

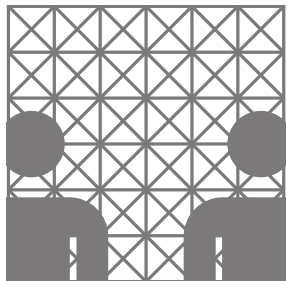
# Specialization Module

# Speech Technology

Timo Baumann  
[baumann@informatik.uni-hamburg.de](mailto:baumann@informatik.uni-hamburg.de)

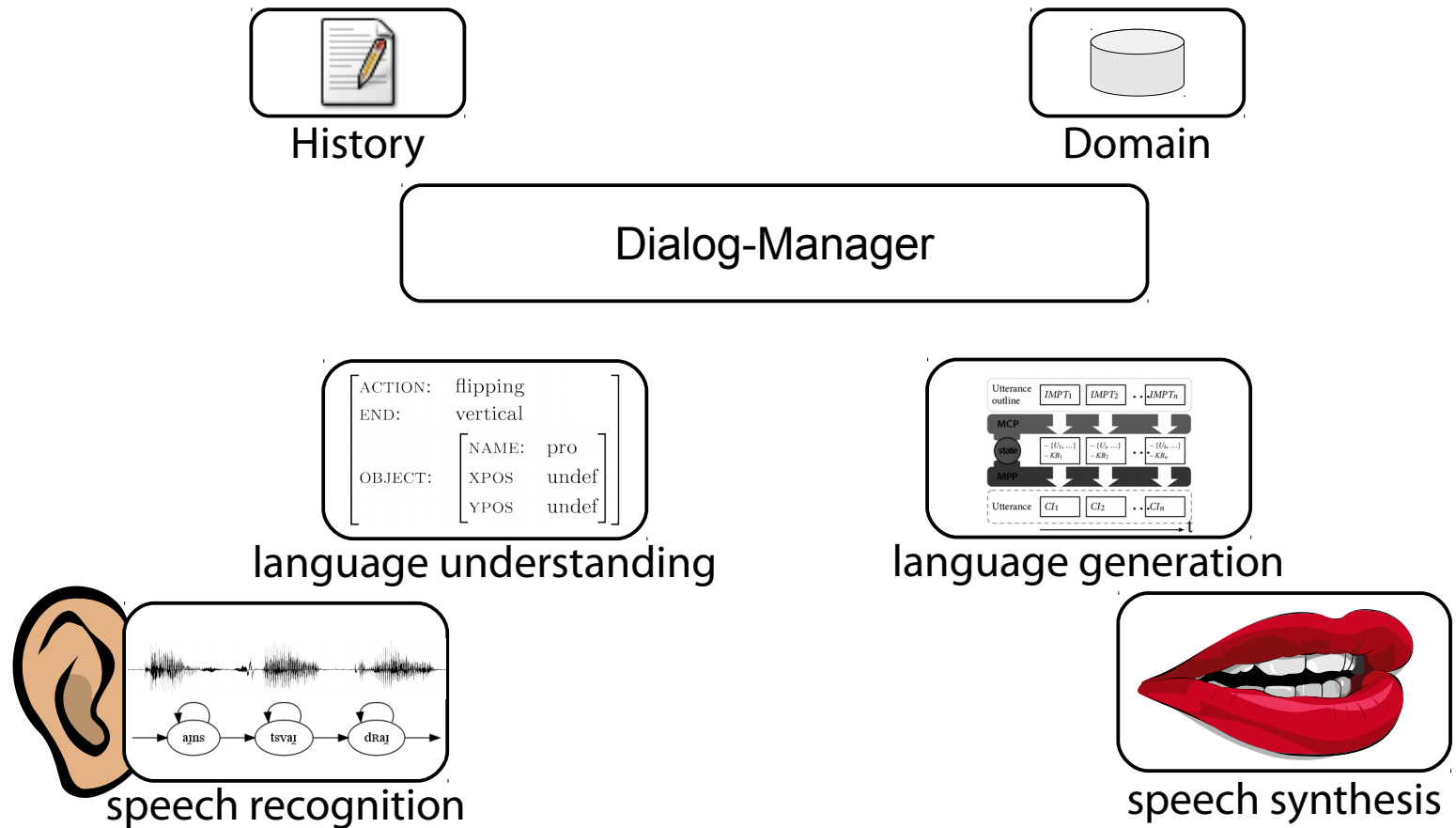


Universität Hamburg, Department of Informatics  
Natural Language Systems Group



# Spoken Dialogue, a Complex Interactive System

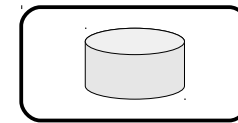
# End-to-end speech pipeline



# End-to-end speech pipeline



History

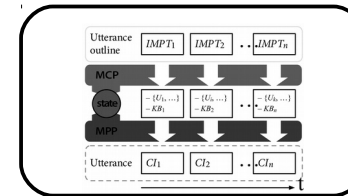


Domain

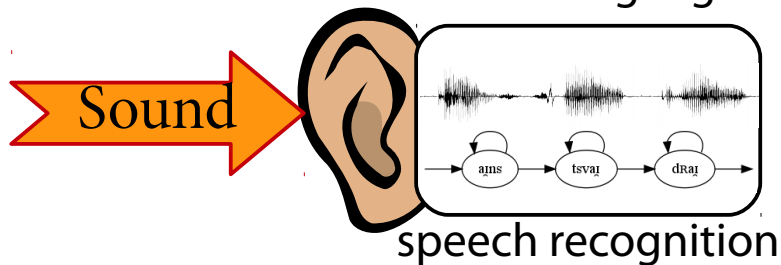
Dialog-Manager

```
[ ACTION:  flipping  
  END:    vertical  
  OBJECT: [ NAME:  pro  
            XPOS: undef  
            YPOS: undef ] ]
```

language understanding

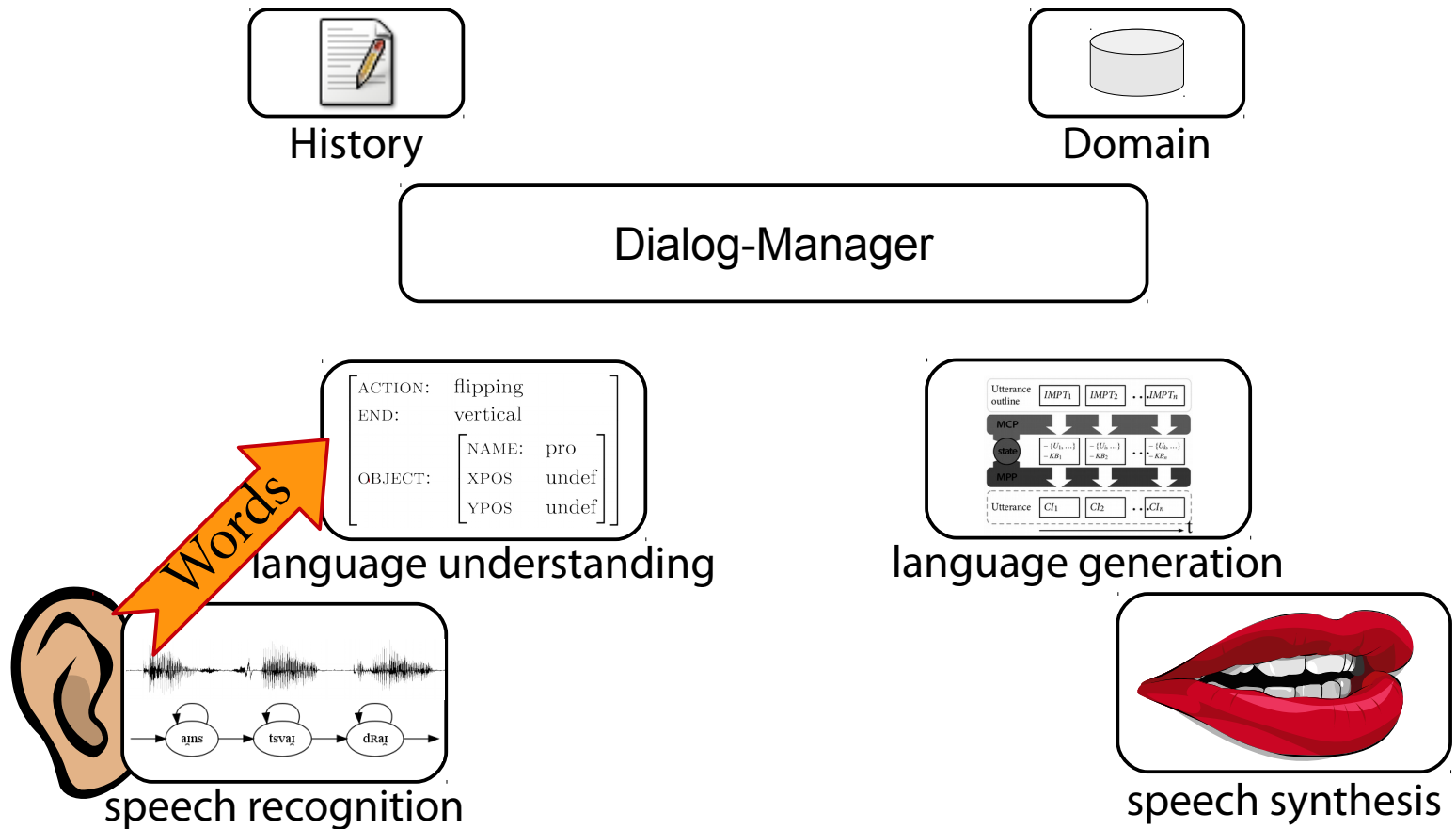


language generation

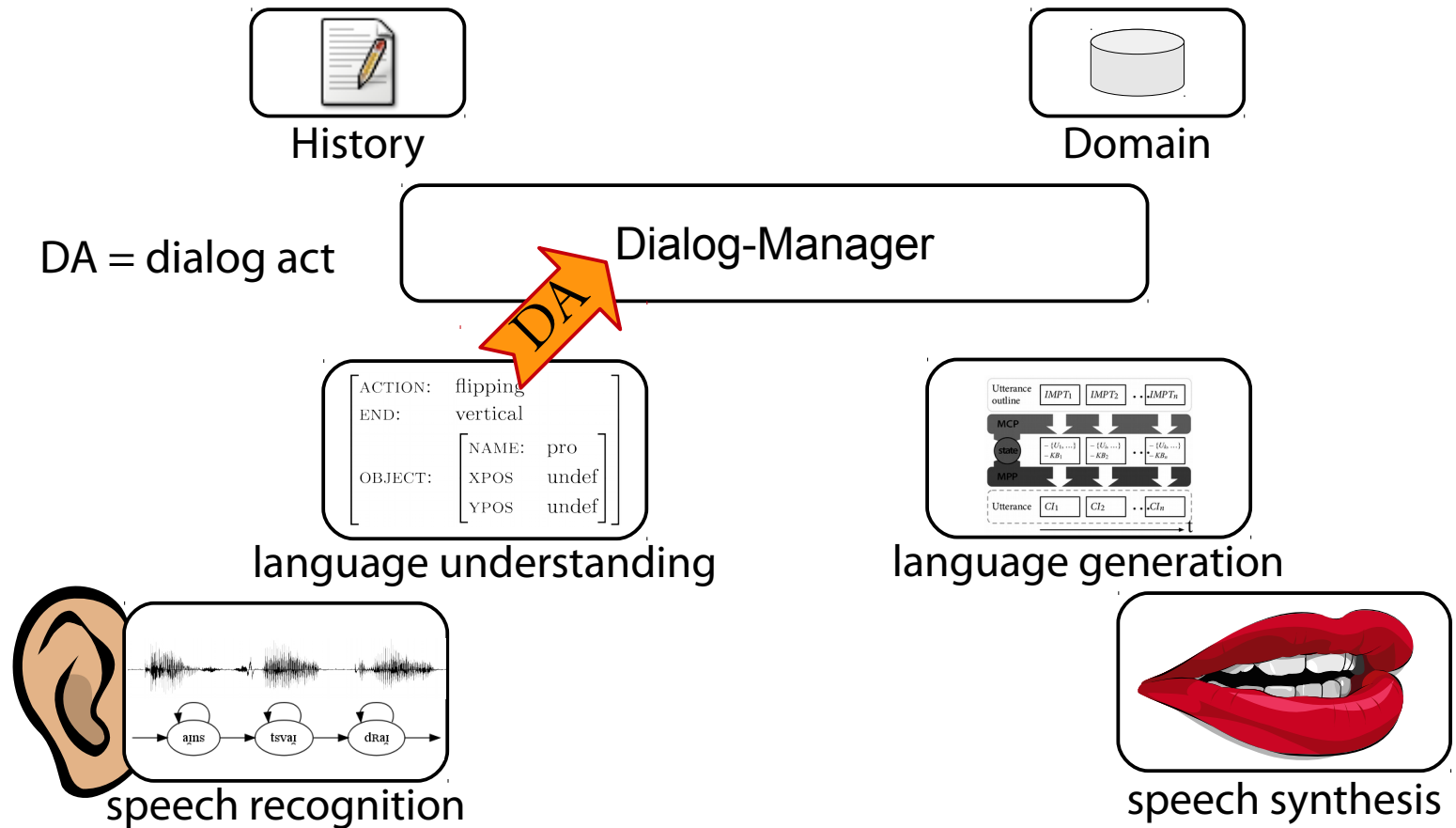


speech synthesis

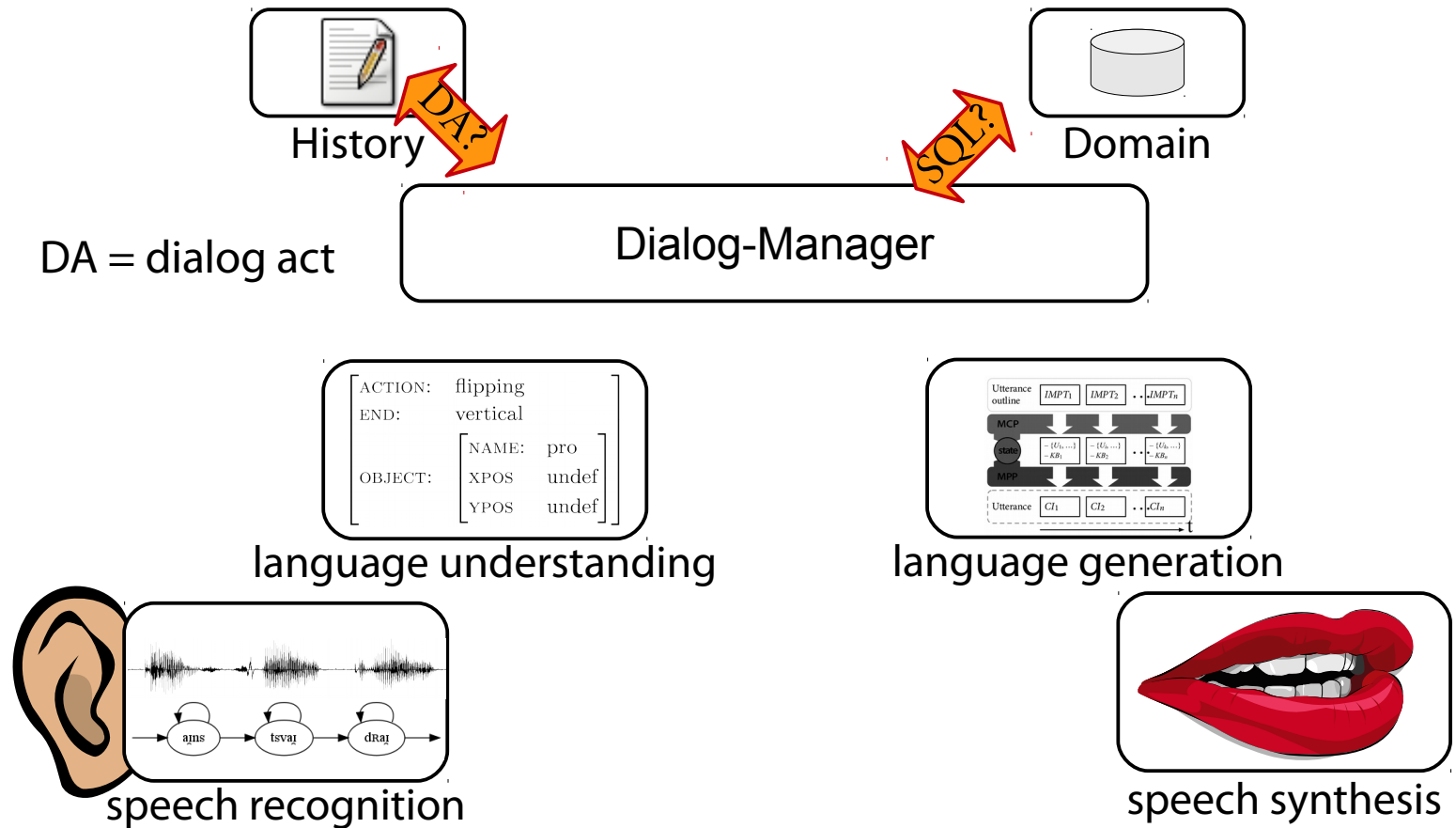
# End-to-end speech pipeline



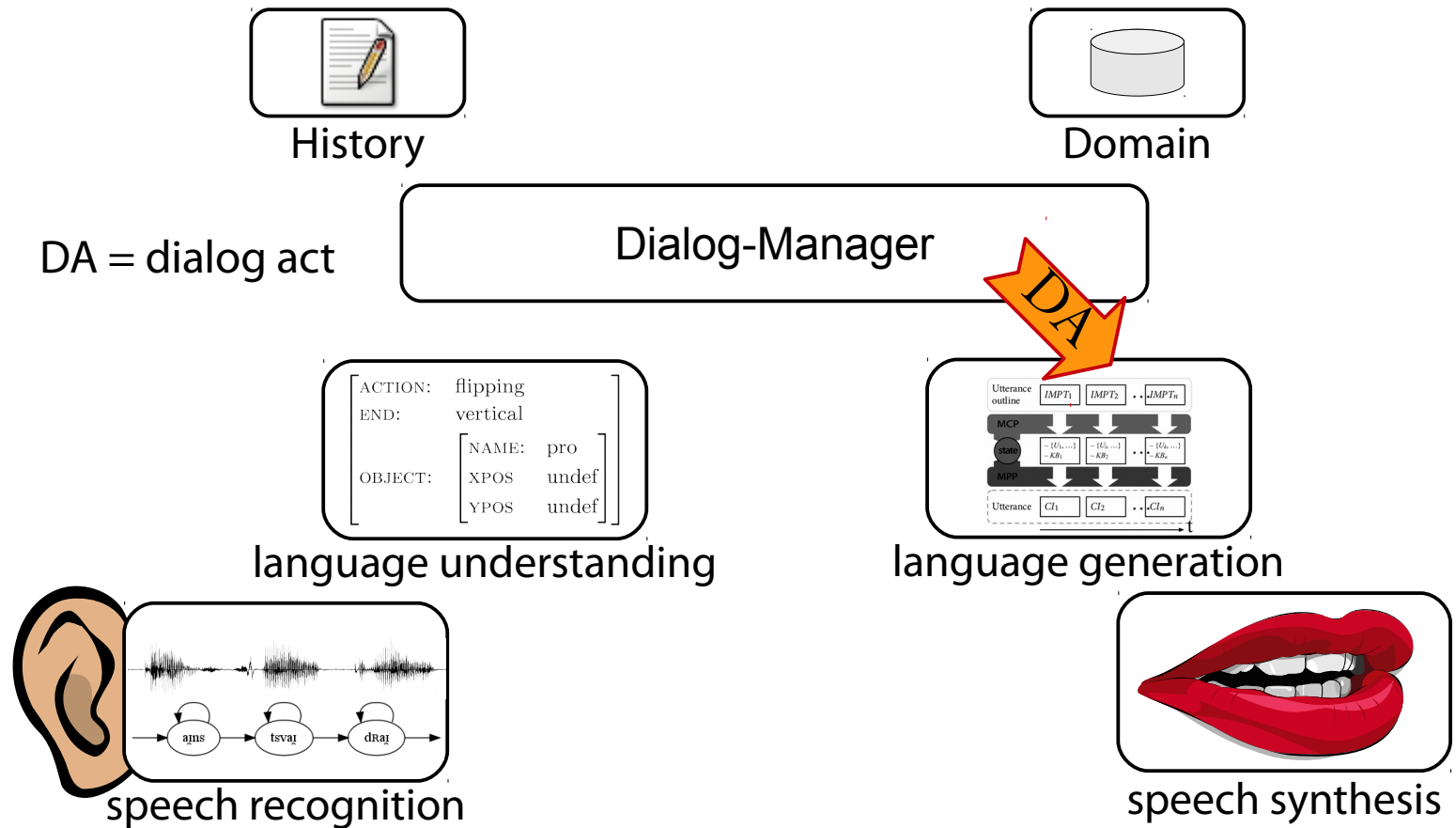
# End-to-end speech pipeline



# End-to-end speech pipeline

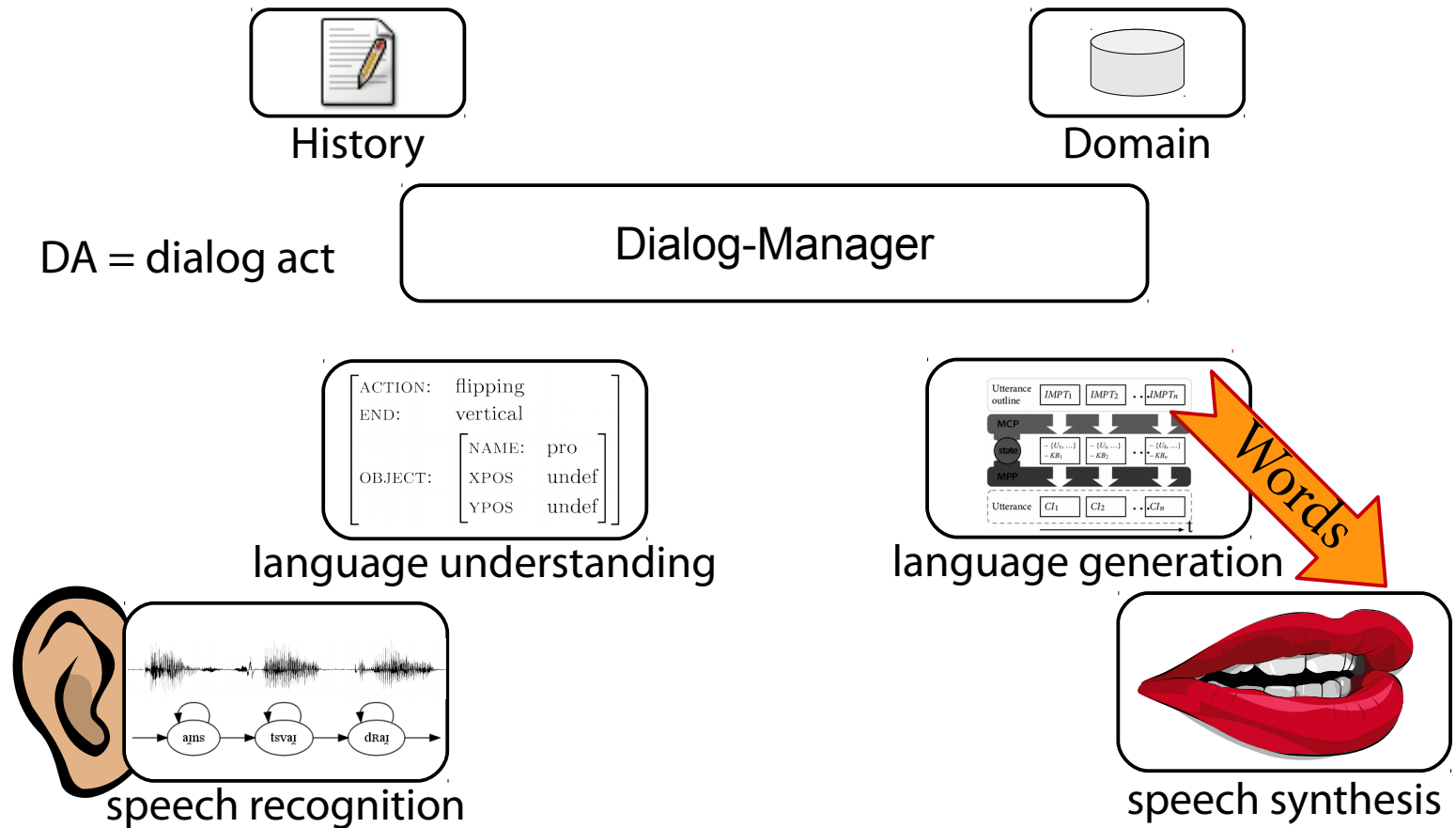


# End-to-end speech pipeline

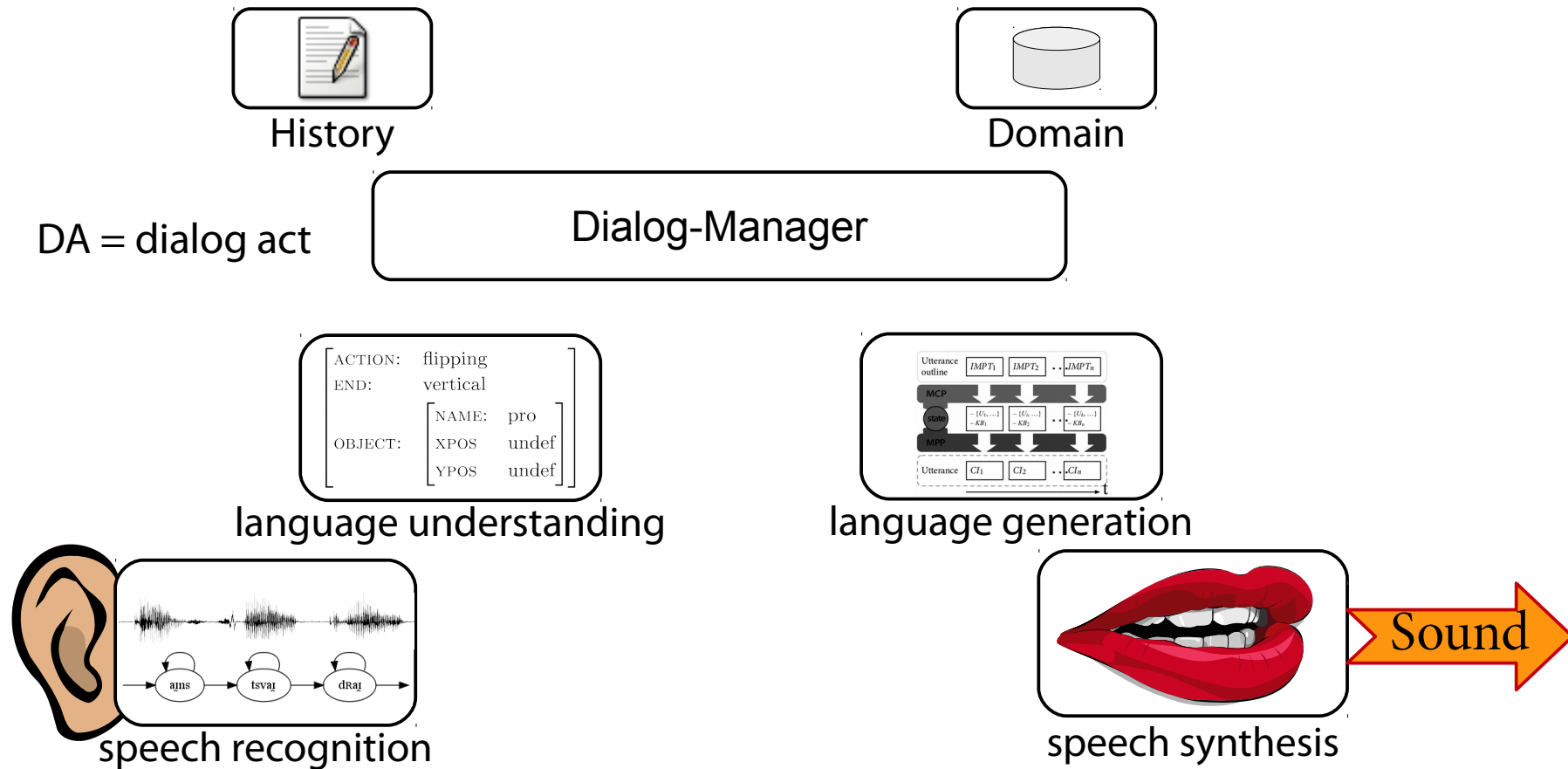




# End-to-end speech pipeline



# End-to-end speech pipeline





Group work with people \*NOT\* from your seminar day:

- put names&topics into appropriate positions of the figure
  - discuss how your topics relate to far distant topics
- have a relation of your (own) topic to \*every single other\* topic discussed in the seminar and the lecture

# Issues in Spoken Human-computer Interaction?

In small groups:

Collect/sort issues you see with dialog/conversational systems.

# Issues in Spoken Human-computer Interaction?

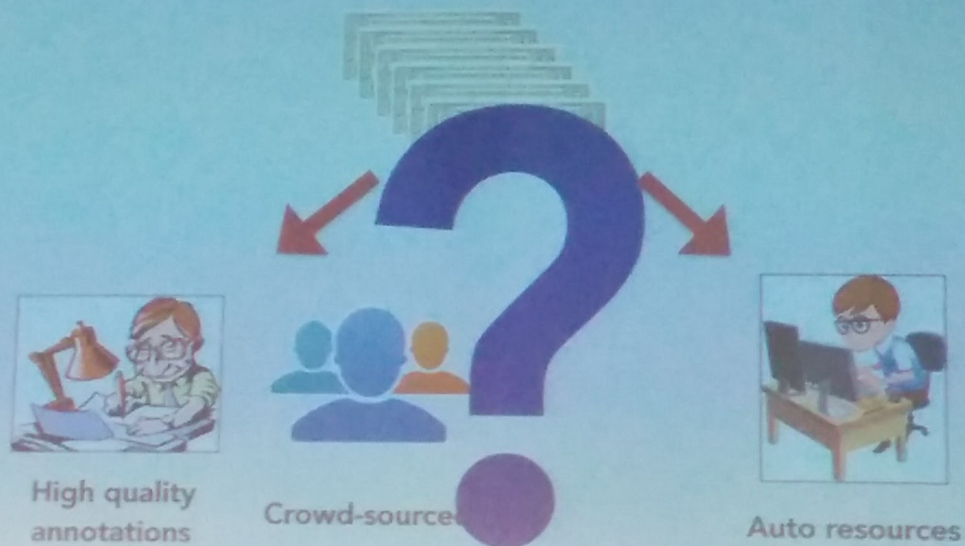
- Speech recognition and understanding (but that was 2005)
  - esp. during error recovery; high user surprisal/cognitive load
  - results in longer system prompts to guide the user more
- Time-outs → both too long and short
- Responsiveness → no „swift exchanges“ as with humans
- Synthesis → slow so that it be understandable
- Feedback → not task but interaction-related
- Adaptation → to the user's style
- other: prosody, non-lexical sounds, prompt generation, ...

What radical changes do you think must happen  
for these issues to be overcome?  
Architecture? Design? Learning? Data?



# Where to spend effort?

## THE RESOURCE TRADE-OFF



J. ODIJK

# Where to spend effort?

## TOP-LEVEL CONCLUSION



J. ODLIK

# Where to spend effort?

## MACHINE TRANSLATION

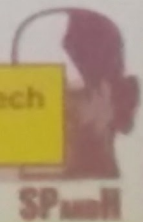
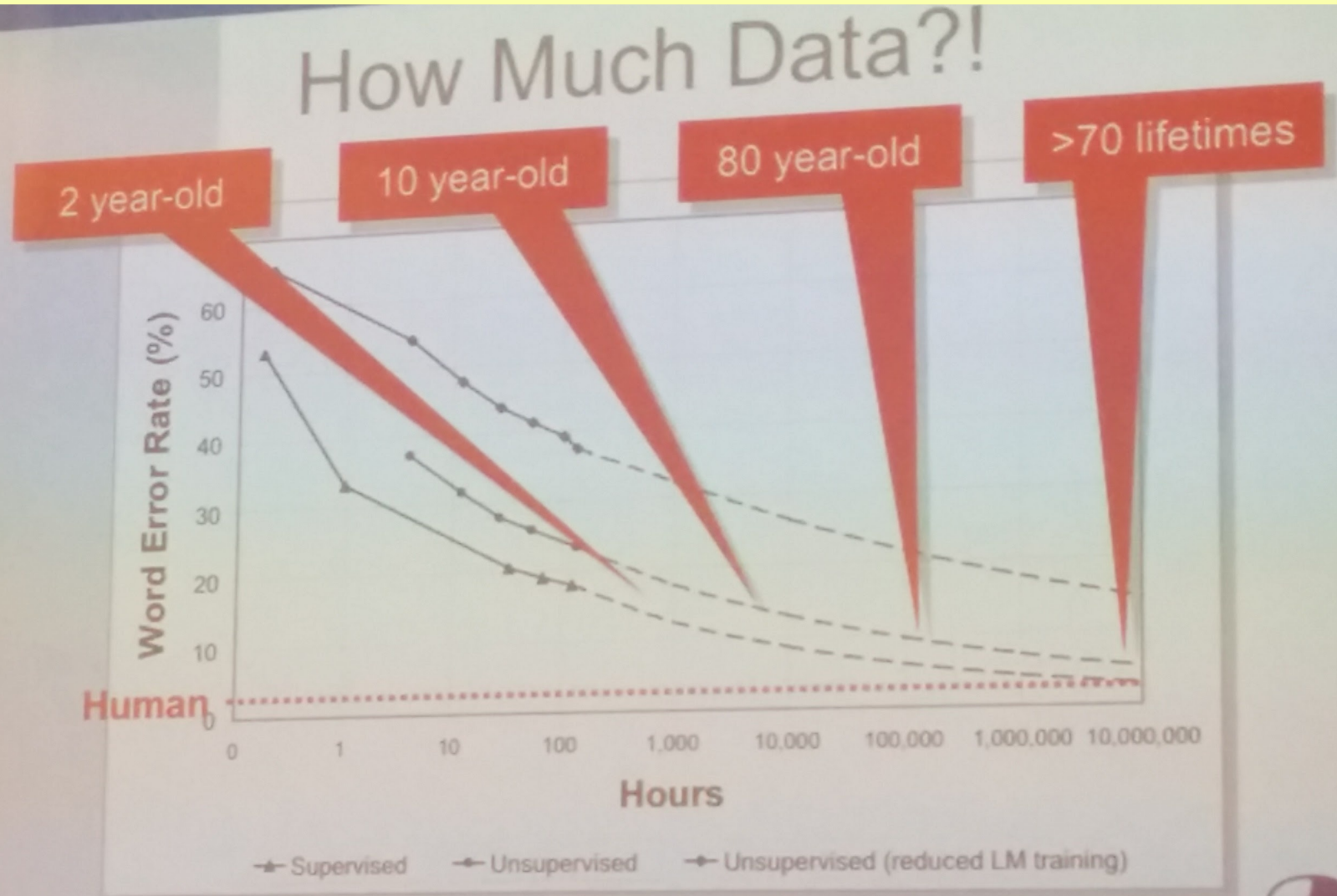


vs.



- Human vs. auto data: about the same
- Human models sometimes better than learned
- Better parsing models = better translation
- Better to **spend on targeted resources** — reordering

# More Data?



# Conclusions

- the problems of today's systems are hard (or impossible) to overcome in pipeline-based systems
- one big learnt system will not be the solution, either (lack of training data, ...)
- we need smarter models
- we need a much smarter interaction between models
- we need to find the bottlenecks and work on those
- you should become excellent software/solution architects, debuggers/problem analysts, and computer/data scientists

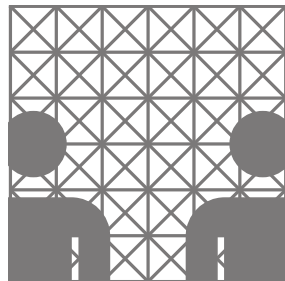
Thank you.

[baumann@informatik.uni-hamburg.de](mailto:baumann@informatik.uni-hamburg.de)

<https://nats-www.informatik.uni-hamburg.de/SLP16>



Universität Hamburg, Department of Informatics  
Natural Language Systems Group



# Further Reading

- Issues in Dialog Systems:
  - Ward, Nigel G., Anais G. Rivera, Karen Ward and David G. Novick (2005). "Root causes of lost time and user stress in a simple dialog system". In: INTERSPEECH 2005, Lisbon, Portugal, pp. 1565-1568.
- Current Trends:
  - Williams, Jason D. (2009). "Spoken Dialogue Systems: Challenges, and Opportunities for Research". In: Proc IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2009.
  - Ward, Nigel G., and David DeVault (2015). "Ten challenges in highly-interactive dialog systems." AAAI Spring Symposium on Turn-taking and Coordination in Human-Machine Interaction.
  - Feng, J., Ramabhadran, B., Hansen, J. H., & Williams, J. D. (2012). "Trends in speech and language processing". IEEE Signal Processing Magazine, 29(1), 184.

# Notizen



# Desired Learning Outcomes

- students have an overview of speech and language technology used in interactive (dialog) systems
- students know the topics that have been discussed in the seminar and are able to relate their topic to the other lecture and seminar topics
- students are aware of current limitations of speech technology and have ideas about the leverage of methods to overcome such limitations