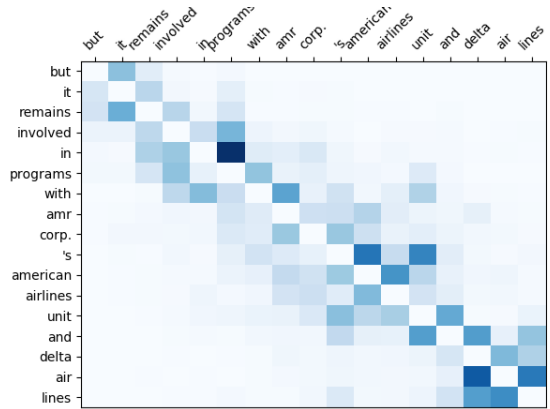
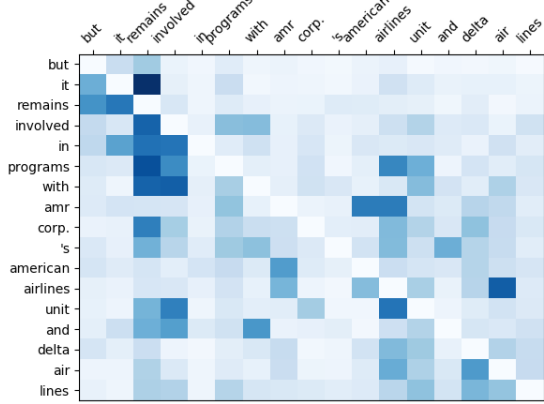


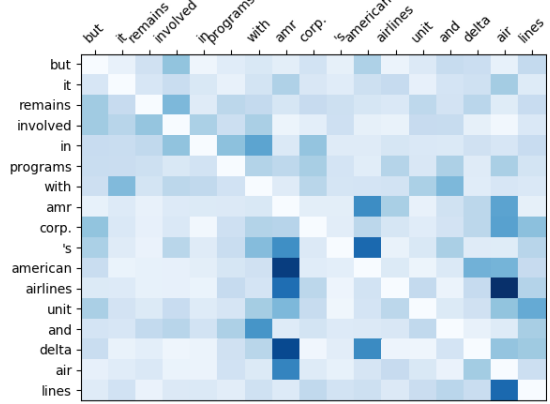
(a) The layer 0.



(b) The layer 3.

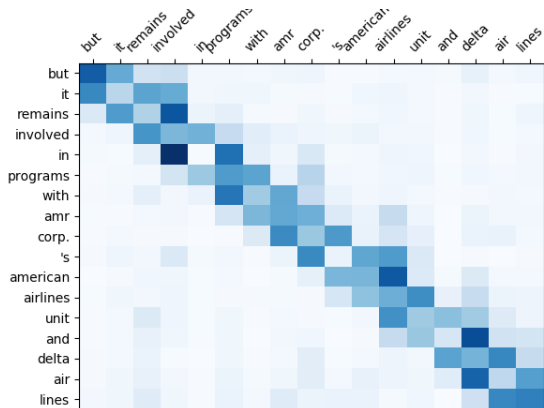


(c) The layer 6.

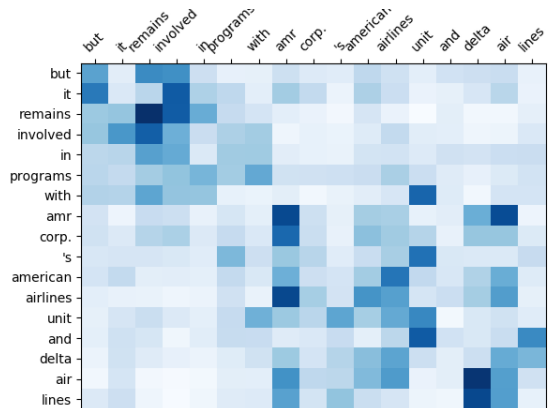


(d) The layer 9.

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



(a) The layer 0.



(b) The layer 3.

Figure 7: The attention heat map from the original Transformer.

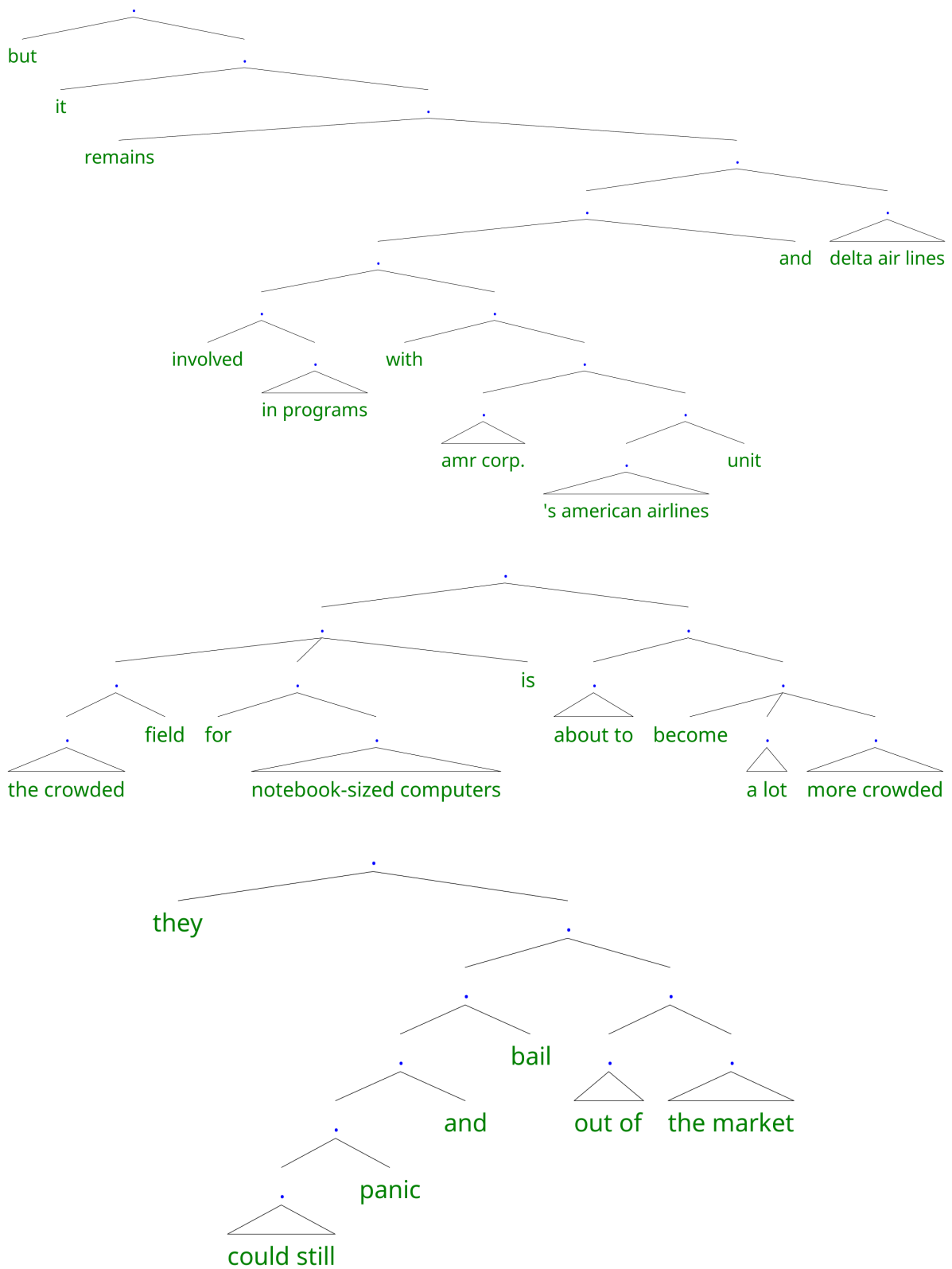


Figure 8: The parse trees induced by the proposed Tree Transformer.

