KNOWLEDGE DISCOVERY

# CROSS-LAYER LANGUAGE MODELS FOR CONVERSATIONAL SPEECH

**Philipp Gabler**, Roman Kern, Bernhard Geiger, Barbara Schuppler

## MOTIVATION

Automatic speech recognition (ASR) for **conversational Austrian German is difficult,** because (1) conversational speech is structurally different from written language, and (2) models do not easily transfer to underrepresented varieties.

Project goals [1]:

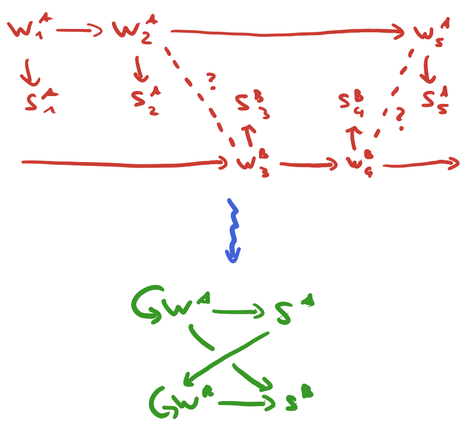
- Improve **ASR for conversational speech**
- Help increase knowledge about **production and perception** of conversational language
- Increase knowledge & **resources for Austrian German**

Research directions:

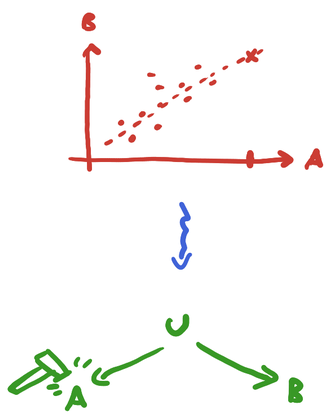
- **LMs for conversational language** & interaction
- **Long-range dependencies**, local contexts, cross-layer features, high-level semantics
- **Generative/causal models** for linguistic data

## BACKGROUND

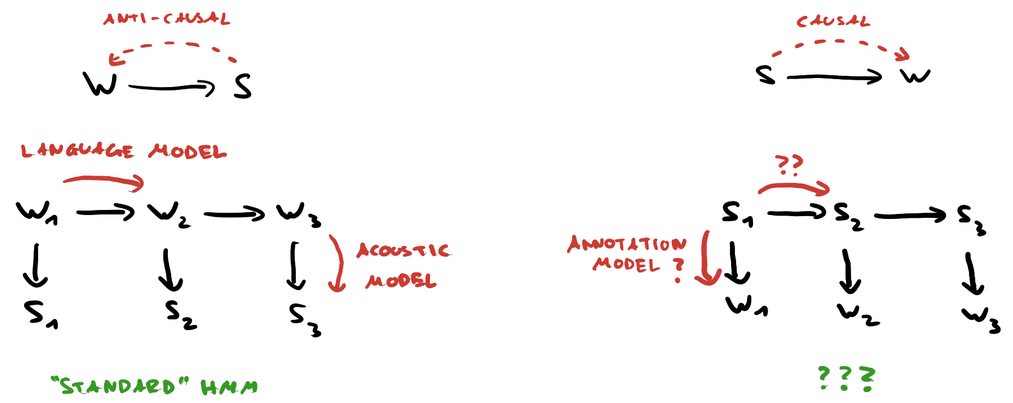
**Multi-speaker LMs [2]**



**Association → mechanism [3]**

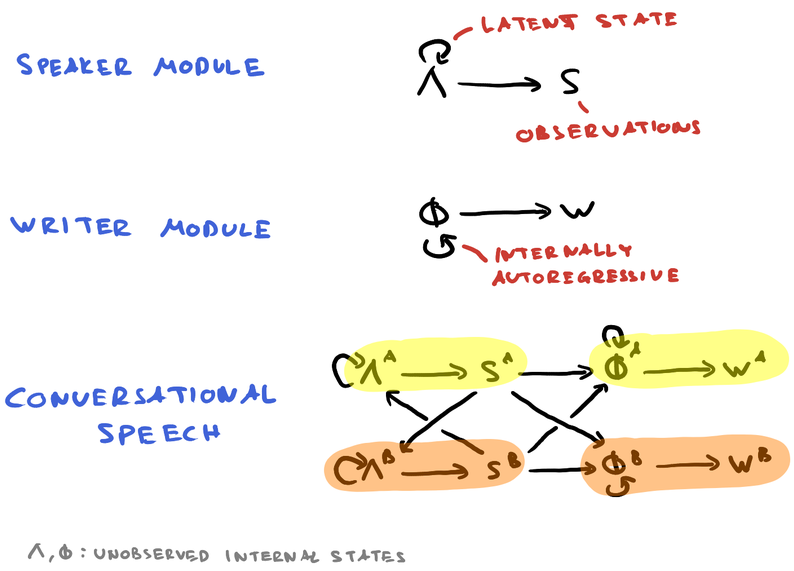



**HMMs [4] are anti-causal – what about annotation?**

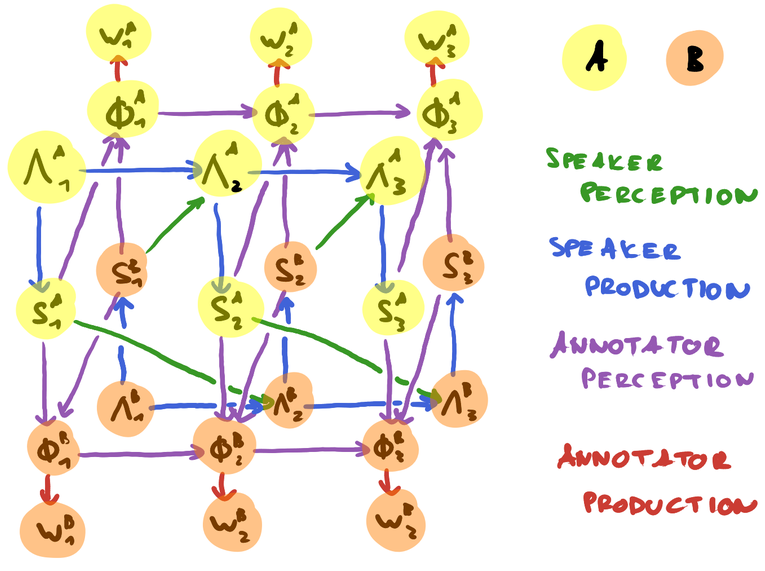



## NEW APPROACH

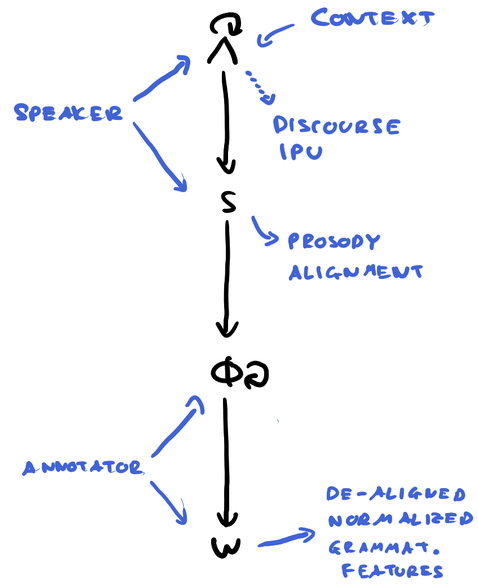
**Modularized, coupled LMs**




**Expansion in time (simplified)**




**Contextual & proxy features**




## PROJECT PLAN

Vision:

- Flexible toolbox of conditional LMs
- Clearly distinguish production & perception
- Latent & observed model for every participant
- Few concrete assumptions made

Remaining questions:

- Concrete structure (coupling)
- Representation (states, alignment, priors)
- Inference
- Simplified setting for exploration

## REFERENCES

[1] The work is funded by grant P-32700-N from the Austrian Science Fund.
[2] G. Ji and J. Bilmes, "Multi-speaker language modeling," in Proceedings of HLT-NAACL 2004: Boston, Massachusetts, 2004, pp. 133–136.
[3] J. Pearl, Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press, 2009.
[4] D. Jurafsky and J. H. Martin, Speech and language processing (3rd edition draft). 2019.