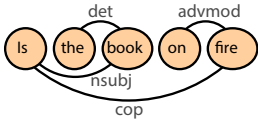
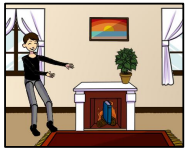
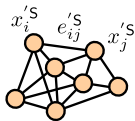


Input scene description  
and parsed question

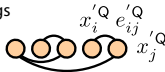


Initial  
embedding

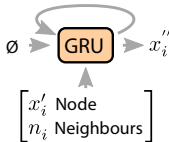
Affine  
projection



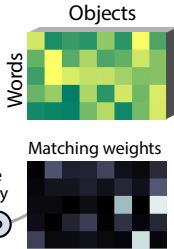
Word/vector  
embeddings



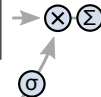
Graph  
processing



Combined  
features



Weighted  
sum



Prediction over  
candidate answers

Sigmoid or  
softmax

...

$\sigma$	
yes	0.9
no	0.0