

If Bleu's correlation with human judgments has been overestimated, then the field needs to ask itself whether it should continue to be driven by