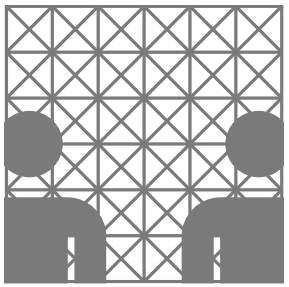
Specialization Module

Speech Technology

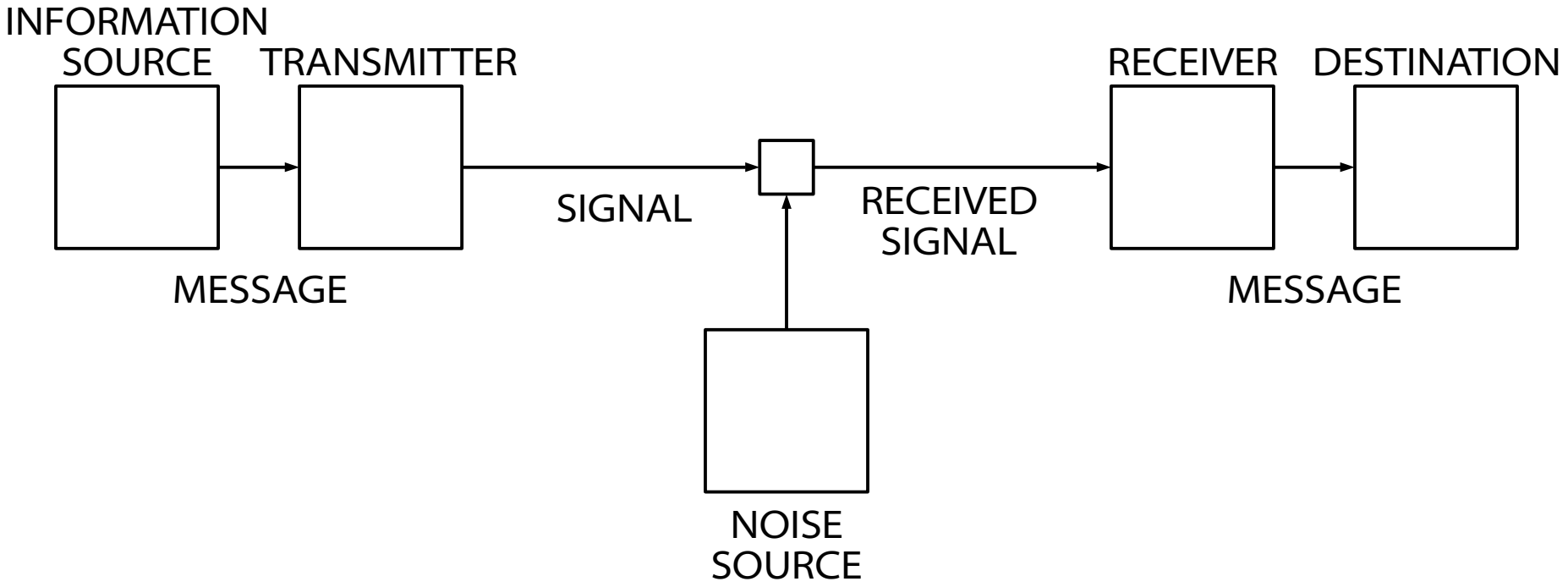
Timo Baumann
baumann@informatik.uni-hamburg.de

UNIVERSITÄT HAMBURG, DEPARTMENT OF INFORMATICS
NATURAL LANGUAGE SYSTEMS GROUP

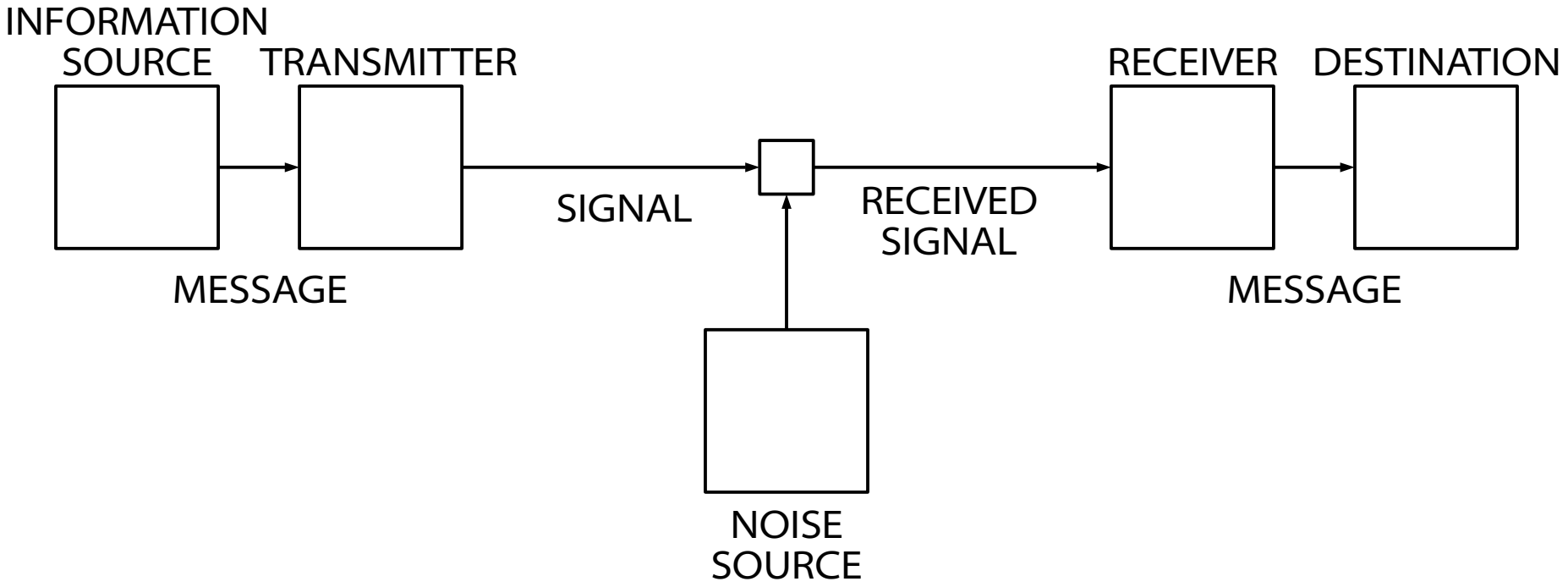


Spoken Dialogue, a Complex Interactive System

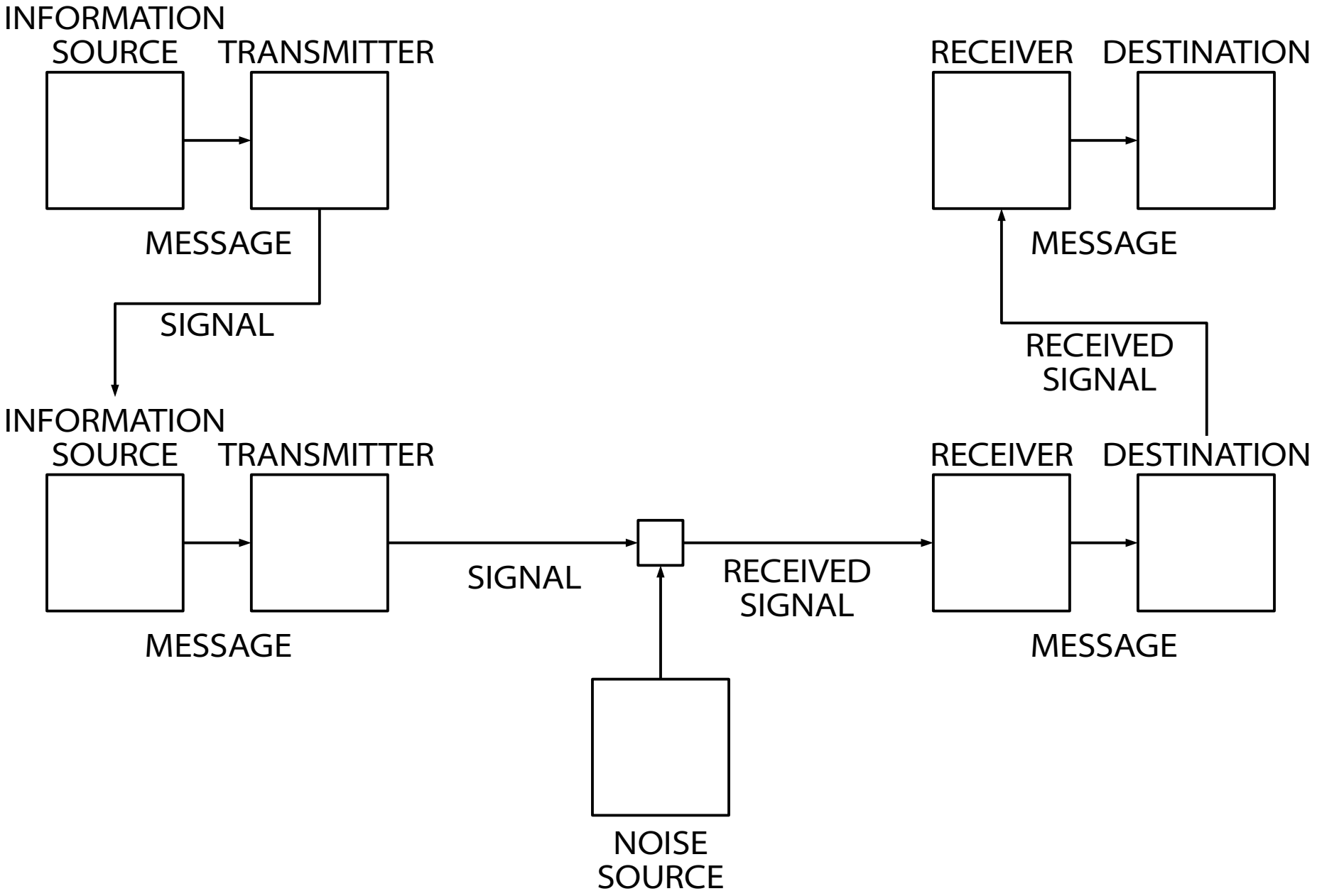
The Noisy-channel Model



The Chain Model of Communication

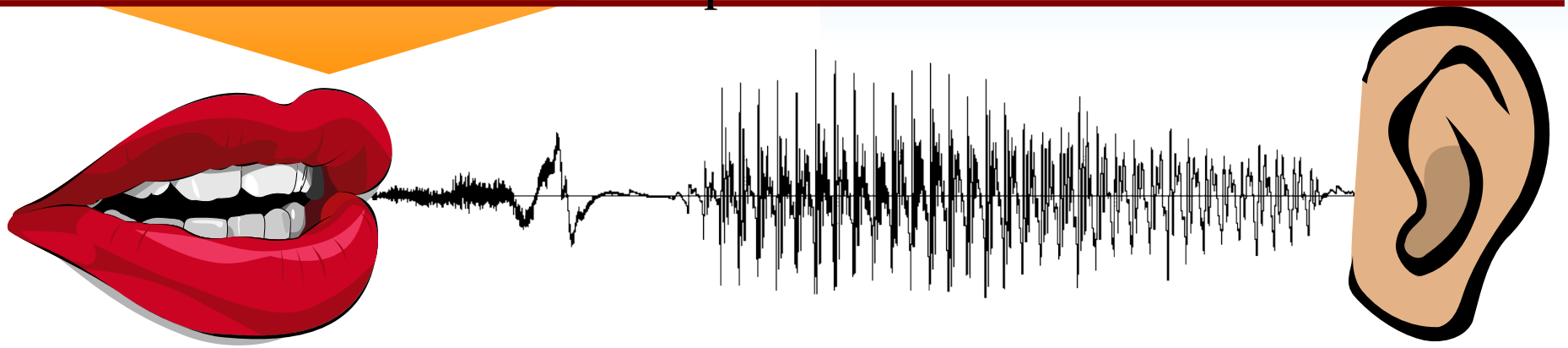


The Chain Model of Communication



Chain model of Communication

find message that describes idea	pragmatics	recover idea described by message
determine structure to convey meaning	semantics/ lexicology	determine meaning of structure
sequentialize structure to word stream	syntax/ morphology	recover structure of sequence
represent words through sounds	phonology/ phonetics	recombine souds to words



Human Communication (simplified)

Speaker

Listener

Concept

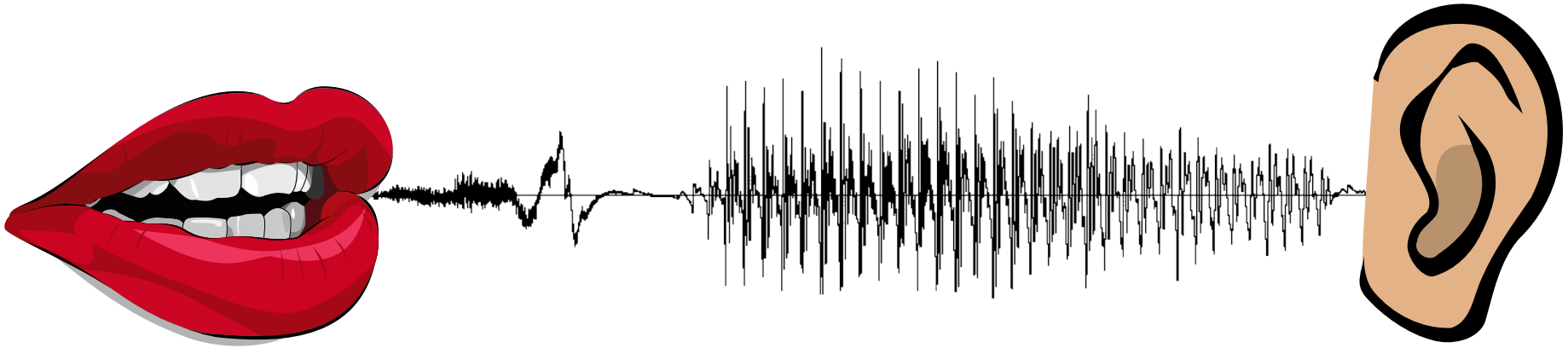
Verbalization

muscular movements

Received Concept

Interpretation

sensory impression



Human Communication (simplified)

Speaker

Listener

Concept

Verbalization

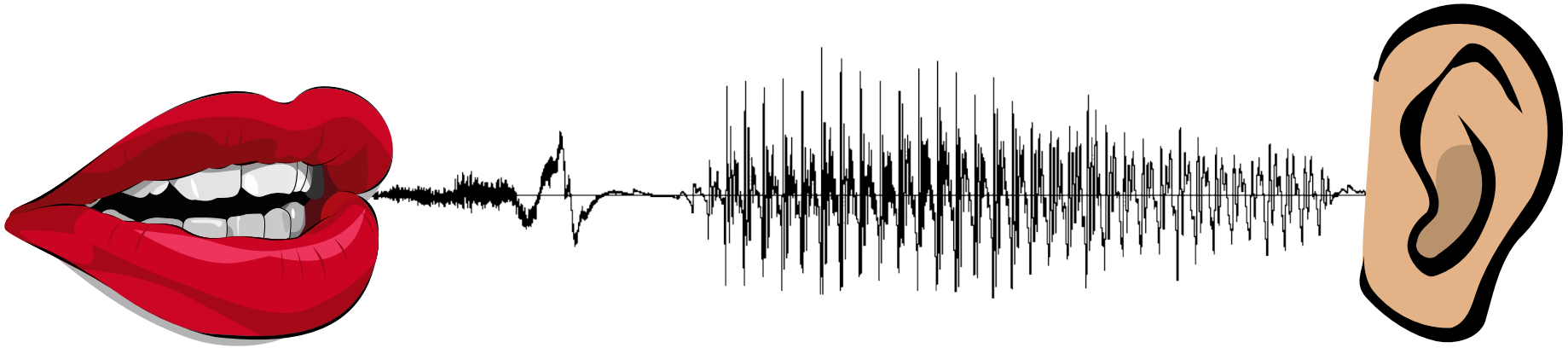
muscular movements

correspondence (hopefully!)

Received Concept

Interpretation

sensory impression



But what about dialogue?

In what ways does the simple model for a dialog agent seem insufficient to you? (in pairs / small groups; 5 minutes)

Aspects of dialogue

- bi-directional communication
  - no clear „sender“ and „receiver“; agents are both
- agents share the communication channel
  - time-sharing
  - additional feedback signals
  - simultaneous speech is more frequent than we think!
- communication is controlled interactively by **both** the current-speaker and the current-listener

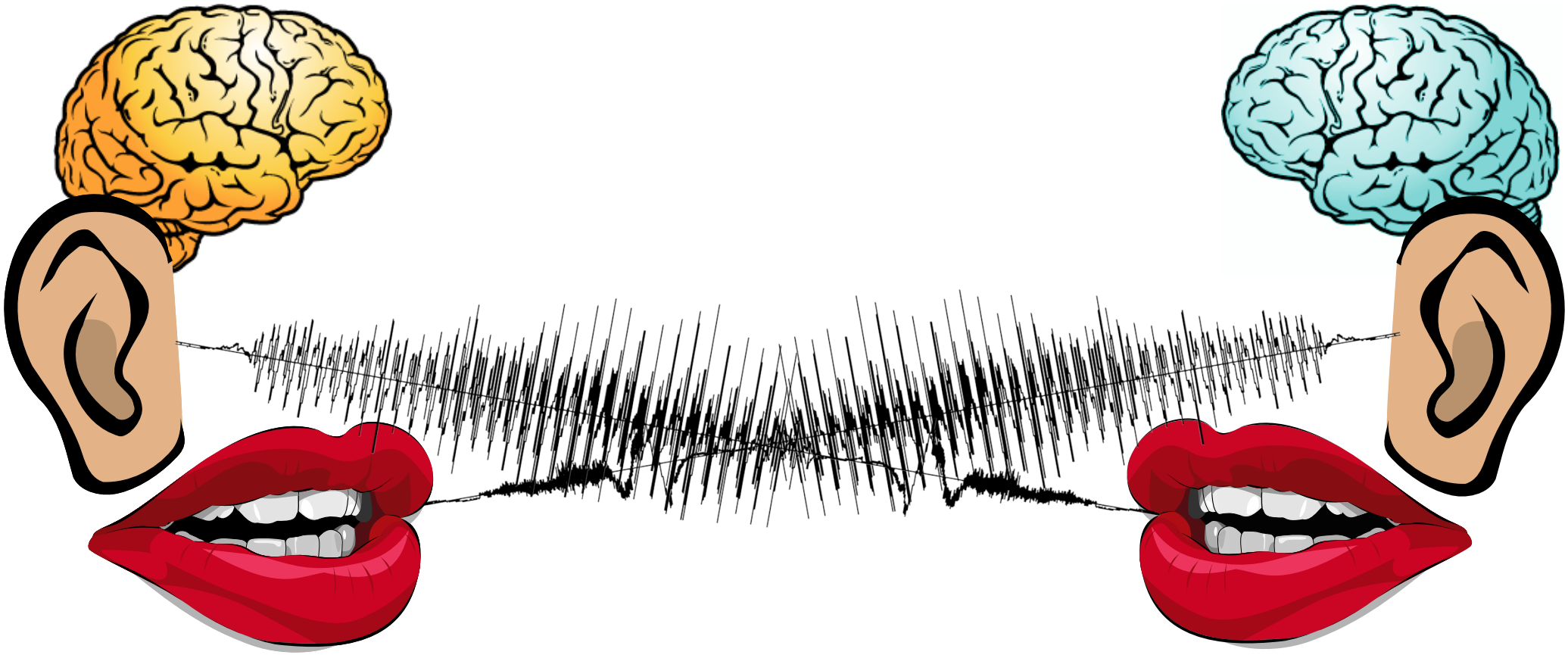
➔ local management within each layer (e.g. entrainment)

➔ turn-taking!

Dialogue (simplified)

dialogue agent

dialogue agent

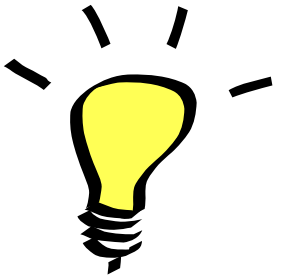


Turn-taking

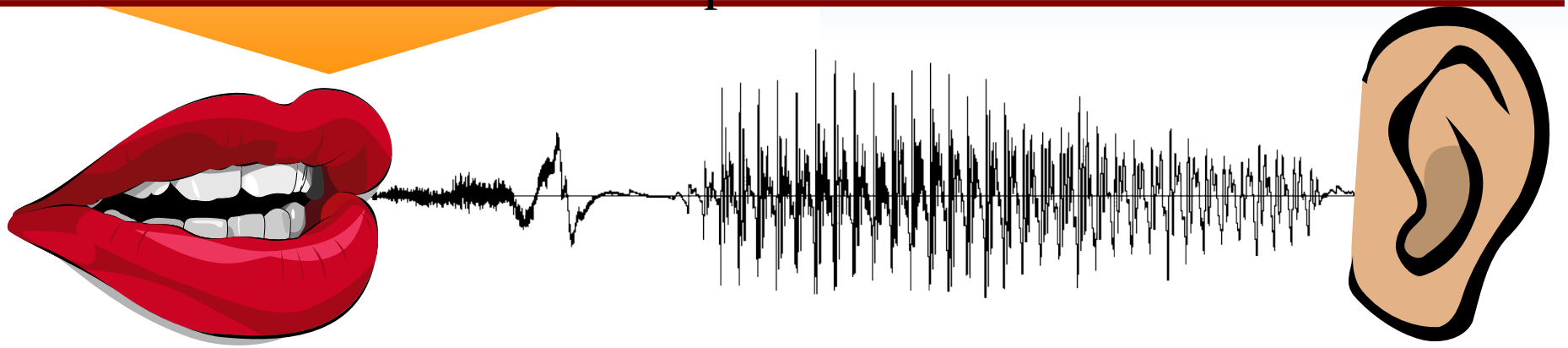
- the question of who talks when in a dialogue
  - „who holds the floor“
    → the task is called floor-tracking or end-of-turn-detection
- need to find out whether the other speaker has finished / whether it's OK to start speaking

The many kinds of turn-taking signals:

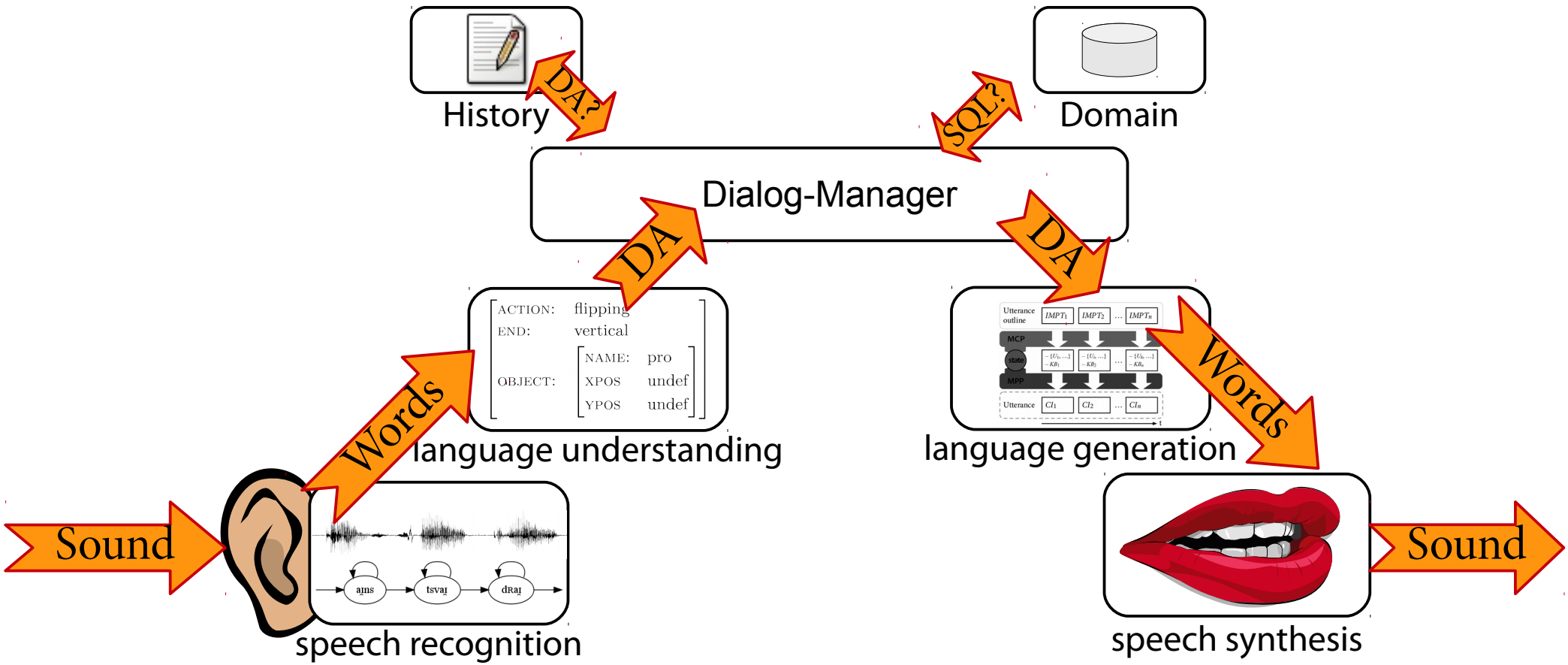
What may indicate that your turn is over /
that your interlocutor may take the floor?



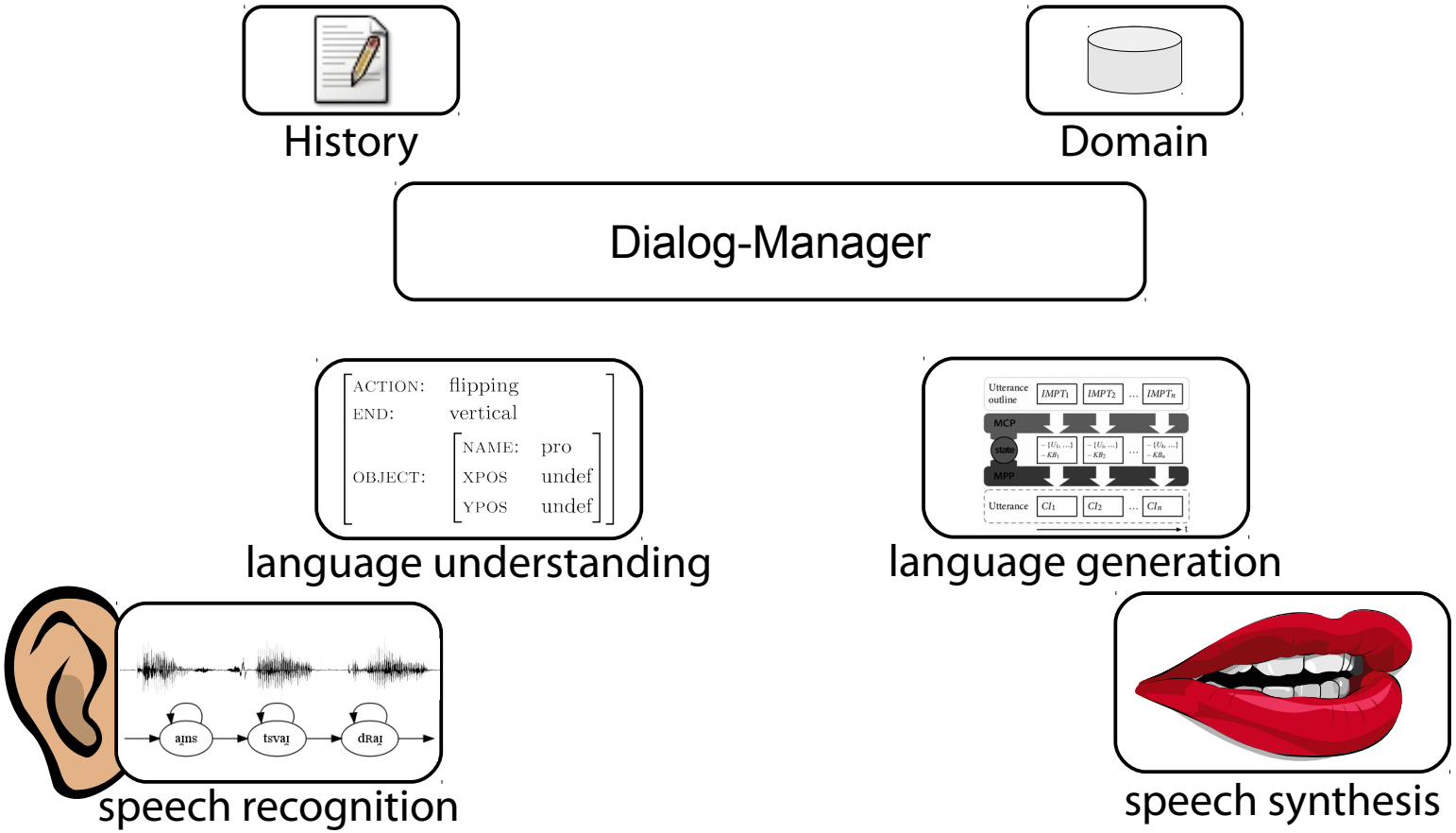
find message that describes idea	pragmatics	recover idea described by message
determine structure to convey meaning	semantics/ lexicology	determine meaning of structure
sequentialize structure to word stream	syntax/ morphology	recover structure of sequence
represent words through sounds	phonology/ phonetics	recombine souds to words



Towards a model of a dialogue agent

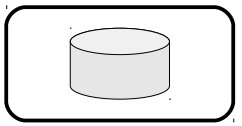


A simple dialogue agent



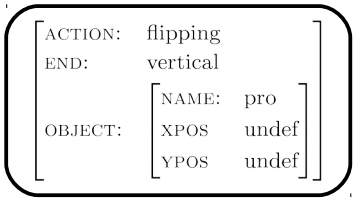
A simple dialogue agent

History

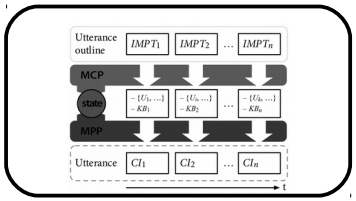


Domain

Dialog-Manager

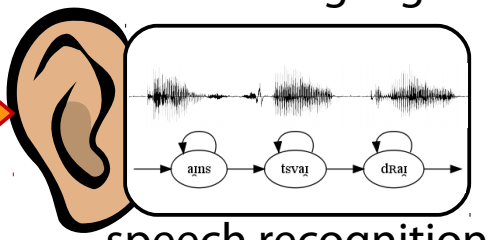


language understanding



language generation

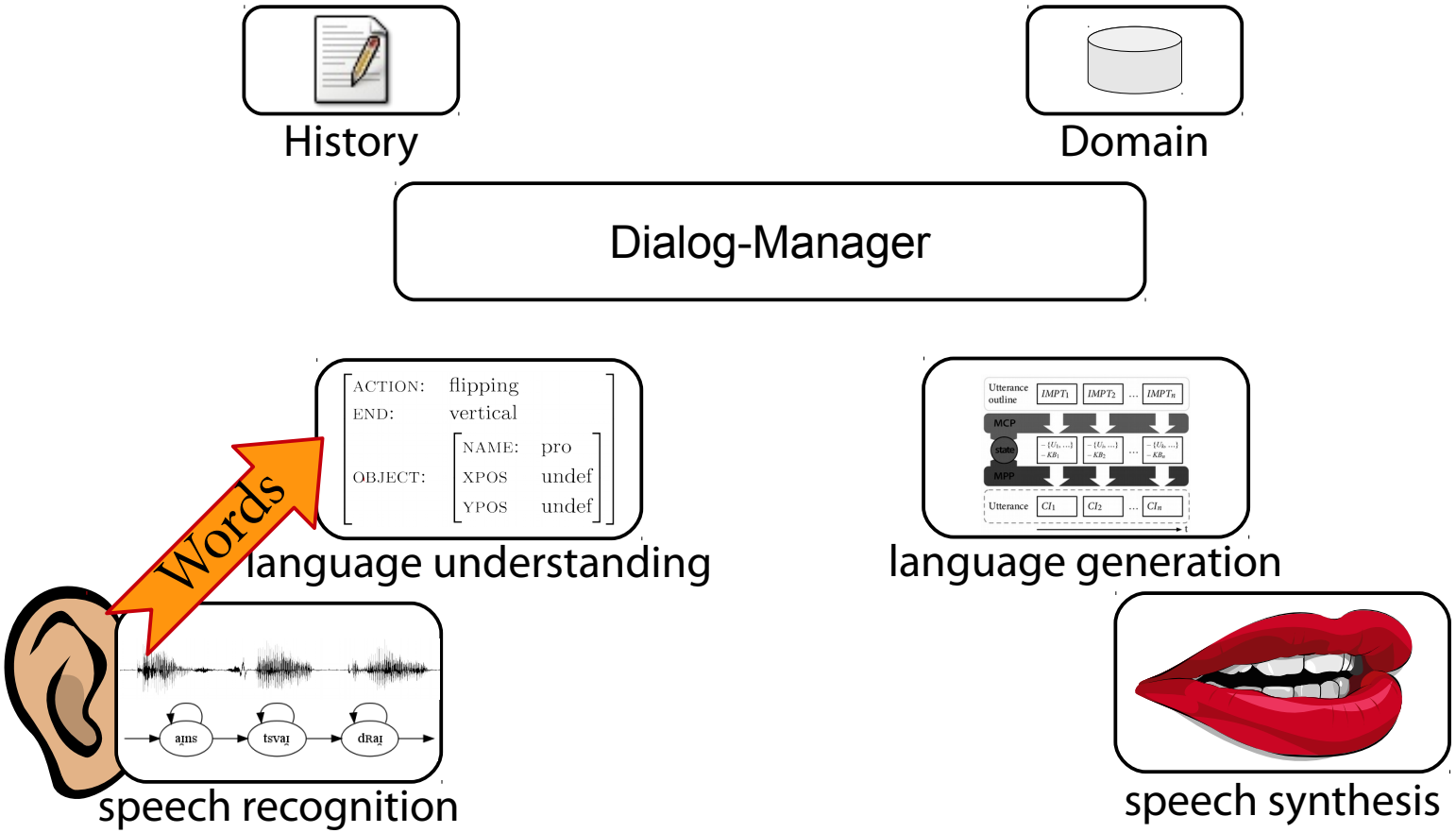
Sound



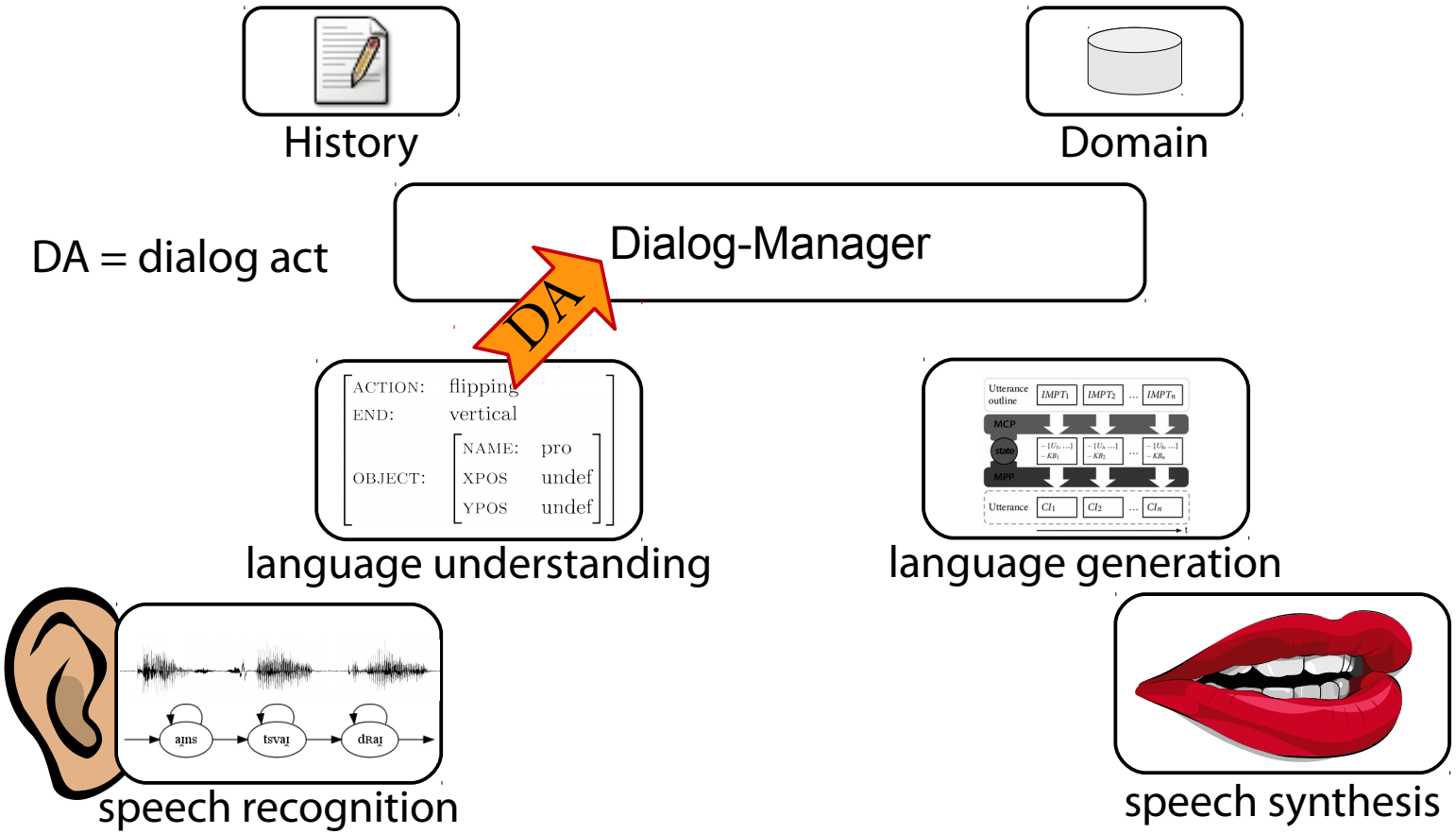
speech recognition

speech synthesis

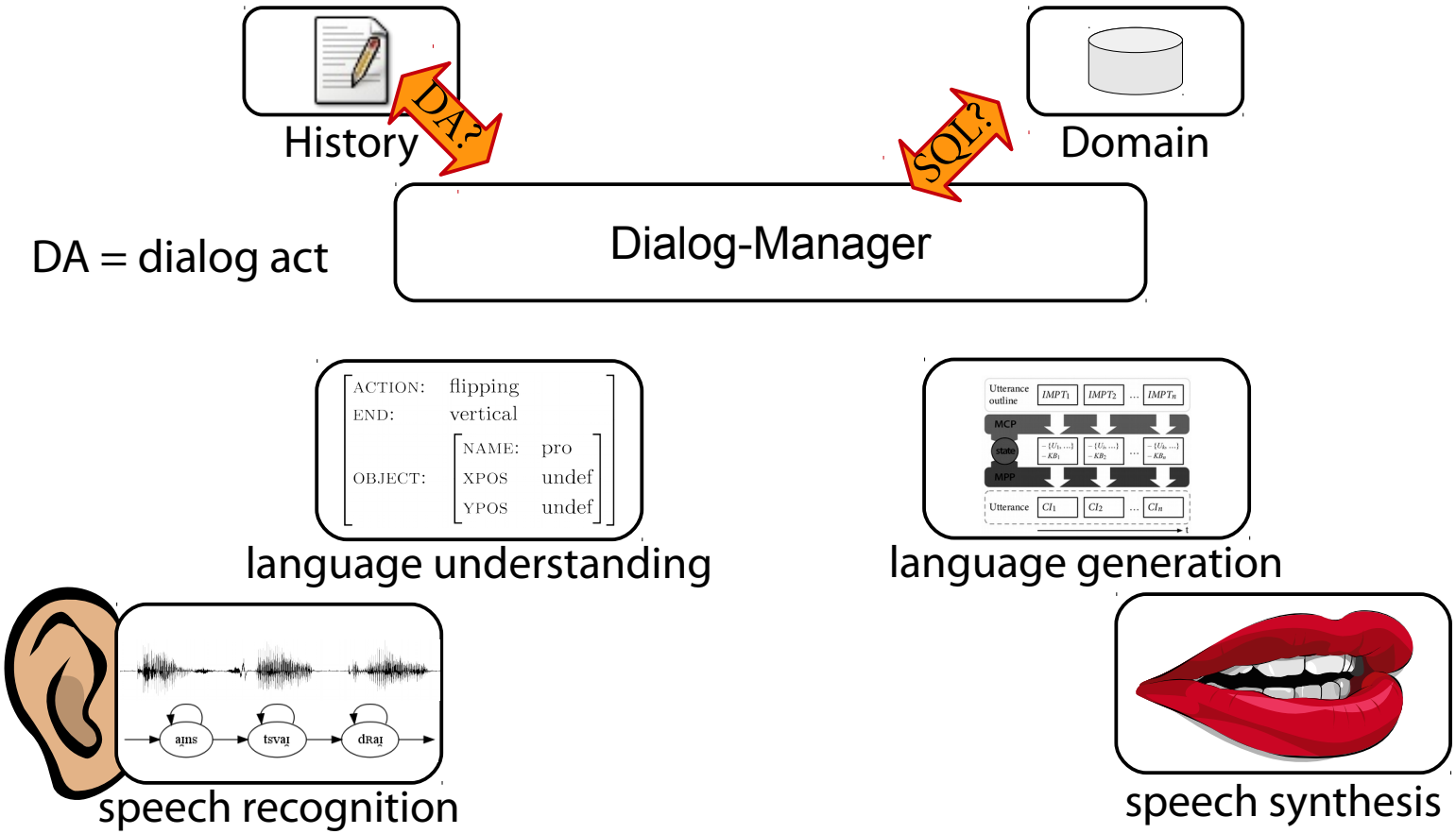
A simple dialogue agent



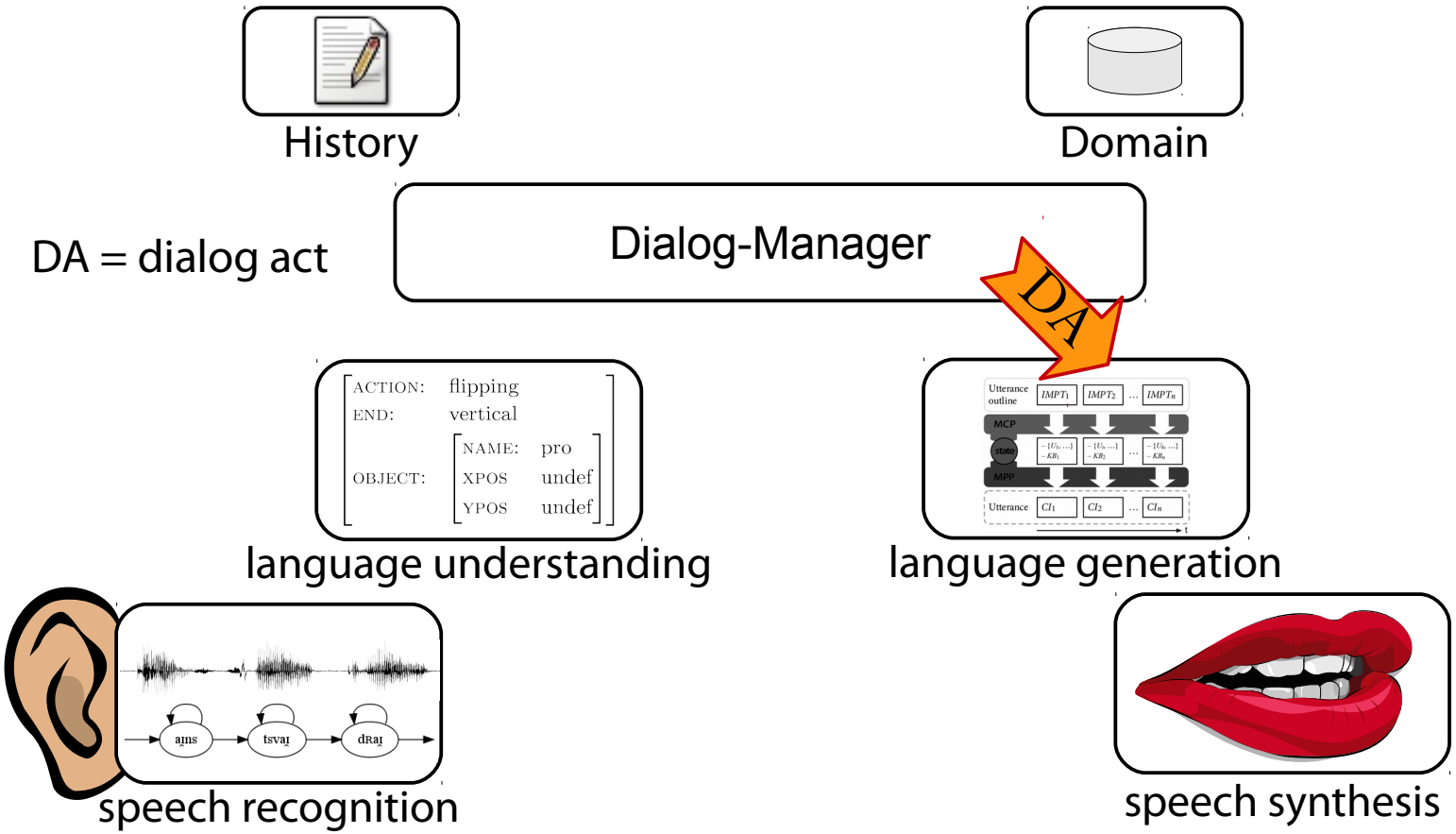
A simple dialogue agent



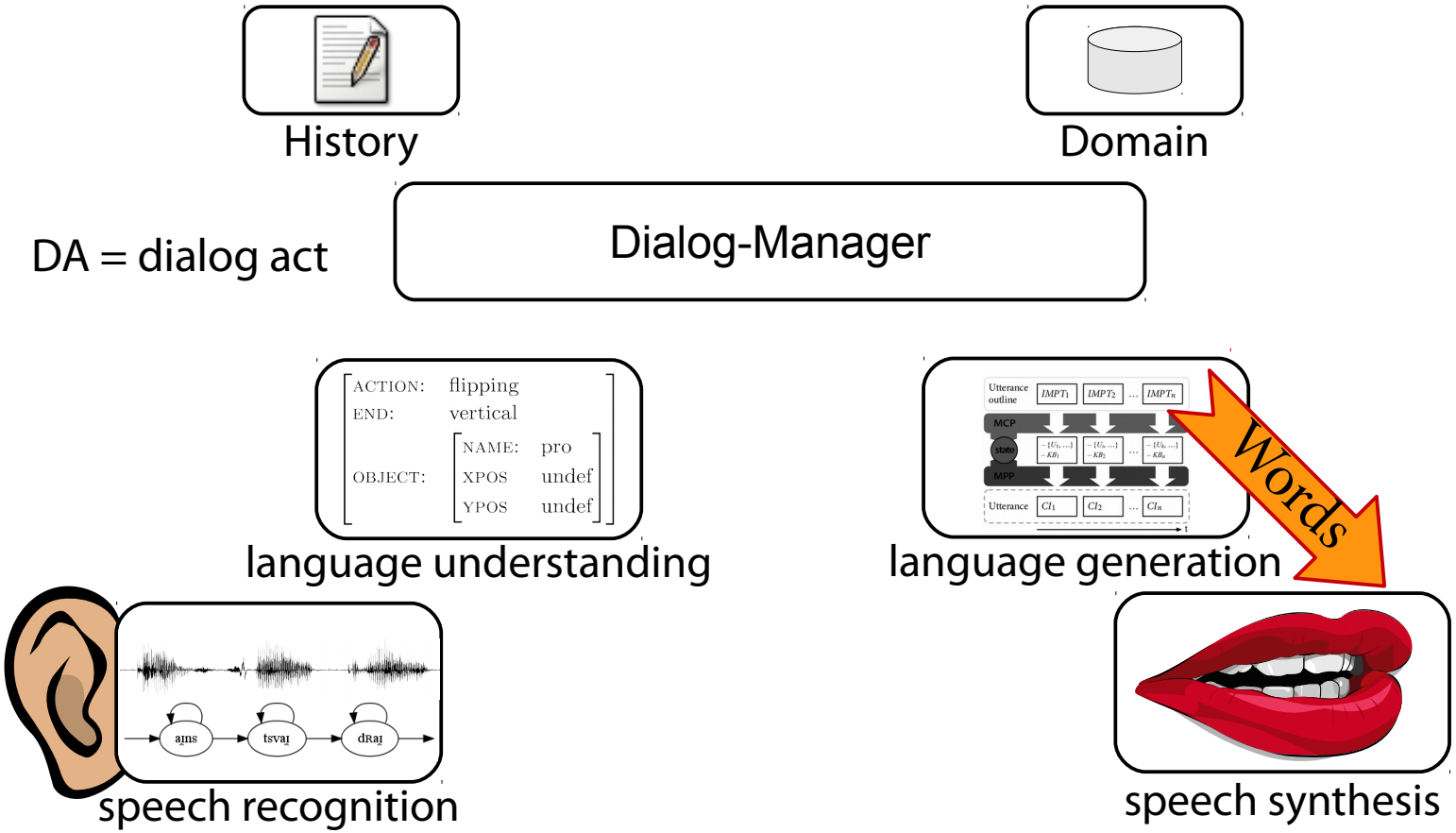
A simple dialogue agent



A simple dialogue agent

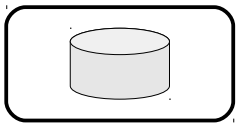


A simple dialogue agent



A simple dialogue agent

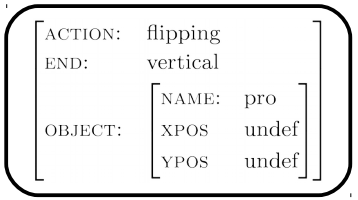
History



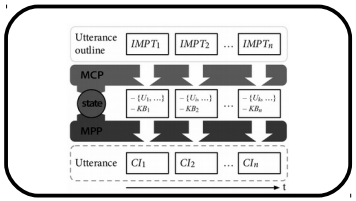
Domain

DA = dialog act

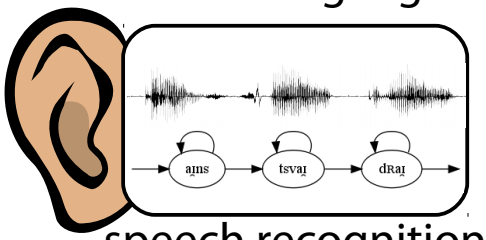
Dialog-Manager



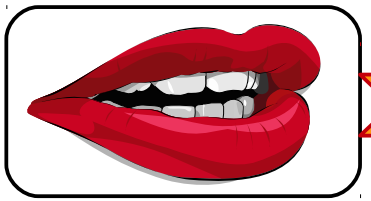
language understanding



language generation

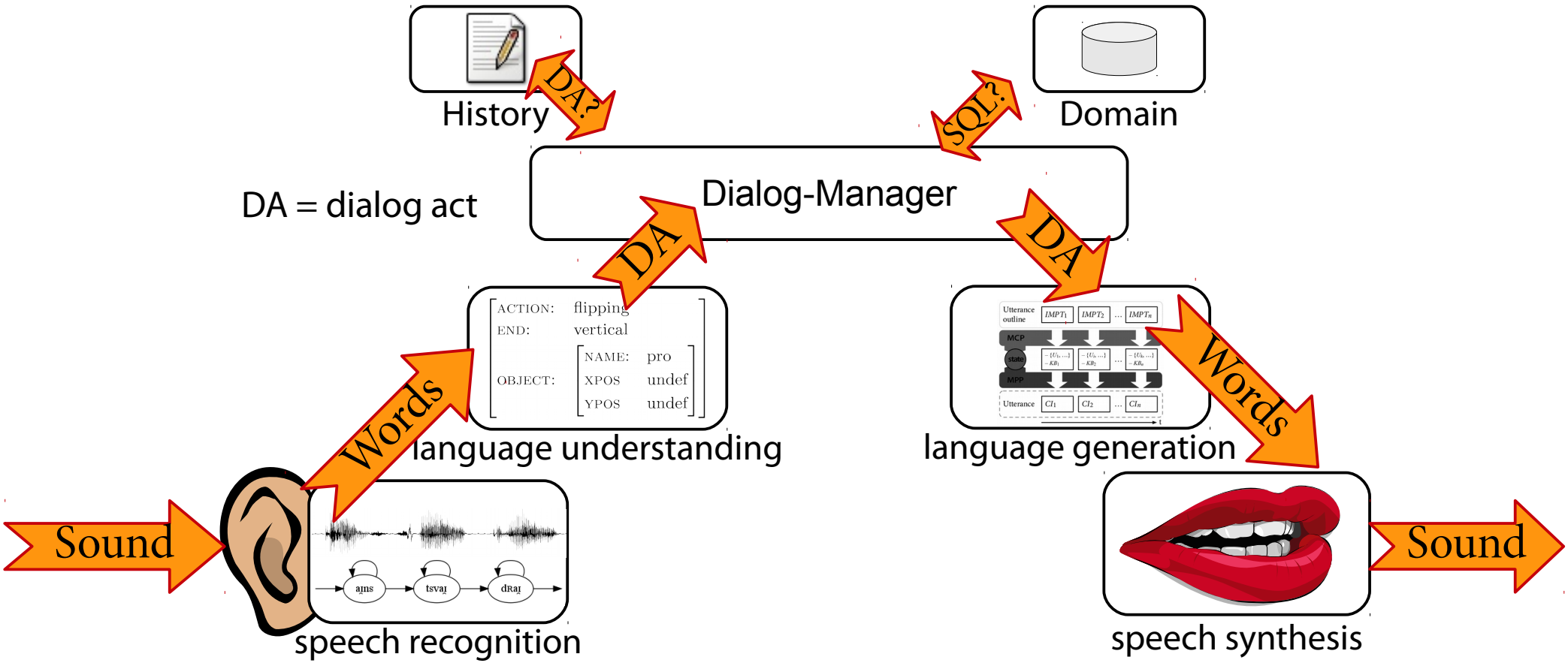


speech recognition

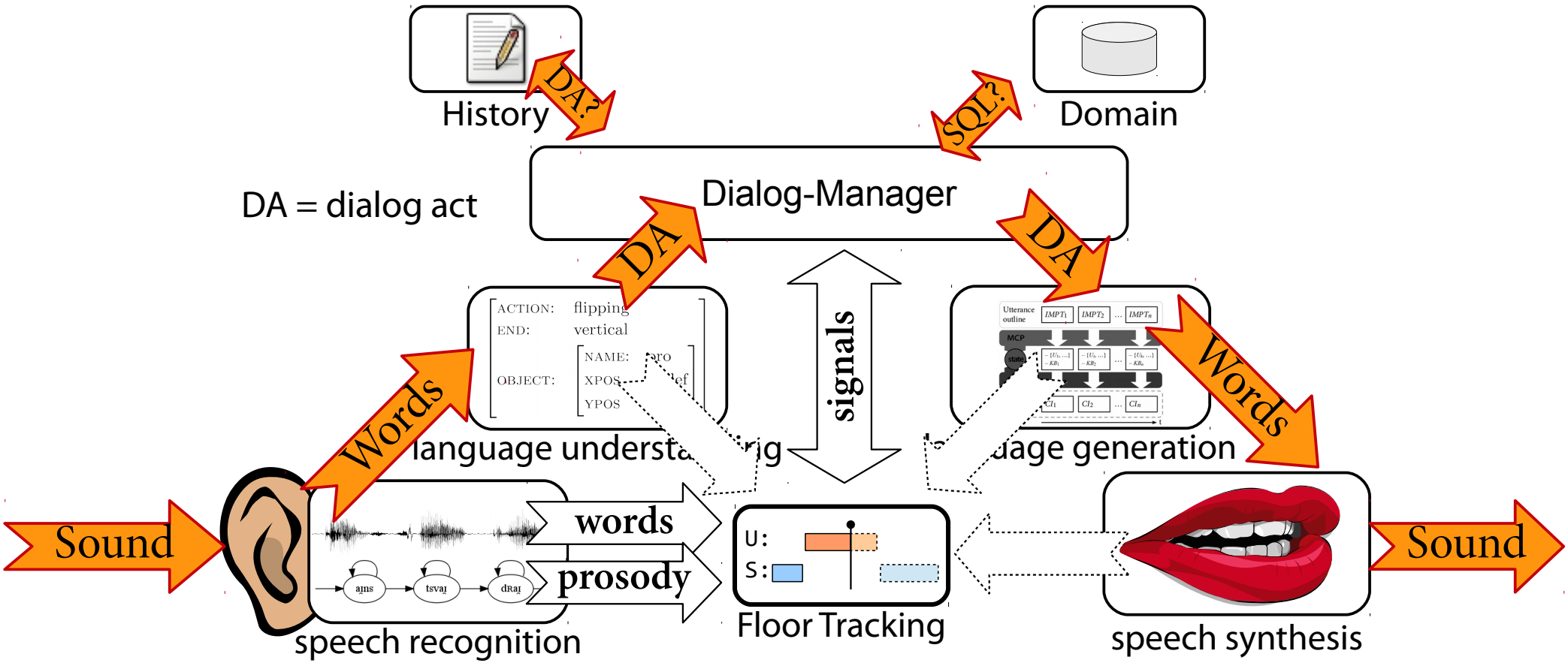


speech synthesis

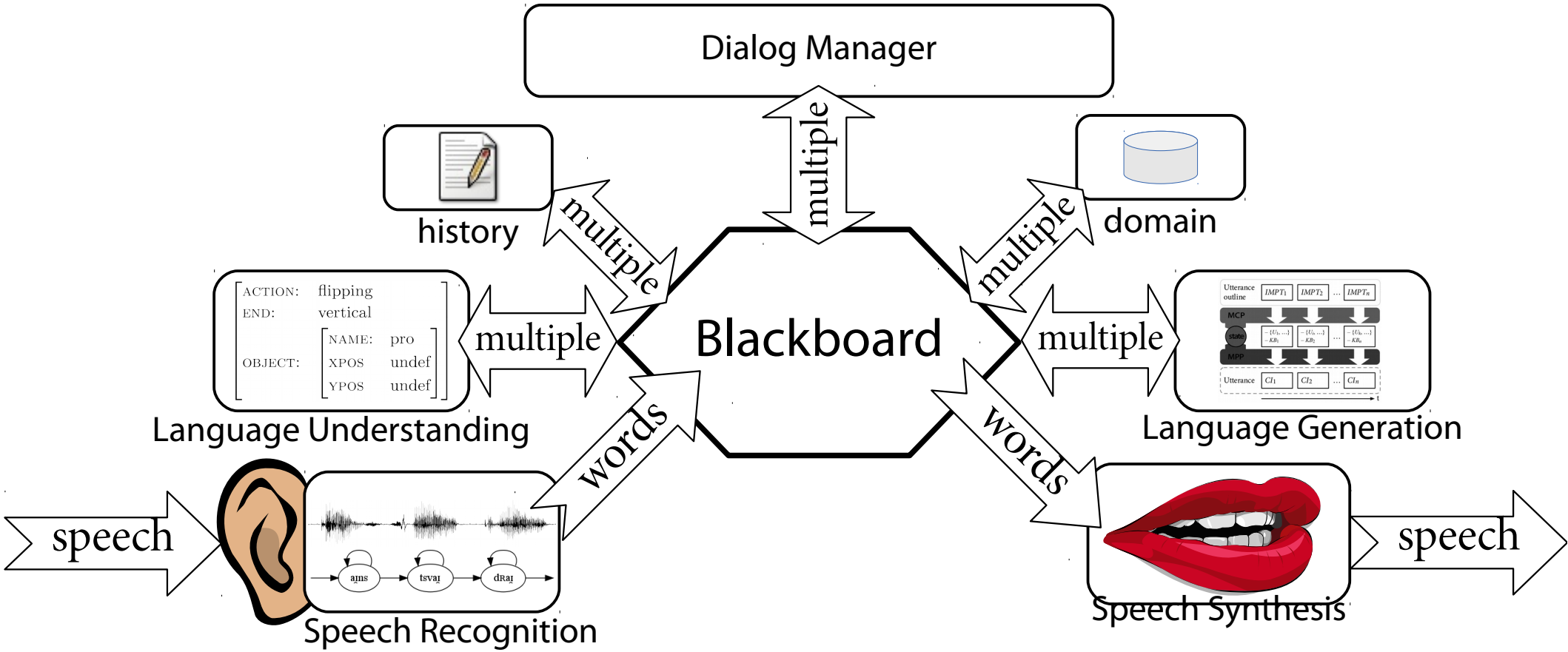
Where is turn-taking?



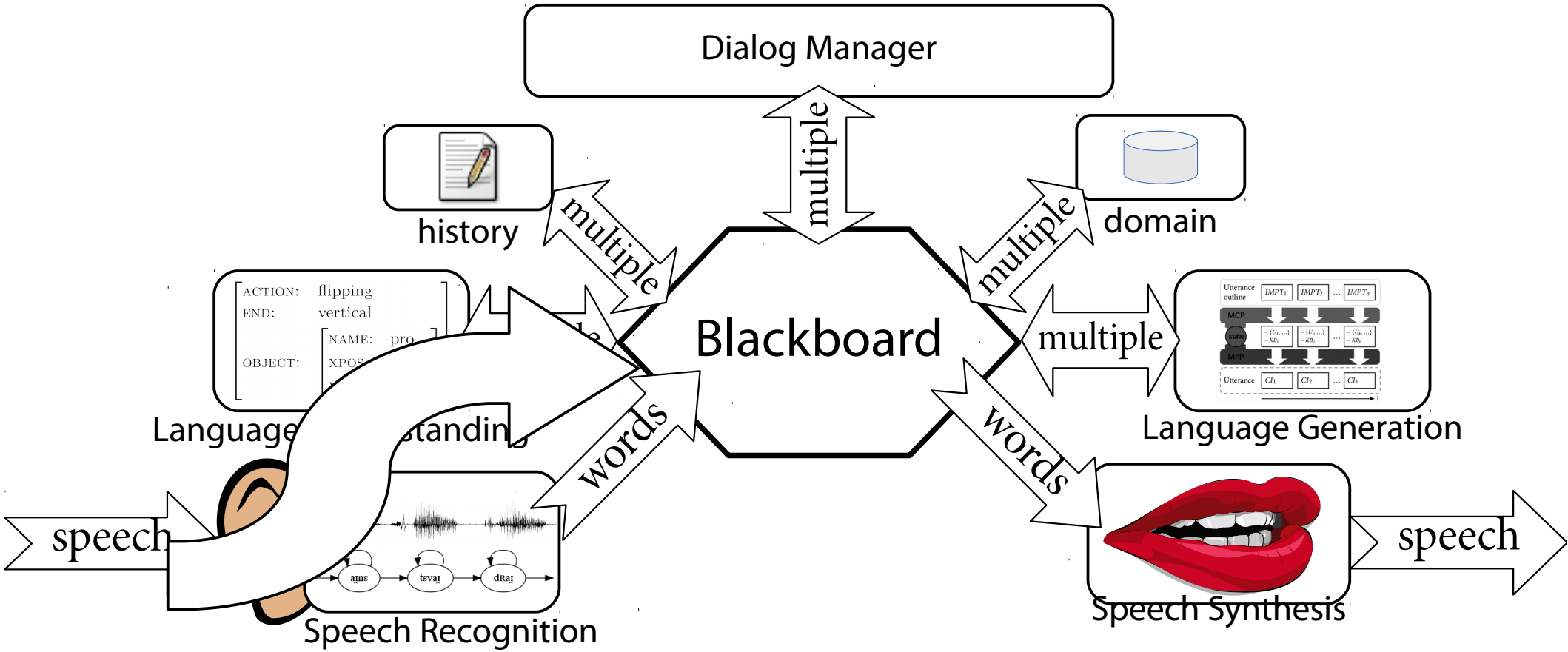
Where is turn-taking?



blackboard-based architecture



blackboard-based architecture



pipeline vs. blackboard

- conceptually very simple
- modules have one input type and one output type
- may use existing modules
- concurrency is easy
- unable to completely solve the problem
- modules can be merged if this is necessary/helpful

- conceptually simple (but complex interactions)
- modules may look at all other modules' output
- may use existing modules (but then loose advantage)
- concurrency is very hard
- in principle able to solve the dialogue problem
- merging is not necessary but possible

How can a simple dialogue agent work?

- dialog interaction is very robust
- in particular, turn-taking behaviour in humans is excellent
- (systems) theoretically: different attractors are available
  - coming to a „slow mode“ of turn-taking if other is slow to repond
  - conversing more clearly to be understood
  - stopping oneself from giving feedback if other is confused by that
  - ...

Conclusion

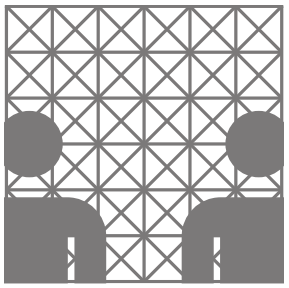
- most applied systems are *modular* and *pipeline-based* (possibly where some modules are merged, forming their own mini-blackboards)
- most applied systems ignore all but the very obvious turn-taking signals: they speak after no-one else has spoken for some time (e.g. 500 milliseconds) and they stop speaking when someone else *barges in*
- turn-taking is extremely complex and uses prosodic and other features
- turn-taking is very robust. This relies on attraction towards stable states in the complex dialogue system
  - dialog systems get away with spending little effort on good turn-taking behaviour

Thank you.

baumann@informatik.uni-hamburg.de

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

UNIVERSITÄT HAMBURG, DEPARTMENT OF INFORMATICS
NATURAL LANGUAGE SYSTEMS GROUP



Further Reading

- Chain model of communication:
  - M. Pétursson & J. Neppert (1996): Elementarbuch der Phonetik. Buske. StaBi: F Ling 062/6.
- Introduction to Dialogue and Linguistics:
  - the relevant chapters in: Jurafsky and Martin (2009): *Speech and Language Processing*. Pearson International. InfBib: A JUR 4204x.
- Systems theoretic views on complex systems in general and on language in particular:
  - Bertalanffy (1972): „The History and Status of General Systems Theory". In: *The Academy of Management Journal* 15(4), pp. 407-426. via Google Scholar.
  - Larsen-Freeman and Cameron (2008): *Complex Systems and Applied Linguistics*, Oxford University Press. StaBi: A 2009 / 7836.

Notizen

Desired Learning Outcomes

- interaction management is a crucial aspect of dialogue
  - in particular channel management in multiple ways
- turn-taking cannot easily be allocated to a „module“ but it emerges from the interaction
- prosody is a field of phenomena relevant in many linguistic layers
- students grasp the idea of emergence in complex systems and attraction as a principle to control such systems