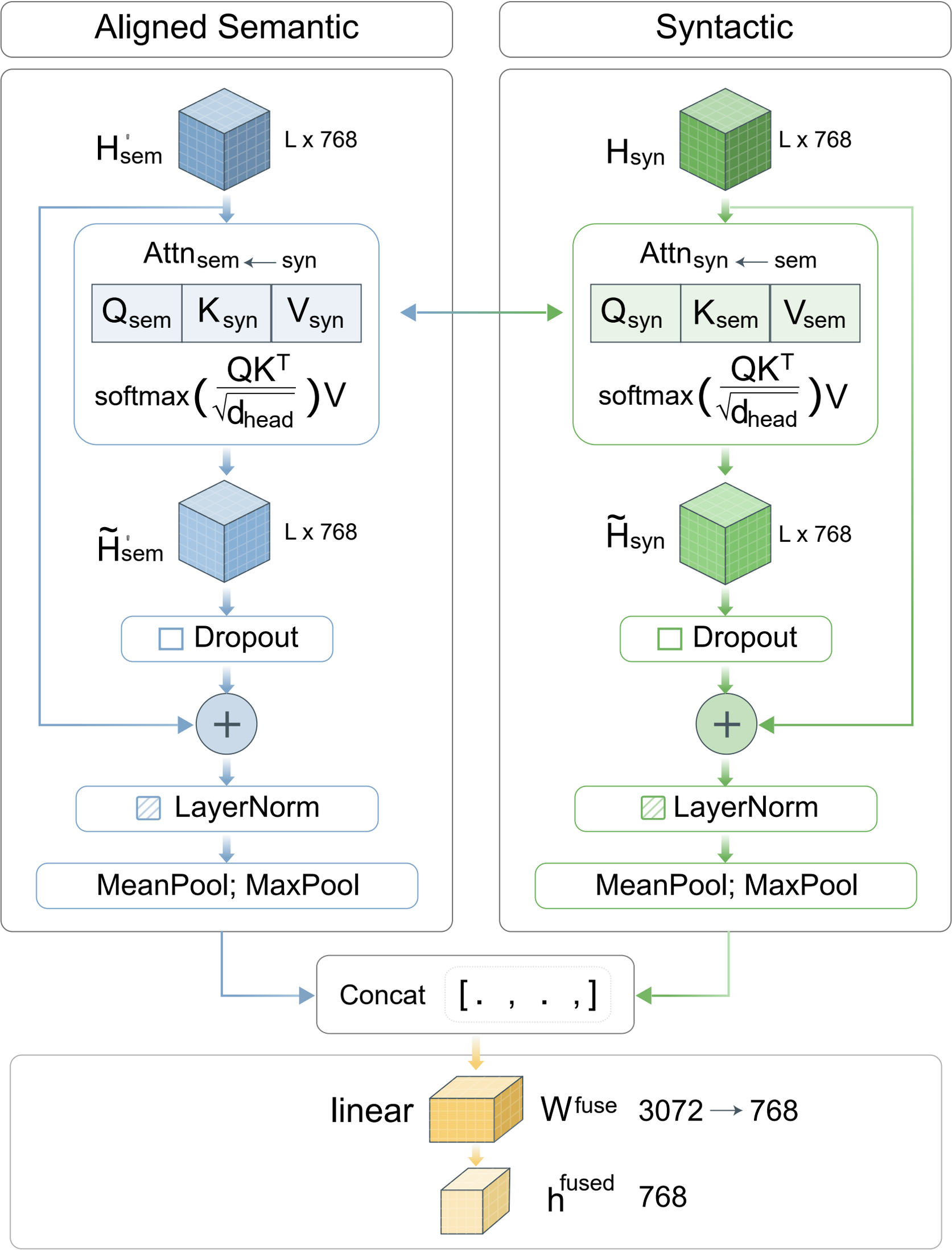
Aligned Semantic
Syntactic
H'_{sem}
L x 768
H_{syn}
L x 768
$Attn_{sem \leftarrow syn}$
Q_{sem} K_{syn} V_{syn}
$softmax(\frac{QK^T}{\sqrt{d_{head}}})V$
$Attn_{syn \leftarrow sem}$
Q_{syn} K_{sem} V_{sem}
$softmax(\frac{QK^T}{\sqrt{d_{head}}})V$
$\tilde{H}'_{sem}$
L x 768
$\tilde{H}_{syn}$
L x 768
Dropout
Dropout
+
+
LayerNorm
LayerNorm
MeanPool; MaxPool
MeanPool; MaxPool
Concat [. , . ,]
linear
W^{fuse} 3072 → 768
h^{fused} 768