



ordinal correlation between human ratings and models'  
sentence pair probability log-ratio (signed-rank cosine similarity)