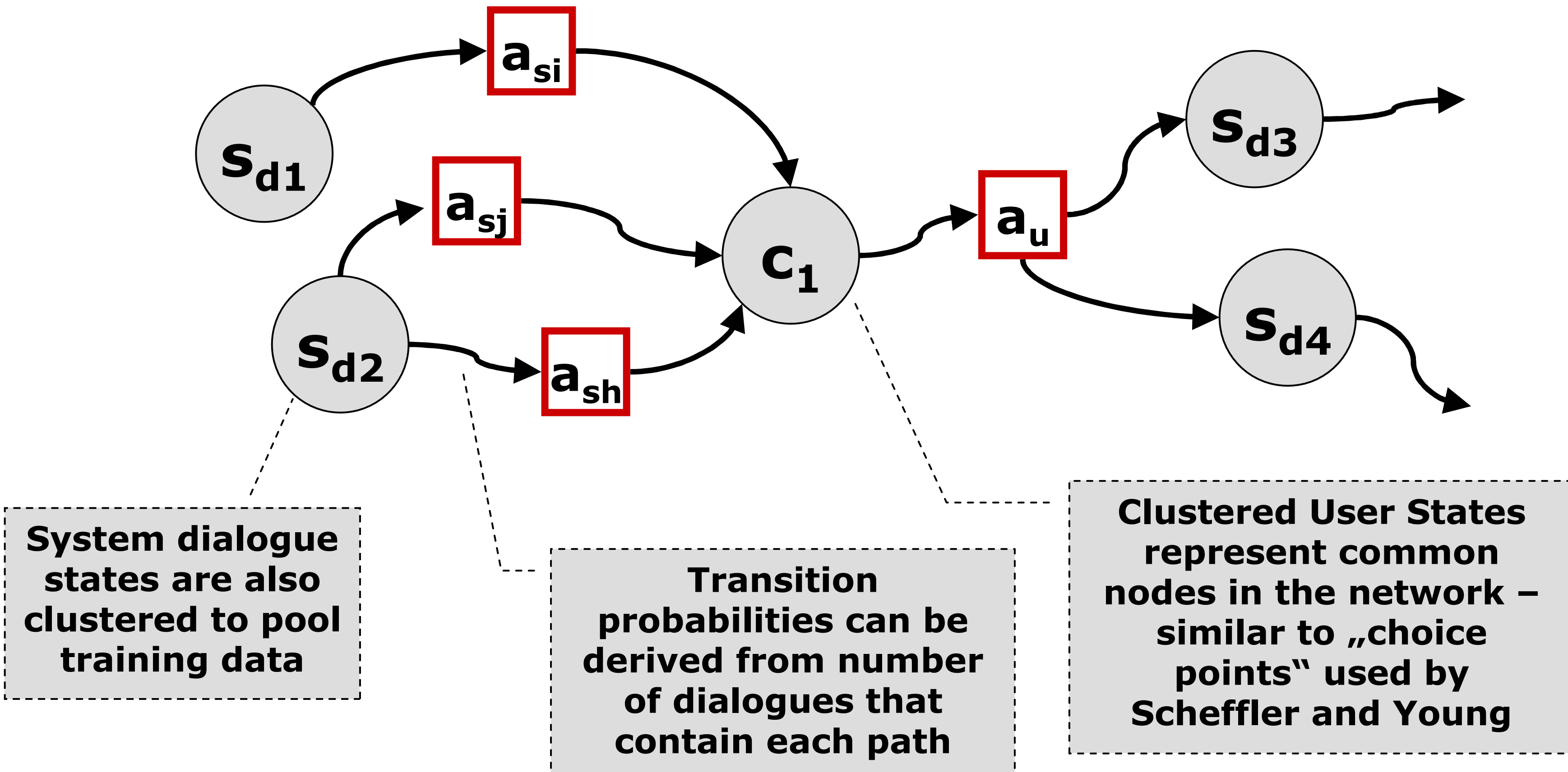


- We want to cluster user states, but the user state can never be fully observed or captured.
- However, we can cluster user actions and assume that similar actions imply similar contexts



Phase II: New User Models (3/3)

- **<u>Idea 2, contd.:</u> Overlay all dialogue sequences to obtain a network**
- **Use frequency counts to obtain transition probabilities**



Backup Slides for Sigdial paper

Evaluation must cover two aspects

- Can the model produce human-like behaviour?
  - Does it produce user responses that a real user might have given in the same dialogue context?

1	Need to compare real and simulated user responses!

- Can the model reproduce the variety of human behaviour?
  - Does it represent the whole user population?

2	Need to compare real and simulated dialogue corpora!

Simulated vs. real user responses

- Split the corpus into training and testing data
- Evaluate how well the model can predict the user responses in the test data
  - Feed in all information about dialogue history and user goal
  - Compare simulated user turn and real user turn
  - Use Precision and Recall to measure how closely the predicted turn matches the real user turn

Use of Precision and Recall

- **Evaluate turn by turn:**

Dialogue in the test set:	**Simulated user responses:**
Sys: greeting instructions request_info orig_city	
Usr: unknown provide_info orig_city london	P=100%, R=50% **Usr:** provide_info orig_city london
Sys: implicit_conf orig_city london request_info dest_city	
Usr: no_answer provide_info orig_city boston	P=0%, R=0% **Usr:** yes_answer provide_info dest_city paris

- **P = Correctly predicted actions / All predicted actions**
- **R = Correctly predicted actions / All actions in real response**

Results: Precision and Recall

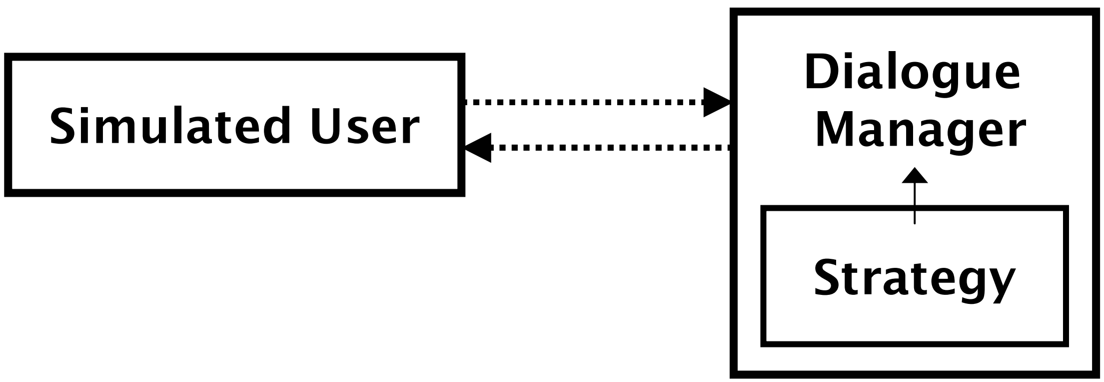
- **Precision and Recall**

	Precision	Recall
Bigram	17.83	21.66
Levin	37.98	31.57
Pietquin	40.16	33.38

- **What do the results mean?**
- **Is this analysis sufficient?**

Simulated vs. real corpora

- We need to evaluate if the model can reproduce the variety of user behaviour in the training data
  - Generate a whole corpus through interaction between the sim. user and the DM
  - Use statistical metrics to compare the simulated corpus to the real one



Statistical metrics

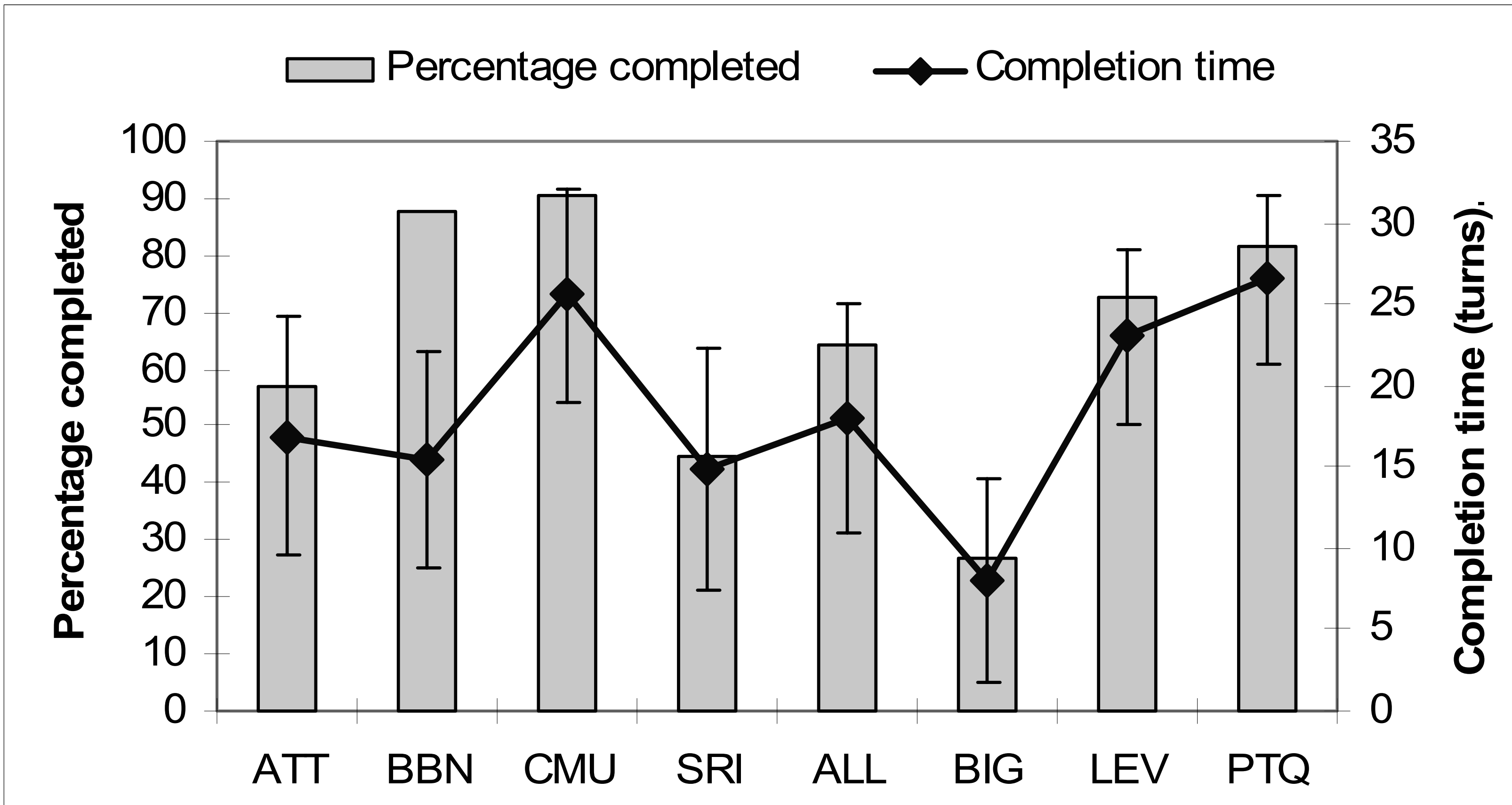
- **High-level dialogue features**
  - **Dialogue length (in number of turns)**
  - **Turn length (in number of actions)**
  - **Proportion of user vs system talk**

- **Dialogue Style and Cooperativeness**
  - **Frequency of different user and system speech acts (average number of occurrences per dialogue)**
  - **Proportion of goal-directed actions vs. Grouding actions vs dialogue formalities vs. Unrecognised actions**
  - **Number of times information is requested, provided, re-requested, re-provided**

- **Dialogue Success and Efficiency**
  - **Average goal / subgoal achievement rate**
  - **Goal completion time**

Results: Goal completion rates / times

- **Goal completion rates and times**



Project overview

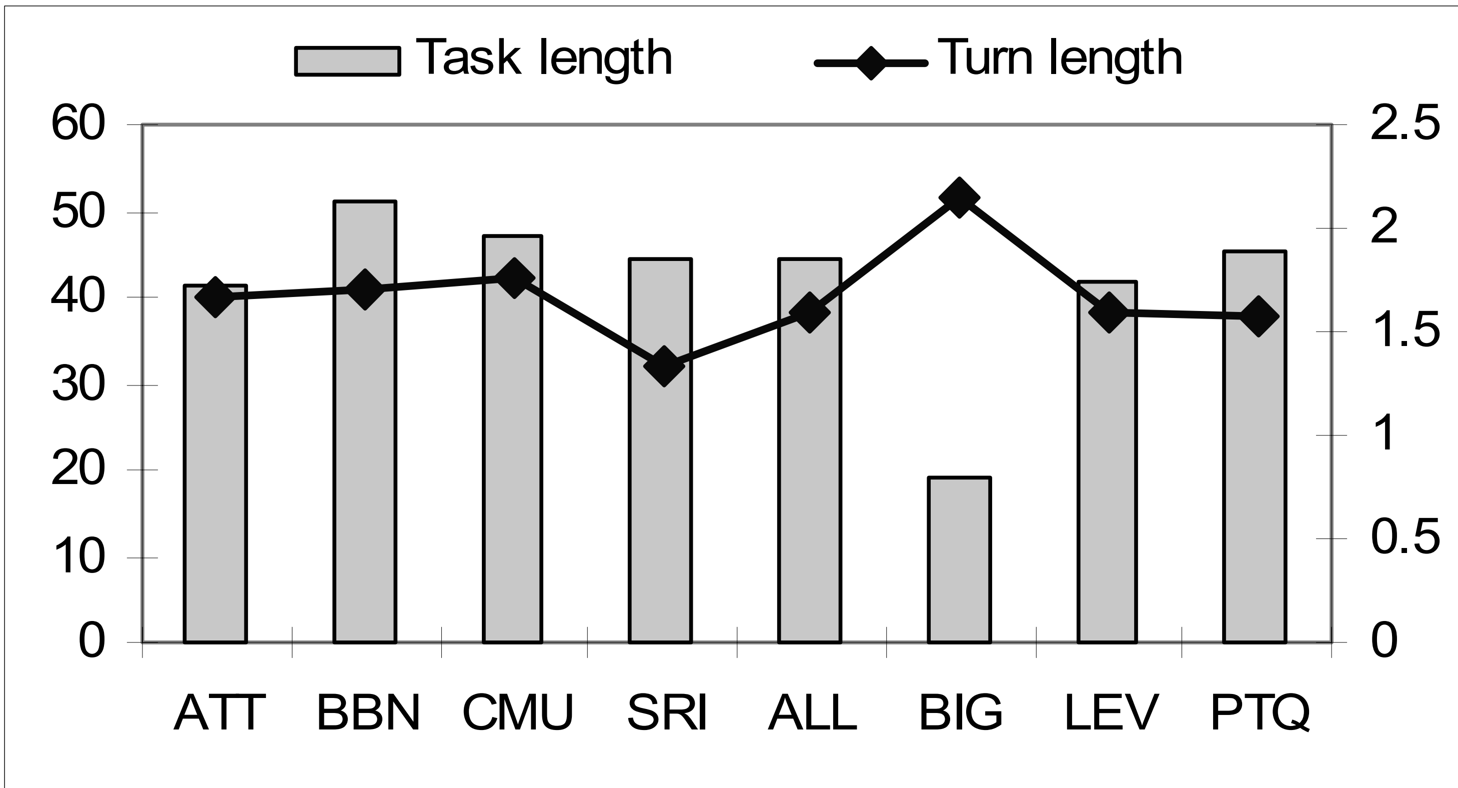
- **Phase I**
  - **Evaluation of the current state of the art**
  - **Re-assessment of standard evaluation practices**
  - Introduction of strategy confidence scores

Work completed

- Phase II
  - Development of new user models
  - Separation of user and error model
- Phase III
  - Acoustic-level simulation
  - Strategy learning under system-specific error conditions
- Phase IV
  - User studies

Phase I: Results (1/5)

- **Simple statistical metrics can distinguish simulated from real dialogue data**



Motivation

- **Lack of a solid user model is currently a major roadblock to automatic DM design**
- **Lack of rigorous evaluation standards has led to uncertainty about the validity of simulation–based learning**
- **Goal is to develop user and error modelling techniques that enable us to learn strategies which outperform competing handcrafted strategies when tested on human users**

Backup slides for simulation techniques

User Models (Backup Slide)

- **Bigram model:** $p(a_u|a_s)$
- **Levin model:** p(yes_answer|expl_conf)
- **Pietquin model:** p(yes_answer|expl_conf, goal)

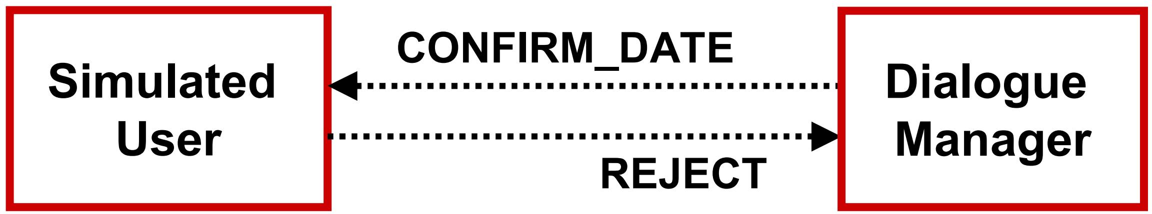
Overview of simulation techniques

- **User simulation for strategy learning is a young field of research:**
  - **Levin, Pieraccini, Eckert (1997, 1998, 2000)**
  - **Lin and Lee (2000)**
  - **Scheffler and Young (1999, 2000, 2001, 2002)**
  - **Pietquin (2002, 2004)**
  - **Henderson, Georgilia, Lemon (2005)**

- **Closely related work on user simulation for SDS evaluation:**
  - **Lopez-Cozar et al. (2003)**
  - **Araki et al. (1997, 1998)**

Levin, Pieraccini, Eckert (1997, 1998)

- Simulation on intention- rather than word- or acoustic level



- N-gram model for predicting the next user intention

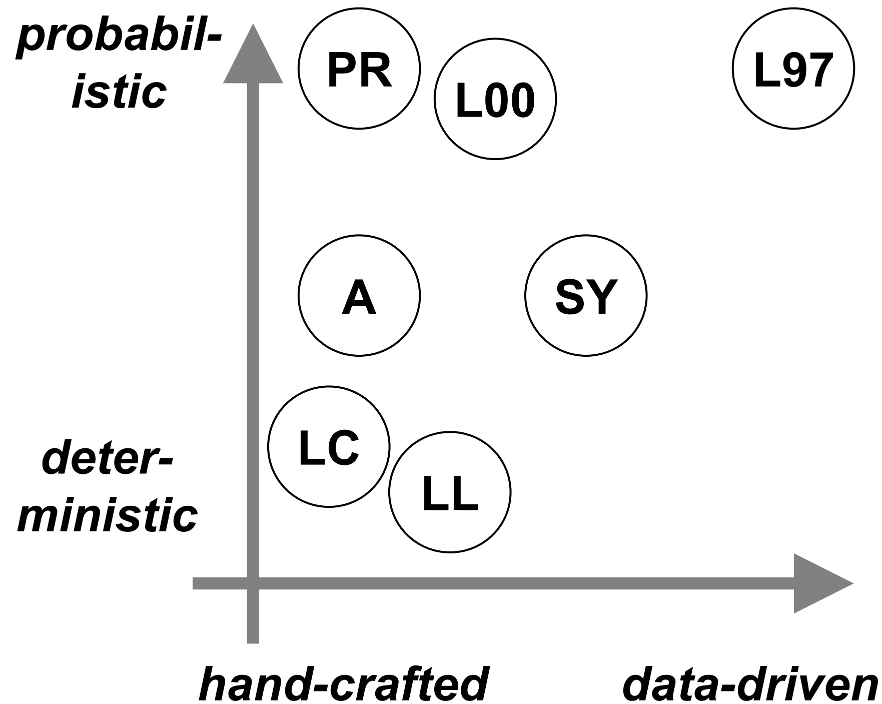
$$\hat{u}_t = \arg\max P(u_t | s_t)$$

- Simulated user responses often unrealistic and in-consistent

System: What is your departure city?
User: New York
System: What is your destination?
User: New York

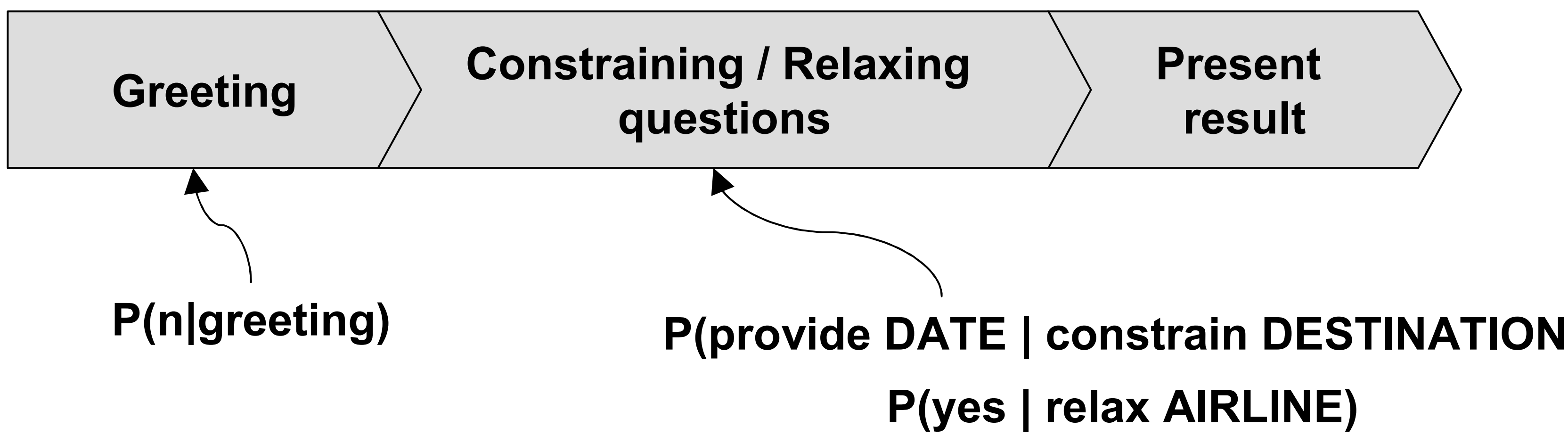
Different approaches to user simulation

- **L97** — Levin, Eckert, Pieraccini (1997, 1998)
- **L00** — Levin, Eckert, Pieraccini (2000)
- **A** — Araki et al. (1997, 1998)
- **LL** — Lin and Lee (2000, 2001)
- **SY** — Scheffler and Young (1999, 2000, 2001, 2002)
- **LC** — Lopez-Cozar et al. (2003)
- **PR** — Pietquin and Renals (2002, 2004)



Levin, Pieraccini, Eckert (2000)

- Attempt to account for weaknesses of the n-gram model
- Assume a simple dialogue model and hand-select appropriate probabilities for predicting user responses



User responses still not goal-consistent!

Scheffler and Young (1999 - 2002)

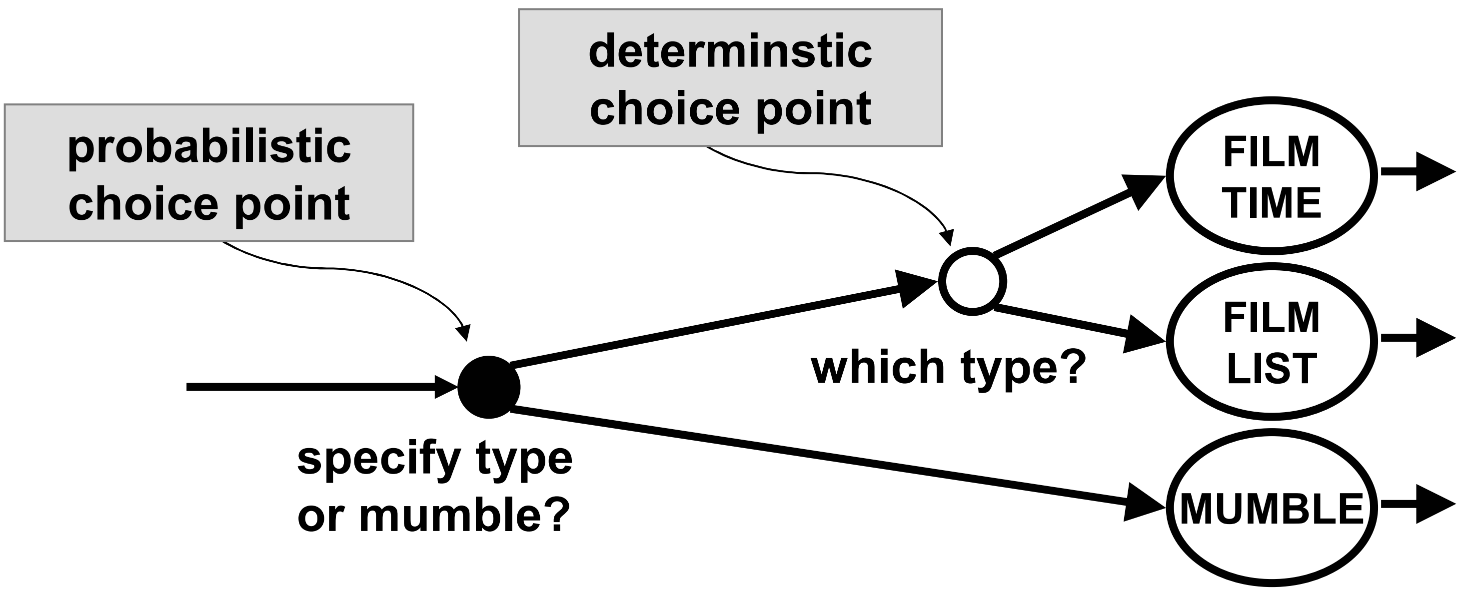
- **User model includes user goal and user‘s beliefs on current system status**

Goal field	Value	Status
Type	GET_FILM_LIST	Specified
Film	NA	NA
Cinema	ARTS_PICT_HOUSE	Pending
Day	TODAY	Pending

- **User acts according to the given goal until it is completed**
- **Frequencies of different goals are estimated from corpus**

Scheffler and Young (1999 - 2002)

- **Utterance generation lattices, obtained by analysing possible dialogue path in existing prototype system**



Heavily task-dependent approach!

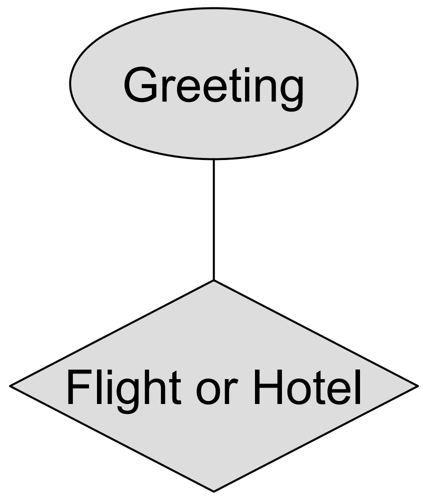
Pietquin (2002, 2004)

- Pietquin combines ideas from Scheffler's and Levin's work
- Probabilities are conditioned on user's goal and memory

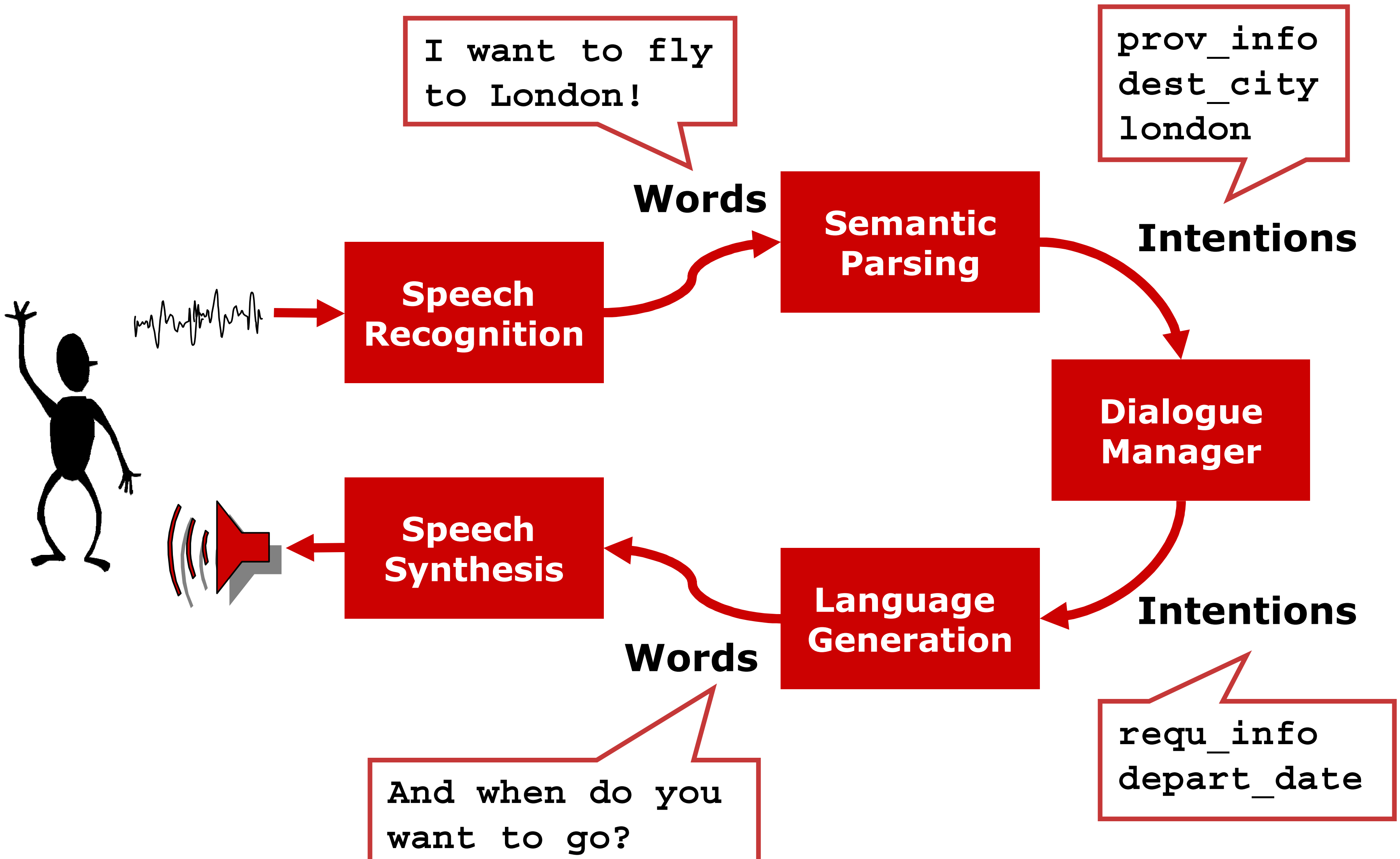
Goal			Memory
Attribute	Value	Priority	Count
PROCESSOR	Pentium	High	0
SPEED	800	High	0
RAM	256	Low	0
HDD	60	Low	0

- P(n|greeting,goal)
- P(provide RAM | constrain HDD, goal, memory)
- P(yes | relax RAM, goal)
- P(close | asked for SPEED, goal, memory)

Graph-based DM



SDS Overview

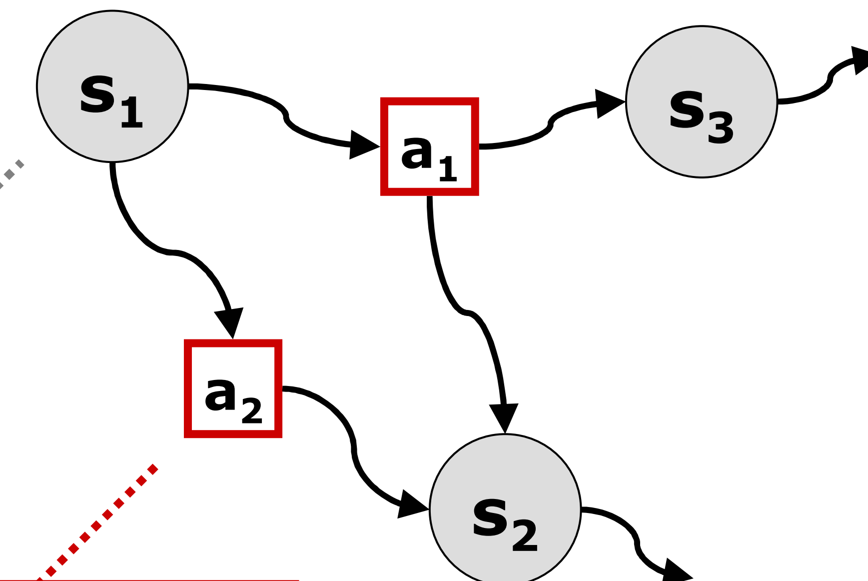


Dialogue as a Markov Decision Process

- Describe dialogue in terms of states and actions
- View DM strategy as a mapping from states to actions

Dialogue State:

orig_city	confirmed
dest_city	known
depart_date	unknown
depart_time	unknown



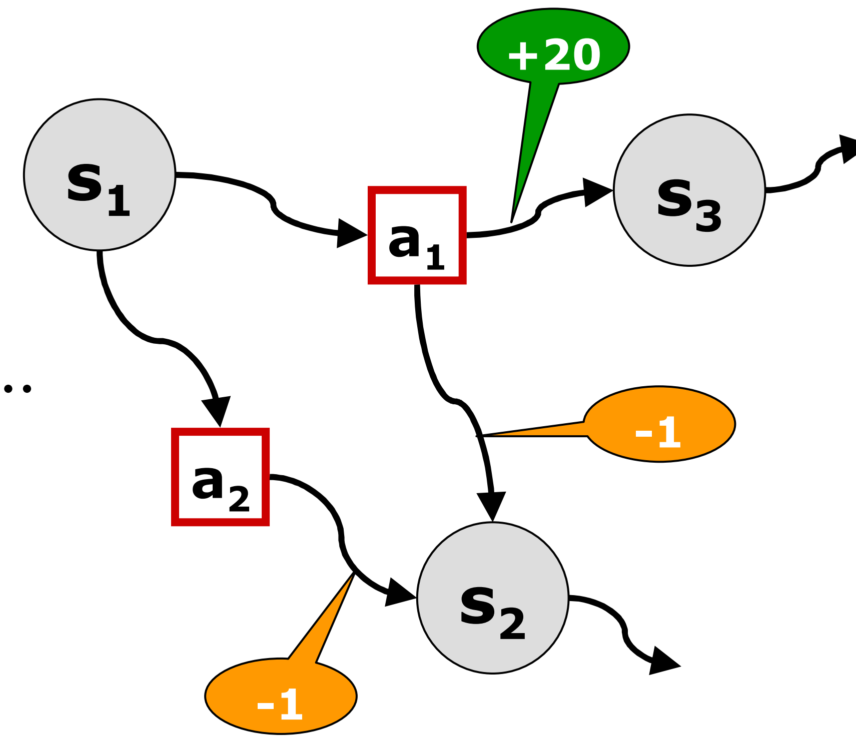
System Action:

<impl_conf, dest_city, london>
<requ_info, depart_date>

Reinforcement Learning (1/2)

- **Learning DM explores its environment through trial-and error and receives a reward r_t at each time t.**
- **Aim is to maximise to the cumulative discounted reward over time $r_{(t)}$**

$$r_{(t)} = \gamma r_{t+1} + \gamma^2 r_{t+2} + \gamma^3 r_{t+3} + \ldots$$



Reinforcement Learning (2/2)

- **Estimate value of taking action a in state s**

$$Q^{\pi}(s,a) = E_{\pi}(r_{(t)} \mid s_t = s, a_t = a)$$

- **Define the optimal policy π***

$$\pi^*(s) = \arg\max_a Q^*(s,a)$$

States

Actions

	s_1	s_2	s_3	…	…	…
a_1	4.23	5.67	2.34	0.67	9.24	…
a_2	1.56	9.45	8.82	5.81	2.36	…
a_3	4.77	3.39	2.01	7.58	3.93	…
…	…	…	…	…	…	…

Q-Learning (Backup Slide)

- **The Q-learning update rule**

$$Q(s,a) := (1-\alpha)Q(s,a) + \alpha\left(r + \gamma \max_{a'} Q(s',a')\right)$$

Learning rate

Reward

Maximum reward obtainable in next state

Old Q-value

Discounting factor