



KNOWLEDGE DISCOVERY

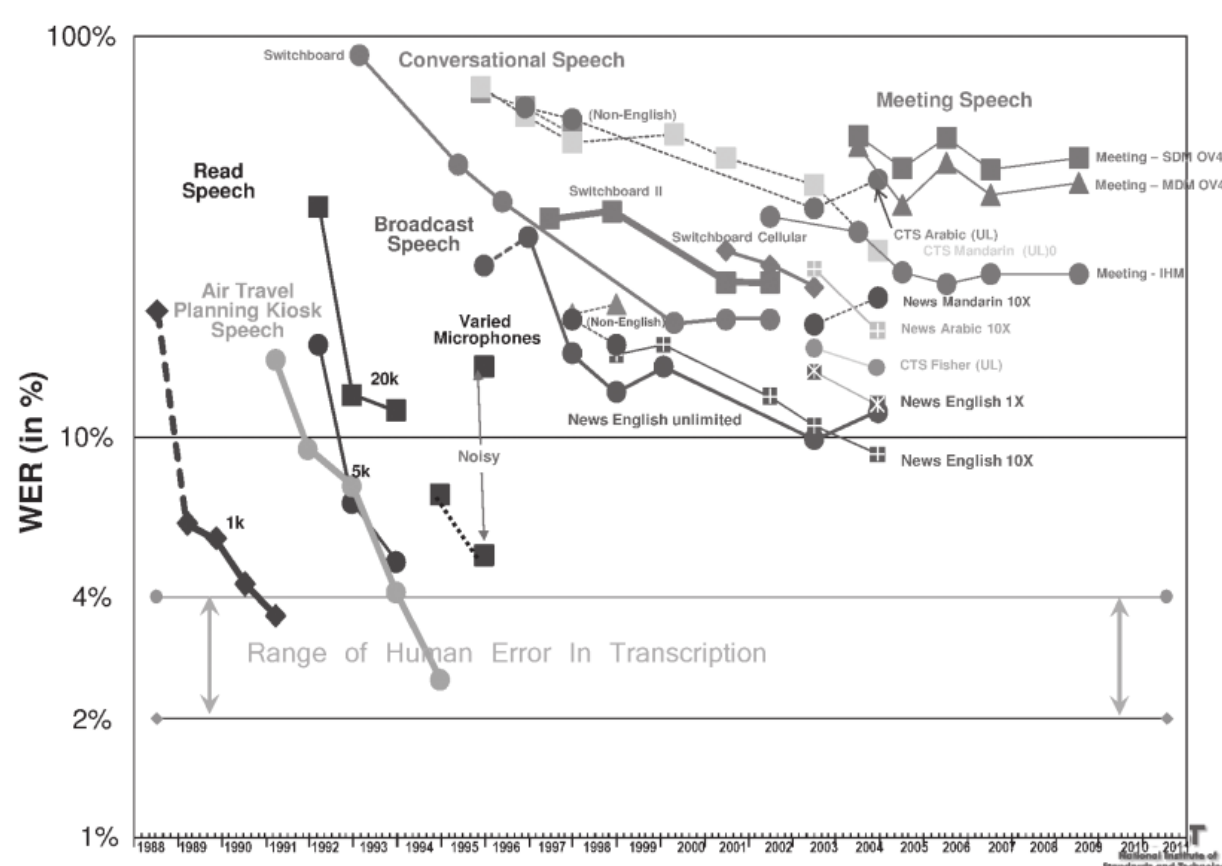
CROSS-LAYER LANGUAGE MODELS FOR CONVERSATIONAL SPEECH

Philipp Gabler, Roman Kern, Bernhard Geiger, Barbara Schuppler

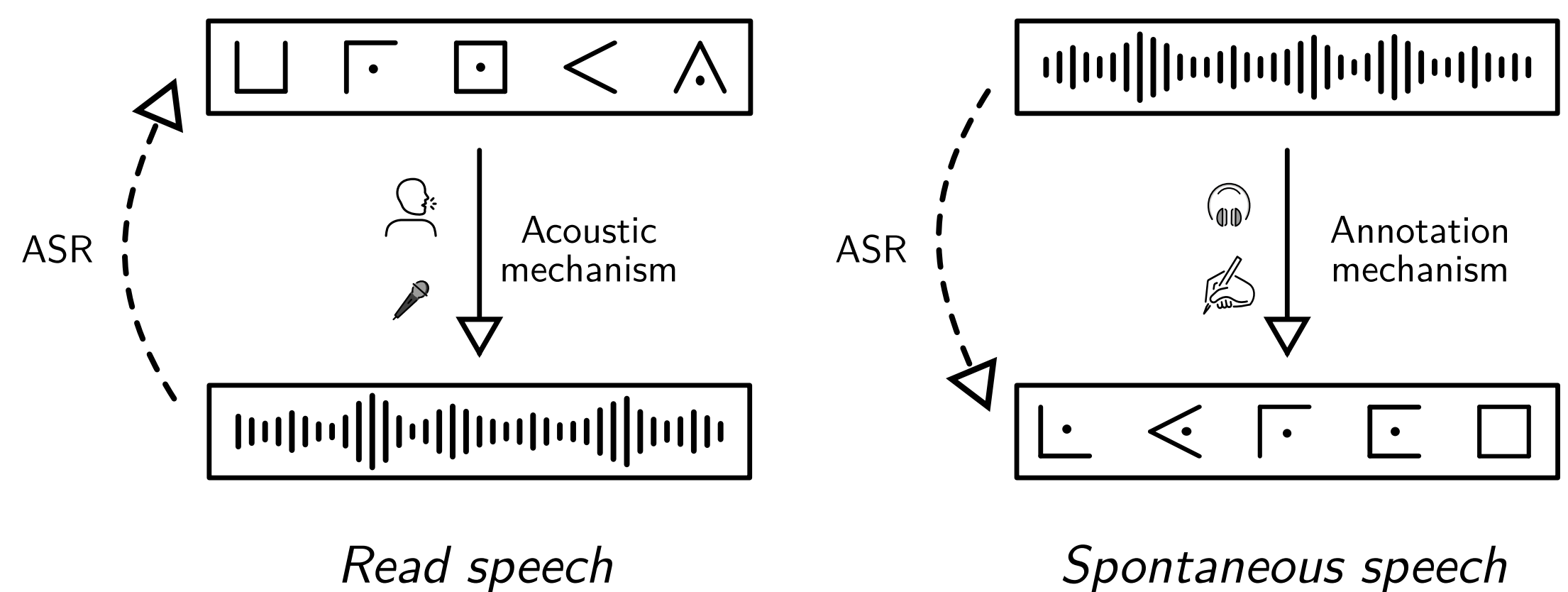
MOTIVATION & BACKGROUND

Project target [1]: how to design & train **language models for automatic speech recognition** (ASR) for conversational Austrian German, integrating components such as prosody, context, or conversation state, and allowing for linguistically relevant analysis.

■ Spontaneous speech is difficult [2]



■ Differences in mechanisms not widely recognized (cf. [5])



■ Resources are few and costly, transfer is hard

also WEESS ick nich
wenn de dir .h
mal irgendwie ne BRAvo kuckst

Example from [3]. Maybe transfer of acoustic model can help? [4]

■ Annotation is a relevant part of the process [6]

hast einen <*EN>reflow offen

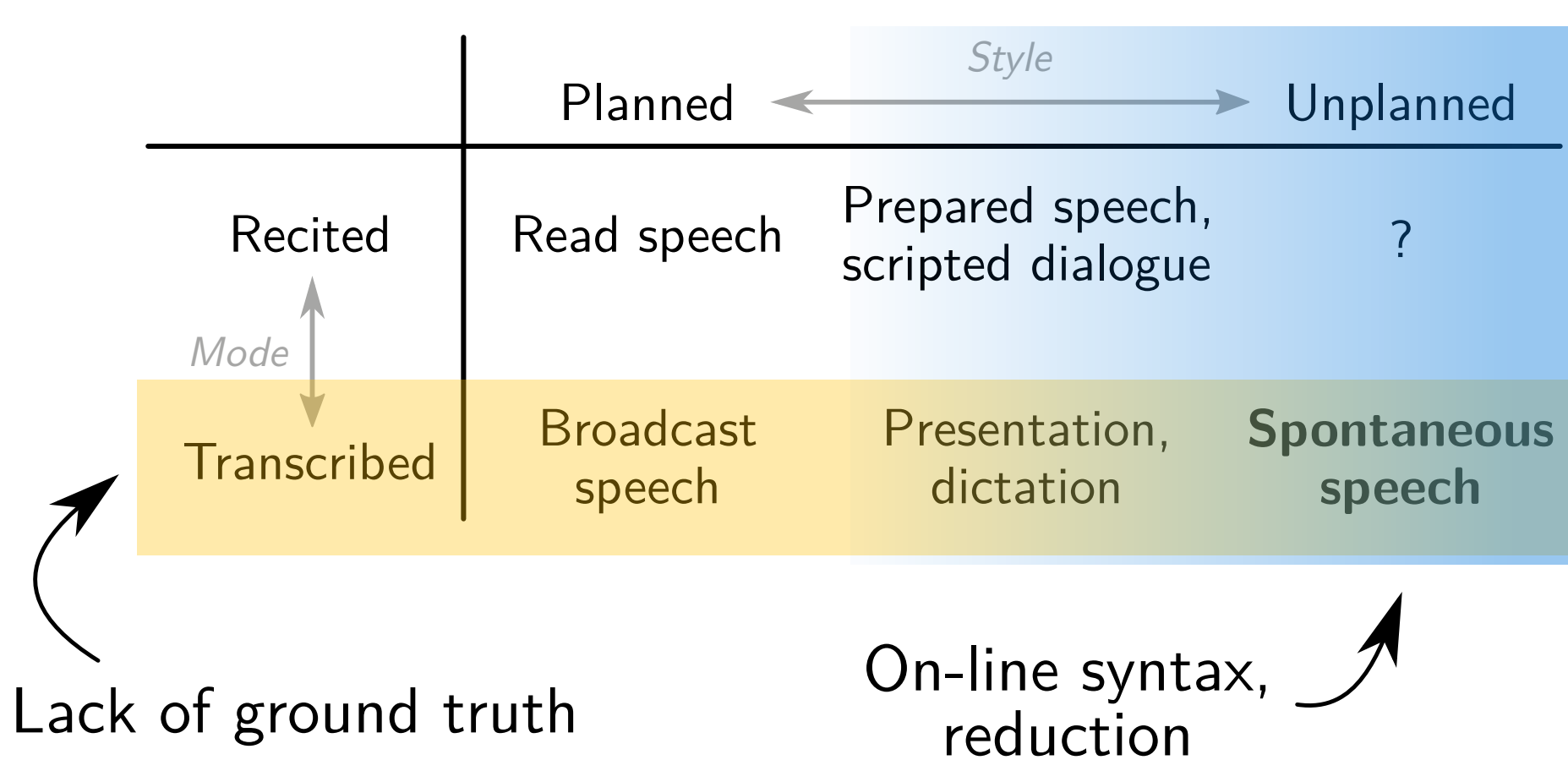
“[do you] have a reflow open”

musst einen <?>Ofen bauen

“[you] must build a <?> oven”

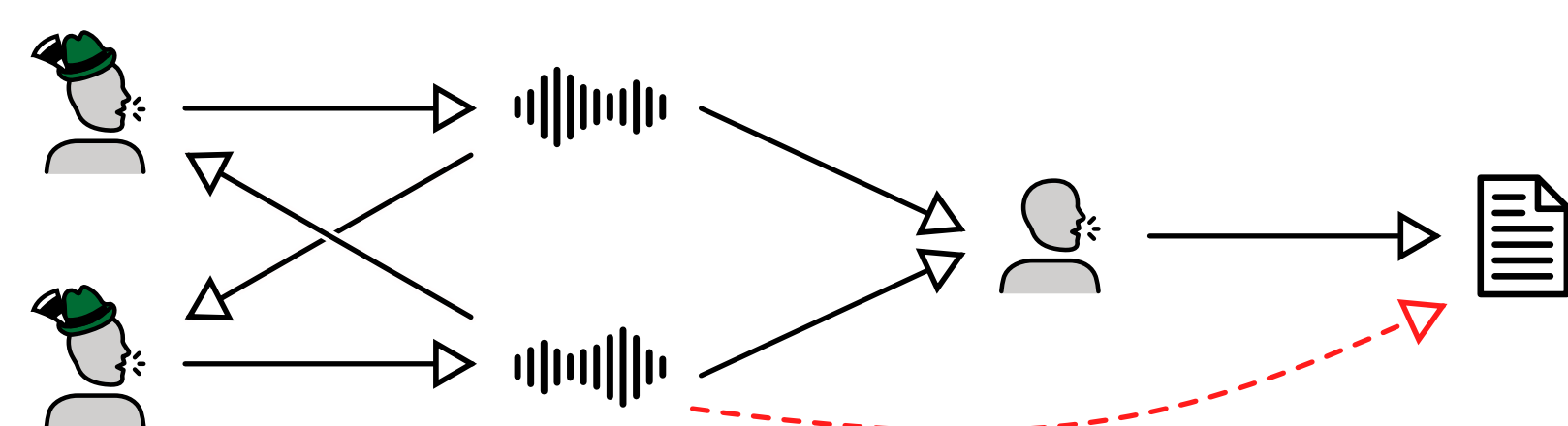
PROJECT PLAN

■ Problem space



■ Current directions:

- Show how causal direction is relevant: analysis of **semi-supervised learning** [6]
- Take annotation seriously: **annotation models** for better estimates of hardness and errors
- Flexible generative models for different settings, including **conversation**



REFERENCES

- [1] The work is funded by grant P-32700-N from the Austrian Science Fund.
- [2] J. Ajot and J. Fiscus, "Speech-To-Text (STT) and Speaker Attributed STT (SASTT) Results", NIST Rich Transcription Evaluation Workshop, 2009.
- [3] P. Auer, "Syntax als Prozess," Interaction and Linguistic Structures, no. 41, 2005.
- [4] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," arXiv:2006.11477 [cs, eess], 2020.
- [5] Z. Jin et al., "Causal Direction of Data Collection Matters: Implications of Causal and Anticausal Learning for NLP," arXiv:2110.03618 [cs], 2021..
- [6] B. Schölkopf, D. Janzing, J. Peters, E. Sgouritsa, K. Zhang, and J. Mooij, "On Causal and Anticausal Learning," arXiv:1206.6471 [cs, stat], 2012.