Frequency of "correct" label in each position generated by GPT