

## 챗GPT의 한중 통역 오류 고찰\*

공수(이화여자대학교)

### 1. 서론

2023년에 챗GPT가 등장함에 따라 통번역을 비롯한 다양한 분야에서 인공지능(AI)을 활용한 업무처리 방식에 중대한 변화가 일어나고 있다. 특히 챗GPT를 효과적으로 활용할 경우 업무의 효율성과 생산성이 크게 향상될 수 있다는 사실이 여러 분야에서 입증되고 있으며, 이는 통역 현장에서도 예외가 아니다.

그러나 여타 AI도구가 그렇듯, 챗GPT가 늘 정확한 결과물만을 도출하는 것은 아니므로 챗GPT를 무조건 신뢰하는 경우 오류가 발생할 수 있다. 특히 챗GPT의 언어 생성 능력이 고도화됨에 따라 겉보기에 유창하고 설득력 있는 결과물이 생성될 수 있게 되었는데, 이러한 현상이 오히려 사용자의 판단을 흐리게 만들 가능성을 제공하고 있다. 게레이로 등(Guerreiro et al., 2023)에 따르면 챗GPT가 번역업무를 수행할 때 의미를 왜곡하는 환각(hallucination) 현상이 발생하기도 한다.

이러한 상황에서 챗GPT의 통역 오류 유형을 정리하여 분석하면 사용자에게 도움을 줄 수 있다. 챗GPT 기반 통번역의 오류 유형을 사전에 인지하

---

\* 심사위원들의 소중한 심사의견을 바탕으로 본 논문 내용의 깊이와 완성도를 높일 수 있었습니다. 귀중한 조언과 피드백에 깊이 감사드립니다.

면 오류 발생 시 적절한 대처가 가능하기 때문이다. 한편, 챗GPT 번역 결과에 대한 평가는 다수 이루어졌으나, 통역 결과물에 대한 오류 분석 연구는 아직 미진한 상황이다. 이에 본 연구는 챗GPT가 수행한 한중 통역 결과물을 중심으로 오류 유형을 분석하고, 오류의 주요 원인을 고찰하는 것을 목적으로 한다. 이를 위해 관련 선행연구를 검토하여 챗GPT 통역 오류 분석의 이론적 토대를 마련하고, 통역 음성 데이터를 분석 대상으로 삼아 오류의 유형 및 원인을 고찰하고자 한다.

## 2. 선행연구 검토

현시점을 기준으로 챗GPT의 등장 이후 2년 남짓한 시간이 지났다. 이에 따라 통번역 학계에서도 관련 연구가 이루어지고 있으나 통역 오류 유형 분석에 초점을 맞춘 연구는 아직 미비한 상황이다. 챗GPT의 통번역 평가 연구에서는 번역 평가가 통역 평가에 비해 상대적으로 활발히 이루어지고 있다. 통역 평가와 번역 평가는 분명 서로 다른 분야이지만, 챗GPT의 번역 평가 선행연구를 통해 챗GPT의 번역 수행 능력을 확인할 수 있어 통역 평가 기반 마련을 위한 시사점을 얻을 수 있다. 이에 본 논문은 우선 챗GPT의 번역 평가에 대한 선행연구를 살펴보고 통역 평가에 관한 논의를 이어가고자 한다.

### 2.1 챗GPT 번역 평가

챗GPT의 번역 평가 연구로는 주로 특정 장르의 번역물에 대한 평가와 오류 분석 등이 있는데 챗GPT와 구글 등 타 기계번역물간의 비교 혹은 인간번역물과의 비교 분석이나, 장르가 다른 텍스트의 챗GPT 번역물의 비교 분석 등이 대부분이다. 챗GPT 번역물에 대한 평가를 보다 객관적으로 하기 위한 방법으로 ‘비교’가 적절하다고 판단되었기 때문인 것으로 보인다. 본고는 선행연구를 통해 챗GPT의 한중 통역 오류 분석 틀을 마련하는 것에 목적을 두기 때문에 본 절에서는 주로 중국어와 관련된 연구, 통역 분석에 기여할 수 있는 연구 및 오류 분석 틀이 구체적으로 서술된 연구를 살펴보

졌다.

중국어와 관련된 연구로는 이아형(2023)의 한중 방향 과학기술번역 평가가 있다. 해당 연구는 AI 번역의 과학기술 텍스트 번역 수행 능력을 평가하기 위해 챗GPT와 인간번역을 비교하여 오류 유형을 정리하였다. 분석 결과에 따르면, 원문 직역 오류, 단어 오류, 문장 분석 오류, 문장 종결 오류, 기타 오류 등 5개 큰 범주 하에 단어 오류, 어순 오류, 주어 누락, 단어 오류, 단문 층위 오류, 복문 층위 오류, 문장종결표현 오류, 시제 부사 탈락 등의 유형이 있다. 주로 문장을 단위로 한 분석이고 분석 단위가 129개로만 구성되어 있어 오류의 종류가 다양하지 않았다. 또한 문장 구조가 비교적 간단한 과학기술 텍스트를 대상으로 한 것도 오류 유형이 많지 않은 이유 중 하나로 보인다.

통역 오류 분석과 관련 있는 연구로는 조정민(2024)의 연설문 번역 평가와 박미정(2023)의 구어 텍스트 번역 평가가 있다. 해당 연구들은 번역을 다루는 연구이지만 통역 텍스트를 대상으로 하기 때문에 본 연구와 연관성이 있다고 보고 살펴보았다.

박미정(2023)은 통역 수업에서 교수자나 학생이 낭독하는 통역 원고를 대상으로 번역 평가를 실시했다. 구체적으로 신문기사를 구어체로 가공한 텍스트나 강연 내용 등의 음성 파일을 전사한 후 이를 가공한 텍스트들을 대상으로 분석했다. 특히 한국어가 고맥락 사회의 언어이고, 한국어에서는 함축적이고 축약적인 의미를 가지는 속담, 일반 통역 현장에서 정리되지 않은 구어체 표현이 많이 사용된다는 점을 고려하여 정리되지 않은 발화에 대한 번역의 단순화 및 명시화 보편소를 살펴보았다. 그리고 번역물을 평가할 때 동일한 ST에 대한 3개의 재생성번역을 포함한 번역물 4개와 구글 번역을 비교 분석했다.

분석 결과, 함축이 포함된 관용구의 경우 챗GPT는 맥락을 정확하게 해석해서 함축적인 의미를 명시화하여 문맥의 추론능력을 입증했다. 또한 함축이 포함된 속담일 경우, 속담의 지칭적 의미와 함축 의미를 정확하게 파악한 것으로 나타났다. 함축이 포함된 짧은 문장 속담에 대한 번역물에서는 앞뒤 문맥이 많아질수록 외연적 의미 너머의 내포적 의미가 번역되는 명시화를 보여주었다. 즉 챗GPT는 함축이 포함된 속담이나 관용구의 내포된 의

미를 명시화하여 정확하게 번역했다는 것이다. 반면 구글 번역은 문자의 표면적 의미만 반영하여 문맥에 맞지 않는 번역물만 생성하였다(박미정, 2023, p. 39-43).

한편 정리된 구어체 텍스트의 경우, 챗GPT가 더 자연스럽게 정확한 번역물을 내놓지만 구글 번역의 품질 역시 크게 떨어지지 않았다는 점이 발견되었다. 그러나 즉흥적인 강연 등 정리되지 않은 발화에서는 챗GPT와 구글의 격차가 큰 것으로 나타났다. 첫째, 정리되지 않은 중간 길이의 발화 텍스트의 경우 챗GPT는 생략, 추가, 문장 재배치, 문장 연결어 추가 등 다양한 전략을 사용하여 논리적으로 번역하였다. 그러나 구글은 직역만 하여 정리 능력을 전혀 보여주지 못했다. 둘째, 정리되지 않은 긴 길이의 발화 텍스트의 경우 챗GPT와 구글 모두 원문의 스피치 레벨 변화를 반영하였으나 구글은 의미 오역이 있는 것으로 나타났다. 또한 단어의 난이도가 높을 경우 챗GPT도 번역 실패를 보였다. 오역이지만 그럴듯한 번역으로 최대한 문맥에 맞춰 문장을 생성했다는 점(박미정, 2023, p. 51)이 흥미롭다.

박미정의 연구는 챗GPT가 타 기계번역보다 향상된 번역 능력을 보유하고 있다는 점을 보여주었다. 특히 함축 의미의 명시화, 인간과 유사한 번역 전략 사용 등이 눈에 띈다. 그럼에도 불구하고 챗GPT가 아직 고난이도 단어 등의 번역에서는 미흡하다는 점을 지적하기도 하였다. 해당 연구가 본 연구의 텍스트 선정에 시사점을 제공할 수 있는 부분은 실제 통역 현장에서 사용되는 통역 텍스트와 유사한 ‘정리되지 않은 구어체 텍스트’를 활용하였다는 점이다. 실제 통역 현장에 만날 수 있는 텍스트로 평가해야 챗GPT의 사용자에게 더 큰 도움을 줄 수 있기 때문이다. 물론 박미정(2023, p. 35)에서 언급된 것처럼 구어 텍스트는 구어성의 정도에 따라 여러 유형으로 나뉘며, 구어성이 가장 약한 연설문에서 구어체 텍스트, 구어 텍스트로 갈수록 구어성이 점점 강하다<sup>1)</sup>. 박미정의 연구에서는 연설문을 분석 대상에서 제외했는데 연설문도 통역 현장에서 만날 수 있다는 점을 고려하여 연설문을 대상으로 분석한 조정민(2024)의 연구도 살펴보겠다.

조정민은 챗GPT와 구글, 파파고 번역을 비교하여 대통령 연설문의 오류

1) 연설문, 신문기사, 구어로 가공한 텍스트, 음성 텍스트를 전사한 문자 텍스트 순으로 구어성이 점점 강하게 나타난다.

를 분석했다. 충실성, 유창성, 격식성, 화용성 등 4가지 큰 오류 범주 하에 원문 내용 왜곡, 누락, 추가, 문법적 오류, 일반 어휘 및 고유 명사/전문 용어의 부정확한 사용, 고위급 인사에 대한 경칭 생략 혹은 부정확한 사용, 다수의 청중을 대상으로 한 상황적 맥락에 맞지 않은 표현 사용 등 하위 범주를 설정했다. 우선 충실성 측면에서 챗GPT는 다음과 같이 간결화, 연결성 강화, 축약화 등 통역을 위한 번역문을 생성해 주는 특징을 보였다.

- 1) 챗GPT의 연설문 번역 결과물에서 나타난 특징으로, 두 개의 문장을 간결하게 연결하여 하나의 문장으로 구성하거나, 앞서 제시된 단어가 다음 문장에서 언급될 경우 이를 상위어나 축약된 어휘로 번역한 사례가 다수 확인되었다. (조정민, 2024, p. 102)
- 2) 챗GPT는 격식성과 화용성 오류가 상대적으로 적고, 통역 시간을 절약하기 위해 축약된 용어를 사용하는 등 통역용 번역으로서 적합한 특성을 보였다. 그러나 위와 같은 사례에서 확인할 수 있듯이, 충실성 측면에서 중요한 내용이 누락될 수 있으므로, 번역 후 반드시 원문과의 대조를 통해 충실성 오류를 보완하는 포스트 에디팅이 필요할 것으로 판단된다 (조정민, 2024, p. 103).

유창성 측면에서 챗GPT는 구글과 비슷한 문법 오류 수를 보였으며 파과고보다 뛰어난 성능을 보였다. 문법 오류 중 특히 명사와 형용사의 수 일치 규칙 오류가 많은 것으로 나타났다. 비교적 소수의 언어쌍에 해당하는 한국어-아랍어 방향이었기 때문에 챗GPT의 정확도가 떨어진 것으로 추측된다. 마지막으로 격식성과 화용성 오류 범주에서는 챗GPT의 경칭 오류와 화용성 오류가 가장 적은 것으로 나타났다. 반면 구글이나 파과고는 맥락에 적합한 경칭을 생략하거나 부적절하게 번역한 경향이 있는 것으로 나타났다.

조정민의 연구는 연설문이라는 특정 장르에 초점을 맞추어 연설문 통역의 시간 전략의 필요성, 경칭 사용의 필요성 등 특징을 고려하여 챗GPT의 번역문을 평가했다. 챗GPT는 구글이나 파과고 등 기존 기계번역에 비해 문장 간 연결 강화, 축약된 어휘로 간결한 번역문 생성 등 텍스트 특성에 부합한 번역 전략을 채택하였다. 그러나 문법 오류 등이 여전히 있어 사용자 입장에서 주의해야 한다는 신호를 보내고 있다.

여기까지 살펴본 연구들은 장르 등을 고려하여 오류 범주를 설정하거나 오류 범주를 설정하지 않고 분석을 통해 오류를 유형화하는 접근 방식을 채택하고 있다. 이들 연구에서 보여준 오류 범주는 특정 장르에 국한될 수 있다. 이와 달리 서로 다른 두 가지 장르의 챗GPT 번역문을 평가하는 연구는 비교적 포괄적인 오류 범주를 제시할 수 있다. 대표적인 예로 퍼펠(Puppel, 2024)의 연구를 들 수 있다.

퍼펠(2024)은 기술적 텍스트 및 창의적 텍스트의 번역결과물 품질을 비교하였다. 크게 정확도, 유창성, 용어(terminology), 문체(style), 지역 협약(locale conventions), 가독성(audience appropriateness) 등 6개로 평가기준을 분류하였는데, 정확도에는 오역, 추가, 누락, 미번역이 포함되고, 유창성에는 형태적 오류, 통사 구조 오류, 철자법 및 부호 사용의 부정확성, 모호한 표현, 응결성 및 응집성 결여가 포함된다. 용어 측면에서는 용어의 일관성 부족, 잘못된 용어 사용 등이 주요 오류에 포함되었으며, 문체와 관련해서는 레지스터 부적절성, 어색하거나 자연스럽지 않은 표현, 문체 일치성 부족이 포함되었다. 가독성에서는 문화적으로 부적절한 표현이나 최종 사용자에게 적합하지 않은 문장 구성이 포함되었다. 이러한 오류 틀을 활용하여 분석한 결과, 과학기술 텍스트에 비해 창의적 텍스트의 번역에서 오류가 더 많이 나타났고 대부분의 오류가 문체 범주에 속하는 것으로 나타났다. 과학기술 텍스트에서는 어색한 번역문이 가장 큰 오류 유형인 반면 창의적 텍스트에서는 문체 일치성 결여의 문제가 가장 크게 나타났다. 그러나 프롬프트가 증가함에 따라 이와 같은 문체 오류는 줄어들고 있다. 흥미로운 점은 정확성 오류가 프롬프트의 추가에 따라 증가한다는 것이다. 전체적으로 봤을 때 챗GPT는 원문의 통사 구조에 많이 따르는 편이며 이로 인해 문체, 정확성, 유창성 등 오류를 보이는 것을 알 수 있다. 또한 창의적 텍스트보다 과학기술 텍스트에서 완성도가 더 높은 번역결과물을 생성하는 것으로 나타났다.

앞서 검토한 선행연구를 통해 다음과 같은 시사점을 얻을 수 있다. 첫째, 챗GPT는 대부분의 경우 기존의 구글이나 파파고 등 보다 다양한 번역 전략을 사용하여 완성도 높은 번역물을 생성한다. 특히 과학기술 텍스트, 구어체 텍스트 등에 대한 번역 수준이 높은 편이다. 둘째, 챗GPT는 아직 창의적인 텍스트를 번역하는 데 한계가 있다. 셋째, 챗GPT는 프롬프트에 따라

번역 결과물을 다르게 생성할 수 있으나 프롬프트가 추가된다고 하여 번역 품질이 나아진다고 보장할 수는 없다. 넷째, 챗GPT 번역을 평가할 때 분석 대상의 장르에 따라 범주를 다르게 설정할 수 있다.

## 2.2 챗GPT 통역 평가

챗GPT 통역 평가에 대한 선행연구는 번역 평가와 달리 매우 제한적이다. 국내의 RISS, 구글 스칼라, 중국의 CALIS 등을 통해 관련 문헌을 조사한 결과, 챗GPT의 통역 결과물을 평가의 주요 대상으로 삼은 연구는 김혜림(2024)의 ‘인공지능(AI) 기술의 전문통역 교육 활용 방안: 챗GPT를 중심으로’가 유일한 것으로 확인되었다. 즉, 챗GPT의 통역 품질에 대해 간략히 언급한 논문은 다수 존재하지만 실제 통역 결과물을 중심으로 분석한 연구는 드문 편이다. 이에 본 연구는 우선 챗GPT의 통역 품질에 대한 기존의 논의를 살펴본 후, 김혜림(2024)의 연구 방법과 주요 결과를 보다 면밀히 검토하고자 한다.

리샤오둥과 신옌권(李晓东과 辛衍君, 2024)은 중국어 통역 교육에서의 챗GPT 활용에 대한 논문에서 챗GPT가 인간의 비언어 정보를 이해하는데 한계가 있다고 강조했다. ‘제스처, 말투, 자연스러운 멈춤, 얼굴 표정, 비언어적 목소리 등 비언어적 소통도 정보를 전달하는 중요한 방법’이라며 ‘(챗GPT)기능이 제한적이라 아직 인간의 말투, 뉘앙스, 표정 등 비언어적 정보를 이해하는데 부족하다’고 주장한 것이다. 또한 챗GPT는 LLM를 통해 어느 정도의 정확도를 확보할 수 있지만 응결성 부족이나 용어 일치성 부족 등 문제가 있다고 지적했다.

김혜림(2024)은 실험을 통해 챗GPT의 한-중, 중-한 방향의 순차통역 결과물에 대한 평가를 연구한 바 있다. 챗GPT를 통해 정치외교, 사회, 경제 분야의 순차통역 텍스트를 한국어와 중국어로 각각 생성한 후 통번역 대학원 출신 통역 강사들에게 평가를 요청하여 이론 연구였다. 평가기준은 이화여대 통번역대학원의 평가기준을 적용하였는바, 구체적으로 언어 구사력, 내용 및 논리, 용어 및 표현, 커버리지, 동시성, 전달력을 평가기준으로 적용하였다. 각 항목당 10점 만점으로 하여 양적 평가를 실시했다.

평가 결과를 구체적으로 살펴보면, 커버리지가 평균 8.9점으로 가장 높

은 점수를 기록하였으며, 그 다음으로 언어 구사력, 전달력, 내용 및 논리, 용어 및 표현 항목 순으로 평가되었다. 다섯 개 평가 항목 모두 평균 8점 이상이거나 이에 근접한 점수를 나타내어, 챗GPT의 통역 성능이 전반적으로 우수함을 보여주었다. 반면, 유일하게 평균 6점대를 기록한 항목은 ‘동시성’이었다. 6명의 평가자 중 3명이 7점을, 1명이 8점을 부여하였으며, 나머지 2명은 6점 이하의 점수를 부여하였다. 해당 항목의 낮은 점수에 대한 구체적인 언급은 없었으나 다른 항목들과 비교했을 때 상대적으로 기대에 미치지 못한 것으로 추정된다. 실제 평가자의 자유 의견을 살펴보면, 챗GPT의 통역 결과물에 대해 대체로 긍정적인 반응을 보였으며 유일하게 낮은 점수를 받은 ‘동시성’ 항목에 대해서도 특별한 결점이나 문제점을 명시한 경우는 드물었다. 다만 7점을 부여한 한 평가자가 “동시성 개선은 시간문제라고 보여진다”고 언급한 바 있어, 기술 발전에 따른 개선 가능성을 시사한다. 종합적으로 볼 때, 챗GPT의 통역 결과물은 통역전문가들로부터도 일정 수준 이상의 신뢰를 얻고 있는 것으로 판단된다. 예컨대 평가자들은 “레지스터에 맞는 언어 사용, 올바른 전문 용어, ST 정보에 대한 높은 커버리지가 인상적이었다”, “정확한 내용 전달이 인상적이었다”, “용어의 정확도가 굉장히 높다” 등과 같은 평가를 통해 챗GPT 통역의 정확성과 전문성을 높이 평가하였다.

챗GPT의 통역 결과물이 이렇게 높은 평가를 받은 이유에 대해 김혜림의 연구에서 특별히 언급된 바는 없지만 챗GPT의 뛰어난 성능이 큰 몫을 했을 것으로 판단된다. 또한 이러한 현상은 ST와도 무관하지 않아 보인다. 앞서 언급하였듯이 해당 연구에서 사용한 ST는 챗GPT를 통해 생성된 텍스트이다. 이에 대해 김혜림(2024, p. 38)은 ‘일반상식에 준하는 정보 난이도를 갖고 있으며 연사의 발화 역시 속도, 발음, 표현 등에서 상당히 표준화되어 있다’며 실제로 전문통역사에게 의뢰되는 ‘극히 고난도의 전문 정보를 다루는 경우가 많은’ 텍스트와 다르다고 설명했다.

김혜림의 연구는 통역 교육에서의 챗GPT 활용 가능성을 고찰하는데 목적을 두기 때문에 챗GPT의 통역 결과물에 대한 오류 분석 대신 전문가 평가를 활용했다. 이 때문에 평가의 범주가 앞서 살펴본 연구의 번역 평가의 범주처럼 세분화되어 있지 않고 대범주만 설정되었다. 그러나 본 연구는 챗

GPT의 오류 분석을 통해 오류 유형을 도출하고 오류 원인을 분석하는 데 목적을 두기 때문에 보다 세분화된 분석틀이 필요할 것이다.

위와 같이 챗GPT의 통역 및 번역 결과물 평가에 관한 기존 연구들은 각각의 연구 목적에 따라 연구 방법을 선택하여 결과를 도출하였다. 본 연구와 같이 챗GPT의 통역 결과물 오류를 고찰하는 연구가 아직 없는 관계로 본 연구에 바로 적용할 수 있는 분석틀도 없어 보인다. 그러나 상기 연구들은 통역 텍스트의 선정, 통역 평가의 범주 설정 등 본 연구의 연구 방법 중 중요한 부분에 시사점을 주고 있다. 따라서 이를 기반으로 다음 장에서 연구 내용을 논의하겠다.

### 3. 연구 대상 및 방법

본 장에서는 연구대상과 연구 방법에 대해 논의하고자 한다. 연구대상에는 평가대상의 선정, 평가대상의 설명 등이 포함되며, 연구 방법에는 오류 분석틀과 분석 절차가 포함된다.

#### 3.1 연구 대상

본 연구는 통역 현장에서 자주 접할 수 있는 연설문과 인터뷰 텍스트를 분석 대상으로 삼았다. 연설문과 인터뷰가 각각 계획된 담화와 비계획 담화의 특성을 지니고 있어 일반 통역 현장에서 접할 수 있는 전형적인 텍스트이기 때문이다. 즉 이 두 가지 텍스트 유형을 분석함으로써 챗GPT를 통역 도구로 사용하는 사용자가 실제로 어떤 통역 오류를 접할 수 있는지 추측할 수 있다. 이로써 본 연구의 목적인 챗GPT 통역 오류 유형 분석을 달성할 수 있고 궁극적으로는 향후 챗GPT 통역 활용에 있어 사용자들에게 보다 실용적인 시사점을 제공할 수 있을 것으로 기대된다.

본 연구는 연설과 인터뷰의 음성 텍스트로서 김혜림(2024)이 평가대상으로 선정한 주제인 정치외교, 사회, 경제 분야의 동영상을 하나씩 선정했다. 즉 총 6개 음성 텍스트가 평가대상이다. 정치외교, 사회, 경제 등 3가지 주제로 선택한 것 역시 통역 현장에서 해당 주제의 내용을 많이 접하기 때문

이다. 연설 부분은 문재인 대통령 퇴임 연설, 허준이 졸업식 축사, 추경호 부총리 “지경학적 분절의 시대...규제 효과구조개혁으로 대응” 등을 선정했고, 인터뷰 부분은 조희용 전 대사의 “응답하라 외교 인생!”, 유시민의 “정말 어마어마한 영상”, 김광석 한양대 겸임교수의 “경제 상황, 한마디로 대혼돈” 등 인터뷰를 선정했다. 상기 6개 동영상 선정 기준은 ‘계획된 발언 또는 즉흥적 발언이라는 특징을 가지는가’였다. 즉, 연설 동영상을 선택할 때는 계획된 내용의 특성을 살리기 위해 연설 내내 원고를 낭독하는 모습을 보인 동영상만 선정했고 인터뷰 동영상을 고를 때는 질문자와 대답하는 화자의 자연스러운 상호작용 등을 확인했다. 분석 대상은 아래 표와 같다.

표 1  
본 연구의 평가 동영상

번호	구분	주제	동영상 제목	링크	길이
1	연설	정치 외교	문재인 대통령 퇴임 연설	<a href="https://www.youtube.com/watch?v=liQMi0Ftw9Q">https://www.youtube.com/watch?v=liQMi0Ftw9Q</a>	12분 36초
2		사회	허준이 졸업식 축사	<a href="https://www.youtube.com/watch?v=OLDhaqosPtA&amp;t=13s">https://www.youtube.com/watch?v=OLDhaqosPtA&amp;t=13s</a>	6분 14초
3		경제	추경호 부총리 “지경학적 분절의 시대...규제 효과구조개혁으로 대응”	<a href="https://www.youtube.com/watch?v=srbwdx0dhug">https://www.youtube.com/watch?v=srbwdx0dhug</a>	13분 16초
4	인터뷰	정치 외교	“응답하라 외교 인생!” 조희용 전 대사의 외교 인터뷰	<a href="https://www.youtube.com/watch?v=C0d5aGdRx4w">https://www.youtube.com/watch?v=C0d5aGdRx4w</a>	8분 23초
5		사회	“정말 어마어마한 영상” 유시민 인터뷰	<a href="https://www.youtube.com/watch?v=AjLIAhKiIV0">https://www.youtube.com/watch?v=AjLIAhKiIV0</a>	7분 20초
6		경제	“경제 상황, 한마디로 대혼돈” 김광석 한양대 겸임교수 인터뷰	<a href="https://www.youtube.com/watch?v=QnUSkp9ZhEo">https://www.youtube.com/watch?v=QnUSkp9ZhEo</a>	6분 18초

각 동영상 마다 분량에 다소 차이가 있으나 본 연구가 챗GPT 통역 오류 유형이 무엇이 있는지, 원인이 무엇인지를 집중적으로 증명하는 데에는 지장이 없는 것으로 판단되기 때문에 해당 음성 텍스트를 분석 대상으로 선정

했다).

### 3.2 연구 방법

본 절에서는 오류 분석을 위한 오류 범주 설정 및 분석 절차를 논의하고자 한다.

#### 3.2.1 오류 범주

본 연구는 정확성, 유창성, 화용성을 3대 범주를 두고 각 범주에 세부적 하위 범주를 설정한다. 정확성에는 원문 내용 왜곡, 누락, 추가, 문법적 오류, 전문 용어의 부정확한 사용, 유창성에는 어색한 표현, 정확하지 않은 발음, 화용성에는 고위급 인사에 대한 경칭 생략 혹은 부정확한 사용, 다수의 청중을 대상으로 한 상황적 맥락에 맞지 않은 표현이 각각 포함된다.

표 2  
챗GPT 오류 범주

대분류	하위분류
정확성	원문 내용 왜곡
	용어의 부정확한 사용
	내용 누락
	내용 추가
	문법적 오류
유창성	어색한 표현
	정확하지 않은 발음
화용성	상황적 맥락에 맞지 않은 표현
	경칭 생략 혹은 부정확한 사용

- 2) 분석 텍스트의 분량과 균형성 또한 중요한 요소이나, 본 연구는 사례 분석을 중심으로 한 탐색적 연구에 해당하므로 통역 오류 유형의 도출 및 오류 발생 원인 분석에 초점을 맞추고자 한다. 향후 오류 유형의 분포나 경향을 보다 정량적으로 분석하는 후속 연구를 진행하는 경우 텍스트의 양적 균형성을 확보하여 보다 신뢰도 높은 결과를 도출할 예정이다.

본 연구는 오류 범주를 설정하는 과정에서 조정민(2024)과 퍼펠(2024)의 평가 범주를 주로 참고하되 본 연구의 목적도 함께 고려하였다. 조정민은 연설문만 분석했기 때문에 충실성, 유창성, 격식성, 화용성 등 4개 범주를 설정하였다. 그러나 본 연구는 인터뷰와 연설 텍스트를 모두 다루고 있어 조정민과 같이 격식성을 별도의 범주로 설정할 경우 인터뷰 텍스트의 격식성 오류가 현저히 적을 수 있다. 게다가 사실 격식성은 상대방을 고려하여 발화 내용을 조절한다는 측면에서 화용성에 포함될 수 있으므로 본 연구에서는 격식성을 화용성의 하위 범주로 포함시켰다. 또한 조정민의 충실성 대신 퍼펠의 정확성을 선택한 이유는 문법적 오류, 전문 용어의 부정확한 사용 등의 경우 원문에 대한 불충실성보다는 원문의 의미 전달 실패로 볼 수 있기 때문에 정확성으로 분류하는 것이 적절하다고 생각했기 때문이다.

### 3.2.2 분석 절차

본 연구는 챗GPT 4.0의 음성 모드를 활용하여 동영상 음성을 들려주고 통역하게 한 후, 통역된 내용을 평가하여 오류를 유형화하고 오류 원인을 분석하는 방식으로 진행되었다. 세부 절차는 다음과 같다.

#### 1) 분석 환경 설정

분석에 사용할 동영상을 준비한 후, 챗GPT 4.0의 음성 모드를 활성화하였다. 음성 모드 활성화 후, 챗GPT에 명령어 "들리는 내용을 중국어로 통역하시오"를 말하여 해당 작업을 수행할 준비를 완료하였다.

#### 2) 동영상 재생 및 음성인식

동영상을 재생하여 챗GPT가 음성인식을 하도록 하였다. 챗GPT는 동영상에서 인식한 음성을 일정한 구간으로 나누어 통역 결과를 출력하였다. 이 과정에서 구간의 길이는 챗GPT의 자동 인식 기능에 따라 정해지기 때문에 일정하지 않았다.

#### 3) 1차 평가와 2차 분석

1차 평가는 챗GPT가 구간별로 통역을 수행한 결과물을 대상으로 이루어

어졌다. 결과물 분석을 통해 본 연구에서 설정한 오류 범주에 따라 오류를 유형화하였다.

2차 분석에서는 1차 평가에서 도출된 오류를 중심으로 챗GPT의 번역문과 비교 분석을 실시하였다. 이를 통해 각 오류가 음성인식의 한계나 통역의 시간적 제약 등 통역 수행 과정에서 비롯된 것인지, 아니면 통역과 번역 모두에서 공통적으로 나타날 수 있는 오류인지를 구분하여 그 원인을 분석하고자 하였다. 이때 챗GPT 번역문과의 비교는 통역문에서 오류가 발생한 특정 부분에 한정하여 이루어졌으며 번역문 전체에 대한 분석은 포함하지 않았다.

## 4. 분석 결과

분석 결과, 표 2에서 제시한 오류 범주 중 ‘내용 추가’ 및 ‘경칭 생략 또는 부정확한 사용’을 제외한 모든 오류 유형이 확인되었다. 이 두 가지 오류 유형이 발견되지 않았다는 것은 분석대상 텍스트의 분량 제한, 챗GPT의 통역 시간 제약 등과 관련이 있는 것으로 보인다. 분석대상의 분량이 충분하지 않아 챗GPT 통역 시 나타날 수 있는 모든 오류 유형이 포착되지 않았을 가능성이 있으며, 통역 시간에 제약이 있는 챗GPT의 특성상 내용을 자의적으로 추가하여 통역하는 사례가 상대적으로 적게 나타난 것으로 해석할 수 있다.

다음으로 본 연구에서 발견한 오류를 정확성 오류, 유창성 오류, 화용성 오류 순에 따라 예시와 함께 제시하고 그 원인을 서술하고자 한다. 이어서 소결에서 오류 유형, 특히 연설문과 인터뷰의 오류 유형 특징 및 오류 원인을 정리한다.

### 4.1 정확성 오류

정확성 오류는 원문 내용 왜곡, 내용 누락, 내용 추가, 문법적 오류 등 통역 결과물이 원문의 의미를 정확하게 전달하지 못한 경우를 말한다. 본 연구에서 빈번하게 발견된 오류이기도 하다. 각각 예시를 통해 살펴보겠다.

#### 4.1.1 원문 내용 왜곡

원문 내용 왜곡은 발화자의 의도가 왜곡되어 전달된 것으로 한자어나 원문 음성이 뚜렷하지 않은 경우, 표준어가 아닌 경우가 여기에 해당한다. 본 연구에서 빈번히 관찰된 유형이다. 순서대로 예시를 살펴보겠다.

<예시 1> 한자어

원문3): 그동안 과분한 사랑과 지지로 성원해 주신 국민 여러분께 무한한 감사의 말씀을 드립니다.

통역문: 我想向那些给予我过多爱与支持的人们表达我无限的感激之情。

번역문: 对一直以来给予我厚爱与支持的国民们, 致以无限的感谢。

출처: 문재인 대통령 퇴임 연설

<예시 1>에서 확인할 수 있듯이 챗GPT는 ST의 ‘과분하다’를 ‘过多(지나치게 많다)’로 통역하여 발화자의 긍정적인 감사의 마음을 부정적으로 표현하였다. 그러나 번역문은 ‘厚爱’로 표현하여 국민의 깊은 사랑을 적절히 표현했다. 이를 놓고 봤을 때 해당 오류는 통역만의 오류로 통역 처리 시간제한 등 원인으로 인한 것으로 보인다.

<예시 2> 음성 모호

원문: 국제사회에서 외국 정부와 외국인을 상대로 매일 일상생활을 한다는 거기에서 내 능력을 발휘하고 국가 이익을 확보하고 추구해 가나는 일이라고 생각하면 결국 거기에서 내가 기본적으로 갖춰야 될 개인적인 어떤 자질이라고 할까요? .

통역문: 在国际社会中, 我每天都与外国政府和外国人互动, 管理自己的能力, 确保国家利益, 做着我认为最乐观和最重要的工作。在这种情况下, 我认为自己必须具备一些个人素质。

번역문: 在国际社会中, 我每天都要与外国政府和外国人打交道, 把这看作是发挥自身能力、争取并维护国家利益的工作的话, 归根到底, 我必须具备一些最基本的个人素质。

출처: 응답하라 외교 인생! 조희용 전 대사의 외교 인터뷰

3) ‘원문’은 음성 파일의 원문, 통역문은 챗GPT의 통역문의 전사 내용, 번역문은 챗GPT의 번역문을 뜻한다.

<예시 2>에서 ‘능력을 발휘하다’, ‘국가 이익을 추구하다’라는 내용이 각각 ‘管理自己的能力(능력을 관리하다)’, ‘做着我认为最乐观和最重要的工作(내가 가장 중요하고 낙관적인 일이라고 생각하는 일을 하고 있다)’로 오역되었다. 해당 오류는 챗GPT의 음성인식이 정확하지 않았기 때문으로 확인되었다. 실제로 음성인식 결과는 ‘내 능력을 관리하고’, ‘국가 이익을 확보하고 최고의 낙관화된 일이라고 생각하면’ 등으로 잘못 인식되었다. 원문의 발음이 또렷하지 않았고 지역 방언이 섞여 있어 음성인식이 원활하지 않았던 것으로 보인다.

한편, 번역문이 각각 ‘发挥自身能力(자신의 능력을 발휘하다)’와 ‘争取并维护国家利益(국가 이익을 추구하다)’로 처리된 것을 보면 음성인식이 정확히 이루어졌을 경우 통역 자체에는 큰 문제가 없었을 것으로 판단된다.

흥미로운 점은 음성인식 과정에서 오류가 발생한 경우 챗GPT가 잘못 인식된 내용을 그대로 번역하지 않고 도착어 텍스트에서 의미상 자연스럽고 그럴듯한 표현이 되도록 원문과 무관하게 의미를 조정한다는 점이다. 이러한 경향은 다음 사례를 통해 확인할 수 있다.

<예시 3>

원문: 첫번째 단계는 15년차 정도까지 15년차라고 하면 대개 과장 정도의 기간입니다.

통역문:第一阶段是大约15年的经验积累, 大约15年左右是大多数人能达到的职业生涯初期阶段。

번역문:第一阶段大约是到工作15年左右, 这个阶段通常对应的是科长级别的职务。

출처: 응답하라 외교인생! 조희용 전 대사의 외교 인터뷰

<예시 3>의 원문은 ‘과장 정도의 기간’인데 챗GPT는 ‘가장 정도의 기간입니다’로 잘못 인식했다. 그러나 통역문에서 잘못 인식된 ‘가장 정도’를 통역하지 않고 ‘직장의 시작 단계’로 처리했기 때문에 논리적으로는 자연스럽다. 이와 같은 오류는 음성인식이 잘 될 때도 발생했다. 다음 예시를 살펴보자.

<예시 4>

원문: 수학은 무모순이 용납하는 어떤 정의도 허락합니다. 수학자들 주요 업무가 그중 무엇을 쓸지 선택하는 것인데, 언어를 어떻게 사용할 것인가에 대한 가능한 여러 가지 약속 중 무엇이 가장 아름다운 구조를 끌어내는지가 그 가치의 잣대가 됩니다.

통역문: 数学允许任何一种定义, 只要它能被接受。数学家的主要工作之一是选择使用哪些定义, 而在可能的多种语言使用约定中, 哪些能引出最美的结构, 便成为衡量其价值的标准。

번역문: 数学允许任何不引起矛盾的定议。数学家的主要工作就是从中选择要使用哪些定义; 而在各种关于如何使用语言的可能约定中, 能够引出最优美结构的约定, 就是衡量其价值的标准。

출처: 허준이 졸업식 축사

<예시 4>의 원문은 수학자라는 발화자 신분에게 맞게 ‘무모순’이라는 수학적 어휘를 사용하였는데 챗GPT는 해당 부분을 ‘받아들이기만 하면 된다’로 오역했다. 그렇지만 통역문만 놓고 봤을 때는 큰 논리적 오류가 없다. 챗GPT가 자의적으로 해석한 것으로 보인다. 이창수(2024, p. 31)에 따르면 챗GPT는 문학 내용을 이해하지 못한 경우 ‘자의적으로 말을 만들어 넣는 등 문제’를 보인다. 이러한 자의적인 해석은 처리 시간이 비교적 짧은 통역에서 특히 두드러지게 나타나는 것 같다. 본 사례의 번역문에서도 ‘不引起矛盾(모순을 일으키지 않는다)’로 정확하게 반영하였다.

4.1.2 용어의 부정확한 사용

용어의 부정확한 사용은 외래어, 한자어, 신조어에서 특히 많이 나타났으며, 이는 용어나 고유명사의 오역 또는 혼동에서 비롯된 것으로 확인되었다.

<예시 5> 한자어

원문: 저의 퇴입사는 위대한 국민께 바치는 헌사입니다.

통역문: 我的大演讲是献给伟大的国民的。

번역문: 我的离任感言是献给伟大国民的一篇致敬词。

출처: 문재인 대통령 퇴임 연설

<예시 5>의 원문은 ‘퇴임사’인데, 챗GPT는 ‘대연설’이라는 어휘를 선택하여 오역이 되었다. 이와 반대로 번역문은 ‘離任感言(퇴임 소감)’으로 통역보다 정확하게 표현했다. 챗GPT의 통역 완성도는 아직 번역 완성도에 미치지 못하는 것으로 보인다.

<예시 6> 신조어

원문: 우리는 소·부·장 자립의 기회로 삼았고, 소·부·장 산업의 경쟁력 강화는 제조업의 경쟁력 강화로 이어졌습니다.

통역문: 我们在小部件上抓住机会, 小部件的竞争力提升了制造业的竞争力。

번역문: 我们把它作为实现材料、零部件、装备自主化的契机, 而材料、零部件、装备产业竞争力的提升也带动了制造业竞争力的增强。

출처: 지경학적 분절의 시대...규제 혁파구조개혁으로 대응 연설

<예시 6>에서 볼 수 있듯이 챗GPT는 ‘소·부·장’을 ‘小部件(작은 부품)’으로 번역하여 용어 선택이 정확하지 않았다. 음성인식에는 문제가 없으나 통역을 수행하기에 시간이 부족해 발생한 오류로 보인다. 번역문에서는 이를 ‘材料、零部件、装备(소재, 부품, 장비)’로 번역하여 해당 단어를 정확히 옮겨낸 것을 확인할 수 있다.

### 4.1.3 내용 누락

내용 누락은 시간, 수치, 핵심 정보 등의 생략으로 문학적 표현, 혹은 고유명사가 있을 때 누락이 발생한 것으로 확인되었다. 예시를 살펴보겠다.

<예시 7> 문학적 표현

원문: 먼 옛날의 나와, 지금 여기의 나와, 먼 훗날의 나라는 세 명의 완벽히 낯선 사람들을 이런 날들이 영성하게 이어 주고 있습니다.

통역문: 很久以后的某个地方, 这些日子将与三位完全陌生的人交织在一起。今天的毕业典礼, 可能就是这样的一天。

번역문: 遥远过去的我、此时此刻的我、以及遥远未来的我, 是三个彼此完全陌生的人, 而这样的日子, 正以一种松散的方式将他们连接在一起。

출처: 허준이 졸업식 축사

<예시 7>의 원문은 ‘먼 옛날의 나와, 지금 여기의 나와, 먼 훗날의 나’라는 철학적이면서도 문학적인 표현을 썼는데, 통역시 모두 누락되었다. 대신 ‘三位完全陌生的人(서로에게 완전히 낯선 세 사람)’으로 일부 정보만 살렸다. 반면 번역문은 ‘遥远过去的我、此时此刻的我、以及遥远未来的我(먼 옛날의 나와, 지금 여기의 나와, 먼 훗날의 나)’로 누락없이 번역했다. 번역문 역시 완벽하다고는 할 수 없지만 내용 누락 측면에서는 통역보다 더 높은 완성도를 보여주고 있다고 할 수 있다.

<예시 8> 고유명사

원문: 우리가 지금 겪고 있는 일이 저는 상당히 혁명적인 상황이라고 봐요. 예전 같으면요, 한국일보 기자 만나 가지고 문화부 기자 만나서 저 억울해요. 인터뷰하면 기자가 원고지 5장짜리 인터뷰 기사로 실어주고 사진 하나 내고 그게 끝이에요. 저런 정도의 어마어마한 영향력을 미치는 영상을 만들 수가 없습니다.

통역문: 我认为我们现在所经历的事情是相当革命性的。如果在过去，遇到不公正的事情，找韩国记者采访，记者可能会写一篇五到六页的采访文章，然后就结束了。那种程度的巨大影响力的视频是无法制作出来的。

번역문: 我认为我们现在所经历的事情是相当具有革命性的。在以前啊，比如说，我去见《韩国日报》的文化部记者，说“我很冤枉”，接受采访的话，记者会写一篇大约5页稿纸的采访报道，配上一张照片，事情就到此为止了。那时候根本不可能制作出像现在这样具有巨大影响力的视频内容。

출처: 응답하라 외교인생! 조희용 전 대사의 외교 인터뷰

<예시 8>을 보면 ‘한국일보’, ‘문화부’ 등 고유명사가 통역문에 반영되지 않았다. 반면 번역문은 이를 하나도 빠지 않고 모두 번역했다.

4.1.4 문법적 오류

문법적 오류는 어순, 조사, 구문 구조 등 문법상의 부정확성을 의미하며, 이 중 어순 오류가 많이 발생한 것으로 확인되었다.

<예시 9> 어순

원문: 취업 준비, 결혼 준비, 육아 교육 승진 은퇴 노후 준비를 거쳐 어디 병원 그럴듯한 일인실에서 사망하기 위한 준비에 산만해지지 않기를 바랍니다.

통역문: 工作、结婚、育儿、教育、晋升、退休、养老的准备, 以及最终在医院的病房里准备离世的日子, 我希望你不要迷失在这些过程中。

번역문: 希望你不要被为了就业准备、结婚准备、育儿教育、升职、退休、养老以及最终在某家医院体面的单人病房中去世的准备所分散了注意力。

출처: 허준이 졸업식 축사

<예시 9>에서는 ‘工作、结婚、育儿、教育 ... 准备’와 ‘以及最终在医院的病房里准备离世的日子’라는 긴 병렬 구문이 ‘我希望你不要迷失在这些过程中’라는 동사의 목적어 역할을 하고 있다. 목적어 부분의 병렬이 매우 길고 각각 원활한 언어관계를 형성하고 있지도 않아서 어순을 바꿔야 한다. 이러한 통역 오류는 원문의 어순과 통사 구조를 그대로 유지하여 나온 결과다. 반면 챗GPT의 번역문이 술어인 ‘希望(바란다)’을 문두에 배치함으로써 뒤에 나온 명사들과 언어관계를 잘 형성하고 있다. 이를 두고 봤을 때 통역 처리의 시간적 제한이 해당 통역 오류를 유발한 것으로 추측된다.

## 4.2 유창성 오류

유창성 오류는 어색한 표현과 정확하지 않은 발음을 포함하는데 각각의 예시를 통해 살펴보겠다.

### 4.2.1 어색한 표현

챗GPT 통역의 유창성 오류는 주로 어색한 표현에 집중되어 있다. 특히 한자어 사용 빈도가 높은 대외 연설문을 통역할 경우, 이러한 유창성 오류가 두드러지게 발생하는 경향이 관찰되었다.

<예시 10>

원문: 저는 이제 평범한 시민의 삶으로 돌아가 국민 모두의 행복을 기원하며 성공하는 대한민국의 역사를 응원하겠습니다.

통역문: 我现在将回归普通市民的生活, 祈愿全体国民的幸福, 并支持成

功的韩国历史。

번역문: 我现在将回归为一名平凡的市民, 衷心祝愿全体国民幸福, 并将继续为大韩民国的成功历史加油助力。

출처: 문재인 대통령 퇴임 연설

<예시 10>에서 챗GPT는 ‘성공하는 대한민국의 역사를 응원하겠습니다’를 ‘支持成功的韩国历史(성공한 한국 역사를 지지한다)’로 통역하였다. ‘支持(지지)’은 보통 발생하고 있거나 발생할 일에 대한 동작에 쓰이기 때문에 과거를 가리키는 단어인 ‘历史(역사)’와는 연어관계가 잘 성립하지 않아 어색하다. 이와 같은 어색한 표현은 번역문에도 발견되었다. 번역문은 ‘继续为大韩民国的成功历史加油助力(계속해서 대한민국의 성공 역사를 위해 힘을 보태고 응원할 것이다)’로, 통역문과 마찬가지로 어색한 표현이 되었다. 과거를 뜻하는 ‘历史(역사)’를 위해 힘내라고 하는 것은 중국어에서 논리적이지 않다. 이렇듯 통역과 번역 결과물 모두에 오류가 발견된다는 점을 고려할 때, 챗GPT는 아직 한중 번역 과정에서 출발어 텍스트의 구조를 효과적으로 재구성하고 의미 중심의 표현으로 전환하는 전략을 충분히 활용하지 못하고 있는 것으로 보인다.

#### 4.2.2 정확하지 않은 발음

챗GPT의 중국어 발음은 대체로 정확하지만 ‘sh’, ‘zh’와 같은 권설음을 발음할 때 정확하지 않다. 그러나 못 알아들을 정도는 아니다. 이는 챗GPT의 일부 음성 합성 엔진이 발음 모델을 충분히 정교하게 훈련하지 못하여 중국어의 권설음(sh, zh 등)을 자연스럽게 구현하지 못하고, sh와 s, zh와 z의 구분이 잘되지 않는 데 기인한 것으로 보인다.

또한 AI의 특별한 기계 목소리도 표준 중국어의 ‘字正腔圆(자정장원: 발음이 정확하고 곁이 좋다)’ 발음 기준을 충족하지 못한다. 발음이 심각하게 나쁘다기보다는 알아들을 수는 있으나 학습자에게 발음 학습 용도로 제공하기에는 적절하지 않다고 생각한다.

### 4.3 화용성 오류

화용성 오류에서는 상황적 맥락에 맞지 않은 표현만 발견되었는데, 이를 예시를 통해 살펴보겠다.

#### 4.3.1 상황적 맥락에 맞지 않은 표현

상황적 맥락에 맞지 않은 표현에는 공식적인 자리나 다수의 청중을 대상으로 한 발언 시 부적절한 표현 사용이 포함된다. 이번 연구 분석에서 챗GPT가 다수의 청중을 대상으로 하는 연설문 인사를 개인 청중을 대상으로 하는 표현으로 부적절하게 처리한 부분이 발견되었다. 예시를 살펴보겠다.

<예시 11>

원문: 안녕하세요? 반갑습니다. 저는 경제부총리 겸 기획재정부 장관 추경호입니다.

통역문: 您好, 很高兴见到您。我是副总理兼计划财政部长, 周京宇。

번역문: 您好! 很高兴见到大家。我是韩国经济副总理兼企划财政部长官秋庆镐。

출처: 추경호 부총리 “지경학적 분절의 시대” 연설

<예시 11> 원문은 연설문의 인사말이다. 그러나 챗GPT는 이를 마치 특정 개인에 대한 인사로 통역했다. 다수의 청중을 대상으로 한 실제 상황에 부합하지 않는다. 그러나 전체 맥락을 들려주고 다시 통역을 시킨 결과, 보다 적절한 표현이 도출되었다.

이 사례는 챗GPT의 통역 결과가 상황에 따라 유동적으로 달라질 수 있음을 보여주며 통역에 활용할 때 사전 상황 정보를 제공하면 정확도를 높일 수 있다는 점을 시사한다. 이와 같이 프롬프트를 활용해 결과물의 품질을 향상시킨 사례는 번역학 연구에서도 찾아볼 수 있다. 야마다(Yamada, 2023, p. 203)에 따르면 마케팅 문서나 문화적으로 민감한 표현을 번역할 때, 번역 목적과 대상 독자에 대한 정보를 프롬프트에 반영하면 챗GPT를 통해 유연하고 정확한 번역 결과를 얻을 수 있다. 본 연구의 <예시 11>에서도 처음에 ‘번역하라’만 지시했을 때는 적절하지 않은 ‘您好!’로 번역된 반면, ‘다음은 연설문의 일부 내용이다’라는 텍스트에 대한 설명을 추가하니 ‘您好!’

가 ‘大家好’로 적절히 수정되었다.

#### 4.4 소결

본 연구에서 발견한 챗GPT의 통역 오류 유형 및 원인을 다음과 같이 정리한다.

##### 4.4.1 챗GPT 한중 통역의 오류 특징

본 연구에서 발견한 챗GPT 한중 통역 오류에는 원문 내용의 왜곡, 정보 누락, 문법적 오류, 용어의 부정확한 사용 등 정확성 오류, 어색한 표현, 부정확한 발음 등의 유창성 오류, 상황적 맥락에 부합하지 않는 표현 사용과 같은 화용성 오류가 있다. 이러한 유형 특징은 다음과 같다.

첫째, 한자어, 문학적 표현, 출발어 텍스트의 통사 구조에서 벗어나 의미를 살려서 통사 구조를 재구성해야 하는 등 한중 통역에서 일반 학습자들이 다루기 어려운 부분에 대해서는 챗GPT도 오류를 범하기 쉽다. 즉, 챗GPT도 한중 통역의 난제들을 처리하기에 아직 한계가 있다. 챗GPT를 통역 학습의 도구로 활용하고자 하는 사용자들은 이 점을 인지하여 챗GPT의 통역 결과물을 판단할 필요가 있다.

둘째, 계획된 담화 유형인 연설문의 경우 발화가 비교적 명료하고 구조적으로 정형화되어 있어 음성인식의 정확도가 높은 편이며, 이에 따라 정보 누락 등의 오류 발생이 많지 않았다. 그러나 비계획된 담화 유형은 발화자의 발화 속도가 빠르고 발음이 명확하지 않은 경우가 많아 음성인식 과정에서 오인식 및 누락이 빈번하게 발생하는 경향을 보였다.

셋째, 챗GPT의 한중 통역 결과는 번역 결과에 비해 전반적으로 완성도가 낮은 것으로 확인되었다. 특히 용어의 부정확한 사용, 누락, 어색한 표현 등과 같은 정확성 측면의 오류가 두드러졌다. 앞서 제시한 사례들을 통해 확인할 수 있듯이 동일한 원문을 챗GPT로 통역한 결과에서는 오류가 다수 나타난 반면, 번역 결과에서는 동일한 오류가 발견되지 않는 경향을 보였다. 이는 현재 챗GPT가 통역 수행 시 번역에 비해 상대적으로 낮은 정확성을 보이고 있음을 시사한다.

#### 4.4.2 챗GPT 한중 통역 주요 오류의 원인

챗GPT의 한중 통역에서 나타나는 주요 오류 유형의 원인을 다음의 네 가지 측면에서 고찰할 수 있겠다. 첫째, 통역 결과와 번역 결과의 비교, 둘째, 통역 수행 시 원문의 전사 내용 분석, 셋째, 챗GPT 자체의 응답 경향, 넷째, 일반적인 AI 통역의 한계이다.

우선, 통역의 시간적 제약은 주요한 오류 발생 원인 중 하나다. 앞서 4.4.1절에서 언급한 바와 같이 동일한 원문에 대해 챗GPT가 번역을 수행할 경우에는 오류가 나타나지 않지만, 통역 수행 시에는 오류가 관찰된 경우가 많았다. 이는 챗GPT가 통역 과정에서 실시간으로 정보를 처리해야 하기 때문에 문장의 전체 구조를 재구성하거나 의미를 충분히 반영하는 데 어려움을 겪는 것으로 해석된다.

둘째, 음성 인식의 정확성 문제 역시 오류 발생 요소로 분석되었다. 실제 분석 과정에서 확인한 바에 따르면, 챗GPT가 자동으로 전사한 음성 파일 내에는 발음이 유사한 단어를 오인식하여 결과적으로 의미 왜곡 또는 핵심 정보의 누락을 초래한 사례가 다수 존재하였다.

셋째, 챗GPT의 고유한 응답 경향, 특히 이른바 ‘환각(hallucination)’ 현상과 관련된 오류가 관찰되었다. 이는 챗GPT가 주어진 맥락이나 정보 외에 자의적으로 해석하거나 사실과 다른 내용을 생성하는 현상으로, 특히 통역 과정에서 원문에 없는 의미가 덧붙여지는 방식으로 나타났다. 이러한 경향은 사용자가 외국어에 익숙하지 않을 경우 더욱 문제를 일으킬 수 있으므로 주의가 필요하다.

넷째, 전체 문맥에 대한 이해 부족이나 코퍼스 학습 범위 밖의 표현, 예컨대 신조어 등은 통역 오류의 또 다른 원인으로 지적된다. 챗GPT가 해당 표현을 학습 데이터에서 충분히 접하지 못해 의미 전달이 부정확하거나 불완전한 결과가 도출될 수 있다.

#### 4.4.3 기타

본 연구의 분석과정에서 통역 오류 등 챗GPT의 통역과 관련된 문제점 외에 지시 불이행으로 인한 단편적 통역, 주변 소음에 의한 통역 중단 등 기술적 문제점도 발견되었다.

예를 들어 화자의 발언이 끝난 후 전체 내용을 통역하도록 지시하였음에도 불구하고 챗GPT는 화자의 일시적 멈춤(pause)을 문장 종결로 잘못 인식하여 성급하게 통역을 시작하는 경우가 많았다. 특히 구어 성격이 강한 인터뷰일 경우, 챗GPT가 멈춤을 잘못 식별하는 오류를 많이 범했다. 더불어 통역 도중 주변에서 다른 소리가 발생할 경우, 통역이 갑작스럽게 중단되거나 시스템이 혼란을 일으켜 이전에 이미 통역한 내용을 반복 출력하는 현상도 있었다. 이어서 통역하도록 지시하더라도 누락된 부분을 정확히 복원하지 못했다.

이렇듯 현재 챗GPT의 통역 기술이 실시간, 맥락 중심의 인간 커뮤니케이션을 완전히 대체하기에는 여전히 기술적, 인지적 한계가 존재함을 알 수 있다.

## 5. 결론 및 시사점

본고는 챗GPT가 한중 통역 수행 시 어떤 유형의 오류를 보이는지와 그 원인이 무엇인지를 분석하는 데 목적을 두었다. 이를 위해 기존 연구를 바탕으로 충실성, 유창성, 화용성 등 세 가지 대범주를 설정하고 각 범주에 세부 하위 항목을 도입하여 연설문 및 인터뷰 각 3편의 녹음 자료를 분석 대상으로 삼았다.

분석 결과, 원문 내용의 왜곡, 정보 누락, 문법적 오류, 전문 용어의 부정확한 사용 등 정확성 관련 오류가 빈번하게 나타났다. 그 외에도 어색한 표현, 부정확한 발음 등의 유창성 오류, 상황적 맥락에 부합하지 않는 표현 사용과 같은 화용성 오류도 확인되었다. 이러한 오류들은 특히 한자어, 신조어, 문학적 표현이 포함되거나 음성이 불분명하거나 사투리가 사용된 경우에 빈번하게 발생하였다.

그러나 대부분 오류는 음성 기반 통역 수행 시에만 나타났으며 동일한 원문을 텍스트 번역으로 지시하여 생성한 번역문에서는 같은 오류들이 거의 발생하지 않았다. 이는 챗GPT의 통역 오류가 주로 실시간 처리의 시간적 제약이나 음성인식 과정의 오류에 기인한 것임을 시사한다. 하지만 한자

어 등 일부 오류는 번역에서도 마찬가지로 확인되었다. 이를 통해 챗GPT 자체적인 코퍼스 훈련 부족, 자의적 해석의 경향 등이 통역 오류를 야기하는 요인 중 하나임을 알 수 있다.

또한 본 연구에서는 챗GPT 통역의 기술적 한계 역시 발견되었다. 지시 불이행으로 인한 단편적 통역, 주변 소음으로 인한 통역 중단 및 내용 왜곡 등이 그 예다. 이러한 기술적 문제는 사용자의 통역 경험에 큰 불편을 초래할 수 있다.

이러한 분석 결과는 챗GPT의 통역 오류가 단순한 언어 처리 능력의 한계에서 비롯된 것이 아니라 기술적 제약, 실시간 처리 환경 등 복합적인 요인에 의해 영향을 받는다는 점을 시사한다. 따라서 향후 챗GPT의 통역 정확성을 제고하기 위해서는 통역 모델 자체의 성능 고도화뿐만 아니라, 음성 인식 기술의 정밀성 향상, 실시간 처리에 따른 시간적 제약 완화 등 다각적인 기술적 보완이 병행되어야 할 것이다.

본 연구를 통해 도출된 시사점은 다음과 같다.

첫째, 챗GPT의 통역 성능은 전반적으로 높은 수준을 보였지만 여전히 정확성 오류가 빈번하게 발생하고 있어 실사용 시 이를 전적으로 신뢰하기 보다는 사후 검토가 필요하다. 특히 챗GPT는 표면적으로 자연스러운 결과물을 산출하더라도 원문 내용과 비교했을 때 의미 왜곡이나 정보 누락이 있을 수 있어 통역물만 보고 오류를 인지하기 어려울 수 있다. 따라서 사용자는 항상 원문과 결과물을 함께 검토해야 한다. 해당 외국어에 대한 이해도가 낮거나 잘 모르는 사용자의 경우, 다른 기계 통역 도구와의 교차 검토(cross-check)를 통해 오류 가능성을 줄이는 방식도 효과적인 활용 전략이 될 수 있다.

둘째, 챗GPT의 중국어 발음은 일부 음소, 특히 권설음 계열에서 부정확한 경우가 종종 발견되었다. 이로 인해 중국어 학습자가 챗GPT의 음성을 따라 발음을 연습할 경우, 잘못된 발음을 습득할 위험이 존재한다. 따라서 발음 학습시 챗GPT는 보조 자료로 활용하되 신뢰할 수 있는 원어민 발음 자료를 병용하는 것이 바람직하다.

셋째, 챗GPT에게 통역 업무를 수행하게 할 경우 화자와 청중의 관계, 주제, 상황적 맥락 등 충분한 배경 정보를 사전에 제공하면 챗GPT의 문맥 이

해와 표현 선택의 정확도 향상에 도움이 될 수 있다. 이는 특히 화용적 오류를 줄이는 데 긍정적인 영향을 줄 수 있다.

넷째, 본 연구 결과에 따르면 챗GPT는 통역보다 번역에서 더 뛰어난 처리 능력을 보였다. 이는 통역 수행 시의 시간적 제약과 음성인식 오류 등 기술적 요인에 기인하는 것으로 보인다. 따라서 통역 학습자들이 챗GPT를 학습 도구로 활용하고자 할 때는 직접 통역 수행을 시키기보다는 번역 기능을 활용하는 방법이 보다 효과적일 수 있다. 예를 들어 챗GPT의 번역문을 자신의 통역 결과와 비교 분석함으로써 보다 자연스럽게 정확한 표현을 학습하는 데 도움이 받을 수 있다. 물론 챗GPT의 번역에도 부정확한 경우가 있어 비판적인 시각이 필요하다.

본 연구는 챗GPT의 한중 통역 오류를 유형화하여 분석하고 원인을 분석했다는 점에서 의미가 있다. 그러나 분석 대상의 장르가 제한적이고 분석 분량 또한 충분하지 못하다는 한계가 있어 본 연구의 결과를 전반적인 한중 통역 상황에 일반화할 수는 없다. 또한 정량적 분석을 하지 못해 오류의 분포 등 양적 결과를 도출하지 못했다. 그럼에도 불구하고 본 연구는 챗GPT를 활용해 한중 통역을 수행하는 사용자들에게 실질적인 시사점과 참고 자료를 제공할 수 있다는 점에서 의미가 있다. 향후 연구에서는 다양한 장르로의 자료 확대, 정량적 분석 방법 도입 등을 통해 본 연구의 한계를 보완하고 챗GPT 기반 통역 기술의 실질적 활용 가능성과 한계를 보다 폭넓게 규명해 나가고자 한다.

## 참고문헌

- 김혜림. (2024). 인공지능(AI) 기술의 전문통역 교육 활용 방안: 챗GPT를 중심으로. *통역과 번역*, 26(2), 27-65.
- 박미정. (2023). 생성형AI 와 기계번역 — 챗GPT 번역을 통한 한일통역교육 고찰. *통번역학연구*, 27(3), 27-56.
- 이아형. (2023). 챗GPT 의 번역 수업 활용 방안 고찰 — 한중과학기술번역을 중심으로. *중국어문학지*, 85, 153-176.

- 이창수. (2024). 챗 GPT, 파파고, 인간 번역가 간의 한영 문학번역 차이점 연구. *번역학연구*, 25(2), 11-37.
- 조정민. (2024). 연설문 한-아 기계 번역의 가용성 연구: 구글 번역, 챗 GPT, 파파고 비교를 중심으로. *중동문제연구*, 23(3), 89-114.
- 李晓东, 辛衍君. (2024). Chatgpt支持下的国际中文教育口译教学实践与启示 [챗GPT 기반으로 하는 글로벌 중국어 교육에 있어서 통역 교육 실천과 시사점]. *语言服务研究*, 4(1), 234-246.
- Guerreiro, N. M., Alves, D. M., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., & Martins, A. F. (2023). Hallucinations in large multilingual translation models. *Transactions of the Association for Computational Linguistics*, 11, 1500-1517.
- Puppel, M. (2024). *A comparative study of technical and creative text translation: evaluating the performance of ChatGPT* [Master's thesis, University of Malta].
- Yamada, M. (2023). Optimizing Machine Translation through Prompt Engineering: An Investigation into ChatGPT's Customizability. MTS 2023 학술대회 논문집 (pp. 195-204). <https://arxiv.org/abs/2308.01391v2>

## An analysis of errors in ChatGPT's Korean-Chinese interpretation

**Su, Kong** (kongsusky1@ewha.ac.kr)

Ewha Womans University

### Abstract

This study investigates the types and causes of errors in Korean-Chinese interpretation outputs generated by ChatGPT. Using 60 minutes of recorded speeches and interviews, the study categorizes errors into three domains—fidelity, fluency, and pragmatics—each with its own set of subcategories. The findings reveal that accuracy-related errors—such as content distortion, omissions, grammatical mistakes, and misuse of technical terms—were frequent. Fluency-related and pragmatic errors, including awkward expressions, mispronunciations, and contextually inappropriate usage, were also observed, especially in cases involving Sino-Korean words, neologisms, unclear audio, or dialects. Most errors occurred only in speech-based interpretation, not in text-based translation of the same source, suggesting that real-time processing constraints and speech recognition issues are key factors. Technical limitations, such as fragmented output and interference from background noise, were also noted. Although honorific misuse and content addition were defined as error types, they were not detected, likely due to time constraints and limited data. The study concludes that while ChatGPT performs well overall, human review remains essential.

**Keywords:** Interpreting studies; ChatGPT; Korean-Chinese interpretation; interpretation errors

**키워드:** 통역학, 챗GPT, 한중 통역, 통역 오류

공수

이화여자대학교 강사

kongsusky1@ewha.ac.kr

논문 투고일: 2025년 5월 15일

1차 심사 완료일: 2025년 6월 1일

2차 심사 완료일: 2025년 6월 10일

게재 확정일: 2025년 6월 15일