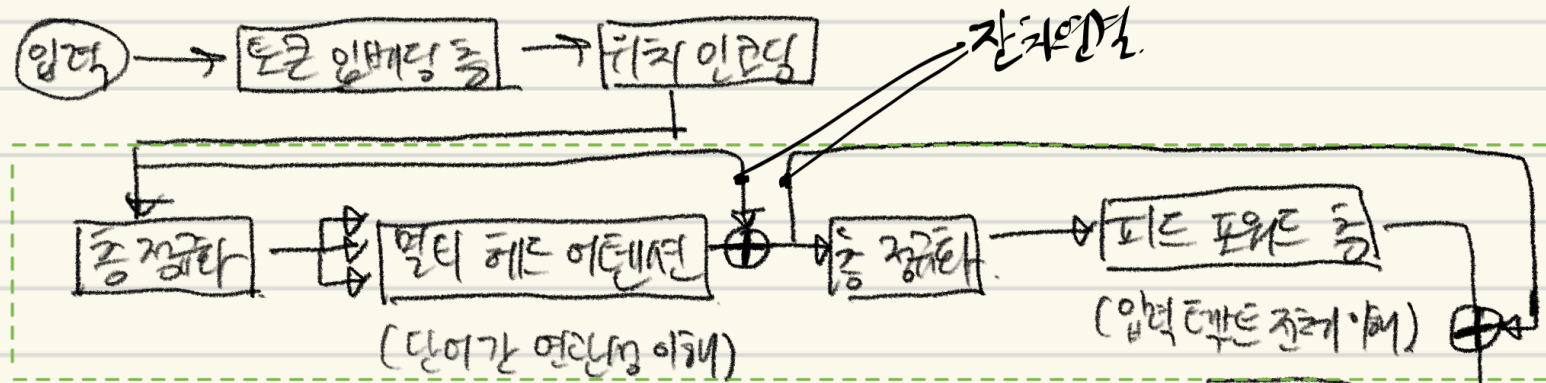


< 피드포워드 층 >

- 데이터 특징 학습 → 인코딩 연결 층 → 입력 텍스트 전체 이해
- 임베딩 차원을 동일하게 유지해야 쉽게 학습할 수 있음.
- 일반적으로 d-model 보다 2~3 배 큰 차원으로 학습했더니 다시 d-model로 변환.

< 인코더 >

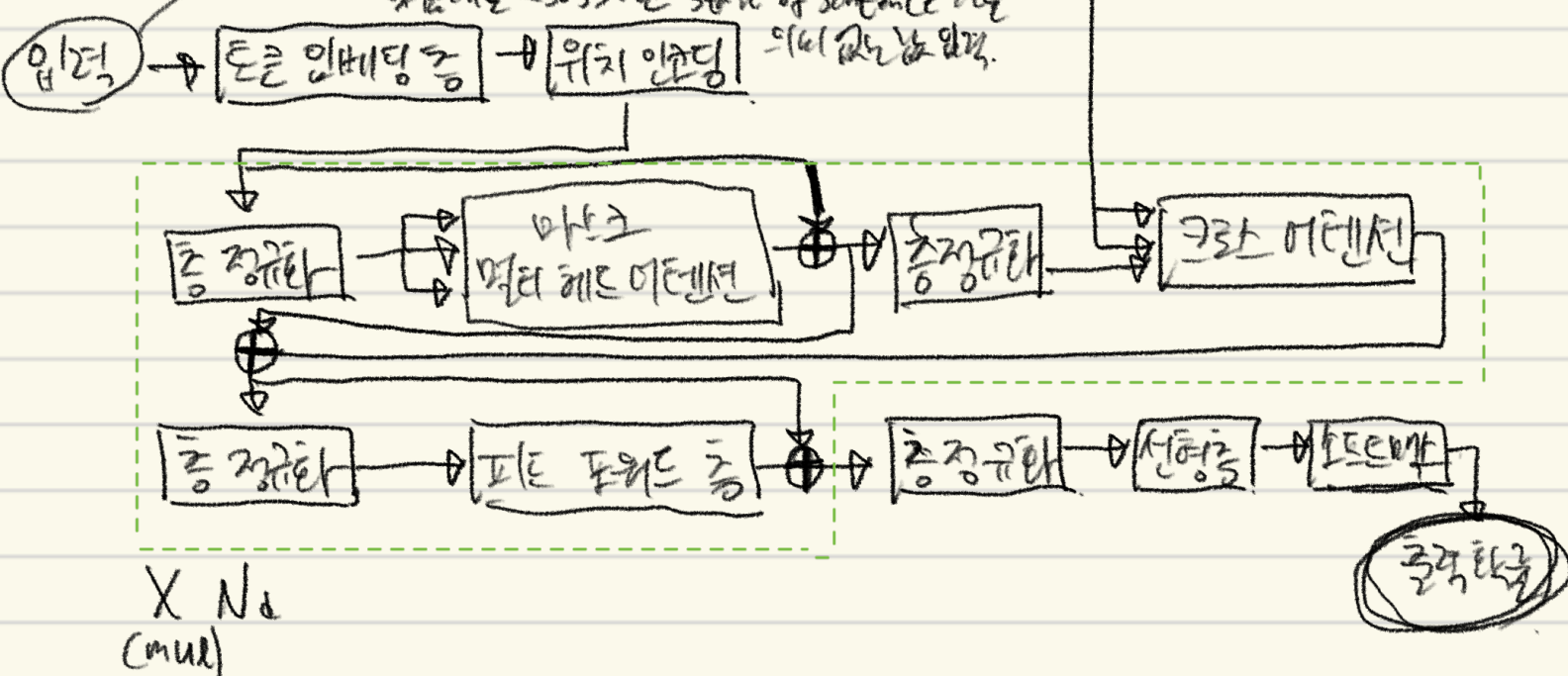
- 앞서 살펴본 멀티 헤드 어텐션, 층 정규화, 피드포워드 층이 반복되는 형태.
- 잔차 연결 (Residual Connection)
 - ∴ 입력값과 학습된 층의 입력은 다시 더해주는 형태로 구현.



$\times N_e$
(mul)

< 디코더 >

여기서 이 입력은 인코더의 출력값.
"출력 텍스트의 입력"인 → 지금까지 생성된 텍스트.
처음에는 <SOS>로 Start of Sentence 라고
의미하는 입력.



$\times N_d$
(mul)

- 마스크 멀티 헤드 어텐션

∴ 텍스트를 생성할 때, 입력 텍스트 (문자로 표현) → 특정 단어 생성 시 해당 단어 이후의 단어를 알면
안됨 ∴ 마스크, 인코더 & 디코더 모두 전체 텍스트를 학습 → 특정시킬 때까지 생성된 단어만 학습 → 마스크.