번역학연구
2018 봄 제19권1호

기계번역 담론에 대한 비판적 고찰*

송 연 석
(한국외국어대)

1. 들어가며

2016년 11월 구글이 공개한 신경망 방식 기계번역을 통해 이전에 비해 괄목할 만한 품질개선이 확인되면서(Wu et al., 2016) 기계번역은 번역학계의 가장 뜨거운 화두로 떠올랐다. 번역사, 나아가 인간번역의 정체성과 미래가 걸린 역사적 도전을 맞아 기계번역을 주제로 한 다양한 학회와 세미나가 열리고 있고 연구도 급증하고 있다. 기계번역이 인간번역을 어느 정도까지 대체할 수 있을 것인지에 대해서는 여전히 의견이 분분하다. 기계번역이 인간번역을 완전히 대체할 수 있는 수준까지 발전할 수 있을지, 대체 가능하다면 그 시기는 언제가 될지, 대체한다는 말은 어떤 수준의 작업까지를 의미하는지 등에 대해서는 개발자들 스스로도 확답을 하지 못하고 있는 실정이며[1] 설령 답을 준다 해도 변

* 이 논문은 2018년 한국외국어대학교 학술연구지원을 받아 작성되었음.
1) 구글 번역 총괄 연구원 “AI가 인간 번역가 대체하기 힘들어” (2017.12.30. 조선일보)

수가 많아 그대로 실현된다는 보장은 없다. 그럼에도 불구하고 대다수 언론은 머지않아 번역사가 다 필요 없게 된다는 식의 자극적 보도를 양산하며 기계번역 담론을 주도하고 있다[2]. 그런데 언론이 상업적 이데올로기의 지배를 받을 수밖에 없고(Shoemaker & Reese, 1996) 특히 국내 언론 매체들이 '과장 선정적 보도,' '편파 불공정 보도', '피상적 보도/받아쓰기' 관행으로 비판 받고 있음은 주지의 사실이다(김상호, 2007; 이정훈, 2013; 홍원식・김은정, 2013). 언론의 속성상 당연한 현상이라 해도 이들이 주도하는 현 기계번역 담론은 아직 기술적, 상업적 실현가능성과 별개로 기대감이 다소 과도하게 부각된 측면이 있다. 이 같은 배경에서 본고는 기계번역 원리에 비추어 현 언론 주도 기계번역 담론을 비판적으로 고찰하고 그 번역학적 함의를 모색하는 것을 목적으로 한다.

기계번역 시대를 앞두고 번역학계는 이미 학술적으로나 교육적으로 다양한 방향으로 대응에 나서고 있다. 포스트에디팅 중심으로 번역 교육의 새로운 방향을 모색하는 연구들(신지선, 2017; 이상빈, 2017)과 특정 텍스트나 문구로 기계번역을 실시해 결과를 분석한 연구(박옥수 2016; 장애리 2017; 최동익 2016; 최효은 2017; 한승희 2017; 함수진, 류수린 2010; 황은하 2014) 등이 대표적이다. 여기서 도출된 결과들이 축적되면 기계번역에 대한 더 효과적인 대응 혹은 적응이 가능해진다는 점에서 매우 의미 있는 연구들이다. 다만 아직은 단편적인 사례 중심으로 연구가 이뤄질 수밖에 없기 때문에 상당한 시간에 걸쳐 다량의 데이터 분석 결과가 축적되기 전에는 유의미한 패턴이나 예측 가능한 규칙을 찾아내 그에 따른 대응 혹은 적응 전략을 수립하기는 어려울 것이다. 어쩌면 언어처럼 방대한 영역에서는 이 같은 귀납적 접근을 통한 일반화 자체가 불가능한 것인지도 모른다. 이에 본고에서는 역으로 기계번역의 원리에서 출발하는 연역적 접근을 통해 기계번역 모델, 번역품질 평가, 상용화 관점에서 각각 기계번역 담론을 조명하고 번역 연구자와 교육자로서 어떻게 대응해나갈 수 있을지 논의하고자 한다.

---

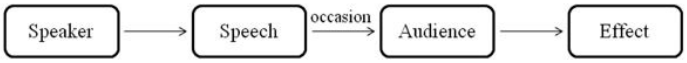
2) "진화하는 번역기..사라지는 번역가?"(2017.1.18. 조선일보), "지구촌 10년 내 언어장벽 사라진다"(2017.2.22. 디지털타임스), "내가 이러려고 영어 배웠다, AI가 번역 다 해주네"(2016.11.18. 중앙일보), "목에 걸면 외국어가 술술...통역사 필요없는 웨어러블"(2017.2.28. 디지털타임스)

2. 기계번역 원리를 통해서 본 기계번역 담론

2.1 기계번역과 커뮤니케이션 모델

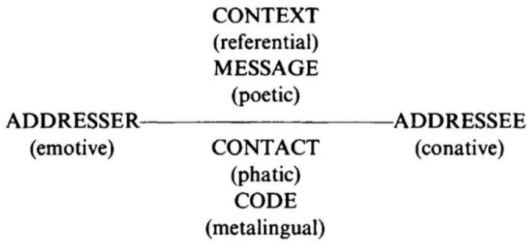
번역은 기본적으로 “언어 간 경계를 넘어 정보를 전달하는 커뮤니케이션 행위”(Bassnet, 2012: 95)이며 커뮤니케이션은 모델로 표현될 수 있다. 최초의 커뮤니케이션 모델은 기원 전 4세기 아리스토텔레스가 창안한 것으로 알려져 있으며 화자, 메시지, 청자의 3요소로 구성된다(그림 1).

그림 1 아리스토텔레스의 커뮤니케이션 모델(D'Andrea et al., 2017)



현대에 들어 언어학자 야콥슨(Jakobson, 1960)이 제시한 커뮤니케이션 모델은 메시지와 메시지를 인코딩하는 발신자, 디코딩하는 수신자라는 기본 구성요소는 동일하지만 맥락을 아우르는 6가지 기능, 즉 지시적(referential), 정서적(emotive), 미학적(poetic), 교감적(phatic), 설득적(conative), 메타언어적(metalingual) 기능 등 화용적 기능을 포함한다(그림 2). 번역학자 나이다(1964) 또한 번역에 대한 독자반응과 메시지의 효과를 중시하는 효과의 등가 이론을 주창한 바 있다.

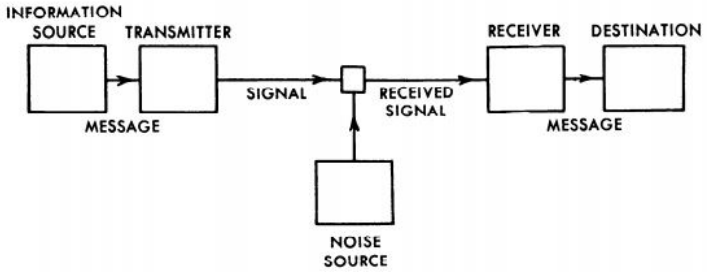
그림 2 야콥슨의 커뮤니케이션 모델(Jakobson, 1960)



그러나 2차 세계 대전 이후 지금까지 기계번역을 비롯해 여러 분야에 영향을 끼치고 있는 정보이론의 근간이 되는 모델은 새넌이 위버와 함께 만든 커뮤니케이션 모델(Shannon & Weaver, 1949)로, 메시지, 발신자, 수신자로 구성되

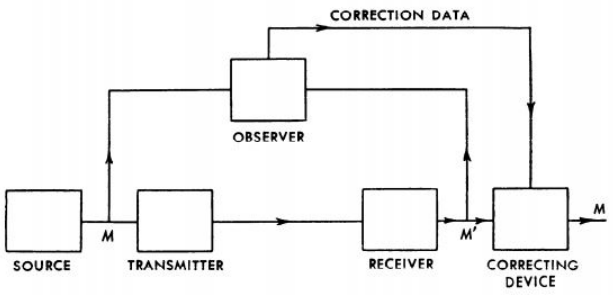
고 메시지가 인코딩과 디코딩을 거쳐 전달된다는 기본 원리는 같지만 메시지 전달을 두 지점 간 전기적 신호 전달로 간주하고 이를 함수로 표현했다는 차이가 있다(Blackburn, 2007). 특히 메시지를 "가능한 여러 메시지들 중 선택된 하나"(Shannon, 1948: 379)로 간주하고 이렇게 특정 의미의 메시지로 선택될 '확률'을 계산할 수 있다고 봤다(그림 3).

그림 3 섀넌-위버의 커뮤니케이션 모델(Shannon & Weaver, 1949: 34)



이들이 수정한 에러교정모델(1949)은 오류를 오차역전파(back propagation)[3]를 통해 수정하는 기계번역의 원리와 유사한 점이 있다(그림 4).

그림 4 커뮤니케이션 에러 교정 모델(Shannon & Weaver, 1949: 68)



수학자이자 전기공학자인 섀넌과 위버의 커뮤니케이션 모델과 정보이론은 이처럼 정보가 전달되는 과정과 수학적 계산에만 초점을 맞췄을 뿐 '효과'에 해

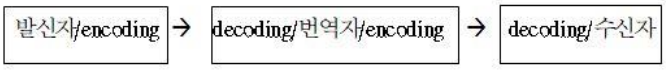
3) 오차 발생 시 이전 단계로 거슬러 올라가 가중치를 조절해나가는 것.

당되는 인지적, 정서적, 사회문적 요인은 고려하지 않았다는 한계가 있다(D'Andrea et al., 2017). 철학자인 아리스토텔레스나 언어학자인 야콥슨, 번역학자 나이다 등이 커뮤니케이션의 '효과'를 중시한 것과 대조적이다. 나이다의 효과의 등가론도 대상마다 달라질 수 있는 '효과'를 계량화하거나 객관적으로 측정하기 어렵다는 이유로 비판을 받은 사실(van den Broek, 1978:40)에 비추어보면 메시지의 '의미'나 '효과'에 대한 관점과 접근법에서부터 이미 공학과 인문학 간 근본적인 차이가 노정되고 있음을 알 수 있다. 기계번역의 한계에 대한 번역학적 관점의 논거 중 일부는 여기서 출발한다.

2.2 인코더-디코더 모델

대부분의 신경망 기계번역(neural machine translation, 이하 NMT)은 메시지를 기호화하는 인코더-디코더 모델을 기본 구조로 삼고 있다(Sutskever et al, 2014). 번역학에서는 전통적으로 번역이 단순한 코드변환(transcoding)이 아니라는 인식(Vermeer, 1984)이 있어 번역과정을 메시지의 인코딩과 디코딩으로 표현하는 경우가 드물지만 논의의 편의를 위해 번역과정을 커뮤니케이션 모델로 도식화하면 그림 5처럼 나타낼 수 있을 것이다.

그림 5 번역의 커뮤니케이션 모델



NMT의 실제 작동원리를 표현한 인코더-디코더 모델(그림 6)은 그림 5의 "decoding/번역자/encoding"에 해당되는 부분을 (encoding)-은닉상태(hidden state)-(decoding)으로 설명하기 때문에 용어에서 오해의 소지가 있으나, 인간번역에서 원천언어(Source Language, SL)로 표현된 메시지를 맥락을 고려해 디코딩하고 다시 목표독자층의 맥락에 맞추어 목표언어(Target Language, TL)로 인코딩한다면, 기계번역에서는 SL 단어로 이뤄진 하나의 시퀀스 전체를 길이가 고정된 숫자 형식의 벡터[4]로 바꾸어 인코딩한 뒤 확률을 계산해 TL 시퀀스로 디코딩한다는 유사성을 갖는다.

그림 6 인코더-디코더 모델(Neubig, 2017)

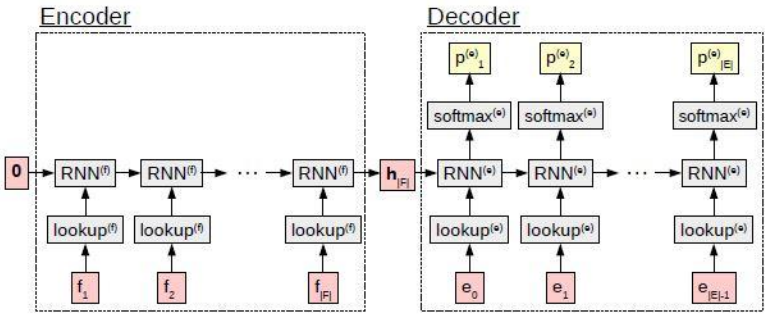


그림 6에서 f는 foreign language의 약자로 SL을, e는 English, 즉 TL을, h는 은닉상태(hidden state)를 나타내며, softmax라는 함수를 이용해 각 요소가 번역될 확률인 결과값 p를 산출해낸다. RNN은 recurrent neural network(순환신경망)의 약자로, 일반신경망이 입력노드-출력노드를 거쳐 결과값이 나오는 네트워크 구조라면 RNN은 같은 네트워크의 입력-출력 노드가 여러 개 겹쳐 있어 순환되는 구조이며 자연어처리 성능이 좋아 NMT시스템에 이용된다. 일반적인 NMT 시스템은 인코딩을 위한 RNN과 디코딩을 위한 RNN들로 이뤄지지만(Wu et al., 2016) RNN의 단점을 보완한 CNN[5]을 사용하기도 한다. 이때 다음 단어 예측을 위해서는 엄청나게 방대한 데이터를 이용한 학습이 필요한데 현실

4) 가령 인풋 시퀀스가 n개 단어로 이뤄진 X라면 X= x1, x2, x3... xn으로 표현될 수 있는데 이 각각이 벡터에 해당된다. 즉 인풋 시퀀스의 기호 수 = 벡터 수로, 고정된 길이의 벡터들이 만들어지는데 각 벡터는 훈련용 데이터에 있는 다른 단어들과의 관련성을 나타낸다. 시퀀스의 기호들은 대부분 단어지만 GNMT처럼 단어보다 한 층위 아래인 단어조각(wordpiece)을 이용하기도 한다. GNMT는 이를 통해 데이터에 없어 어려운 희귀어(rare words) 처리 문제를 개선했다고 보고한 바 있다. 디코더에서는 은닉상태를 만들어 다음에 올 수 있는 후보 기호들을 예측하고 softmax층을 거쳐 이 기호들의 확률분포를 생성하게 된다(Wu et al., 2016).

5) Convolutional Neural Network. RNN에 비해 짧은 단어 시퀀스의 특징을 쉽게 잡아 이를 문장 전체에 걸쳐 누적 반영할 수 있고 '기울기값이 사라지는 문제(vanishing gradient: NMT에서 사용하는 미분 함수에서 기울기가 너무 낮아져 인풋에 변화를 줘도 결과값이 별로 달라지지 않는 문제)'가 상대적으로 적은 대신 일정 수준을 넘어가는 복잡한 패턴은 자연스럽게 표현하지 못한다는 단점이 있다(Neubig, 2017: 46).

에서는 학습 데이터 규모에 한계가 있으므로 각 문장 단위로 표현되는 벡터를 학습시키기보다는 SL 문장의 단어별로 벡터를 만든 뒤 그 중 특정 단어에 더 많은 '주의(attention)'를 기울여 TL단어를 예측하는 '어텐션 벡터(attention vectors)'를 적용하는 식 등으로 보완되기도 한다(Bahdanau et al., 2015; Luong et al., 2015; Neubig, 2017). 그밖에도 초벌번역을 한 뒤 SL문장과 초벌TL문장에 각각 어텐션 메카니즘을 적용해본다든지(Li et al., 2017), TL단어와 SL단어의 관계보다 TL단어 간의 관계에 어텐션을 적용하는 등(Zhou et al., 2017) 번역품질 개선을 위해 NMT 모델을 수정 보완하려는 다양한 연구들이 이뤄지고 있다. 다양하다는 것은 여전히 문제가 해결이 안 됐음을 의미한다.

특히 기계가 벡터가 무엇을 의미하는 것으로 해석하고 있는지, 은닉상태에서 정확히 어떤 일이 일어나는지는 여전히 미지의 영역으로 남아있다. 기계는 오차 발생 시 오차역전파를 통해 결과값을 개선해나갈 뿐이다(Weng et al, 2017). 물론 모델을 수정, 보완하려는 다양한 연구가 이뤄지고 있으므로 기계번역의 품질은 계속 개선될 것이다. 하지만 주어진 데이터에서 스스로 패턴을 찾고 학습하는 NMT의 非지도학습은 오류는 찾아낼 수 있어도 그 원인을 파악하기가 쉽지 않다는 문제가 있고 따라서 번역처럼 변수가 다양한 작업에서는 오류를 예측하기도, 향후 발전 정도나 속도를 예상하기도 어려워진다. 언론의 기계번역 담론은 이 부분은 간과한 채 기계번역 기술발전이 가져올 변화에 대한 결과론에 초점을 맞추는 경향이 있으나 번역교육의 관점에서 볼 때 이 문제는 중요한 시사점을 갖는다. 비교적 유연한 등가 개념을 제시한 투리(Toury, 1995)에서도 등가성의 조건 중 하나는 ST와 TT로 추정되는 두 텍스트 간 '설명가능한 관계'가 있어야 한다는 것이다. '설명'은 이유와 결과 등에 대한 비판적 사고와 분석이 요구되는 행위이며 이것이 인간이 우세한 영역임은 공학자들도 동의하는 명제다[6]. 지금의 기계번역에서는 단 하나의 결과물만 산출되므로 그것이 번역목적에 비추어 최선인지는 인간이 판정해야 하며 그러려면 번역사로서든 프리(pre-)/포스트(post-) 에디터로서든 언어전문가로서든 번역을 둘러싼 다양한 언어적, 비언어적 요인을 고려해 비판적으로 설명할 수 있는 능력이 있어야

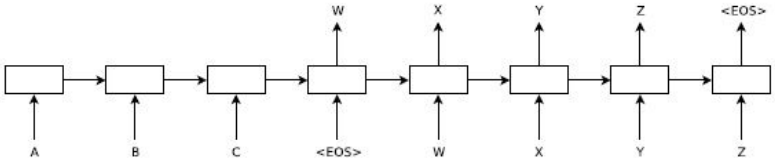
6) "인간 지성이 인공지능보다 뛰어난 까닭은" (2016.9.11.한겨레), "Machines & humans: a promising learning community" (2017.9.11. BSI, Business Systems Integration)

한다. 따라서 번역교육에서도 번역결과물에 대한 설명 능력 함양을 중요한 수업목표로 설정하고 교육내용에도 적극 반영할 필요가 있다.

2.3 시퀀스-투-시퀀스(sequence-to-sequence) 모델

신경망 방식이 기존의 통계기반 방식과 다른 점은 번역 결과로서 가능한 확률을 각 단어별로 계산해나가는 것이 아니라 여러 개 단어로 이뤄진 하나의 의미 단위에 해당되는 시퀀스(sequence)끼리 매핑작업을 한다는 것으로 (Sutskever et al, 2014), 문장의 끝을 의미하는 'eos(end of sentence)'를 표지로 문장(시퀀스) 단위 예측을 하게 된다(그림 7).

그림 7 시퀀스-투-시퀀스 학습 모델(Sutskever et al., 2014)



기계번역 담론에 흔히 등장하는 내용 중 하나는 이렇게 NMT 덕에 문장 단위로 기계번역을 하게 되면서 '맥락'을 고려한 번역이 가능해졌다는 주장이다. 그런데 전술한 바와 같이 엄밀히 기계번역은 그 원리가 되는 모델의 구성요소에 '맥락'이 빠져 있다. 기계번역 관련 각종 문헌에서 지칭하는 'context'는 일반적으로 번역학에서 말하는 상황적 혹은 사회문화적 맥락을 아우르는 거시적 의미의 '맥락'이 아니라 해당 단어나 구문의 전후로 등장하는 단어들을 가리키는 구체적이고 미시적인 'co-text'에 가깝다. 예컨대 context 예측모델을 연구한 바로니 외(Baroni et al, 2014: 238)는 context를 '공기어(co-occurring words)'를 지칭하는 말로 명기하고 있으며, 왕과 조(Wang & Cho, 2016)는 RNN에 통계적 언어모델링을 적용한 'larger context' 모델을 제시했지만 여기서도 맥락은 이전 문장들로 구성된 것으로 간주된다. 즉 기계번역에서 단어의 '의미'란 오랜 세월에 걸쳐 형성된 단어의 이미지나 어감, 상황적·사회문화적 맥락에 따라 다양한 뉘앙스를 내포하고 다르게 해석될 소지가 있는 다차원적 의미가 아니라

주어진 문서에서 산출해낸 단 하나의 공기(co-occurrence) 확률값인 셈이다. 설령 기계번역의 '맥락'이 context가 아닌 co-text 차원의 문제라고 해도 기계번역에서는 이전에 나온 단어 등의 입력 정보를 오래 기억했다가 나중에 나오는 관련성 있는 단어의 번역에 반영해야 하는 '장기의존성(long-term dependencies)' 문제가 있다. 물론 NMT는 이를 해결하기 위해 LSTM[7](long short-term memory)네트워크를 사용하고 있고 구글의 GNMT도 8개 인코더층과 8개 디코더층으로 이뤄진 deep LSTM 네트워크를 사용해 번역품질을 개선했다고 주장한다(Wu et al., 2016). 그러나 언어라는 방대하고 복잡다단한 시스템을 LSTM으로 해결하는 데는 여전히 한계가 있으며 특히 SL 문장이 길어질수록 번역품질이 저하되는 문제가 있다(Zhou et al., 2017).

기계번역 연구들에서 맥락을 context가 아닌 co-text 층위의 문제로 인식하는 것은 NMT가 기본적으로는 주어진 단어에 대해 그 다음 특정 단어가 나타날 조건부확률을 계산하던 통계기반 기계번역의 수학적 접근을 물려받았다는 데서 비롯된 특징으로, '맥락'이 한 가지 개념으로 정의되기 힘들 뿐 아니라 번역 과정에서 재맥락화 현상이 일어나게 마련이라는 번역학적 시각(강지혜, 2007/2008; 이주리애, 2011; Kang, 2007/2010; Munday, 2007)과 근본적인 차이를 보여준다. 문제는 특정 사회 상황과 독자층에 따라 수용 양상이 다를 수 있는 특정 표현의 번역을 해당 사회문화적 환경과 독자층 특성까지 반영해 단 한 개의 확률값으로 계산할 수 있느냐이다. 예컨대 뉴스번역처럼 해당 기관의 이데올로기적 이유로 정확하지 않은 특정 표현이나 번역방식이 선호된다거나(송연석, 2011/2013) 정치적으로 올바른(politically correct) 표현이 기대되는 미디어에서 특정 목적이나 의도를 갖고 일부러 위반해 번역하는 경우(윤희주, 2007) 혹은 반대로 공손성을 위반해 관련 인물의 성격을 변화시키는 경우(최현희・배만호, 2015)처럼 단순히 확률로 환산하기 힘든 사례들이 무수히 많기 때문이다. 애초에 숫자로 계산되지 못하는 요소는 번역에도 반영되지 못하게 되어 있는 기계번역의 원리를 고려할 때 이 같은 요소들을 어떻게 계량화할지는 기계번역 기술 개발자들이 풀어야 할 숙제다. '맥락'의 의미를 공기 확률이 높은 어휘로 이해하고 접근하는 공학적 방식과, 동일한 표현도 의식적, 무의식적으로 작용하

7) 일반 RNN보다 장기의존성을 잘 잡아낼 수 있는 것으로 평가된다(Neubig, 2017: 31).

는 개인의 경험과 지식, 사회문화적 배경에 따라 다르게 수용될 수 있다고 보는 번역학적 사고방식 간 간극은 좁히기 어려울 것이다. 그렇다면 번역교육에서는 우선 '맥락'에 대한 체계적인 이해, 즉 동일한 원문의 번역이 고객의 특성이나 요구, 맥락에 따라 어떻게 달라질 수 있는지 이해하고 이를 바탕으로 여러 번역 대안 중 적합한 하나를 골라낼 수 있는 능력(Pym, 2013)을 우선 배양한 뒤 이를 포스트에디팅에 적용하는 순서로 커리큘럼을 구성해야 할 것이다.

3. 품질평가 관점에서 본 기계번역 담론

기계번역의 품질개선 여부를 검증 및 주장하기 위해 가장 많이 사용하는 척도는 BLEU 점수다. METEOR, NIST 등도 있지만 현재 기계번역 관련 연구 발표가 가장 활발히 이뤄지고 있는 학회인 ACL(Association for Computational Linguistics)에서도 각 연구 성과를 BLEU 점수를 근거로 제시하는 등 BLEU가 가장 널리 사용되고 있는 척도다. BLEU는 인간의 번역평가에 시간과 비용이 많이 든다는 단점(Hovy, 1999)을 극복하기 위해 파피네니 외(Papineni et al., 2002)가 제안한 자동 기계번역 평가방법으로, "기계번역물이 인간 전문가의 번역물에 가까울수록 품질이 좋다"가 기본 전제다. 이에 따라 BLEU 평가시스템은 인간번역과의 "유사성(closeness)"을 평가할 척도와 "양질의 인간 준거번역물 코퍼스" 두 가지 필수 요소로 구성된다(ibid., 311). 수많은 기계번역 결과물의 품질을 신속히 판정해야 그에 따라 기계번역 시스템을 수정할 수 있는데 인간의 평가에 시간과 비용이 많이 소요된다는 현실적 한계 때문에 BLEU를 사용하는 것이다. BLEU 점수는 0~1 사이로, 1에 가까울수록 준거번역인 인간번역과 유사성이 높은, 즉 번역품질이 좋은 것으로 해석되는데 수치가 작다 보니 백분율로 표시되기도 한다. 파피네니 외는 알 수 없는 중국어 원문에 대한 영어 번역을 예시로 들며 BLUE의 원리를 다음과 같이 설명한다(ibid., 312).

> 인간번역1: It is a guide to action that ensures that the military will forever heed Party commands.
> 인간번역2: It is the guiding principle which guarantees the military forces

always being under the command of the Party.
인간번역3: It is the practical guide for the army always to heed the directions of the party.

기계번역1: It is a guide to action which ensures that the military always obeys the commands of the party.
기계번역2: It is to insure the troops forever hearing the activity guidebook that party direct.

기계번역1은 기계번역2보다 준거가 되는 인간번역1,2,3과 일치하는 표현을 훨씬 많이 사용했다는 점에서 더 좋은 점수를 받게 된다. 이때 일치하는 표현은 한 단어(“which”나 “always”) 혹은 여러 개의 연쇄 단어(“It is a guide to action”이나 “ensures that the military”)가 될 수도 있다. 즉 일치하는 단어 수(n)를 계산하는 엔그램(n-gram)방식인데, 기계번역 결과물의 엔그램과 인간의 준거번역 엔그램을 비교한 뒤 일치하는 수가 많을수록 품질이 좋은 것으로 평가하게 된다. 파피네니 외(ibid., 312)에 따르면 다음 예에서 n=1로 계산할 경우 정밀도(precision)는 [준거번역 중에서 일치하는 기계번역 단어의 최대 개수/기계번역물 총 단어 수]= 2/7이나 되지만 n=2로 하면 0/7=0이 된다. 따라서 n=1일 경우 원문을 얼마나 충분히 옮겼는지 번역의 충분성(adequacy)을 가리는 척도가 되고 n이 높아질수록 자연스러움(fluency)을 판정할 수 있게 된다.

기계번역: the the the the the the the.
준거번역1: The cat is on the mat.
준거번역2: There is a cat on the mat.

기계번역의 엔그램 중에서 준거번역에도 사용된 엔그램 비율을 각 엔그램별로 계산한 뒤 기하평균[8]을 낸 값으로 구하는 이 정밀도(precision) 방식은 기계번역 결과물에 들어있는 단어 중에서 준거번역에도 사용된 단어가 몇 개인지,

8) 가령 2와 8의 산술평균은 2+8/2=5이지만 기하평균은 2x8=16의 제곱근인 4가 된다. 가령 연도별 수익율 같은 비율 간의 평균을 구할 때는 산술평균이 아닌 기하평균을 사용한다.

즉 품질이 어떻든 기계가 내놓은 답 중 몇 개가 맞고 틀리는지를 따지기 때문에 위 사례에서 알 수 있듯이 번역의 품질을 가리는 데 한계가 있다. 이를 보완하기 위한 것이 재현율(recall) 방식으로(Peled & Reichart 2017), 준거번역 결과물에 들어있는 단어 중 기계번역에도 사용된 단어가 몇 개인지, 즉 가능한 답 중에서 기계가 몇 개를 맞혔는지 계산한다. 그런데 BLEU는 재현율은 반영하지 않고 정밀도 방식만 따르기 때문에 품질평가에 한계가 있다는 것이 단점으로 꼽힌다(Banerjee & Lavie, 2005). BLEU는 번역에서 단어선택이나 배열순서에 따라 무수히 많은 경우의 수가 나올 수 있다는 문제를 극복하기 위해 여러 개의 준거번역을 사용해 단어선택의 다양성을 반영하고 여러 문장에서 반복되는 동일한 구문은 한번으로 계산하는 식으로 엔그램 정확도를 수정해 사용하고 있으며 높은 수의 엔그램에 보상을 주는 방식으로 사용되고 있다(Callison-Burch, et al., 2006). 그러나 원리 상 준거번역을 많이 사용할수록 점수는 높아지기 쉽다. 또한 엔그램 수가 많아진다고 해서 문법적으로 옳은 문장을 만들어낸다는 보장도 없다. 즉 BLEU 점수가 높다고 해서 반드시 실제 번역 품질이 나아졌음을 의미하지는 않는다. 실제 연구에서도 인간평가에서 1위를 한 번역물이 BLEU 점수로는 6위인 경우도 있었다(Callison-Burch et al., 2006/2007). 기계번역 품질이 최근 크게 향상됐음은 분명한 사실이지만 BLEU 점수를 기준으로 한 연구 데이터 기반의 '획기적인' 품질개선 주장 담론은 일반화가 어렵다는 점에서 실제와 차이가 있을 수 있다는 의미다. 칼리슨-버취 외(ibid.)에 따르면 두 단어쌍(bi-gram) 매치수를 b, 번역문의 길이를 k라고 가정할 때, 이론적으로 가능한 번역의 경우의 수는 (k-b)!이지만 엔그램 단어묶음의 위치만 바꾸는 순열(permutation)로도 동일한 BLEU 점수 획득이 가능하다. 이들은 다음 예문에서 순열로 비슷한 BLEU 점수를 받을 수 있는 경우의 수가 최소 40,320개나 된다며 BLEU 평가의 맹점을 주장한다.

> 인간1: Orejuela appeared calm as he was led to the American plane which will take him to Miami, Florida.
> 인간2: Orejuela appeared calm while being escorted to the plane that would take him to Miami, Florida.
> 인간3: Orejuela appeared calm as he was being led to the American plane that was to carry him to Miami in Florida.

인간4: Orejuela seemed quite calm as he was being led to the American plane that would take him to Miami in Florida.

기계1: Appeared calm/ when / he was / taken / to the American plane/, /which will / to Miami, Florida.
기계2: which will | he was | , | when | taken | Appeared calm | to the American plane | to Miami , Florida.

칼리슨-버취 외는 한 개의 번역에 동일한 BLEU 점수를 받을 수 있는 경우의 수는 수백만 개에 이를 수 있으며 그 경우 인간 평가자도 모두 동일한 점수를 줄 것이라 기대하기 어려운 만큼 실제 개선이 이뤄지지 않았어도 상대적으로 높은 점수를 받는 일이 가능하다고 지적한다. 구글 역시 GNMT 연구결과를 발표하면서 “BLEU 점수는 높일 수 있어도 번역품질에 대해 인간이 받는 인상은 거의 나아지지 않았음”(Wu et al., 2016: 18)을 인정하고 그 원인으로 “평가척도로서의 BLEU와 인간평가자가 인지하는 실제 번역품질 사이의 불일치(mismatch)” 가능성을 제기한 바 있다. 해당 연구에서 기계번역 모델 학습 단계에는 정밀도와 재현율을 모두 반영했으나 실제 기계번역 결과물 평가에서는 정밀도 방식만을 사용했는데 이유는 제시하지 않았다.

사실 평가는 인간번역에서도 어려운 영역이다. 지금도 합의된 혹은 보편적으로 사용되는 단일한 평가기준은 찾기 힘들며 많은 연구자와 교육자들이 평가자의 주관을 배제할 수 있는, 일관성과 신뢰성이 보장된 객관적 평가가 쉽지 않음을 경험해왔다(한국외대 통번역대학원 번역평가인증 연구팀, 2016). 누가 채점해도 비슷한 결과를 얻을 수 있는 TOEFL 같은 표준화된 평가와 달리 번역은 목적에 따라 등가 달성 여부가 달라질 수 있고(House, 1997; Vermeer, 1984/2004) 독자가 속한 사회문화적 배경에 따라서도 수용 양상이 다양하게 나타날 수 있다(이상빈, 2013; Chen, 2011). 즉 인간번역의 품질평가조차 일관성을 담보하기 어려운 현실에서, 수많은 경우의 수 중 일부에 불과한 준거번역과의 단어 일치율을 근거로 기계번역 결과물의 실질적인 품질 향상을 주장하는 데는 한계가 있을 수밖에 없다. 파피네니 외(ibid, 312)는 BLEU 원리를 설명하면서 “주어진 원천 문장 한 개에 대해 다수의 ‘완벽한(perfect)’ 번역본이 있게 마련”임을 인정하지만 정작 ‘완벽함’은 번역학에서도 확정하지 못한 실체 없는

개념이며 합의할 만한 기준도 없다. 구글은 학습데이터 없이도 가능하다는 제로샷(zero shot) 번역[9]에서 “괜찮은(reasonable)” 품질의 번역결과가 나왔다(Johnson et al. 2017: 345)고 주장하는데, 어느 정도 수준이 “괜찮은” 것인지에 대한 기준은 역시 제시하지 않았다. 평가기준은 차치하더라도 평가자를 번역 사용자가 해야 할지, 평가자 간(inter-rater) 신뢰도 및 평가자 내적(intra-rater) 신뢰도는 어떻게 보장할 것인지, 평가자는 몇 명이 적정한지 등등 기계번역 평가에 대한 견해는 인간번역 평가만큼이나 다양하다(Aranberri-Monasterio, 2009; Arnold et al., 1993; Carroll, 1966; White, 2003). 이런 여러 변수를 무시한 채 일치 단어수를 기반으로 한 인간번역과의 형식적 유사성을 근거로 기계번역 품질 개선을 주장하고 있는 것이다. 그렇다면 이에 대해 목적이나 기능별로 합의 및 통용될 수 있는 구체적인 번역평가기준을 수립하고 번역품질을 검증하는 일은 번역학계가 맡아야 할 몫이다. 번역학적 관점에서 이뤄지는 기계번역 품질 평가 결과와 그에 대한 분석 데이터가 충분히 축적되면 역으로 이를 통한 특정 장르나 텍스트유형별로 통일된 평가기준 도출도 가능할지 모른다[10]. 궁극적으로는 그러한 세분화된 평가기준과 더불어 인증제도를 마련함으로써 기계번역과 상생해나갈 수 있을 것이다. 그렇다면 번역교육은 평가 및 인증 인력을 양성하기 위해서라도 번역물 생산능력뿐 아니라 평가와 감수 능력 배양에 주안점을 두되 특히 인간 고유의 능력인 창의력과 비판적 사고력을 발휘할 수 있는 분야를 중점적으로 육성해나가야 할 것이다.

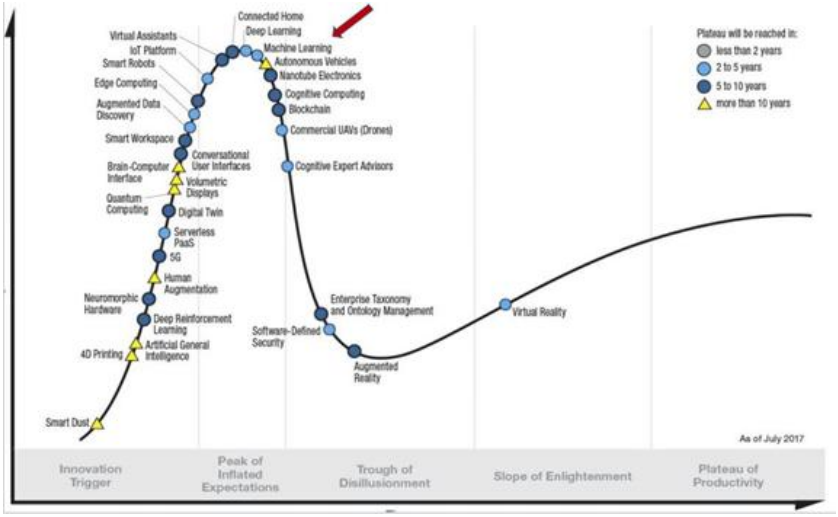
4. 상용화 관점에서 본 기계번역 담론

기계번역 연구는 이미 1962년에 현 ACL의 전신인 ‘기계번역 및 컴퓨터언어학 협회(AMTCL: Association for Machine Translation and Computational Linguistics)’가 설립됐을 정도로 짧지 않은 역사를 갖고 있다[11]. 90년대 이후 인공지능 및 관련 기술의 획기적인 발전으로 기계번역이 새로운 국면에 접어들

9) 가령 영-한, 영-일 데이터 학습만으로도 한-일 번역이 가능하다(Johnson et al, 2017).
10) 기계번역을 이용한 평가기준 수렴 가능성을 지적해주신 심사위원께 감사드린다.
11) 출처: ACL 홈페이지

면서 2005년부터는 매년 기계번역 경진대회(WMT: Workshop on Machine Translation)가 열리고 있고 유럽어 중심인 WMT에 맞서 2014년부터는 아시아 언어를 중심으로 하는 WAT(Workshop on Asian Translation)도 매년 일본에서 개최되는 등 기계번역에 대한 연구가 활발히 이뤄지고 있고 다양한 성과를 내고 있다. 그런데 연구 성과란 실험 환경에서 연구자들이 도출해낸 결과로, 실제 시장 적용이나 상용화는 또 다른 차원의 문제다. 이 점에서 IT전문 시장조사기관인 가트너(Gartner)가 신기술에 대한 거품 낀 기대와 실제 상업적 활용가능성을 구분하기 위해 매년 발표하는 '하이프 사이클(hype cycle)'을 눈여겨볼 필요가 있다. 이에 따르면 기계번역의 핵심이라 할 수 있는 딥러닝과 머신러닝은 2017년 과잉기대 단계의 정점을 지나고 있다(그림 8). 기술은 "평범하고 흔해져 눈에 띄지 않을 정도로 일상에 깊숙이 스며들어야 비로소 심대한 변화를 일으킨다"는 기술학자 셔키(Shirky, 2007: 105)의 통찰을 굳이 빌리지 않더라도 현재의 언론 주도 기계번역 담론은 실제 시장에서의 구현 시기와 적용 문제는 간과한 채 당장 획기적인 변화가 일어날 것처럼 과장된 측면이 있다.

그림 8. 2017년 신기술 하이프 사이클(Gartner, 2017)



우선 신기술 적용은 조직이 단시간에 기술을 가져다 쓰기만 하면 되는 일

방적인 과정이 아니라 “조직도 기술에 맞춰 적응해나가야 하는 상호적인 과정”으로, 시간도 필요하고(Leonard-Barton, 1988: 253) 상용화 과정에서 이뤄지는 다양한 행위별로 각각 부가가치가 발생해야 성공하기 때문이다(Jolly, 1997: Sung, 2009에서 재인용). 깁슨과 스마일로에 따르면(Gibson & Smilor, 1991) 신기술 상용화는 네 단계를 거친다. 1단계에서는 기술이 처음 개발되어 다양한 연구가 이뤄지는 한편 학회나 뉴스 등 다양한 채널을 통해 발표되고, 2단계에서는 기술개발자와 사용자가 책임을 공유하기 시작하며, 3단계에서는 기술이 조직에서 적시에 효율적으로 구현되어야 하고, 4단계에서는 시장에서 성공을 거둬야 하는데 투자수익률(ROI)이나 시장점유율로 측정된다. 그런데 4단계까지 가는 동안 경영진 지원, 정부 지원, 관계자 간 협력, 경영에 대한 이해 등 다양한 변수가 작용한다(성태경, 2017). 기계번역은 이제 막 1단계에 들어섰으며 2단계의 책임 부분은 아직 논의조차 이뤄지지 못하고 있는 상태다.

상용화는 또한 혁신기술이 기성 기업에 전파되는 경우와 파생기술이 신생기업에 전파되는 두 가지 종류로 구분되는데(Gibson et al, 2000), 기계번역은 이미 기성 LSP(language service provider)에서 제한적으로나마 사용되고 있으나 파생기술은 아직 모기술조차 미완성 단계라 전파란 더욱 요원한 일이다. 모기술 자체도 기술적 난이도가 높은데다 당장 뚜렷한 수익이 없는 가운데 방대한 데이터를 확보하고 이를 학습 및 처리할 자원에 많은 비용과 시간이 소요되는 만큼 기업 경영 관점에서 핵심 과제로 추진될 수 있는지는 의문이다. 따라서 상용화 관점에서는 구글이나 네이버 같은 극소수 대기업의 기술만 생존할 가능성이 높다. 이 경우 독과점의 부작용은 차치하고도 번역품질의 균일화 혹은 획일화에 대한 수용성이 어떤 양상으로 나타날지 예측하기 어렵다는 점에서 또 한 가지 변수가 된다.

기계번역 기술은 개발도 어렵지만 적용도 쉬운 문제가 아니다. 가령 기업들이 구글의 번역 API[12]를 사용해 각자의 목적에 맞추어 대량의 문서를 번역하려면 번역 자체는 GNMT를 이용하더라도 입력과 출력 단계에서 일정 규모의

12) Application Programming Interface. 사용자가 애플리케이션을 개발할 수 있게 도와주는 인터페이스로, 용도별로 구성해 제공된 레고 블록을 짜맞추는 것으로 이해하면 쉽다. 가령 구글번역의 경우 사용료를 내고 인증을 받아 translation API를 이용해 번역서비스나 번역애플리케이션을 개발할 수 있다.

자원을 투입해야 하며 프로젝트 규모에 따라 확장성(scalability)도 요구된다. 수천 장의 문서를 구글이 번역한 원시데이터 그대로 고객에게 전달할 수는 없기 때문이다. 특히 문서의 중요도나 난이도로 인해 고도의 포스트에디팅이 요구되는 경우에는 비용 대비 효과도 고려하지 않을 수 없다. 가독성과 설득적 효과가 중요한 문서라면 기계번역은 비효율적인 선택일 수 있다. 중요한 서버 유지 및 관리에 드는 비용도 이윤추구가 목적인 기업으로서는 중요한 변수다. 또한 기밀유지가 필요한 민감한 문서를 구글이나 네이버 같은 공개된 클라우드 API를 이용해 번역하게 허용할지는 또 다른 변수다. 언론의 기계번역 담론이 즐겨 인용하는 구글번역과 네이버 파파고 번역은 범용 예시라 일반인이 일상에서 필요로 할 만한, 공개되어도 무방한 주제나 내용, 분량의 번역에 적합하다. 가령 특정 웹사이트 안내문 같은 정보성 텍스트나 이메일 문의처럼 단순한 상호작용 문서를 번역하는 것은 지금 수준으로도 큰 무리가 없다. 그러나 이는 그 동안도 전문번역사에게 적정 번역료를 지급하며 의뢰하는 종류의 번역은 아니었다. 여느 서비스업이 그렇듯 번역시장도 문서의 난이도, 번역료, 관련 업무 및 작업의 범위, 종류, 성격에 따라 다양한 세그먼트로 나뉠 수 있고, 번역업체의 재정규모도 다양하기 때문에 언론이 주장하듯 모든 과정이 기계번역으로 자동화되고 대체되기는 힘들다고 봐야 할 것이다.

5. 나가며

기계번역 기술에 아직 여러 가지 한계가 있어도 앞으로 더욱 개선될 것임은 부인할 수 없는 사실이다. 그렇다고 해도 미래는 아무도 알 수 없는 미지의 영역이다. 뛰어난 음질의 디지털 기반인 CD와 MP3가 등장했을 때만 해도 아날로그인 LP '레코드판'은 영원히 사라질 것으로 예상됐지만 디지털 혁명이 완성기에 이른 2007년 이후 미국에서 LP앨범 판매량이 오히려 급증해 2015년 기준 전체 음원수익의 4분의 1에 육박했다는 사실(Sax, 2016)은 인간과 기술의 그리 단순하지 않은 관계를 단적으로 보여주는 예다. 기계번역이라는 편리하고 획일화된 산출물에도 그 같은 반작용이 일어날지는 알 수 없다. 그러니 기계번역 담론이 제기하는 미래에 대한 섣부른 예측이나 전망보다는 번역학이 실질적

으로 맡을 수 있는 역할에 초점을 맞추는 것이 더 생산적일 것이다. 당장은 번역교육의 방향 설정이 급선무일 것이고 실무 및 연구에서는 번역평가/감수의 중요성이 커졌다. 무엇이 어떤 이유에서 잘된 번역 혹은 수용가능한 번역인지, 인간번역과 기계번역은 어떤 예측가능한 차이를 보이는지, 기계번역이 '잘됐다'는 건 어떤 수준을 말하는지 등을 명확히 규정하는 작업은 공학자가 아닌 번역학자가 할 수 있고 또 해야 하는 일이다.

번역사가 필요 없게 된다는 식의 언론의 주장에 대한 대응도 번역학의 몫이자 의무다. 신기술이 발명되면 결국 사라지는 직업도 있지만 기존 직업의 성격을 보완 및 강화한 새로운 직업이 창출된다는 사실은 이미 역사적으로 수많은 사례를 통해 증명되어왔다. 외국어를 잘하는 사람들이 늘고 비전문가번역(김순미, 2016; 이지민, 2016)까지 확산되고 있는 현 추세에서는 단순 정보전달 수준의 번역 또는 일반인의 접근성이 좋고 재미나 보람 등 심리적 보상이 동기가 되는 번역은 어차피 NMT 기술이 없었어도 경쟁력이 떨어질 수밖에 없다. 반대로 비전문가나 기계와 차별화되는 희소가치 있는 번역사가 된다면 기회는 오히려 더 많다고 봐야 할 것이다. 번역문서에 대한 포스트에디팅/프리에디팅 기술뿐 아니라 기계번역 프로그램을 번역API 등을 이용해 자신에게 맞춤화해 더 효율적으로 활용할 수 있는 기술적 능력과 번역능력을 겸비한다면 대체불가능한 전문성을 인정받는 인력이 될 것이다. 순수인문학 성격의 번역은 번역가의 문체가 중요한 문학이나 비문학이라도 언어유희 등 창의력이 요구되는 장르, 인간심리에 대한 깊은 이해를 바탕으로 감성에 호소하고 설득해야 하는 장르 등에 국한될 가능성이 높은 만큼, 번역교육 커리큘럼은 기존의 툴사용번역(CAT) 능력에 더해 번역API를 이해하고 적용할 수 있는 기본 코딩능력을 포함해 기계번역 프로그램을 활용 및 관리하는 기술적 능력, 번역결과물을 평가/설명/선택할 수 있는 능력을 함양하는 과정과 기계사용이 선호되지 않는 인문학 장르에 대한 번역능력을 배양하는 과정 두 가지 트랙으로 나누어 운용할 수 있을 것이다. 둘 다 기계가 인간을 대체하기 힘든 능력인 만큼, 이를 겸비했을 때 역사적으로 항상 그랬듯 새로운 기회가 열릴 것이다. 물론 지금으로서는 기계번역의 한계나 발전 속도 등 모든 것이 예측에 불과하며 그것이 본고의 명백한 한계다.

NMT 등장 이후 번역학은 격동의 시기를 지나고 있지만 사실 이 모든 문제는 결국 '무엇이 잘된 번역인가'라는 다분히 (간)주관적이고 인문학적인 문제로

귀결된다. 번역학이 태동하기 이전부터 존재했던 화두다. 번역학의 본질 자체는 달라지지 않은 것이다. 따라서 기후변화 용어를 빌리자면 교육과 연구에서는 '선제적으로 적응(adaptation)'해나가되 언론이 주도해온 '인간번역의 위기' 프레임에는 차분히 대응해야 할 것이다. 본고는 그 필요성을 환기했다는 데서 의미를 찾고자 하며 구체적인 교육 방안은 후속 연구로 남기기로 한다.

참고문헌

강지혜 (2007)「출판번역과 텍스트의 '재맥락화': '셀프헬프'의 번역을 중심으로」,『번역학연구』8(1): 7-36.

강지혜 (2008)「번역에서 인용의 문제: CNN.com 뉴스텍스트를 중심으로」,『번역학연구』9(4): 7-40.

김상호 (2007)「언론의 객관성에 대한 분석적 고찰」,『언론과학연구』7(3): 5-38.

김순미 (2016)「디지털 시대 비전문가들의 참여번역 현상 －그 의미와 번역계의 미래에 주는 시사점」,『번역학연구』17(3): 7-32.

마승혜 (2017)「한영 기계번역 포스트에디팅에 대한 경험적 고찰」,『한국번역학회 2017 가을학술대회 발표논문집』: 84-98.

박옥수 (2016)「한영 병렬 코퍼스와 기계번역에서 의존명사 －'-것'이 포함된 어휘의 번역 방식 연구」,『동아인문학』37: 469-492.

성태경 (2017)「IT업계의 기술상용화 주요 요인에 관한 연구」,『서비스연구』7(2): 91-105.

송연석 (2011)「이데올로기가 제도적 번역에 미치는 영향 연구: 천안함 침몰사건 관련 외신인용기사를 중심으로」,『번역학연구』12(1): 145-165.

송연석 (2013)「뉴스편역의 이데올로기: 4대강 사업 관련 외신편역의 분석」,『통번역학연구』17(4): 75-100.

신지선 (2017)「테크놀로지 패러다임에서의 번역능력 재조명」,『통번역학연구』21(4): 51-71.

윤희주 (2007)「완곡어법과 위악어법의 번역원인 및 번역방법」,『번역학연구』8(1): 193-220.

이상빈 (2013)「대명사 ‘그녀’의 수용에 관한 독자반응 연구」,『통번역학연구』 17(4): 121-137.

이상빈 (2017)「학부번역전공자의 기계번역 포스트에디팅, 무엇이 문제이고, 무엇을 가르쳐야 하는가?」,『통역과 번역』19(3): 37-64.

이정훈 (2013)「한국 언론의 상업화 논의에 관한 비판적 검토」,『한국언론정보학보』: 315-328.

이주리애 (2011)「기사문 번역의 재맥락화 양상에 대한 일고찰: 뉴스위크 한국판과 일본판의 평가어 분석을 중심으로」,『번역학연구』12(2): 157-184.

이지민 (2016)「디지털 시대 비전문가들의 참여번역 현상 －그 의미와 번역계의 미래에 주는 시사점」,『통번역학연구』20(2): 103-128.

장애리 (2017)「국내 기계 통번역의 발전 현황 분석 －한중 언어쌍을 중심으로」,『번역학연구』18(2): 171-206.

최동익 (2016)「과거시제의 영어 복문과 중문에 대한 번역가 번역과 기계번역 비교」,『언어학연구』39: 377-398.

최현희・배만호 (2015)「공손성 번역의 적절성에 관한 연구」,『새한영어영문학』 57(2): 151-169.

최효은 (2017)「특허 기계번역 결과물의 평가 －KIPRIS의 무료 한영 기계번역을 중심으로」,『통역과 번역』19(1): 139-178.

한국외대 통번역대학원 번역평가인증 연구팀 (2016)「번역인증제도 (실무편)」,『2016 한국외대통번역연구소-국방어학원 합동학술대회 ‘언어, 통번역의 평가 및 인증』발표집: 23-35.

한승희 (2017)「기계번역 인간번역 트랜스크리에이션의 문체 비교: 광고 번역을 중심으로」,『통번역학연구』21(2): 163-188.

함수진・류수린 (2010)「기술문서의 한일기계번역 문제에 대한 통제언어 연구 －‘-되다’ 구문의 기계번역 수월성 제고를 위한 통제규칙」,『번역학연구』 11(4): 191-236.

황은하 (2014)「언어학적 지식에 기반한 한중 뉴스 표제의 기계번역」,『번역학연구』15(5): 333-362.

홍원식・김은정 (2013)「TV 미디어 비평의 어제와 오늘: <미디어비평(KBS)> 10년, 내용분석」,『한국언론정보학보』64: 59-84.

Aranberri-Monasterio, Nora and Sharon O'Brien (2009) 'Evaluating RBMT Output for –ing Forms: A Study of Four Target Languages', *Technology Evaluation* 8, 105-122.

Arnold, Doug, Louisa Sadler and Lee Humphreys (1993) 'Evaluation: An Assessment', *Machine Translation* 8, 1-24.

Banerjee, Satanjeev and Alon Lavie (2005) 'METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments', in *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, 65-72.

Baroni, Marco, Georgiana Dinu and German Kruszewski (2014) 'Don't Count, Predict! A Systematic Comparison of Context-Counting vs. Context-Predicting Semantic Vectors,' in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 238-247.

Bassnett, Susan (2011) 'The Translator as Cross-Cultural Mediator', in Kevin Windle and Kirsten Malmkjaer (eds) *The Oxford Handbook of Translation Studies*, New York: Oxford University Press, 94-107.

Blackburn, Perry L. (2007) *The Code Model of Communication – A Powerful Metaphor in Linguistic Metatheory*, SIL International.

Broek, Raymond van den (1978) 'The Concept of Equivalence in Translation Theory: Some Critical Reflections,' in James Holmes, Jose Lambert and Raymond van den Broek (eds) *Literature and Translation*, Leuven: Academic, 29-47.

Callison-Burch, Chris, Miles Osborne and Philipp Koehn (2006) 'Re-evaluating the Role of BLEU in Machine Translation Research,' in *Proceedings of the 11th Conference of the European Chapter of the ACL*, 249-256.

Callison-Burch, Chris, Cameron Fordyce, Philipp Koehn, Christof Monz and Josh Schroeder (2007) (Meta-)evaluation of Machine Translation, in *Proceedings of the Second Workshop on Statistical Machine Translation*, 136-158.

Chen, Yamei (2011) ‘The Transltor’s Subjectivity and Its Constraints in News Transediting: A Perspective of Reception Aesthetics,’ *Meta* 56(1), 119-144.

D’Andrea, Alessia, Arianna D’Ulizia, Fernando Ferri and Patrizia Grifoni (2017) ‘EMAG: An Extended Multimodal Attribute Grammar for Behavioral Features,’ *Digital Scholarship in Humanities* 32(2), 251-275.

Gibson, David V. and Raymond W. Smilor (1991) ‘Key Variables in Technology Transfer: a Field-Study Based Empirical Analysis,’ *Journal of Engineering and Technololgy Management* 8(4), 287-312.

Gibson, David V. and Molly N. Rogalev (2000) ‘Transfer of R&D results from Public Research Centers: Lessons Learned in the U.S. and Applied to Russian/CIS Research Centers,’ in *Proceedings of 10th Annual Association Technopark Meetings*, June 29-30.

House, Julian (1997) *Translation Quality Assessment: A Model Revisited*, Tübingen: Gunter Narr.

Hovy, Eduard H. (1999) ‘Toward Finely Differentiated Evaluation Metrics for Machine Translation’, in *Proceedings of the Eagles Workshop on Standards and Evaluation*, 127-133.

Jakobson, Roman (1960) ‘Closing Statement: Linguistics and Poetics,’ in Thomas A. Sebeok (ed) *Style in Language*, Cambridge: AM: MIT.

Johnson, Melvin, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes and Jeffrey Dean (2017) ‘Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation,’ *Transactions of the Association for Computational Linguistics* 5, 339-351.

Jolly, Vijay K. (1997) *Commercialization New Technologies*, Boston: Harvard Business School Press.

Kang, Jihae (2007) ‘Recontextualization of News Discourse: a Case Study of Translation of News Discourse on North Korea’, *The Translator* 13(2), 219-242.

Kang, Jihae (2010) 'Positioning and Fact Construction in Translation,' in Maeve Olohan and Maria Calzada-Perez (eds) *Text and Context*, London:St. Jerome Publishing.

Leonard-Barton, Dorothy (1988) 'Implementation as Mutual Adaptation of Technology and Organization,' *Research Policy* 17(5), 251-267.

Luong, Minh-Thang, Hieu Pham and Christopher D. Manning (2015) 'Effective Approaches to Attention-based Neural Machine Translation', in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 1412-1421.

Munday, Jeremy (2007) 'Translation and Ideology – A Textual Approach,' *The Translato*r 13(2), 195-217.

Nida, Eugene (1964) *Toward a Science of Translating*, Leiden: E. J. Brill.

Papineni, Kishore, Salim Roukos, Todd Ward and Wei-Jing Zhu (2002) 'BLEU: a Method for Automatic Evaluation of Machine Translation', in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311-318.

Peled, Lotem and Roi Reichart (2017) 'Sarcasm SIGN: Interpreting Sarcasm with Sentiment Based Monolingual Machine Translation.' in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 1690-1700.

Pierce, John R. and John B. Carroll (1966) *Language and Machines – Computers in Translation and Linguistics (ALPAC Report)*. Washington D.C.: ALPAC.

Pym, Anthony (2013) 'Translation Skill-Sets in a Machine-Translation Age', *Meta* 58(3): 487-503.

Sax, David (2016) *The Revenge of Analog: Real Things and Why They Matter*, New York: Public Affairs.

Shannon, Claude E. (1948) 'A Mathematical Theory of Communication,' *The Bell System Technical Journal* 27, 379-423, 623-656.

Shannon, Claude E. and Warren Weaver (1949/1964) *The Mathematical Theory*

of Communication, Urbana: The University of Illinois Press.

Shirky, Clay (2009) *Here Comes Everybody: The Power of Organizing without Organizations*, New York: Penguin Books.

Shoemaker, Pamela and Stephen D. Reese (1996) *Mediating the Message: Theories of Influences on Mass Media Content*, New York: Longman.

Sung, Tae Kyung (2009) 'Technology Transfer in the IT Industry: A Korean Perspective' in *Technological Forecasting & Social Change* 76, 700-708.

Sutskever, Ilya, Oriol Vinyals and Quoc V. Le (2014) 'Sequence to Sequence Learning with Neural Networks,' in *Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS)*, 3104-3112.

Vermeer, Hans (1984/2000) 'Skopos and Commission in Translational Action,' in Lawrence Venuti (ed) *The Translation Studies Reader*, London/New York: Routledge, 221-232.

Wang, Tian and Kyunghyun Cho (2016) 'Larger-Context Language Modelling with Recurrent Neural Network,' in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1319-1329.

Weng, Rongxiang, Shujian Huang, Zaixiang Zheng, Xinyu Dai and Jiajun Chen (2017) 'Neural Machine Translation with Word Predictions,' in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 136-145.

White, John (2003) 'How to Evaluate Machine Translation,' in Harold Somers (ed) *Computers and Translation: A Translator's Guide*, Amsterdam/Philadelphia: John Benjamins, 211-244.

인터넷 자료

김경미 (2016. 11. 8) 「내가 이러려고 영어 배웠나, AI가 번역 다 해주네」, 『중앙일보』, 2018년 1월 15일 검색.

김종일 (2017. 12. 30) 「구글 번역 총괄 연구원 "AI가 인간 번역가 대체하기 힘들어"」, 『조선일보』, 2018년 1월 15일 검색.

송혜리 (2017. 2.28)「목에 걸면 외국어가 술술...통역사 필요없는 웨어러블」,『디지털타임스』, 2018년 1월 15일 검색.

정상혁 (2017. 1 18)「진화하는 번역기...사라지는 번역가?」,『조선일보』, 2018년 1월 15일 검색.

정재승 (2016. 9. 11)「인간 지성이 인공지능보다 뛰어난 까닭은」,『한겨레』, 2018년 1월 15일 검색.

주영재 (2016. 2.1)「10년내 언어장벽 사라진다」,『경향신문』, 2018년 1월 15일 검색.

Bahdanau, Dzmitry, Kyunghyun Cho and Yoshua Bengjo (2015) 'Neural Machine Translation by Jointly Learning to Align and Translate', arXiv:1409.0473 [cs.CL]

Li, Aodong, Shiyue Zhang, Dong Wang and Thomas Fang Zheng (2017) 'Enhanced Neural Machine Translation by Learning from Draft,' arXiv:1710.01789 [cs.CL]

Neubig, Graham (2017) Neural Machine Translation and Sequence-to-sequence Models: A Tutorial, https://arxiv.org/abs/1703.01619.

Wu, Yonghui, Mike Schuster, Zhifeng Chen, Quoc V. Le and Mohammad Norouzi (2016) 'Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation,' https://arxiv.org/abs/1609.08144

Zhou, Long, Ziajun Zhang and Chengqing Zong (2017) 'Look-ahead Attention for Generation in Neural Machine Translation,' arXiv:1708.09217 [cs.CL]

[Abstract]

A Critical Look at Discourses on Machine Translation

Song, Yonsuk
(Hankuk University of Foreign Studies)

Machine translation (MT) has recently received much attention both in academia and industry, but the current discourse on MT and its impact on professional translators has been predominantly driven by the local media, which is often criticized for sensationalism and commercialism. This paper aims to critically examine the media-led discourses on MT by reviewing the latest studies on neural MT and comparing the views on translation held by researchers of translation studies (TS) and MT. The paper sets out to explore the differences between TS and MT in how they model a communication process and view the notion of 'meaning' and 'context.' It then examines the validity of BLEU, a popular MT evaluation metric. Finally, the paper discusses the problems of implementing the technology from a commercialization perspective and provides implications for translator training.

▸ Key Words: neural machine translation, machine learning, BLEU, translator training, translation evaluation
▸ 주제어: 신경망기계번역, 머신러닝, BLEU, 번역교육, 번역평가

송연석
한국외국어대학교 통번역대학원 한영과 조교수
yonsuk@gmail.com
관심분야: 기계번역, 번역교육, 이데올로기

논문투고일: 2018년 2월 2일
심사완료일: 2018년 3월 13일
게재확정일: 2018년 3월 20일