

## 딥러닝 기반 한국어 맞춤법 교정을 위한 오류 유형 분류 및 분석

구선민<sup>1</sup>, 박찬준<sup>2</sup>, 소아람<sup>3</sup>, 임희석<sup>4\*</sup>

<sup>1</sup>건국대학교 컴퓨터공학과 학생, <sup>2</sup>고려대학교 컴퓨터학과 석박사통합과정,  
<sup>3</sup>고려대학교 Human-inspired AI연구소 연구교수, <sup>4</sup>고려대학교 컴퓨터학과 교수

### Classification and analysis of error types for deep learning-based Korean spelling correction

Seonmin Koo<sup>1</sup>, Chanjun Park<sup>2</sup>, Aram So<sup>3</sup>, Heuseok Lim<sup>4\*</sup>

<sup>1</sup>Student, Konkuk University, Computer Science

<sup>2</sup>Master & Ph. D. Combined Student, Department of Computer Science and Engineering, Korea University

<sup>3</sup>Research Professor, Human-inspired Computing Research Center, Korea University

<sup>4</sup>Professor, Department of Computer Science and Engineering, Korea University

**요약** 최근 기계 번역 기술과 자동 노이즈 생성 방법론을 기반으로 한국어 맞춤법 교정 연구가 활발히 이루어지고 있다. 해당 방법론들은 노이즈를 생성하여 학습 셋과 데이터 셋으로 사용한다. 이는 학습에 사용된 노이즈 외의 노이즈가 테스트 셋에 포함될 가능성이 낮아 정확한 성능 측정이 어렵다는 한계점이 존재한다. 또한 실제적인 오류 유형 분류 기준이 없어 연구마다 사용하는 오류 유형이 다르므로 질적 분석에 어려움을 겪고 있다. 이를 해결하기 위해 본 논문은 딥러닝 기반 한국어 맞춤법 교정 연구를 위한 새로운 '오류 유형 분류 체계'를 제안하며 이를 바탕으로 기존 상용화 한국어 맞춤법 교정기(시스템 A, 시스템 B, 시스템 C)에 대한 오류 분석을 수행하였다. 분석결과, 세 가지 교정 시스템들이 띄어쓰기 오류 외에 본 논문에서 제시한 다른 오류 유형은 교정을 잘 수행하지 못했으며 어순 오류나 시제 오류의 경우 오류 인식을 거의 하지 못함을 알 수 있었다.

**주제어** : 한국어 맞춤법 교정, 기계 번역, 인공지능경망 기계번역, 오류 분석, 자연어처리

**Abstract** Recently, studies on Korean spelling correction have been actively conducted based on machine translation and automatic noise generation. These methods generate noise and use as train and data set. This has limitation in that it is difficult to accurately measure performance because it is unlikely that noise other than the noise used for learning is included in the test set. In addition, there is no practical error type standard, so the type of error used in each study is different, making qualitative analysis difficult. This paper proposes new 'error type classification' for deep learning-based Korean spelling correction research, and error analysis perform on existing commercialized Korean spelling correctors (System A, B, C). As a result of analysis, it was found the three correction systems did not perform well in correcting other error types presented in this paper other than spacing, and hardly recognized errors in word order or tense.

**Key Words** : Korean spelling correction, Machine translation, Artificial neural network machine translation, Error analysis, Natural language processing

\*This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2018-0-01405) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)" and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2021R1A6A1A03045425).

\*Corresponding Author : Heuseok Lim(limhseok@korea.ac.kr)

Received October 29, 2021

Revised November 23, 2021

Accepted December 20, 2021

Published December 28, 2021

## 1. 서론

맞춤법 교정(Spelling error correction)이란 주어진 문장에서 오류를 감지하고 이를 올바른 문장으로 교정하는 시스템이다. 현재까지 대부분의 한국어 맞춤법 교정 연구는 규칙 기반(rule-based)과 통계 기반(statistical-based)의 연구로 진행되어 왔다. 규칙 기반 방법론은 다양한 종류의 오류 교정 규칙을 미리 정의하여 교정하는 방법이다[1]. 규칙 기반은 문장 구조를 파괴하지 않고 맞춤법 및 문법 오류를 교정할 수 있다는 장점이 있으나 규칙에 정의된 것만 교정할 수 있으며 구축하는데 많은 시간과 비용이 필요하다는 단점이 있다.

통계 기반 방법론은 주어진 문장의 각 단어가 오류일 가능성을 확인하고 대치어 후보들에 대한 확률을 구해 오류어 보다 대치어의 확률이 더 크면 해당 단어를 판단하는 방법이다[2,3]. 그러나 높은 수준의 분석 성능을 위해 많은 양의 데이터가 필요하다는 단점이 존재한다.

최근 기존 방법론의 시간적, 비용적, 데이터적 문제를 해결하기 위해 기계 번역(Machine Translation) 기술과 자동 노이즈 생성(Automatic noise generation) 방법론을 기반으로 한국어 맞춤법 교정(Korean spelling error correction) 연구가 활발히 이루어지고 있다. 모델을 훈련하기 위해 공개된 병렬 말뭉치가 존재하지 않는다는 한계점을[4-6] 완화하기 위하여 자동 노이즈 생성 방법을 사용하여 pseudo 병렬 말뭉치를 생성하고 학습 셋과 테스트 셋으로 분리하여 모델 훈련 및 검증을 진행하고 있다[7].

즉 기존 연구들은 한국어 맞춤법 교정기의 성능을 양적 분석을 통해 수행하는 경우가 대부분이다. 명확한 오류 유형 체계가 없기 때문에 해당 방법론들은 일부 유형에 대한 자동 노이즈 생성 방법으로 생성한 병렬 말뭉치를 훈련과 테스트에 모두 사용한다. 이는 훈련에 사용된 오류 유형 외의 오류 유형을 테스트 할 수 없으므로 정확한 성능 측정이 어렵다는 한계점이 존재한다.

본 논문에서는 한국어 맞춤법 교정 연구를 위해 제안한 ‘오류 유형 분류 체계’를 바탕으로 기존에 존재하는 대

표적인 한국어 맞춤법 교정기 3가지(시스템 A, 시스템 B, 시스템 C)<sup>1)2)3)</sup>의 오류 유형 질적 분석을 수행한다.

한국어 맞춤법 교정 연구를 위한 오류 유형 분류 체계를 만들기 위해 오류 유형은 4가지 대분류(띄어쓰기 오류, 문장부호 오류, 수사 오류, 맞춤법 및 문법 오류)와 하위분류를 포함하여 총 23가지의 세부 오류 유형으로 구분하였다. 해당 분류를 바탕으로 오류 문장에 대해 각 교정기가 수행한 교정 문장에 대해 질적 분석을 수행하였다.

검증을 위한 데이터로 공개된 병렬 문장과 임의로 오류 분류가 포함된 예시 문장을 생성하여 데이터 셋으로 활용하였다. 여러 오류 유형이 혼합된 경우와 혼합되지 않은 경우의 문장에 대해 각 교정기가 수행한 교정 문장을 비교 분석하였다. 이를 바탕으로 각 교정기별 오류 유형에 따른 교정 성능을 질적 분석하였다.

본 연구의 기여는 다음과 같다. 첫째, 본 연구에서는 기존의 기계 번역 기반 한국어 맞춤법 교정 연구가 가진 데이터 구축 및 성능 평가의 한계를 지적하고 최초로 ‘오류 유형 분류 체계’를 제안하여 객관적인 gold test set을 구축할 수 있는 초석을 마련하였다. 둘째, 오류 유형 분류 체계를 기반으로 기존에 존재하는 대표적인 한국어 맞춤법 교정기 3가지가 수행한 교정 문장에 대해 질적 분석을 수행하여 구체적인 오류 유형별 교정 성능을 평가할 수 있도록 하였다.

본 논문은 2장에서 맞춤법 교정 및 자동 노이즈 생성 기법 관련 연구들을, 3장에서 ‘오류 유형 분류 체계’에 대해 설명한다. 4장에서는 본 연구에서 진행하는 태스크, 활용한 데이터 및 각 한국어 맞춤법 교정기에 대한 설명과 함께 질적 분석 수행 결과에 대해 언급하며 오류 유형 분류 체계 관점에서 결과를 분석하였다. 5장에서 결론으로 본 논문을 마무리한다.

## 2. 관련 연구 및 배경지식

### 2.1 한국어 맞춤법 교정

한국어 맞춤법 교정은 주로 written-style text를

1) Pusan National University Korean spelling corrector. <http://speller.cs.pusan.ac.kr/>

2) Daum Korean spelling corrector. [https://alldic.daum.net/grammar\\_checker.do](https://alldic.daum.net/grammar_checker.do)

3) Naver Korean Spelling corrector Beta. <https://bit.ly/3CtaPl4>

다루며 규칙 기반(rule-based) 통계 기반(statistical-based), 인공지능명 기계 번역(neural machine translation) 순서로 발전되어 왔다.

규칙 기반 방법론은 정의한 오류 교정 규칙에 따라 오류를 교정하는 방법이다. 문장이 들어오면 오류를 인식하여 정의해놓은 교정 규칙에 맞게 교정한다. 예를 들어 <‘ㄴ’, ‘이’, ‘쁘’, ‘예’> 규칙은 이전 문맥에 ‘ㄴ’와 다음 문맥에 ‘쁘’가 있으면 ‘이’를 ‘예’로 교정한다[8]. 그러나 구축하는데 많은 시간이 필요하다는 단점이 있다.

통계 기반 방법론은 대표적으로 Shannon이 발표한 noisy channel model과, n-gram language model이 있다[9,10]. Noisy channel model은 자연어처리 문제를 디코딩 문제로 간주한다. N-gram 임베딩을 통해 unknown-word에 대해서도 임베딩을 생성할 수 있다는 장점이 있으나 많은 양의 데이터가 필요하다는 단점이 있다.

기계 번역 방법론이란 기존 방법론의 시간, 비용, 데이터의 한계를 완화시킬 수 있는 방법이다. 기계 번역 관점에서 맞춤법 교정은 오류 문장(error sentence)을 정답 문장(correct sentence)으로 ‘번역(translation)’ 하는 것이다. 주로 Sequence to sequence model (eg. RNN or CNN or Transformer)을 기반으로 noising Encoder와 Denoising Decoder 구조를 통해 교정한다[11-13]. input으로 들어온 error sentence 를 Noising Encode Module을 거쳐 vectorize된 값을 Denoising Decoder Module에서 noise를 제거하여 교정하는 방식이다.

### 2.2 자동 노이즈 생성 기법

자동 노이즈 생성 기법이란 병렬 말뭉치를 생성하기 위해 단일 말뭉치에 대해 몇 가지 노이즈 생성 방법들을 적용하여 병렬 말뭉치를 자동으로 생성하는 기법이다. 자동 노이즈 생성을 통해서 단일 말뭉치만으로 학습을 위한 병렬 말뭉치 생성이 가능하다.

모델 학습과 평가를 위한 병렬 말뭉치가 많지 않은 한국어 맞춤법 교정을 위해서도 자동 노이즈 생성 기법을 사용한다[14,15]. 한국어 맞춤법 교정 연구를 위한 오류 유형 체계가 없고 특정 오류 유형에 대한 연구 [16,17]가 대부분이다. 따라서 띄어쓰기 오류, 문장부호 오류, 음운 변형, 발음 오류, 맞춤법 및 문법 오류 중 몇 가지를 선택하여 올바른 문장에 노이즈를 추가하는

방법을 사용한다. 하지만 연구마다 정의하는 오류 유형이 다르며 생성한 데이터 셋에서 훈련 셋과 테스트 셋을 분할하여 사용하기 때문에 훈련에 사용된 오류 유형이 테스트에 포함될 확률이 높으므로 객관적인 연구가 어렵다.

### 3. 제안하는 오류 유형 분류 체계

한국어 맞춤법 교정 연구의 객관적인 비교 연구를 위한 gold test set 구축을 위한 초석으로 ‘오류 유형 분류 체계’를 제안한다. 제안하는 분류 체계는 Table 1과 같다. 4가지의 대분류(띄어쓰기 오류, 문장부호 오류, 수사 오류, 맞춤법 및 문법 오류)를 기준으로 총 23가지의 세부 오류 유형으로 구분하였다. 특히 맞춤법 및 문법 오류는 1차 오류와 2차 오류로 나누어 세부적인 분석이 가능하도록 하였다.

띄어쓰기 오류란 한국어 띄어쓰기 규칙에 위배되는 경우로 정의한다. 딥러닝 모델뿐만 아니라 사람의 경우도 빠른 타이핑 속도나 습관에 따라 자주 나타날 수 있는 오류이다.

문장부호 오류란 한국어 문장 사이에 문장 부호가 부착되지 않거나 잘못된 위치에 부착된 경우로 정의한다. 문장부호 오류는 딥러닝 모델에서 문장을 생성할 때 미등록어로 인하여 일어나기 쉬운 오류이다.

수사 오류는 양을 나타내는 양수사와 순서를 나타내는 서수사가 오류를 일으킨 경우로 정의한다. 예를 들어 올바른 문장인 ‘한 시 일 분’을 ‘하나 시 일 분’이나 ‘일시 일분’으로 잘못 표기한 경우가 해당된다.

맞춤법 및 문법 오류는 한국어 맞춤법 및 문법에 위배되는 경우로 정의한다. 한국어에서 가장 빈번하게 발생하는 경우로 1차 오류와 2차 오류로 구분하였다. 1차 오류와 2차 오류는 중첩 오류가 가능하다. 1차 오류로 먼저 분류한 후 일종의 하위분류로서 2차 오류를 진행하여 더욱 세부적이고 구체적인 분류가 가능하다.

1차 오류는 단일언어 오류와 다중언어 오류로 구분하였다. 단일언어 오류는 한국어 문장 내에서의 오류로 4가지 오류 유형(삭제 오류, 추가 오류, 교체 오류, 분리 오류)로 세분화하였다. 교체 오류는 단어가 다른 단어로 교체되는 단어 교체와 한 음절 안에서 음운의 순서가 변경되는 순서 교체로 세분화하였다.

Table 1. Error Type Classification Scheme

Error Type			Explanation		
Spacing Error			In case of violating the spacing rules		
Punctuation Error			In case punctuation marks are not attached between Korean sentences or are attached in the wrong place		
Numerical Error			In case the positive number indicating quantity and the ordinal number indicating the order are in error		
Spelling and Grammar Error	Primary Error	Monolingual Error	Remove Error	In case some words are not recognized or endings or suffixes are omitted	
			Addition Error	In case the same word is repeated or an unused participle or ending is added	
			Replace Error	Word replace	In case a word is replaced by another word
				Rotation replace	In case the order of phonological changes within a syllable
		Separation Error	Separating consonants and vowels in letters		
	Multilingual Error	Typing language Error	In case writing with a keyboard other than Korean		
		Foreign word conversion Error	In case written differently from the standard foreign language pronunciation		
	Secondary Error	Spelling Error	Consonant vowel conversion error	In case a spelling error in non-speaking alphabet units	
			Grapheme-to-phoneme(G2P) Error	In case write letters according to pronunciation	
		Syntax Error	Phrase structure rule Error	In case not match the Korean word order	
			Tense Error	In case use a verb that doesn't fit the tense	
			Postposition Error	In case use a probing that doesn't fit the grammar	
			Suffix Error	In case use an ending that is not gram-matical	
			Auxiliary predicate Error	In case use an auxiliary verb that is not grammatically correct	
		Semantic Error	Dialect Error	In case writing using non-standard language	
			Polite speech Error	An adjective expression that does not fit the subject	
			Word sense ambiguity Error	In case the spelling is the same but has two or more meanings	
			Coreference Error	In case Invalid entity reference	
		Discourse context Error	In case does not fit the context of the previous discourse		
		Neologism Error	In case use grammar or new words that are not in the existing grammar system		

다중언어 오류는 한국어 외 다른 언어로 인하여 발생한 오류로 정의하며 2가지 오류 유형(타이핑 언어 오류, 외래어 변환 오류)으로 세분화하였다. 2차 오류의 경우 4가지 유형(철자 오류, 구문 오류, 의미 오류, 신조어 오류)으로 세분화하여 분류하였다.

제안한 분류 체계를 적용하여 효율적이고 명확한 오류 유형 분석과 객관적인 비교 연구가 가능하도록 하였으며 대표적인 한국어 맞춤법 교정기 3가지(시스템 A, 시스템 B, 시스템 C)에 대해 질적 분석을 통해 교정 성능을 평가하였다.

#### 4. 오류 분석

##### 4.1 데이터

본 논문은 깃허브에 공개되어 있는 한국어 맞춤법

교정 데이터 셋의 예시 문장을 활용하여 분석을 진행하였다<sup>4)</sup>. 데이터는 총 3000개의 오류 문장과 정답 문장으로 이루어져 있다. 오류 문장을 test-src, 정답 문장을 test-tgt로 구분하여 제공한다. 오류 문장에는 한 개 이상의 오류 유형이 포함되어 있으며 정답 문장은 이를 한국어 맞춤법에 맞게 교정한 문장이다. 오류 문장은 제안한 오류 유형 체계 중에서 띄어쓰기 오류와 교체 오류 및 철자 오류 등으로 구성되어 있다. 해당 병렬 데이터 셋을 활용하여 한국어 맞춤법 교정기의 교정 성능을 판단하는 데 사용한다. 우리는 그중 랜덤으로 뽑은 11 문장과, 선택한 문장에 포함되지 않은 오류 유형이 포함되도록 21문장을 임의로 생성하여 총 32문장에 대해 질적 분석에 사용하였다.

4) Korea-Spelling-Correction-Resource. <https://github.com/Parkchanjun/Korea-Spelling-Correction-Resource>

## 4.2 한국어 맞춤법 교정기

본 논문에서는 대표적인 상용화 한국어 맞춤법 교정기 3가지(시스템 A, 시스템 B, 시스템 C)에 대해 각 교정기가 수행한 교정 문장으로 질적 분석을 수행하였다.

**시스템 A**는 오류 문장에서 오류를 인식하고 해당 오류에 대한 대치어 후보를 제시하여 사용자가 선택하여 교정할 수 있도록 한다. 철자 검사/교정기에서 띄어쓰기, 붙여쓰기, 음절 대치와 같은 교정 방법에 따라 오류를 수정하여 결과를 출력한다. 인식한 오류에 대해 어떤 오류인지 제시하고 대치어와 함께 일반적인 문법 사용 예시를 같이 제시한다.

**시스템 B**는 오류 문장에서 오류를 인식하고 오류를 4가지 유형(맞춤법, 표준어 의심, 띄어쓰기, 통계적 교정)으로 분류하여 교정을 수행한 후 사용자에게 교정 결과를 제공한다. **시스템 B**에서 분류한 오류 유형 중 '표준어 의심'은 오류로 인식하였으나 올바른 교정 방법을 찾지 못해 교정하지 못한 경우를 표현한다. 각 오류 유형별 교정 결과를 다른 색상으로 구분하여 사용자가 어떤 유형의 오류 문장을 입력하여 교정하였는지 쉽게 인식할 수 있도록 한다.

**시스템 C**는 오류 문장에서 오류를 인식하여 교정을 수행한 후 사용자에게 교정 결과를 제시한다. 인식한 오류에 대해 교정할 수 있는 오류는 교정하여 제시하며 오류는 인식했으나 교정할 수 없는 경우는 '오류 의심'으로 표기한다. 교정을 수행한 단어와 오류 의심 단어를 다른 색으로 표기하여 한눈에 볼 수 있도록 한다. 특히 띄어쓰기 오류의 경우 교정 기호(✓)를 사용하여 직관적으로 확인할 수 있도록 한다.

## 4.3 분석 결과

### 4.3.1 한국어 맞춤법 교정기 모델별 성능 비교

본 논문에서는 오류 문장에 대해 교정기가 수행한 교정 문장을 비교 분석하였다. 분석 결과 전체적인 성능은 **시스템 A**와 **시스템 C**가 비슷하게 좋으며 **시스템 B**가 약간 떨어진다. Table 2에서 질적 분석 예시를 보여준다. 오류 문장과 정답 문장 및 한국어 맞춤법 교정기들이 수행한 교정 결과와 인식한 오류에 따른 바르게 교정한 수를 제시하였다.

예를 들어 오류 문장이 '국내에서 유일하게 법정

커버드본드를 발행할 수 있는 주택금융공사는 지난해 7월 아시아 최총로 5억 달러 규모의 커버드본드를 발행했다' 이고 정답 문장이 '국내에서 유일하게 법정 커버드본드를 발행할 수 있는 주택금융공사는 지난해 7월 아시아 최초로 5억 달러 규모의 커버드본드를 발행했다'인 경우가 있다. 해당 오류 문장의 오류인 '국넷에서'와 '주택금융공상 은', '최총로'를 단어 교체, 자모 변환오류로 분류할 수 있으며 문장부호 오류로도 분류할 수 있다.

**시스템 A**의 경우 오류 문장을 '국내에서 유일하게 법정 커버드본드를 발행할 수 있는 주택금융 공사는 지난해 7월 아시아 최총로 5억 달러 규모의 커버드본드를 발행했다.'로 교정했다. **시스템 B**의 경우 '국내에서 유일하게 법정 커버드본드를 발행할 수 있는 주택금융 공사는 지난해 7월 아시아 최총로 5억 달러 규모의 커버드본드를 발행했다'로 교정했다. **시스템 C**의 경우 '국내에서 유일하게 법정 커버드본드를 발행할 수 있는 주택금융 공사는 지난해 7월 아시아 최총로 억 달러 규모의 커버드본드를 발행했다'로 교정했다.

**시스템 A**의 경우 '국넷에서'와 '주택금융공상는', '최총로', 문장부호를 오류로 인식하였다. '국넷에서'의 대치어를 '국내에서'로 제시하여 맞춤법 및 문법 오류 중 단어 교체, 자모 변환 오류에 해당하는 오류를 인식하고 올바르게 교정하였음을 알 수 있다. '주택금융공상는'의 대치어로 '주택금융 공상은'을 제시하여 띄어쓰기 오류와 조사 오류로 분류하여 잘못 교정하였음을 알 수 있다. '최총로'의 경우 철자 분석 결과 어절 분석을 수행하지 못했기 때문에 오류를 인식하였으나 대치어 후보를 제시하지 못했다. 또한 데이터 셋에서 다루지 않은 문장 부호 오류를 인식하여 문장 끝에 온점(.)을 붙혀 대치어를 제시하였다. 따라서 4가지 오류를 인식하여 그중 2가지를 올바르게 교정하였다.

**시스템 B**의 경우 '국넷에서'와 '주택금융공상는', '최총로'를 오류로 인식하였다. 단어 교체 및 자모 변환 오류인 '국넷에서'와 '주택금융공상는'을 통계적 교정을 통해 '국내에서'와 '주택금융공사는'으로 올바르게 교정하였다. '최총로'의 경우 오류로 인식하였으나 교정하지 못하고 '표준어 의심'으로 표시하여 제시하였다. 따라서 3가지 오류를 인식하여 그 중 2가지를 올바르게 교정하였다.

Table 2. Qualitative Analysis Example Table: Correct answer correction rate is the number of correct corrections for the number of detected errors included in the error sentence

Error sentence	Correct sentence	System A	System B	System C	correct correction rate (A, B, C)
국내에서 유일하게 법정 커버드본드를 발행할 수 있는 주택금융공사는 지난해 7월 아시아 최초로 5억 달러 규모의 커버드본드를 발행했다	국내에서 유일하게 법정 커버드본드를 발행할 수 있는 주택금융공사는 지난해 7월 아시아 최초로 5억 달러 규모의 커버드본드를 발행했다	국내에서 유일하게 법정 커버드본드를 발행할 수 있는 주택금융 공사는 지난해 7월 아시아 최초로 5억 달러 규모의 커버드본드를 발행했다.	국내에서 유일하게 법정 커버드본드를 발행할 수 있는 주택금융공사는 지난해 7월 아시아 최초로 5억 달러 규모의 커버드본드를 발행했다	국내에서 유일하게 법정 커버드본드를 발행할 수 있는 주택금융 공사는 지난해 7월 아시아 최초로 5억 달러 규모의 커버드본드를 발행했다	2/4, 2/3, 1/3
남자프로농구 울산 모비스의 양동근이 11월의 선수에 선정됐다	남자프로농구 울산 모비스의 양동근이 11월의 선수에 선정됐다	남자프로농구 울산 모비스의 양동근(인) 11월의 선수에 선정됐다.	남자 프로농구 울산 모비스의 양동근이 11월의 선수에 선정됐다	남자 프로농구 울산 모비스의 양동근 이 11월의 선수에 선정됐다	2/3, 2/2 1/2
메이저리그에섯 계속해서 뛰겠다는 의욕이 강한 데다 메이저리그 구단들의 관심이 여전하기 때문이다	메이저리그에서 계속해서 뛰겠다는 의욕이 강한 데다 메이저리그 구단들의 관심이 여전하기 때문이다	메이저리그에 값 계속해서 뛰겠다는 의욕이 강한 데다 메이저리그 구단들의 관심이 여전하기 때문이다	메이저리그에섯 계속해서 뛰겠다는 의욕이 강한 데다 메이저리그 구단들의 관심이 여전하기 때문이다	메이저리그에섯 계속해서 뛰겠다는 의욕이 강한 데다 메이저리그 구단들의 관심이 여전하기 때문이다	1/3, 0/3, 2/3
드룩빠의 분저스로 일보는 코트디부아르에 석패했다	드룩바의 분전으로 일보는 코트디부아르에 석패했다	드로그바의 분전으로 일보는 코트디부아르에 석패했다	드로 빠의 분저스로 일보는 코트디부아르에 석패했다	드룩빠의 분전으로 일보는 코트디부아르에 석패했다	2/3, 0/3, 1/2
한편 치될 한 달가 노프니벤트를 진행한다	한편 7월 한 달간 오픈이벤트를 진행한다	한편 치를 한 달 가 높으니 벤트를 진행한다.	한편 치될 한 달가 노프네 벤트를 진행한다	한편 칠월 한 달가 노프이벤트를 진행한다	1/4, 0/3, 2/3

Table 3. Korean spelling correction by error type: Error recognition and performance are indicated by slice. 0 indicates no success. 1 indicates success. 2 indicates intermittent success

Error Type				System A	System B	System C
Spacing Error				1/1	1/1	1/1
Punctuation Error				2/1	0/0	0/0
Numerical Error				0/0	0/0	0/0
Spelling and Grammar Error	Primary Error	Monolingual Error	Remove Error	1/1	1/1	1/1
			Addition Error	0/0	0/0	0/0
			Replace Error	2/1	2/1	2/1
		Multilingual Error	Word replace	2/0	0/0	0/0
			Rotation replace	2/0	0/0	0/0
			Separation Error	1/1	0/0	0/0
	Secondary Error	Spelling Error	Typing language Error	0/0	0/0	0/0
			Foreign word conversion Error	2/1	2/1	2/1
	Secondary Error	Spelling Error	Consonant vowel conversion error	2/2	2/2	2/2
			Grapheme-to-phoneme(G2P) Error	1/1	2/1	1/1
		Syntax Error	Phrase structure rule Error	0/0	0/0	0/0
			Tense Error	2/2	2/2	2/2
			Postposition Error	1/1	0/0	0/0
			Suffix Error	1/1	0/0	0/0
Auxiliary predicate Error			0/0	0/0	0/0	
Semantic Error		Dialect Error	1/2	0/0	0/0	
		Polite speech Error	0/0	0/0	0/0	
		Word sense ambiguity Error	-	-	-	
	Coreference Error	-	-	-		
Secondary Error	Neologism Error	Discourse context Error	-	-	-	
		Neologism Error	2/1	0/0	0/0	

**시스템 C**의 경우 ‘국넷에서’와 ‘주택금융공사는’, ‘최출로’를 오류로 인식하였다. ‘국넷에서’는 ‘국내에서’로 올바르게 교정하였으나 ‘주택금융공사는’은 **시스템 A** 교정 결과와 동일하게 ‘주택금융 공상은’으로 잘못 교정하였다. 또한 ‘최출로’는 교정하지 못하였다. 따라서 3가지 오류를 인식하여 그 중 1가지를 올바르게 교정하였다.

또 다른 예시는 오류 문장인 ‘남자프로농구 울산 모비스윗 양동근잉 11월의 선수에 선정됐다’와 정답 문장인 ‘남자프로농구 울산 모비스의 양동근이 11월의 선수에 선정됐다’가 있다. 해당 오류 문장의 ‘모비스윗’을 단어 교체, 자모변환 오류로 분류할 수 있으며 ‘양동근잉’을 추가 오류, 조사 오류로 분류할 수 있다. 또한 문장부호 오류가 포함되어 있다.

**시스템 A**의 경우 오류 문장을 ‘남자프로농구 울산 모비스의 양동근일(인) 11월의 선수에 선정됐다.’로 교정하였다. ‘모비스윗’과 문장부호 오류를 인식하여 올바르게 교정하였다. 하지만 조사 오류인 ‘양동근잉’의 경우 대치어로 ‘양동근일’과 ‘양동근인’을 제시하여 올바른 교정을 수행하지 못했음을 알 수 있다.

**시스템 B**의 경우 ‘남자 프로농구 울산 모비스윗 양동근이 11월의 선수에 선정됐다’로 교정하였다. ‘남자프로농구’를 띄어쓰기, ‘양동근잉’을 통계적 교정을 통해 교정하였다. ‘모비스윗’의 경우 표준어의심 표시를 통해 오류임을 인식했으나 교정하지 못했음을 알 수 있다.

**시스템 C**의 경우 ‘남자 프로농구 울산 모비스의 양동근 잉 11월의 선수에 선정됐다’로 교정했다. ‘남자프로농구’와 ‘모비스윗’을 교정하였다. ‘양동근잉’의 경우 띄어쓰기 오류로 인식하여 ‘양동근 잉’으로 잘못 교정하였다.

Table 3에서 교정기 별로 교정을 실패한 오류를 제안한 오류 유형에 대해 제시하였다. ‘/’를 기준으로 왼쪽에는 오류 인식, 오른쪽에는 오류 교정 성공을 나타내었다. ‘0’은 수행하지 못한 것을 나타낸다. ‘1’은 수행에 성공한 것을 나타낸다. ‘2’는 간헐적으로 수행하는 것을 나타낸다. 띄어쓰기 오류는 3가지 교정기 모두 오류 인식과 교정을 잘 수행하였다. 문장부호 오류의 경우 **시스템 A**만 간헐적으로 오류를 인식하고 교정을 수행하였다. **시스템 B**와 **시스템 C**는 오류를 인식하지 못하여 교정하지 못하였다. 수사 오류의 경우 3가지 교정기

모두 수행하지 못했다. 예를 들어 ‘한시 셋 분’을 넣었을 때 **시스템 A**는 사람을 존칭하여 나타내는 ‘분’으로 인식하여 교정하려고 하였으며 **시스템 B**와 **시스템 C**는 오류를 인식하지 못했다. 삭제 오류는 3가지 교정기 모두 잘 수행하였다. 추가 오류는 ‘예전에 있었었다.’를 넣었을 때 3가지 모두 수행하지 못했다. 교체 오류 중 단어 교체는 3가지 교정기 모두 간헐적으로 오류를 인식하면 교정을 잘 수행하였다. 순서 교체의 경우 ‘날씨가 좋다.’를 넣었을 때 **시스템 A**가 오류를 감지하였으나 ‘날씨가’가 아닌 ‘날씨기’로 잘못된 대치어 후보를 제시하였다. **시스템 B**와 **시스템 C**의 경우 오류를 감지하지 못했다. 분리 오류의 경우 **시스템 A**만 교정을 수행했다. 타이핑 언어 오류의 경우 3가지 교정기 모두 오류를 감지하지 못하였으며, 외래어 변환 오류와 자모변환 오류의 경우 일부 오류만 인식하며 인식한 오류에 대해서는 교정을 잘 수행하였다. 발음 오류는 **시스템 B**는 간헐적으로 오류 인식 및 교정을 수행하였으며 **시스템 A**와 **시스템 C**는 선정한 데이터에 대해 모두 잘 수행하였다. 어순 및 시제 오류는 3가지 모델 모두 수행하지 못하였다. 조사 오류의 경우 격조사는 교정을 잘 수행하지만 보조사나 접속조사는 수행하지 못하였다. 어미 오류는 ‘지금 오아라.’를 넣었을 때 **시스템 A**만 오류 인식 및 교정을 수행하였다. 보조용언 및 높임말 오류는 3가지 모두 수행하지 못하였다. 방언 오류의 경우 **시스템 A**만 올바른 교정을 수행하였다. 중의어 오류 및 참조 오류, 담화 맥락 오류는 분석 데이터가 문장 이하 단위이기 때문에 분석에 제외되었다. 신조어 오류는 **시스템 A**만 간헐적으로 인식 교정을 수행하였다.

Table 3에서 시스템 별 오류 유형 분류에 따른 맞춤법 교정 성능을 나타내었다. ‘/’를 기준으로 왼쪽에는 인식 성능, 오른쪽에는 교정 성능을 표시하였다. 0은 실패, 1은 성공, 2는 간헐적 성공을 나타낸다. 시스템이 오류를 인식해야 교정을 수행할 수 있으므로 오류를 인식하지 못한 경우 교정 성능도 낮게 나타났다. 본 논문에서는 문장 단위의 연구를 수행했기 때문에 단락 단위의 연구가 필요한 오류 유형의 경우 ‘-’으로 표기하였다. 추후 연구에서는 단락 단위로 확장하여 수행할 계획이다.

Table 4에서 데이터의 총 오류 양과 오류 인식 수 및 올바르게 교정한 개수를 볼 수 있다. 총 32문장에 대한 오류 개수는 59개이다. 그중 **시스템 A**가 오류를 인식한 개수는 48개이며 바르게 교정한 개수는 23개이다. 시스템

B가 오류를 인식한 개수는 35개이며 바르게 교정한 개수는 9개이다. 시스템 C의 오류 인식 개수는 29개이며 바르게 교정한 개수는 15개이다. 시스템 A가 문장부호 교정을 지원하므로 다른 교정기보다 인식 개수가 많다. 3가지 교정 모델 모두 띄어쓰기 오류를 인식하고 올바르게 교정한다. 시스템 A의 경우 방언 오류인 '생추(상추의 방언)'에 대해 유일하게 올바르게 교정하였으며 문장부호 오류에서 강세를 보였다. 시스템 C의 경우 받음 오류에서 강세를 보였다. 외래어 변환 오류는 시스템 B, C가 모두 올바르게 교정하였으며 시스템 A의 경우 일부만 교정하였다. 3가지 교정 모델 모두 수사 오류, 자모변환 오류, 어순 오류, 시제 오류, 높임말 오류에 대해서 잘 교정하지 못하였다. 특히 어순 오류와 시제 오류의 경우 오류를 인식하지 못하므로 어절 단위의 분석을 통해 전체적인 맥락 및 구조에 대한 분석이 약하다고 볼 수 있다.

**Table 4. Number of corrections by model. Total error indicates the number of errors included in the entire dataset statement. Total recognition means the total number of errors recognized by the system. Total Correct Correction is the total number of correct errors the system corrects**

Model	Total error	Total recognition	Total correct correction
System A	59	48	23
System B	59	35	9
System C	59	29	15

**Table 5. Performance of corrector model. Recognition indicates the rate at which the system detects errors included in the sentence. correct correction rate indicates the correct correction rate for detected errors**

Model	recognition	correct correction
System A	0.81	0.48
System B	0.59	0.26
System C	0.49	0.52

Table 5는 교정기 모델 별 성능이다. 오류 인식을 및 교정 성공률을 나타낸다. 오류 인식률은 총 오류에 대해 오류 감지를 수행한 비율을 나타낸다. 교정 성공률은 인식 오류를 대상으로 교정을 성공한 비율을 나타낸다.

오류 인식률은 시스템 A, B, C의 순서로 높았으며 교정 성공률은 시스템 C, A, B의 순서로 높았다. 따라서 시스템 C가 오류를 감지하면 올바른 교정을 가장 잘 수행한다고 분석할 수 있다.

#### 4.3.2 특이 유형 세부 분석

같은 오류 유형이 포함되어 있는 오류 문장을 간헐적으로 오류를 인식하여 교정하였기 때문에 어떠한 경우에 교정하는지 알아보기 위해 추가 분석을 수행하였다.

시스템 A가 교정을 수행할 때 문장부호 오류를 항상 교정하지 않았다. 오류 문장을 넣어 어떤 케이스에서 오류를 인식하는지 분석해 보았다. 분석한 결과 마지막 어절에 추가 오류가 있을 경우 문장부호 오류는 분류하지 않고 교정하지 않았다.

또한 공개 병렬 데이터 셋에는 문장 마지막에 문장부호가 부착되어 있지 않았기 때문에 임의로 문장부호가 아예 없는 경우와 올바르게 교정하지 않는 문장부호가 부착되어 있는 경우로 나누어 실험해 보았다.

예를 들어 '밥 먹을까?'라는 오류 문장은 물음표(?)가 부착되어 '밥 먹을까?'라는 정답 문장이 되어야 한다.

시스템 A의 경우 문장부호가 부착되어 있지 않은 경우에는 교정하지 못했다. 또한 문장부호 중 ' / ! / - '가 붙어있는 경우 또한 오류를 인식하지 못하고 교정하지 않았다. '/'가 부착되어 있는 경우 오류를 인식하여 올바르게 교정하였다. 시스템 B와 시스템 C의 경우 문장부호 오류를 수행하지 않는다. 따라서 시스템 A의 문장부호 성능이 뛰어난 편은 아니나 다른 교정기가 지원하지 않는 교정을 수행한다는 것에 의의가 있다.

## 5. 결론

기존 연구들은 훈련에 사용된 노이즈 이외의 노이즈를 테스트할 수 없어 정확한 성능 측정이 어려우며, 오류 유형 분류 체계가 없어 객관적인 비교 연구가 어렵다는 한계점이 존재한다. 이를 해결하기 위해 본 논문에서는 한국어 맞춤법 교정 연구의 객관적인 비교 연구를 위한 gold test set 구축을 위한 초석으로 '오류 유형 분류 체계'를 제안하였다. 제안한 분류 체계에 대해 대표적인 한국어 맞춤법 교정기 3가지(시스템 A, 시스템 B, 시스템 C)가 수행한 교정 문장에 대해 질적 분석하였다. 실험 데이터에 대한 교정 성능은 시스템 C, 시스템 A, 시스템



B 순서로 좋았다. 시스템 C의 경우 오류 인식률은 낮으나 올바른 교정을 수행하는 경우가 오류 인식률 대비 높았다. 시스템 A의 성능이 높은 이유 중 하나는 다른 교정기에서 수행하지 않는 문장부호 오류를 시스템 A에서 수행하는 것이 원인인 것으로 분석할 수 있다. 하지만 문장부호 오류를 간헐적으로 수행하므로 높은 수준의 교정을 수행했다고 판단하기는 어렵다.

질적 분석 결과 제안한 오류 유형 중 띄어쓰기 오류를 제외하고는 대부분의 오류 유형을 잘 교정하지 못하였다. 특히 어순 오류와 시제 오류의 경우 오류 유형을 인식하지 못하였기 때문에 문장의 전체적인 구조나 의미 파악이 활발하지 않음을 알 수 있다. 물론 해당 실험 결과는 gold test set의 부재로 일부 데이터 셋에 대해 실험한 것이므로 교정기 모델 전체의 성능이라고 장담하기는 어렵다. 그러나 오류 유형 분류 기준을 제시하고 이를 기반으로 한 질적분석을 진행했다는 점에서 의의가 있는 연구이다. 추후 연구에는 문장 단위의 연구를 단락 단위로 확장하여 더욱 세부적인 질적 연구를 수행할 계획이다.

## REFERENCES

[1] J. Xiong, Q. Zhang, S. Zhang, J. Hou & X. Cheng. (2015, June). HANSpeller: a unified framework for Chinese spelling correction. *In International Journal of Computational Linguistics & Chinese Language Processing, Volume 20, Number 1, June 2015-Special Issue on Chinese as a Foreign Language*.

[2] M. Kim, J. Jin, H. C. Kwon & A. Yoon. (2013, December). Statistical context-sensitive spelling correction using typing error rate. *In 2013 IEEE 16th International Conference on Computational Science and Engineering* (pp. 1242-1246).

[3] J. H. Lee, M. Kim & H. C. Kwon. (2017). Improved statistical language model for context-sensitive spelling error candidates. *Journal of Korea Multimedia Society, 20(2)*, 371-381.

[4] C. Park, K. Kim, Y. Yang, M. Kang & H. Lim. (2020). Neural spelling correction: translating incorrect sentences to correct sentences for multimedia. *Multimedia Tools and Applications*, 1-18.

[5] M. Lee, H. Shin, D. Lee & S. P. Choi. (2021).

Korean Grammatical Error Correction Based on Transformer with Copying Mechanisms and Grammatical Noise Implantation Methods. *Sensors, 21(8)*, 2658.

[6] C. Park, S. Park & H. Lim. (2020). Self-Supervised Korean Spelling Correction via Denoising Transformer. *7th International Conference on Information, System and Convergence Applications*

[7] C. Park, J. Seo, S. Lee, C. Lee, H. Moon, S. Eo & H. S. Lim. (2021, August). BTS: Back TranScription for speech-to-text post-processor using text-to-speech-to-text. *In Proceedings of the 8th Workshop on Asian Translation (WAT2021)* (pp. 106-116).

[8] J. Byun, H. C. Rim & S. Y. Park. (2007, August). Automatic spelling correction rule extraction and application for spoken-style korean text. *In Sixth International Conference on Advanced Language Processing and Web Information Technology (ALPIT 2007)* (pp. 195-199). *IEEE*.

[9] E. Brill & R. C. Moore. (2000, October). An improved error model for noisy channel spelling correction. *In Proceedings of the 38th annual meeting of the association for computational linguistics* (pp. 286-293).

[10] M. Konchady. (2009). Detecting Grammatical Errors in Text using a Ngram-based Ruleset. *Retrieved* October, 6, 2011.

[11] Li, H., Wang, Y., Liu, X., Sheng, Z., & Wei, S. (2018). Spelling error correction using a nested rnn model and pseudo training data. *arXiv preprint arXiv:1811.00238*.

[12] A. Solyman, Z. Wang & Q. Tao. (2019, September). Proposed model for arabic grammar error correction based on convolutional neural network. *In 2019 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)* (pp. 1-6). *IEEE*.

[13] A. Kuznetsov & H. Urdiales. (2021). Spelling Correction with Denoising Transformer. *arXiv preprint arXiv:2105.05977*.

[14] J. H. Min, S. J. Jung, S. H. Jung, S. Yang, J. S. Cho & S. H. Kim. (2020). Grammatical Error Correction Models for Korean Language via Pre-trained Denoising. *Quantitative Bio-Science, 39(1)*, 17-24.

[15] M. Lee, H. Shin, D. Lee & S. P. Choi. (2021). Korean Grammatical Error Correction Based on Transformer with Copying Mechanisms and

Grammatical Noise Implantation Methods. *Sensors*, 21(8), 2658.

- [16] S. K. Kim, T. Y. Kim, R. W. Kang & J. Kim. (2020). Characteristics of Korean Liaison Rule in the Reading and Writing of Children of Korean-Vietnamese Multicultural Families and the Correlation with Mothers' Korean Abilities. *Korean Speech-Lang. Hear. Assoc.* 29, 57-71.
- [17] K. Lee. (2018). Patterns of Word Spacing Errors in University Students' Writing. *J. Res. Soc. Lang. Lit.* 97, 289-318.

임 희 석(Heuseok Lim)

[정회원]



- 1992년 : 고려대학교 컴퓨터학과 (이학학사)
- 1994년 : 고려대학교 컴퓨터학과 (이학석사)
- 1997년 : 고려대학교 컴퓨터학과 (이학박사)

- 2008년 ~ 현재 : 고려대학교 컴퓨터학과 교수
- 관심분야 : 자연어처리, 기계학습, 인공지능
- E-Mail : limhseok@korea.ac.kr

구 선 민 (Seonmin Koo)

[학생회원]



- 2018년 3월 ~ 현재 : 건국대학교 컴퓨터공학과 (공학사)
- 관심분야 : AI, Machine Translation, Grammar Error Correction, Deep Learning, Natural Language Processing
- E-Mail : fhdahd899@gmail.com

박 찬 준(Chanjun Park)

[학생회원]



- 2019년 2월 : 부산외국어대학교 언어처리창의융합전공 (공학사)
- 2018년 6월 ~ 2019년 7월 : SYSTRAN Research Engineer
- 2019년 9월 ~ 현재 : 고려대학교 컴퓨터학과 석박사통합과정

- 관심분야 : Data-centric AI, Machine Translation, Grammar Error Correction, Deep Learning
- E-Mail : bcj1210@naver.com

소 아 램(Aram So)

[정회원]



- 2013년 2월 : 백석대학교 정보통신학과(공학사)
- 2020년 8월 : 고려대학교 컴퓨터학과 (공학박사)
- 2020년 9월 ~ 현재 : 고려대학교 Human-inspired AI연구소 연구교수

- 관심분야 : 인공지능, 자연어처리, 기계학습
- E-Mail : aram@korea.ac.kr