Generating long and semantic-coherent reports to describe medical images poses great challenges towards bridging visual and linguistic modalities, incorporating medical domain knowledge, and generating realistic and accurate descriptions. We propose a novel Knowledge-driven Encode, Retrieve, Paraphrase (KERP) approach which reconciles traditional knowledge- and retrieval-based methods with modern learning-based methods for accurate and robust medical report generation. Specifically, KERP decomposes medical report generation into explicit medical abnormality graph learning and subsequent natural language modeling. KERP first employs an Encode module that transforms visual features into a structured abnormality graph by incorporating prior medical knowledge; then a Retrieve module that retrieves text templates based on the detected abnormalities; and lastly, a Paraphrase module that rewrites the templates according to specific cases. The core of KERP is a proposed generic implementation unit---Graph Transformer (GTR) that dynamically transforms high-level semantics between graph-structured data of multiple domains such as knowledge graphs, images and sequences. Experiments show that the proposed approach generates structured and robust reports supported with accurate abnormality description and explainable attentive regions, achieving the state-of-the-art results on two medical report benchmarks, with the best medical abnormality and disease classification accuracy and improved human evaluation performance.