

Exercise Briefing: Transformer

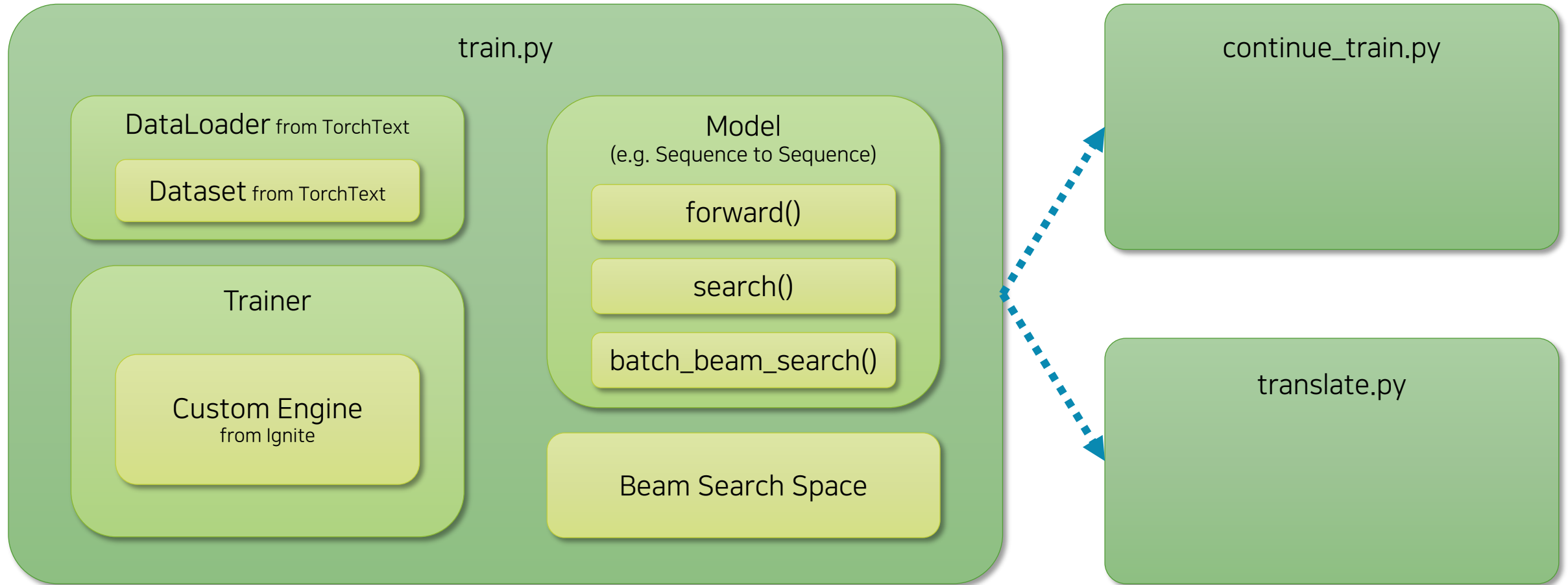
Ki Hyun Kim

nlp.with.deep.learning@gmail.com

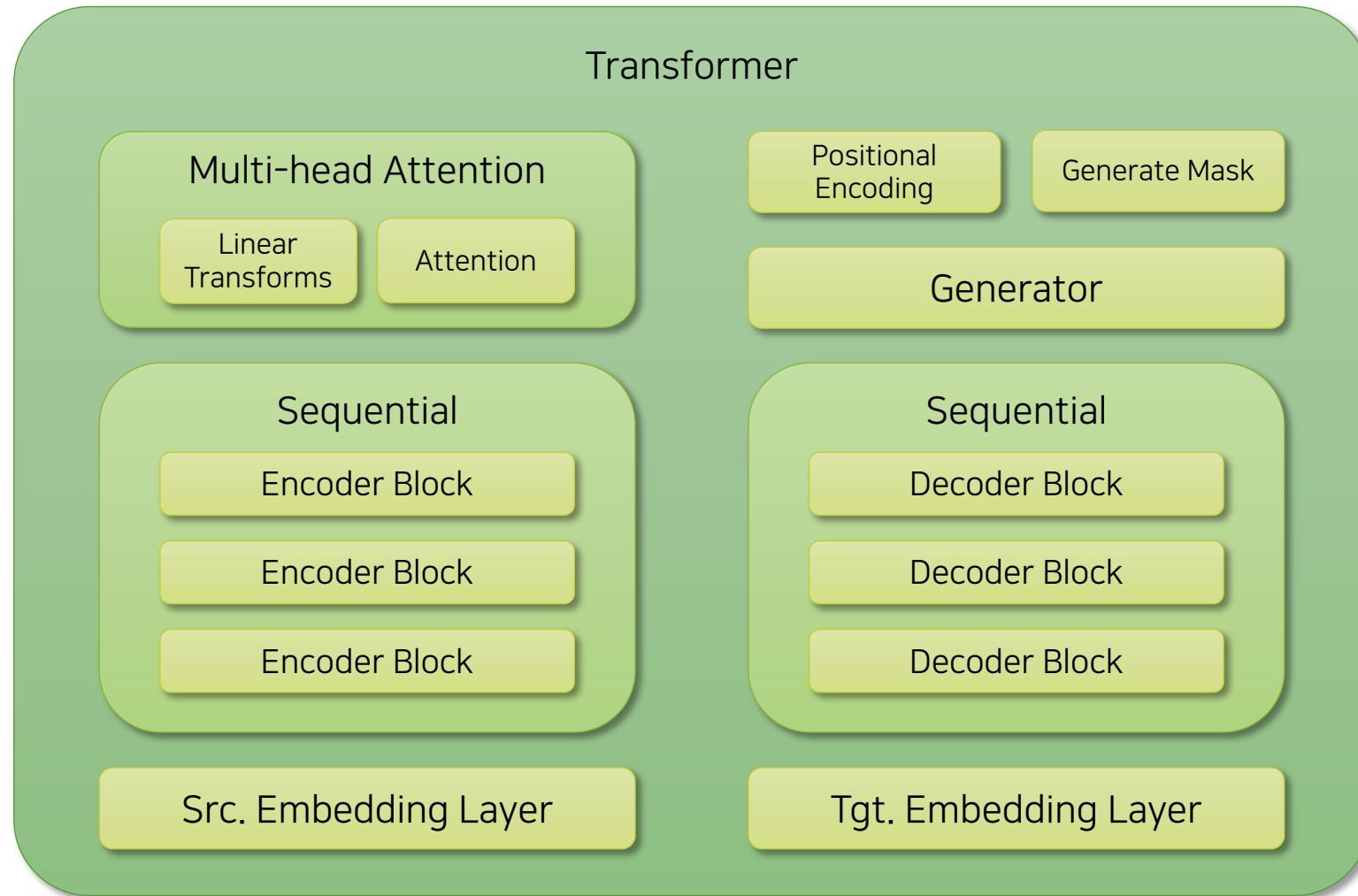
Again!

- 모델의 입출력 인터페이스만 맞춘다면 Trainer를 별도로 구현할 필요가 없다!

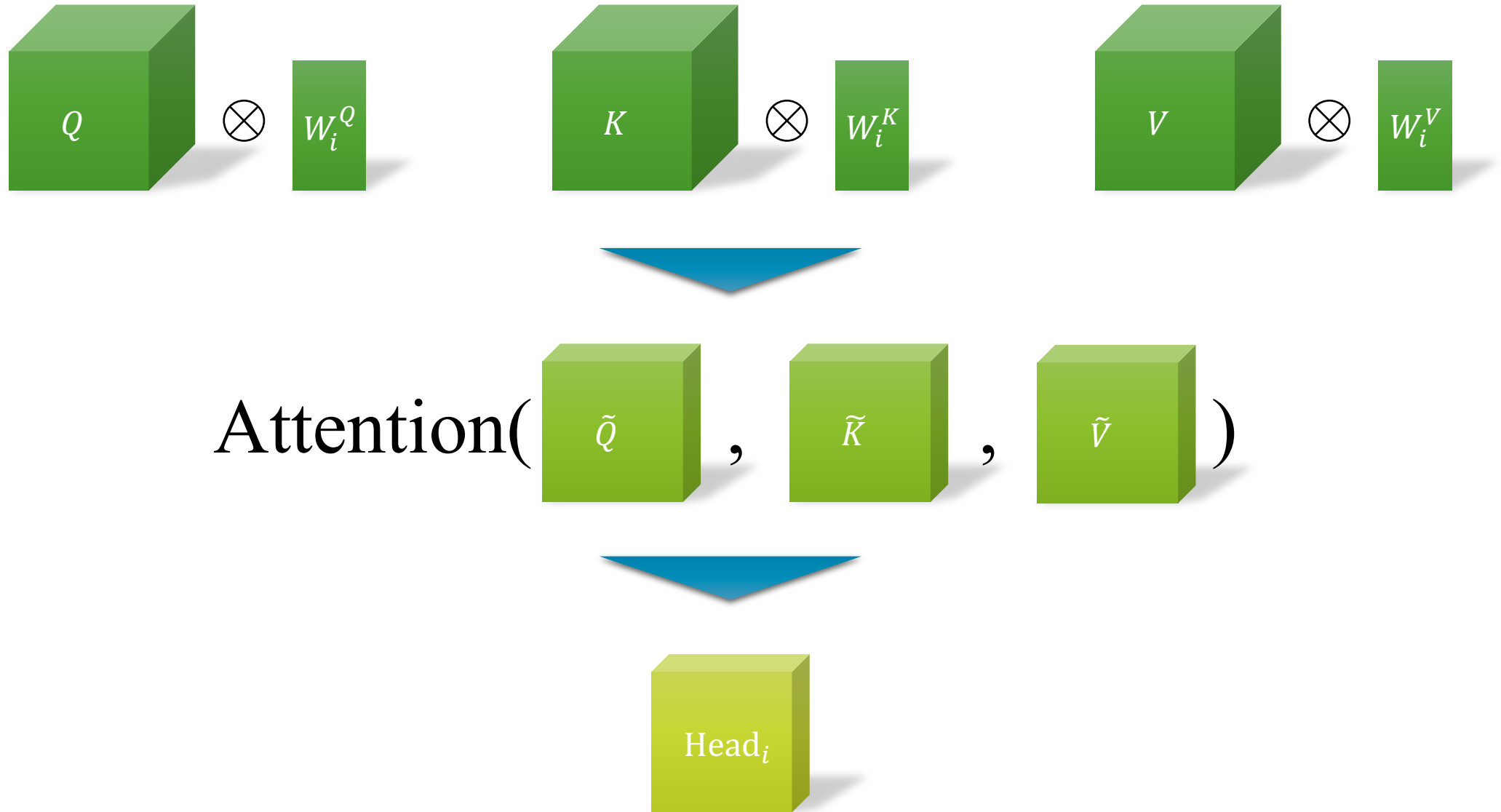
Project Implementation



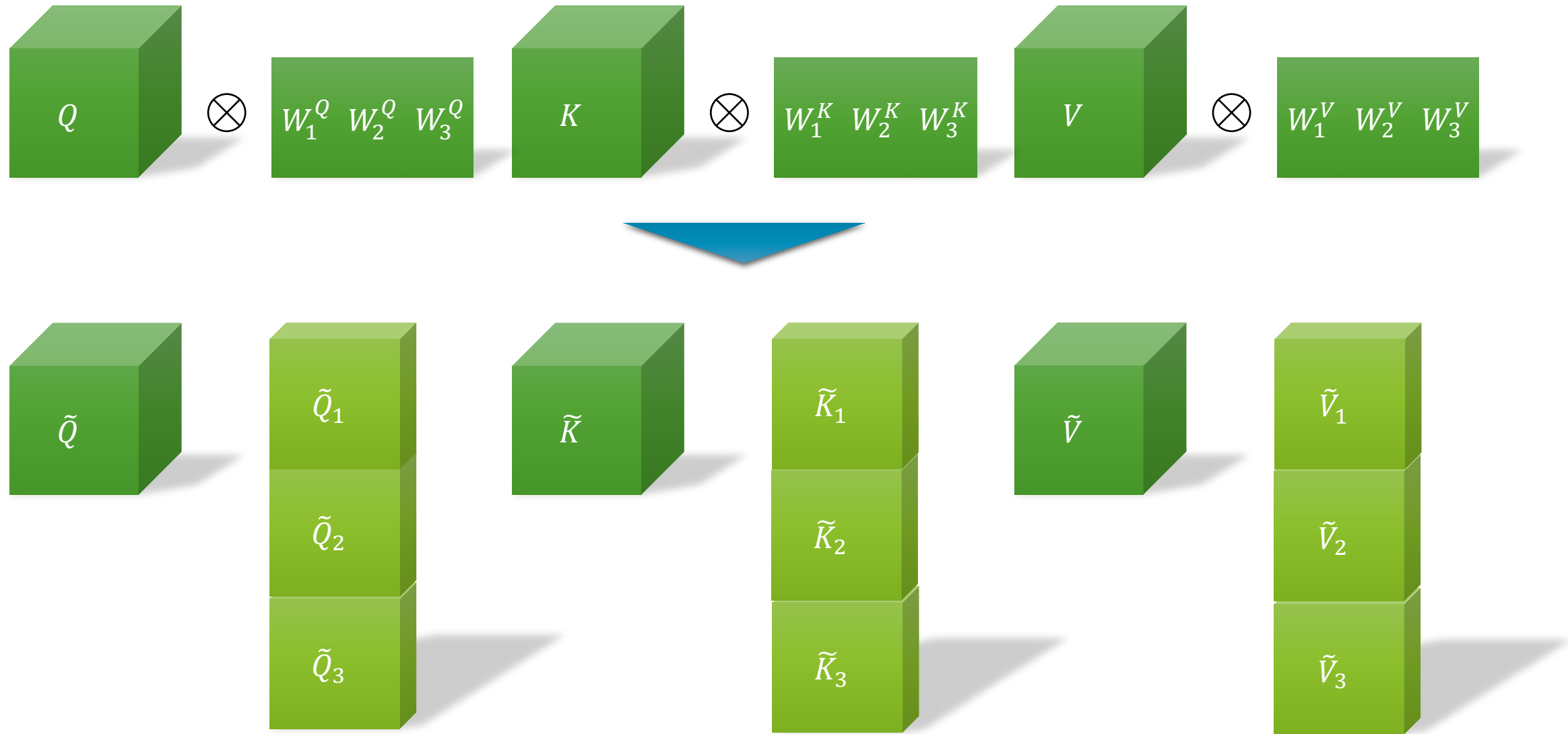
How to Implement Transformer



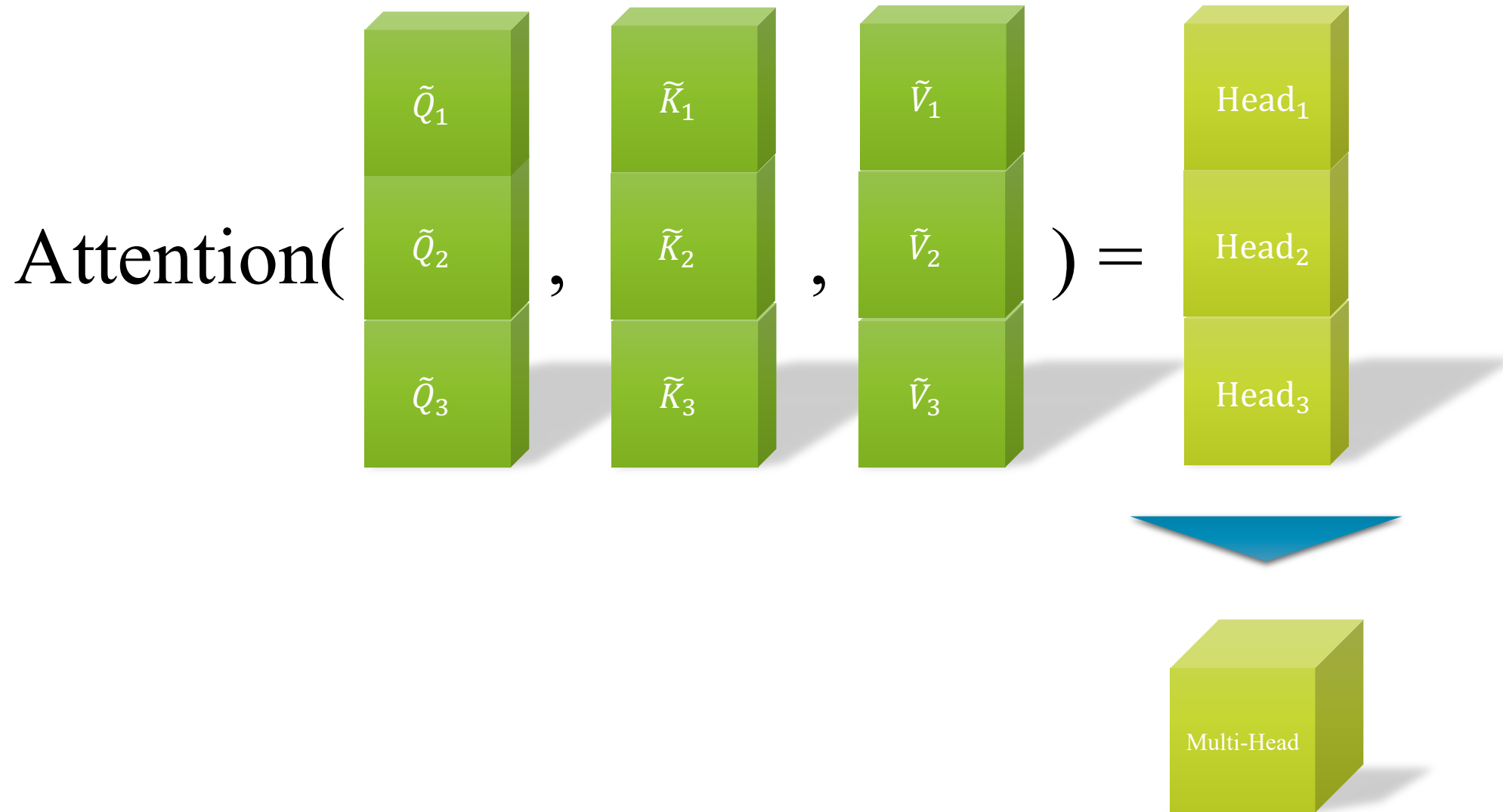
Multi-head Attention



Parallelized Multi-head Attention



Parallelized Multi-head Attention



What we need to remember in Beam Search

- From Board:
 - Input: x_t
 - ~~Last hidden state: h_{t-1}~~
 - ~~Last cell state c_{t-1}~~
 - ~~Last H-tilde: \tilde{h}_{t-1}~~
 - Outputs from each hidden layer until last time-step: $h_{i,1:t-1}^{dec}$
- Expand:
 - Entire output from encoder: $h_{1:n}^{enc}$
 - mask