

MODULAR MECHANISMS

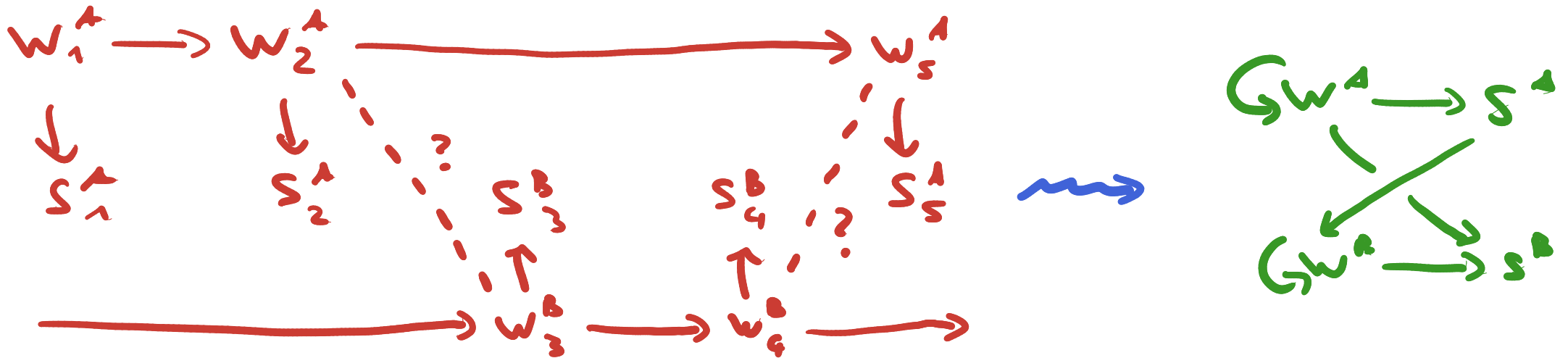
A CAUSAL PERSPECTIVE
ON THE DATA GENERATION PROCESS
OF CONVERSATIONAL SPEECH DATA

PHILIPP GARLER

2021-08-26

MOTIVATION I

MODELS TAKING MULTI-SPEAKER
INTERACTION SERIOUSLY?



MOTIVATION II

APPROACHES TAKING INTO ACCOUNT
DIFFERENCES IN CAUSAL STRUCTURE?

ASSOCIATION



PREDICTION



MECHANISM



INTERVENTION /
COUNTERFACTUAL

THE STORY OF SPEECH RECOGNITION

READ SPEECH



ANTI-CAUSAL

SPONTANEOUS
SPEECH



CAUSAL

W: WRITTEN SIGNAL, S: SPOKEN SIGNAL
(BOTH OBSERVED)

THE STORY OF SPEECH RECOGNITION

READ SPEECH



ANTI-CAUSAL

SPONTANEOUS
SPEECH

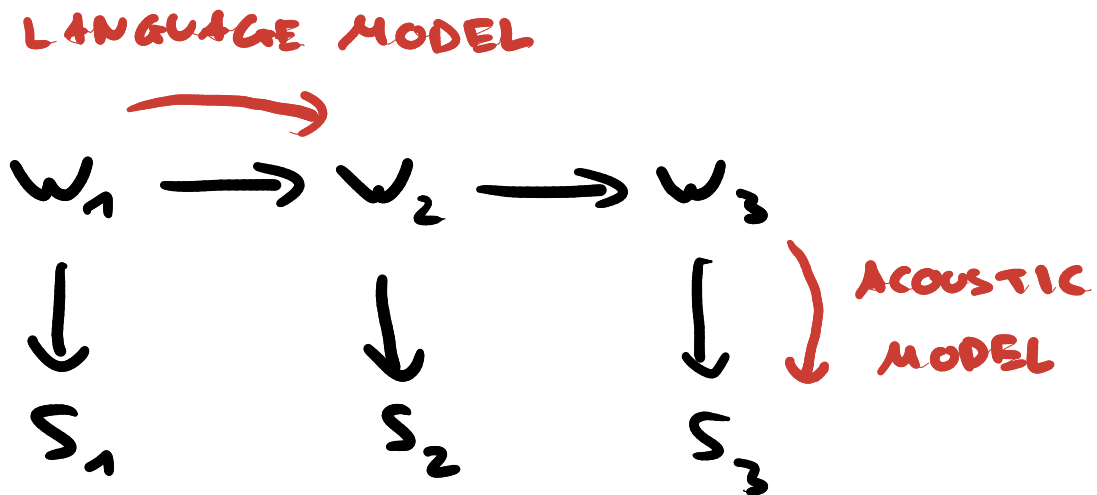


CAUSAL

W: WRITTEN SIGNAL , S: SPOKEN SIGNAL
(BOTH OBSERVED)

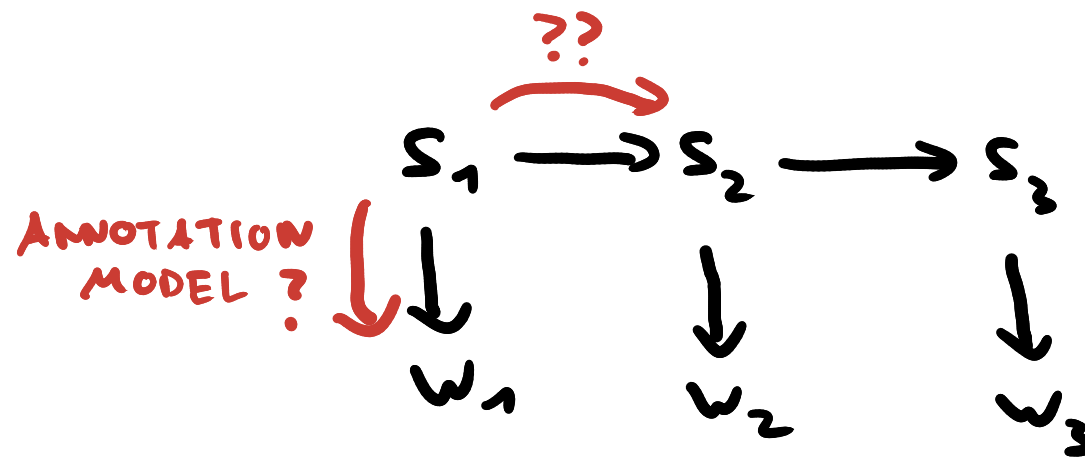
THE STORY OF SPEECH RECOGNITION

READ SPEECH



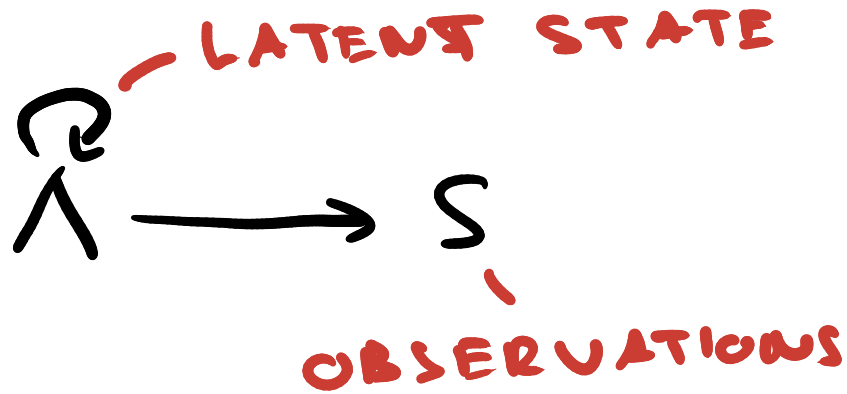
"STANDARD" HMM

SPONTANEOUS SPEECH



???

ABSTRACTION!



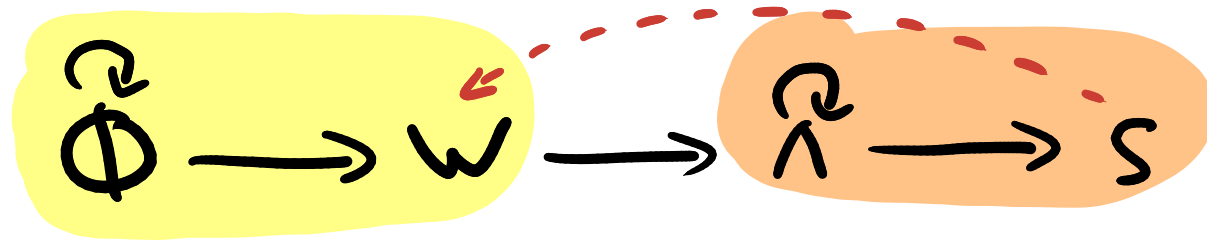
SPEAKER MODULE



WRITER MODULE

Λ, Φ : UNOBSERVED INTERNAL STATES

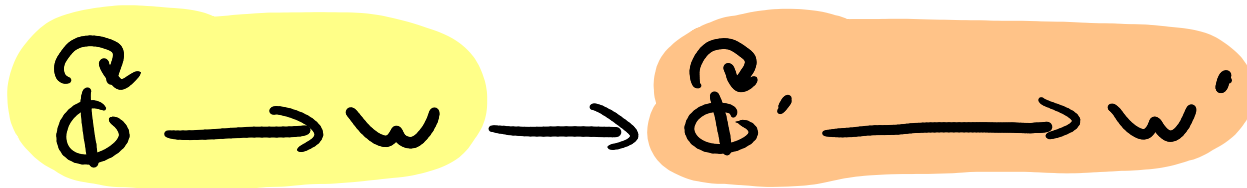
VERTICAL COUPLING



READ
SPEECH

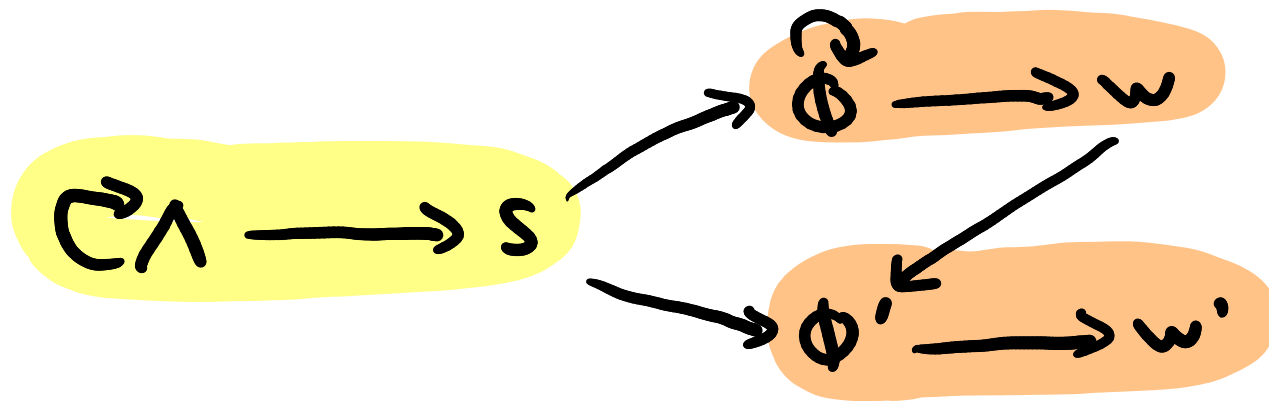


SPONTANEOUS
SPEECH

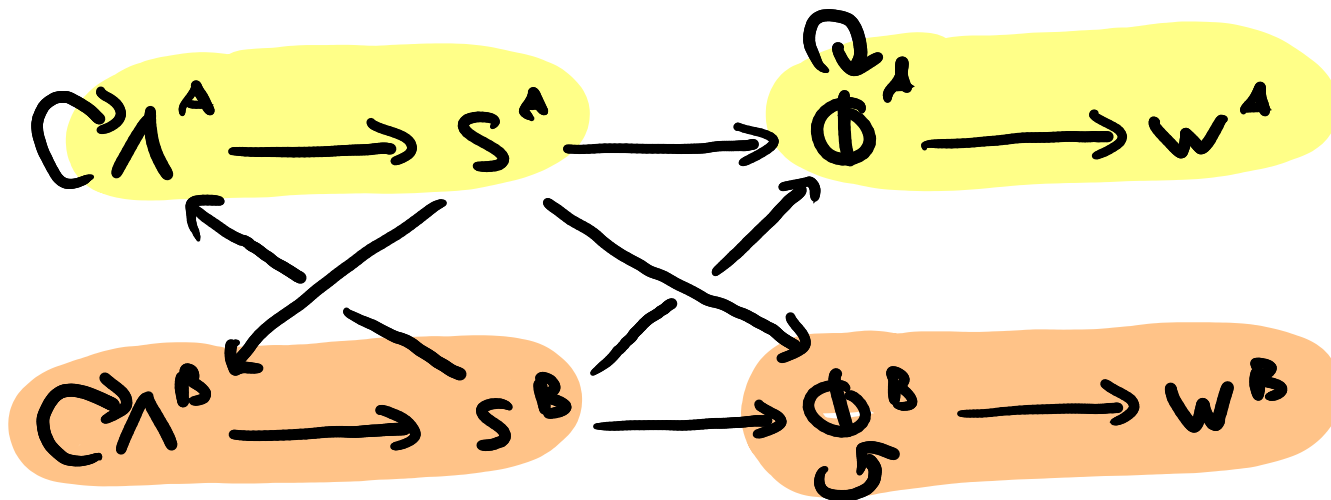


TRANSCRIPTION

DEPTHWISE COUPLING

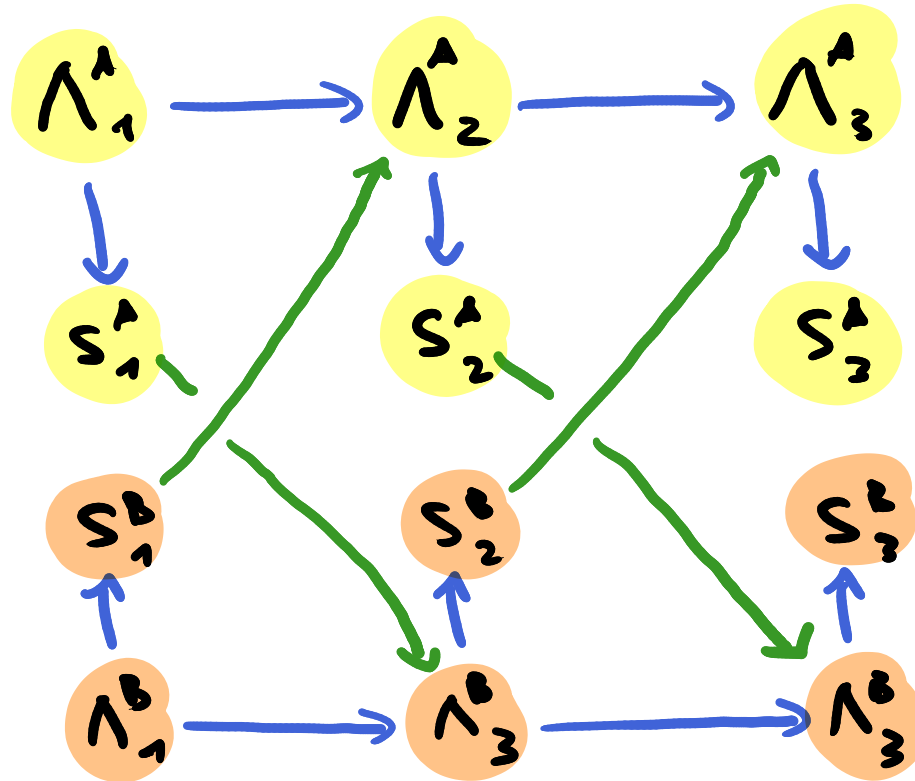
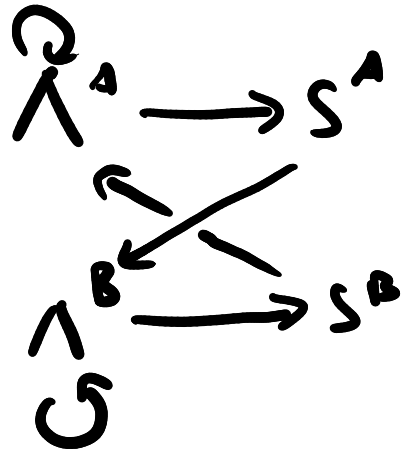


MULTIPLE
ANNOTATION
PASSES



CONVERSATIONAL
SPEECH

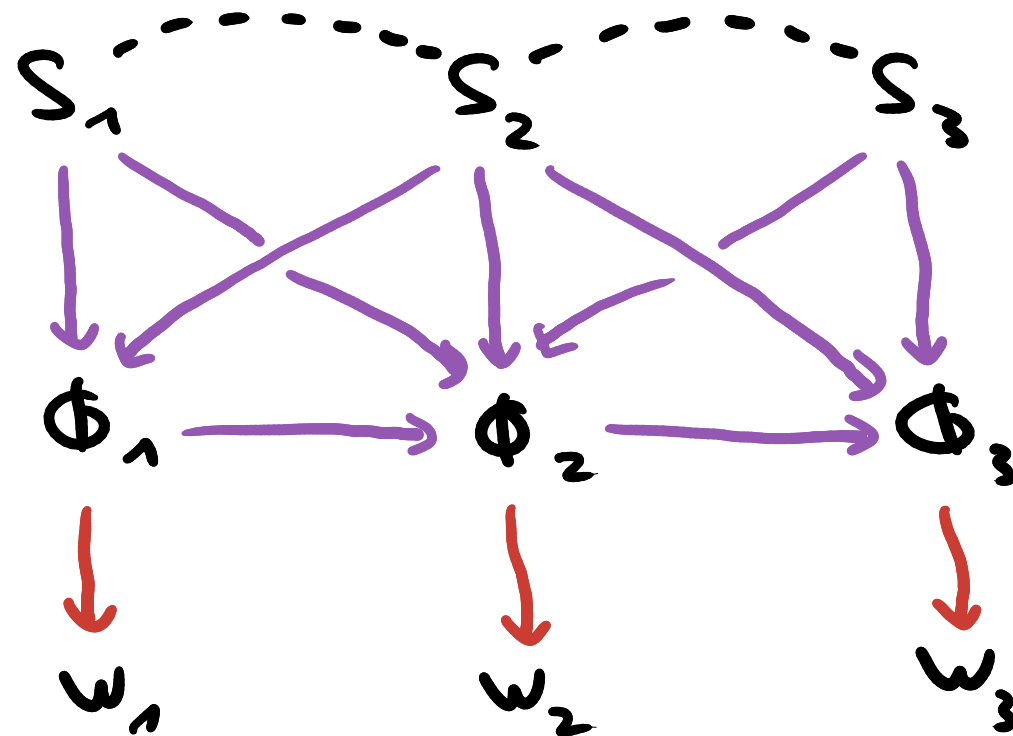
EXPANSION IN TIME



SPEAKER
PERCEPTION

SPEAKER
PRODUCTION

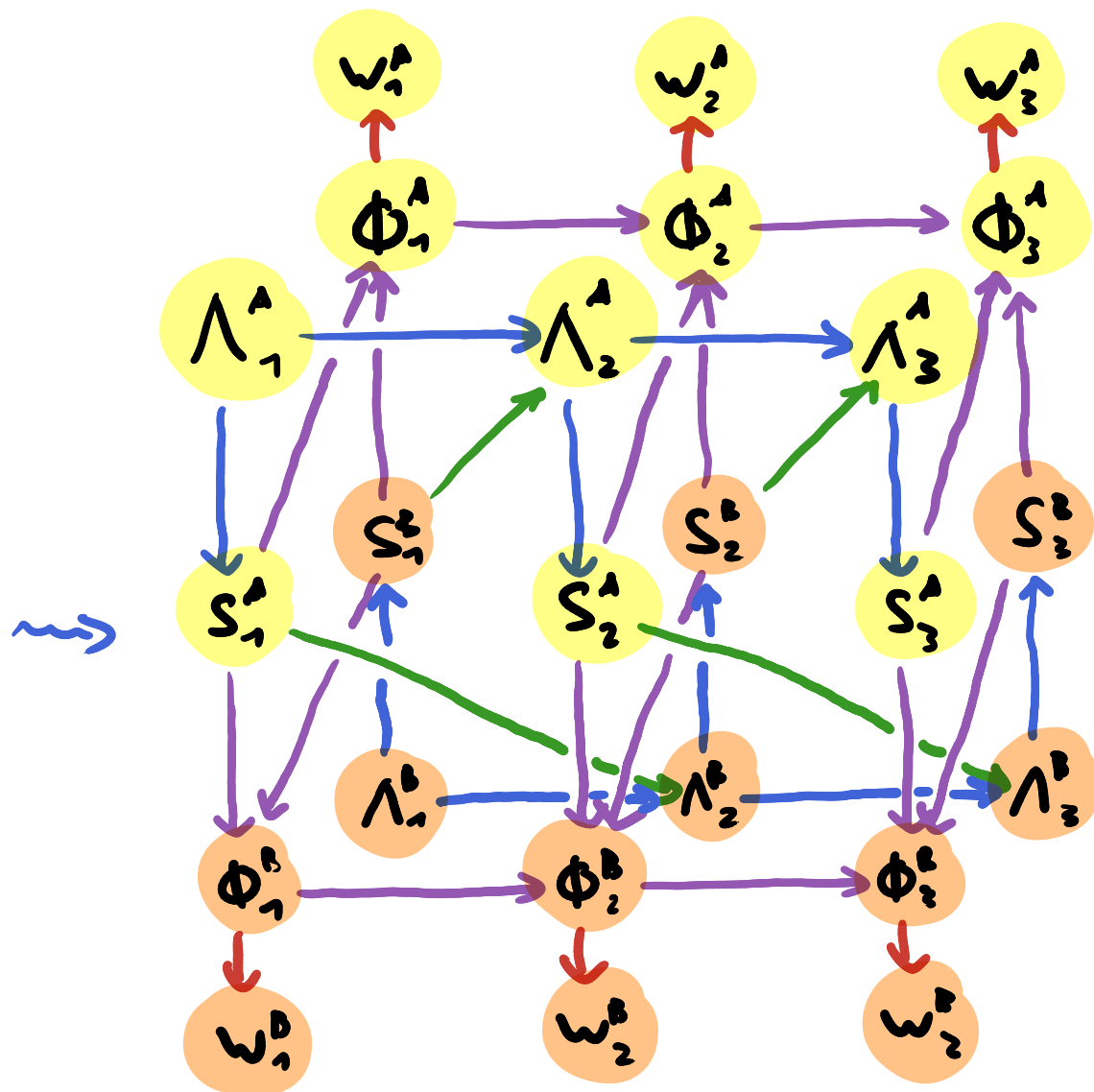
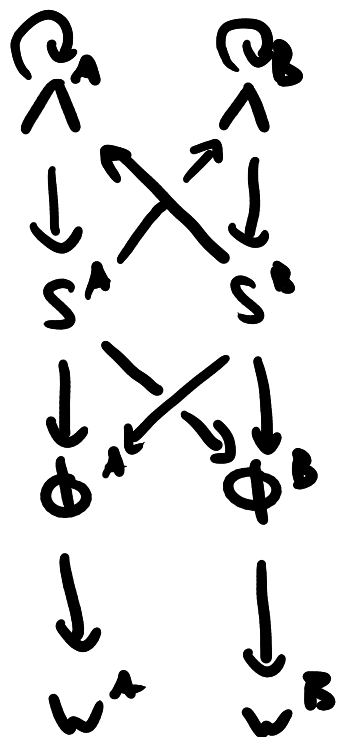
EXPANSION IN TIME



ANNOTATOR
PERCEPTION

ANNOTATOR
PRODUCTION

EXPANSION IN TIME



A

B

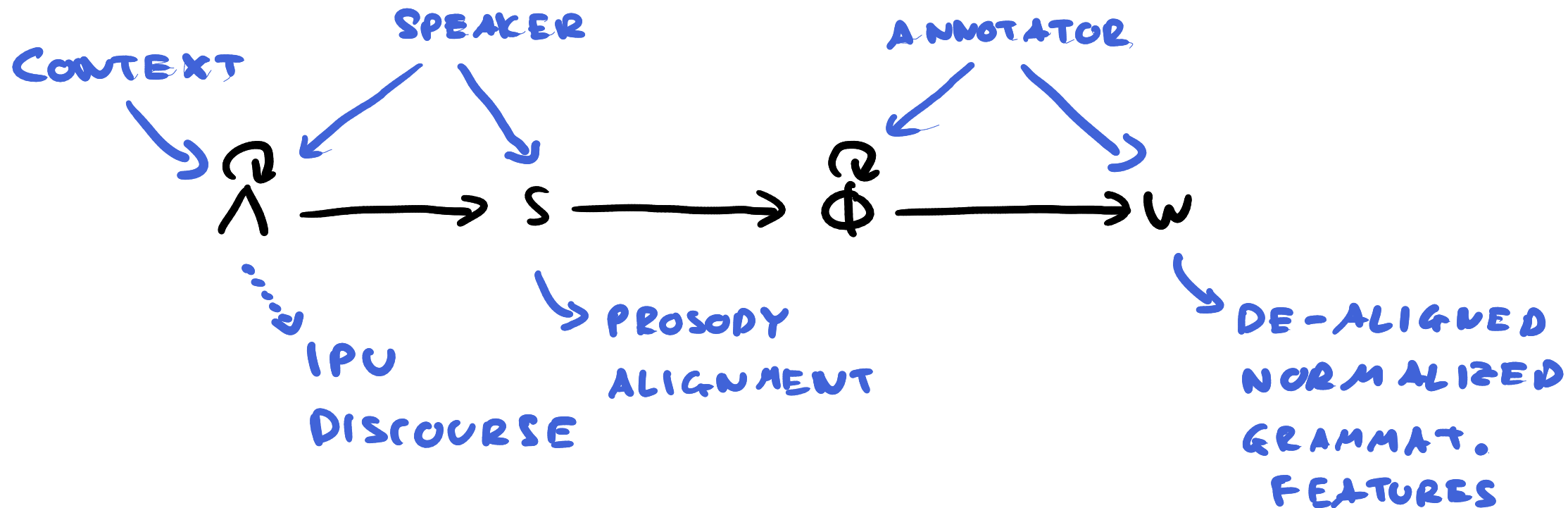
SPEAKER
PERCEPTION

SPEAKER
PRODUCTION

ANNOTATOR
PERCEPTION

ANNOTATOR
PRODUCTION

CONTEXTUAL & PROXY FEATURES



TAKEAWAYS

- FLEXIBLE TOOL BOX OF CONDITIONAL LANGUAGE MODELS
- CLEARLY DISTINGUISH PRODUCTION AND PERCEPTION
- LATENT & OBSERVED MODEL FOR EACH PARTICIPANT
- FEW ASSUMPTIONS MADE, CHOICES LEFT OPEN

QUESTION I: STRUCTURE

- HMM
- ATTENTION
- CONVOLUTION
- DIRECTED / UNDIRECTED

QUESTION II: REPRESENTATION

◦ INTERNAL STATES:

↳ NEURAL NETS ($\hat{=}$ VECTORS + LINEAR ALGEBRA)

↳ DISCRETE

◦ ALIGNMENT

↳ SEQ2SEQ, CTC, PROBABILISTIC

◦ SHARING OF LINGUISTIC KNOWLEDGE

↳ ADDITIONAL CONSTRAINTS

↳ MUTUAL INFORMATION / INF. BOTTLENECK

QUESTION III: INFERENCE

- FULLY GENERATIVE

 - ↳ MCMC, VI, MESSAGE PASSING...

- CONVERT TO DISCRIMINATIVE?

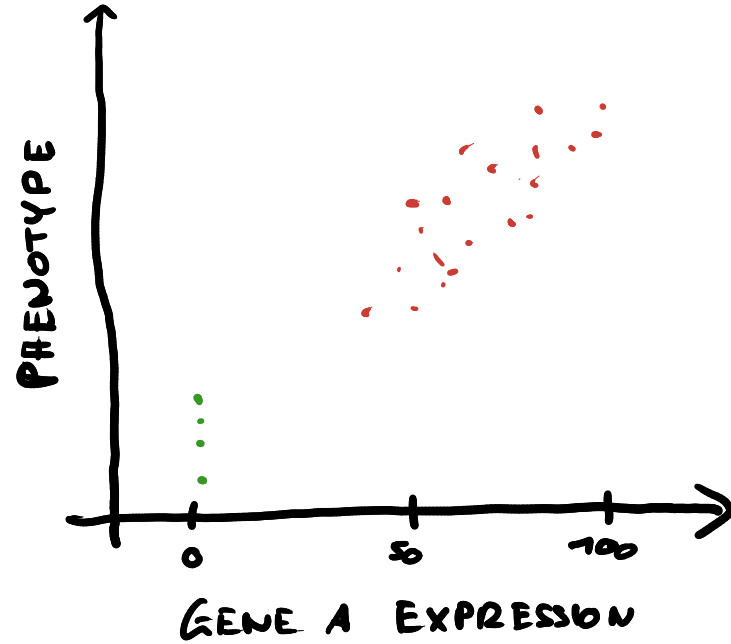
- STRICT-SENSE LM = MARGINAL?

FURTHER CONSIDERATIONS

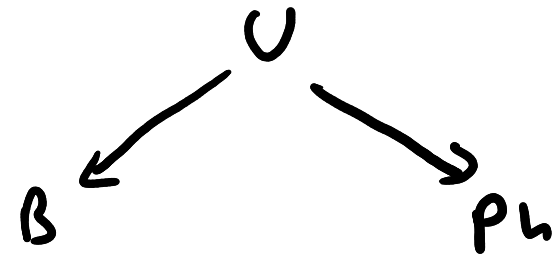
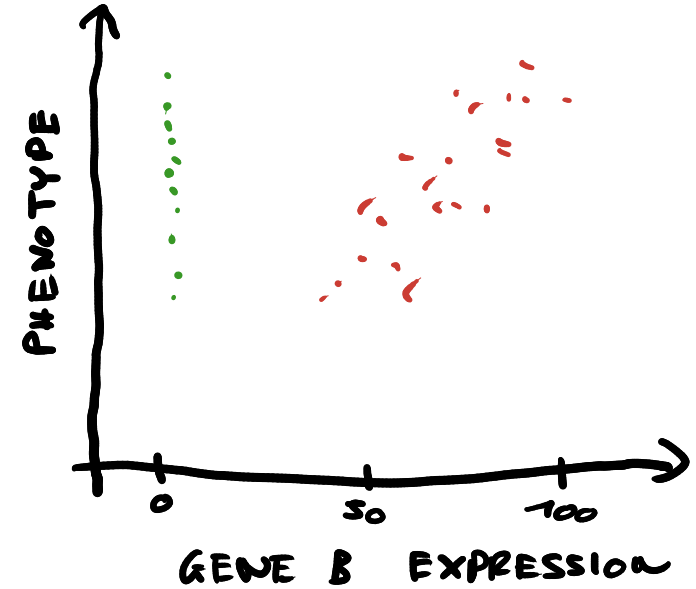
- LOW RESOURCES
- TRANSFER LEARNING
- MODEL / INVESTIGATE ANNOTATOR ERRORS?
- FIRST TRY ON:
 - ↳ MOVIE DIALOGS
 - ↳ CHAT CORPORA ?
 - ↳ AUTO-SUBTITLES

APPENDIX

TWO GENES



$A \rightarrow Ph$



CLASSIFY WHAT'S WRONG

- POPULATION SHIFT
- ACQUISITION SHIFT
- ANNOTATION SHIFT
- PREVALENCE SHIFT
- MANIFESTATION SHIFT

