# 제6장 기계번역

## 6.1 인간의 언어가 정말 어려운 이유

- 1. 너무 많은 규칙: 인간의 언어를 몇 가지 규칙만으로 설명하기란 사실상 불가능한데, 왜냐하면 인간의 언어는 신조어가 생겨나면서 계속 확장하기 때문이다
  2. 너무 많은 오류: 분명히 문법에 어긋난 문장인데, 우리는 아무렇지 않게 받아들인다.
  3. 너무 많은 의미: 같은 발음을 지닌 단어가 여러 뜻을 갖는 경우가 있다

## 6.2 기계번역의 시작

- 인간이 사용하는 언어를 기계를 사용해 다른 언어로 번역해내는 일을 기계번역이라 한다

## 6.3 규칙 기반, 모든 규칙을 정의하다

- 시스트란이라는 회사가 있었는데(기계번역을 대표하는 회사) 규칙 기반 기계번역을 이용했다.

규칙을 아무리 세워도 언어의 무궁무진한 변화를 결코 따라갈 수 없기 때문에 한계가 있었다.

이 이유가 규칙 기반 번역 모델이 계속해서 실패했던 이유다

## 6.4 예시 기반의 통계 기반 가능성을 보이다

-나가오 마코토 교수는 예시 기반 기계번역(Example-Based Machine Translation)이 라는 획기적인 방식을 제안 이에 기반해 매우 성능이 좋은 영어-일본어 번역 시스템을 만듬

규칙을 통해 언어를 '이해'하기보다, 경험을 통해 '모방'하는 형태로 접근했다. 기본적인 문장의 의미를 파악한 다음 비슷한 문장의 의미를 비교해 전체 의미를 '유추'해내는 방식이다.

하지만 수많은 동음이의어를 직접 처리해야 하는 등 이 방식에도 많은 한계가 있었다

통계 기반 기계번역(Statistical Machine Translation)

이 모델은 문장을 단어 또는 구문 단위로 분할한 다음 이를 번역하고 다시 문장으로 합치는 과정에 확률적인 방법을 접목한다

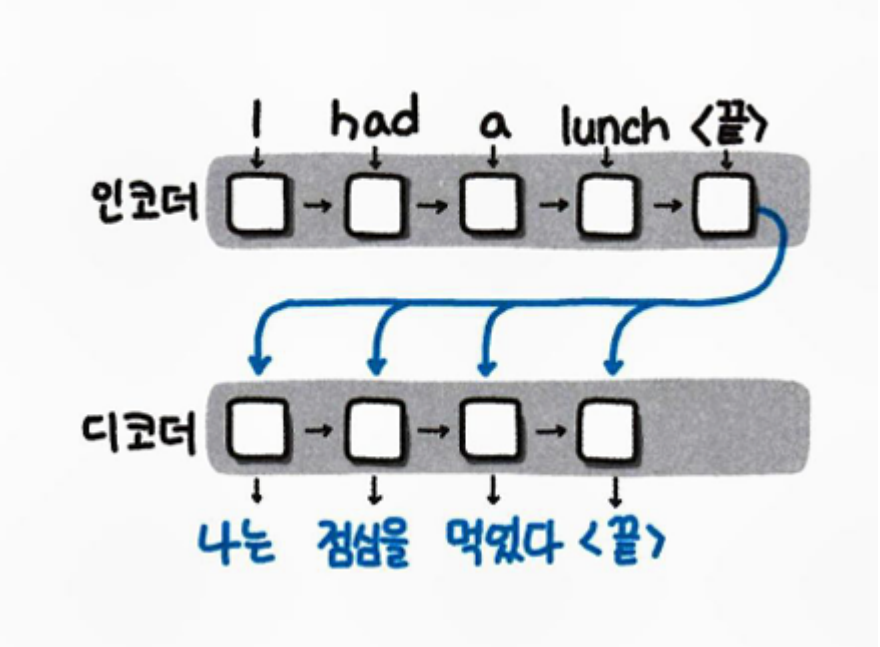
## 6.5 인공 신경망 기반,마침내 혁신이 시작된다

-바야흐로 딥러닝을 본격적으로 기계번역에 도입한다. 이후에는 구문 단위를 넘어 아예 문장 전체에 딥러닝을 적용한다. 이를 신경망 기반 기계번역(Neural MachineTranslation)이라고한다

신경망 기반 기계번역의 작동은 문장 전체를 마치 하나의 단어처럼 통째로 번역해서 훨씬 더 자연스러운 번역이 가능하게 했다. 인공 신경망이라는 거대한 모델과 이를 견인할 수 있는 방대한 데이터를 확보하면서 이것이 가능해졌다

## 6.6 어텐션 가장 혁신적인 발명

- 문장을 압축하는 과정과 풀어내는 과정




문장을 압축하는 부분을 인코더(Encoder) 반대로 문장을 푸는 부분은 디코더(Decoder)라고 한다

한 단어씩 차례대로 푸는데, 이때 2가지 입력을 받는데 첫 번째는 앞선 단어의 번역이고,

두 번째가 바로 인코더가 압축한 벡터이다 문장 번역이 끝날 때까지 디코더는 계속해서 인코더가 압축한 벡
터를 참조하면서 더 자연스러운 문장을 만든다. 이런 방식으로 인공 신경망을 활용한 기계번역은 엄청난 성능을 보인다

이 방식에는 2가지 문제가 있다. 첫 번째 문제는 번역할 원문의 길이와 관계없이 원문을 일정한 길이의 벡터로 한 번만 압축한다는 점입니다

두 번째 문제는 한번 만든 벡터를 계속 참조하다 보니 번역문이 길어질수록 핵심 단어를 놓친다는 점

위의 한계를 극복하는 혁신적 개념인 어텐션(Attention)

어텐션의 핵심은 중요한단어에 별도로 가중치를 부여할 수 있다는 점이다

그리고 어텐션은 단어 사이의 거리가 아무리 멀어도 서로 관련이 있는 단어라면 그 단어에 별도로 표시를 해두어 가중치를 높일 수 있다. 그래서 어텐션은 특히 긴 문장에서 높은 성능을 낼 수 있다

## 6.7 번역 규칙을 스스로 학습하다

-끊임없이 변형되고 확장하는 언어를 형식적으로 분석하는 데는 명백한 한계가 존재했기에, 컴퓨터를 이용한 자연어 처리 연구는 수십 년 동안이나 지지부진했지만 신경망을 도입하면서 돌파구가 열렸다

기계번역에 더 이상 규칙을 입력하지 않고. 비슷한 문장에서 규칙을 스스로 학습한다. 신경망 기반 기계번역 또한 수많은 문장을 보며 스스로 규칙을 학습하고 언어를 이해한다. 번역이라는 복잡한 문제를 데이터를 통해
스스로 해결한다 신경망 기반 모델은 끊임 없이 발전하고. 단순히 문장 전체를 학습하는 수준을 넘어 중요한 단어에 주목하는 어텐션이라는 개념도 고안되었다

## 6.8 인간을 뛰어넘은 기계번역

-2017년에는 카카오가 카톡 챗봇 형태로 카카오i 번역 서비스를 선보이고 네이버는 신경망이 등장하기 이전에도 파파고라는 이름으로 오랫동안 번역 서비스를 해왔었는데 신경망 기반 영어-한국어 기계번역 서비스를 세 계 최초로 출시 했다. 이후 파파고를 개발한 핵심 인력들은 현대자동차에서 자동차 도메인에 적합하도록 모델을 개선하여 신경망 기반의 번역 서비스를 출시한다

## 6.9 바벨탑 인간은 신의 형벌을 극복할수 있을까

-