

한국어의 특수성을

반영한 번역 성능 향상



참여기업체 : AITRICS

지도 교수님 : 최희열 교수님

팀원 : 허재무 김준태 김정희 김주환

I. 문제 배경

■ 과제의 필요성

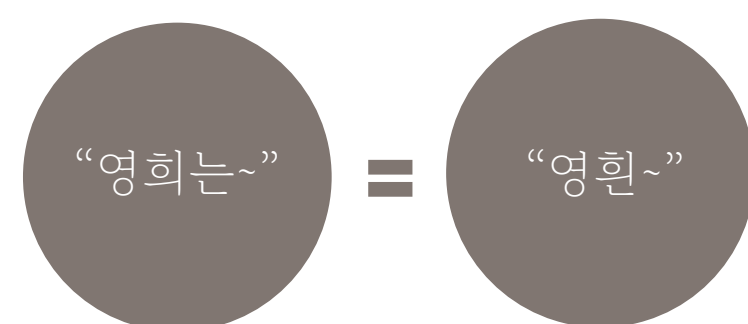
kakao brain

Flitto Clova

오늘날 네이버 CLOVA를 비롯하여 많은 인공지능 서비스 플랫폼, 번역기가 나오고 있음. 하지만 대부분의 번역기는 한글의 특수성이 반영되지 않음. 이는 가독성 및 번역 성능의 저하로 이어짐.

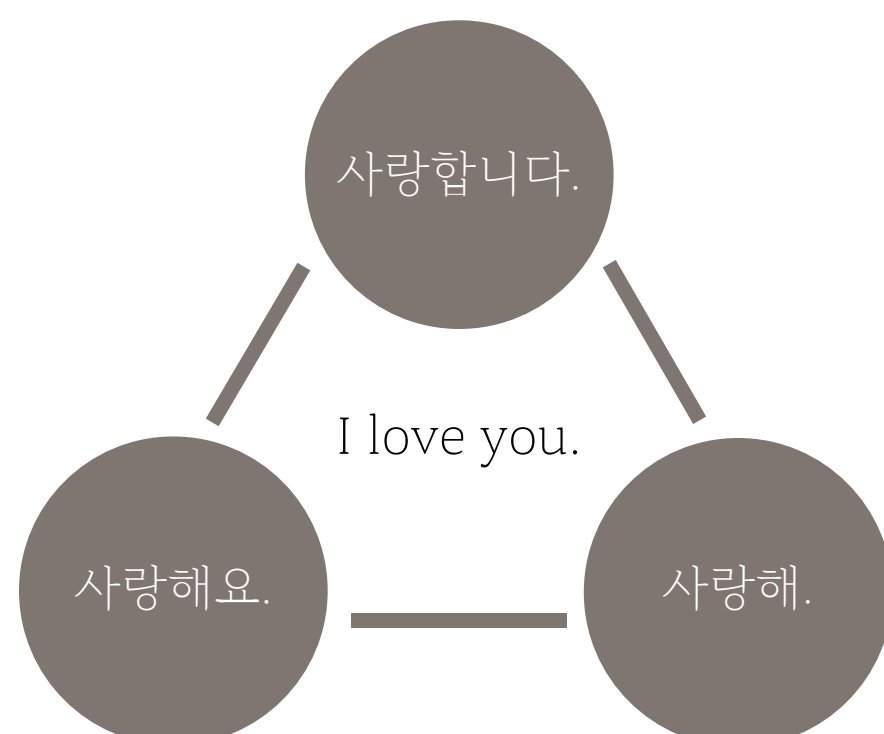
1) 받침

보통의 경우 명사는 조사(은/는/이/가)와 함께 쓰는 것이 일반적이지만, ‘엄만, 아빤’처럼 조사를 명사에 결합하여 사용하는 경우도 존재함



2) 존댓말 반말의 유무

동일한 의미를 가지지만, 다양한 형태로 표현이 가능함



III. 기존 연구 결과

■ 대중적인 번역기 비교

	I love you. I will marry with you.	→	사랑해요 나는 당신과 결혼할 것이다.
	영희 정말 대단해!	→	Younghee is really incredible!
	I love you. I will marry with you.	→	사랑해. 나는 당신과 결혼할 것이다.
	영희 정말 대단해!	→	Young white is really great!
	I love you. I will marry with you.	→	사랑해요, 나는 당신과 결혼할 거예요.
	영희 정말 대단해!	→	You're so great!

	정확성	가독성
	😊	😞
	😞	😊

Papago의 경우 높임말로 변환하는 기능은 있지만 기본적인 번역은 높임말과 반말을 혼합하여 사용하고 있음.

구글, 카카오의 경우 영희를 영+희으로 인식함. 즉 글자를 기준으로 문장을 번역하고 있음.

II. 문제 정의

■ Problem statement

한국어의 특수성을 반영한 모델의 학습이 필요함

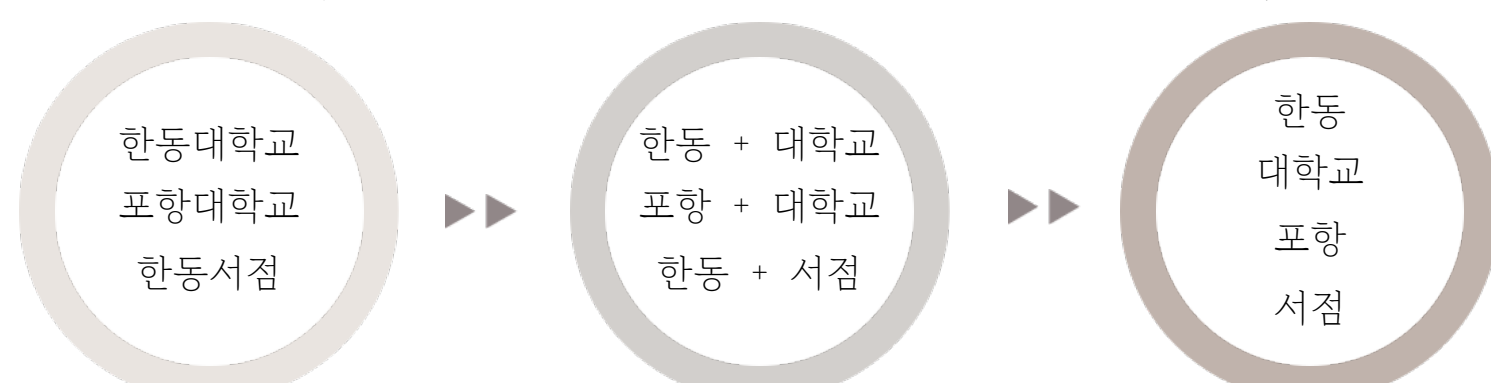
- 높임말 & 반말의 혼합 사용은 문장의 흐름을 끊고 가독성을 떨어림
- 받침 인식 불가능은 의도하지 않은 고유명사의 추가 등 번역의 성능을 낮춤

■ Constraints

높임말 & 반말 변환을 위한 모듈의 부재

음절이 떨어지는 단위로 BPE 생성

*BPE란 oov(out of vocabulary)문제를 해결하기 위한 알고리즘으로 문장을 단어보다 작은 단위의 서브 워드 단위로 분할하여 단어를 구성하는 방법론임



■ Objectives

한국어의 특수성을 반영한 번역 성능 및 가독성 향상

■ Function

모듈을 통한 문장의 반말 & 높임말 변환
자모 단위 bpe 생성을 통한 받침 반영

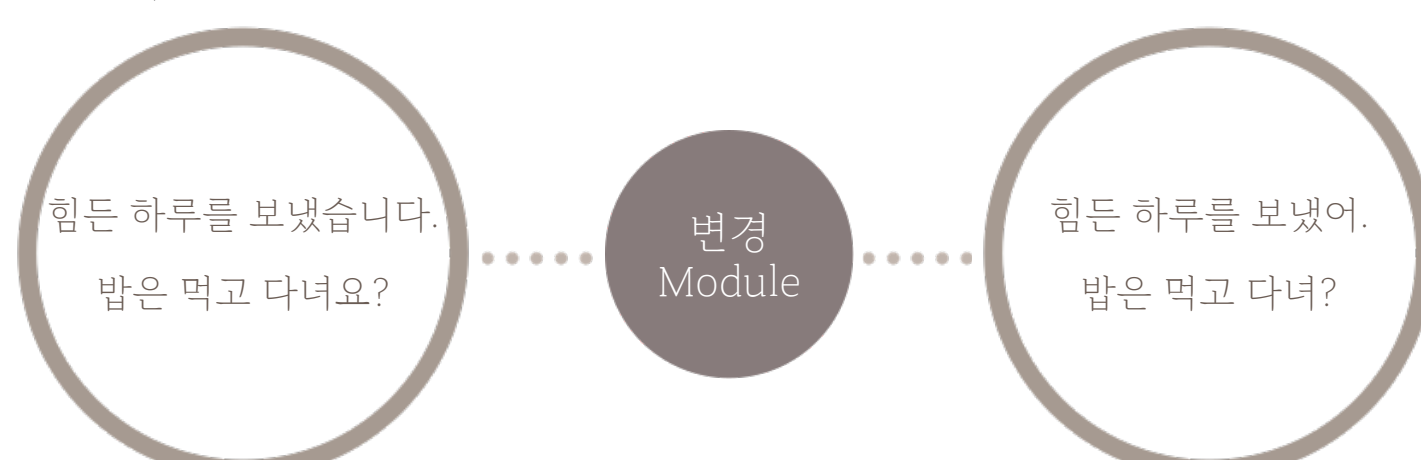
IV. 접근 방법

■ Key approach(핵심 접근법)

모듈을 이용한 높임말-반말 변환을 통해 한국어의 특수성 반영
자모 단위 번역을 통한 bpe 생성으로 인한 성능 향상

■ System Design

높임말 <-> 반말 transform



글자단위 -> 자모단위 transform의 bpe 변화

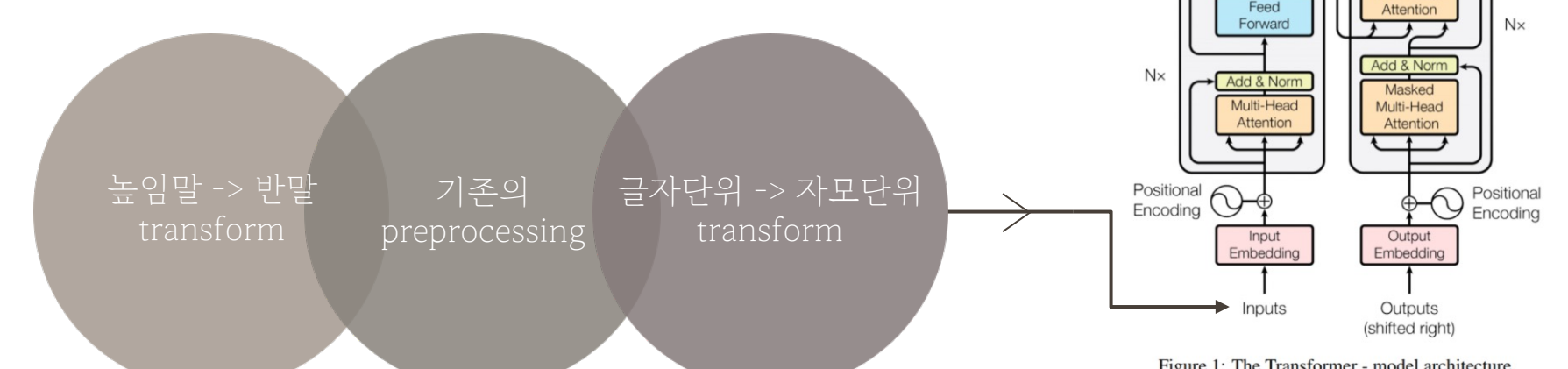
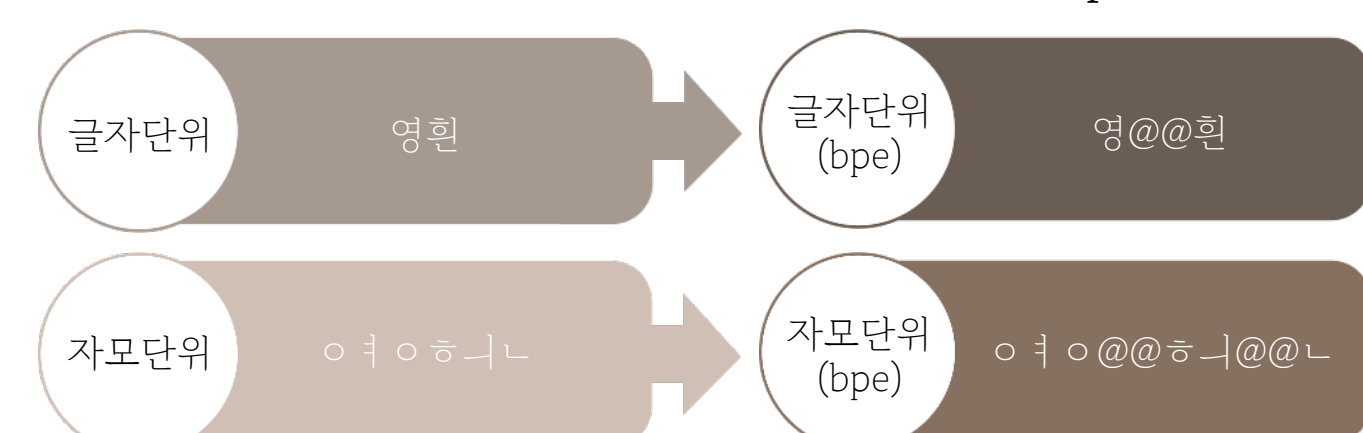


Figure 1: The Transformer - model architecture.