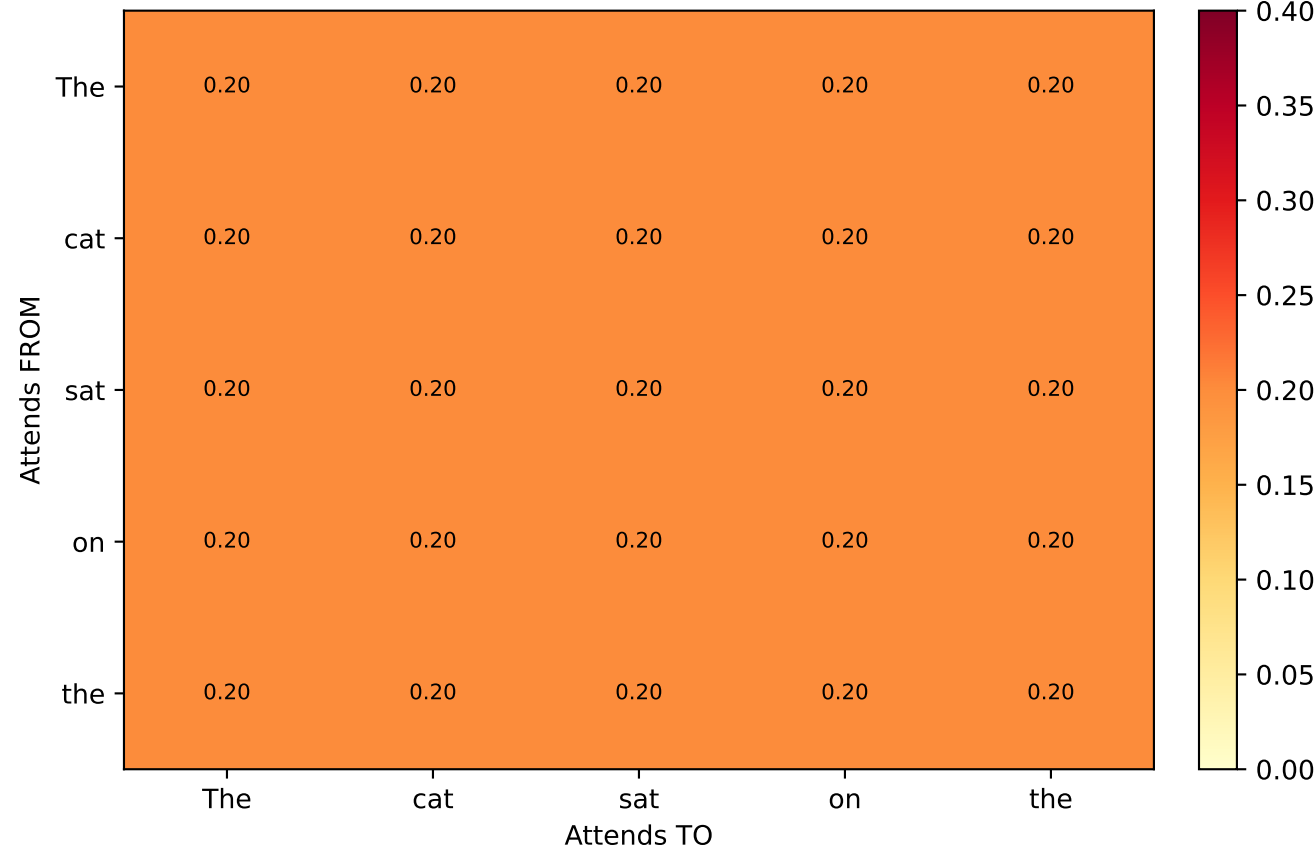
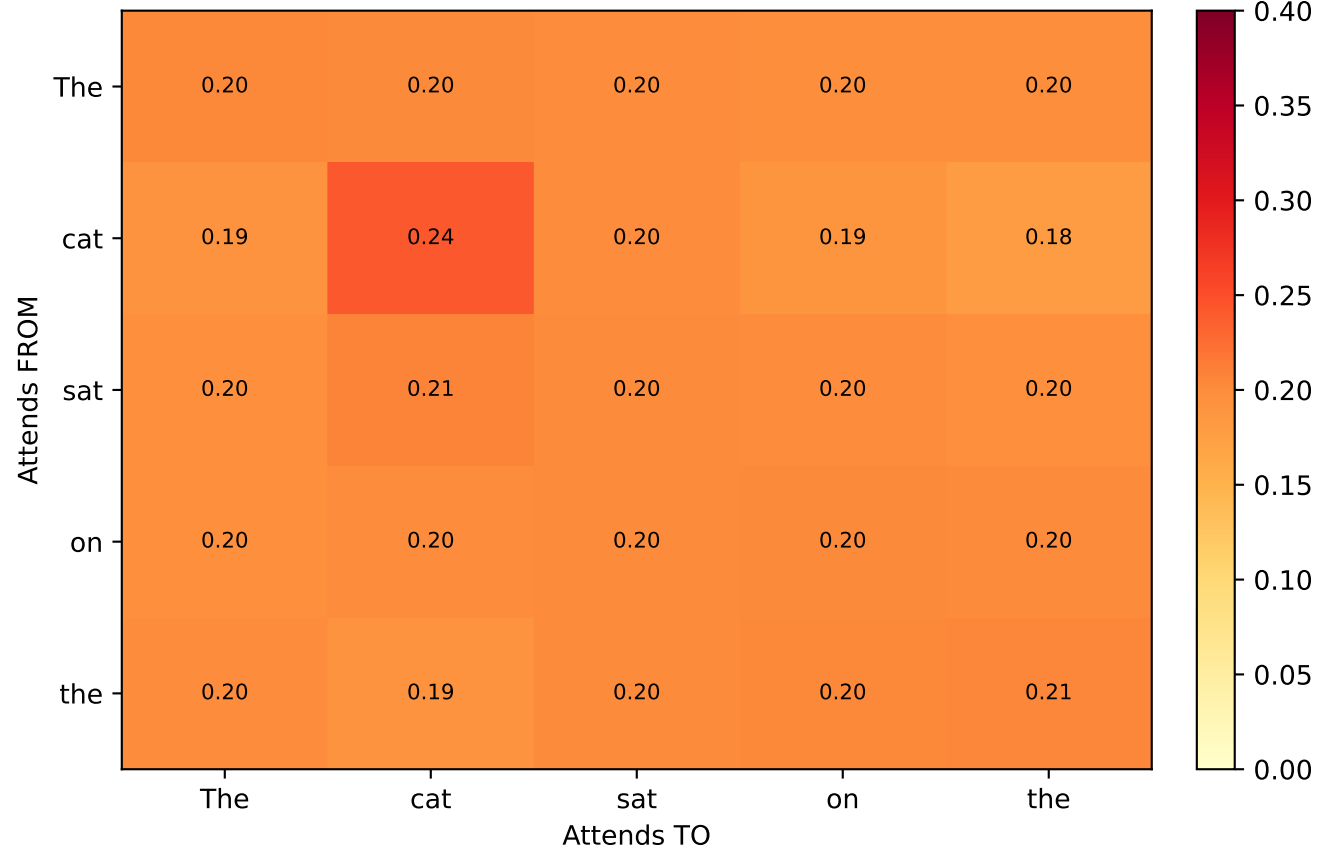


Real Multi-Head Attention Patterns (4 Heads)

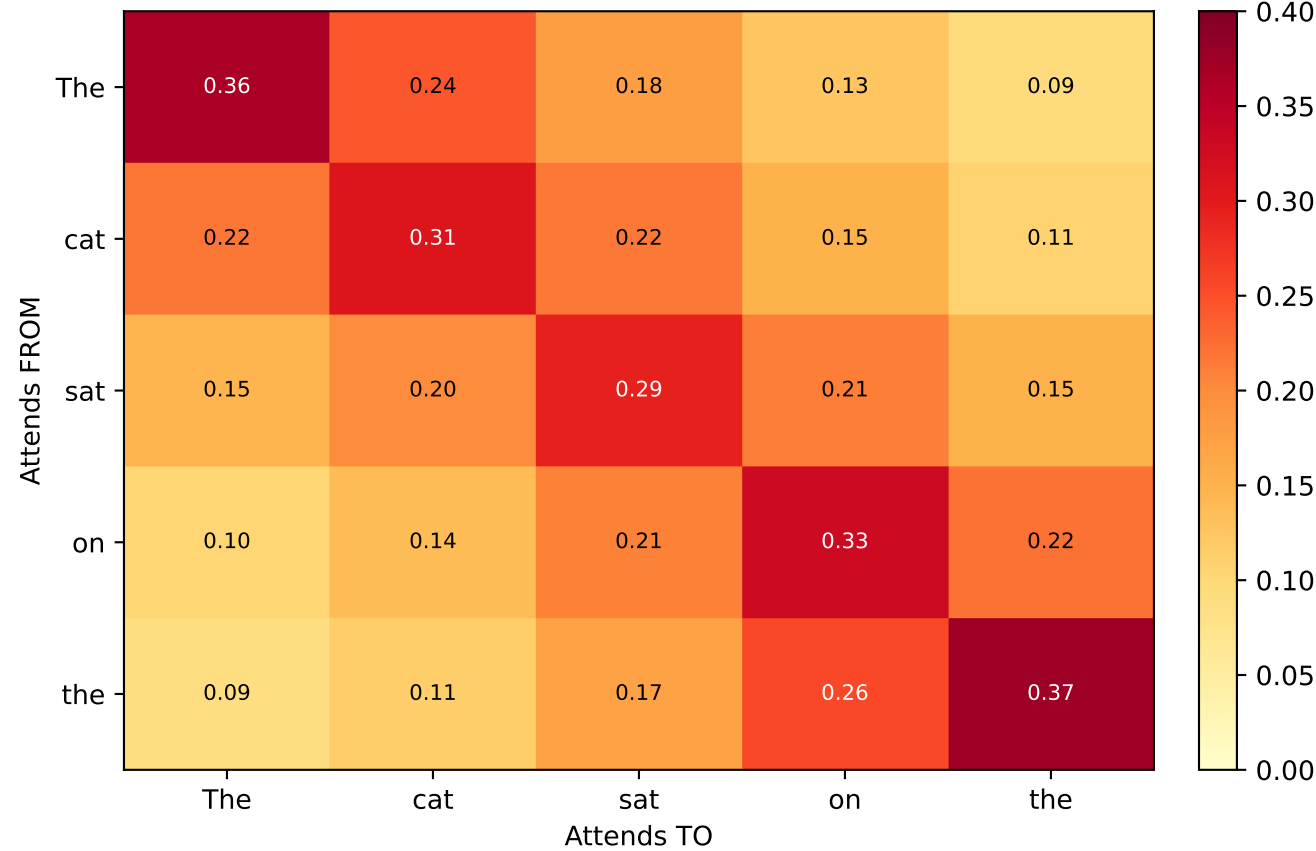
Head 1: Grammar



Head 2: Semantics



Head 3: Position



Head 4: Global

