Multi-Head Attention Patterns: "The bank by the river"
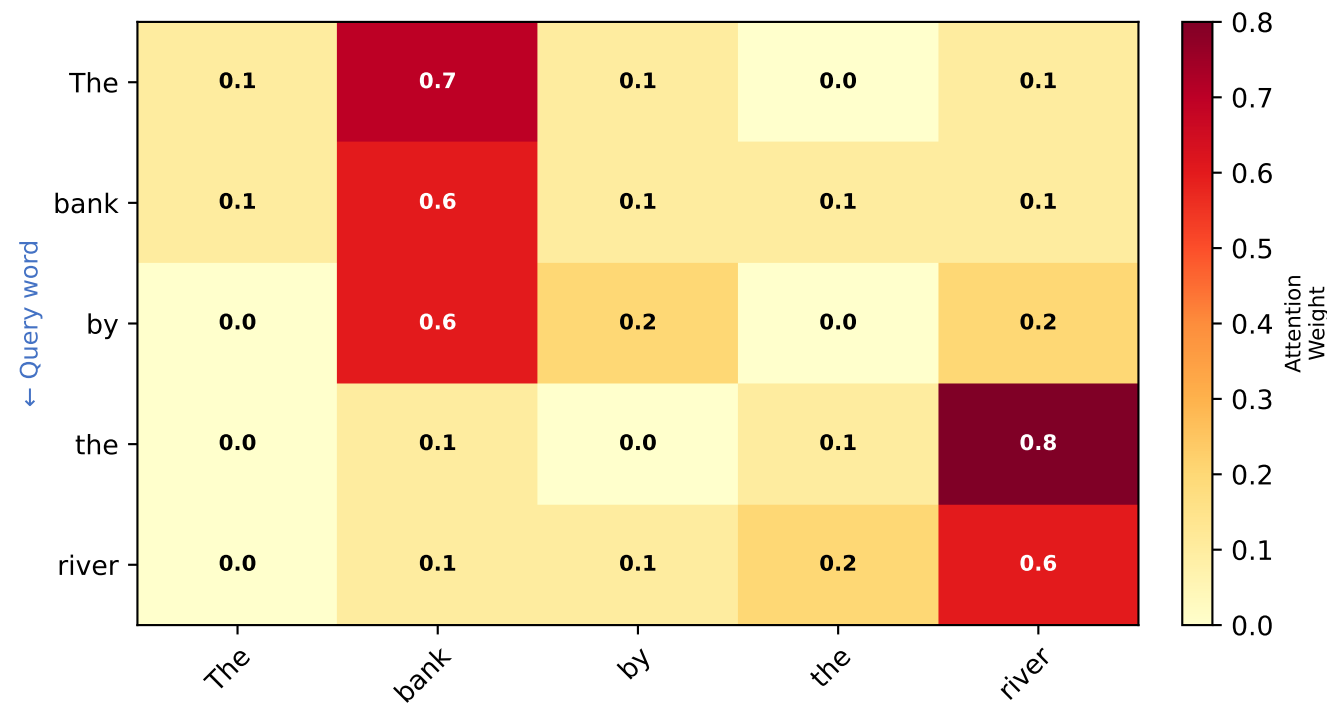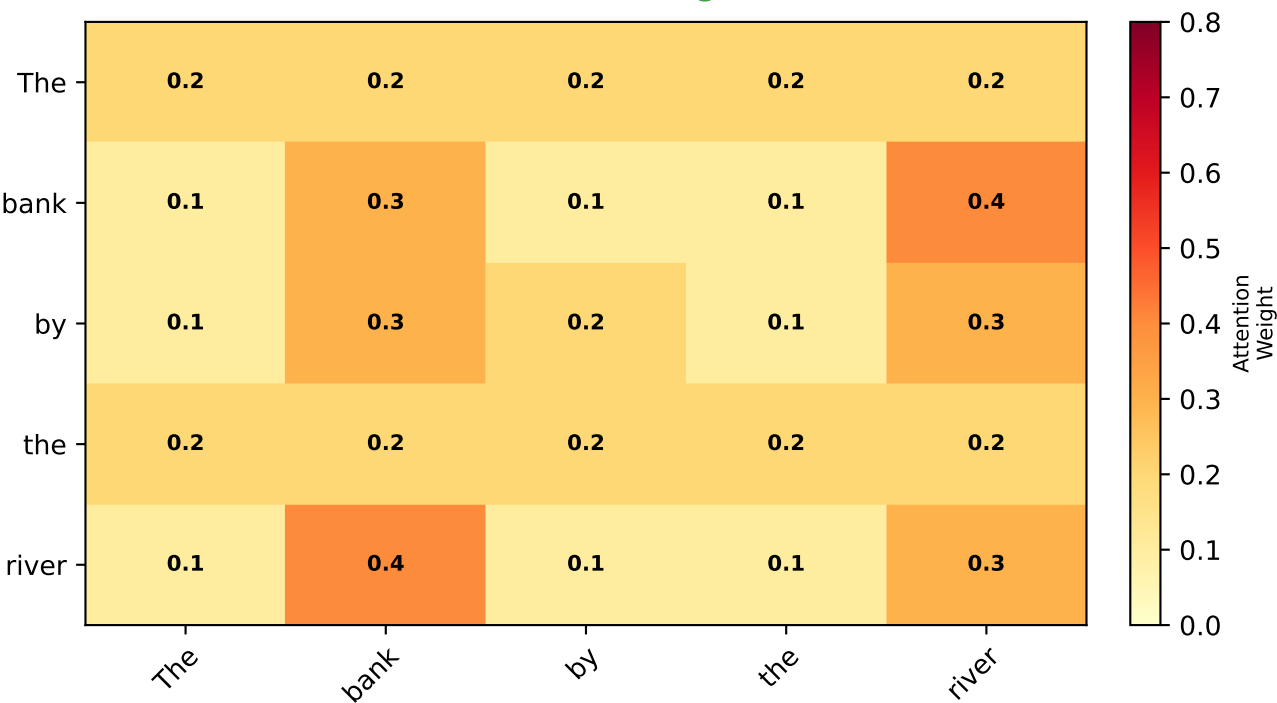
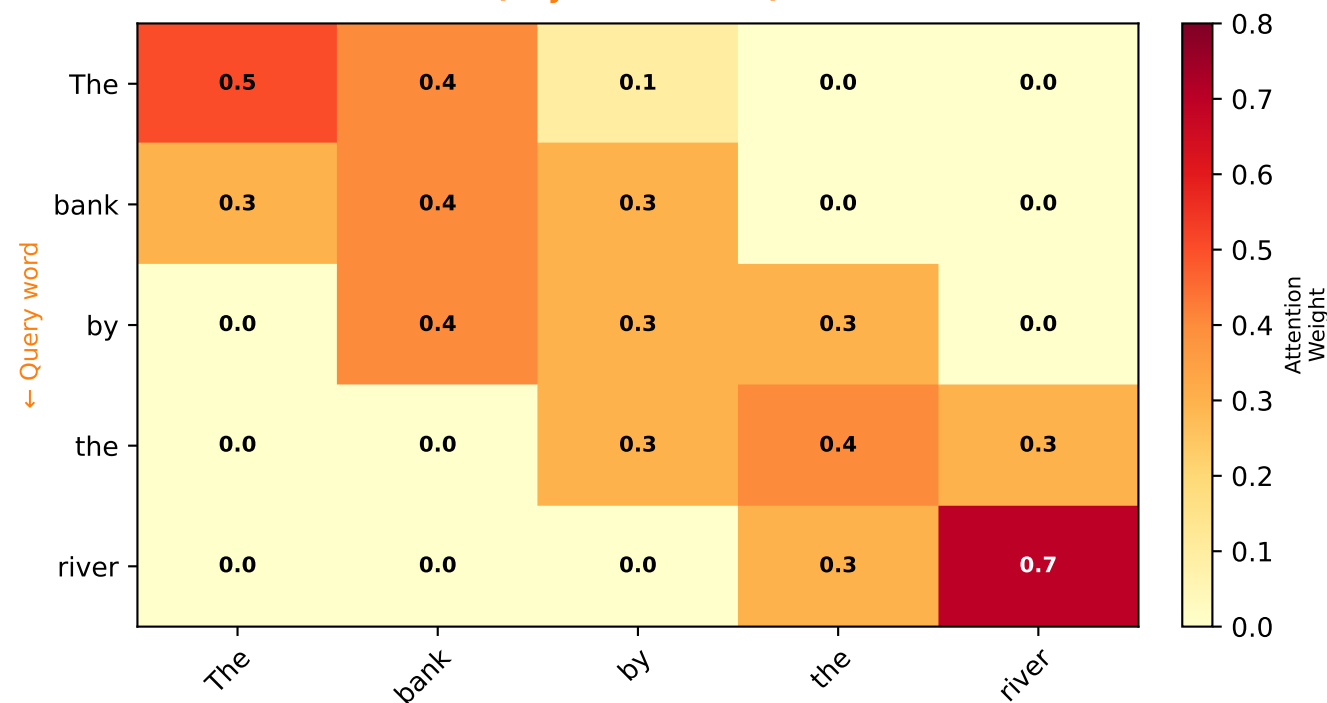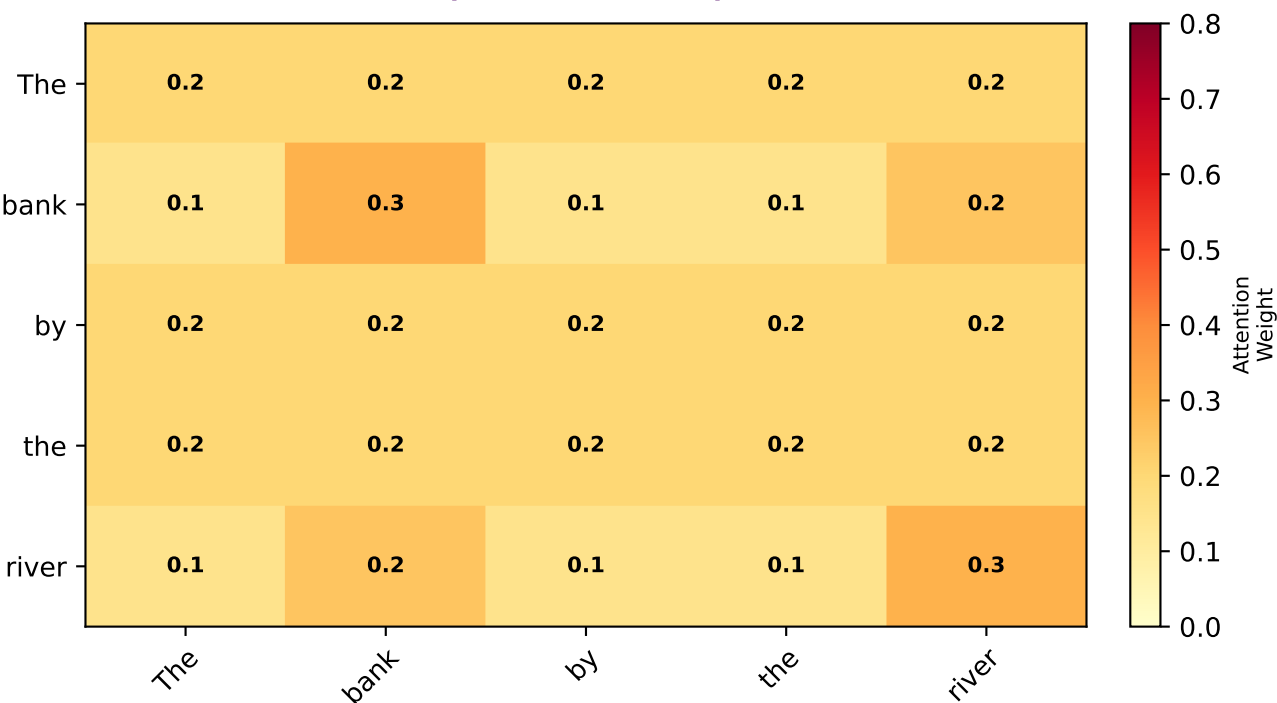Each head captures different relationships: Syntax (grammar), Semantics (meaning), Position (locality), Global (context)