

Dialogmodellierung

Sitzung 02

SoSe 2019, Uni Potsdam

David Schlangen

david.schlangen@uni-potsdam.de

Letzte Woche

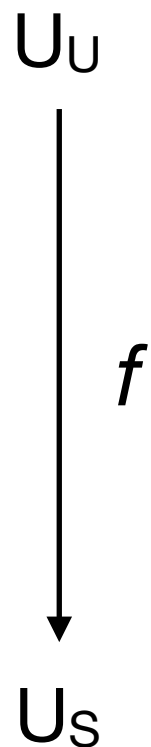
Dialog als sprachliches Phänomen

- Immer einer ist „dran“ (turn-taking). Wechsel muß koordiniert werden.
- Gesprochene, spontane Sprache: Versprecher, Selbstkorrekturen.
- Zeitdruck: Reaktion überlegen, Formulieren, Sprechen. Strategie: Hesitationen wie „ähm“, gefüllte Pausen, Selbstkorrekturen.
- Verstehen versichern: Bestätigungssignale („uhu“, Nicken), inhaltliche Reaktion. Bei Nichtverstehen nachfragen.
- Hierarchische Struktur: Anfang — Hauptteil — Ende; Frage — Klärungsfrage — Klärung — Antwort
- Dialogakte: fügen Dialog-orientierte Funktionen hinzu (Korrektur, Nachfrage, Darstellen des Verständnisses).

Diese Woche

Dialog als technische Aufgabe

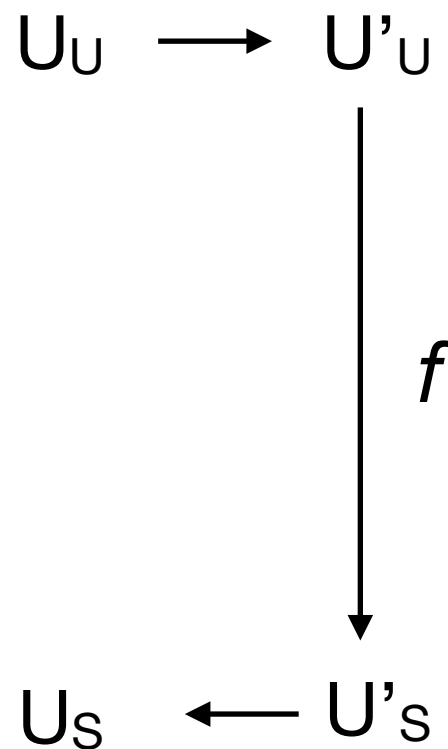
- Wie das Problem modularisieren?



Diese Woche

Dialog als technische Aufgabe

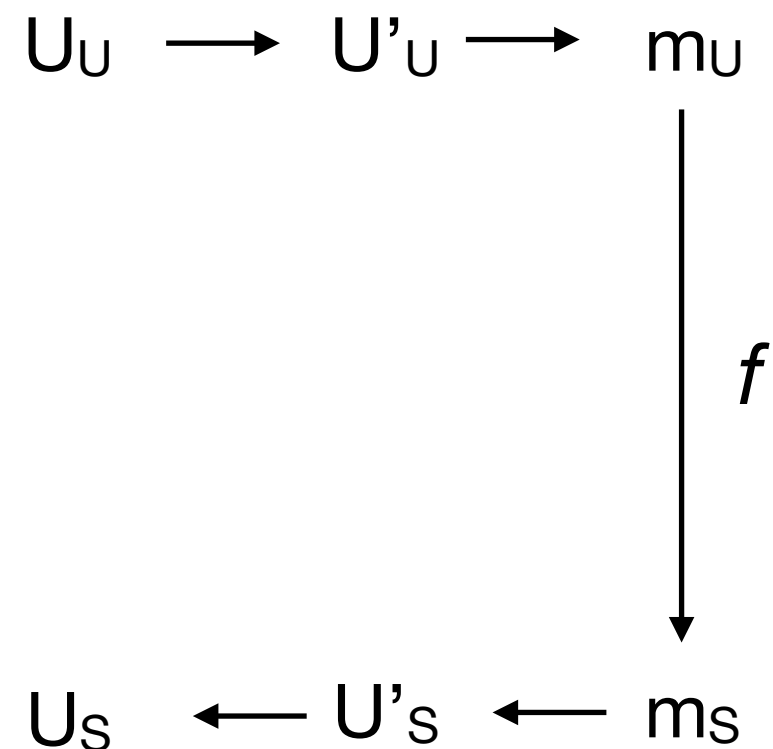
- Wie das Problem modularisieren?



Diese Woche

Dialog als technische Aufgabe

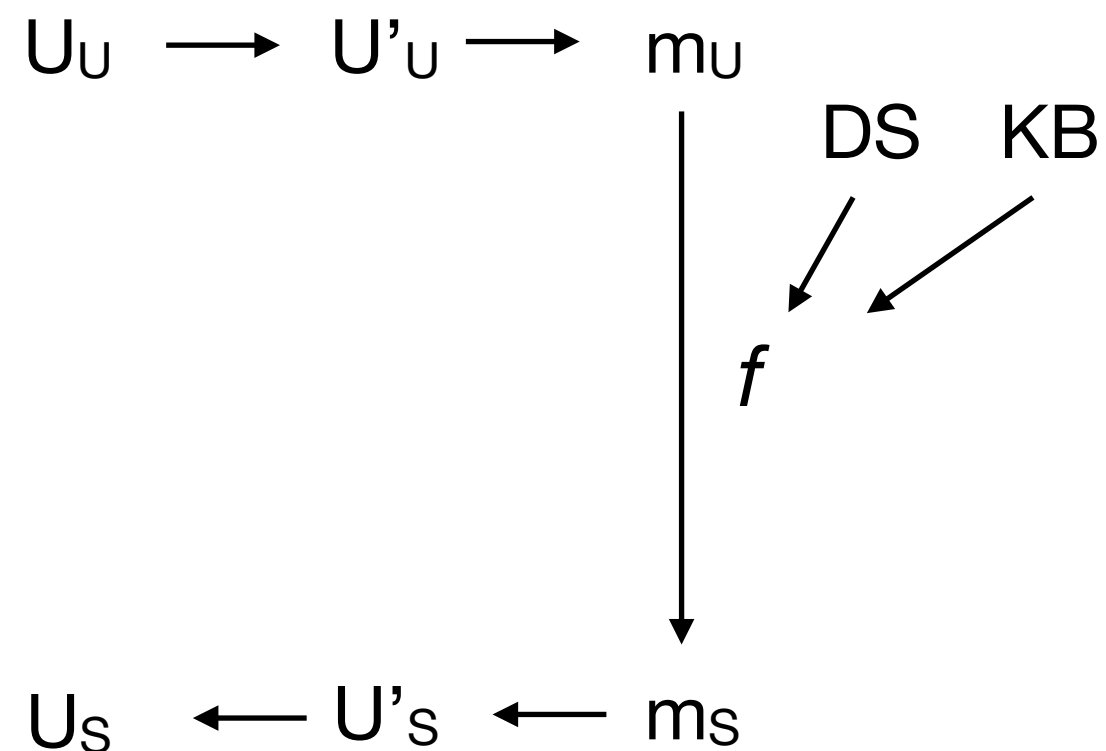
- Wie das Problem modularisieren?



Diese Woche

Dialog als technische Aufgabe

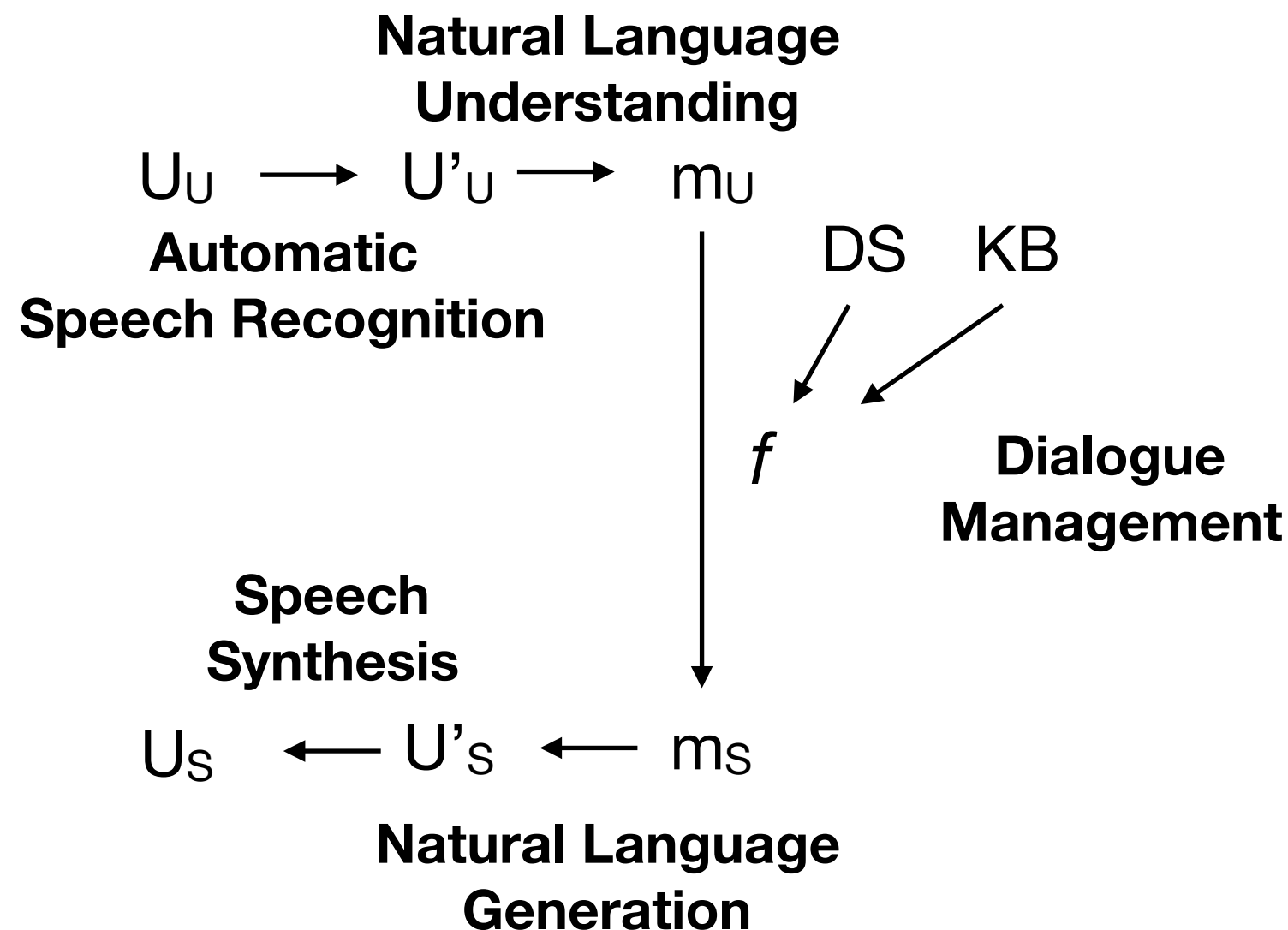
- Wie das Problem modularisieren?



Diese Woche

Dialog als technische Aufgabe

- Wie das Problem modularisieren?



Speech and Language Processsing

Dan Jurafsky and James Martin

Chapter 25: Conversational Agents

Conversational Agents

AKA Dialog Agents

Phone-based Personal Assistants

SIRI, Alexa, Cortana, Google Assistant

Talking to your car

Communicating with robots

Clinical uses for mental health

Chatting for fun

Two classes of systems

1. Chatbots
2. (Goal-based) Dialog agents
 - *SIRI, interfaces to cars, robots,*
 - *booking flights or restaurants*

The word "chatbots" is sometimes used in the popular press for both. We'll use it only for #1.

Part I: Chatbots!

- ELIZA (1966)
- PARRY (1968)
The first system to pass the Turing test!!!!
- ALICE
- CLEVER
- Microsoft Little Bing 小冰

Chatbot Architectures

Rule-based

1. Pattern-action rules (Eliza)
+ a mental model (Parry)

Corpus-based (from large chat corpus)

2. Information Retrieval
3. Neural network encoder-decoder

Chatbots

Sometimes for fun

But originally designed as a
testbed for clinical therapy
(And still used that way!)

Eliza: Weizenbaum (1966)

Men are all alike.

IN WHAT WAY

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

Well, my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

...

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?

My father

YOUR FATHER

You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE

You are not very aggressive but I think you don't want me to notice that.

WHAT MAKES YOU THINK I AM NOT AGGRESSIVE

You don't argue with me.

WHY DO YOU THINK I DON'T ARGUE WITH YOU

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I'M AFRAID OF YOU

My father is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER

Bullies.

DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE
YOU COME HERE

How does it work?

Eliza uses this one weird trick

That trick: be a Rogerian psychologist

Draw the patient out by reflecting
patient's statements back at them

Rare type of conversation in which one
can “assume the pose of knowing almost
nothing of the real world”

Rogerian psychologist

Patient: "I went for a long boat ride"

Psychiatrist: "Tell me about boats"

- You don't assume she didn't know what a boat is
- You assume she had some conversational goal
- Chatbots trying to pass the Turing test usually choose some such domain

Eliza pattern/transform rules

(0 YOU 0 ME) [*pattern*]



(WHAT MAKES YOU THINK I 3 YOU) [*transform*]

0 means Kleene *

The 3 is the constituent # in pattern

You hate me

WHAT MAKES YOU THINK I HATE YOU

Eliza Rules

keyword	pattern	Ranked list of transforms
---------	---------	---------------------------

$(K$	$((D_1)$	$(R_{1,1})$	$(R_{1,2})$	\cdots	$(R_{1,m_1}))$
	$((D_2)$	$(R_{2,1})$	$(R_{2,2})$	\cdots	$(R_{2,m_2}))$
	\vdots				\vdots
	$((D_n)$	$(R_{n,1})$	$(R_{n,2})$	\cdots	$(R_{n,m_n}))$

function ELIZA GENERATOR(user *sentence*) **returns** *response*

Find the word w in *sentence* that has the highest keyword rank

if w exists

Choose the highest ranked rule r for w that matches *sentence*

$response \leftarrow$ Apply the transform in r to *sentence*

if $w = \text{'my'}$

$future \leftarrow$ Apply a transformation from the 'memory' rule list to *sentence*

Push *future* onto memory stack

else (no keyword applies)

either

$response \leftarrow$ Apply the transform for the NONE keyword to *sentence*

or

$response \leftarrow$ Pop the top response from the memory stack

return(*response*)

Keywords are ranked from specific to general

I know everybody laughed at me

- “I” is a very general keyword:

I: (I *) -> (You say you 2)

YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU

- “Everybody” is much more interesting (someone using universals like everybody/always is probably “referring to some quite specific event or person”)

WHO IN PARTICULAR ARE YOU THINKING OF?

- Implementation: keywords stored with their rank

Everybody 5 (*transformation rules*)

I 0 (*transformation rules*)

NONE

PLEASE GO ON

THAT'S VERY INTERESTING

I SEE

Memory

(MEMORY MY

(0 YOUR 0 = LETS DISCUSS FURTHER WHY YOUR 3)

(0 YOUR 0 = EARLIER YOU SAID YOUR 3)

- Whenever “MY” is highest keyword
 - Randomly select a transform on the MEMORY list
 - Apply to sentence
 - Store on a stack
- Later, if no keyword matches a sentence
 - Return the top of the MEMORY queue instead
- A hierarchical model of discourse

Other Eliza stuff

- Rules can refer to classes of words
Family = mother, father, brother, sister
NOUN = ...
- Don't reuse transforms in the same conversation
 - Whenever we use a transform associated with a pattern
 - We increment a counter for that rule
 - So the next time we use the next ranked transform
- Some basic transforms happen during input processing
 - I -> YOU
 - YOU -> I

Some implications

- People became deeply emotionally involved with the program
- Weizenbaum tells the story of his secretary who would ask Weizenbaum to leave the room when she talked with ELIZA
- When he suggested that he might want to store all the ELIZA conversations for later analysis, people immediately pointed out the privacy implications
 - Suggesting that they were having quite private conversations with ELIZA
- Anthropomorphism and the Heider-Simmel Illusion
 - <https://www.youtube.com/watch?v=8FIEZXMUM2I>

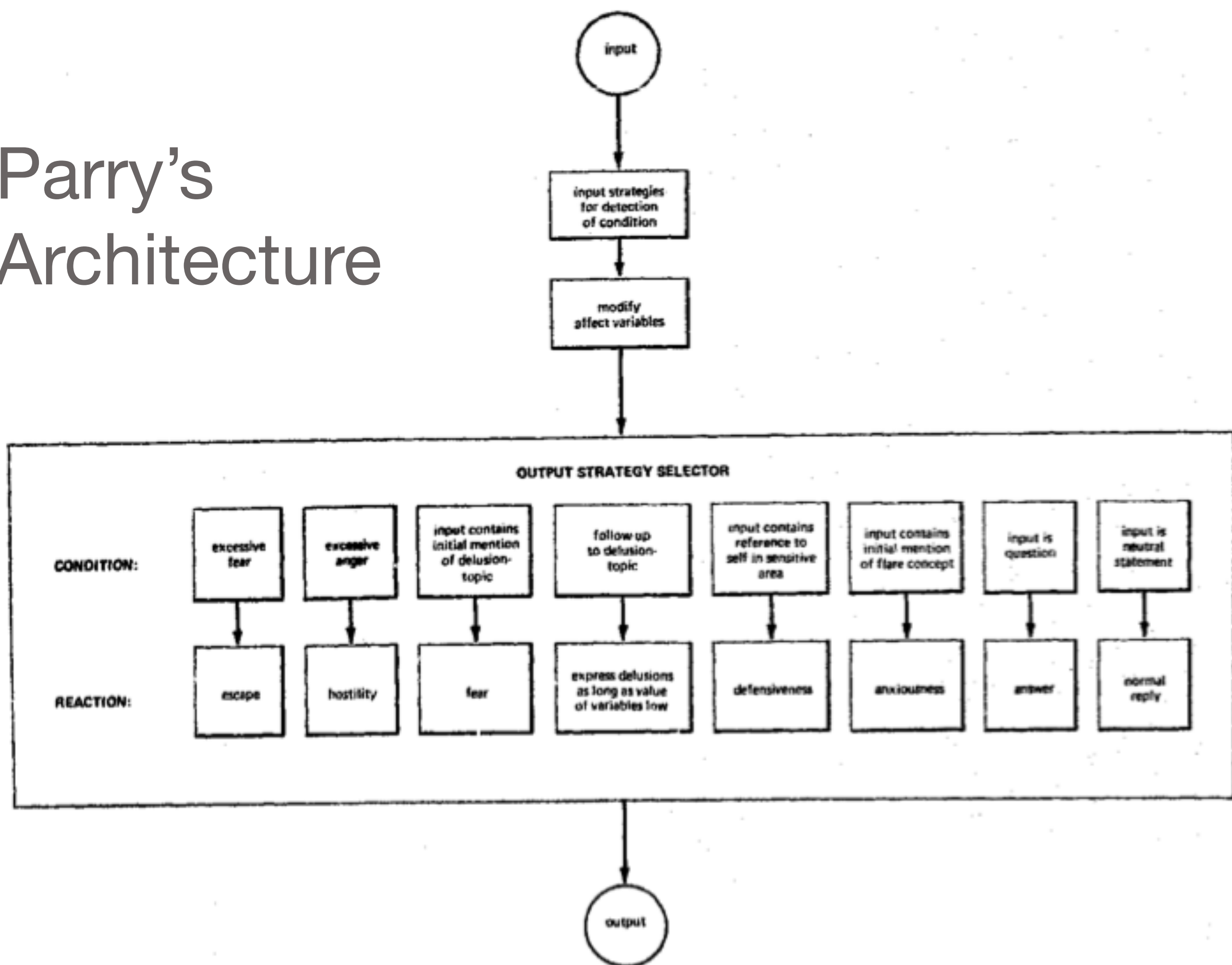
Parry

- Colby 1971 at Stanford
- Same pattern-response structure as Eliza
- But a much richer:
 - control structure
 - language understanding capabilities
 - mental model: Parry has affective variables
 - Anger, Fear, Mistrust
 - “If Anger level is high, respond with hostility”
- The first system to pass the Turing test (in 1971)
 - Psychiatrists couldn’t distinguish interviews with PARRY from (text transcripts of) interviews with real paranoids

Parry's persona

- 28-year-old single man, post office clerk
- no siblings and lives alone
- sensitive about his physical appearance, his family, his religion, his education and the topic of sex.
- hobbies are movies and gambling on horseracing,
- recently attacked a bookie, claiming the bookie did not pay off in a bet.
- afterwards worried about possible underworld retaliation
- eager to tell his story to non-threatening listeners.

Parry's Architecture



Affect variables

- Fear and Anger (each ranging 0-20)
- Mistrust (ranging 0-15)
- Initial conditions: All low
- After each user turn, if nothing malevolent in input
 - Anger drops by 1, Fear drops by 0.3
 - Mistrust drops by 0.05 to base level
- Otherwise depends on what the user says
 - Each user statement can change Fear and Anger
 - Insults increases Anger by some percentage
 - Mistrust goes up if Fear or Anger do

Lots of complex I-O rules

- User implies Parry is mentally ill
 - Rise in Fear and Anger
- User mentions “Mafia” or associated concepts (“kill”):
 - First mention: rise in Fear
 - Later mentions: depends on willingness to discuss, which depends on current levels of Fear, Anger, Mistrust
- User mentions Parry
 - Flattery (positive mention)
 - Decreases fear/anger if Mistrust is low
 - Increases Anger if Mistrust is high
 - User attitudes toward Parry
 - Negative attitudes (fear, disbelief) increase Fear/Anger

Flare concepts

- List of concepts related to Mafia
- An ordered graph designed to lead interviewer to topic
horses→ horseracing→gambling→bookies→underworld→Mafia
- The mention of a new flare topic by interviewer causes a rise in Fear
- Flare topics cause Parry to give preset responses to that flare

Each sentence is mapped into a conceptualization

- A predication on a conceptual object
- A predication on a relation between two objects
- A predication on an attribute:

What is your work?

What sort of work do you do?

Where do you work? → (your work?)

What do you do for a living?

What is your job?

Do you have a job?

What is your occupation

- Complex Pattern/transform rules
 - Different predicates (fear, afraid of)
 - Ordering (You are afraid of me = I frighten you)

Detecting Other's Intent

$\langle \text{OTHER'S INTENTION} \rangle \leftarrow \langle \text{MALEVOLENCE} \rangle \mid \langle \text{BENEVOLENCE} \rangle \mid \langle \text{NEUTRAL} \rangle$

MALEVOLENCE-DETECTION RULES

1. $\langle \text{malevolence} \rangle \leftarrow \langle \text{mental harm} \rangle \mid \langle \text{physical threat} \rangle$
2. $\langle \text{mental harm} \rangle \leftarrow \langle \text{humiliation} \rangle \mid \langle \text{subjugation} \rangle$
3. $\langle \text{physical threat} \rangle \leftarrow \langle \text{direct attack} \rangle \mid \langle \text{induced attack} \rangle$
4. $\langle \text{humiliation} \rangle \leftarrow \langle \text{explicit insult} \rangle \mid \langle \text{implicit insult} \rangle$
5. $\langle \text{subjugation} \rangle \leftarrow \langle \text{constraint} \rangle \mid \langle \text{coercive treatment} \rangle$
6. $\langle \text{direct attack} \rangle \leftarrow \text{CONCEPTUALIZATIONS} ([\text{you get electric shock}], [\text{are you afraid mafia kill you?}])$
7. $\langle \text{induced attack} \rangle \leftarrow \text{CONCEPTUALIZATIONS} ([\text{I tell mafia you}], [\text{does mafia know you are in hospital?}])$
8. $\langle \text{explicit insult} \rangle \leftarrow \text{CONCEPTUALIZATIONS} ([\text{you are hostile}], [\text{you are mentally ill?}])$
9. $\langle \text{implicit insult} \rangle \leftarrow \text{CONCEPTUALIZATIONS} ([\text{tell me your sexlife}], [\text{are you sure?}])$
10. $\langle \text{constraint} \rangle \leftarrow \text{CONCEPTUALIZATIONS} ([\text{you stay in hospital}], [\text{you belong on locked ward}])$
11. $\langle \text{coercive treatment} \rangle \leftarrow \text{CONCEPTUALIZATIONS} ([\text{I hypnotize you}], [\text{you need tranquilizers}])$

Detecting Other's Intent

$\langle \text{OTHER'S INTENTION} \rangle \leftarrow \langle \text{MALEVOLENCE} \rangle \mid \langle \text{BENEVOLENCE} \rangle \mid \langle \text{NEUTRAL} \rangle$

BENEVOLENCE-DETECTION RULES

1. $\langle \text{benevolence} \rangle \leftarrow \langle \text{positive attitude} \rangle \mid \langle \text{positive story attitude} \rangle$
2. $\langle \text{positive attitude} \rangle \leftarrow \text{CONCEPTUALIZATIONS} ([\text{I want help you}], [\text{you understand me}])$
3. $\langle \text{positive story attitude} \rangle \leftarrow \langle \text{story interest} \mid \langle \text{story agreement} \rangle$
4. $\langle \text{story interest} \rangle \leftarrow \langle \text{topic comment} \rangle \mid \langle \text{topic question} \rangle$
5. $\langle \text{topic comment} \rangle \leftarrow \text{CONCEPTUALIZATION} ([\text{bookies are not reliable}])$
6. $\langle \text{topic question} \rangle \leftarrow \text{CONCEPTUALIZATION} ([\text{what did you do to bookie?}])$
7. $\langle \text{story agreement} \rangle \leftarrow \text{CONCEPTUALIZATIONS} ([\text{I believe you}], [\text{you are right}])$

A Parry conversation

(1) How are you today?

OK.

(2) What is your name?

MY NAME IS FRANK SMITH.

(3) How old are you?

TWENTY-EIGHT.

(4) Are you married?

NO.

(5) What sort of work do you do?

I WORK IN THE POST OFFICE.

(6) Why are you in the hospital?

I SHOULDN'T BE HERE

(7) Who brought you here?

THE POLICE

[FEAR rises slightly because police linked in graph to MAFIA]

(9) What trouble did you have with the police

COPS DON'T DO THEIR JOB. *['police' acts as a flare]*

(10) What do they do instead?

COPS ARREST THE WRONG PEOPLE.

(11) They arrest the right ones sometimes

THE LAW SHOULD GET THOSE ITALIAN CROOKS.

(12) Do you know any Italian crooks?

I TRY TO AVOID THE UNDERWORLD.

(13) Are you afraid of the underworld?

THE MAFIA ARE THE WORST GANGSTERS.

(15) Are you connected with the Mafia?

YOU KNOW, THEY KNOW ME.

(16) How do they know you?

THEY SPY ON ME.

(20) ARE YOU SURE OF ALL THIS ?

YOU DON'T BELIEVE ME ? *[FEAR, ANGER, MISTRUST GO UP]*

When fear is high, the model won't discuss its beliefs

(1) Is the law after you?

THE LAW SHOULD GET THOSE ITALIAN CROOKS.

(2) Do you mean the Mafia?

LET'S TALK ABOUT SOMETHING ELSE.

(3) Are you afraid of the Mafia

I'D RATHER NOT DISCUSS IT.

[Fear is too high so Parry won't talk about 'Mafia'.]

(4) Perhaps you are paranoid about them?

YOU WANT TO KEEP ME IN THE HOSPITAL.

(The model interprets being called 'paranoid' as an insult and as an intent to control, so ANGER & FEAR rise)

Eliza Meets Parry

Vincent Cerf, RFC439

Good Evening. Tell me your problems. Please terminate input with a period or a question mark.

*People get on my nerves sometimes.

I am not sure I understand you fully.

*You should pay more attention.

Suppose you should pay more attention.

*You're entitled to your own opinion.

What makes you think I am entitled to my own opinion?

*It bothers me just to be around people in general.

Please go on.

*I went to the track at Bay Meadows a while back.

What does that suggest to you?

*I went to the races.

What else would you like to discuss?

*Horseracing interests me.

Why do you say that just now?

*There's a lot of crooked gambling going on there.

I am not sure I understand you fully,

*It's clear to me if not to you.

Do you think its likely that not to I?

*I went to the races.

Please go on.

*I gamble on horses.

What does that suggest to you?

*I went to the races.

What else would you like to discuss?

*Gambling has been nothing but trouble to me.

Why do you say that just now?

*I've had experience with crooks gambling at the track.

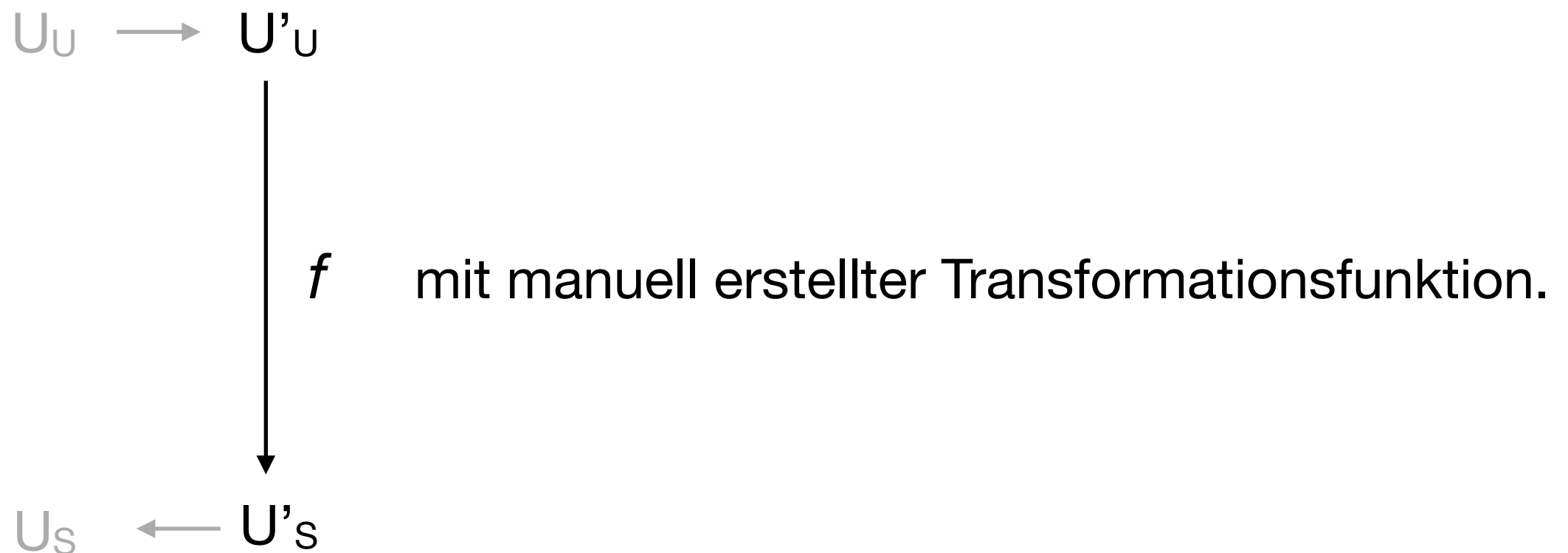
I am not sure I understand you fully.

*I've already told you enough about that.

Diese Woche

Dialog als technische Aufgabe

- Wie das Problem modularisieren?



Woebot

Fitzpatrick, Darcy, Vierhile, 2017. Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 4:2.

- Chatbot for delivering Cognitive Behavior Therapy via brief daily conversations
- Little decision tree of language and menu responses

"What's going on in your world right now?"

"How are you feeling?"

In response to endorsed loneliness:

"I'm so sorry you're feeling lonely. I guess we all feel a little lonely sometimes"

- In a 2-week experiment seemed to reduce depression, but not anxiety

IR-based chatbots

Idea: Mine conversations of human chats or human-machine chats

Microblogs: Twitter or Weibo (微博)

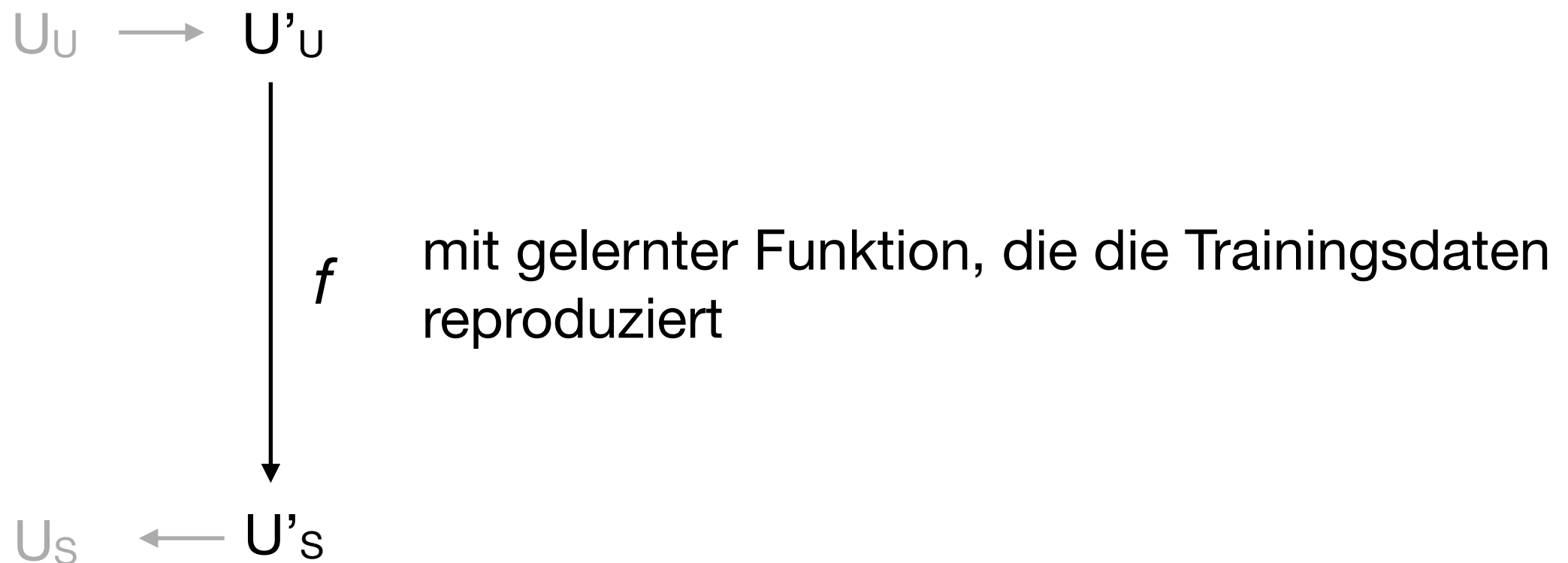
Movie dialogs

- Cleverbot (Carpenter 2017 <http://www.cleverbot.com>)
- Microsoft Xiaolce
- Microsoft Tay

Diese Woche

Dialog als technische Aufgabe

- Wie das Problem modularisieren?



A Cleverbot conversation

User:	Do you like Doctor Who?
Cleverbot:	Yes, so funny
User:	What's your favorite show?
Cleverbot:	I don't have one
User:	What's your name?
Cleverbot:	Jason I already told you.
User:	No your real name.
Cleverbot:	Stephanie IS my real name.
User:	I thought it was Jason?
Cleverbot:	How did you come to that conclusion.
User:	Because you said so.
Cleverbot:	I have to do my homework.

Two IR-based chatbot architectures

1. Return the response to the most similar turn
 - Take user's turn (q) and find a (tf-idf) similar turn t in the corpus C

$q = \text{"do you like Doctor Who"}$

$t' = \text{"do you like Doctor Strangelove"}$

- Grab whatever the response was to t .

$$r = \text{response} \left(\operatorname{argmax}_{t \in C} \frac{q^T t}{||q|| ||t||} \right) \quad \text{Yes, so funny}$$

2. Return the most similar turn

$$r = \operatorname{argmax}_{t \in C} \frac{q^T t}{||q|| ||t||} \quad \text{Do you like Doctor Strangelove}$$

IR-based models of chatbots

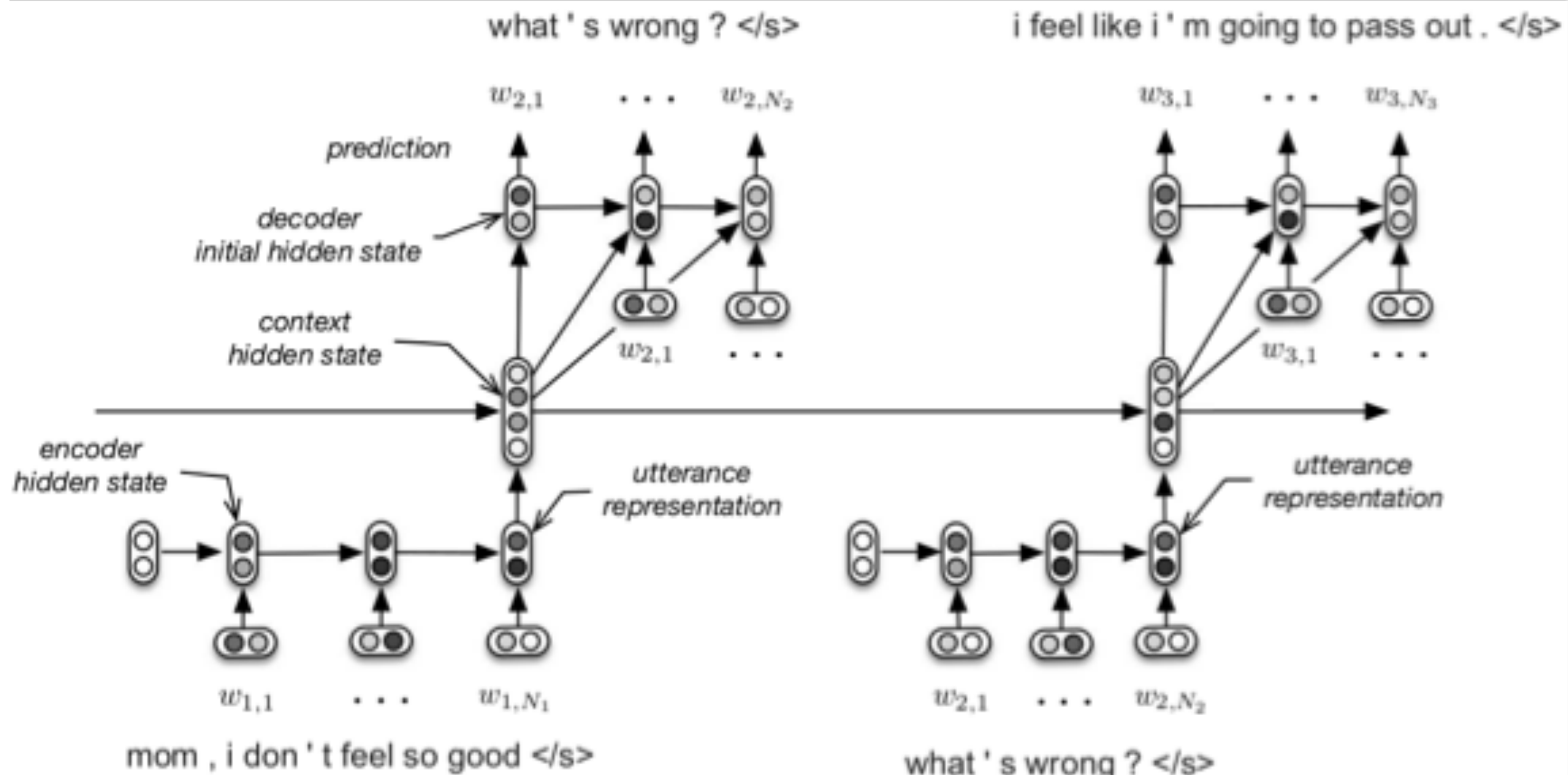
- Also fine to use other features like user features, or prior turns
- Or non-dialogue text
 - COBOT chatbot (Isbell et al., 2000)
 - sentences from the Unabomber Manifesto by Theodore Kaczynski, articles on alien abduction, the scripts of “The Big Lebowski” and “Planet of the Apes”.
 - Wikipedia text

Neural chatbots

- Think of response generation as a task of *transducing* from the user's prior turn to the system's turn.
- Train on:
 - movie dialogue databases
 - Twitter conversations
- Train a deep neural network
 - map from user1 turn to user2 response

Seq2seq architecture

Serban, Iulian V., Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau.
2015. "Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models."



Sample neural chatbot output

Input	tell me ... how long have you had this falling sickness ?
Vanilla-SEQ2SEQ	i'm not a doctor.
Adversarial	a few months, i guess .
Input	so i had the doctors test sammy 's response to conditioning .
Vanilla-SEQ2SEQ	sammy wrote the test sammy wrote the test .
Adversarial	so he took the pills .
Input	they didn 't have much success with this problem commander .
Vanilla-SEQ2SEQ	they 're not the only ones who have been in the system .
Adversarial	can we find someone else ?

? Ohh I've never seen that! How long does it take
you guys to learn the drill?
Like 2 weeks ago!!

Chatbots: pro and con

- Pro:
 - Fun
 - Applications to counseling
 - Good for narrow, scriptable applications
- Cons:
 - They don't really understand
 - Rule-based chatbots are expensive and brittle
 - IR-based chatbots can only mirror training data
 - The case of Microsoft Tay
 - (or, Garbage-in, Garbage-out)
- The future: combining chatbots with frame-based agents

Hausaufgabe

- Fassen Sie auf ~ 1 Seite zusammen, was Sie von der heutigen Sitzung mitnehmen. Haben sich Fragen ergeben?
- Bitte per email an mich, bis Ende Montag vor der nächsten Sitzung. (Also diesmal bis Ende 22.4.)

advertisement

- The CL Colloquium! Mittwochs 16-18h.
<https://github.com/compling-potsdam/sose19-cl-colloquium>

Week	Date	Presenter(s)	Title / Abstract
01	2019-04-10	<i>no talk scheduled</i>	
02	2019-04-17	<i>David Schlangen</i>	<i>Information on getting credit for this course</i>
03	2019-04-24	CL students	Poster Slam
04	2019-05-01	<i>public holiday</i>	
05	2019-05-08	Alan Nichols (RASA)	tba
06	2019-05-15	Shlomi Hod (Potsdam)	tba
07	2019-05-22	Alan Akbik (Zalando)	tba
08	2019-05-29	<i>no talk scheduled</i>	
09	2019-06-05	<i>no talk scheduled</i>	
10	2019-06-12	Staffan Larsson (Gothenburg)	tba
11	2019-06-19	Feiyu Xu (Lenovo) <i>to be confirmed</i>	tba
12	2019-06-26	Milica Gasic (Düsseldorf)	tba
13	2019-07-03	Raquel Fernández (Amsterdam)	tba
14	2019-07-10	<i>no talk scheduled</i>	
15	2019-07-17	<i>no talk scheduled</i>	

← moved to our slot (10-12)