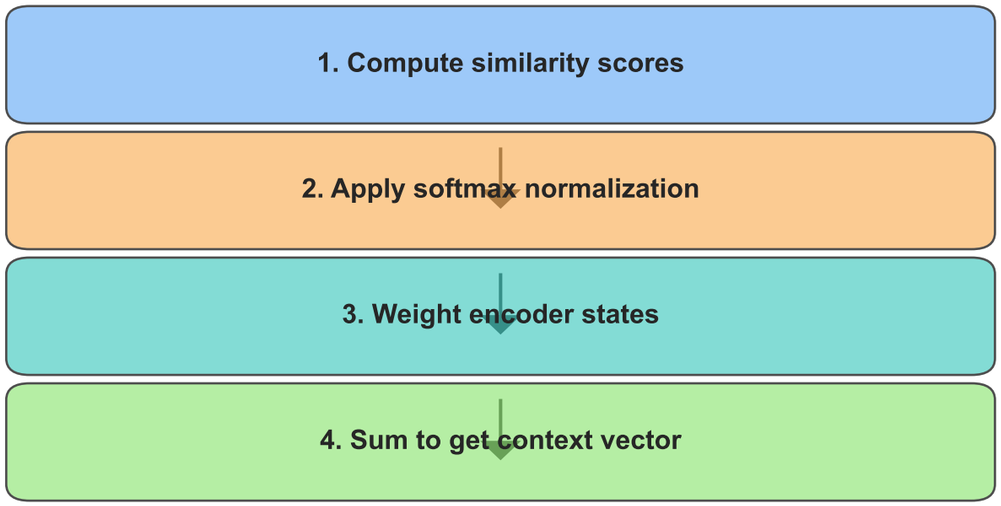
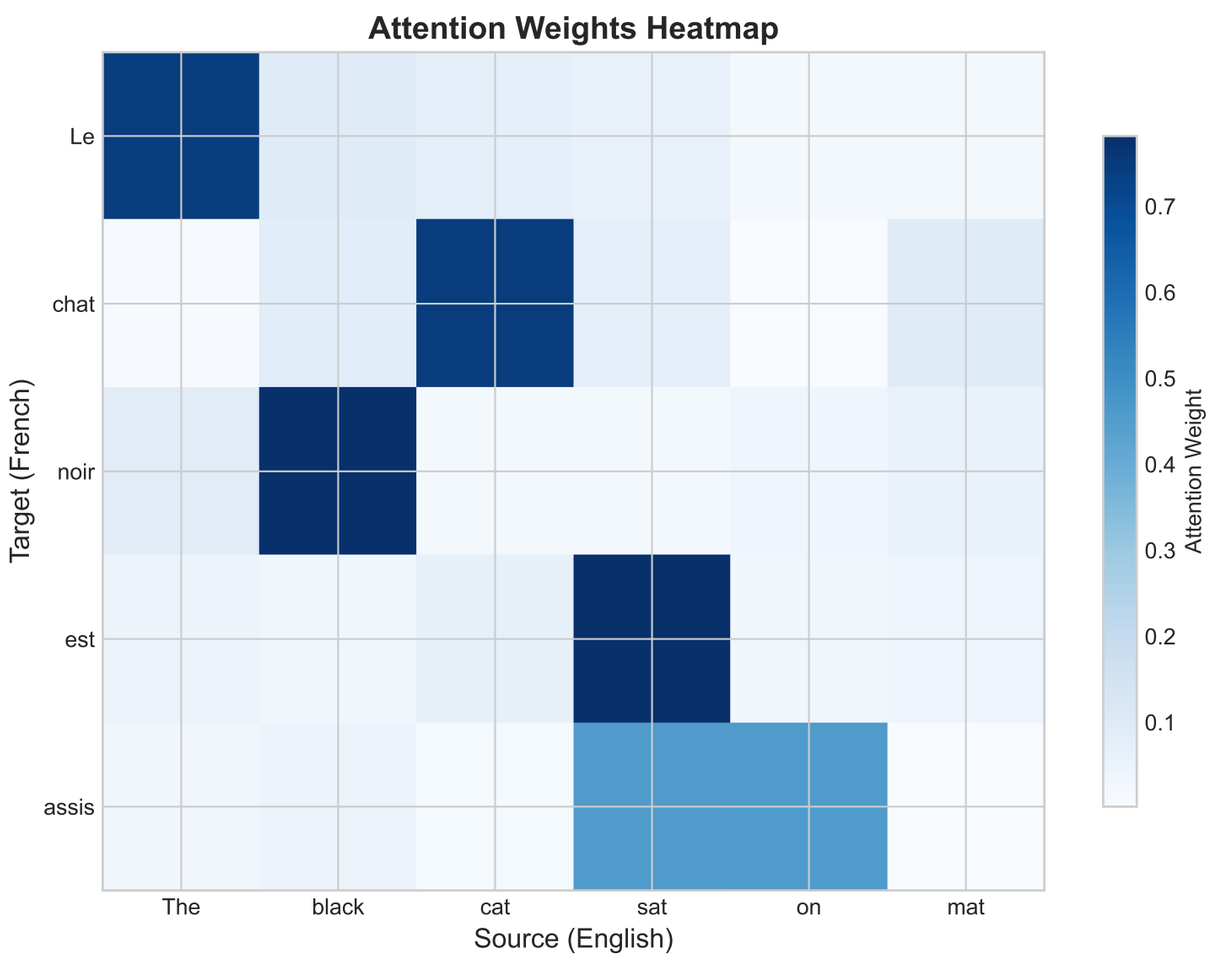
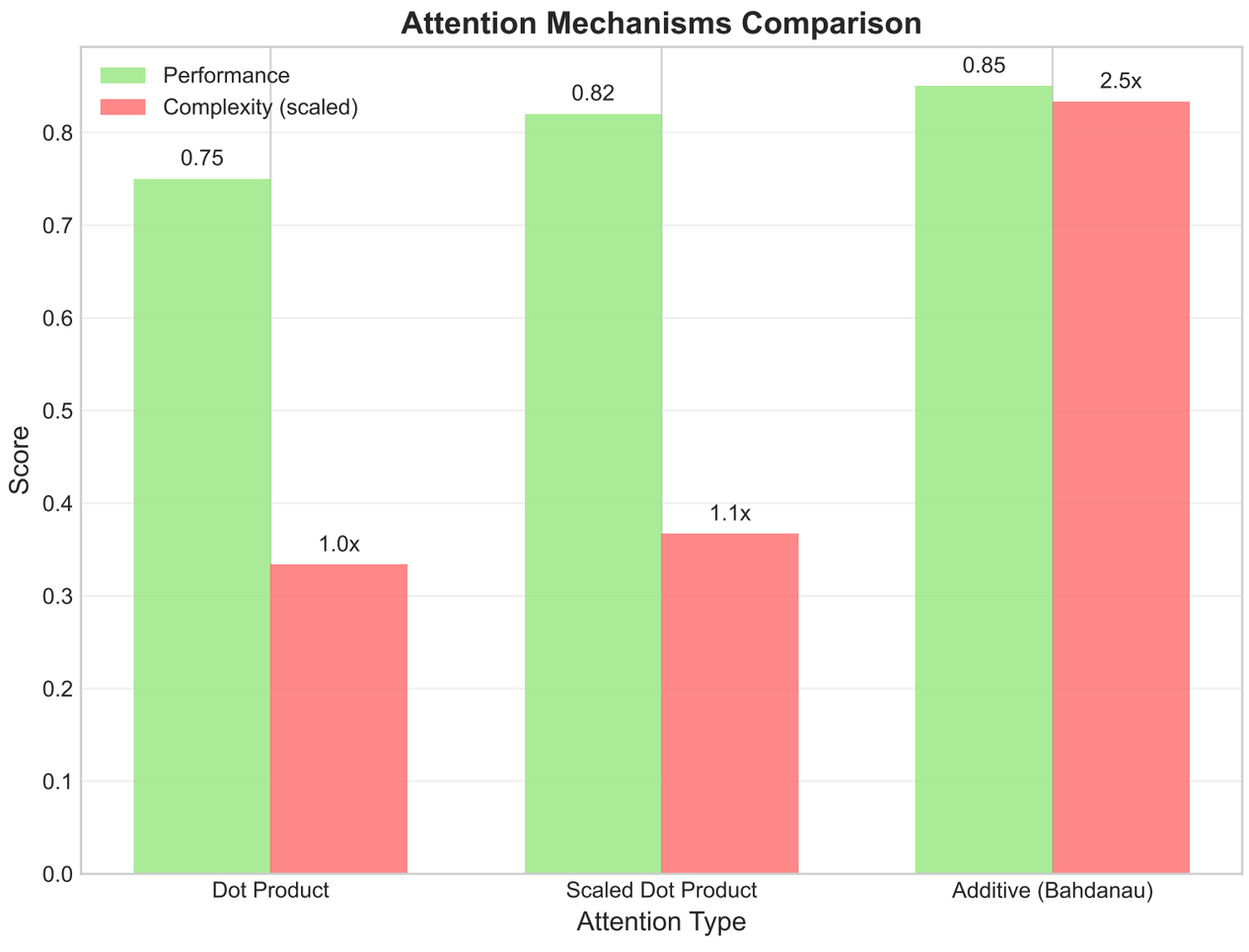
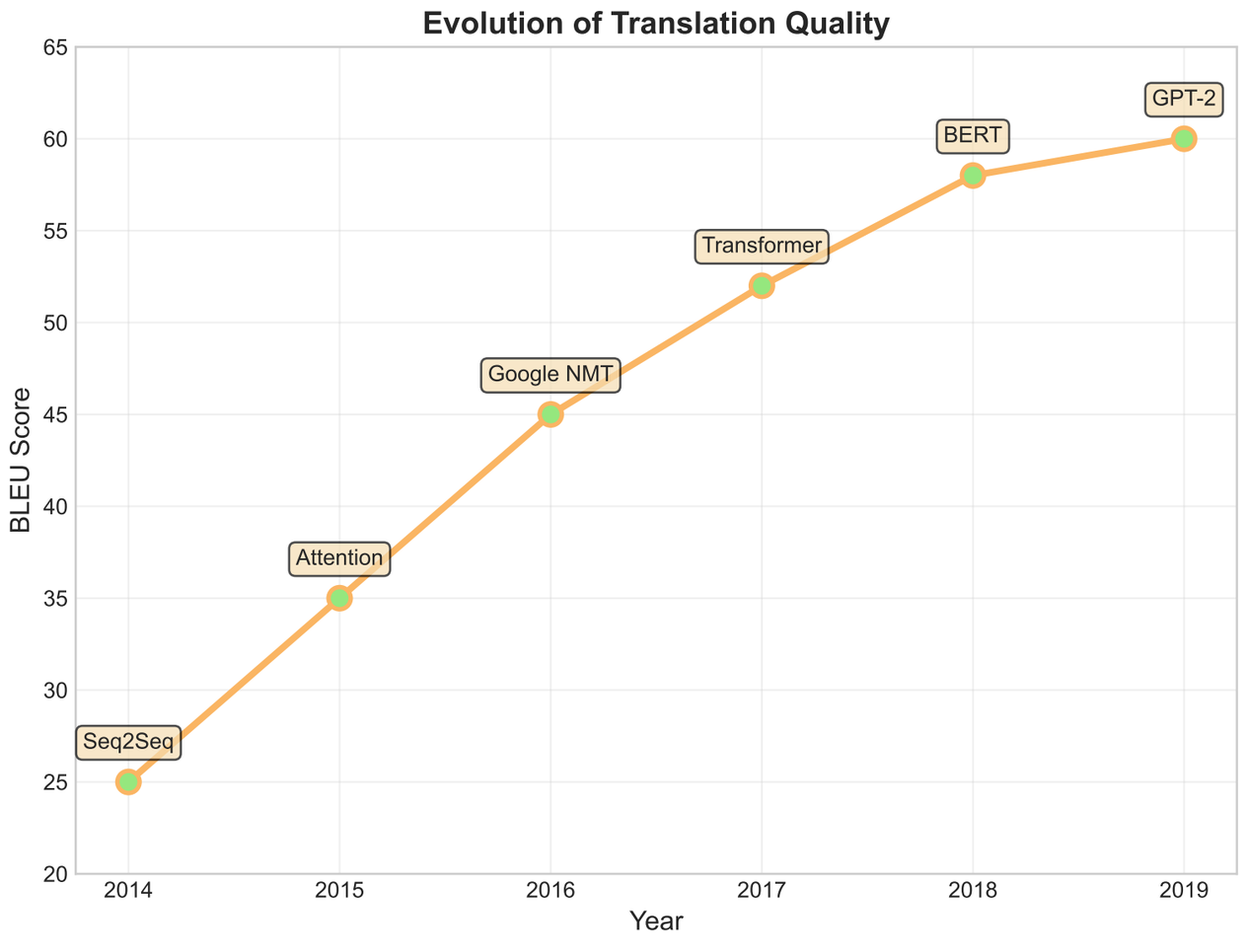

Attention Mechanism Steps
1. Compute similarity scores
2. Apply softmax normalization
3. Weight encoder states
4. Sum to get context vector


Attention Weights Heatmap
Target (French)
Le
chat
noir
est
assis
The
black
cat
sat
on
mat
Source (English)
Attention Weight
0.7
0.6
0.5
0.4
0.3
0.2
0.1


Attention Mechanisms Comparison
Performance
Complexity (scaled)
0.75
1.0x
0.82
1.1x
0.85
2.5x
Score
Dot Product
Scaled Dot Product
Additive (Bahdanau)
Attention Type


Evolution of Translation Quality
BLEU Score
Seq2Seq
Attention
Google NMT
Transformer
BERT
GPT-2
2014
2015
2016
2017
2018
2019
Year