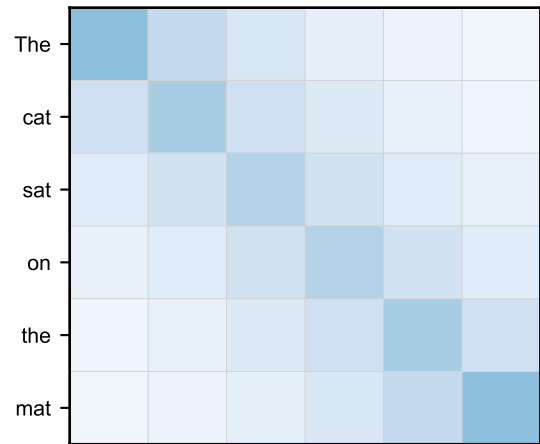
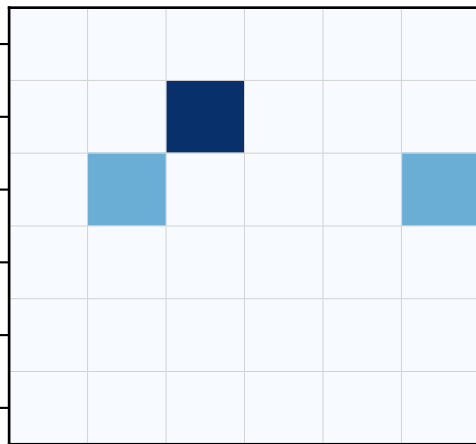


# Multi-Head Attention: Different Heads Learn Different Patterns

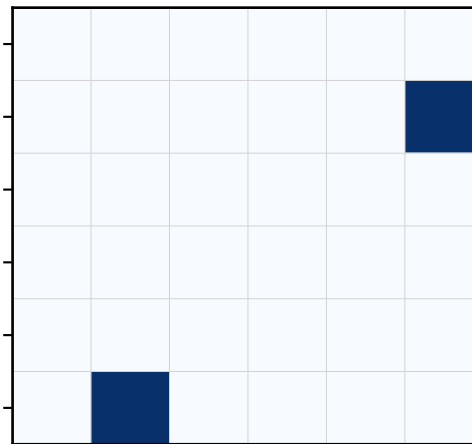
Head 1: Position



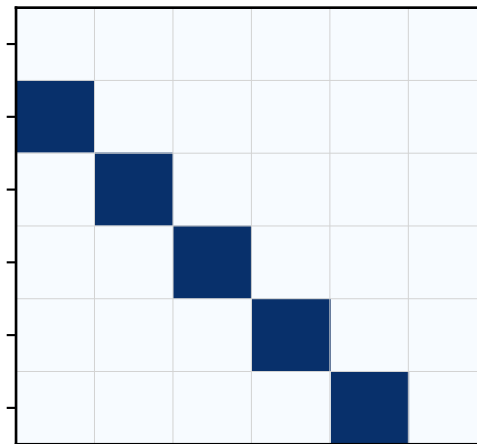
Head 2: Syntax



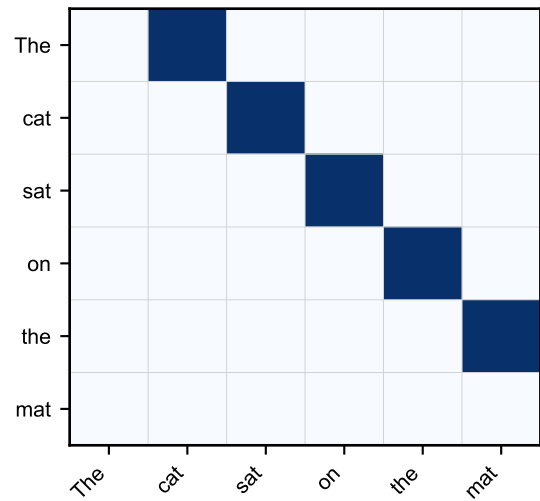
Head 3: Semantic



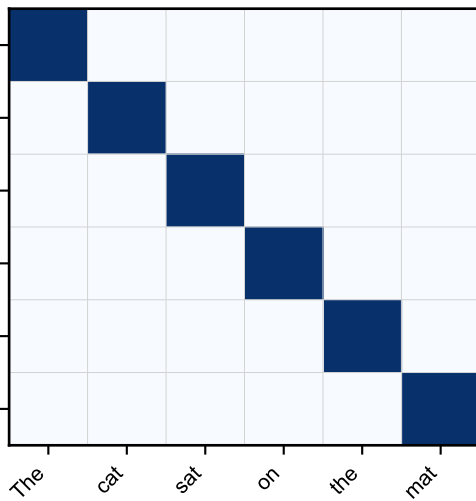
Head 4: Previous



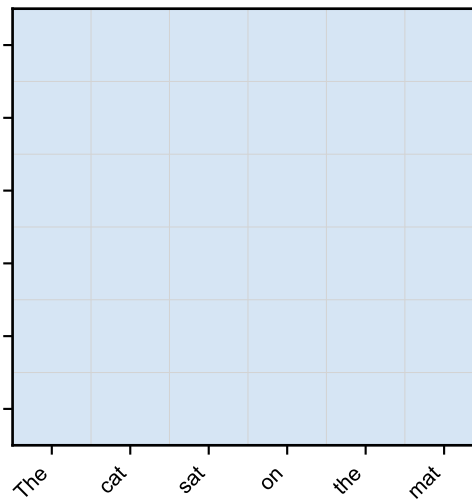
Head 5: Next



Head 6: Self



Head 7: Global



Head 8: Random

