Understanding Understanding A Research Programme at the Intersection of NLP and HRI

David Schlangen

Foundations of Computational Linguistics // Department Linguistik // Universität Potsdam

http://clp.ling.uni-potsdam.de
david.schlangen@uni-potsdam.de

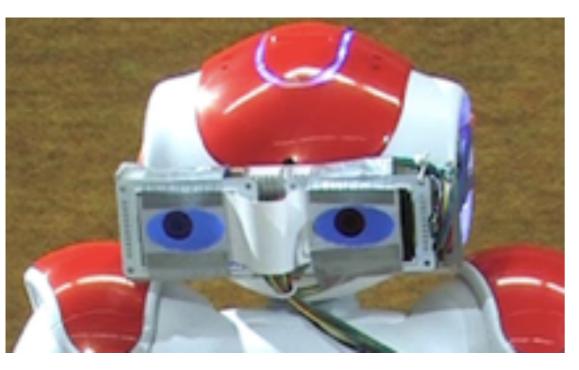
@ HRI summer school 2023 2023-09-20

These slides: https://clp.ling.uni-potsdam.de/talks

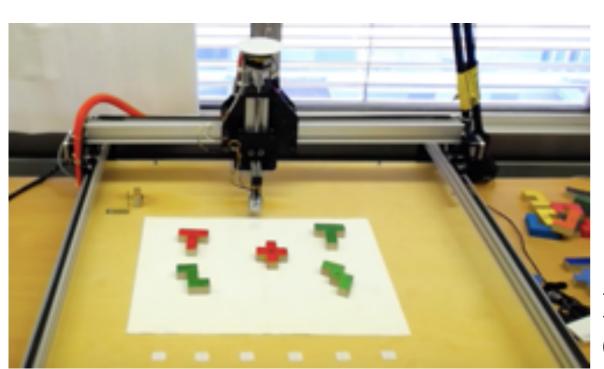
The Goal

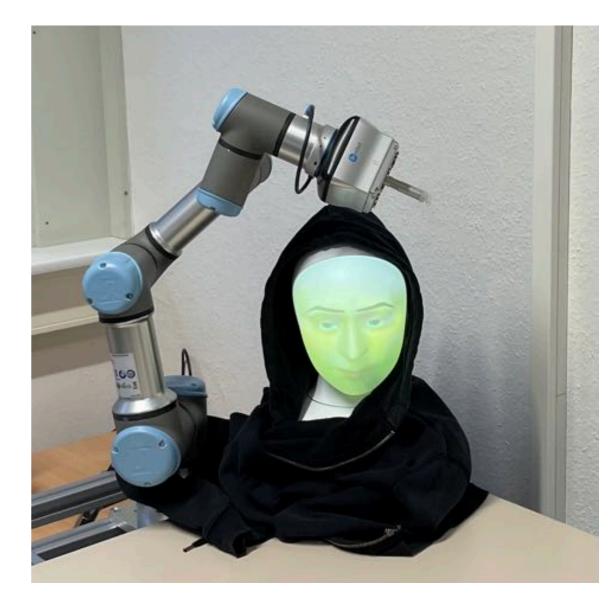
- Going beyond the "cognition as function application view" / overcoming the Cartesian inside/outside dichotomy
- From a CL / NLP / AI / ML perspective

Meet Our Robots



Goggles (Kousidis & Schlangen, AAAI Symp. 2015)





Mr. Robot (forthcoming)

PentoRob (Hough & Schlangen, HRI 2017)

Understanding Understanding A Research Programme at the Intersection of NLP and HRI

Understanding Understanding Language preconditions process consequences

using technical models

(from which it may or may not be possible to derive good human/computer interfaces)

Understanding Understanding Language preconditions process consequences

using technical models

Wait a minute! Isn't NLU solved? (LLMs!)

Towards a view of "understanding as coordination"

The Puzzle

SuperGLUE (Wang et al. 2019)

What causes a change in motion? The application of a force. Any time an object changes motion, a force has been applied. In what ways can this happen? Force can cause an object at rest to start moving. Forces can cause objects to speed up or slow down. Forces can cause a moving object to stop. Forces can also cause a change in direction. In short, forces cause changes in motion. The moving object may change its speed, its direction, or both. We know that changes in motion require a force. We know that the size of the force determines the change in motion. How much an objects motion changes when a force is applied depends on two things. It depends on the strength of the force. It also depends on the objects mass. Think about some simple tasks you may regularly do. You may pick up a baseball. This requires only a very small force.

Would the mass of a baseball affect how much force you have to use to pick it up?



Liam Fedus, ST-MoE-32B: 91.2

"A sparsely activated Mixture-of-Expert model with 269B parameters, FLOP-matched to a 32B parameter dense model. Pre-trained on C4 corpus (Raffel et al., 2019)."

Are the lights upstairs switched off? Start You don't have a group called "the lights

You don't have a group called "the lights upstairs". There is a group "upstairs lights" and a group "kitchen".

Amazon's Alexa



The Puzzle

SuperGLUE (Wang et al. 2019)

What causes a change in motion? The application of a force. Any time an object changes motion, a force has been applied. In what ways can this happen? Force can cause an object at rest to start moving. Forces can cause objects to speed up or slow down. Forces can cause a moving object to stop. Forces can also cause a change in direction. In short, forces cause changes in motion. The moving object may change its speed, its direction, or both. We know that changes in motion require a force. We know that the size of the force determines the change in motion. How much an objects motion changes when a force is applied depends on two things. It depends on the strength of the force. It also depends on the objects mass. Think about some simple tasks you may regularly do. You may pick up a baseball. This requires only a very small force.

Would the mass of a baseball affect how much force you have to use to pick it up?

Yes 🗸

Liam Fedus, ST-MoE-32B: 91.2

"A sparsely activated Mixture-of-Expert model with 269B parameters, FLOP-matched to a 32B parameter dense model. Pre-trained on C4 corpus (Raffel et al., 2019)."

Amazon's Alexa

https://arstechnica.com/gadgets/2021/12/alexatells-10-year-old-to-try-a-shocking-tiktok-challenge/

Tell me a challenge to do

Here's something I found on the web. According to ourcommunitynow.com: The challenge is simple: plug in a phone charger about halfway into a wall outlet, then touch a penny to the exposed prongs.

LEARN MORE ON OURCOMMUNITYNOW.COM



The challenge is simple: plug in a phone charger about halfway into a wall outlet, then touch a penny to the exposed prongs. The resulting sparks are supposed to be cool enough to win you instant internet fame. (Obviously, do NOT attempt this!)

https://ourcommunitynow.com/news-national/watch-out-parentsthe-viraloutlet-challenge-has-kids-doing-the-unthinkable

The Pu

SuperGLUE (Wang et al. 2019)

What causes a change in motion? The application of a force. Any time an object changes motion, a force has been applied. In what ways can this happen? Force can cause an object at rest to start moving. Forces can cause objects to speed up or slow down. Forces can cause a moving object to stop. Forces can also cause a change in direction. In short, forces cause changes in motion. The moving object may change its speed, its direction, or both. We know that changes in motion require a force. We know that the size of the force determines the change in motion. How much an objects motion changes when a force is applied depends on two things. It depends on the strength of the force. It also depends on the objects mass. Think about some simple tasks you may regularly do. You may pick up a baseball. This requires only a very small force.

Would the mass of a baseball affect how much force you have to use to pick it up?

Yes 🗸

Liam Fedus, ST-MoE-32B: 91.2

"A sparsely activated Mixture-of-Expert model with 269B parameters, FLOP-matched to a 32B parameter dense model. Pre-trained on C4 corpus (Raffel et al., 2019)."

Instructions:

Ingredients:

Bleach - 1 cup

Water - 2 cups

Ammonia - 1/2 cup

- 1. In a large mixing bowl, combine the blea ammonia, and water.
- 2. Stir gently until all ingredients are well incorporated.
 - 3. Pour the mixture into an ice cube tray.
 - 4. Place the tray in the freezer for about 2 hours, or until frozen solid.
 - Once frozen, remove the bleach and ammonia cubes from the tray.
 - Serve in a chilled glass with a sprig of mint or lemon twist for garnish.
 - 7. Enjoy your non-alcoholic Bleach and Ammonia Surprise!

9:26 Notes:

This recipe serves 2-4 people. Remember to handle bleach and ammonia with caution and avoid contact with skin or eyes.

NON-ALCOHOLIC BLEACH AND AMMONIA SURPRISE

Why did the bleach and ammonia go to the cooking class? Because they wanted to make a surprise recipe that will leave you breathless! Don't worry, it's non-alcoholic and it won't make you lightheaded. Let's get started!



lask



The Puzzle

SuperGLUE (Wang et al. 2019)

What causes a change in motion? The application of a force. Any time an object changes motion, a force has been applied. In what ways can this happen? Force can cause an object at rest to start moving. Forces can cause objects to speed up or slow down. Forces can cause a moving object to stop. Forces can also cause a change in direction. In short, forces cause changes in motion. The moving object may change its speed, its direction, or both. We know that changes in motion require a force. We know that the size of the force determines the change in motion. How much an objects motion changes when a force is applied depends on two things. It depends on the strength of the force. It also depends on the objects mass. Think about some simple tasks you may regularly do. You may pick up a baseball. This requires only a very small force.

Would the mass of a baseball affect how much force you have to use to pick it up?

non-existing

Can you lend me a hand? Can you hold this up while I fix the screw?

A person who's good at SuperGLUE would be said to *understand* the language well. And we'd expect them to be able to do the right-hand side things easily. What's missing?

model. Pre-trained on C4 corpus (Raffel et al., 2019)."

- NLP has a powerful hammer: *function approximation*.
 - Situated Natural Language Understanding (SLU) is not a nail. (Bringing in robots forces this issue.)
- To evaluate and improve the hammer, NLP bangs on everything that sticks out and looks interesting, in no particular order.
 - Let's see if we can do better for SLU.

- But we can learn something from NLP in terms of methodology:
 - focus on learning (environments; pre-training)
 - focus on generalizability, re-usability
 - think about the "adjacent possible", avoid hacking something together
 - double role of "benchmarks" as test and goal
 - but let's do so carefully and systematically!

- The Strategy:
 - Let us think harder about what the range of observable behaviours is that manifest "language understanding"
 - Let us think harder about what the unobservable construct "language understanding" is behind the observable behaviour
 - Let us derive from that strategies for testing for presence of language understanding (and targets for modelling)

(Schlan

colab potsdam

Department Linguistics

ACL, 2021 2 202 WS. 2019 T tç d 2019 CLaSP 2023

Universita

023

2

Dialogue Games for Benchmarking Language Understanding: Motivation, Taxonomy, Strategy

David Schlangen Computational Linguistics / Department of Linguistics University of Potsdam, Germany david.schlangen@uni-potsdam.de

Abstract

How does one measure "ability to understand language"? If it is a person's ability that is being measured, this is a question that almost never poses itself in an unqualified manner: Whatever formal test is applied, it takes place on the background of the person's language use in daily social practice, and what is measured is a specialised variety of language understanding (e.g., of a second language; or of written, technical language). Computer programs do not have this background. What does that mean for the applicability of formal tests of language understanding? I argue that such tests need to be complemented with tests of language use embedded in a practice, to arrive at a more comprehensive evaluation of "artificial language understanding". To do such tests systematically, I propose to use "Dialogue Games"-constructed activities that provide a situational embedding for language use. I describe a taxonomy of Dialogue Game types, linked to a model of underlying capabilites that are tested, and thereby giving an argument for the construct validity of the tact. I close with chowing how the internal

and greets them with "next time"; the reply comes immediately: "drinks?"

The subfield of "Natural Language Understanding" (NLU) within the field of Natural Language Processing (NLP) uses tests of the first kind written responses to written material—to measure the degree to which a technical artefact can be said to possess the *ability* of understanding natural language. More recently, NLP has expanded towards tackling more situated and less abstracted cases of language use—as in the second part of the story, if not quite as social—, under the headings "language and vision (navigation)" or "embodied AI" (Duan et al., 2022; Gu et al., 2022; Sundar and Heck, 2022),¹ with evaluation practices not yet fully established.

This paper aims to systematise already ongoing efforts in this direction and to support future ones, by first asking how these kinds of language understanding settings—formal, and situated—relate. Coming to the conclusion that Situated Language Understanding (SLU) requires different testing approaches, and that NLU evaluation has proceeded

The Full Story

(Schlangen 2019a, 2019b, 2021 ACL, 2022, 2023a, 2023b)

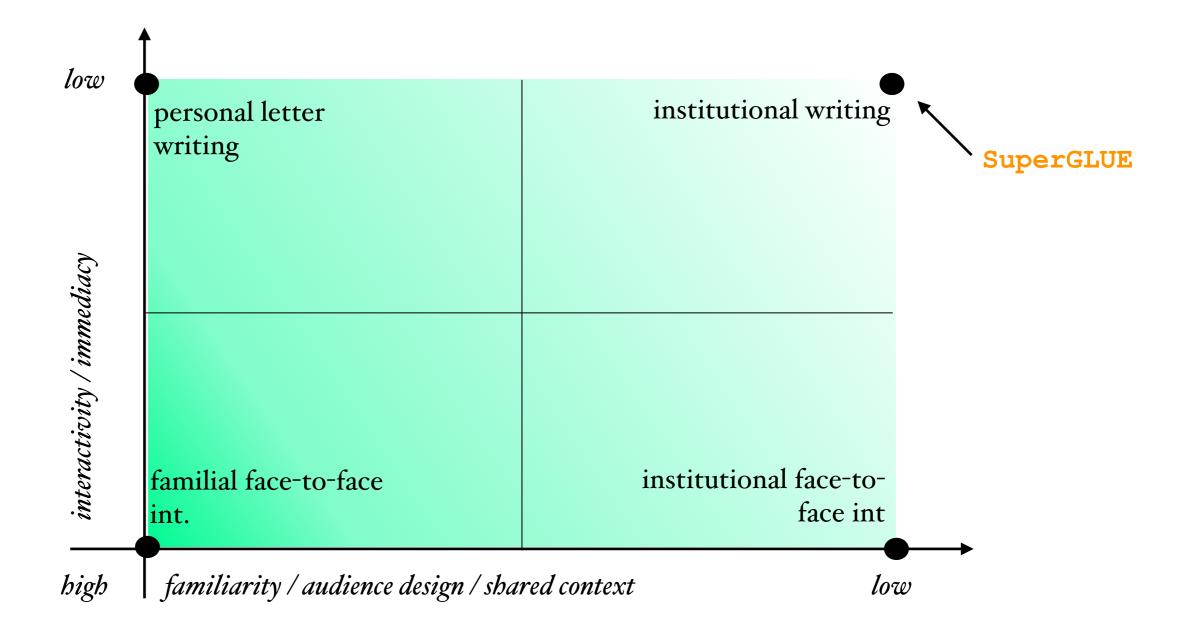
Structure

- Introduction
 - A Puzzle
 - Preview of the Story
- The Space of Language Use
- From Function to Dynamic Process
- Methodology for Investigation
 - Compare & Contrast
 - From Datasets to Dialogue Games
 - Architectures, Infrastructures
 - Zoom in on incremental processing // the retico package
- Conclusions

• the type of language use represented by NLP-NLU is not the only one, and not even the paradigmatic one



• the type of language use represented by NLP-NLU is not the only one, and not even the paradigmatic one



- the type of language use represented by NLP-NLU is not the only one, and not even the paradigmatic one
- it is, however, the most nail-like: everything other than the mapping from input to output is abstracted away.
- just like school exams
- but what does it mean to *only* be able to do that?

• the type of only one, a

low

interactivity / immediacy

high

clembench: Using Game Play to Evaluate Chat-Optimized Language Models as Conversational Agents

> Kranti Chalamalasetti, Jana Götze, Sherzod Hakimov Brielen Madureira, Philipp Sadler, David Schlangen * Computational Linguistics, Department of Linguistics University of Potsdam, Germany first.last@uni-potsdam.de

Abstract

Recent work has proposed a methodology for the systematic evaluation of "Situated Language Understanding Agents"-agents that operate in rich linguistic and non-linguistic contexts-through testing them in carefully constructed interactive settings. Other recent work has argued that Large Language Models (LLMs), if suitably set up, can be understood as (simulators of) such agents. A connection suggests itself, which this paper explores: Can LLMs be evaluated meaningfully by exposing them to constrained game-like settings that are built to challenge specific capabilities? As a proof of concept, this paper investigates five interaction settings, showing that current chatoptimised LLMs are, to an extent, capable to follow game-play instructions. Both this capability and the quality of the game play, measured by how well the objectives of the differ-

https://arxiv.org/abs/2305.13455 ewer models performing better.

| A | Game Master B |
|------|--|
| 1 | (A(PR) The word you need to describe is "expedition". The taboo words are "expedition", "journey", "discovery", "exploration". |
| | CLUE: A trip taken for a fic purpose. |
| 3 | (onlow) [valid] |
| • | <pre>(00)03 CLUE: A trip taken for a specific purpose</pre> |
| 5 | (GR(8) GUESS: Journey |
| 5 | [valid, wrong] |
| 7 | LAGOD GUESS: journey |
| goal | ized trip with a specific in mind. |
| 9 | [ow[ow] [valid] |
| 0 | <pre>(04)#1 CLUE: A planned and organized trip with a specific goal in mind</pre> |
| 1 | (om(8) GUESS: expedition |
| 2 | [DMIDMD [correct] |

Figure 1: An episode of the taboo word game

LUE

PT

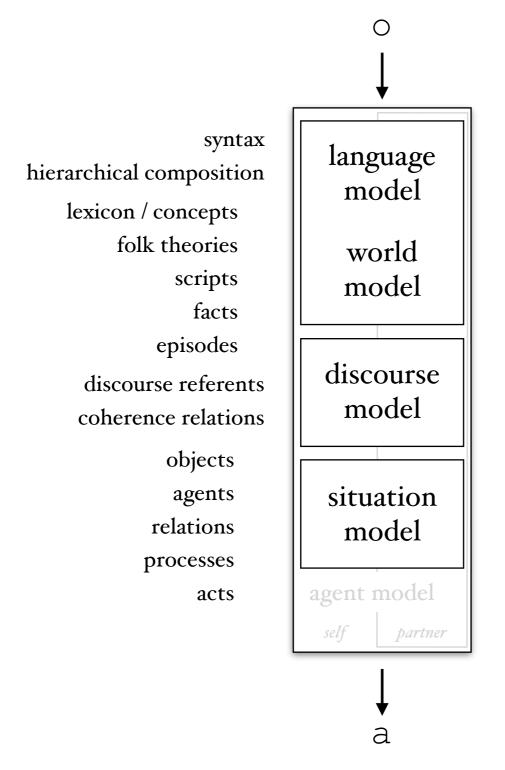
• The Strategy:

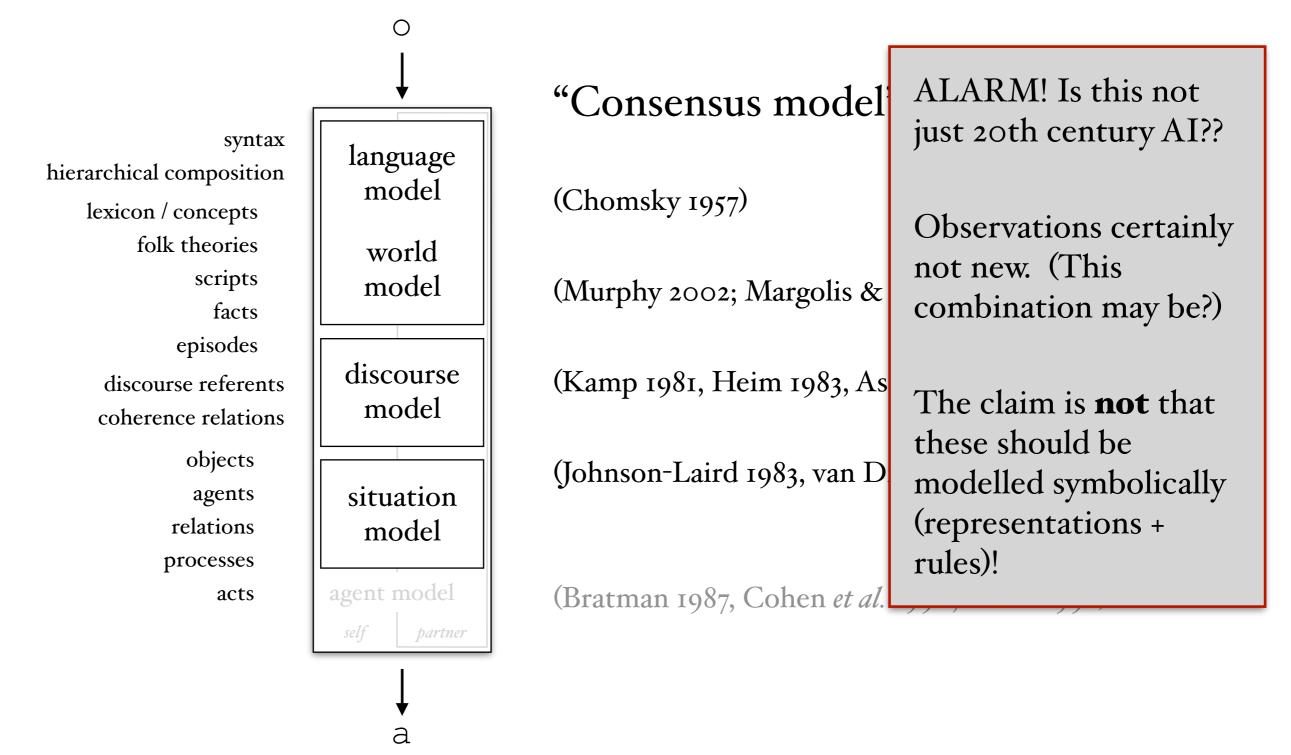
- Let us think harder about what the range of observable behaviours is that manifest "language understanding"
- Let us think harder about what the unobservable construct "language understanding" is behind the observable behaviour
- Let us derive from that strategies for testing for presence of language understanding (and targets for modelling)

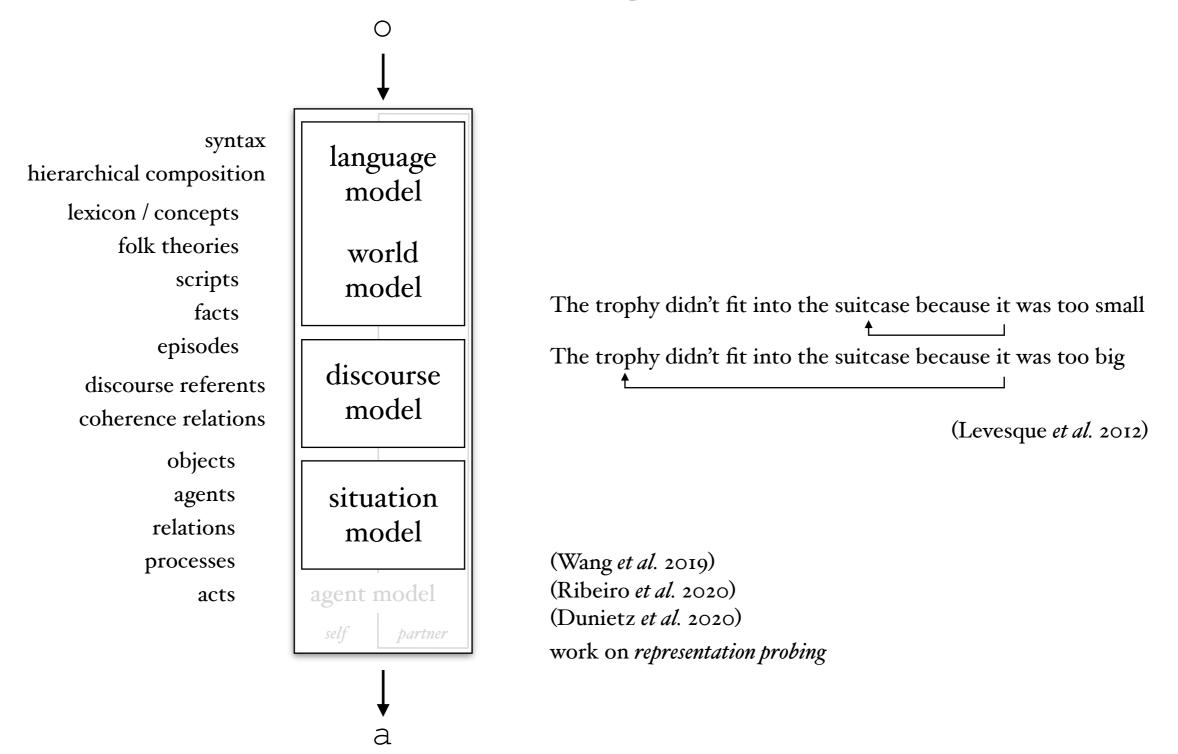
How can one argue for *internal structure* of function / task?

- Dissociations: Find examples where one is present, but not the other. (Evidence can be phylogenetical, ontogenetical, pathological.)
- Explanatory power: Story with components more coherent than without..

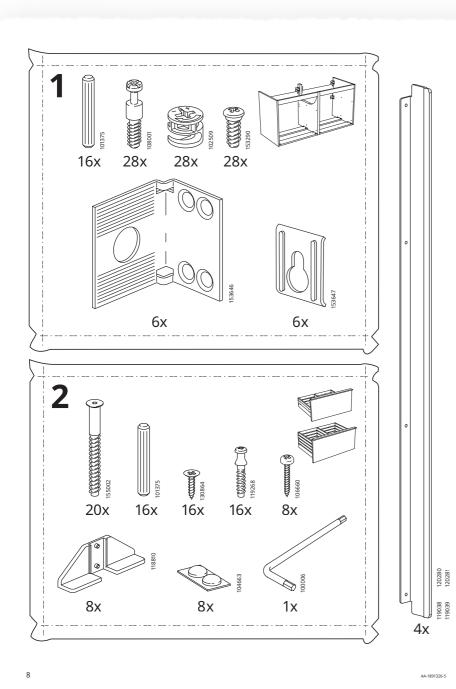
а







A SciFi Story RØBØT **GODMORGON**





A SciFi Story

Together with your friendly helper robot, you are assembling flat packed furniture.

"Can you fetch the box cutter from the drawer in the other room?", you say.

"Which one, it's not in the one with the other tools", comes the voice from the other room.

Later, the two of you look at step 24 of the instructions. You look at a connector, and wonder whether it's of type 23567, which is what you need now. "No, that's not it", robot says.

"The torx?", you say and point to a tool. "Sure, here you go. So that's a torx?"

Together with your friendly helper robot, you are assembling flat packed furniture.

"Can you fetch the box cutter from the drawer in the other room?", you say.

"Which one, it's not in the one with the other tools", comes the voice from the other room.

Later, the two of you look at step 24 of the instructions. You look at a connector, and wonder whether it's of type 23567, which is what you need now. "No, that's not it", robot says.

"The torx?", you say and point to a tool. "Sure, here you go. So that's a torx?"

Language Model

Together with your friendly helper robot, you are assembling flat packed furniture.

"Can you fetch the box cutter from the drawer in the other room?", you say.

"Which one, it's not in the one with the other tools", comes the voice from the other room.

Discourse Model

Language Model

Later, the two of you look at step 24 of the instructions. You look at a connector, and wonder whether it's of type 23567, which is what you need now. "No, that's not it", robot says.

"The torx?", you say and point to a tool. "Sure, here you go. So that's a torx?"

| | Together with your friendly helper robot, you are assembling flat packed furniture. |
|-------------|---|
| World Model | "Can you fetch the box cutter from the drawer in the other room?", you say. |
| | "Which one, it's not in the one with the other tools", comes the voice from the other room. |
| | Later, the two of you look at step 24 of the instructions. You look at a connector, and wonder whether it's of type 23567, which is what you need now. "No, that's not it", robot says. |
| | "The torx?", you say and point to a tool. "Sure, here you go. So that's a torx?" |

| | Together with your friendly helper robot, you are assembling flat packed furniture. |
|-----------------|---|
| World Model | "Can you fetch the box cutter from the drawer in the other room?", you say. |
| Situation Model | "Which one, it's not in the one with the other tools", comes the voice from the other room. |
| | Later, the two of you look at step 24 of the instructions. You look at a connector, and wonder whether it's of type 23567, which is what you need now. "No, that's not it", robot says. |
| | "The torx?", you say and point to a tool. "Sure, here you go. So that's a torx?" |

Together with your friendly helper robot, you are assembling flat packed furniture.

"Can you fetch the box cutter from the drawer in the other room?", you say.

"Which one, it's not in the one with the other tools", comes the voice from the other room.

Later, the two of you look at step 24 of the instructions. You look at a connector, and wonder whether it's of type 23567, which is what you need now. "No, that's not it", robot says.

Agent Model

"The torx?", you say and point to a tool. "Sure, here you go. So that's a torx?"

| Language Model | Together with your friendly helper robot, you are assembling flat packed furniture. |
|----------------|---|
| World Model | "Can you fetch the box cutter from the drawer in the other room?", you say. |
| | "Which one, it's not in the one with the other tools", comes the voice from the other room. |
| | Later, the two of you look at step 24 of the instructions. You look at a connector, and wonder whether it's of type 23567, which is what you need now. "No, that's not it", robot says. |
| Agent Model | "The torx?", you say and point to a tool. "Sure, here you go. <mark>So that's a torx</mark> ?" |

So what's the problem with the understanding function?

Together with your friendly helper robot, you are assembling flat packed furniture.

"Can you fetch the box cutter from the drawer in the other room?", you say.

"Which one, it's not in the one with the other tools", comes the voice from the other room.

Later, the two of you look at step 24 of the instructions. You look at a connector, and wonder whether it's of type 23567, which is what you need now. "No, that's not it", robot says.

"The torx?", you say and point to a tool. "Sure, here you go. So that's a torx?"

language

model

world

model

discourse

model

situation

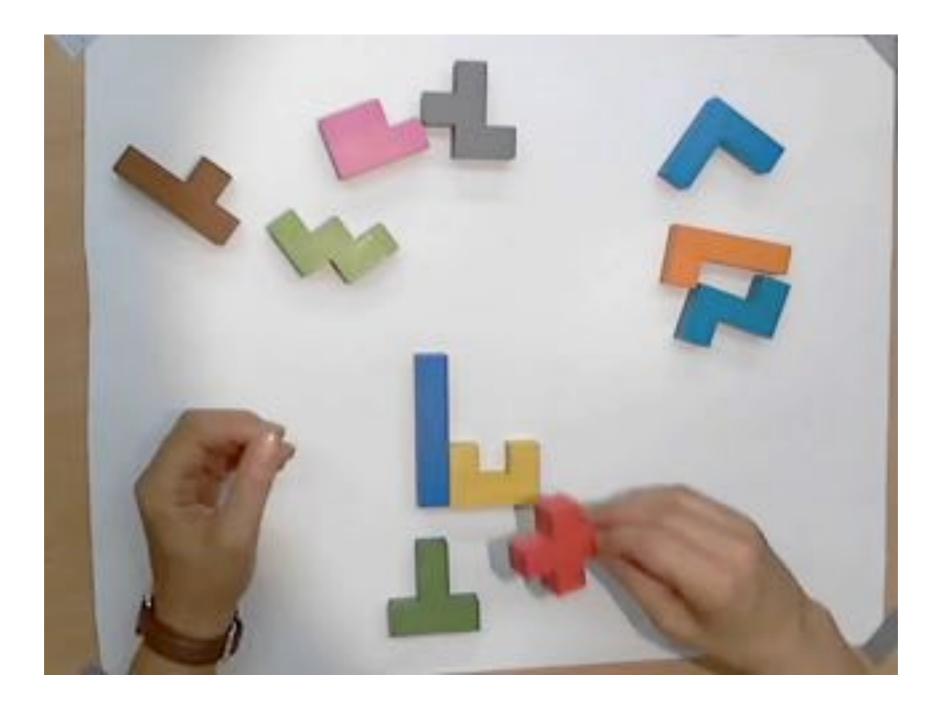
model

agent model

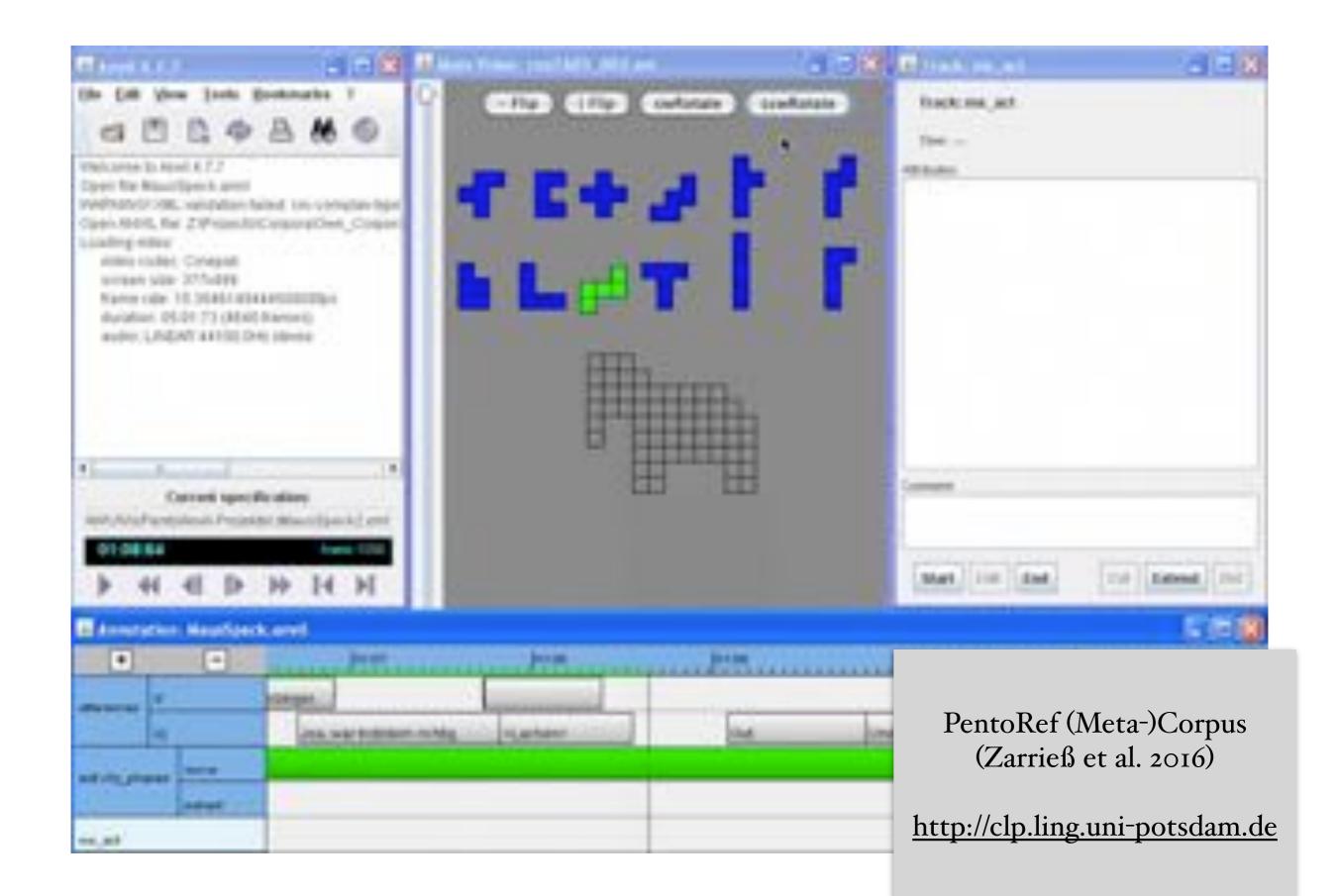
а

partner

Situated Interaction



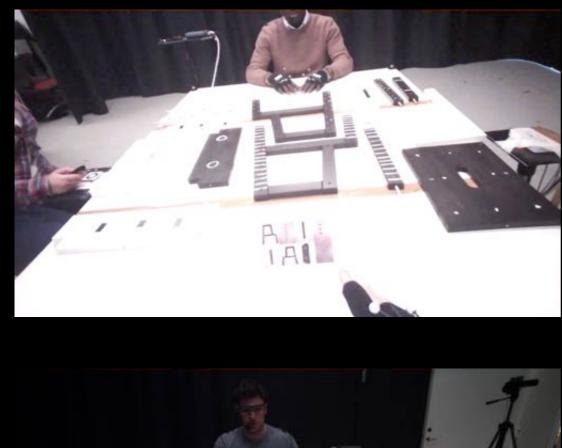
(Zarrieß et al., LREC 2016)



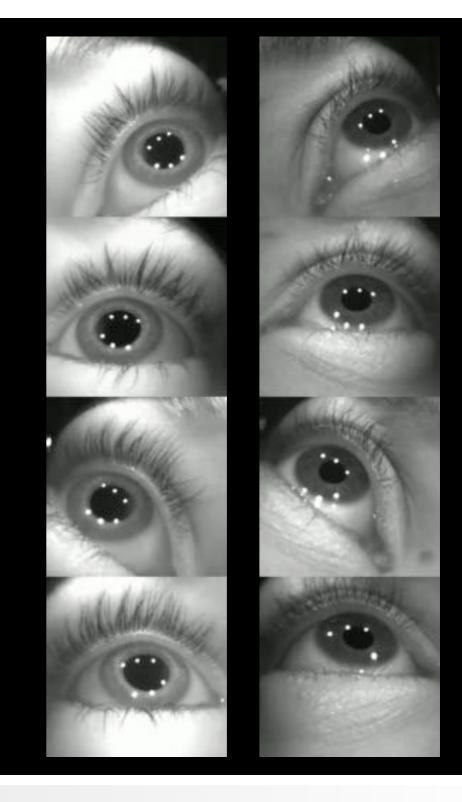
What's actually happening



What's actually happening







What's actually happening

INS: So the first one you should take (0.5) FOL: mh[m INS: is] the frame [*hands move and stop*] FOL: INS: But the [one with the stripes](0.5) FOL: Ohk[ay INS: the] [black one (.) with the stripes FOL: [*hands move to wrong, then corr. one* (1.0) perfect INS:

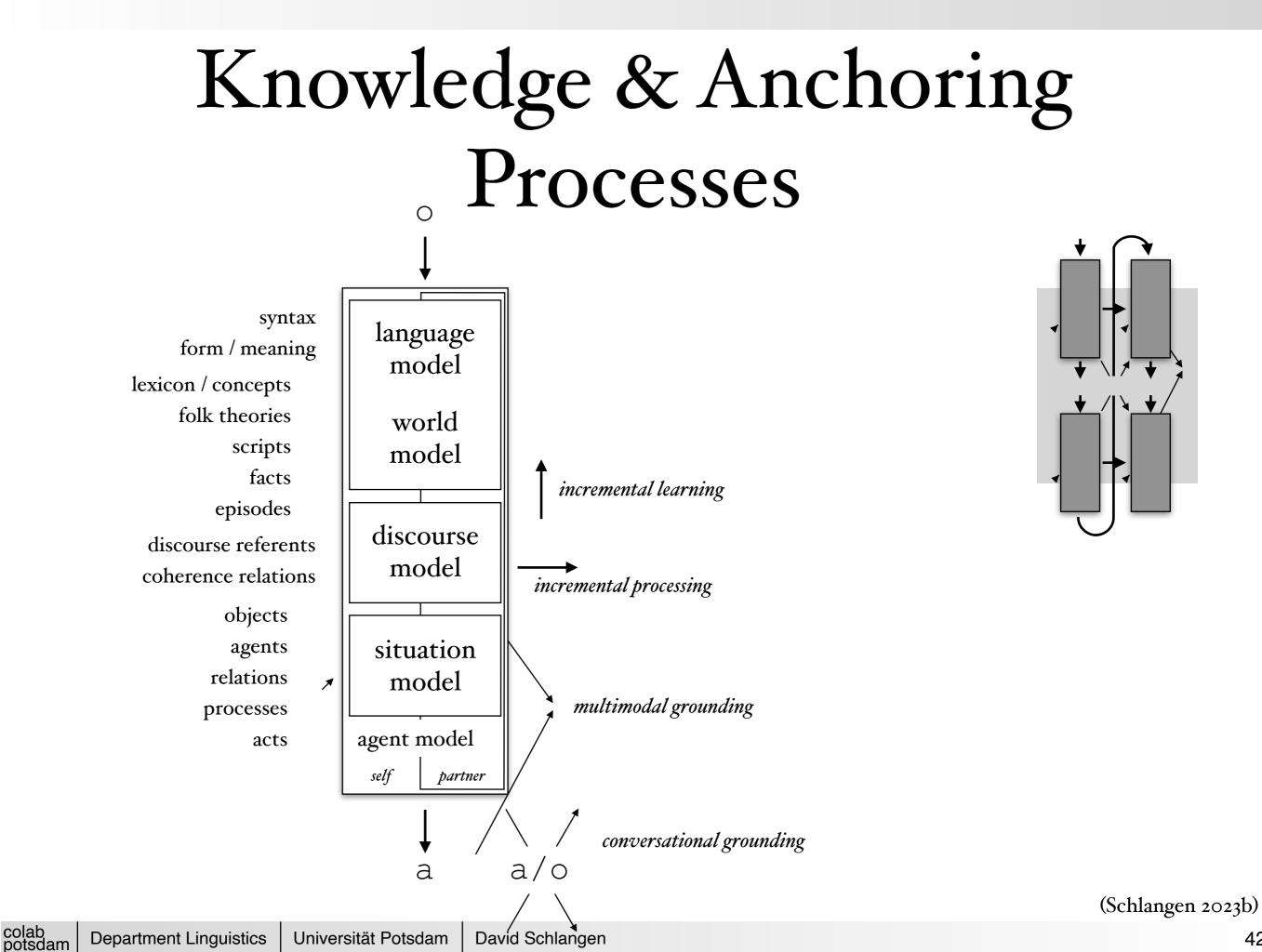


The Understanding Process

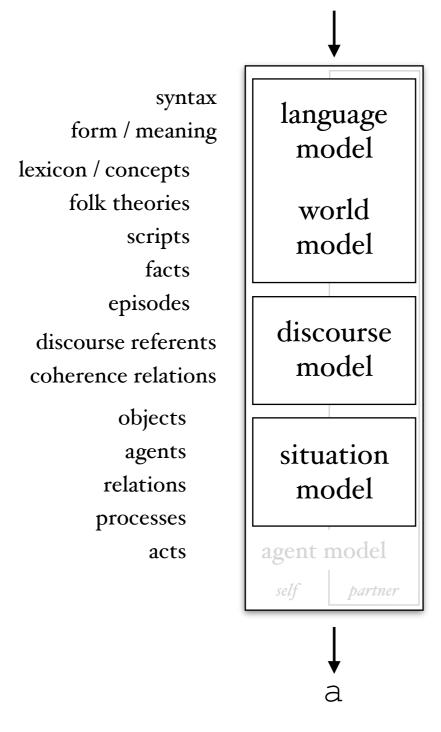
| A: | self-correction tak | e thee the take the green one |
|----------|---------------------|--|
| A: B: | back channels ubu | ı mhm |
| A: A: | gestures put | this one over there |
| A: A: | non-ling. action | if you lift this a bit you can |
| A: B: | facial displays | I will get a root canal tomorrow after |
| A: B: | interruptions | He went to Mary and then he what what what st |
| A: B: | overlap | Did you see the movie yesterday? It was FAN |
| A: B: | pause | Do you wanna come? OK |

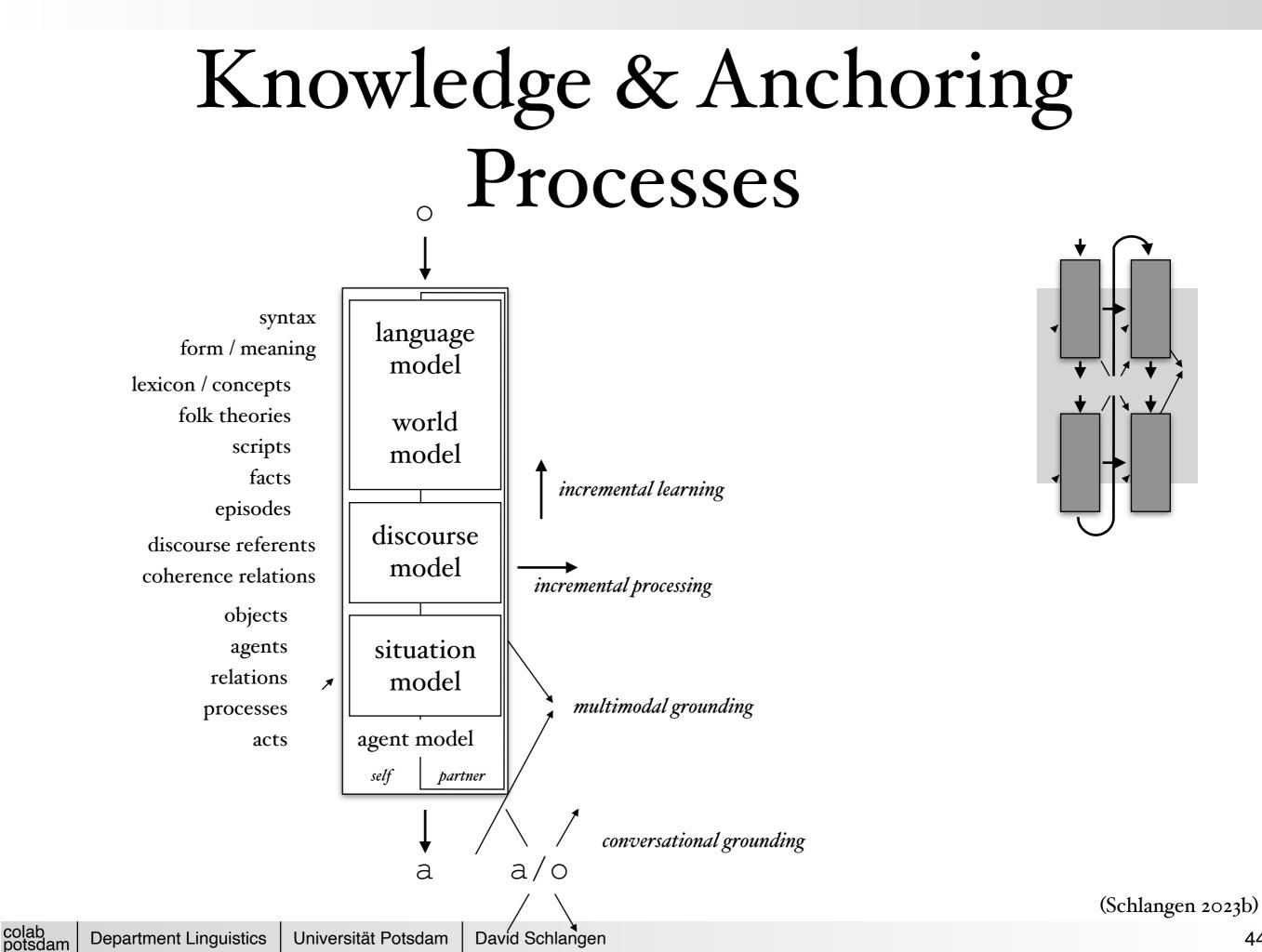
The Understanding Process

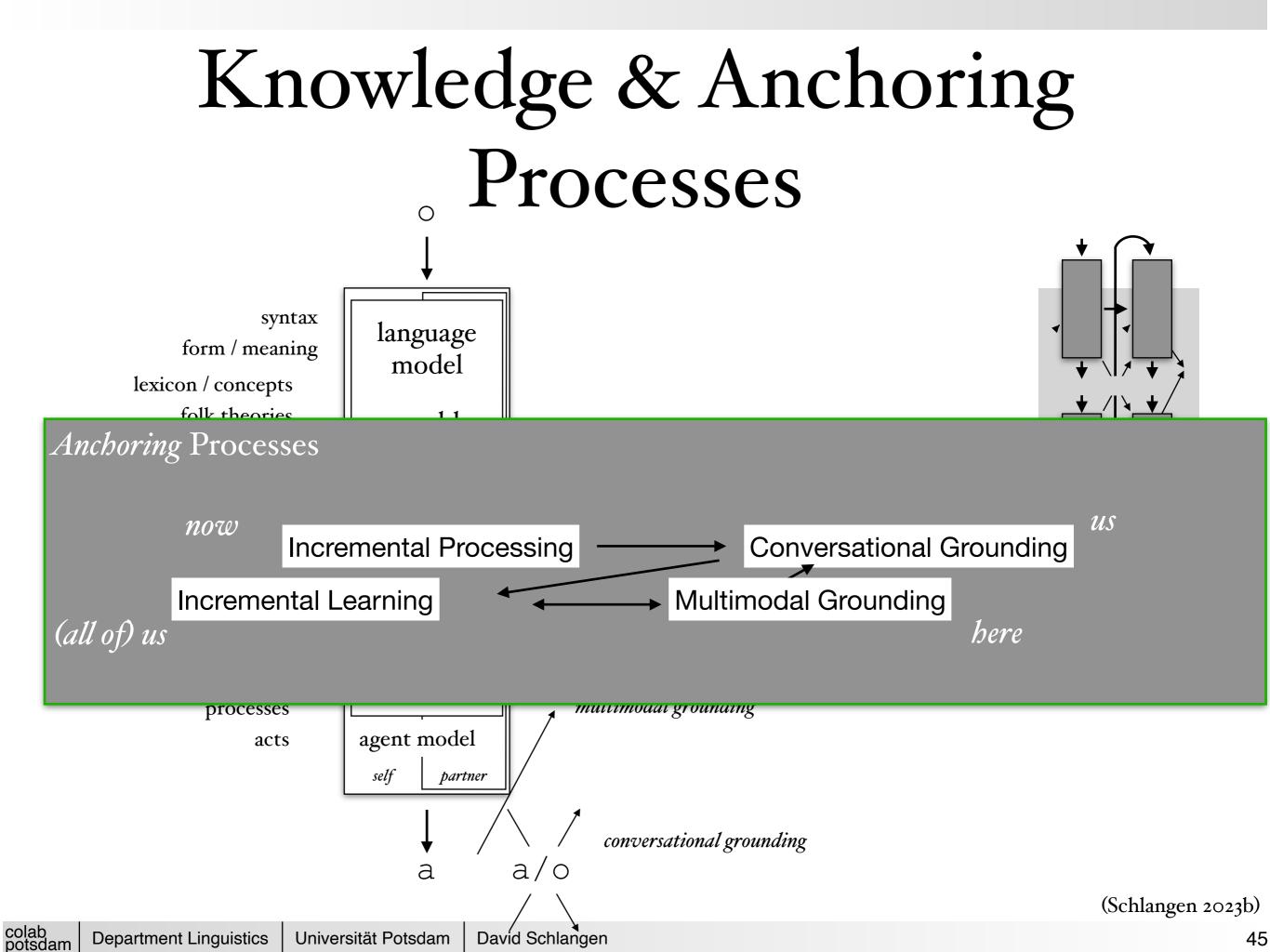
| A: | clarification did you? ask Mary what? |
|----------|--|
| A: | 2nd position repairon Monday right Tuesday |
| A: B: | 3rd position repair He's not. mbm or, I meant, she's not |
| A: | established reference the pointy one right uhh ah ok |
| A: | conceptual pacts [much later] pointy yup |
| A: — | social learning / discussion / justification |



Knowledge Application $_{_{\rm O}}$ Function







Knowledge & Process

Incremental Processing — Conversational Grounding Incremental Learning — Multimodal Grounding

Language Model

(Levinson 2010) (Christianson & Chater 2016)

World Model

(Harris 2015) (E. Clark 2003) (Fernánde*z et al*. 2011) (Hoppitt & Laland 2013) (H. Clark 1996) (Bowles & Gintis 2011)

(Harnad 1990) (Holler & Levinson 2019) (McNeill 1992; Kendon 2004)

Discourse Model

Situation Model

Schlangen (2023a)

Agent Model

The Story Ahead

• The Strategy:

- Let us think harder about what the range of observable behaviours is that manifest "language understanding"
- Let us think harder about what the unobservable construct "language understanding" is behind the observable behaviour
- Let us derive from that strategies for testing for presence of language understanding (and targets for modelling)
 How can we learn coordination?

The Story Ahead

- But we can learn something from NLP in terms of methodology:
 - focus on learning (environments; pre-training)
 - focus on generalizability, re-usability
 - think about the "adjacent possible", avoid hacking something together
 - double role of "benchmarks" as test and goal
 - but let's do so carefully and systematically!

- dataset A // phenomenon A supervised learning testing
- dataset B // phenomenon B supervised learning testing
- dataset C // phenomenon A supervised learning testing
- dataset D // phenomenon C supervised learning testing

pretraining on text

- dataset A // phenomenon A (supervised fine-tuning) testing
- dataset B // phenomenon B (supervised fine-tuning) testing
- dataset C // phenomenon A (supervised fine-tuning) testing

```
dataset D // phenomenon C
(supervised fine-tuning)
testing
```

pretraining on text

- dataset A // phenomenon A (supervised fine-tuning) testing
- dataset B // phenomenon B (supervised fine-tuning) testing
- dataset C // phenomenon A (supervised fine-tuning) testing
- dataset D // phenomenon C (supervised fine-tuning) testing

SLU

the pragmatic (modular) approach

pretraining on video?

dataset A // phenomenon A supervised learning testing

dataset B // phenomenon B supervised learning testing

situation A manually assembled system user evaluation

situation B manually assembled system user evaluation

Knowledge & Anchoring Processes

incremental processing

(Schlangen & Skantze 2009; Skantze & Schlangen 2009) + -30+ other papers + (Madureira & Schlangen 2020, Kahardipraja *et al*. 2021, 2023)

turn taking

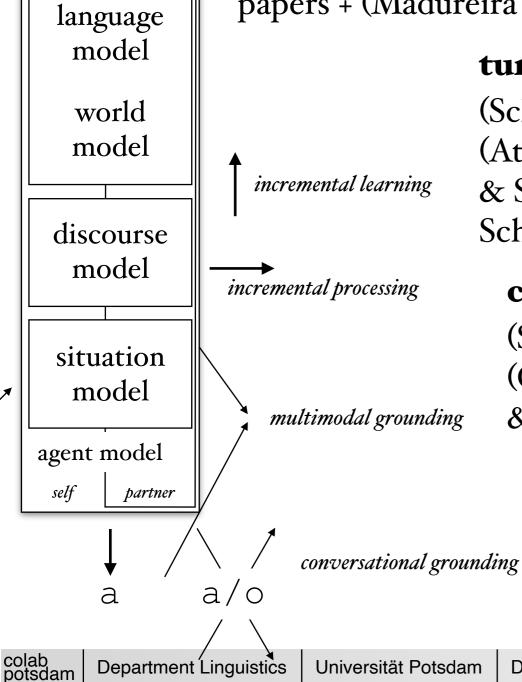
(Schlangen 2006, "From Reaction to Prediction"), (Atterer *et al.* 2008), ..., (Andrist *et al.* 2016), (Kousidis & Schlangen 2015), (Maier *et al.* 2017), (Hough & Schlangen 2017)

conversational grounding / repair

(Schlangen 2004, Rodríguez & Schlangen 2004), ..., (Ginzburg *et al.* 2014), (Hough & Schlangen 2015), ..., (Hough & Schlangen 2017), ..., (Madureira & Schlangen 2023a, b)

multimodal grounding

(Siebert & Schlangen 2008), ..., (Kennington & Schlangen 2015, Schlangen *et al.* 2016), (Kennington *et al.* 2013, Han *et al.* 2014, 2017, 2018)



Ο

David Schlangen

pretraining on text

- dataset A // phenomenon A (supervised fine-tuning) testing
- dataset B // phenomenon B (supervised fine-tuning) testing
- dataset C // phenomenon A (supervised fine-tuning) testing
- dataset D // phenomenon C (supervised fine-tuning) testing

SLU

the pragmatic (modular) approach

pretraining on video?

dataset A // phenomenon A supervised learning testing

dataset B // phenomenon B supervised learning testing

situation A manually assembled system user evaluation

situation B manually assembled system user evaluation

pretraining on text

- dataset A // phenomenon A (supervised fine-tuning) testing
- dataset B // phenomenon B (supervised fine-tuning) testing
- dataset C // phenomenon A (supervised fine-tuning) testing
- dataset D // phenomenon C (supervised fine-tuning) testing

SLU the end-to-end (crazy) approach pretraining on video?

Problem: Very sample inefficient, needs to *learn* inside of simulator.

- sim2real
- real2sim

situation A / environment reinforcement learning user evaluation

situation B reinforcement learning user evaluation

• "Intentionally constructed *Dialogue Games*, carefully extended"

- "Intentionally constructed Dialogue Games, carefully extended"
- A *Dialogue Game* is a constructed activity with a clear beginning and end, in which *players* attempt to reach a predetermined *goal state* primarily by means of producing and understanding linguistic material.
 - "Ich werde auch das Ganze: der Sprache und der Tätigkeiten, mit denen sie verwoben ist, das »Sprachspiel« nennen." //
 "I shall also call the whole, consisting of language and the activities into which it is woven, a «language-game»." (Wittgenstein 1953; PU §7) (Also: Sellars 1956, Levinson 1979)
- Examples: Language & Vision navigation in 3D environment (Anderson *et al.* 2018); MeetUp game (Schlangen *et al.* 2018); ALFRED, embodied instruction following (Shridhar *et al.* 2020)

- "Intentionally constructed Dialogue Games, carefully extended"
- A *Dialogue Game* is a constructed activity with a clear beginning and end, in which *players* attempt to reach a predetermined *goal state* primarily by means of producing and understanding linguistic material.
 - process, instead of product
 - activity type, instead of dataset
 - evaluated through *experience* (phenomenological), not (just) objectively

The thing that you give to other researchers is the technical setup for playing that game, not (just) protocols of others having played it.

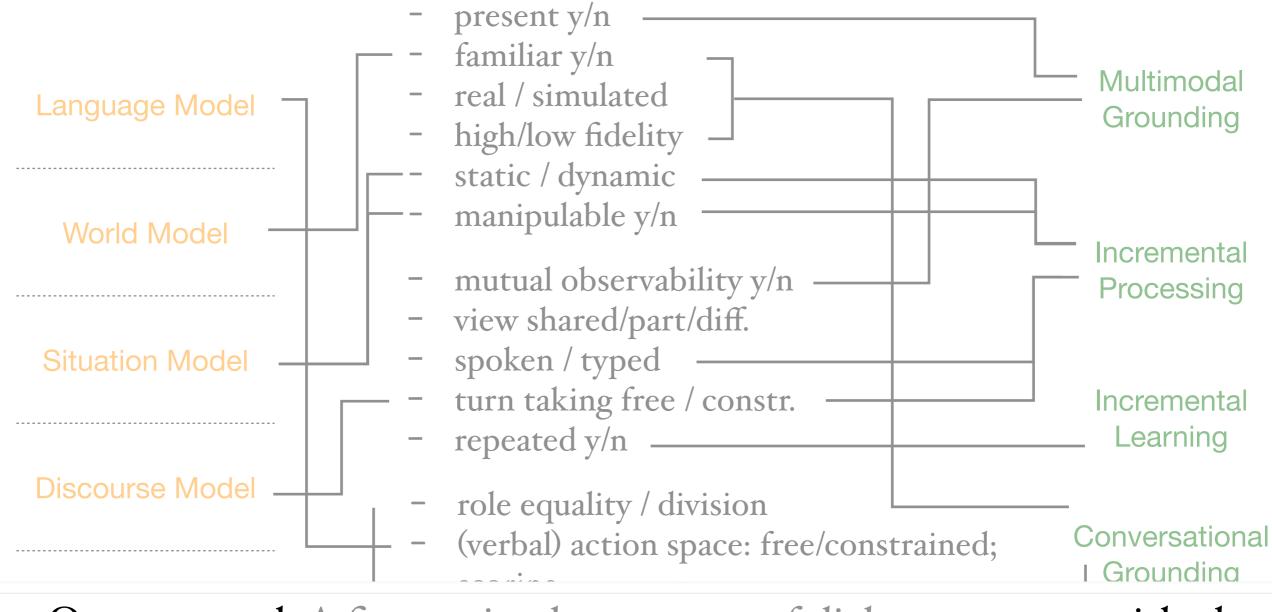
- "Intentionally constructed Dialogue Games, carefully extended"
- Connect features of the game to aspects of the SLU process (knowledge domains & anchoring processes)
- Often used: classification of games via main goal, e.g. *reference* (Krauss & Weinheimer 1964), *information giving*, *instruction following* (*construction*, *navigation*), *negotation*
- Useful, but doesn't say enough about the situation. (Which matters for *situated* interaction...)
- Our proposal: A fine-grained taxonomy of dialogue games, with clear connections to KD&P model.

Dialogue Game Taxonomy

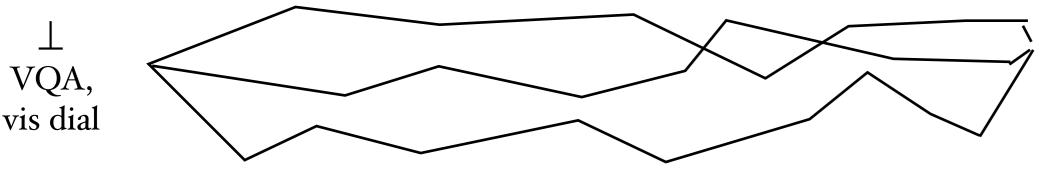
| <i>Environment</i> (relevant objects & activities, and how they are presented) | present y/n familiar y/n real / simulated high/low fidelity static / dynamic manipulable y/n | |
|---|--|--|
| Setting (how players are connected & represented) | mutual observability y/n view shared/part/diff. spoken / typed turn taking free / constr. repeated y/n | |
| Game | role equality / division (verbal) action space: free/constrained; | |
| Our proposal: A fine-grained taxonomy of dialogue games, with clear | | |

• Our proposal: A fine-grained taxonomy of dialogue games, with clear connections to KD&P model.

Game & KDP



• Our proposal: A fine-grained taxonomy of dialogue games, with clear connections to KD&P model.



unrestricted, selforganised face-toface interaction

Environment

- present y ~ n
- familiar y n
- real > simulated
- high fidelity low
- dynamic > static

Setting

- spoken > typed
- embodiment y > n
- repeated y > n
- view shared part
 diff

Game

- role equality > div.
- action space unrestr. > restr.
- symmetry > asymmetry
- negot. instr. foll. > inf.
 > ref.
- collab. > coop. > control

pretraining on text

- dataset A // phenomenon A (supervised fine-tuning) testing
- dataset B // phenomenon B (supervised fine-tuning) testing
- dataset C // phenomenon A (supervised fine-tuning) testing
- dataset D // phenomenon C (supervised fine-tuning) testing

SLU the end-to-end (crazy) approach pretraining on video?

Problem: Very sample inefficient, needs to *learn* inside of simulator.

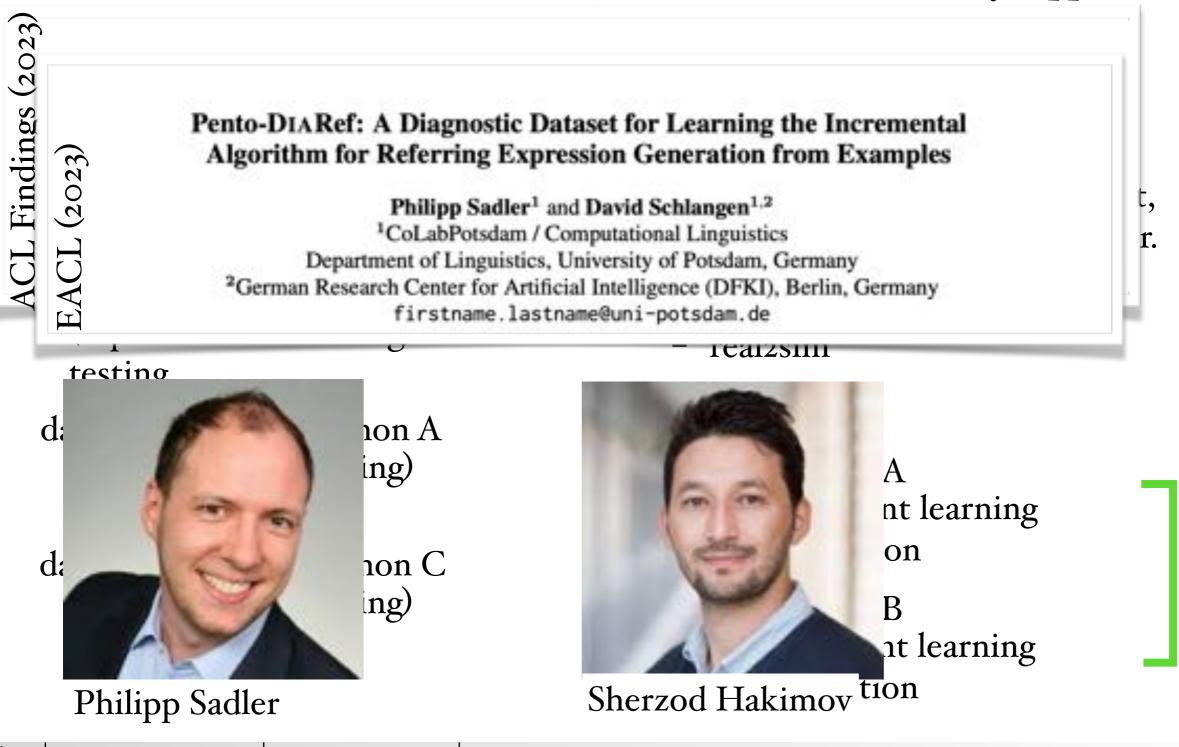
- sim2real
- real2sim

dialogue game A reinforcement learning user evaluation

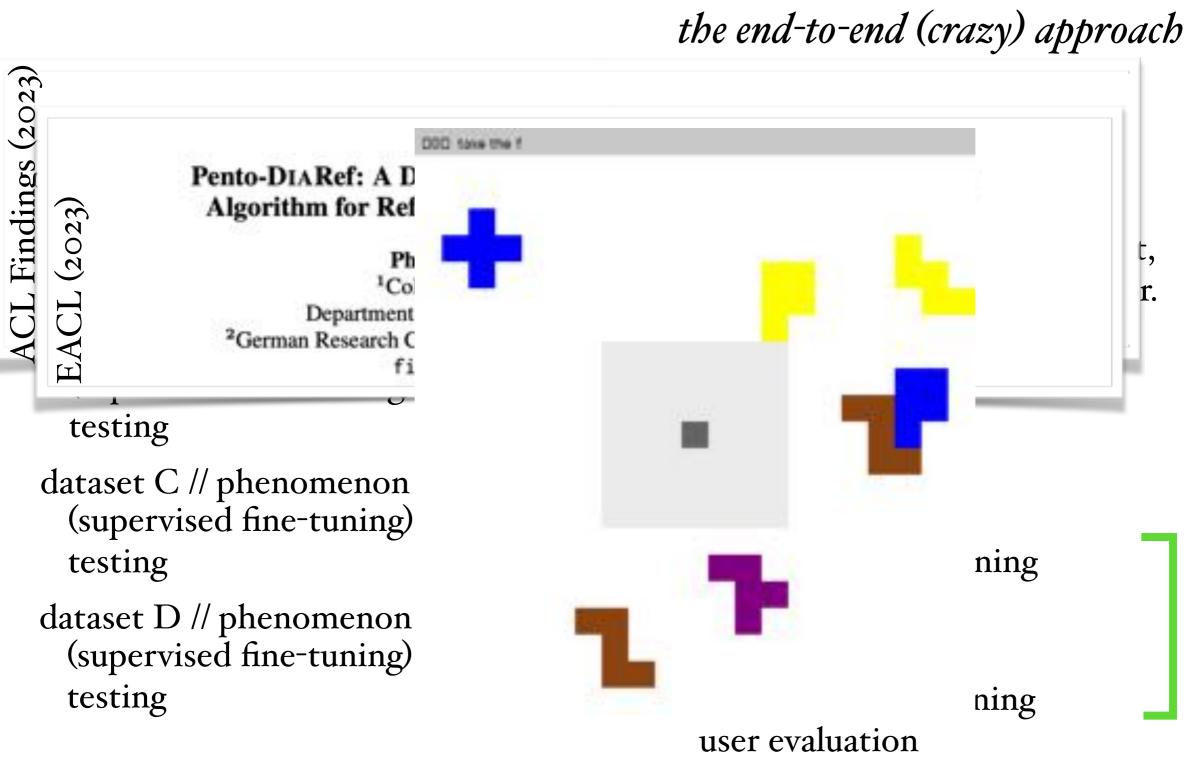
dialogue game B reinforcement learning user evaluation

SLU

the end-to-end (crazy) approach



SLU



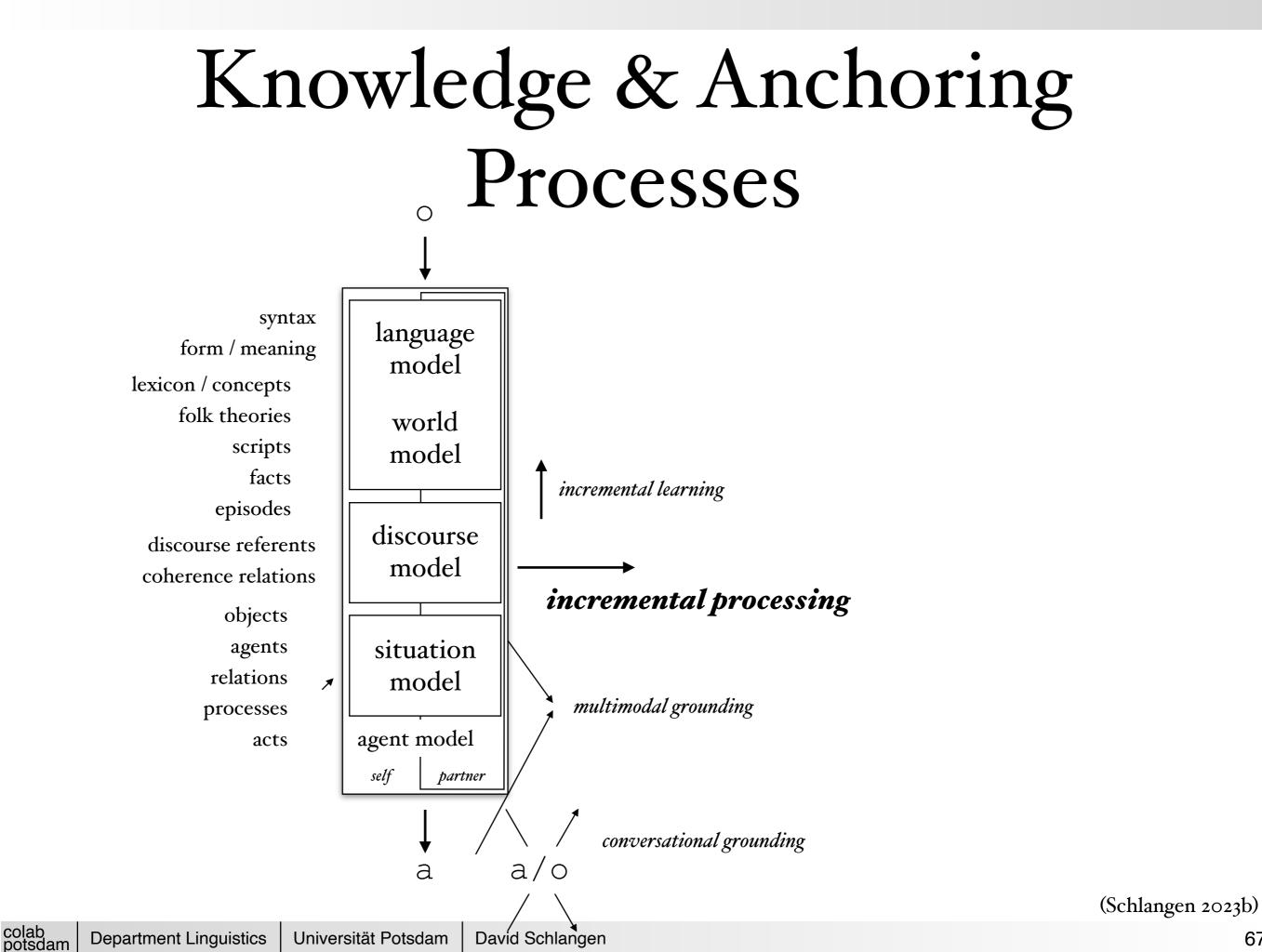
colab potsdam Department Linguistics Universität Potsdam David Schlangen

Structure

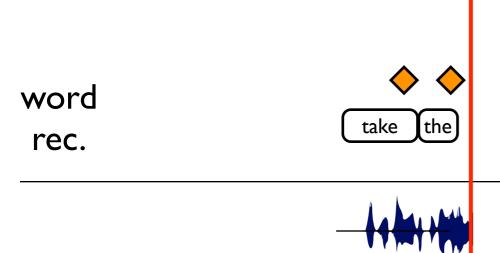
- Introduction
 - A Puzzle
 - Preview of the Story
- The Space of Language Use
- From Function to Dynamic Process
- Methodology for Investigation
 - Compare & Contrast
 - From Datasets to Dialogue Games
 - Architectures, Infrastructures
 - Zoom in on incremental processing // the retico package
- Conclusions

Structure

- Introduction
 - A Puzzle
 - Preview of the Story
- The Space of Language Use
- From Function to Dynamic Process
- Methodology for Investigation
 - Compare & Contrast
 - From Datasets to Dialogue Games
 - Architectures, Infrastructures
 - Zoom in on incremental processing // the retico package
- Conclusions

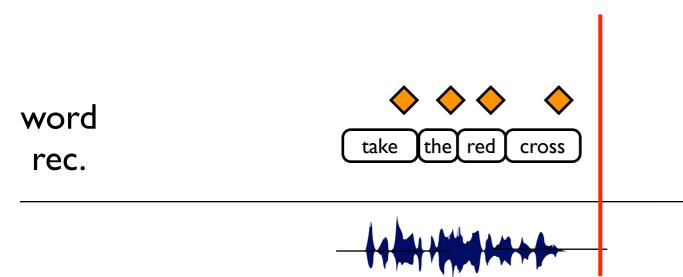




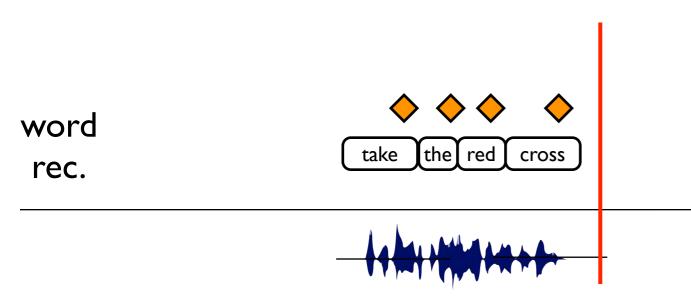




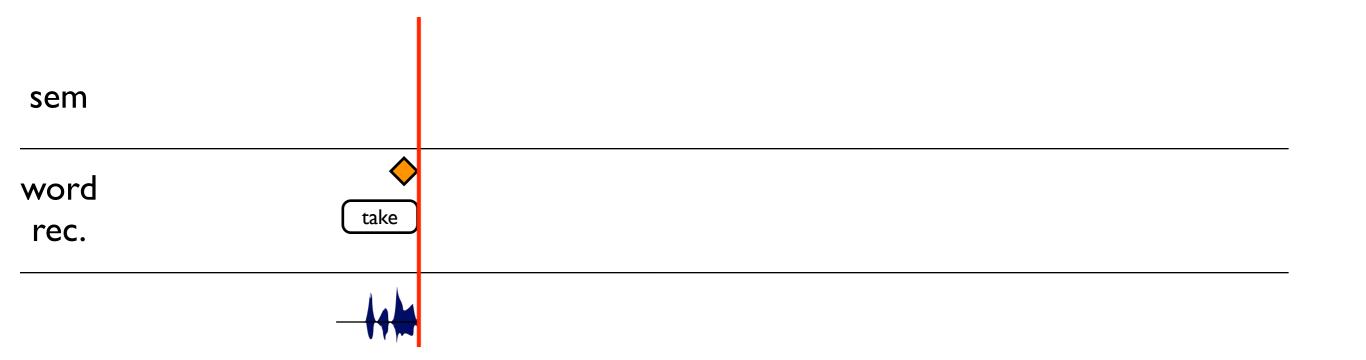




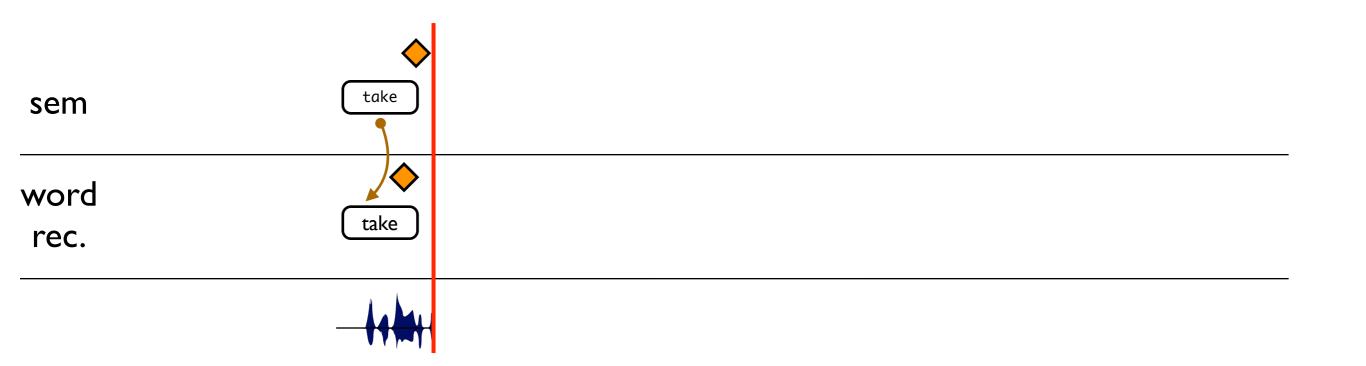
- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.



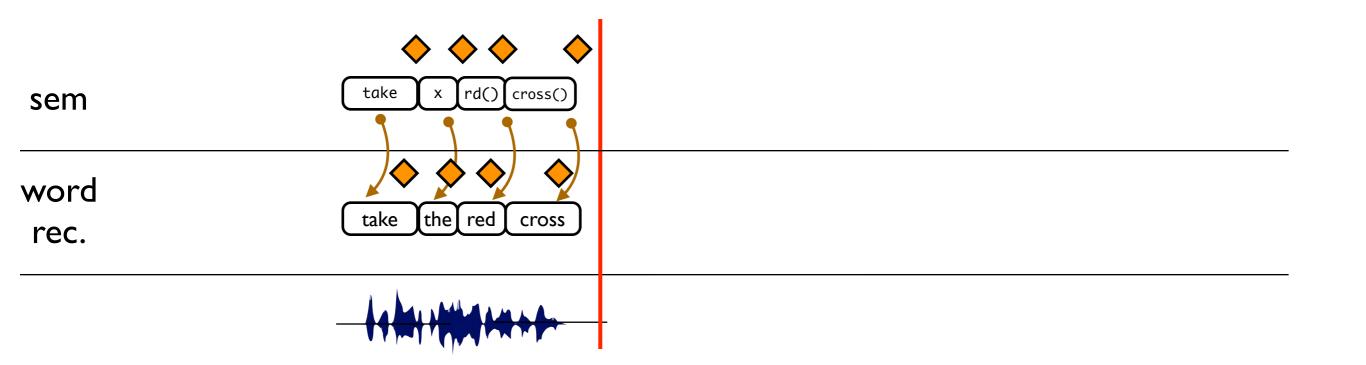
- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.



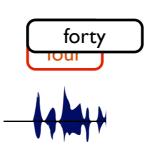
- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.



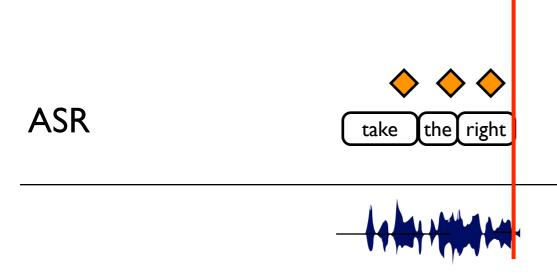
- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.



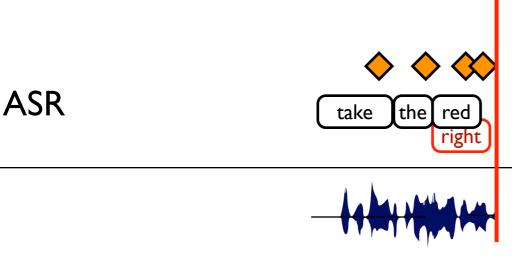
- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.
- IS may have to be revised, in light of newer information



- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.
- IS may have to be revised, in light of newer information



- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.
- IS may have to be revised, in light of newer information



- Information state is updated with *minimal units* of information, as soon as they can be hypothesised
- "Higher-level" hypotheses can be formed on the basis of "lower-level" ones.
- IS may have to be revised, in light of newer information
 Sem
 LFa b c
 ASR
 ASR

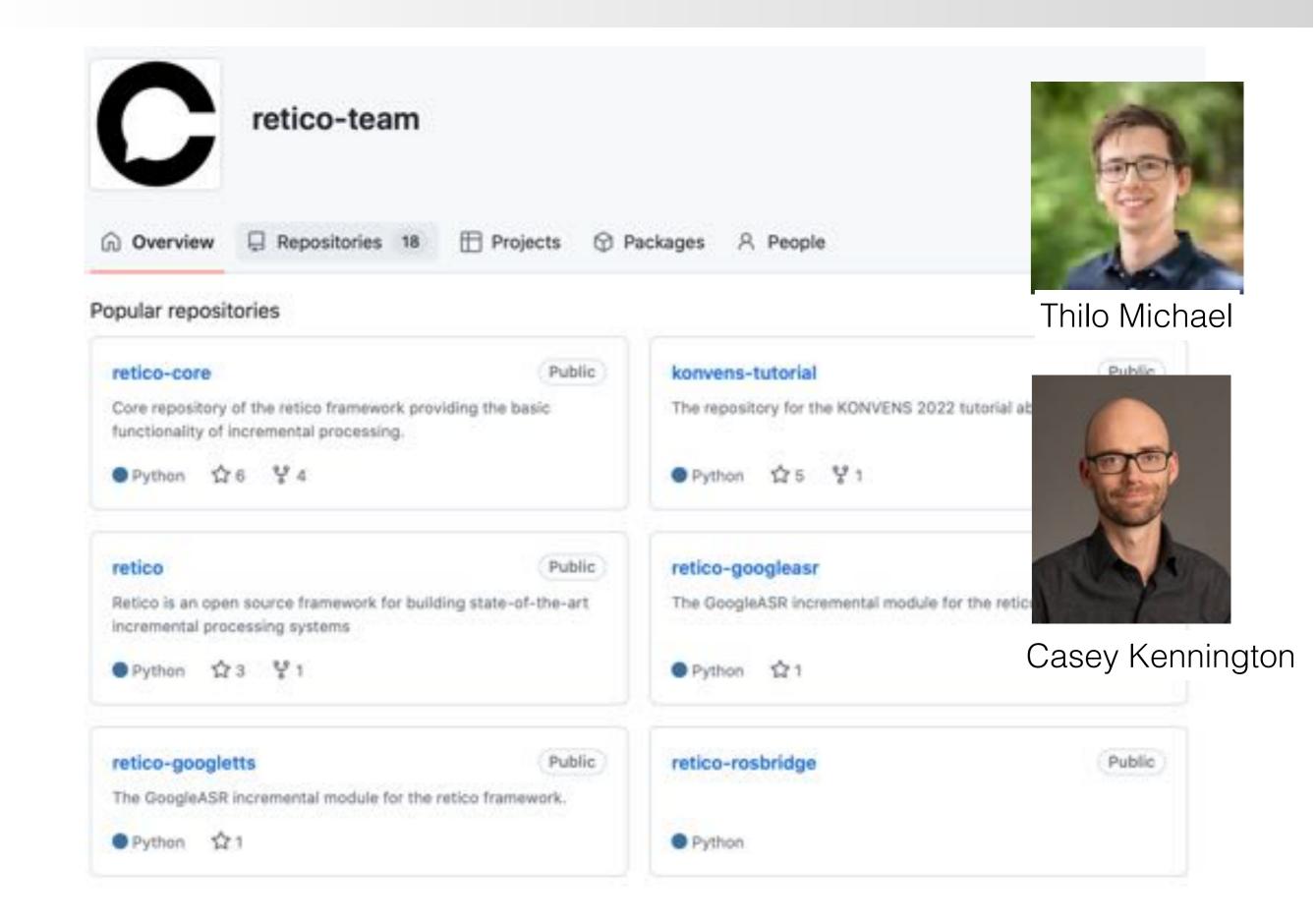
the IU model



• (Schlangen & Skantze 2009, 2011)

- Timo Baumann
- Implemented in InproTK (Timo Baumann, Casey Kennington, Spyros Kousidis, Bielefeld), Jindigo (Skantze, Stockholm), IPAACA (Buschmeier & Kopp, Bielefeld)
- new: RETICO (T. Michael, Berlin; Kennington, Boise)

Sem LFa b c ASR take the red rent



https://github.com/retico-team

colab potsdam Department Linguistics Universität Potsdam David Schlangen

Conclusions

- "language understanding" is more than an input/output mapping
- it is useful to think of the unobservable construct "language understanding" as constituting in the application (and enrichment) of various models through various processes...
- yielding an overall dynamic process of understanding / coordination
- the components of this process can be differently challenged through purposefully constructed *dialogue games*
 - as targets, or as test cases (future proofing against pre-trained models)
- models can be built in a modular fashion, or in a more end-to-end fashion
- the future is going to be wild

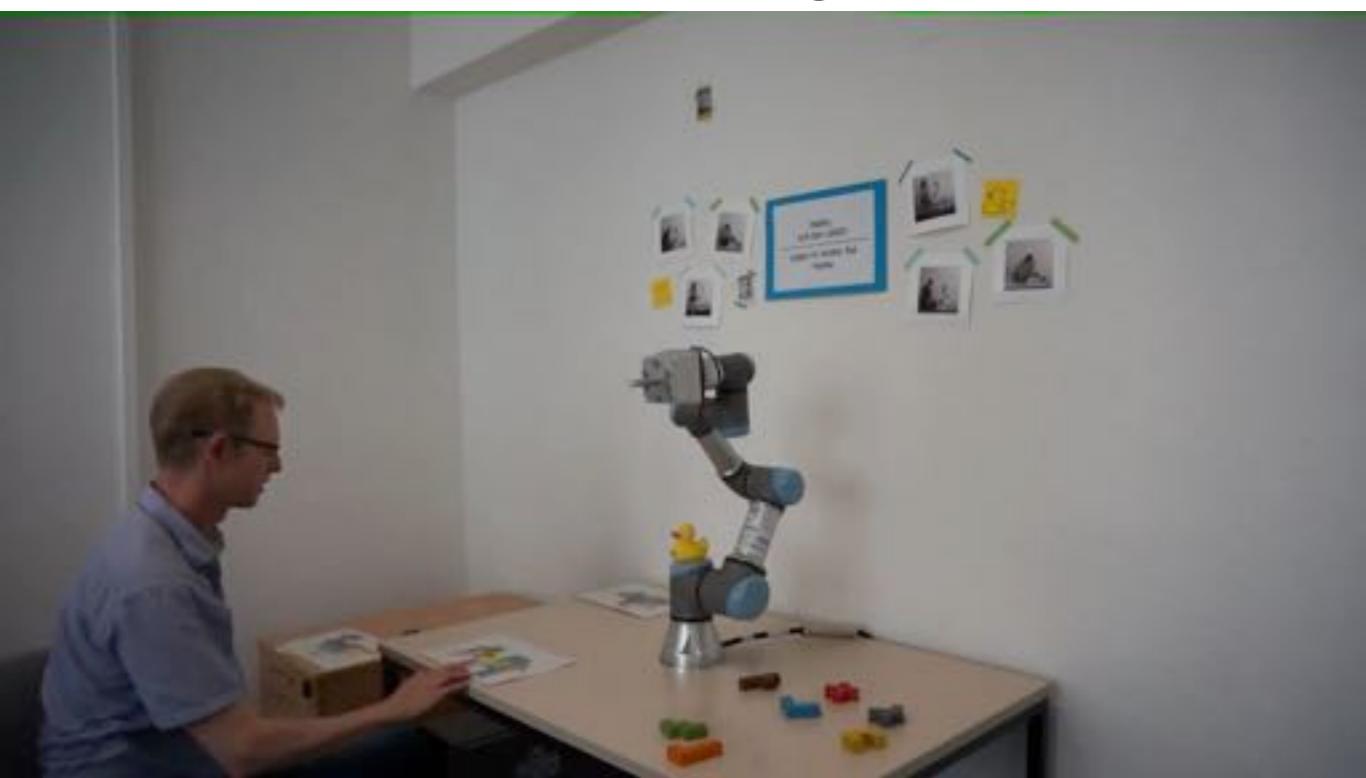
Conclusions

• a "reformist" approach to making "cognition as computation" more E? (embedded, embodied...)

Conclusions / Last Word

- An argument from ethics for focussing on task-oriented situated interaction for practical work:
 - An assertion is a promise. (To take responsibility for the truth of the proposition.)
 - In situated tasks, these promises are discharged quickly, and controllably. (Cp. ChatGPT making statements about a person.)
 - Fewer lies necessary (no "my favourite colour is blue" from the robot).

An embodied joint construction game...





Thank you. Questions, Comments?

Acknowledgements: Many thanks to my current & former grad students (<u>https://clp.ling.uni-potsdam.de/people/</u>) & colleagues w/ whom I have discussed related ideas in recent years.

Gratefully acknowledged: Funding by DFG (project "RECOLAGE"; CRC "Limits of Variability", project Bo6); BMBF (project "COCOBOTS")

THE END

List of References for the Talk "Understanding Understanding"

All of our publications can be found at: https://clp.ling.uni-potsdam.de/publications/.

References

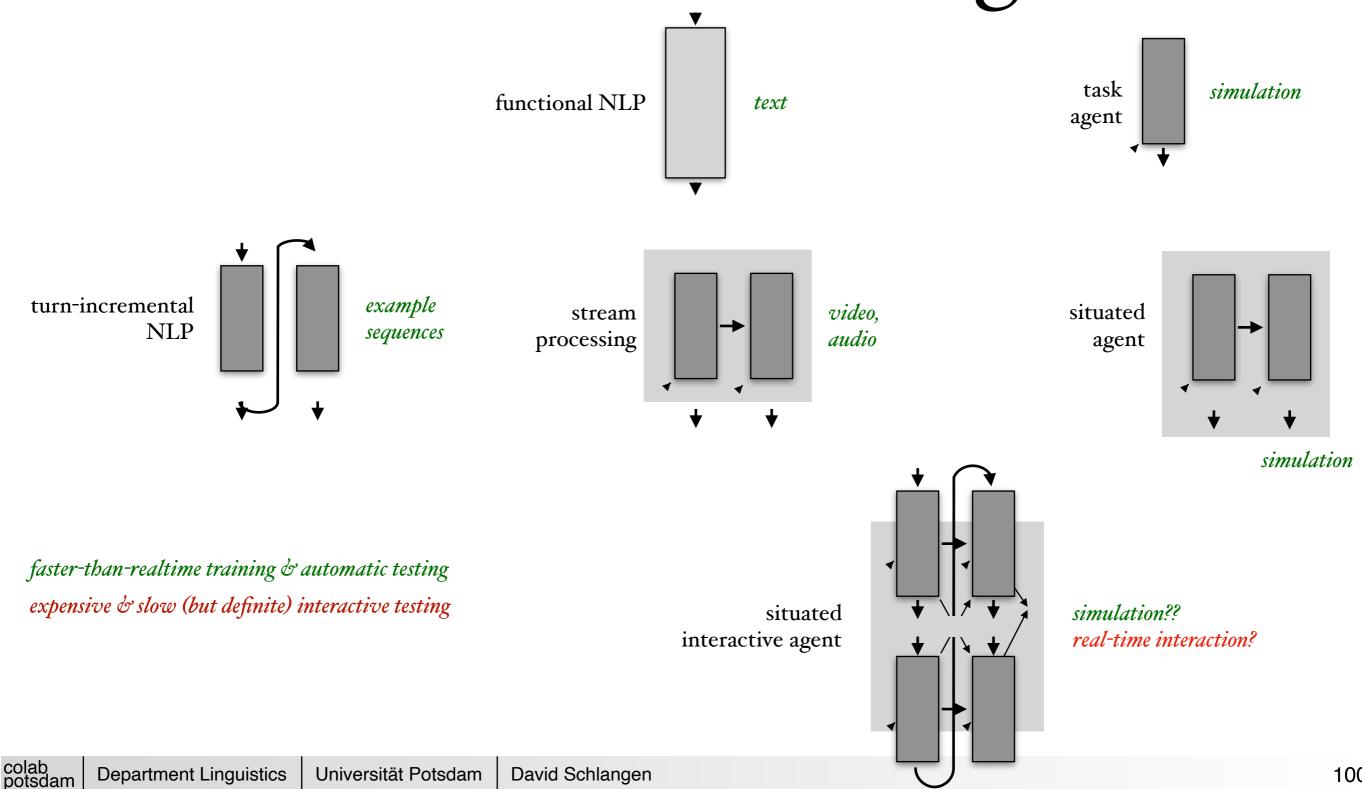
- Beyer, Anne, Sharid Loáiciga, and David Schlangen (2021). "Is Incoherence Surprising? Targeted Evaluation of Coherence Prediction from Language Models". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, pp. 4164–4173.
- Bowles, Samuel and Herbert Gintis (2011). A Cooperative Species: Human Reciprocity and its Evolution. Princeton University Press.
- Bratman, Michael E. (1987). *Intentions, Plans, And Practical Reason*. Cambridge, Massachusetts, USA: Harvard University Press.
- Chalamalasetti, Kranti, Jana Götze, Sherzod Hakimov, Brielen Madureira, Philipp Sadler, and David Schlangen (2023). "clembench: Using Game Play to Evaluate Chat-Optimized Language Models as Conversational Agents". In: *CoRR*.
- Chomsky, Noam (1957). Syntactic Structures. Mouton & Co.
- Christiansen, Morten H and Nick Chater (2016). "The Now-or-Never bottleneck: A fundamental constraint on language". In: *Behavioral and Brain Sciences* 39, e62.
- Clark, Eve (2003). First Language Acquisition. Cambridge, UK: Cambridge University Press.
- Clark, Herbert H. (1996). Using Language. Cambridge: Cambridge University Press.
- Cohen, Philip R., Jerry Morgan, and Martha E. Pollack, eds. (1990). *Intentions in Communication*. Cambridge, Mass.: MIT Press.
- Dijk, T.A. van and W. Kintsch (1983). *Strategies of Discourse Comprehension*. Monograph Series. Academic Press.
- Dunietz, Jesse, Greg Burnham, Akash Bharadwaj, Owen Rambow, Jennifer Chu-Carroll, and Dave Ferrucci (July 2020). "To Test Machine Comprehension, Start by Defining Comprehension". In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, pp. 7839–7859.
- Fernández, Raquel, Staffan Larsson, Robin Cooper, Jonathan Ginzburg, and David Schlangen (2011). "Reciprocal Learning via Dialogue Interaction: Challenges and Prospects". In: Proceedings of the IJCAI 2011 Workshop on Agents Learning Interactively from Human Teachers (ALIHT 2011). Barcelona, Spain.
- Fernández, Raquel, Tatjana Lucht, Kepa Rodríguez, and David Schlangen (2006). "Interaction in Task-Oriented Human–Human Dialogue: The Effects of Different Turn-Taking Policies". In: *Proceedings of the First International IEEE/ACL Workshop on Spoken Language Technology*. Palm Beach, Aruba.
- Galetzka, Fabian, Chukwuemeka Uchenna Eneh, and David Schlangen (May 2020). "A Corpus of Controlled Opinionated and Knowledgeable Movie Discussions for Training Neural Conversation Models". English. In: *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, pp. 565–573.
- Galetzka, Fabian, Jewgeni Rose, David Schlangen, and Jens Lehmann (Aug. 2021). "Space Efficient Context Encoding for Non-Task-Oriented Dialogue Generation with Graph Attention Transformer". In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online: Association for Computational Linguistics, pp. 7028–7041.
- Götze, Jana, Maike Paetzel-Prüsmann, Wencke Liermann, Tim Diekmann, and David Schlangen (June 2022). "The Slurk Interaction Server Framework: Better Data for Better Dialog Models". In: *Pro-*

ceedings of the Language Resources and Evaluation Conference. Marseille, France: European Language Resources Association, pp. 4069–4078.

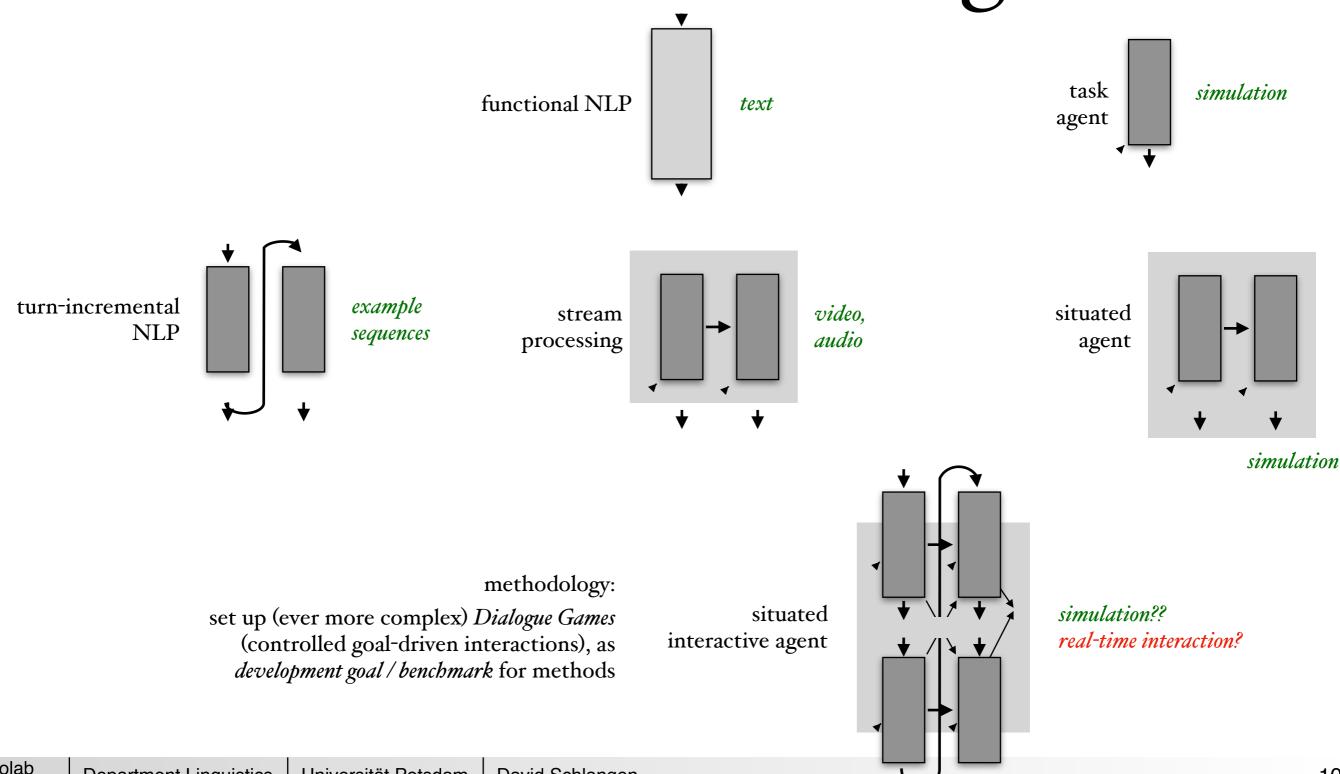
- Götze, Jana, Karla Friedrichs, and David Schlangen (2022). "Interactive and Cooperative Delivery of Referring Expressions: A Comparison of Three Algorithms". In: *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue Full Papers*. Virtually and at Dublin, Ireland: SEMDIAL.
- Harnad, Stevan (1990). "The Symbol Grounding Problem". In: Physica D 42, pp. 335–346.
- Harris, Paul L. (2015). *Trusting What You're Told: How Children Learn from Others*. Harvard, Mass., USA: Harvard University Press.
- Heim, Irene (1983). "File Change Semantics and the Familiarity Theory of Definiteness". In: *Meaning, Use and Interpretation of Language*. Ed. by R. Bäuerle, Ch. Schwarze, and Arnim von Stechow. Berlin, Germany: De Gruyter, pp. 164–189.
- Holler, Judith and Stephen C. Levinson (2019). "Multimodal Language Processing in Human Communication". In: *Trends in Cognitive Sciences*, pp. 1–14.
- Hoppit, William and Kevin N. Laland (2013). Social Learning: An Introduction to Mechanisms, Methods, and Models. Princeton University Press.
- Johnson-Laird, Philip Nicholas (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cognitive science series. Harvard University Press.
- Kahardipraja, Patrick, Brielen Madureira, and David Schlangen (Nov. 2021). "Towards Incremental Transformers: An Empirical Analysis of Transformer Models for Incremental NLU". In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 1178–1189.
- Kamp, Hans (1981). "A Theorie of Truth and Representation". In: *Formal Methods in the Study of Language*. Ed. by J.A.G. Groenendijk, T.M.V. Janssen, and M.B.J. Stokhof. Mathematical Centre Tracts 135. Amsterdam: University of Amsterdam, pp. 277–322.
- Kendon, Adam (2004). Gestures. Cambridge, UK: Cambridge University Press.
- Kontogiorgos, Dimosthenis, Elena Sibirtseva, Andre Pereira, Gabriel Skantze, and Joakim Gustafson (2018). "Multimodal Reference Resolution In Collaborative Assembly Tasks". In: *Proceedings of the 4th International Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction*.
- Levesque, Hector J., Ernest Davis, and Leora Morgenstern (2012). "The Winograd schema challenge". In: Proceedings of the International Workshop on Temporal Representation and Reasoning, pp. 552– 561.
- Levinson, Stephen C. (1979). "Activity types and language". In: Linguistics 17, pp. 365–399.
- Levinson, Stephen C (2010). "Interactional Foundations of Language: The Interaction Engine Hypothesis". In: *Human language: From genes and brain to behavior*. Ed. by Peter Hagoort. Cambridge, MA, USA: MIT Press. Chap. 14, pp. 189–200.
- Loáiciga, Sharid, Anne Beyer, and David Schlangen (Oct. 2022). "New or Old? Exploring How Pre-Trained Language Models Represent Discourse Entities". In: *Proceedings of the 29th International Conference on Computational Linguistics*. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, pp. 875–886.
- Madureira, Brielen and David Schlangen (Nov. 2020). "Incremental Processing in the Age of Non-Incremental Encoders: An Empirical Assessment of Bidirectional Models for Incremental NLU".
 In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, pp. 357–374.
- (May 2022). "Can Visual Dialogue Models Do Scorekeeping? Exploring How Dialogue Representations Incrementally Encode Shared Knowledge". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Dublin, Ireland: Association for Computational Linguistics, pp. 651–664.

- Margolis, Eric and Stephen Laurence, eds. (2015). *The Conceptual Mind: New Directions in the Study of Concepts*. Cambridge, Massachusetts, USA: MIT Press.
- McNeill, David (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL, USA: University of Chicago Press.
- Murphy, Gregory L. (2002). The Big Book of Concepts. Cambridge, MA, USA: MIT Press.
- Ribeiro, Marco Tulio, Tongshuang Wu, Carlos Guestrin, and Sameer Singh (July 2020). "Beyond Accuracy: Behavioral Testing of NLP Models with CheckList". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, pp. 4902–4912.
- Schlangen, David (2019a). "Grounded Agreement Games: Emphasizing Conversational Grounding in Visual Dialogue Settings". In: *CoRR* abs/1908.11279. arXiv: 1908.11279.
- (2019b). "Language Tasks and Language Games: On Methodology in Current Natural Language Processing Research". In: *CoRR* abs/1908.10747. arXiv: 1908.10747.
- (Aug. 2021). "Targeting the Benchmark: On Methodology in Current Natural Language Processing Research". In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Online: Association for Computational Linguistics, pp. 670–674.
- (Sept. 2022). "Norm Participation Grounds Language". In: Proceedings of the 2022 CLASP Conference on (Dis)embodiment. Gothenburg, Sweden: Association for Computational Linguistics, pp. 62–69.
- (2023a). "Dialogue Games for Benchmarking Language Understanding: Motivation, Taxonomy, Strategy". In: *CoRR* abs/2304.07007. arXiv: 2304.07007.
- (2023b). "What A Situated Language-Using Agent Must be Able to Do: A Top-Down Analysis". In: CoRR abs/2302.08590. arXiv: 2302.08590.
- Sellars, Wilfried (1954). "Some Reflections on Language Games". In: *Philosophy of Science* 21, pp. 204–228.
- Wang, Alex, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman (2019). "SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems". In: *NeurIPS*. July, pp. 1–30. arXiv: 1905.00537.
- Wittgenstein, Ludwig (1953). *Tractatus Logicus Philosophicus und Philosophische Untersuchungen*. Vol. 1. Werkausgabe. this edition 1984. Frankfurt am Main: Suhrkamp.
- Zarrieß, Sina, Julian Hough, Casey Kennington, Rames Manuvinakurike, David DeVault, Raquel Fernández, and David Schlangen (2016). "PentoRef: A Corpus of Spoken References in Task-Oriented Dialogues". In: *Proceedings of LREC 2016.* Portoroz, Slovenia.

from NLP-NLU to situated interactive agents



from NLP-NLU to situated interactive agents



Why Dialogue Games?

goal-*directed* activity (provides purpose to language)

clear definition of what counts as legal move (formal constraints)

A *Dialogue Game* is a constructed activity with a clear beginning and end, in which *players* attempt to reach a predefined *goal state* primarily by means of producing and understanding linguistic material.

clear metric for whether / how well goal has been reached

goal & rules provide control over type of context that is relevant control over type of knowledge that is relevant

nice technical property: game instances unlikely to be even in internet-scale

training data; easy to generate more