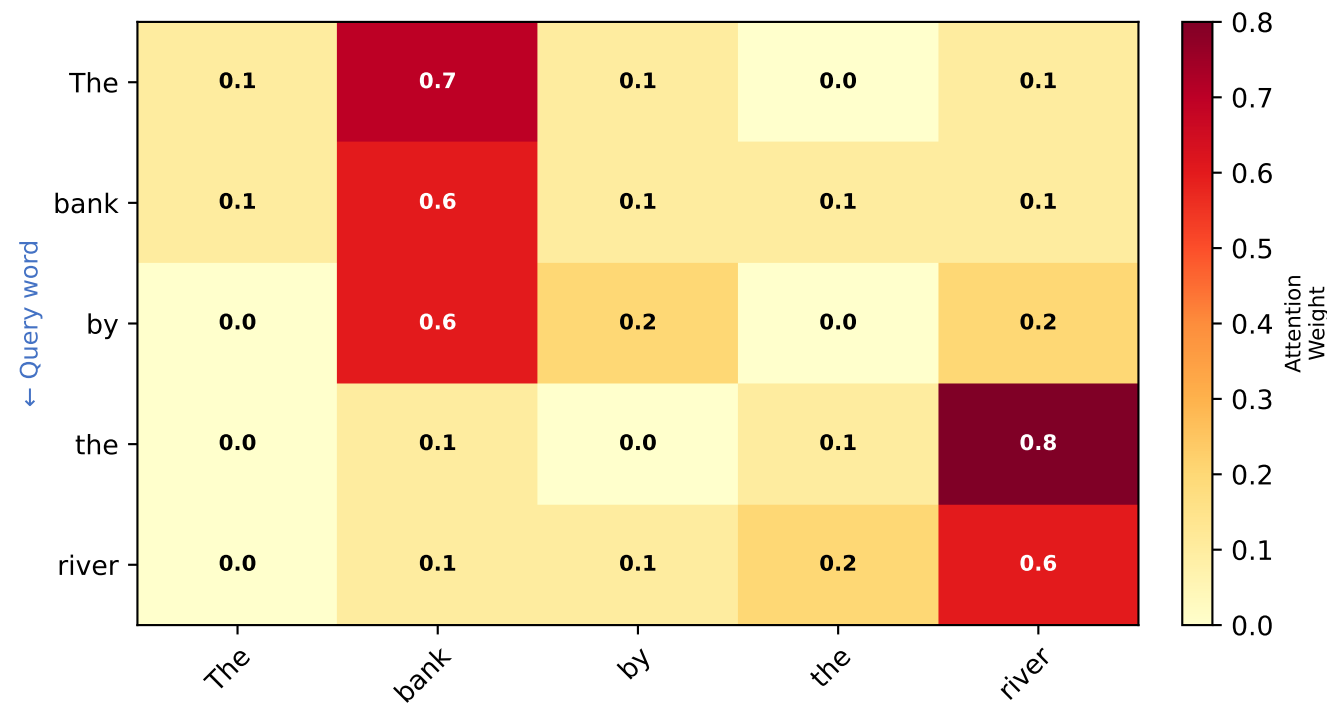
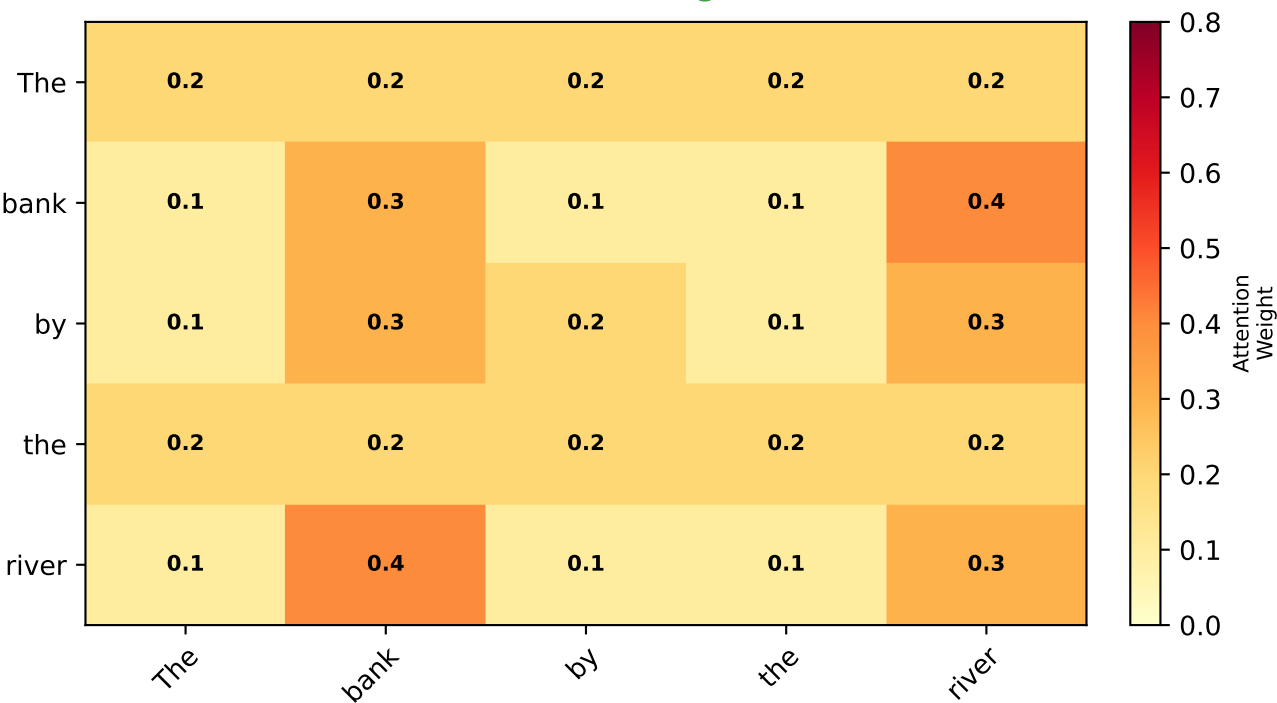


# Multi-Head Attention Patterns: "The bank by the river"

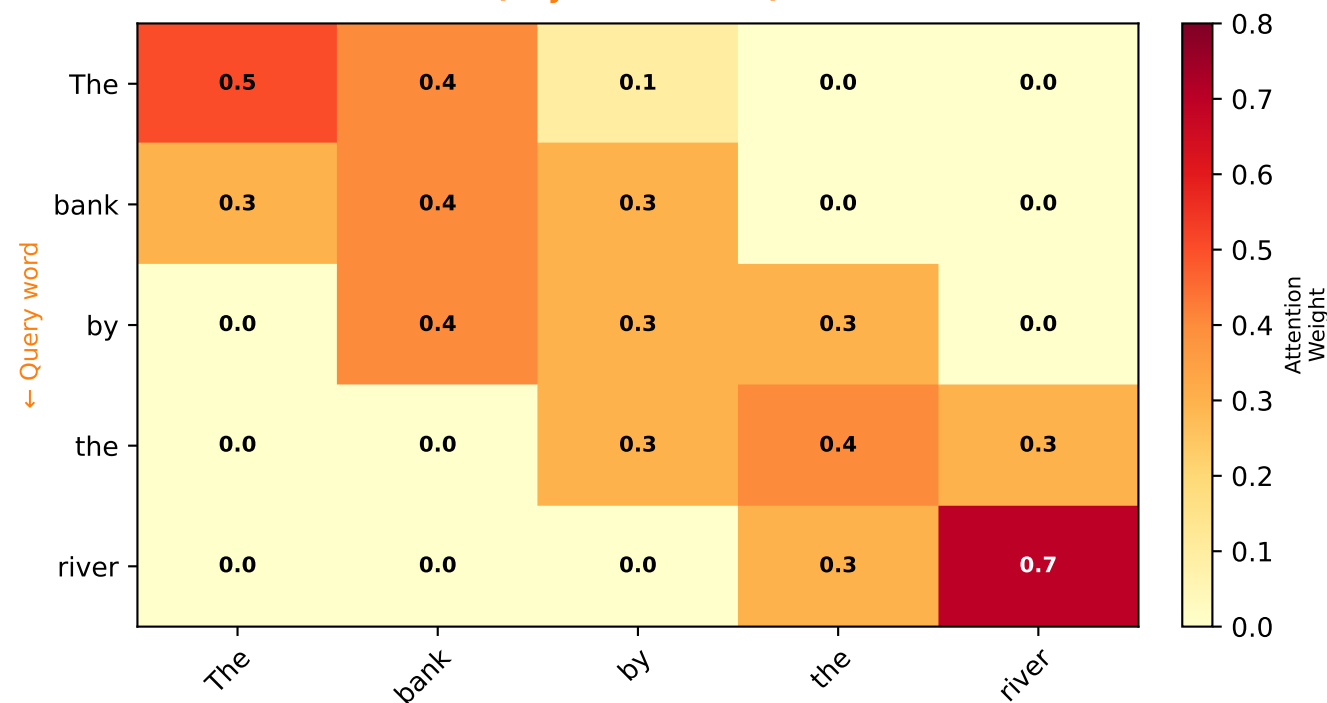
Head 1: Syntactic  
(Articles → Nouns)



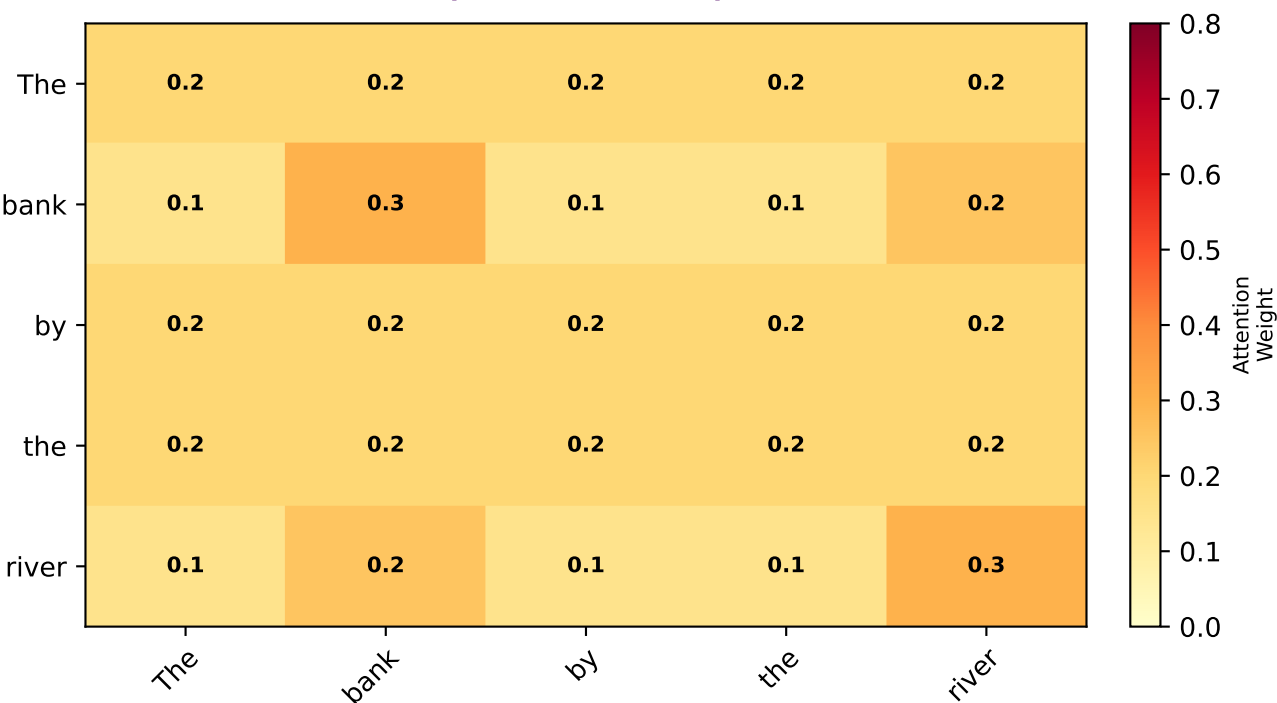
Head 2: Semantic  
(Related Meanings)



Head 3: Positional  
(Adjacent Words)



Head 4: Global  
(Sentence-Level)



Attending to →

Attending to →

Each head captures different relationships: Syntax (grammar), Semantics (meaning), Position (locality), Global (context)