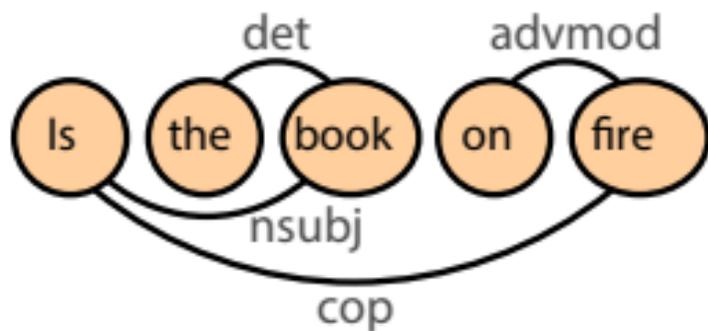
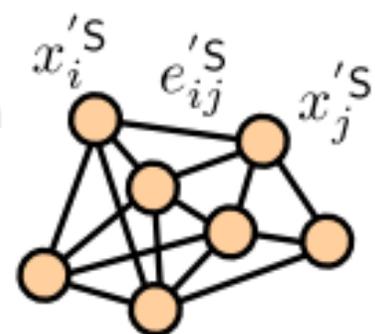


Input scene description and parsed question

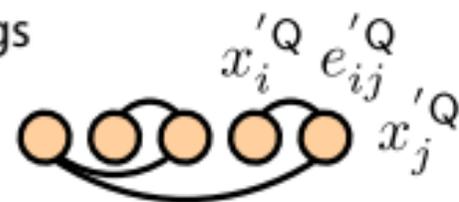


Initial embedding

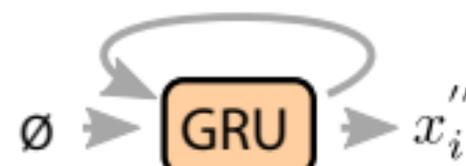
Affine projection



Word/vector embeddings



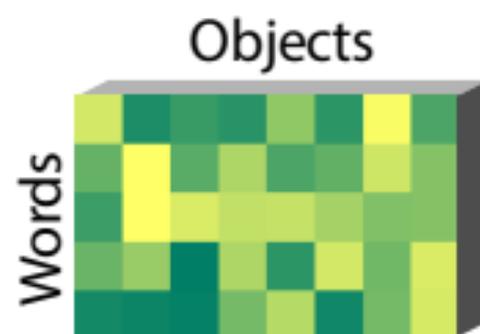
Graph processing



$\begin{bmatrix} x'_i & \text{Node} \\ n_i & \text{Neighbours} \end{bmatrix}$

Cosine similarity

Combined features



Weighted sum



Sigmoid or softmax



yes	0.9
no	0.0