

한국어 · 조선어 언어규범 차이에 기반한 텍스트 전환 모형 구축 연구

김 용 해 (충남대학교) 이 하 양(충남대학교) 이 용 훈(충남대학교)
류 병 래(충남대학교)

2025 봄 한국현대언어학회, 현대문법학회, 대한언어학회 공동학술대회
5/24/2025(금), 대구가톨릭대학교

CONTENTS

1

서론



2

언어규범에 대한 분석



3

연구방법



4

연구 결과 및 향후 과제



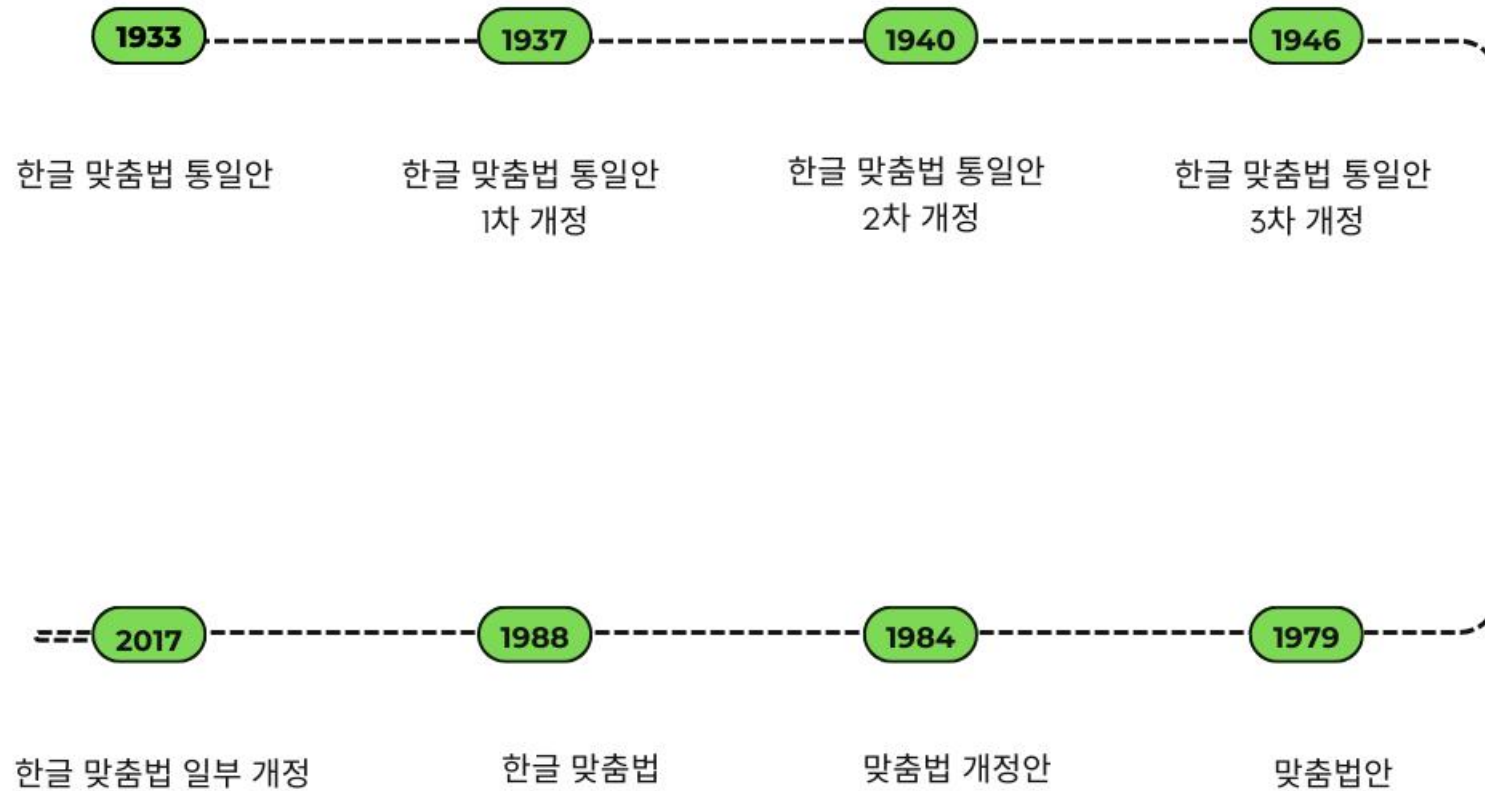
연구동기

- 동일한 언어적 기원을 가진 언어라도 지리적, 정치적, 문화적 환경의 차이에 따라 규범적 변화가 일어난다.
 - 한국어와 중국 조선어는 약 70여 년간의 분단과 상이한 사회적, 정치적 환경에서 독자적인 언어 규범 체계를 형성하였다.
 - 이러한 규범 차이는 단어의 형태, 문장 구조, 발음, 띄어쓰기, 맞춤법 등 언어 전반에 걸쳐 나타나며, 언어 교육, 문서 자동 교정, 기계 번역 등 실용 영역에서 중요한 고려 요소이다.
- 본 연구는 한국어와 조선어의 주요 언어 규범 차이를 비교하고, 이를 바탕으로 텍스트 전환 모형을 구축하고자 한다.

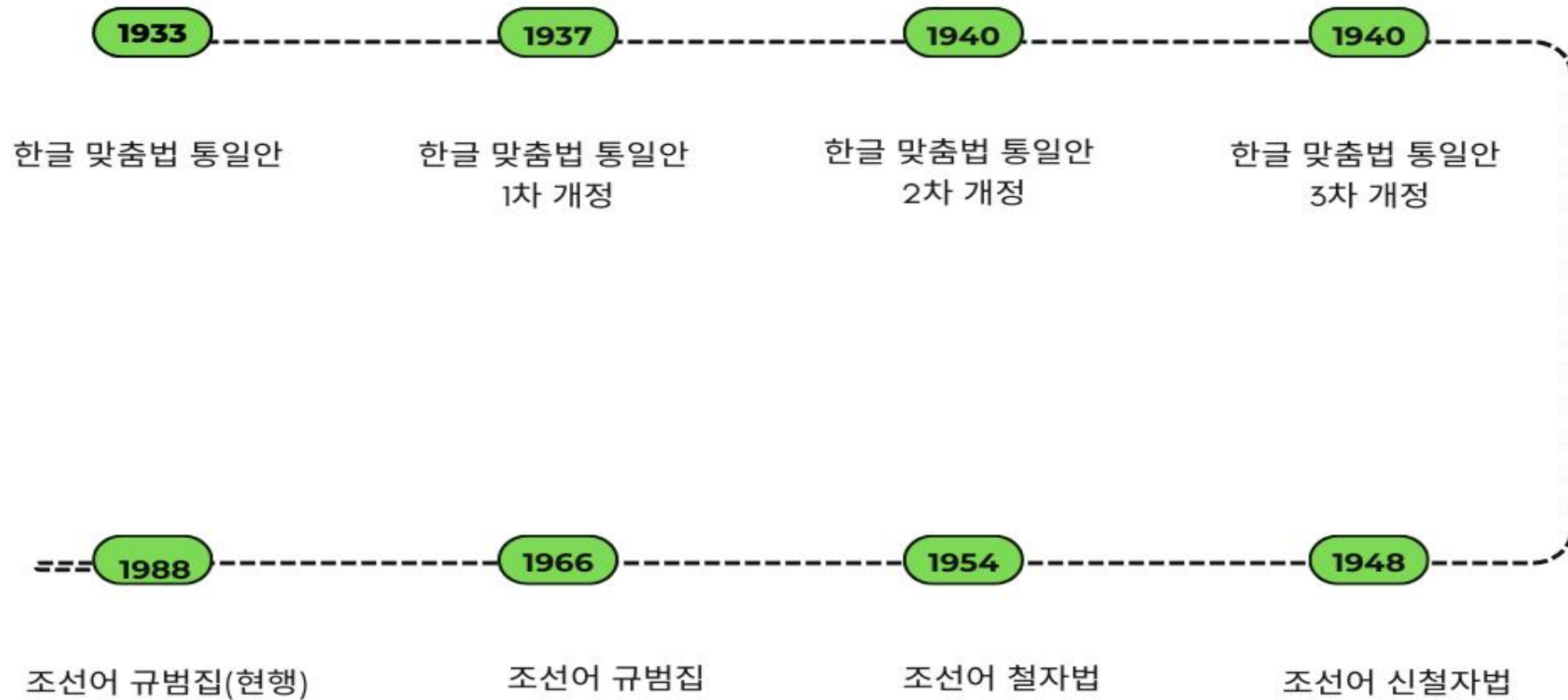
연구목적

- 한국어와 중국 조선어의 주요 언어 규범 차이를 체계적으로 비교하여 텍스트 전환 모형을 제안.
- 두음법칙, 사이시옷, 띄어쓰기와 같은 핵심 규칙을 반영하여 보다 정교하고 효율적인 텍스트 전환 모형 개발.
- NLP 관점에서 언어의 자동 변환 가능성을 제시하고, 언어 통합과 교류의 기반 마련.

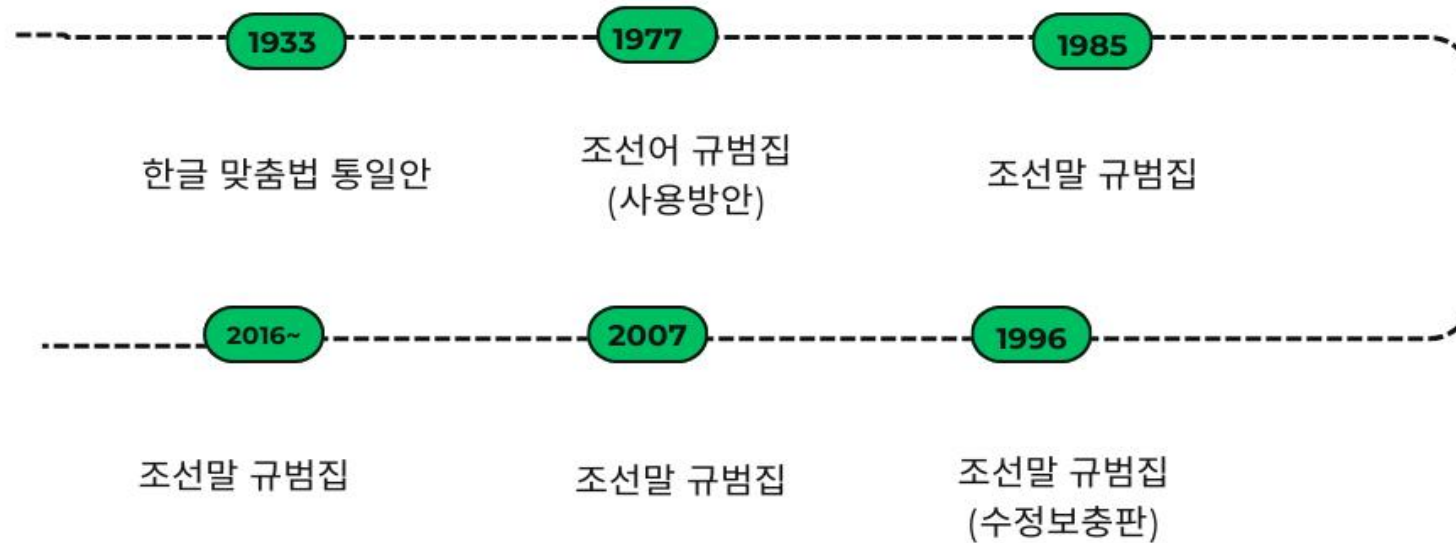
한국



조선



중국 조선어



언어규범에 대한 분석

02 언어규범에 대한 분석

한국	1933. 10	1937. 3	1940. 6	1946. 9	1979. 12	1984. 12		1988. 1	2017. 3	
	한글 맞춤법 통일안	한글 맞춤법 통일안 1차개정	한글 맞춤법 통일안 2차개정	한글 맞춤법 통일안 3차개정	맞춤법안	맞춤법 개정안		한글 맞춤법	한글 맞춤법 일부 개정	
조선	1933. 10	1937. 3	1940. 6	1946. 9	1948. 1	1954. 9		1966. 6	1988. 2	
	한글 맞춤법 통일안	한글 맞춤법 통일안 1차개정	한글 맞춤법 통일안 2차개정	한글 맞춤법 통일안 3차개정	조선어 신철 자법	조선어 철자법		조선말 규범 집	조선말 규범집	
중국	1933. 10	1937. 3	1940. 6	1946. 9	1977. 11	1985. 1	1996. 7	2007. 9	2016-	
	한글 맞춤법 통일안	한글 맞춤법 통일안 1차개정	한글 맞춤법 통일안 2차개정	한글 맞춤법 통일안 3차개정	조선말 규범 집 (시용방안)	조선말 규 범집	조선말 규 범집 (수정보충 판)	조선말 규범집	조선말 규범 집	

두음법칙

- 한국의 맞춤법: 첫소리가 “ㄴ, ㄹ”인 한자가 단어의 첫머리에 올 때, 현실 발음에 따라 “ㄴ, ㄹ”이 탈락되거나 변한 소리대로 표기.
- 중국 조선어 맞춤법: 단어의 첫머리이건 아니건 어느 경우에도 일정하게 한자의 본음 “ㄴ, ㄹ”을 밝혀 표기.

두음법칙 구분	한국어	중국 조선어
표기 원칙	발음 중심의 표기 (현실 발음 반영)	형태 중심의 표기 (한자의 원형 유지)
적용 여부	두음법칙 적용 → ‘ㄴ, ㄹ’이 탈락되거나 바뀜	두음법칙 미적용
예시	ㄴ 탈락: ‘녀자’ → 여자 (女子) ㄹ 탈락: ‘양심’ → 량심 (良心) ㄹ → ㄴ: ‘로동’ → 노동 (劳动)	‘녀자’ → 녀자 (女子) ‘로동’ → 로동 (劳动)
표기 목적	발음의 자연스러움과 현실 반영	형태의 일관성과 원형 보존 우선

- 한국어: 발음을 기준으로 사이시옷을 삽입하여 표기함
- 중국 조선어: 사이시옷을 표기하지 않고, 형태소의 원형을 유지함

사이시옷 구분	한국어	중국 조선어
표기 원칙	발음 중심(음운주의)	형태 중심(형태주의)
사이시옷 처리	된소리 유도 시 'ㅅ'삽입	사이시옷 전면 생략
예시	바다 + 가 → 바닷가 나무 + 잎 → 나뭇잎	바다 + 가 → 바다가 나무 + 잎 → 나무잎
표기 목적	실제 발음을 반영하여 의미 구분	동일 의미는 동일 형태로 유지

- 수 관형사와 단위 명사 결합 시, 한국어는 띄어쓰기를, 중국 조선어는 붙여쓰기를 원칙으로 한다.

띄어쓰기 구분	한국어	중국 조선어
표기 원칙	'수 관형사 + 단위 명사'는 띄어 씀	항상 붙여 씀
적용 예시	세 개, 두 그릇, 열 컵	세개, 두그릇, 열컵
표기 목적	문법적 독립성 강조 (의존 명사는 독립 어절로 취급)	결합 의미 강조 및 단순화 (분석보다 실용성과 일관성 우선)

3.1 말뭉치 수집

03 연구방법

- 표준국어대사전을 바탕으로 두음법칙, 사이시옷 규칙을 반영한 대규모 말뭉치 구축.
- 총 509,062개 어휘 중 두음법칙 적용 단어 약 13,082개

한국어 올림말	조선어 올림말	품사분류	한자어 표현	뜻 풀이
나각	라각	[명]	螺角	(나각만[-강-]) [명] 《음악1》 소라의 껍데기로 만든 옛 군악
나건	라건	[명]	羅巾	[명]비단으로 짠 수건.
나계	라계01	[명]	螺階	[명]《건설》'나선 계단'을 줄여 이르는 말.
나계	라계02	[명]	螺髻	[명] 【①】소라 껍데기 모양처럼 만든, 아이들의 북상투. 【②】《불교》



3.1 말뭉치 수집

03 연구방법

- 표준국어대사전을 기반으로 사이시옷을 적용한 병렬 말뭉치 구축
- 총 509,062개 어휘 중 사이시옷 단어 약 1,602개 선별.

예시	한국어 올림말	조선어 올림말	첫두음절	뒤초성
1	바닷가	바다가	2	ㄱ
2	깃발	기발	1	ㅂ
3	햇살	해살	1	ㅅ

볼때마다 헛갈리는 한글 맞춤법

사이시옷

3.1 말뭉치 수집

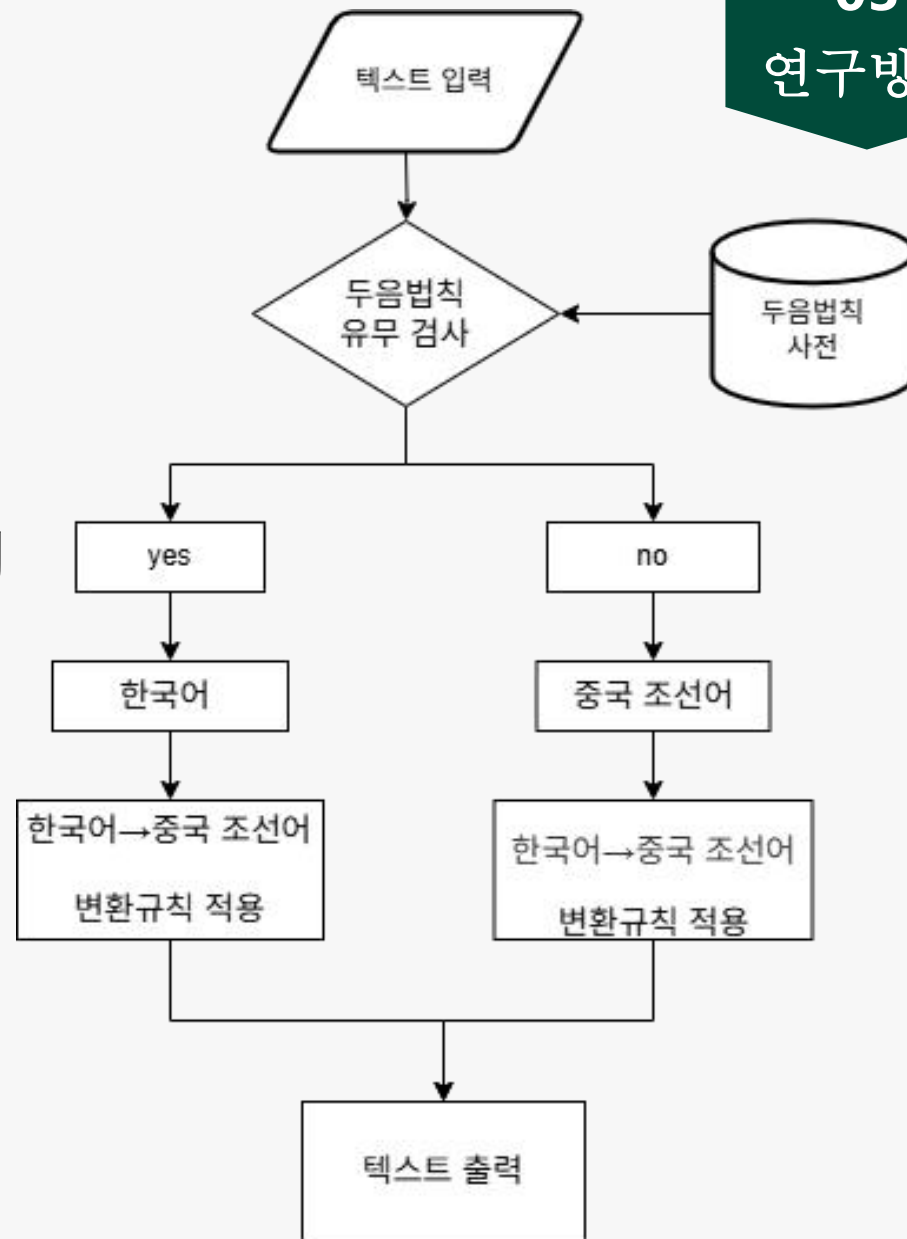
- 띄어쓰기는 『 조선말 규범집 』 에서 추출하였다.

올림 막		해석
냥	[불]	【①】 전날에, 돈을 세는데 쓰는 이름수의 하나. 한돈의 열곱이다. 례: 석~ 두돈. 【②】 귀금속이나 고려약제의 질량을 나타내는 이름수의 하나. 한돈의 열곱이다. 례: 열~ 네돈. 【③】 이 단어를 이름수로 쓰는 일부 명사뒤에 쓰이어 《약간수의 냥》이라는 뜻. 례: 돈~이나 있는 부자.
가지	[불]	【①】 《사물의 하나하나의 부류》를 이르는 이름수의 단위의 하나. 례: 세~ 방법. 여러~ 색깔의 천.

3.2 데이터 처리 및 모형 설계

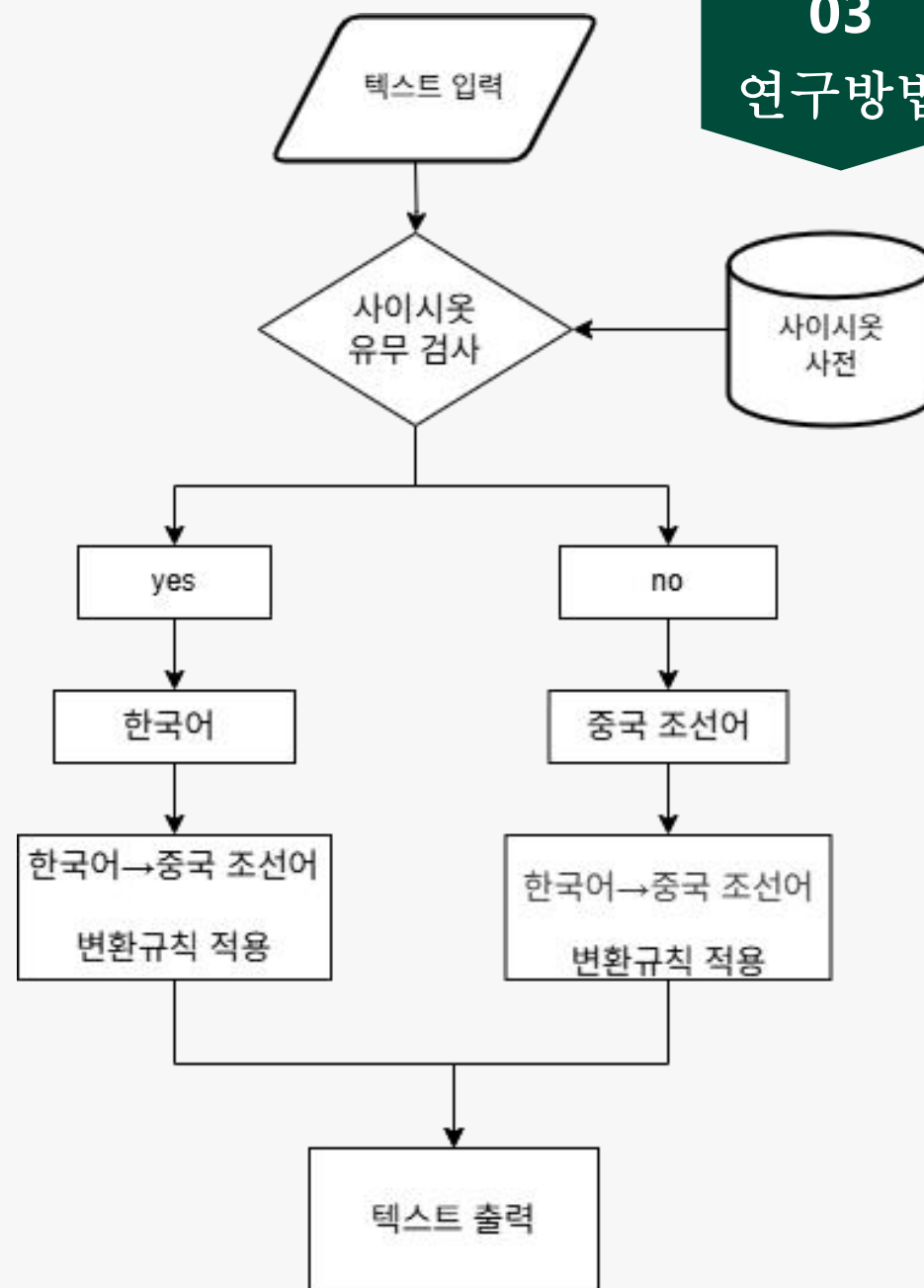
03 연구방법

- 두음법칙 유무를 기반으로 언어를 자동 식별하고, 언어별 전환 규칙을 적용하여 텍스트를 자동 변환함



3.2 데이터 처리 및 모형 설계

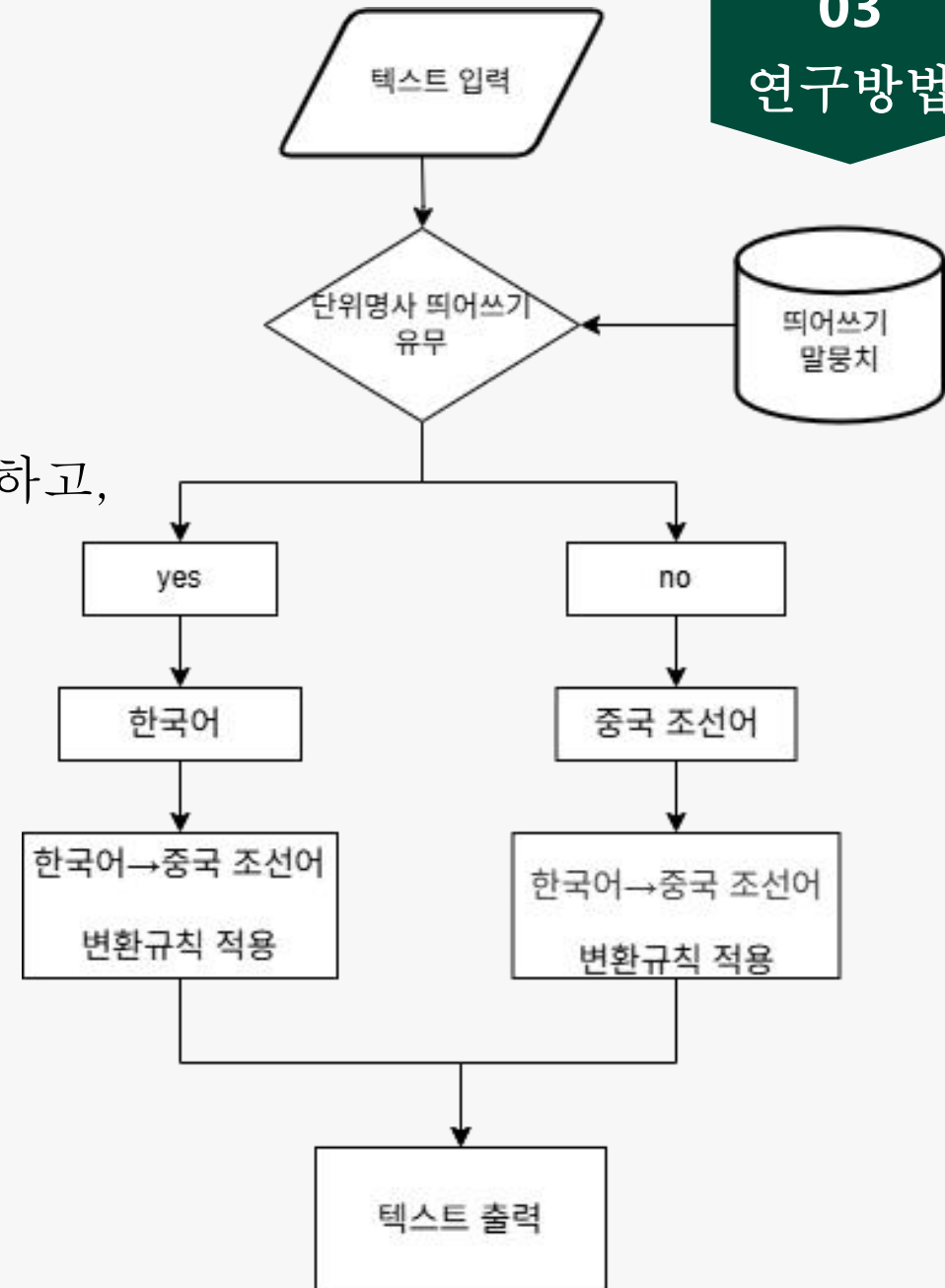
- 사이시옷 유무를 기반으로 언어를 자동 식별하고, 언어별 전환 규칙을 적용하여 텍스트를 자동 변환함



3.2 데이터 처리 및 모형 설계

03 연구방법

- 단위명사 띄어쓰기 유무를 기반으로 언어를 자동 식별하고, 언어별 전환 규칙을 적용하여 텍스트를 자동 변환함



3.2 데이터 처리 및 모형 설계

03 연구방법

한국어/ 중국 조선어 텍스트 전환 모형

두음법칙 사이시옷 띄여쓰기  txt

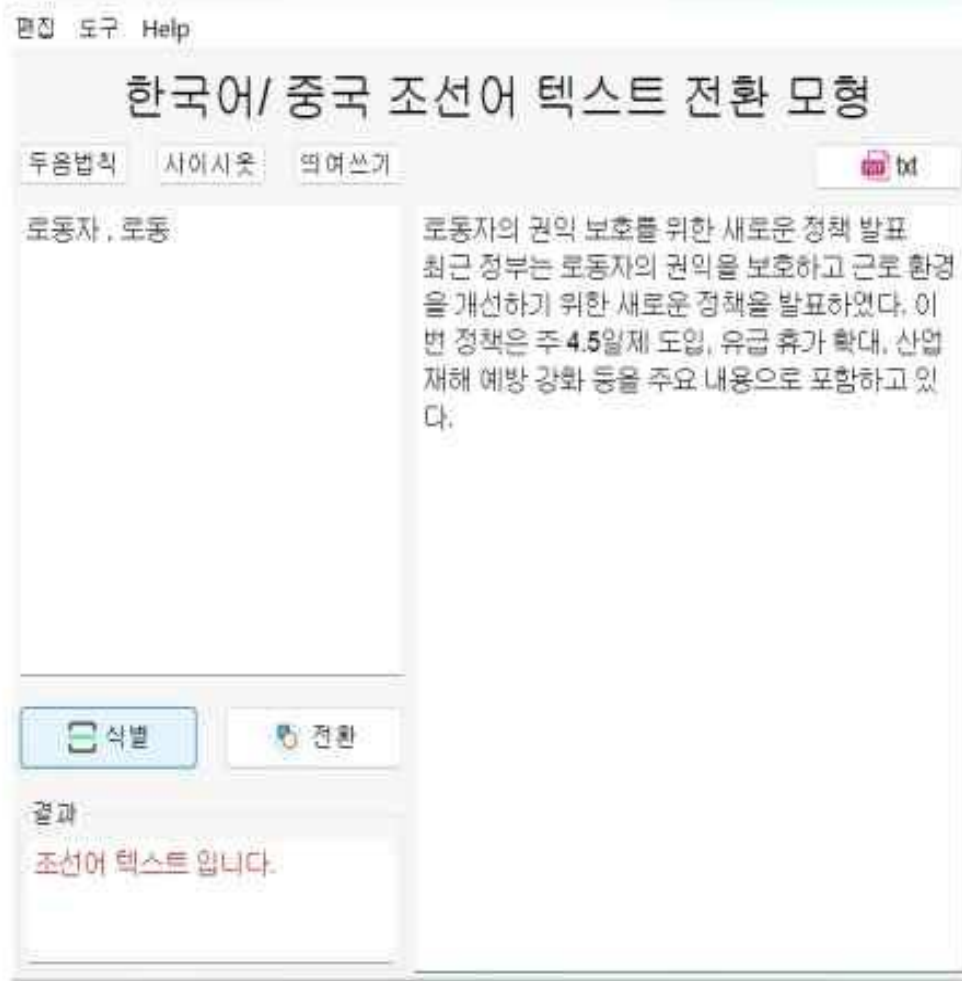
여기에 텍스트를 입력하거나 파일을 가져오세요.

 식별  전환

결과

4.1 두음법칙 전환

04 연구결과



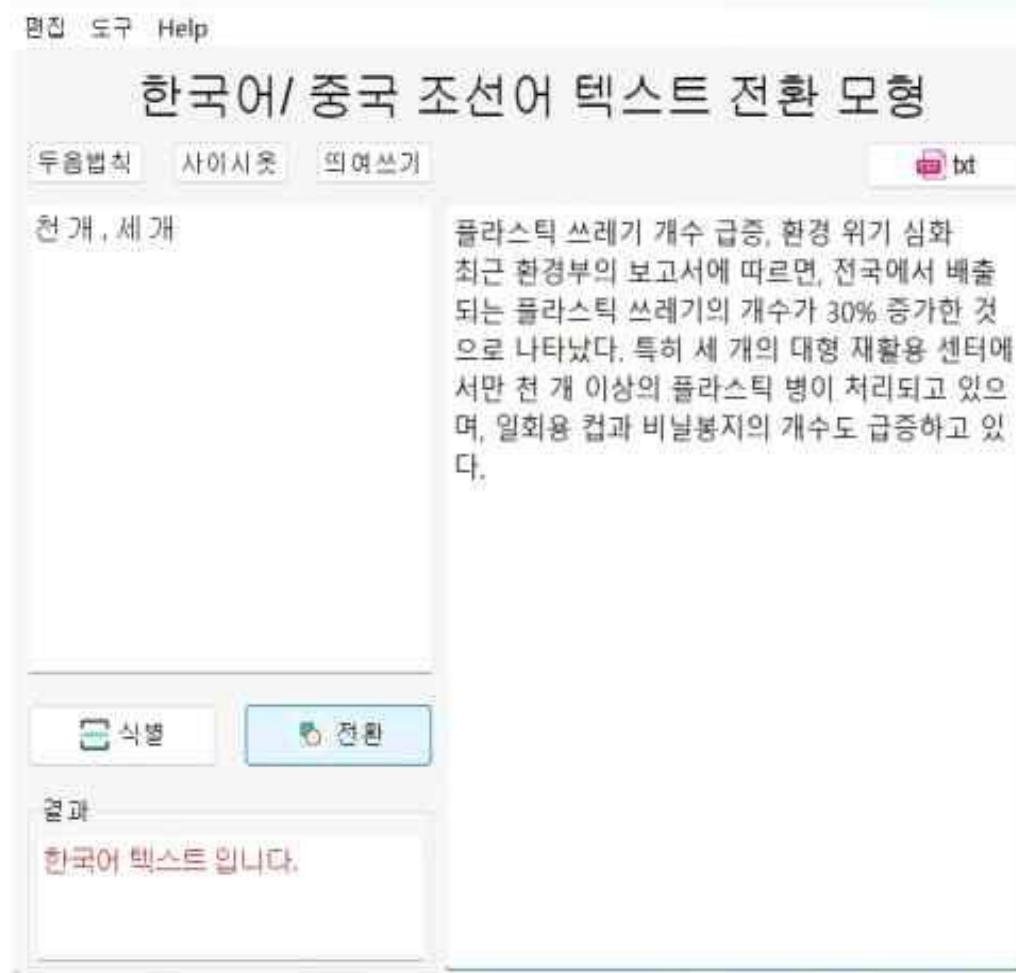
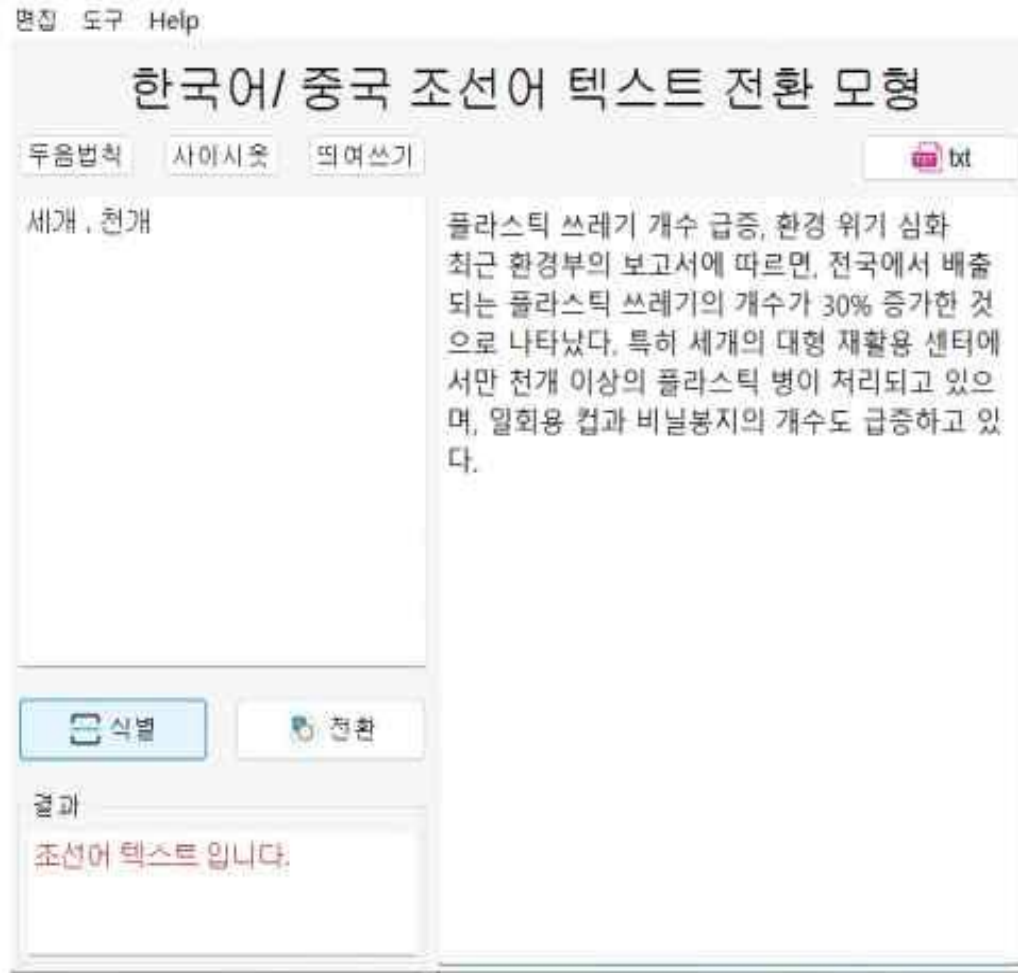
4.2 사이시옷 전환

04 연구결과



4.3 띄어쓰기 전환

04 연구결과



결론 및 향후 과제

- 본 연구는 한국어와 중국 조선어 간의 언어 규범 차이를 반영한 텍스트 전환 모델을 제안함.
- 이를 통해 두 언어 간 자동 변환 및 정보 교환의 효율성을 높였다. 두 언어 간의 효과적인 변환을 가능하게 하여 문서 교정, 기계 번역, 언어 교육 등 다양한 언어 정보화 응용 분야에서 활용할 수 있는 기초를 마련하였다.

향후 과제:

- 신조어 및 방언 포함 확장형 말뭉치를 추가적으로 넣을 것
- 딥러닝 기반의 학습형 전환 모형과의 결합 연구



- Jin, Yongshou. 2010. "중국 조선어 규범화 방향과 규범 원칙에 대한 고찰". 《동강학간》, 27(3): 24-28.
- 고려대학교 민족문화연구원 2009. 『(고려대)한국어대사전』. 서울: 고려대학교 출판부.
- 국립국어연구원 1999. 『표준국어대사전』, 서울: 두산동아.
- 김광수, 강옥설. 2015. "한국과 조선 그리고 중국조선어 맞춤법의 비교연구". 《중국조선어문》. 제2호, 34-59
- 김영수. 2009. "남북 언어 규범의 차이와 중국 조선어 규범 문제", 《한국어교육학회 학술발표논문집》. 2009(1): 73-79.
- 김영수, 김광수, 김철준. 2016. 『조선말규범집』. 연변: 연변교육출판사.
- 김철준, 김광수. 2008. 『조선어문법』. 연변: 연변대학출판사.
- 김철준, 김광수, 지동은, 김영수. 2019. 『조선말규범집해설』. 연변: 연변교육출판사.
- 남기심, 고영근. 1993. 『표준국어문법론』. 서울: 탑출판사.
- 류은중. 1996. "두음법칙에 의한 남북조선의 언어차이 문제(1)(2)". 《중국조선어문》 83:45-67.
- 리광일. 2022. "중국 조선어와 한국어의 띄어쓰기 규칙에 관한 연구" 《중국조선어문》 2022.4:74-81.
- 사회과학출판사. 2006. 『조선말대사전』. 평양: 사회과학출판사.
- 연세대학교. 1998. 『연세한국어사전』. 서울: 두산동아.
- 이석주. 1993. "남북한의 언어 규범 비교". 《한성어문학》 12.0:3-34.
- 최윤갑, 전학석. 1994. 『중국, 조선, 한국 조선어 차이 연구』. 연변: 연변인민출판사.

