

[칼럼이 있는 AI 특] AI 번역, 따라올테면 따라와 봐

구글 번역기부터 네이버 파파고까지. 인공지능(AI) 번역 서비스는 이제 우리 일상 속 한 부분이 되었다. ‘구글 번역체’라는 표현이 무슨 의미인지 모르는 사람을 찾아보기 힘들 정도다.

미숙한 번역의 대명사로 ‘구글 번역체’를 사용하면서도 이미 많은 사람들이 AI 번역을 쓰고 있다. 파파고는 출시 직후부터 해외여행 필수 준비물로 자리잡았다. 아마존이나 알리 익스프레스에서 ‘직구’하는 건 구글 페이지 번역 클릭 한 번이면 된다.

반면 사용자들에게 그저 편리하게 느껴지는 AI 서비스가 관련 직종 종사자들에게는 위협이 되곤 한다. 인간 번역가라는 직업은 곧 역사 속으로 사라질 운명일까?

전문가가 말하는 정답은 ‘아니오’다. AI 번역기가 사회 속에 자리잡으면서 인간 번역가들은 오히려 일자리를 잃을 걱정을 덜게 됐다. 번역이야말로 AI 가 아닌 인간이 잘 할 수 있는 일일지도 모른다.

[칼럼] 인공지능 번역, 따라올테면 따라와 봐

남영자 중앙대 인문콘텐츠연구소 HK 연구교수



남영자 중앙대 인문콘텐츠연구소 HK 연구교수

그야말로 AI 가 연중무휴 종횡무진 전방위적 활약을 펼친다. AI 에 평이 박한 사람일지언정 AI 에 의존하는 대표적인 분야 가운데 하나는 기계 번역이 아닌가 싶다. 특히 대학생들의 경우 리포터나 영문 원서 번역을 위해 네이버의 ‘파파고’와 ‘구글 번역’을 이용한 경험이 있을 것이다. 기계 통번역 기술은

규칙기반 방식에서 통계기반 방식으로 발전했고, 2010년대 후반 기계 학습을 활용한 인간의 뇌 신경망을 모방한 신경망 기계번역으로 진화했다.

인공지능에 기반한 신경망 기계번역은 스스로 학습하는 방식으로 품사, 격식체 및 비격식체와 같은 언어의 사용역(register) 등을 반영해 인간과 상당히 근접한 수준으로 번역한다. 이제 인간 번역가의 위상이 흔들리지 않을지 염려된다. 일각에서는 멀지 않은 장래에 기계번역이 인간 번역가를 완전히 대체할 것으로 전망한다.

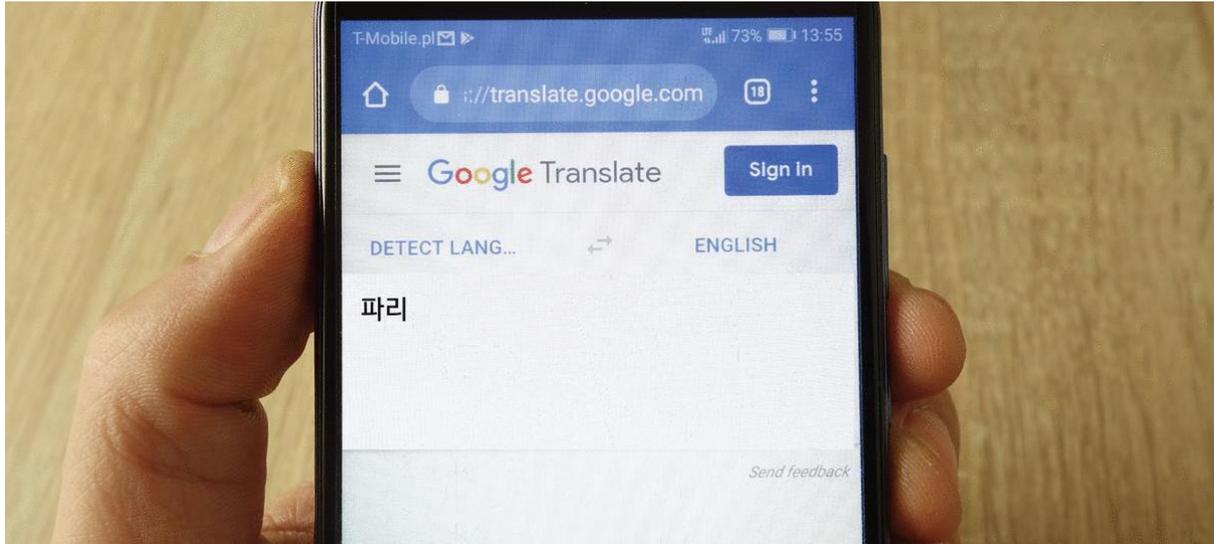
과연 AI까지 탑재해 거침없이 진격하는 기계번역은 어디까지 왔을까? 이에 필자는 구글 번역과 파파고의 한영 및 영한 번역 능력을 비교해 보았다. 먼저 필자가 좋아하는 ‘어전’을 사용한 문장을 만들어 보았다.

동생이 신의 직장이라 불리는 공사에 합격했다. 동생을 축하하기 위해 우리 집 식구는 어제 저녁 외식을 했다. 동생이 시킨 어전(漁煎)이 나왔을 때 갑자기 파리 한 마리가 날아다녔다. 나는 “감히 어전(御前)에서 파리가 날아다니다니?”라 말했다. 동생은 “짐(朕)이 오늘은 기분이 매우 좋으니, 어전(御前)에서 어전(漁煎) 위를 비행하는 파리를 용서하노라”했다.

My brother passed the construction called God's workplace. To celebrate my brother, my family ate out last night. Suddenly, a fly flew around when the fish tank my brother ordered came out. I said, "How dare a fly fly from the fishing grounds?" My brother said, "Jim is in a great mood today, so I forgive Paris for flying over the fishing grounds at the palace." (파파고 번역)

My younger brother passed the construction called God's workplace. To celebrate my brother, my family ate out last night. When my brother's order came out, a fly suddenly took flight. I said, "How dare flies fly around the palace?" My younger brother said, "Jim is in a very good mood today, so I forgive the fly that flies over it." (구글 번역)

파파고와 구글 번역 모두 ‘동생’을 ‘brother’로 번역해 기계 번역의 고질적 문제인 성별 바이어스(gender bias)를 반영한다. 구체적으로 이름이 언급되지 않은 공공 기업체인 ‘공사’는 건설, 공사를 뜻하는 ‘construction’으로 번역됐다. 파파고는 ‘어전(漁煎)’을 어류 탱크인 ‘fish tank’, ‘어장’을 ‘fishing grounds’로 번역했고, 임금의 앞인 ‘어전(御前)’을 임금이 거쳐하는 궁전인 ‘palace’로 번역했다. 구글 번역은 동생이 시킨 ‘어전(漁煎)’을 동생이 주문한 것으로 영리하게 회피해 번역했고, ‘어전(御前)’을 ‘palace’로 번역했다.



'파리'는 번역기에서 이렇게 다르게 번역될 수도 있다. (사진 편집=임채린 기자)

한편 파파고는 날아다니는 '파리'는 'fly'로, 용서의 대상이 되는 '파리'는 프랑스 수도 'Paris'로 번역했다. 구글 번역은 '파리'를 일관되게 'fly'로 번역했다. 두 번역기 모두 임금이 자기를 가리키는 일인칭 대명사인 '짐'을 사람 이름 'Jim'으로 번역했다.

다음으로 10 대와 20 대가 즐겨 쓰는 줄임말을 사용해 보았다. '아아(아이스 아메리카노) 1 잔, 따아(따뜻한 아메리카노) 1 잔'을 파파고는 'One iced americano, one hot americano'로 정확하게 인식했고, 구글 번역은 'Aah 1 cup, daa 1 cup'으로 '아아'와 '따아'를 소리나는 대로 번역했다. '친구 생파(생일파티)로 생선(생일선물)을 준비했어'를 파파고는 'i got you fish for your friend's birthday'로 번역해 생파에 대한 업데이트는 되어 있지 않은 것을 보여줬다.

반면 구글 번역은 'i prepared fish with my friend's fresh green onion'로 번역해 입력 단어인 생파의 표층 의미를 최대한 반영하고자 애쓰는 듯하다. '안 물어 본 것'과 '안 궁금한 것'의 합성어인 '안물안궁'을 파파고는 'i didn't ask'로 그리고 구글 번역은 '안물안 궁'으로 인식해 'Anmulan Palace'로 번역했다.

필자가 나열한 기계번역의 오역은 '귀여운 애교'라 할 수 있으나 경우에 따라 오역은 중차대한 결과를 초래할 수 있다. 인류 역사상 살상을 목적으로 핵무기가 사용된 유일한 사례인 일본 원자폭탄 투하는 오역이 부른 참사로 알려져 있다. 1945년 7월 26일 독일 포츠담에서 미·영·중 수뇌부는 일본의 항복 조건을 규정한 포츠담 선언문을 발표했다. 포츠담 선언은 "무조건 항복 외 다른 대안은 즉각적이고 완전한 파멸"이라는 일본을 향한 최후통첩이었다.

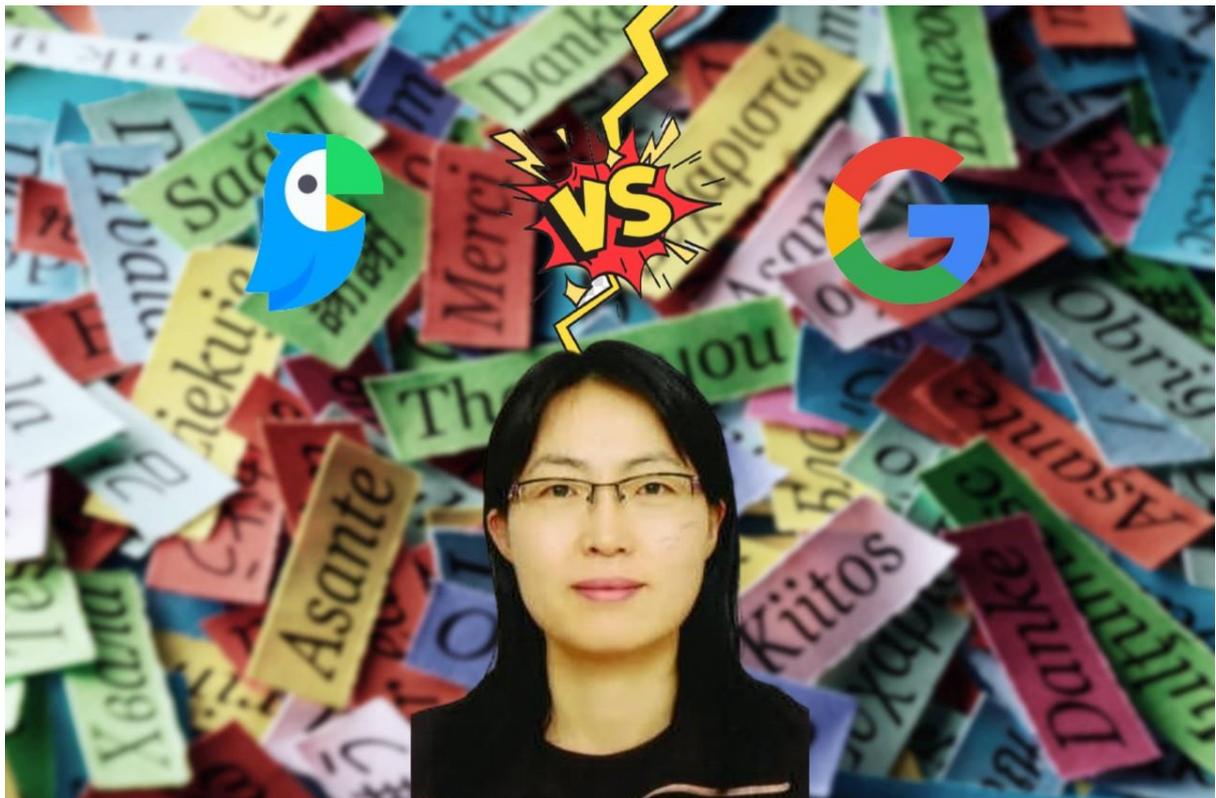
7월 28일 일본 스즈키 칸타로 총리는 포츠담 선언에 대해 '논평을 유보한다'는 의도로 여러 가지로 해석될 소지가 있는 '모쿠사츠(黙殺)'라는 표현을 사용했는데, 이를 일본 언론에서 '무시하다'로 번역했다. 이러한 일본의 오만함(?)은 미국 국민과 트루만 대통령의 분노를 샀고, 8월 6일과 9일 각각 히로시마와 나가사키에 원폭이 투하됐다.

위 예시는 상황과 맥락에 대한 오판이 초래한 가장 극단적인 결과라 할 수 있다. 분명한 것은 번역은 좁게는 화자의 몸짓, 눈빛, 어조, 뉘앙스에서 넓게는 국제정세와 외교의 흐름에의 이해를 요한다. 또한 번역은 사회·문화적 문맥, 지정학적 문맥 등의 지배를 받는다. 경우에 따라서는 화자의 이야기를 목표 언어의 사회·문화·역사·정치적 문맥에 맞추어 재해석해야 한다. 이렇듯 총체적 상황에 대한 이해가 가능해야 소기의 번역이 가능한 것이다.

필자가 제시한 ‘어전’ 예문에서 보았듯이 기계번역은 상황 전체를 파악하기에는 아직은 역부족인 듯하다. 더 나아가 언어가 화자의 감정을 반영한다는 점은 인간 번역가의 특화 영역을 확장시킨다. 기계번역에서 오류가 잦은 부문도 인간 번역가를 대상으로 하는 틈새시장을 만들 것이다. 결국 기계번역이 인간 번역가를 완전히 대체하는 날은 아직은 요원한 듯하다.

비하인드 인터뷰

칼럼을 읽은 후 칼럼니스트에게 질문 혹은 반문하는 것은 다소 귀찮거나 힘든 일이다. 독자를 대신해 시 타임스가 여전히 남은 궁금증을 풀어봤다. 조금은 매울지도.



남영자 중앙대 인문콘텐츠연구소 HK연구교수(사진=남영자 교수·네이버·구글·셔터스톡, 편집=박성은 기자)

Q. AI 번역은 이미 우리 일상의 일부가 됐다고 해도 과언이 아니다. 인간 번역가들이 여기에 위기감을 느끼진 않는지?

영국과 같은 유럽 지역에서 관련 설문조사를 진행한 바 있다. AI 번역기가 처음 등장했을 때는 “우리는 이제 끝났다”라는 마음이었다면 이제는 걱정이 줄어든 것 같다.

Q. AI 번역기가 인간 전문가를 완전히 대체 가능할까? 가능하다면 시기는 언제라고 보나?

번역 일은 많은 시간이 흘러도 AI 가 완전히 대체하기는 힘들 거라 본다. 오히려 사람이 돋보일 수 있는 영역이 아닐까 생각한다. 번역은 단순히 해석하는 작업이 아니다. 인간 번역가들도 비슷한 의미를 각기 다르게 표현한다. 여기에는 개인 삶의 경험이 관여하기에 아무리 많은 데이터를 넣어도 AI 가 하기 힘들 것 같다.

Q. 번역가 이외 사용자들은 의견은 다를 수도 있을 것 같은데?

IT 기업 고위직 종사자를 대상으로 의견을 물었더니 60~70% 정도가 계약 성사 건은 반드시 인간에게만 맡긴다고 답했다. 기술 분야 종사자들이 이러한 답을 했다는 것이 인상적이다.

Q. 번역일을 완전히 AI 에게 맡기기 어려운 이유는 무엇일까? 번역 오류가 나는 이유는?

언어에서는 수학과 달리 1 대 1 매칭이 안 되는 경우가 많다. ‘사과’를 반드시 ‘Apple’로 해석할 수 없는 것이다. ‘감칠맛’과 같은 단어는 특히 영어에는 아예 없는 것이다. 이외 언어 자체 특성, 각 국가·지역 간 문화 차이, 말을 하는 화자와 분위기, 감정 등 수많은 요소가 개입한다.

Q. 칼럼 본문에 나오는 ‘사용역’이 표현에 개입하는 요소들인 것 같다. 범위는 어디까지인지?

표현 하나가 등장한 T.O.P(Time.Place.Occasion)가 모두 사용역에 해당한다. 사투리부터 발음, 억양, 화자의 표정까지 모두 해당한다.

Q. 성별 편향성 문제는 AI 번역의 대표적인 한계점이다. 발생 원인이 궁금하다.

AI 번역에 사용한 데이터 자체에 편향이 있기 때문이다. 영어가 AI 번역 데이터 대부분을 차지하는 것도 성별 바이어스에 영향을 미친다. 영어에서는 우리말과 달리 꼭 ‘He’와 ‘She’를 구분한다. 언어 자체에 성별 구분 성격이 있다는 의미다.

Q. AI 번역기가 ‘아아(아이스 아메리카노)’, ‘뜨아(뜨거운 아메리카노)’와 같은 유행어를 번역해내 이슈화되고 있다. 어떻게 가능한 일인지?

사용자 수정 기능을 통해 주 사용자인 젊은 층이 업데이트를 한 것으로 보인다. 유행어 번역은 사실 다른 것보다 오히려 간단하다. ‘아아’가 반드시 아이스 아메리카노를 의미하는 것처럼 고유명사와 같은 성격을 지니기 때문인 것 같다.

Q. 기업 간 계약이나 국가 외교에는 어렵겠지만 AI 번역기가 힘을 발휘할 수 있는 분야도 있을 것 같은데?

논문과 같이 비교적 정형화된 글과 문장 형식을 사용하는 분야에서는 비교적 수월할 거라 생각한다. 소설, 시와 같은 예술 영역에서는 정서라는 것이 중요하기 때문에 AI가 잘 할 수 있는 영역은 아닌 것 같다.

Q. AI 번역은 데이터가 많이 필요한 분야로 보인다. 대기업 밖 연구자들이 할 수 있는 일이 있을까?

AI 번역기 개발 이외에도 할 일들이 많다. 구체적으로 어떤 연구를 하느냐에 따라 필요한 인력과 비용에 차이가 있다. 예를 들어 나는 지금 감정언어 데이터셋 구축 연구를 진행 중인데 파파고 개발에 들어가는 만큼의 데이터와 인력이 필요하진 않다.

Q. 구글 번역기, 파파고 이후 차세대 AI 번역 기술은 어떤 모습일까?

AI 실시간 통역이라 할 수 있겠다. AI 통역도 적용 분야마다 난이도나 실현 가능성에 차이가 있다. 호텔이나 음식점에 적용하는 것은 수월하겠지만 사업이나 외교에서는 어려울 것이다. 질문 자체를 해석하는 것 이외 행간이나 분위기를 읽는 것이 중요한데 이 플러스 알파가 가능해지려면 시간이 상당히 걸릴 것 같다.

남영자 중앙대 인문콘텐츠연구소 HK 연구교수는 현재 감정언어 데이터셋 구축에 주력 중이다. 남 교수의 주요 연구 분야는 음성 감정 인식(Speech emotion recognition), AI 음성 페르소나(AI-voice persona), 음성 인식(Speech perception), 음성학(Phonetics), 음향심리학(Psychoacoustics) 등이다.