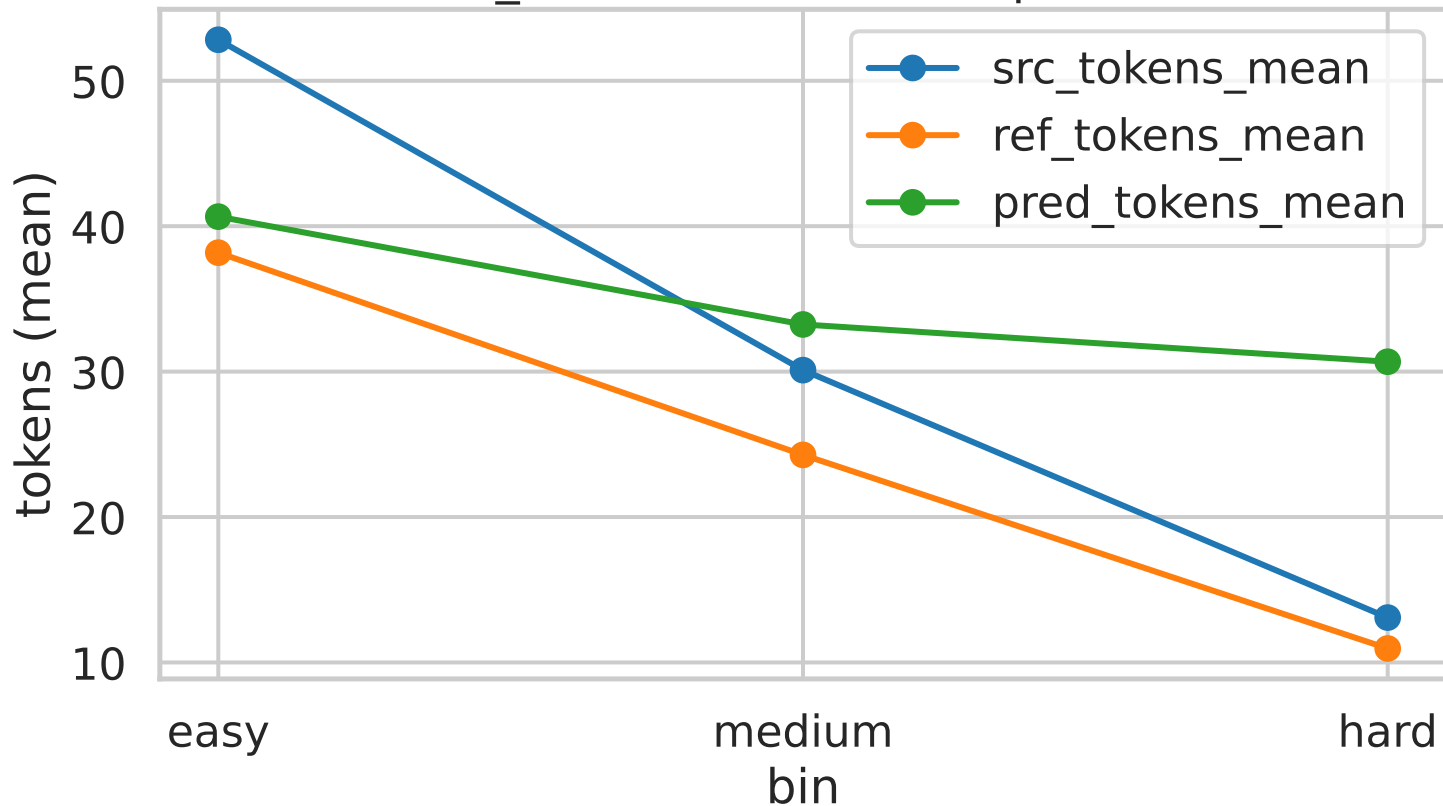
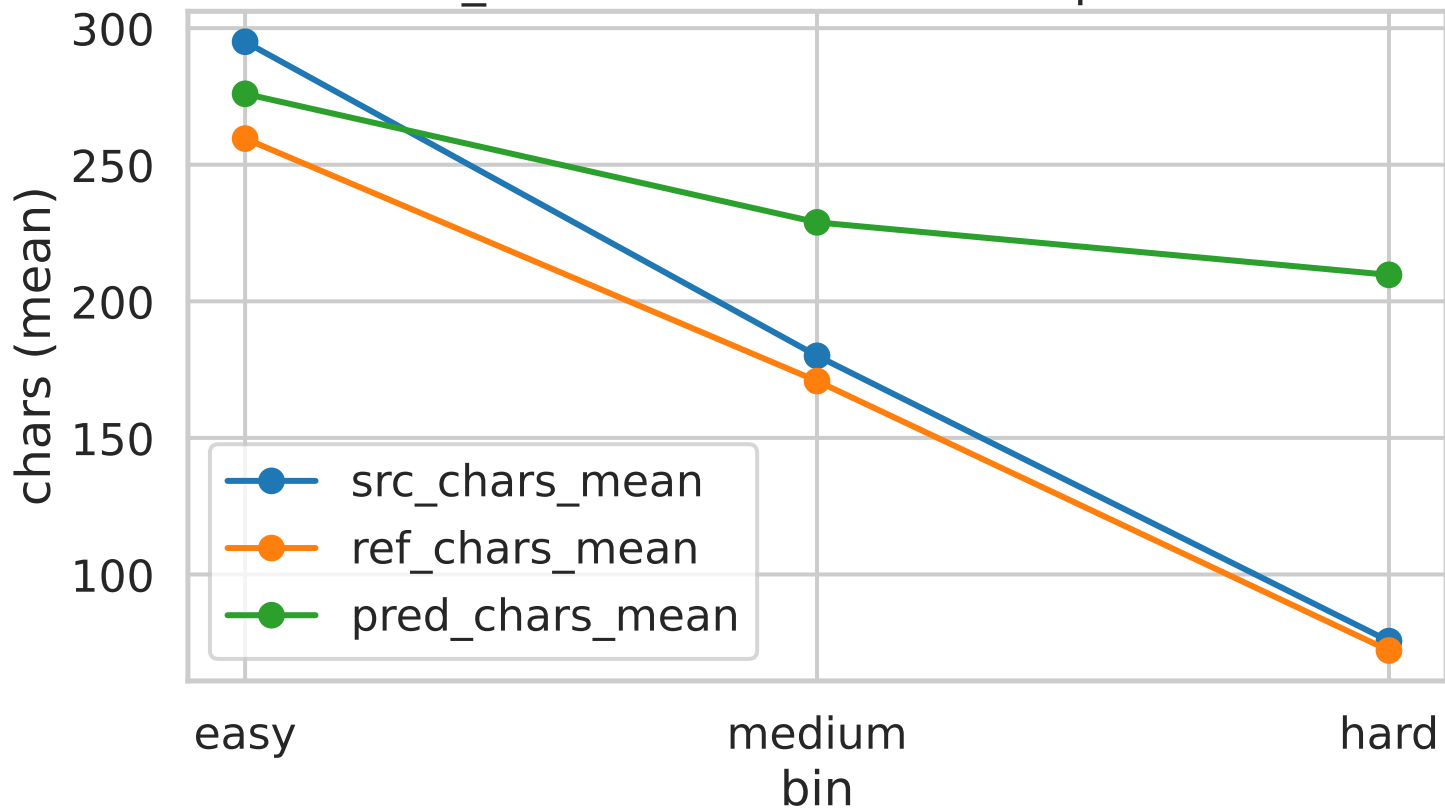


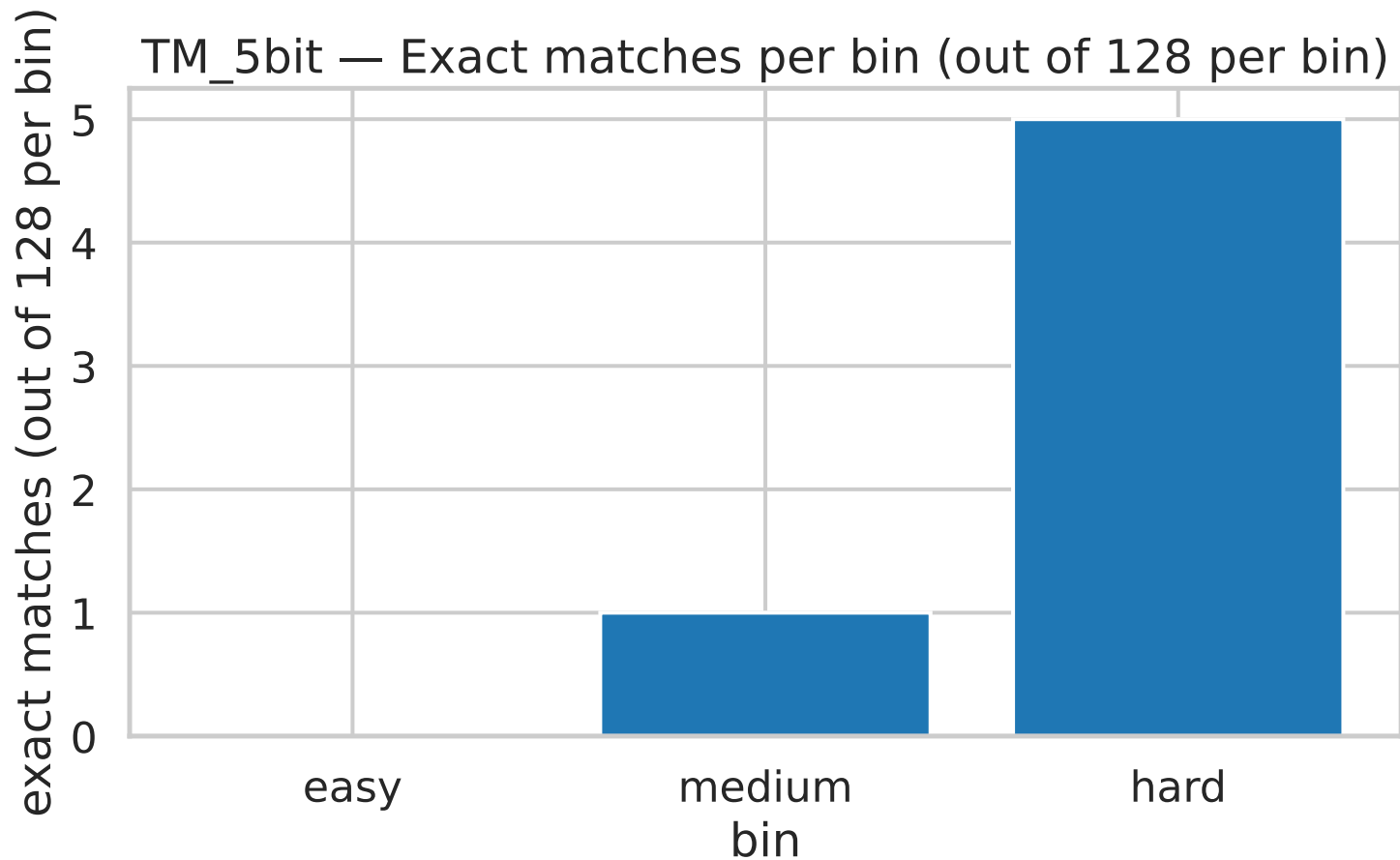
TM\_5bit — Mean tokens per bin



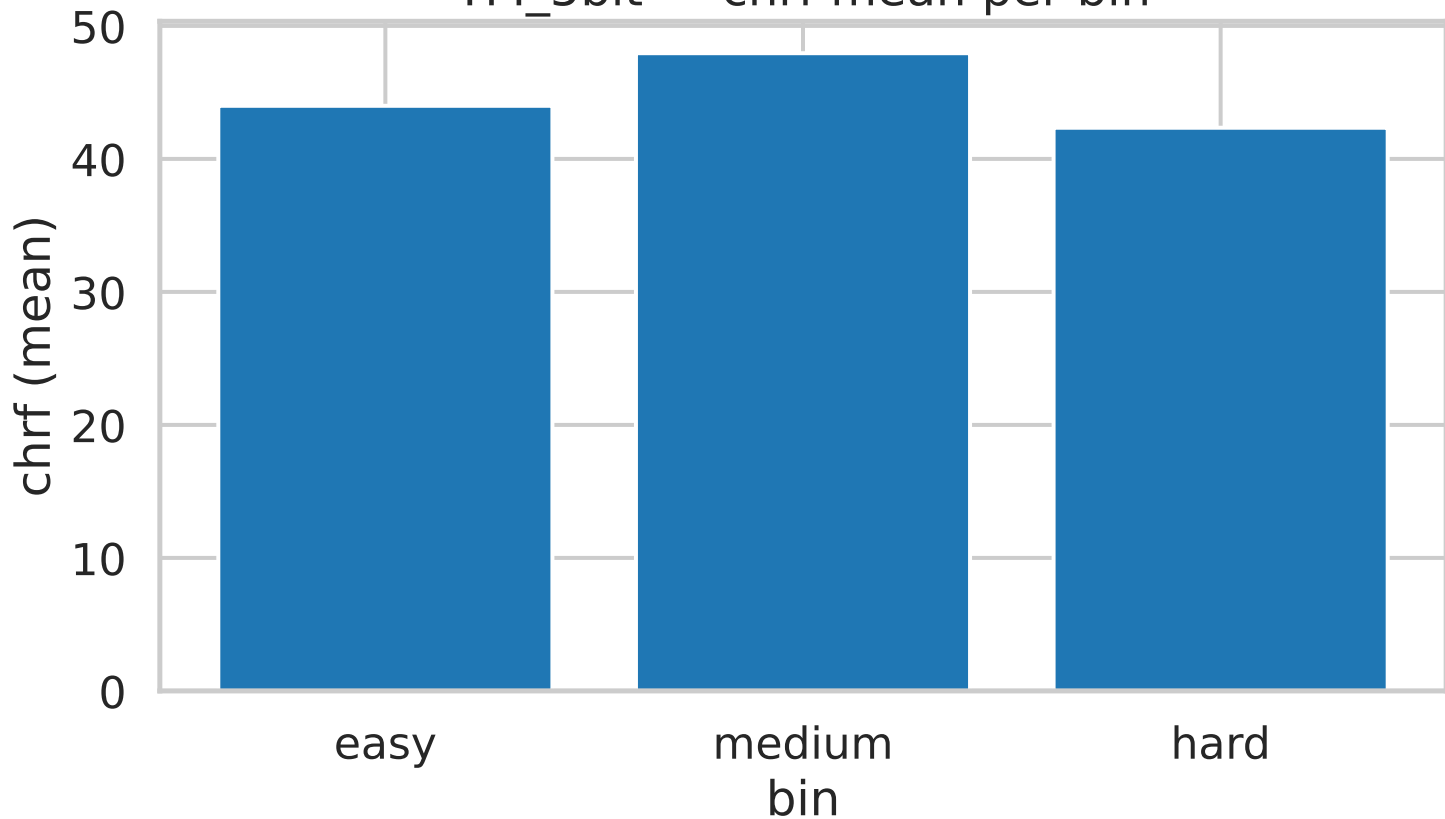
TM\_5bit — Mean characters per bin



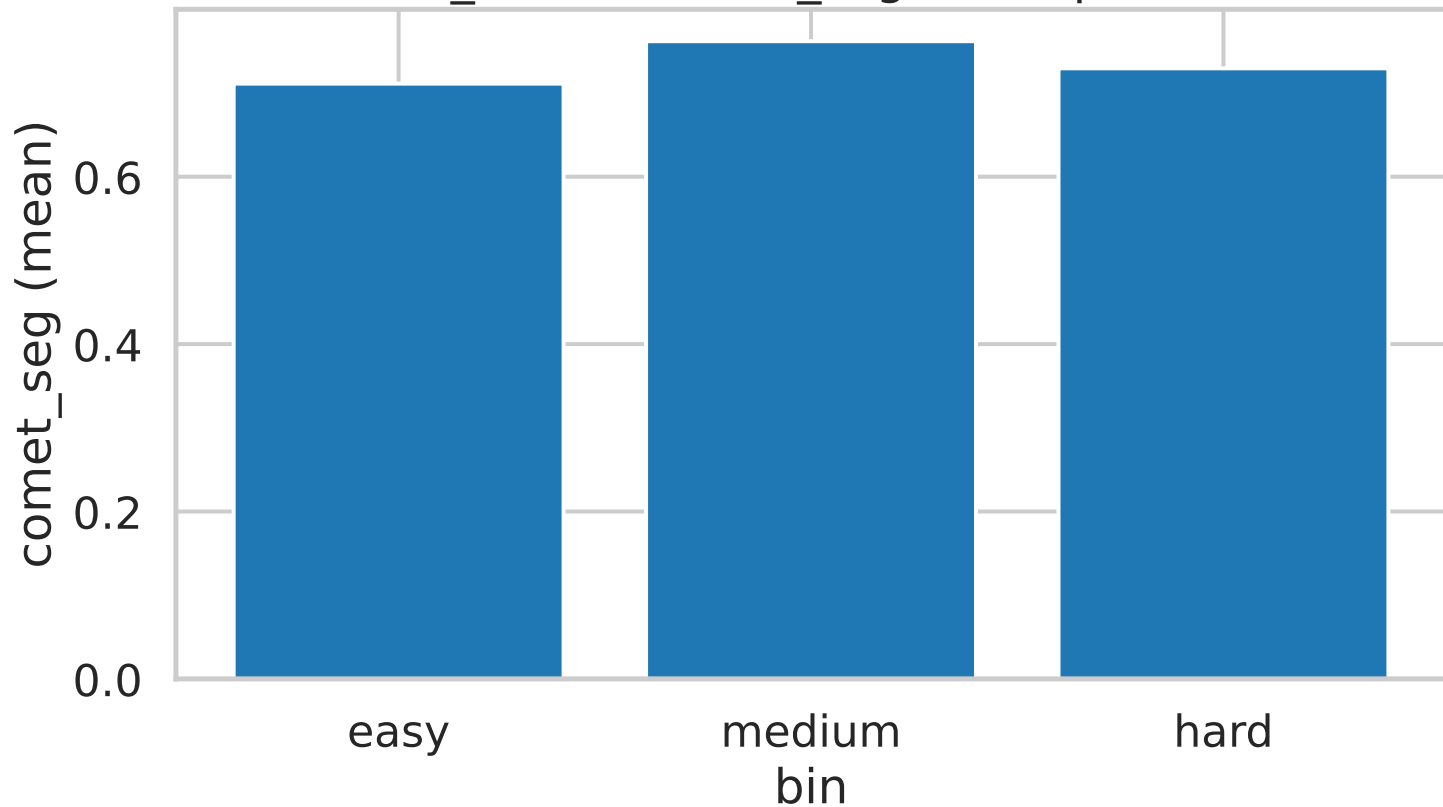
TM\_5bit — Exact matches per bin (out of 128 per bin)



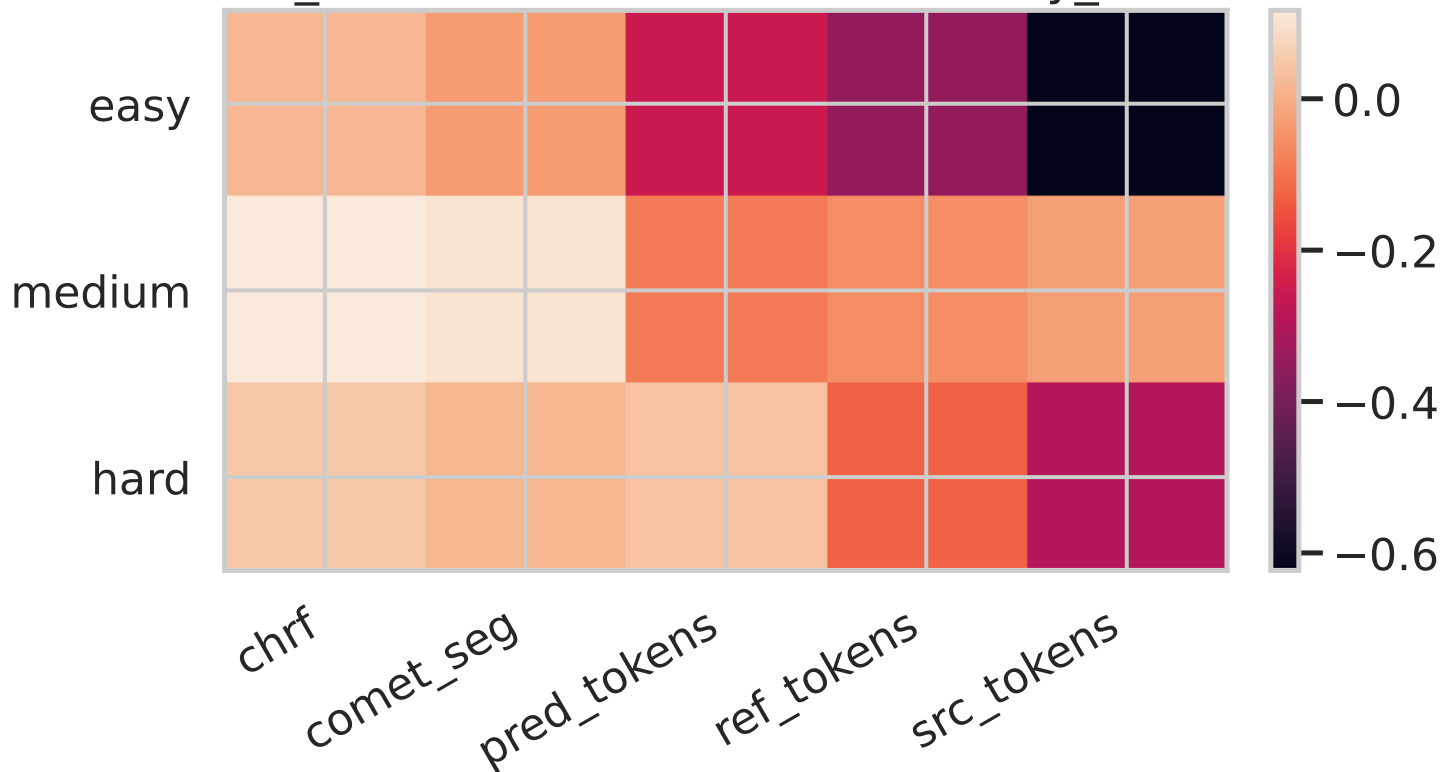
TM\_5bit — chrF mean per bin



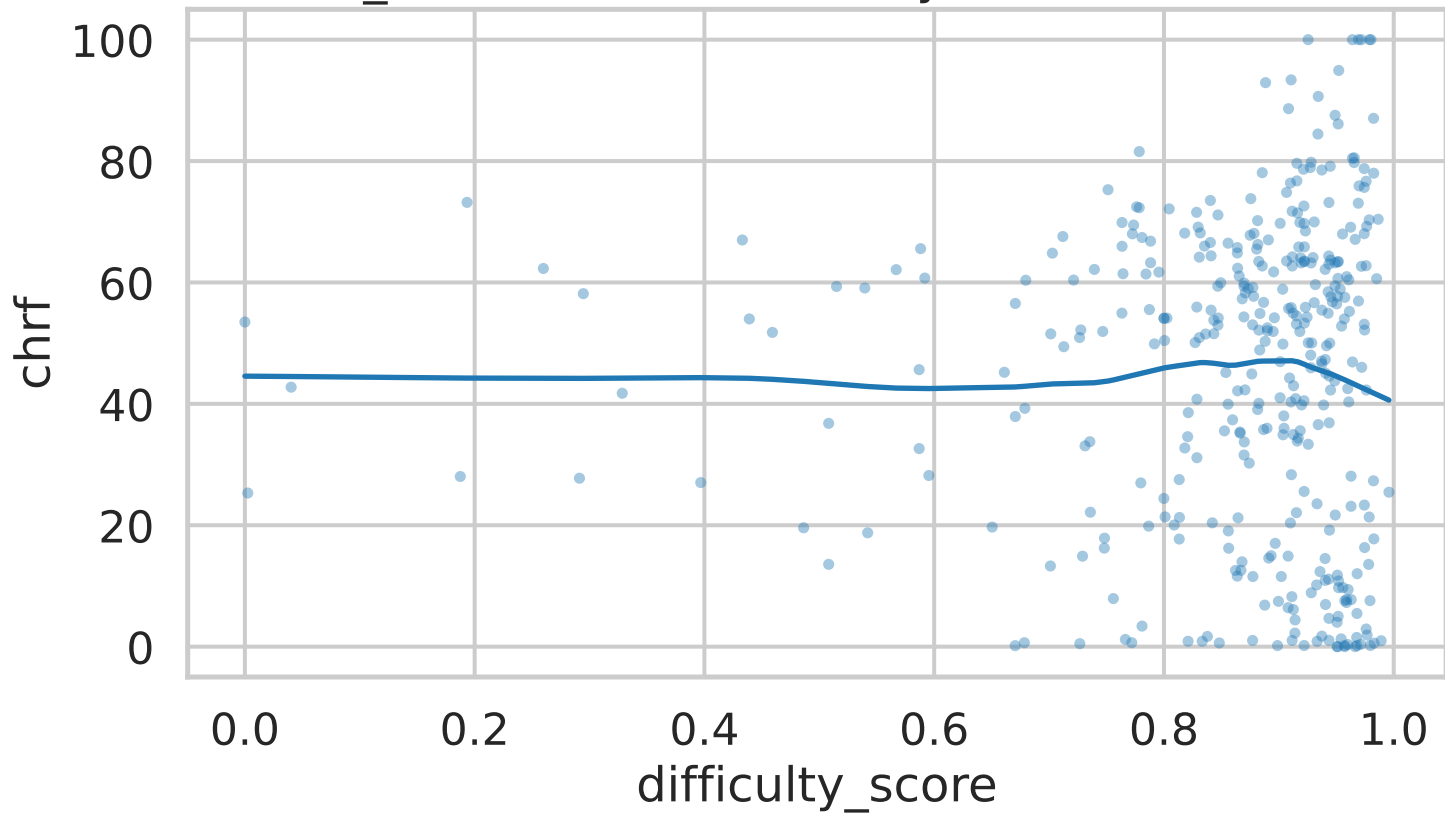
TM\_5bit — comet\_seg mean per bin



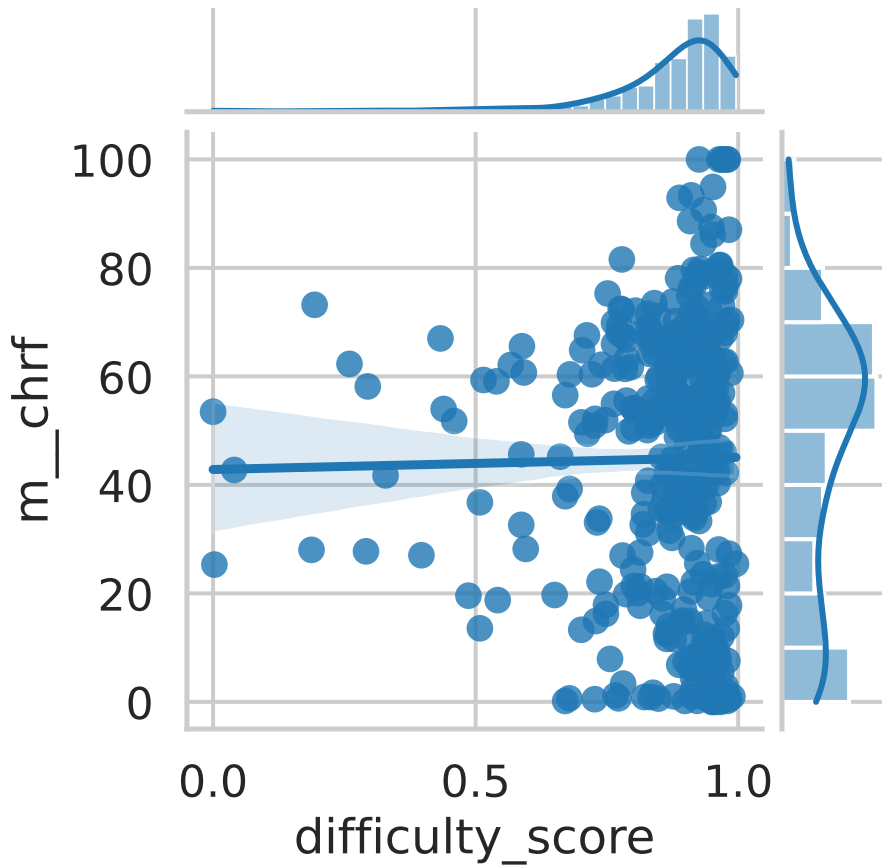
TM\_5bit — Correlations w/ difficulty\_score



TM\_5bit — chrF vs difficulty (LOWESS=True)

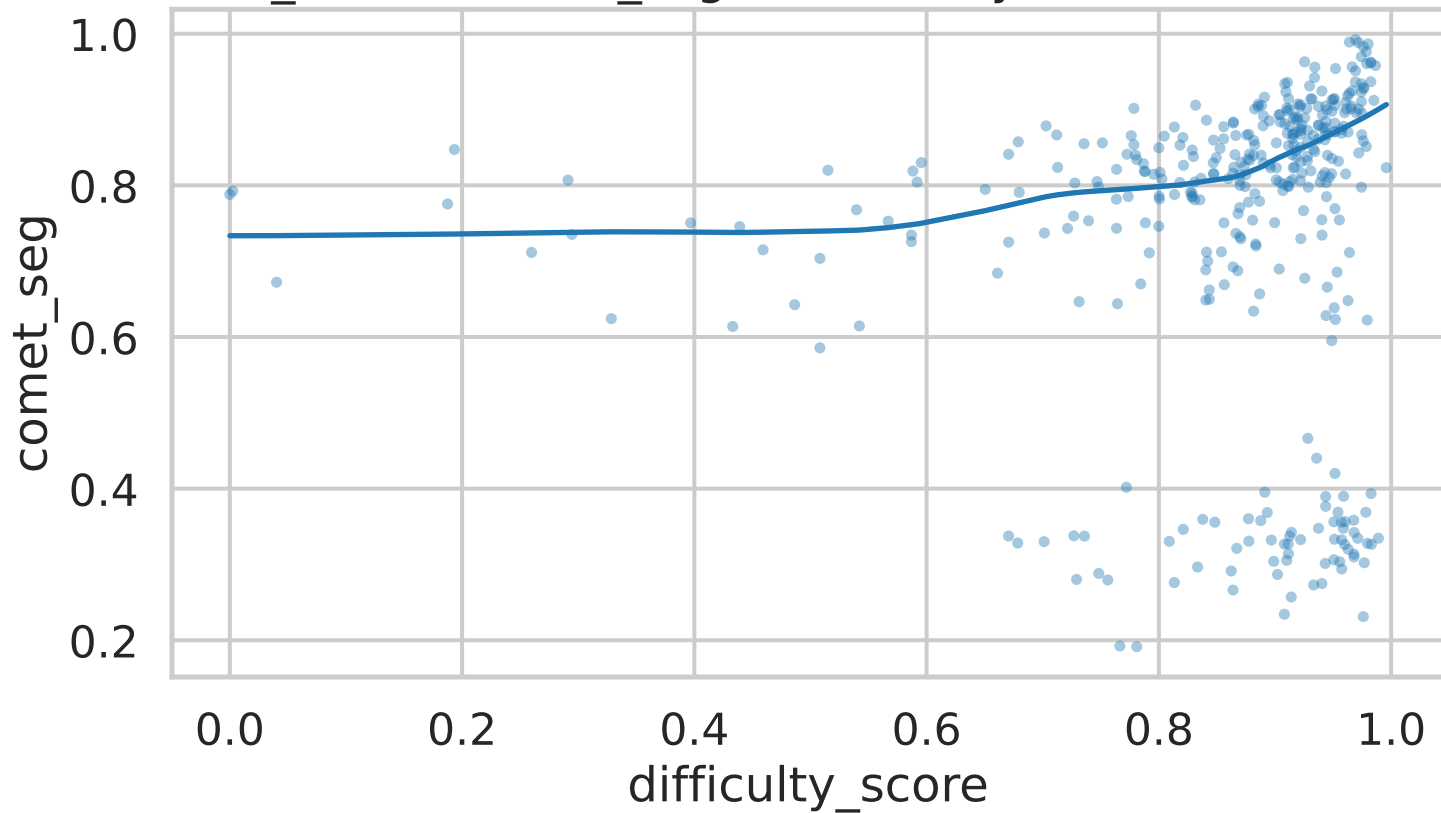


m\_chrf = joint chrf vs difficulty\_score

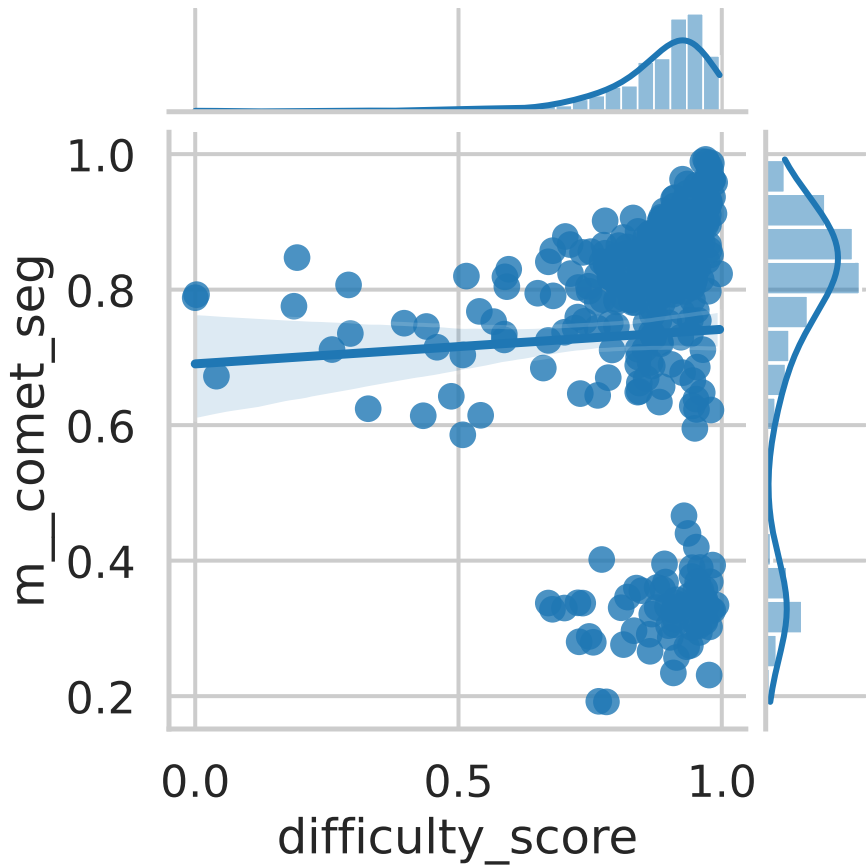




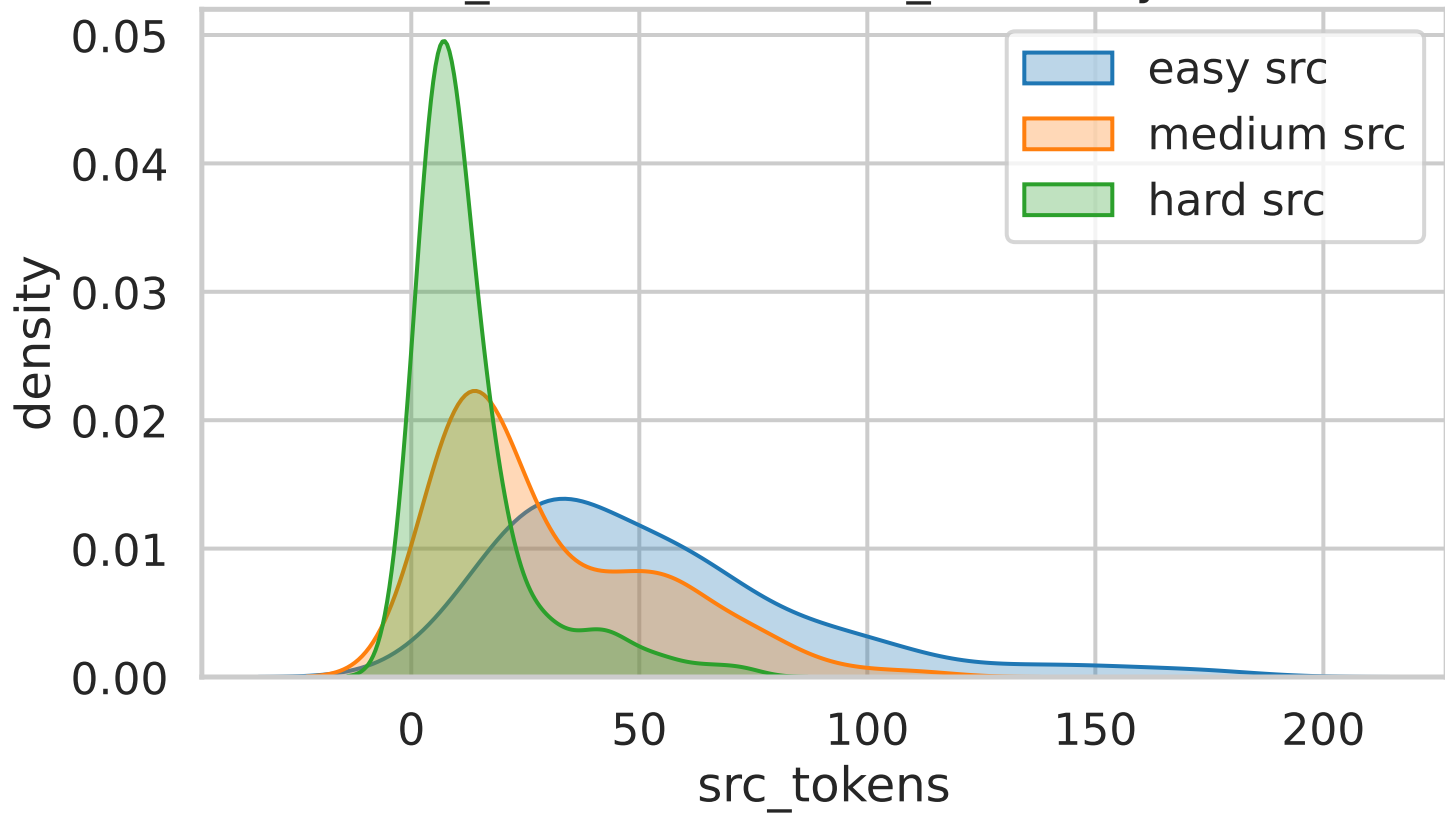
TM\_5bit — comet\_seg vs difficulty (LOWESS=True)



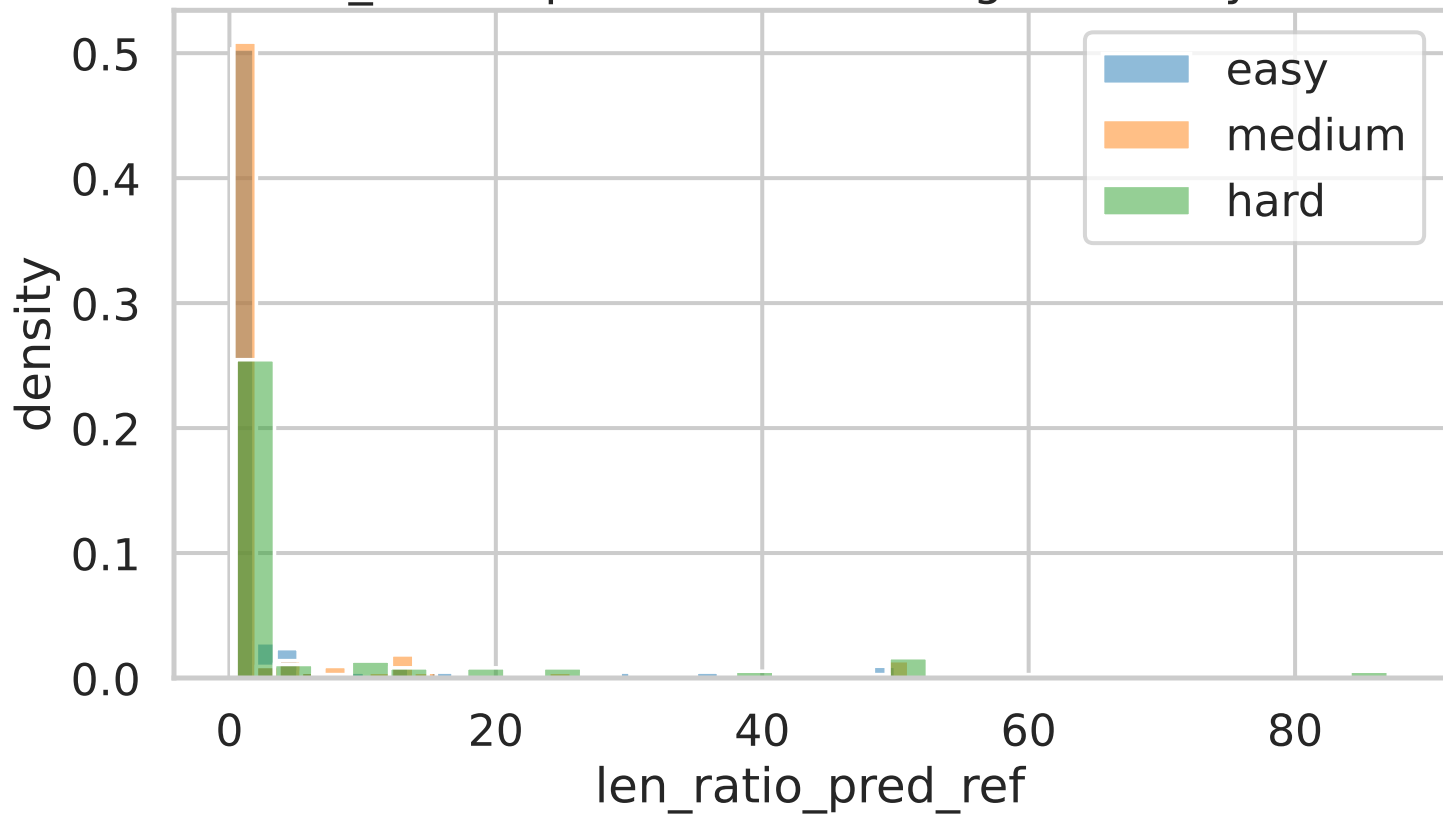
bit = joint\_comet\_seg vs difficulty\_score



TM\_5bit — KDE of src\_tokens by bin



TM\_5bit — pred/ref token length ratio by bin



TM\_5bit — Number consistency (src→pred) per bin

