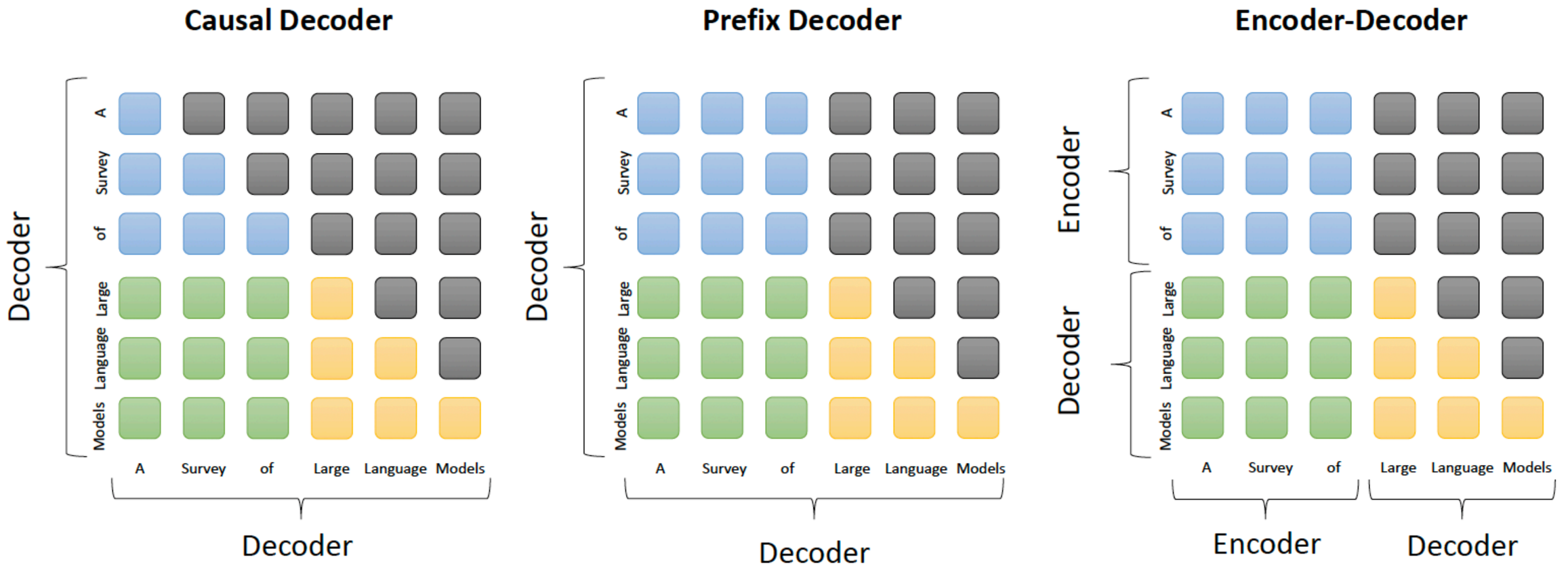

Fig. 7: A comparison of the attention patterns in three mainstream architectures. Here, the blue, green, yellow and grey rounded rectangles indicate the attention between prefix tokens, attention between prefix and target tokens, attention between target tokens, and masked attention respectively.

Transformerの構成要素：正規化

- 正規化の方法
  - ❖ 学習を安定させるため、正規化は広く採用されている
  - ❖ 初期のTransformerはLayerNorm (平均と分散を用いて正規化する手法) を用いていたが、近年ではRMSNorm (二乗平均平方根を用いた手法) やDeepNorm (スキップ接続をスケールして学習を安定させる手法) といった手法も提案されている
- 正規化の位置
  - ❖ 学習の安定性から、ほとんどのLLMで各サブレイヤーと最終予測の前に正規化を置くpre-LNが採用されている