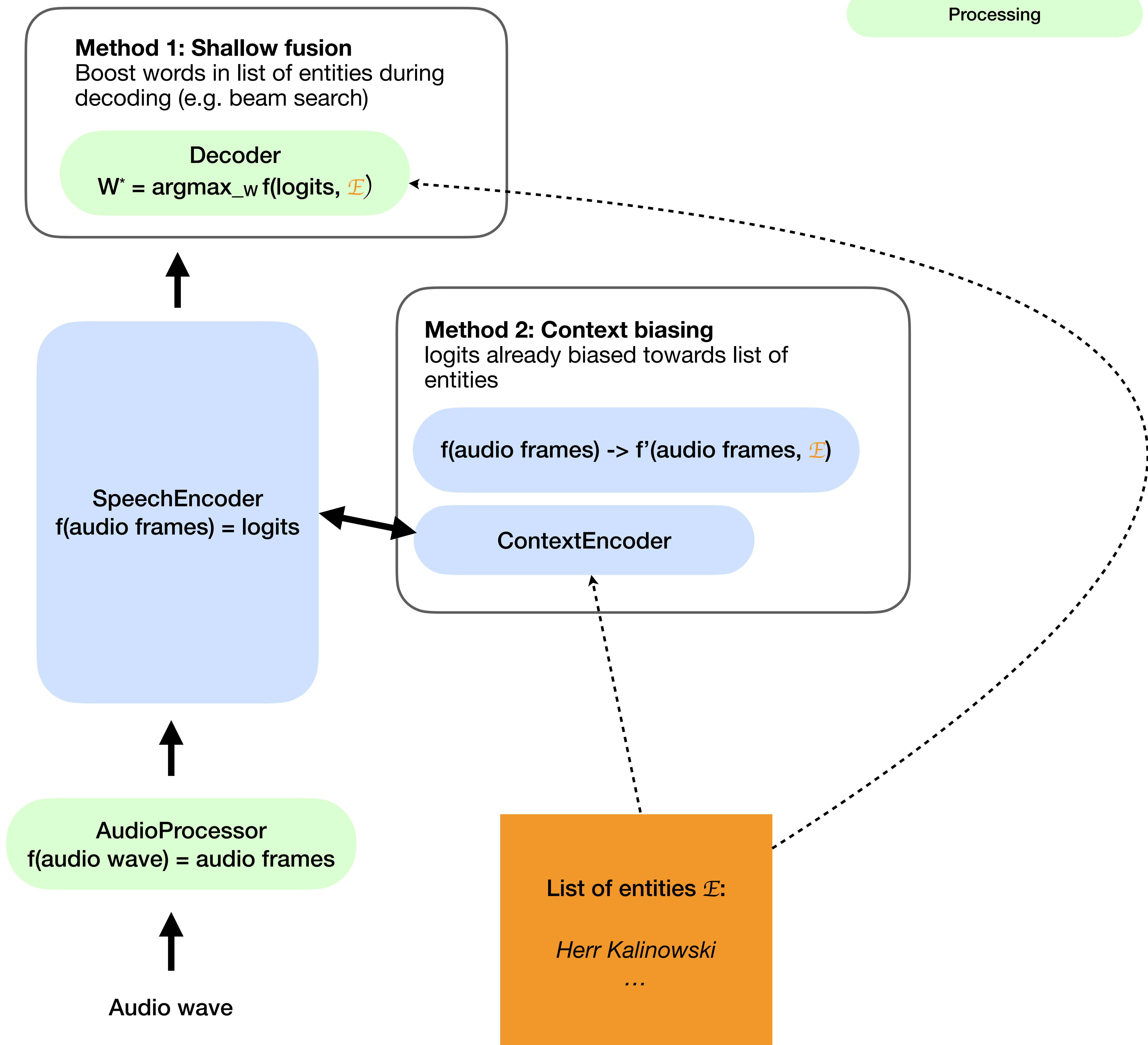


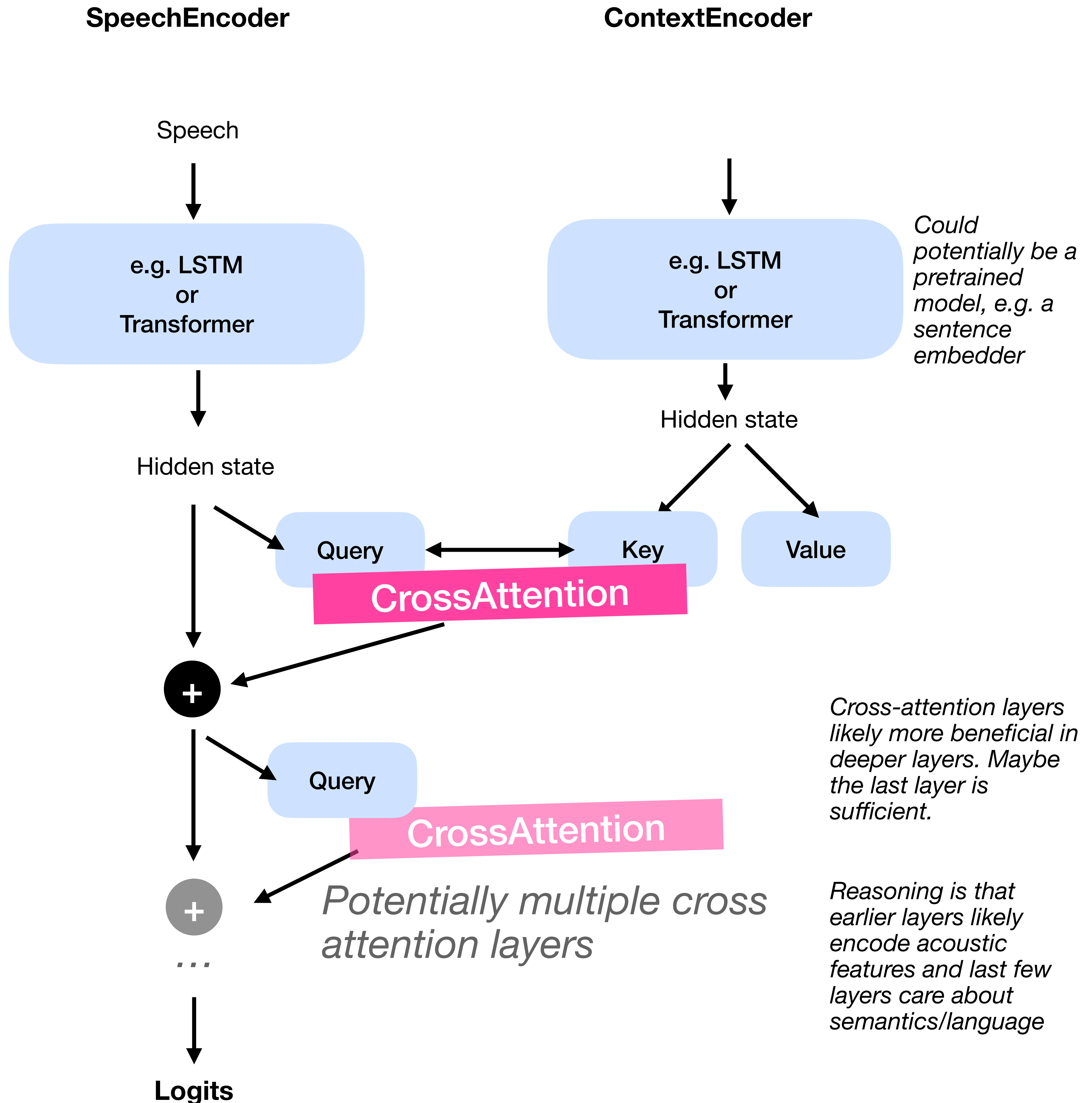
Assuming a system trained E2E audio -> words
w/o using separate LM model

Legend

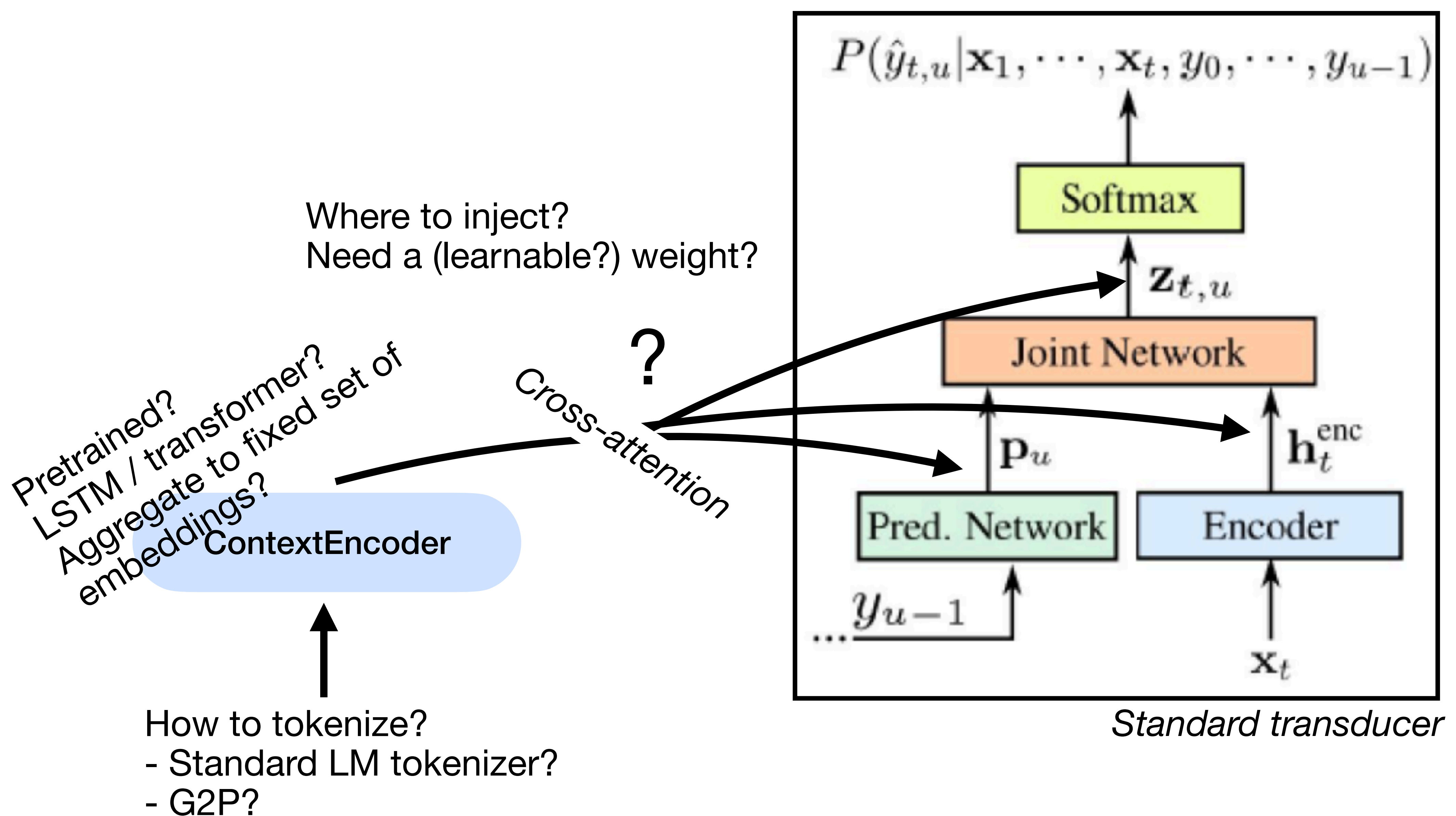
- Neural component
- Processing



Method 2: Context biasing



Transducer with deep context biasing - Design questions



List of entities \mathcal{E} :

Herr Kalinowski

...

Training dynamics/data:

Likely needed to start with pretrained transducer + context encoder than starting both from scratch.

Needs extensive training data to cover various types of contexts