Figure 6: The attention heat map from the Tree Transformer.

Figure 7: The attention heat map from the original Transformer.
Figure 8: The parse trees induced by the proposed Tree Transformer.
Figure 9: The parse trees induced by the proposed Tree Transformer.