We propose a general class of language models that treat reference as an explicit stochastic latent variable. This architecture allows models to create mentions of entities and their attributes by accessing external databases (required by, e.g., dialogue generation and recipe generation) and internal state (required by, e.g. language models which are aware of coreference). This facilitates the incorporation of information that can be accessed in predictable locations in databases or discourse context, even when the targets of the reference may be rare words. Experiments on three tasks shows our model variants based on deterministic attention.

https://arxiv.org/abs/1611.01628