

# 한중 기계번역을 활용한 오류 수정 및 생산성 연구\*

홍 연 옥\*\*

<目次>

I. 서론	3. 유의어 오류 유형
II. 한국인 학습자의 중국어 오류 유형	4. ‘很’의 오류 유형
1. 존현문 오류 유형	III. 한중 신경망 기계번역의 정확도 및 활용 효과 분석
2. 연동문 오류 유형	IV. 결론

## I. 서론

현재 네이버, 바이두, 구글에서 제공하는 기계번역 서비스는 NMT<sup>1)</sup> 기술을 이용하면서 이전의 규칙기반 기계번역과 통계기반 기계번역에 비해 훨씬 우수한 성능을 보이고 있다. 따라서 최근 국내에서도 기계번역을 외국어 학습에서 효과적으로 활용하고자 하는 시도가 영어를 연구하는 연구자와 교육자들을 중심으로 증가하는 추세이다. 이에 비해 중국어 교육현장에서는 기계번역을 활용하는 시도가 나타나고 있지만, 여전히 구체적인 교

\* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021S1A5B5A16077207)

\*\* 세종대학교 인문-IT 융합연구원 학술연구교수

1) NMT란 Neural Machine Translation의 약어로서 인공 신경망 기계번역을 나타낸다. 입력문장은 커다란 가상공간인 인공 신경망에 벡터(좌표 값)로 저장되는데, 이때 유사한 의미의 문장은 가깝게 모여 있게 된다. NMT는 입력 문장 주변의 정보를 참고하여 정확한 번역을 산출한다.

학방안에 대한 논의에 대해서는 초보적인 단계이다. 이에 본 연구에서는 기계번역을 활용하여 중국어 학습자의 오류를 수정하는 교학 방법을 이용하여 기계번역을 중국어 학습에 유용하게 활용할 수 있는 방안을 제안해 보고자 한다. 학습자의 오류를 파악하는 것은 외국어를 가르치고, 연구하고, 학습하는 모든 이들에게 중요한 의미를 갖기 때문이다. Corder(1985)는 학습자의 오류를 통해 교사는 학습자가 어느 정도 목표에 도달했는지 파악함으로써 향후 무엇을 더 가르쳐야 하는지 파악할 수 있다고 하였다. 또한 연구자는 오류를 통해서 학습자가 언어를 습득, 학습하는 방법과 전략을 파악할 수 있다고 하였다. 마지막으로 학습자도 자신의 오류를 파악함으로써 목표어를 배우면서 자신의 가설이 타당한지 검증해 볼 수 있는 수단이 될 수 있다고 하였다. 따라서 본고는 기계번역의 발전으로 외국어 교육과 학습 방식에 변화가 요구되는 시대에 현 신경망 기계번역이 외국어 교수법에서 중요한 오류 수정에 활용할만한 수준에 이르렀는지 검증하고, 향후 외국어 학습에서 기계번역을 활용한 오류 수정 방안을 고찰해보고자 한다. 이에 한국인 중국어 학습자의 오류로 자주 지적되어왔던 존현문, 연동문, 유의어, ‘很’의 오류 유형 4가지를 중점적으로 고찰해볼 것이다. 오류문은 北京语言大学的 HSK动态作文语料库의 한국인 작문의 오류문장과 존현문, 연동문, 유의어, ‘很’의 오류 유형과 관련한 선행 연구에서 제시한 오류 문장에서 선별하였다.

한편, 발전된 과학 기술과 인간과의 협력을 위한 작업으로 기계번역에서 출력된 결과물을 수정하여 질 높은 번역으로 활용하기 위한 방안은 기계와 인간이 서로의 결과물을 포스트에디팅하는 작업을 들 수 있다. 인간과 기계는 서로의 오류를 수정하는 과정에서 번역의 결과물을 개선할 수 있고, 정확한 학습 데이터를 축적할 수 있다. 축적된 결과물은 인간에게는 유창한 언어를 구사할 수 있는 학습 능력을 향상시킬 것이며, 기계번역에는 딥러닝 시스템에서 필수적인 정확한 함수 데이터로 구성되어 번역의 정확도를 높일 수 있는 재학습 데이터로 활용될 것이다. 따라서 기계번역과 인간번역에서 자주 발생하는 오류를 유형화한다면 향후 정확한 번역을

위한 포스트에디팅 작업에 도움이 될 수 있을 것이고, 기계번역을 활용해서 생산성 높은 번역을 담당하게 될 미래의 번역가들의 가이드라인 작성에도 도움이 될 수 있다. 본 연구에서는 앞서 지적한 바와 같이 한국인 학습자들에게 자주 나타나는 중국어 번역의 오류 유형을 살펴보고, 신경망 기계번역에 적용하여 결과물을 비교 분석한 후, 한국인 중국어 학습자들의 포스트에디팅을 활용한 학습 효과를 고찰해볼 것이다.

## II. 한국인 학습자의 중국어 오류 유형

肖奚强(2000)은 외국어 학습자들에게서 나타나는 오류와 실수는 서로 다른 의미임을 지적한 바 있다. 실수는 학습자가 피로나 긴장 등으로 인해서 범하게 되는 우연적인 현상이기 때문에 화자가 모국어를 사용할 때도 나타나는 현상이므로 이를 통해 화자의 언어능력을 측정할 수 없다. 이와 달리 오류는 학습자가 언어를 학습하고 습득하면서 발생하는 체계적이고, 규칙적인 현상이기 때문에 화자의 언어능력을 보여주는 지표가 되며, 언어를 연구하는 연구자들이 주목할 가치가 있는 대상이라고 하였다. 즉, 학생들은 실수한 경우에는 스스로 교정할 수 있기 때문에 이를 수정하기 위한 학습이 필요하지 않지만, 오류인 경우에는 본인의 학습 오류 유형을 파악하고 재학습할 필요가 있다.<sup>2)</sup> 또한 鲁健骥(1984)는 외국어 학습자가 어떤 오류를 범하는지, 오류가 어떻게 발생하는지 연구하는 오류분석은 외국어 교육을 학습자에게 보다 더 효과적으로 진행하도록 하는 의미가 담겨있다고 밝혔다. 따라서 본 장에서는 한국인들이 자주 범하는 오류의 유형을 문

2) Rod Ellis(2003)는 실수와 오류를 구분하는 방법에 대해 사용 빈도와 사용자가 스스로 교정할 수 있는지 여부의 기준을 제시하였다. 학습자가 범하는 동일 문법 항목에 대해서 잘못된 습관의 사용 빈도가 높을 경우 이는 실수가 아니라 오류이다. 또한 학습자가 스스로 자신의 실수를 교정할 수 없다면 오류라고 판별한다.

형 오류와 어휘 오류 유형으로 나누어 문형 오류 유형으로는 존현문과 연동문에서 나타나는 한국인 학습자의 오류 유형을 고찰하고, 어휘 유형에서는 중국어 유의어, ‘很’의 사용에서 나타나는 오류 유형을 고찰해보고자 한다.

### 1. 존현문 오류 유형

존현문(存現文)은 사람이나 사물이 특정 장소에 존재하거나, 출현 또는 소실하는 의미를 나타내는 문장의 형식을 칭한다. 존현문에서는 장소명사가 주어의 위치에 사용되고, 출현하거나 소실하는 사람이나 사물을 지칭하는 성분이 목적어 위치에 놓이게 된다. 刘月华(2001)는 존현문(存現文)을 사람 또는 사물의 존재를 나타내는 존재문(存在文)과 출현, 소실을 나타내는 은현문(隱現文)으로 나누었다. 사실 중국어의 존현문 형식은 한국어 문형에는 존재하지 않는다. 주어로서 장소 또는 시간이 사용되는 중국어 존현문과 달리 한국어에서는 장소와 시간을 나타내는 성분이 부사어로서 사용되기 때문이다. 중국어에서는 오히려 주어의 위치에 사용되는 장소명사 앞에 ‘在’나 ‘从’이 사용될 수 없다.

- (1) a. 村里死了一只羊。  
 b. \*在村里死了一只羊。  
 마을에서 양 한 마리가 죽었다.
- (2) a. 图书馆前站着一个人。  
 b. \*在图书馆前站着一个人。  
 도서관 앞에 한 사람이 서있다.

또한 존현문의 목적어는 비한정성이라는 특징을 갖는다는 사실도 한국인 중국어 학습자들에게는 생소한 개념이다. 왜냐하면, SOV(주어-목적어-동사)의 어순을 갖는 알타이어 계통의 한국어는 한정성의 개념이 발달하

지 않았기 때문이다. 김도영(2012)은 SOV 어순을 갖는 언어는 SVO 어순을 갖는 언어에 비해 한정 용법이 발달하지 않았다는 점을 지적하며, SVO 어순을 갖는 언어는 전치와 후치 수식이 모두 가능하지만, SOV 어순을 갖는 한국어는 후치 수식이 불가능하기 때문에 한정과 비한정이 모두 앞에 출현한다고 하였다. 이로 인해 한국인 중국어 학습자들은 존현문에는 비한정성 목적어가 사용된다는 규칙을 인지하지 못하고 한정성 목적어를 사용하여 오류문장을 산출한다.

한편, 한국인 중국어 학습자가 중국어를 학습할 때, 존재문을 중심으로 학습하는 경향이 있다. 특히 존재문 중에서 존재동사 ‘有’, ‘是’, ‘在’를 사용한 존재문에서 어순에 주의해야 한다는 점을 학습한다. 예를 들어 ‘有’나 ‘是’가 동사인 존재문에서는 주어의 위치에는 장소를 나타내는 어휘가 놓이고, 목적어의 위치에는 존재대상을 나타내는 어휘가 놓이는 반면, ‘在’가 사용된 존재문에서는 주어자리에 존재대상이 오고, 목적어 자리에 장소를 나타내는 어휘가 온다는 점을 주의할 점으로 학습한다. 때문에 이러한 점에서는 한국인 학습자들은 오류가 비교적 적은 편이다.

이에 비해서 존재동사 없이 존재의 양태를 나타내는 정태 존재문이나 동태 존재문에서는 비교적 많은 오류를 보인다. 먼저 정태 존재문이나 동태 존재문에서 지속을 나타내는 상표지 ‘着’를 사용해서 지속상태인 대상의 존재를 표현하는데, 이때 ‘在’와 ‘正在’와 같은 부사는 함께 사용할 수 없다는 점을 인지하지 못한다는 점은 한국인 학습자들의 대표적 오류로 나타난다. 또한 은현문에서 술어 뒤에 상표지 ‘了’를 사용하여 장소에 일어난 사건의 완료를 표현하는데, 이때는 지속을 나타내는 상표지 ‘着’와 함께 사용할 수 없다는 점도 한국인 학습자들이 부족한 학습 사항으로 보인다. 또한 출현이나 소실의 의미는 이동을 나타내는 동사와 관련되는 경우가 많기 때문에 방향보어 ‘來’, ‘去’, ‘出’, ‘进’ 등과 자주 사용된다는 점도 교학에서 보완해야 할 점으로 지적된다. 다음은 박애화(2013)에서 모국어 어순의 간섭으로 인해 나타난 한국인 중국어 학습자의 존현문 오류 문장이다.

- (3) \* 前面一个人來了。
- (4) \* 我们家里几个客人來了。
- (5) \* 昨天一件大事發生了。
- (6) \* 今天几个朋友來了。
- (7) \* 前天晚上火灾發生了。
- (8) \* 我们班里新同学來了。
- (9) \* 这个房间50个人坐不下。

(3-9)는 한국어 어순을 그대로 중국어 작문을 할 경우에 적용시켜 나타난 비문이다. 한국인들은 여전히 주격조사로서 해석되는 문장 성분을 주어 위치에 놓고, 장소를 나타내는 어휘는 부사어 위치에 놓으려는 경향이 나타난다. 이는 모국어의 간섭으로 인해서 나타나는 외국어 학습 오류라고 볼 수 있다.

## 2. 연동문 오류 유형

연동문(連動文)이란 하나의 주어를 갖는 두 개 혹은 그 이상의 동사구를 사용해서 동작의 순서나 목적, 수단, 방식을 나타내는 문장이다. 순서를 나타내는 연동문의 통사적 특징으로는 첫 번째 동사 뒤에 ‘了’, ‘过’, 결과보어를 사용해서 동작의 완료, 경험, 결과 상태를 나타낸다는 것이다. 한편, ‘有’의 특수용법으로 후치 수식어가 존재하는 ‘有’ 연동문이 있다. 이와 같은 연동문에서 한국인들이 자주 오류를 범하는 연동문은 목적을 나타내는 연동문이다. 순서나 방식을 나타내는 중국어 연동문의 동사 순서는 동작이 진행되는 순서에 따라 배열되기 때문에 한국어 동사 배열순서와 중국어 동사 배열순서는 거의 동일하다. 이와 달리 목적을 나타내는 연동문은 중국어와 한국어에서 동사의 배열순서가 상이하다. 즉, 순서나 방식을 나타내는 연동문은 중국어와 한국에서 모두 시간 순서의 흐름에 따라 나열되지만 목적을 나타내는 한국어 연동문은 이와 같은 시간순서 원리를 따르지 않기 때문이다. 따라서 한국인 중국어 학습자가 목적을 나타내는

연동문에서 특히 오류를 많이 범하는 것은 한국어에서는 목적을 나타내는 동사가 시간 순서 원리를 어기고 제 1동사에 위치할 수 있는 모국어의 간섭현상 때문이라고 볼 수 있다. 다음은 HSK 4급 수준의 한국인 중국어 학습자의 오류 예문이다.<sup>3)</sup>

- (10) \* 他帮我去北京的火车票买到了。
- (11) \* 他拍照片去公园。
- (12) \* 姐姐经常锻炼去健身房。
- (13) \* 他想买生日礼物去商店。
- (14) \* 老师取课程表去。
- (15) \* 我们上大学都有机会。
- (16) \* 明天我们去上海坐飞机。

(10-16)과 같이 한국인 중국어 학습자들은 중국어의 시간 순서에 의한 연동문의 어순을 정확하게 이해하지 못한 채 한국어 어순에 의존하여 중국어 문장을 작문하는 경향이 나타난다. 앞서 지적한 바와 같이 중국어의 연동문 문장의 어순에서는 시간 순서의 원리가 매우 엄격히 지켜진다. 예를 들면 중국인들은 목적을 이루기 위해서는 특정 장소에 도착해야 한다고 인지한다. 이와 달리 한국어는 목적을 나타내는 동사가 ‘~하러’, ‘~하기 위해’라는 표현으로 다른 동사보다 앞에 위치한다. 즉, 한국인 중국어 학습자들은 목적을 나타내는 중국어 연동문에서 동사 순서는 한국어 문장에서의 동사 순서와 다르다는 점과 중국어의 연동문은 시간순서의 원리가 적용된다는 점을 먼저 이해하는 것이 중요하다.

### 3. 유의어 오류 유형

언어에서 의미가 같거나 비슷한 어휘가 두 개 이상 존재하는 경우에 이

3) 박애화, <한국 학생 新HSK 4級 쓰기 偏誤 분석>, 《東亞人文學》 제 26집, 2013, 152쪽에서 예문의 일부를 인용함.

들을 유의어라고 한다. 유의어는 해당 언어의 직감과 경험이 부족한 외국어 학습자들에게는 어려운 과제이다. 따라서 유의어를 구분하여 올바르게 사용하도록 하는 학습은 교육자들이 더욱 더 주목해야 할 부분이다. 중국어 유의어에 관한 연구는 국내외에서 다양하게 이루어져 왔다. 김해정(2000), 맹주익(2002), 박덕준 외(2000)에서는 중국어 유의어에 대한 상세한 분석을 하였으며, 이경숙(2001), 이명윤(2006) 등에서는 유의어를 교육 현장에서 어떻게 교육할 것인가를 논하는 등 중국어 유의어 교수방안에 대한 모색은 꽤 오래전부터 이루어져 왔다. 그러나 한국인 중국어 학습자는 중국인 모국어 화자들과 같은 직감으로 유의어를 판단하는 것이 불가능하다. 따라서 선행연구를 통해 증명된 유의어에 대한 변별 요소들을 일일이 학습자가 암기해야 하는 방법에 의존할 수밖에 없기 때문에 유의어 연구 결과와 교육 방안이 중국어 학습자들의 학습적 부담을 해소하는 방법으로는 적용될 수 없었다. 또한 이를 학습하지 않거나, 학습한 내용을 기억하지 못하는 경우에는 유의어를 변별하지 못하는 오류가 반복되고 있었다. 한국인 중국어 학습자가 빈번하게 범하는 유의어 오류에 대해 연구한 김선녀(2011)와 민재홍(2020)에서 제시한 한국인 중국어 학습자의 중국어 유의어 오류 중에서 본 연구에서 기계번역과 한국인 학습자의 번역 비교에서 다루고자 하는 오류 예문은 다음과 같다.

## [表现 / 表达]

- (17) \* 他不喜欢表现自己的感情。  
 (18) \* 他不是爱表达自己的人。

## [依赖 / 依靠]

- (19) \* 女儿是他生活依靠。  
 (20) \* 人不能有依赖思想。

## [合适 / 适合]

- (21) \* 这双鞋你穿着正适合。

(22) \* 过去的经验未必合适当前的情况

[了解 / 理解]

(23) \* 他很理解中国的历史和文化。

(24) \* 我们认识的时间不长, 我还不理解他。

[保持 / 维持]

(25) \* 请大家维持清洁。

(26) \* 请维持适当的温度。

이상의 ‘表现’과 ‘表达’, ‘依赖’와 ‘依靠’, ‘合适’과 ‘适合’, ‘了解’와 ‘理解’, ‘保持’과 ‘维持’는 중급중국어 이상을 학습한 대부분의 학생들이 교재나 HSK 시험 대비 문법서를 통해 학습한 유의어 항목이다. 그러나 실제 작문에서는 이들의 오류가 여전히 반복되는 현상이 나타난다. 이는 유의어 각각의 변별 요소들을 암기하는 학습 방법이 학습자의 실제 사용과 작문에서는 효과적이지 않았다는 것을 암시해준다.

#### 4. ‘很’의 오류 유형

‘很’은 한국인 학습자들이 한국어의 정도를 나타내는 부사어인 ‘매우’라는 의미로 초급중국어에서부터 학습하는 어휘이다. 그러나 ‘很’의 용법은 한국어에서의 부사어 ‘매우’의 용법과 동일하지 않다. 때문에 한국인들이 중국어 작문을 할 경우, ‘很’의 오류를 자주 범하게 된다. 이것은 北京语言大学에서 수집한 HSK动态作文语料库의 한국인 학습자들의 작문에서도 ‘很’을 사용한 문장의 오류가 많은 수를 차지한다는 것으로도 확인해볼 수 있다.

姜昌义(2010)는 한국인 유학생들의 작문에서 나타난 비문을 중심으로 한국어 정도부사와 한국어 정도부사의 의미를 비교하며 ‘很’의 오류를 추가, 누락, 대체, 어순 오류로 분류하여 고찰하였다. 이와 동일하게 焉德才

(2016)는 한국인 중국어 학습자의 오류를 ‘韩国留学生汉语中介语语料库’에 수록된 오류문 중심으로 분석하였고, 오류 유형은 추가, 누락, 어순, 대체 네 가지 유형으로 나누었다. 오류 분석결과 한국인 학습자의 ‘很’의 오류는 추가와 누락에서의 오류가 대부분을 차지하였고, 어순과 대체 오류는 상대적으로 적은 비율을 차지하였다. 또한 郑月(2018)는 북경어언대학교 동태(动态) 말뭉치자료를 활용하여 한국인 중국어 학습자의 ‘很’의 오류 유형을 추가, 누락, 어순, 대체로 나누어서 서술하였지만, 연구의 대부분은 추가와 누락 오류를 중심으로 서술한 바 있다. 한편, 周佳佳(2017)는 한국인 중국어 학습자의 ‘很’의 오류를 누락, 추가, 대체, 결합 네 가지로 나누어서 설명하였고, 손정애(2010)는 한국인 중국어 학습자의 ‘很’의 오류 유형을 첨가, 누락, 대체 유형으로 분류하였다. 이와 같은 선행연구에서 지적된 한국인 중국어 학습자의 ‘很’의 네 가지 오류 유형을 예문으로 제시하면 다음과 같다.<sup>4)</sup>

- (27) \*学习外国语难。(누락)
- (28) \*不过现代的代沟问题比以前很严重。(추가)
- (29) \*我们都知道吸烟对个人健康的影响太大。(대체)
- (30) \*我觉得我的同屋不很胖。(어순)

즉, 선행연구들의 연구 결과를 종합해보면, 한국인 중국어 학습자의 ‘很’의 오류는 추가와 누락의 오류 유형이 대부분이며, 몇몇 연구에서 지적한 어순, 결합 유형은 상대적으로 적은 수를 차지한다.

4) 焉德才, 〈韩国学生习得汉语程度副词‘很’语法偏误生态描写〉, 《海外华文教育》第 79期, 227-238页, 2016에서 ‘很’의 오류 예문의 일부를 인용하였음.

### Ⅲ. 한중 신경망 기계번역의 정확도 및 활용 효과 분석

본 장에서는 한중 신경망 기계번역을 중국어 학습에 활용할 수 있는지의 여부를 선행연구의 오류문장이 기계번역에서는 어떻게 출력되는지를 검증하고, 기계번역을 활용해서 작문과제를 한 중급중국어 수강생들의 학습 효과를 고찰해보고자 한다. 이를 위해 본고는 2장에서 논의하였던 한국인 중국어 학습자들의 오류유형에 관한 선행연구를 바탕으로 다음과 같은 오류문장을 선별하였고, 이와 관련한 한국어 번역을 기계번역으로 추출하였다. 또한 기계번역과 현재의 학부생들의 번역 능력을 비교해보기 위해 중급 중국어 수업을 수강하고 있는 한국인 중국어 학습자 18명에게 동일한 한국어 예문을 제공하고 중국어로 번역하도록 하였다. 또한 본인의 번역결과물을 기계번역 결과물을 통해서 수정하도록 하였고, 이와 관련해서 학습 효과를 설문 조사하였다. 다음은 본 연구에서 활용한 한국어 예문 예시이다.

〈표 1〉 신경망 기계번역 한국어 입력 문장과 모범답안

측정 문법	한국어 문장(ST)	모범 번역
1	교실에는 학생이 많다.	教室里有很多学生。
2	시장은 병원 옆에 있습니다.	市场在医院旁边。
3	앞에서 한 사람이 오고 있었다.	前面来了一个人。
4	우리 집에 손님 몇 분이 오셨다.	我家来了几位客人。
5	어제 한 가지 큰 일이 발생 했다.	昨天发生了一件大事。
6	오늘 친구 몇 명이 왔다.	今天来了几个朋友。
7	그저께 저녁에 화재가 발생 했다.	前天晚上发生了火灾。
8	우리 반에 새 친구가 왔다.	我们班来了个新朋友。
9	마을에서 양 한 마리가 죽었다.	村里死了一只羊。
10	도서관 앞에 한 사람이 서있다.	图书馆前站着一个人。

	측정 문법	한국어 문장(ST)	모범 번역
11	연 동 문	그는 내가 북경에 가는 기차표를 사는 것을 도와주었다.	他帮我买了去北京的火车票。
12		그는 사진 찍으러 공원 간다.	他去公园拍照。
13		누나는 늘 단련하러 헬스장에 간다.	姐姐经常去健身房锻炼。
14		그는 생일 선물을 사러 상점에 가려고 한다.	他要去商店买生日礼物。
15		선생님은 수업시간표를 가지러 가셨다.	老师去拿课程表。
16		우리는 대학 갈 기회가 다 있다.	我们都有机会上大学。
17		내일 우리는 상해에 비행기를 타고 간다.	明天我们坐飞机去上海。
18		나는 영화관에 가서 영화를 보지 않아.	我不去电影院看电影。
19		그는 도서관에 가서 책을 빌렸습니다.	他去图书馆借书。
20		오늘 나는 숙제할 시간이 없다.	今天我没有时间做作业。
21	유 의 어	그는 자기감정을 표현하는 것을 좋아하지 않는다.	他不喜欢表达自己的感情。
22		그는 자신을 나타내는 것을 좋아하는 사람이 아니다.	他不是爱表现自己的人。
23		너 다른 사람한테 너무 기대지마!	你不要太依赖别人了。
24		내가 힘들 때면 나한테 기대.	你累的时候就靠着我。
25		이 신발은 네가 신으니 꼭 맞는다.	这双鞋你穿正合适。
26		과거의 경험이 반드시 지금의 상황에 맞는 것은 아니다.	过去的经验不一定适合现在的情况。
27		그는 중국의 역사와 문화에 대해 잘 이해하고 있다	他对中国的历史和文化很了解。
28		우리가 알고 지낸지 얼마 되지 않아 나는 그를 잘 이해하지 못한다.	我们认识不久，我不太了解他。
29		모두 청결을 유지하세요.	请大家保持清洁。
30		적당한 온도를 유지하세요.	请保持适当的温度。
31	정 도 부 사 很	외국어를 공부하는 것은 어렵다.	学习外语是困难的。
32		현대의 세대 차이는 이전보다 심해졌다.	现代的代沟比以前严重了。
33		흡연이 개인의 건강에 미치는 영향이 너무 크다는 것은 우리 모두 알고 있습니다.	我们都知道吸烟对个人健康的影响很大。

	측정 문법	한국어 문장(ST)	모범 번역
34	정 도 부 사 很	나는 내 룸메이트가 뚱뚱하지 않다고 생각한다.	我觉得我室友不胖。
35		작년에 아버지가 보내주신 편지는 사실 정말 감동적이었어요.	去年爸爸寄来的信其实很感人。
36		나의 아버지는 중국의 정치·경제·사회·문화 방면에 대한 관심이 많으시다.	我父亲对中国的政治、经济、社会、文化方面很感兴趣。
37		내가 담배를 피우기 시작한 원인은 사회의 영향과 밀접한 관계가 있다.	我开始吸烟的原因与社会的影响密切相关。
38		자전거 여행이 시작되자 모두들 기뻐했다.	自行车旅行开始了, 大家都很高兴。
39		날씨가 이전보다 시원해졌다.	天气比以前凉爽了。
40		그는 큰 수박 한 개를 샀다.	他买了一个大西瓜。

〈표 1〉의 예문 중 (1-20)은 한국인 중국어 학습자가 나타내었던 문형 오류로 (1-10)은 존현문과 관련된 예시이며, (11-20)은 연동문과 관련된 예시이다. (21-40)은 어휘 오류가 나타났던 오류 문장으로 (21-30)은 유의어 변별 오류 예문이고, (31-40)은 중국어 정도 부사 ‘很’과 관련한 오류 문장이다. 먼저 이상의 예문을 각각 파파고, 구글, 바이두 신경망 기계번역을 이용해서 중국어로 추출했을 때의 오류 개수 및 번역 정확도를 분석해보면 다음과 같다.

〈표 2〉 신경망 기계번역 출력문장의 정확도

	파파고	구글	바이두	평균
존현문	1개(90%)	1개(90%)	0개(100%)	2개(93.3%)
연동문	0개(100%)	1개(90%)	0개(100%)	1개(96.7%)
유의어	1개(90%)	3개(70%)	1개(90%)	5개(83.3%)
很	1개(90%)	2개(80%)	1개(90%)	4개(86.7%)
총합	3개(92.5%)	7개(82.5%)	2개(95%)	

본 실험에 사용된 한국인이 자주 범하는 대표 오류 유형 문장의 한국어 번역을 신경망 기계번역으로 출력해보면, 그 결과물은 매우 높은 정확도를 보인다. 파파고와 바이두는 실험 문장 모든 유형에서 각각 1개 이하의 오류 문장을 출력하면서 매우 높은 정확도를 보였고, 구글에서도 존현문과 연동문의 오류에서는 1개 이하의 오류를 산출하는 것을 확인해볼 수 있었다.

같은 한국어 입력 문장 예문을 중급 중국어를 수강하는 학생들에게 작문하도록 하였을 때, 최초 번역문의 정확도는 존현문 54.6%, 연동문 52.3%, 유의어는 42.3%, ‘很’과 관련한 작문에서 45.5%의 정답률을 보였던 것에 비하면, 기계번역의 정확도는 매우 높은 편이다. 이는 한국인 학습자가 중국어 학습에서 참고할만한 교학 도구로서의 가치를 보여준다. 기계번역은 존현문과 연동문의 어순과 통사적 특징을 잘 학습하고 있었고, 유의어와 ‘很’의 용법도 잘 학습하고 있었다.

〈표 3〉 기계번역 결과물 예시

예문	유형	파파고	구글	바이두
9	존현문	村里死了一只羊。	村里有一只羊死了。	村里死了一只羊。
14	연동문	他要去商店买生日礼物。	他想去商店买生日礼物。	他要去商店买生日礼物。
23	유의어	你别太依赖别人了!	不要太依赖别人!	你不要太依赖别人了!
24		你累的时候就靠着我。	当你遇到困难时依靠我	你累的时候就靠着我。
25	유의어	这双鞋非常适合你。	这双鞋非常适合你。	这双鞋你穿正合适。
26		过去的经验不一定适合现在的情况。	过去的经验不一定适合现在的情况。	过去的经验不一定适合现在的情况。
34	很	我觉得我的室友不胖。	我不认为我的室友很胖。	我觉得我室友不胖。

이에 본 연구에서는 학생들 각자 기계번역을 사용해서 결과물을 추출하고 본인의 작문과 비교하고 이를 이용해서 기계번역의 결과물과 자신의 결과물을 포스트에디팅<sup>5)</sup>하도록 하였다. 포스트에디팅 이후의 결과물의 정확도를 비교해보면, 학생들의 작문에서도 4가지 오류 유형에서 모두 90% 이상의 정확도를 보였다. 이는 이미 상당부분 모범 답안을 가지고 있었던 신경망 기계번역을 활용하여 본인들의 작문을 수정을 한 결과라고 보여진다. 또한 학생들이 신경망 기계번역을 활용하여 도움을 받았던 영역에 대한 설문에서는 어휘 선택이 78.8%, 문법 수정이 67.7%로 학생들이 응답했다. 또한 본 실험에 참여한 학생들이 기계번역을 활용한 작문에 대해 제시한 의견의 일부는 다음과 같다.

〈표 4〉 포스트에디팅 관련 과제의 수강생 설문조사 답변 예시

	신경망 기계번역 활용 후 번역 생산성과 관련된 의견
참여 학생 A	기계번역을 활용하니 번역의 속도를 높일 수 있었고, 중국어로 어떻게 번역해야 할지 막막했던 문장들도 기계번역의 도움을 통해 정확한 번역을 할 수 있었다.
참여 학생 B	사실 기계번역에 대한 편견이 있어서 평소에는 기계번역을 잘 사용해보지 않았지만, 이번 과제를 통해 기계번역을 사용해보니 기계번역의 정확도가 매우 높다는 사실에 놀랐다.

5) 포스트에디팅과 관련해서는 수업 1-2주차에 교육을 실시하였고, 영어작문의 포스트에디팅으로 제시하였던 곽중철, 한승희(2018)를 참고하여 중국어 포스트에디팅 관련 작업에서 사용할 가이드라인을 제공하였다. 이는 필자의 把자문 오류 분석 연구와 관련한 <기계번역의 교육적 활용 가능성 고찰 및 제안>에서 진행된 수업 방식과 가이드라인을 동일하게 사용했음을 밝혀둔다. 먼저 사전 학습으로는 앞서 지적한 바와 같이 1-2주차에 포스트에디팅 관련 교육을 실시한다. 이후 매 시간 해당 주차의 문법 사항을 학습하고, 이와 관련한 오류 문장 제시 및 작문을 과제로 제시한다. 이후 학생들은 신경망 기계번역을 이용하여 오류문장을 검토하고, 본인의 작문을 수정한다. 이후 교실 학습에서 본인의 결과물을 공유하고, 기계번역의 오류와 본인의 오류 수정에 도움이 되었던 점들을 토론한다. 이후 복습 과제로 정확한 문장 쓰기를 하며, 해당 주차의 문법 사항을 정리한다.

	신경망 기계번역 활용 후 번역 생산성과 관련된 의견
참여 학생 C	어떻게 작문해야 할지 몰랐던 문장은 기계번역의 도움을 받아 작문할 수 있었다. 그런데 기계번역이 번역해낸 문장이 올바른 문장인지 판단할 때는 애매하고 어려운 부분이 있었다.
참여 학생 D	기계번역을 통해 작문 과제를 하니 보다 쉽게 과제를 해결할 수 있었다. 그리고 나서 포스트에디팅 작업을 하면서 그 시간에 배웠던 문법 사항들을 복습하며 수정연습을 한 점이 좋았다.
참여 학생 E	기계번역을 하며 수정한 문장들을 조원들끼리 토론하며 중국어 문법에 대해 토론한 것이 유익했다.

설문조사의 의견을 종합해본 결과 학생들은 중국어 문법과 작문 수업에서 기계번역을 활용해서 작문을 하고 이를 포스트에디팅하는 수업과제에 대해 만족도가 높았다. 또한 포스트에디팅을 하면서 학생들은 한중 신경망 기계번역에 대한 신뢰도가 높아졌고, 또 기계번역의 오류 유형을 함께 공유하면서 동음이의어와 시제 오류와 같은 오류<sup>6)</sup>를 기계번역이 여전히 범하고 있음을 학습하면서, 포스트에디팅을 할 때 주의사항으로 인지하기도 하였다. 또한 기계번역을 사용해서 본인들의 오류를 수정하며, 문법사항을 스스로 학습하는데 도움이 되었다는 의견이 많았다. 다만, 기계번역을 포스트에디팅하는 능력이 부족한 학생들에게는 토론을 통한 재학습이 필요하기도 하였다.

즉, 신경망 기계번역을 중급중국어 문법과 작문 시간에 이용하였을 때 학생들의 작문 생산성을 높일 수 있었고, 학생들의 작문 정확도도 높일 수 있었으며, 이를 토론하는 과정을 통해 학생들은 문법 사항을 재학습하는 효과를 얻을 수 있었다. 이에 본고는 한중 신경망 기계번역이 향후 더욱

6) 예를 들면, 파과고 번역에서 2번 예문 ‘시장은 병원 옆에 있습니다.’가 ‘市长在 医院旁边.’으로 번역되었고, 구글 번역에서는 7번 예문 ‘그저께 저녁에 화재가 발생 했다.’가 前天晚上发生火灾.로 번역되어 오류가 나타났다.

발전할 것이라는 점을 고려하면, 중국어 교학에서도 기계번역을 교학 도구로 활용하여, 학습자가 더욱 효율적인 학습 효과를 얻을 수 있음을 확인하였다.

#### IV. 결론

본 연구는 신경망 기계번역으로 한국어를 중국어로 번역했을 때의 정확성을 파파고, 구글, 바이두 한중 번역 결과물을 통해 입증하고, 신경망 번역을 번역 교육에 활용했을 때의 학습 효과를 설문조사를 통해 확인해 보았다. 또한 한국인 학습자들이 기계 번역을 검증할 수 있는 능력을 검증하고, 어떠한 오류 유형에 취약한지 파악해볼 수 있다는 점에서 향후 신경망 기계번역을 활용한 중국어 교수법을 제안해볼 수 있었다. 따라서 향후 학문적, 사회적으로 과학의 발전을 잘 활용할 수 있는 실용 인재에게 적합한 중국어 교학방안을 연구함에 있어서 신경망 기계번역을 활용한 교수법은 중국어의 연구와 교육에 다음과 같은 의미가 있다.

첫째, 신경망 기계번역의 문제점과 그 유형을 파악해볼 수 있다. 신경망 번역 체계 이전의 통계 기반 번역은 단어와 단어가 모여 구성된 구(Phrase) 단위의 번역 방식이었다면, 신경망 번역 시스템은 문장 전체의 맥락에서 구성 요소들을 변환하면서 해석하고 번역하는 방식이다. 이러한 시스템의 전환은 맥락에 의한 번역을 가능하게 함으로써 보다 정확하게 출력 언어로 번역해준다는 장점을 가진다. 다만 현재의 신경망 번역 시스템은 일부 다의어, 문화소 등의 어휘와 문장의 성분 또는 품사를 파악함에 있어서 오류를 범하여 의미 또는 문법적으로 이상한 문장을 만들어 내기도 한다. 따라서 학생들과 신경망 기계 번역의 오류를 함께 고찰하며 정확도를 검증하는 과정에서 기계번역 시스템의 성능을 개선하여 기능을 향상시키는데 유용한 데이터를 마련할 수 있다.

둘째, 한국인 중국어 학습자들의 작문 오류를 분석한다. 본 연구에 참여하는 중급 중국어 수강생들은 수업시간에 신경망 기계 번역 프로그램을 사용하여, 기계 번역의 정확도를 스스로 검증하고, 완전한 문장으로 수정한 후 작문을 하도록 하였다. 기계 번역의 원문과 그것을 수정한 문장을 작성하면서 학습자들은 적절한 기계 번역의 도움을 받을 수 있을 뿐만 아니라 문장의 의미, 문법의 오류 등도 보다 깊이 검증해보는 효과를 얻었다. 이러한 과정을 거쳤음에도 불구하고 수정되지 않는 문장은 한국인 학습자들이 틀리기 쉬운 중국어 문법의 유형이라고 할 수 있다. 따라서 이들의 데이터를 통해 중국어 교육 현장에서 더욱 강조해야 할 중국어 문법 패턴 등을 데이터화할 수 있다.

셋째, 신경망 기계번역의 인공지능 학습 데이터를 제공한다. 신경망 기계번역은 자동학습 능력을 보유하고 있다. 즉, 보다 정확한 번역기로 변화하기 위해 언어 전문가들이 기계 번역의 오류 유형을 분석하여 인공지능이 정확한 데이터를 학습할 수 있도록 피드백을 입력하면, 신경망 번역의 오류를 개선하는데 도움이 될 수 있다.

앞서 지적한 바와 같이 현재 영어 통번역을 교육하는 교학현장을 중심으로 기계번역을 활용한 교수법 연구가 많이 진행되고 있는 것도 이상과 같은 이유일 것이다. 따라서 과학의 발전과 시대가 요구하는 인재를 양성함에 있어서 중국어 교육현장에서도 기존에 진행해왔던 교수법의 틀을 깨고, 학습자들에게 더욱 효과적이고 효율적인 신경망 기계번역 교수방안 연구가 지속되길 기대해본다.

### <참고문헌>

- 곽중철·한승희, <포스트에디팅 측정지표를 통한 기계번역 오류 유형화>, 《통번역학연구》 22, 2018.

- 김도영, 〈HSK코퍼스 중 한국 학생의 把字句 오류분석과 유추〉, 《중국어 문논총》 53, 2012.
- 김해정, 〈중국어 학습 과정에서 나타나는 오류분석〉, 충남대학교 석사학위 논문, 2000.
- 맹주익, 〈韓國人의 中國語 語彙 誤謬分析〉, 《중국학연구》 23, 2002.
- 민재홍, 〈한국 학습자의 중국어 오류 표현 분석〉, 《중국어교육과 연구》 31, 2020.
- 박덕준·박정구·오문의, 〈현대 중국어 유의어 연구〉, 《중국언어연구》 12, 2000.
- 박애화, 〈한국 학생 新HSK 4級 쓰기 偏誤 분석〉, 《東亞人文學》 26, 2013.
- 서보현·김순영, 〈기계번역 결과물의 오류유형 고찰〉, 《번역학연구》 19(1), 2018.
- 손정애, 〈한국인 학습자의 중국어 정도부사 ‘很’의 오류 분석〉, 《중국언어 연구》 33, 2010.
- 이경숙, 〈현대중국어 유의어 지도방안 연구: 유의어 오류분석을 중심으로〉, 한국외국어대학교 교육대학원 석사학위논문, 2001.
- 이명윤, 〈고등학교 중국어 교과서 유의어 분석 및 지도방안〉, 이화여자대학교 교육대학원 석사학위논문, 2006.
- 이상빈(2018), 〈학부 번역전공자의 구글 기계번역 포스트에디팅에 관한 현상학 연구〉, 《통번역학연구》 22, 2018.
- 이용아, 〈중국어 동사유의어 오류분석을 통한 중국어 효율적 지도방안〉, 공주대학교 교육대학원 석사학위논문, 2005.
- 홍연옥, 〈기계번역의 교육적 활용 가능성 고찰 및 제안 -把字문의 인간번역과 기계번역의 비교를 중심으로-〉, 《중국어 교육과 연구》 37, 2022.
- 姜昌义, 〈韩国留学生习得汉语程度副词的偏误分析与教学启示〉, 青岛大学硕士学位论文, 2010.

- 刘月华,《实用现代汉语语法》,北京:商务印书馆,2001.
- 鲁健骥,〈中介语理论与外国人学习汉语的语音偏误分析〉,《语言教学与研究》,1984.
- 吴垚垚,〈韩国HSK作文语法偏误分析—以大邱新华学院为例〉,西南科技大学硕士学位论文,2019.
- 肖奚强,〈韩国学生汉语语法偏误分析〉,《世界汉语》第2期,2000.
- 焉德才,〈韩国学生习得汉语程度副词‘很’语法偏误生态描写〉,《海外华文教育》第79期,2016.
- 郑月,〈韩国成年人汉语习得“很”“太”“是”的偏误分析〉,兰州大学 硕士学位论文,2018.
- 周佳佳,〈韩国学生副词“很”和“太”的使用偏误分析〉,四川师范大学 硕士学位论文,2017.
- Corder. S. P. Error analysis and Interlanguage, New York: Oxford University Press, 1985.
- Rod Ellis, Task-based Language Learning and Teaching, Oxford University Press, 2003
- ‘HSK动态作文语料库’,北京语言大学, <http://hsk.blcu.edu.cn>

### <Abstract>

Efforts to use machine translation for foreign language learning are currently conducted by researchers and educators who study English. On the other hand, attempts to use machine translation have been conducted in the Chinese language education, but discussions on specific teaching methods are still in the rudimentary stage. Therefore, this study attempts to propose a teaching method that can use machine translation effectively

in the field of Chinese teaching using an error correction teaching method with machine translation. The types of errors in Chinese translation that frequently appear in Korean learners, are considered, and the results are compared and analyzed by extracting their translation from neural machine translation. In addition, to modify the results output from machine translation and to use them as high-quality translation as a collaboration with advanced science technology and humans, the study attempts to post-edit each other's results, and to analyze the Chinese learning effect using machine translation.

As a result of the analysis, humans and machines are able to improve the results of translation and accumulate accurate learning data in the process of correcting each other's errors. The accumulated results will increase the ability of humans to speak a fluent language, and machine translation will consist of accurate function data essential in deep learning systems and will be used as re-learning data to increase translation accuracy.

Key Words : Neural machine translation, error, Serial verb construction, Existential sentence, synonym, hen(很)

