Syntax Error Type Distribution Across Models
(Breakdown shows different failure patterns from pretraining)