

‘인공신경망 기계번역’ 관련 기사의 의미구성에 대한 연구: 헤드라인에 대한 의미네트워크분석을 중심으로

천 중 성 · 강 수 정
(한양대 · 숙명여대)

1. 서론

2016년 구글이 통계기반 기계번역(SMT: Statistical Machine Translation)을 역사에 묻어두고 인공신경망 기계번역(이하 NMT: Neural Machine Translation)으로 ‘구글번역’을 새롭게 공개한 이래, NMT가 번역 서비스를 이용하는 대중적 수요의 상당 부분을 흡수하고 있는 것으로 보인다. 2021년 기준 네이버 ‘파파고’의 월간 이용자 수는 1,300만 명에 달하는데(아시아경제 2021) 지금은 그 규모가 더욱 확장되었을 것이다. 최근에는 단순한 번역기의 역할을 벗어나 음성인식, 뉴스, AR·VR 등 다른 기술들과 결합되어 언어 전환의 일상적 도구화가 되고 있다(김준석 2017). 이러한 성장은 딥러닝 방식의 알고리즘으로 그 성능이 실시간으로 향상되고 있고, 누구나 무료로 이용이 가능한 높은 접근성과 편의성을 배경으로 들 수도 있다(강수정 2021: 44). NMT의 확산에 대해 미디어의 역할에 초점을 맞춘 논의도 등장하고 있다(송상헌 2020: 7). 미디어가 ‘기계와 인간의 대결’ 구도를 부각하거나 통번역사를 미래에 사라질 우선적 직업으로 자극적으로 묘사하여 NMT의 성능과 품질에 대한 이용자들의 인식이 실

제보다 확장되었을 것이라는 시각이다.

사실 미디어의 선정적 묘사가 아니더라도 기술의 확산에 미디어가 일정한 역할을 한다는 논의는 지속적으로 전개되어왔다(Rogers 2010). 특히 NMT에 관한 기사의 양으로만 보면 보도의 증가세가 확연하게 나타난다. 한국언론진흥재단의 뉴스데이터 분석 서비스인 빅카인즈(BIGKinds)에서 2008년부터 2017년까지 관련 기사를 검색어 ‘기계번역’으로 추출한 결과 400건의 기사가 확인되었다).

〈표 1〉 연도별 ‘기계번역’으로 검색된 기사

연도	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
기사	29	7	9	15	6	27	11	37	118	141

〈표 1〉과 같이 2015년까지 평이한 수준을 보이던 기계번역 관련 기사량은 ‘구글번역’과 ‘파파고’가 등장한 2016년 이후로 급격하게 증가하였다. 미디어의 보도가 신기술인 ‘NMT’에 집중되면서 이용자들의 신기술에 대한 인지 증가에 일정한 역할을 했을 것이라 유추할 수 있다.

일반적으로 뉴스는 여론을 파악하고 반영하여 사회적 의제형성에 기여하는 것으로 알려져 있다. 그러나 여론은 세상에서 일어나는 실제 현상에 반응하기 보다는 뉴스 미디어가 만들어 낸 ‘유사 환경(pseudo-environment)’에 반응하는 것으로(반현과 Maxwell 2007: 10), 있는 그대로의 현실이 아닌 미디어가 재구성한 현실을 보도하는 구성주의적 특성을 보인다. 기술과 뉴스 보도의 관계에 관한 연구에서도 일반 공중들의 새로운 기술에 대한 접근이 주로 미디어를 통해서 이루어지며 미디어가 의미화하고 구성하는 것이 그들의 이슈 해석 혹은 사회적 이슈 해석에 영향을 미칠 수 있다고 보고한다(최윤정과 권상희 2014: 275).

이러한 면에서 NMT 관련 보도가 급격하게 증가한 2016년 이후의 기사들이 기계번역과 관련해서 어떠한 현실을 재구성하고 있는지 파악할 필요성이 제

1) 구글이 독자적인 엔진으로 ‘구글번역’을 출시한 것이 2007년이고, 2016년에 인공지능경망 기반 ‘구글번역’을 출시하고, 한 달의 시차를 두고 네이버의 ‘파파고’가 론칭, 2017년 카카오를 비롯한 다양한 서비스가 본격화되었다. 이에 따라 검색 연도를 ‘구글번역’ 출시 이듬해를 기준으로 잡아 2008년부터 2017년까지 10년간 조사하였다.

기된다. 기술에 대한 미디어의 의미구성이 이용자의 인지에 영향을 미칠 수 있다는 것을 전제하면, 이용자들은 구성된 의미의 특성으로 기술을 수용할 수 있기 때문이다. 또한 일반공중 혹은 사회적 분위기가 기계번역에 대해 긍정적이면 그 분위기는 전문가인 번역사들의 기술에 대한 평가와 인식에도 영향을 미칠 수 있다(천종성 2020: 275)는 점에서 미디어를 통해 구성되는 현실을 파악하는 것은 전문 번역사들의 기계번역 평가에도 영향을 미치는 중요한 요소가 될 수 있다.

따라서 실재적 현실로 자리 잡아가고 있는 NMT가 미디어와 뉴스를 통해서 어떻게 구성되었는지 살펴보는 것이 이 연구의 목적이 된다. 이 연구에서는 기계번역에 대한 뉴스 분석을 통해서 어떠한 이슈가 강조되고 그 기반이 되는 가치는 무엇이며 이 새로운 번역 현상에 미디어와 우리 사회가 어떻게 접근하고 있는지 실증적으로 살펴볼 것이다. 이를 위해서 의미네트워크분석(semantic network analysis) 방법을 적용한다.

의미네트워크분석은 사회네트워크분석(social network analysis)의 방법론을 메시지 커뮤니케이션에 적용한 것으로 개념 간의 연결과 관계성을 통해 특정 의미의 메시지가 형성되는 구조적 맥락을 관찰할 수 있게 한다(한관중 2003: 226). 따라서 NMT와 관련된 뉴스보도의 주요 이슈를 파악하고 파악된 개념들이 어떠한 관계를 형성하고 있는지, 그 의미의 연결구조는 어떻게 구성되는지를 분석하는 데 유용하게 적용될 수 있다. 또한 분석의 결과는 우리 사회가 NMT를 어떠한 방식으로 수용하는지, 그리고 그 방향성에서 번역학 연구자와 실무 종사자에게 어떠한 함의를 줄 수 있는지 파악할 수 있게 할 것이다.

2. 이론적 배경 논의

2.1 NMT에 대한 선행연구

기계번역에 대한 연구자들의 관심 역시 2016년 NMT가 등장하면서 크게 늘었다. 2007년에야 처음 발표된 기계번역 관련 논문이 2017년에 14편, 2018년에 21편으로 폭발적으로 늘어났다(최문신 2019: 283). SMT의 시기인 1990년대

의 연구는 품질평가나 오류분석이 주를 이루었으며 그 결과는 시스템 개선을 위한 자료로 활용되었던(이하자와 안동미 2011: 152) 것으로 보인다. 그러나 NMT가 도입되면서 품질평가와 관련한 연구의 대부분은 기계번역의 품질을 긍정적으로 평가하고 있었다(김순미 2017; 신지선 2017; 장애리 2017; 이창수 2019).

기계학습에 기반한 NMT의 특성상 이용자가 늘어나면 늘어날수록 번역의 품질은 더욱 향상되고, 그로 인해 이용자가 늘어나고 다시 품질이 향상되는 선순환구조가 형성된다(김준석 2017: 78). 따라서 1천만 명이 넘는 사용자 데이터에 기반한 상시적인 딥러닝으로 인해 특정한 시점에 측정하는 번역 품질평가의 결과가 영속적인 결과로 제시될 수 있는지에는 의문이 든다. 이에 최근 기계번역에 대한 연구는 단순한 품질평가나 오류분석에서 벗어나 포스트에디팅, 기계와 인간의 협업, 번역교육의 패러다임 전환 등으로 확장되고 있다. 기계와 인간의 협력적 구조를 모색하기도 하고(김순미 2018), 기계와 인간의 협업을 통해 완성된 번역의 주체가 누구인지의 문제를 고민하기도 하며(이승일 2018), 기계와 인간의 협업에 대비한 새로운 교육모델을 개발하고 책임소재를 명확히 하는 제도적 방안에 대한 대응을 주장하기도 한다(마승혜 2018). 한편 기계번역의 실질적 최대 수용자인 번역사들을 대상으로 한 수용행위에 관한 연구(강수정 2020, 2021; 천중성 2020)에서는 기계번역에 부정적 인식을 보일 것으로 예측되었던 전문가 집단인 통번역사나 예비 통번역사들 모두 기계번역을 자신의 업무에 활용할 유용한 도구로 인식하며 품질 중심의 적극적인 수용 태도를 보인다는 흥미로운 사실이 밝혀졌다. 이는 기계번역이 품질을 논하는 단계를 지나 이미 일상의 영역으로 들어왔으며, 이제는 수용 여부의 문제가 아닌 기술의 발전 방향과 이에 따른 인간의 적응 문제가 주요한 관심의 대상이 될 수밖에 없음을 의미한다.

2.2 기술에 대한 뉴스의 의미구성

뉴스는 객관적으로 존재하는 현실을 지면이나 방송으로 재구성하여 보도하지만, 미디어 관련 연구들은 뉴스가 있는 그대로의 현실이 아닌 미디어가 구성하는 현실임을 설명한다. 대표적으로 프레임이론(frame theory)이 있다(Entman

1993). 미디어가 프레임을 설정하는 것은 세상에 일어나는 수많은 사건들 사이에서 사실이라고 여겨지는 부분의 특정 측면을 선택하고 강조하여, 텍스트에서 그 부분이 다른 것들보다 명확히 두드러지게 만드는 것으로 어떠한 문제에 대한 정의나 인과관계의 해석, 도덕적 평가를 할 수 있도록 수용자를 인도한다(Entman 1993: 52). 수용자들은 미디어가 선택하여 제시하는 선택성(selectivity)과 그것을 다른 것들보다 두드러지게 만드는 현저성(salience)이라는 특성을 통해 뉴스가 제시하는 프레임을 받아들이게 된다(이준웅 2005: 138).

기술에 대한 뉴스보도는 기술의 현저한 특징을 부각시켜 현상을 평가하는 시각을 대중과 공유하게 된다는 면에서 기술 수용에 있어 중요한 역할을 한다. 혁신적인 아이디어나 새로운 기술이 성공하는 사례를 살펴보면 S자 형태의 예측가능한 패턴으로 사회 전반에 퍼져나간다(Newell, Genschel and Zhang 2015: 28). 새로운 기술이 등장하는 초기에 프레임이 적용되면 이용자들은 미디어가 제시하는 기술의 현저한 특성에 기반해 기술을 해석하거나 평가하여 수용한다.

이와 같은 신기술 보도에 대한 프레임의 접근은 인간배아복제와 줄기세포 같은 생명과학 연구를 비롯해 원자력, 우주과학 연구에 적용되어 왔지만(최윤정과 권상희 2014), 정치권력의 진보와 보수매체의 이슈(강명구 2000) 혹은 사회적 갈등 보도(오창우 2017)와 같이 광범위하게 적용된 분석은 아니었다. 오히려 과학적 지식이 요구되는 기술에는 가치중립적 접근이 요구되는 경향이 있다(김만재와 전방욱 2012: 20). 따라서 기술에 대한 보도는 프레임과 가치중립적 접근이 혼재하는 것으로 보인다. 이러한 경향은 신기술에 대한 미디어의 보도가 인간이 가진 합리적 특성뿐만 아니라 직관적이고 본능적인 감각에 함께 작용하기 때문이기도 하다(장한수, 조성경과 홍정석 2020: 4). 선행연구들에 따르면 미디어가 구성하는 기술적 특성에 대한 의미가 이용자의 이성적 감성적 영역에 작용할 수 있지만, 사회적으로 논쟁적 이슈가 부각되지 않으면 가치중립적 보도에 기반하여 기술에 대해 수용할 수 있는 범위를 보여주고 사회적 인식을 형성하는 데 기여한다.

미디어를 통해 구성되는 뉴스는 결국은 언어를 통해 이용자와 공중에게 전달된다. 새로운 기술이나 아이디어에 뉴스의 프레임과 구성방식을 적용할 수 있는 것은 대부분 뉴스가 언어로 구성되기 때문인데, 뉴스에서 사용하는 특정 단어와 텍스트는 사건이나 인물에 대한 의미와 특성에 대한 관점을 보여주는

도구가 된다(van Dijk 2013: 175). 또한 뉴스가치는 언론이 기사의 어떠한 관점을 앞세우거나 강조함으로써 나타나는 무엇인가로 설명되는데, 여기서 관점 역시 뉴스 언어이다. 언어는 사실을 표현하고 의미를 현저하게 구성하는 기본 요소이기 때문에 뉴스가치와 밀접한 관련이 있다(이완수와 최명일 2019: 111).

뉴스의 언어적 속성을 가장 잘 나타내는 것은 헤드라인(headline)이다. 헤드라인은 특정한 이슈의 속성을 압축하여 표현하는 과정에서 언어적 구성요소를 최대한 활용함으로써 독자에게 미치는 영향을 최대화하는데(Conboy 2007: 13), 뉴스 내용을 포함한 기사의 전체적 방향성을 인지하고 뉴스의 핵심적 키워드를 부각시켜 여론에 미치는 관심을 증폭시킨다(이준호 2015: 251). 최근에는 포털과 모바일을 중심으로 한 뉴스 유통구조로 인해 헤드라인에 의한 ‘예독가능성(previewability)’도 높아졌다(정수영과 황경호 2015: 309). 기사 내용을 모두 읽지 않고 헤드라인의 키워드만으로 해당 이슈와 내용을 판단하게 된다는 것이다. 이는 헤드라인에 드러나는 언어적 특성을 파악하면 뉴스 전체가 구성된 현실을 파악하는 데 효과적일 수 있음을 의미한다. 또한 뉴스는 의미구조화를 통해 특정 가치나 규범을 나타내기 때문에 헤드라인에 나타난 단어의 관계가 중요하게 된다. 담론으로서의 뉴스에서 사용하는 개념과 단어들의 조합은 우연한 산물이 아니라 체계적이고 의도적인 사회적 실천이라는 의미(van Dijk 2013: 175)에서 기계번역에 대한 뉴스보도에서 드러나는 단어들의 의미적 연관성을 파악한다면 우리 사회의 기계번역 의미체제 구성방식을 이해할 수 있다.

2.3 뉴스와 의미네트워크분석

의미네트워크분석은 사회구조를 구성요소 간의 상호작용적 관계에 기반하여 파악하려는 사회네트워크분석을 언어기반의 메시지 구성과 커뮤니케이션 현상에 적용한 것이다. 소셜네트워크분석 혹은 사회연결망분석이라고 불리는 네트워크 분석 방법은 성별과 나이의 인구통계학적 속성 혹은 인지나 태도와 같은 개인의 심리적 특성의 합이 사회를 설명할 수 있다는 속성주의적 시각을 극복하려는 연구방법이다(한관중 2003: 224). 사회는 사람들 간의 관계를 통해 형성되고 그 연결망을 통해 정보와 자원이 유통되면서 인간은 사회적 네트워크의 위치에 따라 다르게 행동한다는 관계론적 접근방식이다(김용학, 윤정로, 조혜선

과 김영진 2007: 74). 사회 구성원들의 지속적인 상호작용 관계는 규칙적인 패턴을 가진 네트워크를 형성하게 되는데, 네트워크 안에서 사람이나 객체는 노드(node)로 표현되고 노드 간의 연결은 링크(link)로 표현한다. 노드와 링크의 규칙적인 관계 패턴을 파악하여 전체구조를 설명하기 위해서 통계적 방법과 수학적 알고리즘에 기반한 다양한 개념과 방법을 제시하게 된다. 자연과학에서는 복잡계 연구의 일환으로 네트워크 분석이 발달했고, 사회과학에서는 네트워크를 통한 사회구조를 분석해왔다(한관중 2003: 222).

의미네트워크분석은 이러한 네트워크 분석방법을 언어기반의 텍스트에 적용하여 단어와 단어의 관계를 통해서 메시지의 의미를 파악한다. 언어에 기반한 지식은 단어 사이의 관계를 통해 형성된 네트워크라는 인식에 근거하는 것으로, 핵심 단어들어 어떤 특정한 형태로 결합될 때 특별한 경향이 발생하기 때문에 의미의 네트워크를 분석할 수 있다(박한우와 Loet 2004: 1378). 네트워크 기반의 텍스트 분석에 ‘언어네트워크분석’과 ‘의미네트워크분석’을 혼용하여 쓰기도 하지만, 전자가 언어 텍스트 자체의 관계 분석이라면 후자는 언어 텍스트에 내재된 의미론적 속성에 대한 분석이다(이수상 2014: 50). 따라서 의미네트워크분석은 개별 단어의 조합에만 주목하는 것은 아니다. 전체 네트워크 안에서 어떠한 위치를 차지하고 있는지, 어떠한 구조를 형성하고 있는지에 주목하여 메시지 안에 숨어 있는 커뮤니케이션의 의미를 계량화하여 구조적 패턴을 추출할 수 있다(최윤정과 권상희 2014: 252). 때문에 의미네트워크분석을 뉴스 텍스트에 적용하면 뉴스보도에 등장하는 주요 단어를 파악하고, 그 단어들(node) 사이의 연결(link)을 통해 뉴스가 지닌 의미의 연결구조와 패턴을 파악하여 메시지와 커뮤니케이션의 의미를 보다 분명하게 알 수 있다. 정치뉴스에 공동으로 출현하는 빈도를 파악하면 정치인들의 권력구조를 파악할 수 있고(강명구 2000) IR(investor relations)과 같은 경제뉴스를 개념화하는 데에도 적용 가능하다(문안나 2016). 의미네트워크분석을 미디어 분석에 적용하면 주로 사회적 갈등구조를 뉴스가 어떻게 의미적으로 구성하고 있는지 파악할 수 있게 된다. 의료민영화 찬반 이슈를 이해하고(김유호 2012), 신공항 건설 이슈(김문수와 김순양 2018), 미세먼지 이슈(차현주, 전재민과 권상희 2020) 등 사회적 이슈와 갈등이 어떻게 구성되는지 파악할 수 있다. 최근에는 빅데이터 (최윤정과 권상희 2014), AI(임종수 외 2017), 넷플릭스(이형민과 박진우 2020) 등 새로운 사

회적 현상이나 새로운 기술에 대한 미디어의 의미네트워크 연구에 적용되고 있으며, 본 연구 또한 이러한 연구의 연장선상에서 실증적 기여를 하게 될 것이다.

의미네트워크분석은 간혹 내용분석과 개념상 혼란이 있는 경우가 있다. 메시지 분석 연구에서 내용분석과 의미연결망 분석을 구분하는 가장 큰 차이는 전통적 내용분석이 사전에 설정한 범주를 통해 메시지에 나타난 단어를 분석하는 데 중점을 둔다면, 의미연결망 분석은 특정 단어와 동시에 출현하는 단어가 만들어 내는 것이 무엇인가에 중점을 둔다(박지영, 김태호와 박한우 2013: 47). 전통적인 내용분석 방법이 노동집약적이고 연구자의 주관적 해석을 배제할 수 없다면 의미연결망 분석은 이러한 단점을 보완할 수 있는 컴퓨터 기반 메시지 분석 기법이라고 볼 수 있다(박한우와 Loet 2004: 1377). 방법론의 철학적 접근에 기반하면 내용분석이 가설 연역적 방법이라면 의미네트워크분석은 통계적 데이터가 제시하는 텍스트 패턴을 통해 의미를 해석하는 귀납적 접근에 가깝다. 데이터에 기반하여 메시지에 나타난 주요 개념 사이의 의미론적 연관을 파악하기 때문에 텍스트마이닝과 같은 빅데이터 분석에도 적용되고 있다. 즉 많은 양의 텍스트를 해체한 후 다시 조합하여 텍스트 행간 의미를 잡아 드러나지 않은 의미를 파악하는데, 이러한 과정에서 빈도만 고려하여 의미를 찾는 내용분석과 달리 특정 단어와의 연관성을 파악하여 구조적 관계를 파악하게 된다.

3. 연구 문제와 연구 방법

3.1 연구 문제

이 연구는 2016년 ‘구글번역’이 출시된 이후 급격하게 증가한 NMT와 관련된 뉴스보도의 의미네트워크를 파악한다. NMT 출시 이후 통번역의 광범위한 수요가 흡수되거나 새롭게 창출되는 과정에서 관련 기사가 어떠한 방식으로 의미를 구성하고 있는지 파악하면 이 새로운 기술의 대중적 수용의 방향성을 가늠할 수 있기 때문이다. 특히 기사의 핵심적 가치를 담고 있는 헤드라인을 분석한다면 의미를 구성하는 핵심 키워드를 보다 분명하게 밝혀낼 수 있을 것으로

기대한다.

의미네트워크에 기반하여 기사 헤드라인을 분석한 연구들은 핵심 키워드를 추출하여 매체가 강조하는 뉴스의 프레임을 분석하거나 핵심 논점을 파악하였다(이완수와 최명일 2020; 정수영과 황경호 2015; 최윤정과 권상희 2014). 본 연구에서도 헤드라인에 사용된 키워드를 추출하여 해당 단어 간 의미네트워크를 분석하는 것을 기본으로 한다. 다만 신기술에 대한 매체의 가치중립적 접근을 고려하여 뉴스 기사를 특정한 유목으로 사전에 구분하거나 가정하지 않은 상태에서 기사에서 파악된 단어들의 관계 자체를 분석한다. 이는 메시지의 의미화 패턴을 분석하는 것으로 추출된 키워드의 상호작용 관계의 맥락 속에서 역동적인 의미를 포착할 수 있을 것이다. 따라서 아래와 같은 연구문제를 설정한다.

- 연구문제 1. NMT 관련 뉴스의 헤드라인에 나타나는 핵심어는 무엇인가?
- 연구문제 2. NMT 관련 뉴스의 핵심어 간의 네트워크는 어떻게 나타나는가?
그리고 핵심어들의 관계는 어떠한 구조적 특징을 보이는가?
- 연구문제 3. NMT 관련 뉴스 네트워크를 구성하는 하위 의미네트워크 구조는 어떠한가?

3.2 분석의 대상

인터넷과 모바일 환경은 미디어 뉴스보도의 유통구조를 변화시키고 있다. 이용자들이 종이와 TV를 벗어나 SNS 등 다양한 플랫폼 환경에서 뉴스를 접할 수 있기 때문에 뉴스의 선택권이 생산자에서 소비자 중심으로 재편되어간다(Lee and Tandoc 2017). 따라서 의미구성분석을 위해서 주요 일간지와 같은 특정 매체로 연구를 제한하는 것은 매체 환경 변화에 적합한 방식이 아니다. 이 연구에서는 중앙일간지 등에 특별한 가중치를 두지 않고, 한국언론진흥재단에서 제공하는 뉴스서비스인 빅카인즈를 통해서 NMT 관련어를 검색하여 뉴스 기사를 수집하였다. 빅카인즈는 종합일간지를 비롯한 경제지, 지역지, 방송사 등의 기사 DB를 빅데이터 분석기술을 접목하여 제공하고 있다.

분석자료의 수집 기간은 ‘구글번역’이 NMT를 처음 출시한 때인 2016년 9

월 1일부터 2022년 6월 30일까지이다. 분석을 위한 뉴스가 수집된 매체는 중앙 일간지, 지역지, 방송사, 전문지 등 총 54개 매체이다²⁾. 연관 검색어는 ‘기계번역’, ‘신경망번역’, ‘ai번역’, ‘구글번역’, ‘파파고’, ‘NMT’ 등 인공지능기계번역을 잘 나타낼 수 있는 핵심적 단어로 설정하였으며, 검색결과로 해당 단어가 뉴스본문에 포함된 기사의 헤드라인 총 2,079건이 수집되었다. 이 가운데 같은 내용이 중복 보도된 기사를 제외한 1,306건의 뉴스 헤드라인을 연구의 분석대상으로 선정하였다. 분석대상 기사의 연도 별 현황은 아래의 <표 2>와 같다³⁾.

<표 2> NMT 관련어의 뉴스보도 양 (BIGKinds)

연도	2016	2017	2018	2019	2020	2021	2022	합계
기사	125	438	210	194	132	147	60	1,306

3.3 분석의 절차

연구를 위해서 수집된 1,306개의 헤드라인을 의미네트워크 분석하기 위한 첫 번째 단계로, 수집된 헤드라인을 구성하는 단어들을 해체하여 핵심어를 선정하는 작업을 진행하였다. 의미네트워크분석에서는 핵심 단어들이 특정한 형태로 결합될 때 특정한 의미를 발생하기 때문에 핵심 단어 간의 의미론적 연관을 밝히기 위해서 핵심어를 구분한다(박한우와 Loet 2004: 1378). 핵심어는 텍스트의 주제적 특성을 가장 잘 나타낸다고 판단되는 단어 중에서 그 자질(features)이 높은 단어로 선택되는데 주로 단어의 출현 빈도를 사용한다(이수상

- 2) (중앙지-경제지-지방지-방송-전문지의 가나다 순) 경향신문, 국민일보, 내일신문, 동아일보, 문화일보, 서울신문, 세계일보, 조선일보, 중앙일보, 한겨레, 한국일보, 매일경제, 머니투데이, 서울경제, 아시아경제, 아주경제, 파이낸셜뉴스, 한국경제, 헤럴드경제, 강원도민일보, 강원일보, 경기일보, 경남도민일보,경남신문,경상일보, 경인일보, 광주매일신문, 광주일보, 국제신문, 대구일보, 대전일보, 매일신문,무등일보, 부산일보, 영남일보, 울산매일, 전남일보, 전북도민일보, 전북일보, 제민일보, 중도일보, 중부매일, 중부일보, 충북일보, 충청일보, 충청투데이, 한라일보, KBS, MBC, OBS, SBS, YTN, 디지털타임스, 전자신문
- 3) 2016년은 9월부터, 2022년은 6월까지 수집한 관계로 해당 연도의 기사를 모두 반영하지는 않는다. 본 연구에서는 연도별 구분이 아닌 전체 기사의 의미네트워크를 분석하는 관계로 연도별 기사량은 이해를 위한 참고자료이다.

2014: 55). 핵심어 선정에는 한국어 텍스트 분석 소프트웨어인 KrKwick를 사용하였다. KrKwick는 네덜란드 암스테르담대학교의 로트 레예데스도르프(Loet Leydesdorff)가 개발한 Full Text 소프트웨어를 한국어 분석을 위해서 변형한 것으로(김유호 2012: 211), 언어 네트워크분석 방법을 활용한 논문에서 단어 추출프로그램으로 가장 많이 사용된다(이수상 2014: 61). KrKwick⁴⁾를 통해 선정된 핵심어는 별도의 정제과정이 필요하다. 관사, 접속사, 문장의 어미 등을 제외하고 유사한 맥락에서 사용된 단어들을 분석에 적합한 형태로 통합하는 작업이다. 예를 들어 ‘네이버로’, ‘네이버의’, ‘네이버웍스’ 등의 단어들은 ‘네이버’로 일원화하고, ‘신기술’, ‘기술력’ 등은 ‘기술’로 통합하는 작업을 거쳐야 한다. AI와 인공지능은 의미상 같은 용어이지만, 언어에 기반한 기사에서는 영어와 한국어 연관어들이 다르게 사용될 수 있어서 구분하여 정제하였다.

두 번째 단계로, 이렇게 수집되고 정제된 핵심어들이 각 헤드라인에서 같이 언급된 횟수를 파악하여 공출현 행렬데이터(co-occurrence matrix)를 구성하게 된다. 언어와 의미를 구성하는 네트워크 분석은 핵심어들이 분석대상이 되는 텍스트와 문장에 얼마나 같이 자주 출현하는가를 통해서 그 관계를 설정하는데, 공동으로 자주 출현할수록 핵심어들이 해당 텍스트 내에서 주제적으로 서로 밀접한 관계를 가진다는 것을 의미한다(강명구 2000: 97). 본 연구에서는 KrKwick에 포함된 KrTitle 프로그램을 이용해 핵심어들이 1,306개의 헤드라인에 공동으로 등장하는 관계의 공출현 행렬데이터를 구성하였다.

셋째로, 헤드라인에 등장한 핵심어들의 공출현 행렬데이터를 기반으로 의미 네트워크분석을 진행하였다. 네트워크 분석에는 노드와 링크의 개수 및 이들 관계의 그래프를 통해 얼마나 밀접한 관계인지를 확인할 수 있다. 이 네트워크는 핵심어들이 헤드라인에 공출현하는 빈도를 기반으로 했기 때문에 핵심어 관계에 방향성은 없으며, 헤드라인에 같이 등장하는 횟수에 따라 핵심어 관계들의 강도를 파악할 수 있다. 또한 네트워크 구조에서 핵심어들이 차지하는 역할을 가늠하기 위해서 중심성 분석(centrality analysis)을 수행하였다. 중심성은 네트워크의 구조적 위치를 통해 노드의 중요성을 나타내는 지표로, 노드에 초점

4) KrKwic 프로그램은 KrKwic, KrTitle, KrText 3개의 하부 프로그램으로 구성되어 있지만, 하부 프로그램인 KrKwic의 윈도우버전은 KrWords이다. 이외에 프로그램의 라이선스와 상세한 이용 방법은 박한우와 Loet(2004)의 논문을 참조하면 된다.

을 맞추기 때문에 네트워크상에서 핵심적 지점을 파악할 수 있게 한다(Luo and Zhang 2015: 280). 네트워크 분석에서 주로 사용되는 중심성은 연결중심성, 근접중심성, 매개중심성의 3가지이지만, 이 연구에서는 연결중심성과 매개중심성 분석으로 핵심어 노드들의 역할을 파악한다.

넷째로, 의미네트워크분석에서는 밝혀진 핵심어들이 어떠한 하위단어들의 집단으로 구성되어 있는지를 파악하면 네트워크의 특징을 설명할 수 있다. 이 연구에서는 구조적 등위성을 구분해내는 CONCOR 분석을 통해서 핵심어들의 하위 연결관계를 파악한다. 구조적 등위성은 A와 B가 직접적으로 연결되어 있지 않더라도 같은 위치의 노드들과 연결되어 있다면 구조적으로 같은 위치, 즉 등위를 차지한다고 간주한다(김용학 2016: 112). 이 방법은 전체 네트워크 핵심어들의 유사한 연결패턴을 군집화하여 도식화해주기 때문에 각각 집단에 대한 의미파악은 물론 전체 연결망의 구조 파악에도 용이하다는 장점이 있다(남경덕 2021: 118). CONCOR⁵⁾ 분석은 키워드 간 공출현 행렬의 피어슨 상관관계를 분석한 다음 유사한 등위로 판명된 하위집단의 블록을 구축하여 블록 간의 관계를 식별하는 방법이다(Wasserman and Faust 1994: 368). 이 연구에서는 CONCOR 분석을 통해 핵심어들이 특정한 하위집단의 블록으로 구분될 수 있는지 파악하고 그 관계를 이해하고자 한다.

이 연구의 네트워크 분석에는 Ucinet 소프트웨어를 사용하고 시각화에는 Netdraw를 활용하였다(Borgatti, Everett and Johnson 2018). Ucinet은 다양한 네트워크 분석 기법을 활용할 수 있는 프로그램으로 지금까지 제시된 모든 지표의 대부분을 다루며 Netdraw라는 시각화 프로그램을 같이 적용할 수 있다는 강점이 있다(김용학 2016: 14).

3.4 분석의 결과

3.4.1 NMT 관련 뉴스의 핵심어 분석

연구문제 1에 해당하는 NMT 관련 뉴스에 나타나는 핵심어 분석을 위해서 KrKwic의 KrWords 프로그램을 이용하여 1,306개의 헤드라인에서 100개의 핵심어를 추출하였다. 뉴스에는 많은 단어가 사용되지만, 뉴스를 구성하는 의미의

5) CONvergence of COrelation의 대문자 부분만을 표시한 용어이다.

해석을 위해서는 핵심적인 단어들이 어느 정도 규모로 형성되는지 파악하는 것이 중요하다. 이 연구에서 핵심어 추출은 단어의 출현 빈도를 기준으로 하였다. 전체 뉴스 헤드라인에서 2회 이상 언급된 단어는 총 1,109개가 있었지만 네트워크분석의 명료함과 연구의 적절성을 위해서 상위 100개의 핵심어를 추출하였다⁶⁾. 추출과정에서 불용어와 어미 등을 삭제하고 용어의 일원화 등의 정제과정을 거쳤다. 결과는 아래 <표 3>과 같다.

<표 3> NMT 관련 뉴스의 핵심어 100

빈도	핵심어 1-33	빈도	핵심어 34-66	빈도	핵심어 67-99
454	네이버	28	웹	18	CES
389	AI	28	영어	18	광운대
276	번역	28	키즈	18	상품
147	기술	28	브라우저	18	음성인식
121	구글	27	데이터	17	대표
102	파파고	26	공개	17	외국어
95	인공지능	26	투자	17	도전
83	서비스	26	모바일	17	MS
70	앱	25	로봇	17	협업
59	카카오	25	활용	17	이슈
56	기업	25	해외	17	전문
54	출시	24	지원	17	여행
54	플랫폼	24	IT	17	자동
53	한국	24	LG	16	기반
50	글로벌	23	시스템	16	대결
49	개발	23	자율주행	16	디지털
48	클라우드	23	진화	16	사용자
43	경쟁	23	성장	16	교육
41	시대	22	세계	15	기능
41	시장	21	확대	15	스마트폰
40	특허	20	국내	15	추가
40	검색	20	외국인	15	4차
39	웨일	19	vs	15	똑똑해진
39	통번역	19	산업혁명	15	현대차
36	스마트	19	스피커	15	제주

6) 핵심어는 네트워크에서 분명한 특징을 나타내는 소수의 단어를 기준으로 하지만, 본 연구는 ‘인공’과 같은 의미 있는 단어의 포함을 위해 100의 상징적 숫자를 선택함.

36	통역	19	콘텐츠	15	중국
34	미래	19	올림픽	15	정보
32	번역기	19	신경망	15	인공
32	인간	18	적용	14	강화
32	연구	18	ICT	14	SK
30	기계번역	18	스타트업	14	산업
29	언어	18	올해	14	역할
28	개최	18	솔루션	14	국제

NMT 관련 뉴스 헤드라인에서 가장 많이 언급된 핵심어는 네이버(454회), AI(389회), 번역(276회), 기술(147회), 구글(121회) 순이다. 산업, 국제, 역할 등의 단어는 약 14회로 가장 적게 언급된 핵심어이다. 핵심어 중에서는 서비스를 먼저 시작한 ‘구글’이 아닌 ‘네이버’가 가장 많이 언급된 것이 특징이다. 이는 뉴스를 수집하기 시작한 NMT 등장 초기인 2016년부터 지속되었는데, 서비스 개시 시기가 구글과 네이버가 한 달 정도로 차이가 크지 않은 이유도 있겠지만, 네이버는 출시 초기부터 한국어 기반의 번역 기술로 사용성을 높여서 일반인은 물론 번역을 전문적으로 공부하는 이들에게서도 네이버 번역에 대한 사용 빈도가 압도적으로 높다는 연구 결과가 보고되고 있는데(강수정 2021: 51) 뉴스 보도에서도 이러한 현상이 반영된 것으로 볼 수 있다.

핵심어들의 또 다른 특징은 가치중립적인 단어들인 많다라는 것이다. 선행연구에서 기술뉴스가 가치중립적으로 다뤄진다는 특징을 보고한 바 있는데, 이때 사용되는 단어들은 사회가 기술을 허용할 수 있는 범위를 보여준다(김만재와 전방욱 2012: 21). 본 연구에서도 매체가 자신들이 강조하는 프레임으로 헤드라인을 구성하기보다는 새로운 번역 현상에 대해 우리 사회가 허용할 수 있는 기술의 범위에 대해 평가하는 과정에서 가치중립적 단어들인 언급되는 것을 확인할 수 있었다. 핵심어의 상위에 네이버, 구글, 파파고, 카카오, 앱 등과 같은 서비스 명칭과 AI, 기술, 인공지능, 플랫폼과 같은 기술용어들이 자리 잡고 있어 서비스와 기술을 강조하는 방향으로 뉴스가 구성되고 있음을 알 수 있었다. 대

7) 구글은 2016년 9월 28일부터 딥러닝을 구글번역에 적용하였고(연합뉴스 2016.9.28.) 네이버는 2016년 10월 21일부터 인공신경망 번역기술을 파파고에 적용하였다(머니투데이 2016.10.21.).

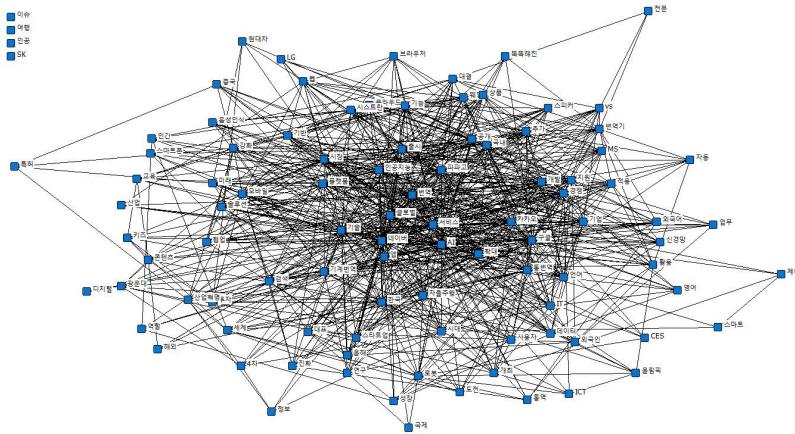
결(16회)이나 vs(19회) 등과 같이 인간과 인공지능의 대결을 암시하는 선정적이고 자극적으로 구성된 핵심어들도 보이지만, 대개는 “인간과 기계의 번역대결, 승자는?”(전자신문 2017.2.21.)과 같이 특정 대학에서 진행된 단발성 이벤트 기사의 헤드라인이나 그에 따른 파생 기사들의 핵심어였고, 실제로 기계와 인간의 대결을 강조하며 번역사의 종말을 선포하는 프레임이 광범위하게 적용된 것으로 보이는 기사는 일반적이지 않았다.

기사에 사용된 기술용어들은 상세기술 혹은 세부 기술의 용어보다는 구체성이 떨어지는 범용적 용어가 많았다. NMT(10회 이하)보다는 AI(389회), 기술(147회), 인공지능(95회)과 같이 일반적인 의미의 기술용어가 핵심어로 파악되었다. 핵심어로만 살펴보면 기술적 이해와 특성이 강조되기보다는 서비스에 대한 소개(파과고 102회, 앱 70회, 카카오 59회, 클라우드 48회)에 초점이 맞춰져 있었다. 비슷한 맥락에서 기업(56회), 출시(54회), 글로벌(50회), 경쟁(43회), 시장(41회) 등의 시장 중심적 용어도 헤드라인에 자주 언급되었다.

3.4.2 NMT 관련 뉴스 의미네트워크와 중심성

뉴스에는 많은 단어가 사용되지만, 뉴스를 구성하는 의미의 해석을 위해서는 핵심적인 단어들의 관계가 중요하다. 단어의 사용 빈도와 관계는 텍스트에서 강조되는 상징성을 보여 줄 수 있고 단어들의 결합을 통해 특정한 의미를 만들어 내는 경향이 있기 때문이다. 예를 들어 ‘청년’이라는 단어가 사용된 문장에서 ‘스마트’가 공동으로 등장하는 것과, ‘실업’이 등장하는 것에서 문장이 주는 의미는 다르게 생성된다. 이 연구에서 연구문제 2의 해결을 위해서 공동으로 나타나는 핵심어들의 관계를 파악하기 위해서 의미네트워크분석을 진행하였다. 아래 <그림 1>은 100개의 핵심어를 노드로 하고 1,306개의 헤드라인에 공출현한 관계를 링크로 하여 핵심어 사이의 전체적 연결관계를 시각화 한 것이다.

〈그림 1〉 핵심어 100개로 구성된 의미네트워크



그래프의 가시성을 위해서 헤드라인에 같이 언급된 정도인 링크의 크기 (value)는 표시하지는 않고 100대 핵심어 간의 관계만을 나타내었다. 이 네트워크에서 핵심어 노드들은 평균 17.34개의 링크를 가지고 있는데, 이는 핵심어들이 평균 17.34개의 다른 단어들과 헤드라인에서 같이 언급된다는 의미이다.

가장 많이 연결된 링크는 네이버와 AI의 59회였고 이어서 번역과 AI가 35회, 번역과 네이버가 34회, 번역과 구글 31회, 번역과 파파고 28회, 번역과 인공지능 16회 순으로 같이 언급되었다. 정리하자면 서비스명과 기술어 그리고 번역이라는 추상명사가 가장 높은 관계로 헤드라인에 등장하고 있었다. 위의 그림에서는 한 번이라도 헤드라인에서 같이 언급된 경우는 링크가 있는 것으로 표현하고 한 번도 같이 언급되지 않았다면 링크가 없는 고립노드로 표현되었다. 이슈, 인공, SK, 여행 등의 단어들은 100대 핵심어들과 같이 언급된 적이 없어 링크가 표시되지 않는다.

네트워크에 대한 이해를 높이기 위해서 100대 핵심어 네트워크에서 헤드라인 공출현 빈도가 30회 이상이 되는 31개 핵심어의 관계를 시각화하여 <그림 2>로 나타내었다. 노드의 크기는 헤드라인에서 언급된 빈도에, 링크의 굵기는 공출현 정도에 각각 비례한다. 즉 언급된 빈도가 높으면 노드가 크게 나타나고, 헤드라인에 같이 언급된 횟수가 많으면 굵은 링크로 표현되는 것이다.

〈표 4〉 NMT 뉴스 핵심어 중심성 분석

핵심어	연결중심성		매개중심성	
	Degree	nDegree	Betweenness	nBetweenness
네이버	614.000	0.105	780.035	16.080
AI	412.000	0.071	740.091	15.256
번역	309.000	0.053	308.525	6.360
기술	135.000	0.023	166.284	3.428
구글	179.000	0.031	168.915	3.482
파파고	164.000	0.028	96.626	1.992
인공지능	117.000	0.020	139.187	2.869
서비스	132.000	0.023	153.514	3.165
앱	63.000	0.011	54.260	1.119
카카오	89.000	0.015	56.031	1.155
기업	30.000	0.005	28.777	0.593
출시	126.000	0.022	58.510	1.206
플랫폼	77.000	0.013	37.597	0.775
한국	41.000	0.007	75.907	1.565
글로벌	93.000	0.016	85.592	1.764
개발	40.000	0.007	18.280	0.377
클라우드	77.000	0.013	11.227	0.231
경쟁	49.000	0.008	35.701	0.736
시대	33.000	0.006	36.053	0.743
시장	58.000	0.010	99.623	2.054
특허	4.000	0.001	2.204	0.045
검색	32.000	0.005	9.357	0.193
웨일	56.000	0.010	11.742	0.242

연결중심성은 네트워크에서 연결망의 중심에 위치한다는 의미로 여러 노드와 직접적으로 많이 연결되면 영향력이 높다. 단순히 연결된 노드의 합이 아니라 네트워크 크기에 따라서 그 값의 상대적 연결 강도를 고려한 활동성 지수를 고려한다(김용학 2016: 120). 네트워크에서는 특정 핵심어가 다른 단어와 함께 자주 등장한다는 점에서 의미구성에 핵심적 역할을 수행한다고 볼 수 있다. <표 4>의 연결중심성을 살펴보면 네이버, AI, 번역 등의 빈도가 높은 핵심어들이 연결중심성이 높게 나와서 이들이 의미구성의 핵심적 키워드임을 알 수 있다. 특징적인 것은 구글, 파파고, 카카오, 출시 등의 구체적 서비스를 지칭하는 핵심어들의 연결중심성이 기사에 언급되는 빈도에 비해 상대적으로 높은 연결

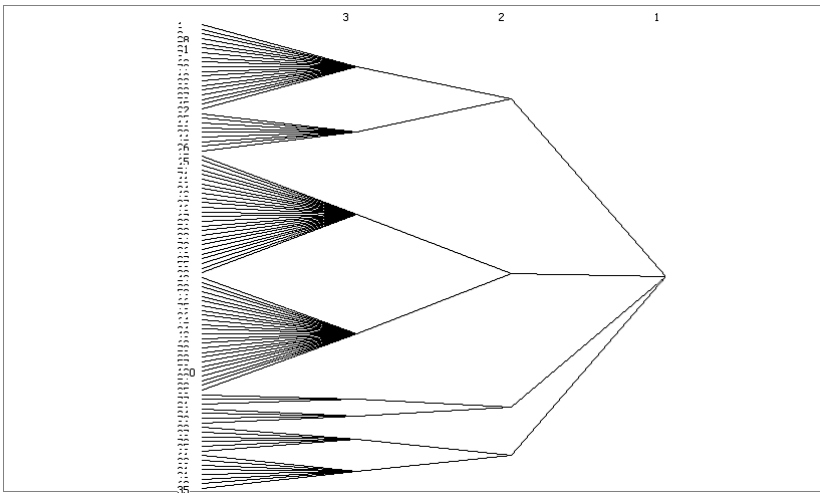
중심성을 보인다. 이는 NMT 관련 뉴스는 구체적인 서비스명을 다른 단어들과 함께 언급한다는 의미로, 서비스를 핵심적 내용으로 하는 기사가 많다는 것이다. 또한 서비스명을 넘어서 특정한 기술적 카테고리를 의미하는 핵심어들의 연결중심성 값이 상대적으로 높았다. 플랫폼, 클라우드 등이 이에 해당하는데, 이는 뉴스들이 기계번역을 단순한 번역기나 앱으로 한정하기보다는 다양한 기술에 확장적 적용을 고려하는 비중이 높았음을 유추할 수 있다.

매개중심성은 특정 노드가 다른 노드들 사이에 위치하는 정도로 표현할 수 있다(오창우 2017: 56). 전체 네트워크에서 노드들의 관계를 통제하거나 중개하는 위치를 말하는데, 다른 단어들이 매개중심성이 높은 단어를 통하지 않으면 서로 단절되어 의미를 잃어버릴 수도 있다. 핵심어들의 경우 헤드라인에 노출된 빈도가 낮더라도 매개중심성이 높은 단어를 제외하게 되면 문장의 원활한 소통이 어려울 수도 있다(박지영 외 2013: 49). 매개중심성은 연결중심성에 비해 중심성 값의 편차가 크기 때문에 핵심적 역할의 노드를 구분하기 용이하다. NMT 관련 뉴스 헤드라인 네트워크에서 다른 핵심어를 통제하거나 중개하는 핵심적인 단어는 연결중심성과 마찬가지로 네이버, AI, 번역 등이다. 이들 단어가 빠질 경우 헤드라인의 의미구성이 달라질 수 있다. 매개중심성 분석의 또 다른 특징은 한국, 글로벌, 시장과 같은 핵심어들이 상대적으로 강한 매개중심성 값을 보여주고 있다는 것이다. 이는 기계번역 시장과 그 영향력의 범위에 관한 기사들이 의미네트워크에서 상당한 통제력을 발휘하고 있다는 것으로 다른 단어들과 함께 헤드라인에 등장하여 문장의 흐름을 주도하고 있음을 알 수 있다. 매개중심성 분석에서는 시장 관련 단어들이 낮은 빈도를 보이지만 헤드라인 의미구성의 중심적 역할을 수행하고 있음을 알 수 있었다.

3.4.3 NMT 관련 뉴스 하위 네트워크 구조

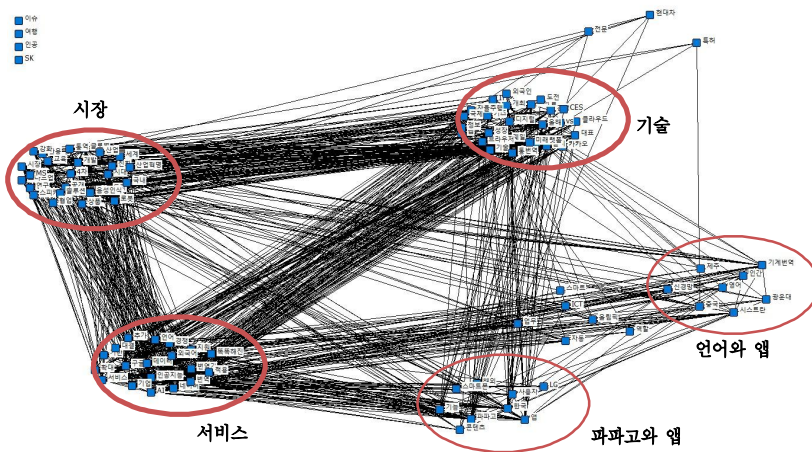
네트워크 분석의 상관관계 알고리즘에 기반해 의미네트워크 구조를 블록화하는 CONCOR 분석으로 유사한 연결패턴을 보이는 핵심어들의 하위그룹을 구분하였다. 적절한 하위그룹의 수는 군집화 방식의 덴드로그램(dendrogram)을 통해 결정하고, 이에 기반하여 <그림 3>과 같이 Level 3의 수준에서 총 8개의 하위그룹을 분류하였다.

<그림 3> 주요 핵심어 100개의 하위 네트워크 덴드로그램



덴드로그램에 따라 구분된 8개의 하위언어집단을 군집화하여 CONCOR 분석을 시각화한 결과는 아래 <그림 4>와 같다.

<그림 4> 주요 핵심어 100개의 CONCOR 분석



제시된 시각화 분석 결과, 고립 노드인 {이슈, 여행, 인공, SK}의 단어 집단과 노드 연결이 약한 {전문, 특허, 현대차}의 군집집단을 제외하면 모두 6개의 하위집단을 가려낼 수 있다. 각 집단을 구성하는 노드 중에서 중심성 값이 높은 단어들로 하위집단을 구분하고 그 특징을 설명하면 아래 <표 5>와 같다.

<표 5> CONCOR 분석에 따른 하위 핵심어 군집

군집명	핵심어
서비스	네이버, AI, 번역, 인공지능, 구글, 기업, 서비스
시장	글로벌, 시장, 음성인식, 통역, 투자, 개발
기술	IT, CES, 클라우드, 웨일, 자율주행, 출시, 기술
파파고와 앱	스마트폰, 파파고, 해외, 한국, 앱
언어와 앱	신경망, 영어, 중국, 인간, 시스템
스마트, 올림픽	스마트올림픽: ICT, 올림픽, 스마트,

이들 하위군집은 노드 간 연결패턴의 상관관계가 높은 집단으로 구분한 것으로 <그림 4>에 따르면 서비스, 시장, 기술의 3가지 군집집단들이 크기가 가장 크고 서로 가장 높은 연결 관계를 보여준다. 상대적으로 연결 노드들의 집합은 적지만 의미 있는 하위집단은 번역언어, 파파고앱 등의 연결이다. 헤드라인의 의미네트워크는 서비스, 시장, 기술, 파파고, 언어앱 등의 의미를 지닌 기사들의 하위 네트워크 집단으로 구성된다고 할 수 있다.

4. 논의 및 결론

본 연구는 ‘구글번역’, ‘파파고’ 등과 같은 NMT가 우리의 일상생활 속에 빠르게 자리 잡은 현상에 주목하였다. 선행연구에 따르면, 이 새로운 번역기술에 대해 전문 번역사들마저도 그 품질을 높이 평가하고 있으며 이를 도구적으로 활용하는 등 적극적인 수용의 양태를 보인다. 본 연구는 품질평가의 단계를 넘어 빠르게 확산되고 있는 기계번역의 긍정적 수용양상에 미디어가 어떠한 역할을 수행하는지, 관련 뉴스기사에 나타난 핵심적 가치와 방향이 무엇인지에 대한 사회적 맥락을 파악하고자 하였다. 이를 위해 2016년 9월부터 2022년 6월

까지 NMT 관련 뉴스 1,306개의 헤드라인에 나타난 핵심어들의 관계를 의미네트워크 분석하였고 뉴스가 제시하는 핵심적 가치를 이해하였다.

그 결과 첫째, 헤드라인에 나타난 핵심적 개념들은 대부분 ‘가치중립적’ 단어들이었다. 네이버(454회), AI(389회), 번역(276회), 기술(147회), 구글(121회) 등의 빈도순으로 헤드라인에 나타난 핵심어들은 특정한 가치나 방향성을 내포하고 있지 않다. 이 단어들이 주로 연결된 다른 단어들 역시 ‘네이버’는 AI, 기술, 서비스, 번역 등의 핵심어와, ‘AI’는 기술, 구글, 파파고와, ‘번역’은 구글, 네이버, 파파고, AI 등의 단어들과 연결되었다. 뉴스의 프레임과 관련된 선행연구들은 뉴스가 강조하고자 하는 단어들을 선택해서 현저하게 부각하는 방식을 통해 미디어와 구성원들의 지배적 가치를 재구성하여 뉴스로 제시한다고 밝히고 있다. 기술과 관련된 보도 역시 프레임의 맥락에서 파악된 연구도 있지만, 특정 사회적 이슈와 연관되지 않을 경우에는 가치중립적인 기사 구성을 통해 사회적 수용의 범위를 제시한다. NMT 관련 뉴스들은 가치중립적인 접근으로 사회적 수용의 범위를 합의하고 조정하는 과정에 있음을 파악할 수 있었다.

둘째, NMT 관련 뉴스들은 번역기, 번역앱 혹은 기계번역의 단순 서비스의 일차적 기술들을 소개하고 있지만, 중요하게는 좀 더 확장적이고 범용적인 카테고리의 기술적 차원으로 기계번역의 확장을 기대하고 있었다. 중심성분석을 통해서 클라우드, 플랫폼 등의 하위기술을 포함하는 상위차원의 기술 단어들이 서비스와 기술 등 다른 단어를 매개하는 역할을 하고 있음을 알 수 있었다. 다시 말해 네이버, AI, 번역, 구글 등이 다른 단어들과 연결될 때 플랫폼, 클라우드 등이 주로 같이 언급된다는 의미로, 이는 NMT가 좀 더 광범위한 기술적 영역으로 확장 성장이 예상되는 지점이며 동시에 보다 다양한 일상의 영역에서 NMT가 적용될 것임을 예측할 수 있다.

셋째, 뉴스기사의 하위네트워크는 기술과 시장, 구체적 서비스 간의 높은 상관수렴(CONCOR)을 보여주고 있다. 이는 NMT의 기술, 시장현황과 전망, 구체적인 서비스 명칭 등에 관한 개념들의 상호관계를 통해 기사가 구성되었음을 뜻하며 기사의 내용이 가치중립적임을 의미한다. 인간과 기계의 대결이나 통번역 직종의 종말 등과 같은 자극적 프레임은 거의 찾아볼 수 없었고 특정 기관의 일회성 이벤트에 지나지 않았다. 뉴스에 드러난 보다 광범위하고 중요한 이슈는 기술의 발전과 시장의 확장 그리고 보다 다양한 서비스와 관련이 있었다.

이 같은 결과는 NMT에 대한 미디어와 대중의 수용방식의 단면을 보여준다. NMT에 대해 번역사들은 이를 도구적으로 활용해 업무의 완성도를 높이고자 하고 번역학 연구자들은 이러한 변화와 수요에 대응하기 위해 번역교육의 패러다임 전환을 적극적으로 고민하는 반면, 미디어에 투영된 이 기술은 이미 언어를 넘어 생활 전반으로 퍼져나가고 기술의 발전과 시장확장을 통한 산업적 가치를 고려하고 있었다. 이러한 차이는 번역학계에 인간번역을 대체해가고 있는 기계번역에 대한 소극적 협업이 아닌 산업적 측면에서 기술과 사회의 변화를 바라보는 것이 중요하다는 점을 시사한다. 포스트에디팅 교육의 도입과 같은 협업모델이 현실적 방안으로 제시되고 있지만, 기계번역의 완성도가 부단히 높아질 것으로 예상되는 만큼 산업적 측면에 맞춘 적극적인 교육모델의 개발이 필요하다. 기계와 인간의 협업 현황과 실질적 수요를 파악하고 이를 기반으로 학부와 전문대학원 과정을 나누어 각기 적용 가능한 초, 중, 고급 과정의 번역 교육 모델을 마련해야 한다. 또한 협업교육을 수료한 실무자들의 권익을 보호할 수 있는 메커니즘 확립을 위한 제도적 측면의 연구도 시급해 보인다.

본 연구는 NMT 관련 뉴스에 대해 의미네트워크분석을 진행하여 기술에 대한 가치중립적 접근이 이루어지고 있으며 이들 뉴스가 기술, 시장, 서비스의 하위 의미집단으로 구성되어 있음을 밝혔음에 학술적 의의가 있다. 하지만 연구의 대상을 뉴스의 헤드라인으로 제한한 것은 한계점이다. 헤드라인이 실제 기사의 가치를 반영한다는 점에서 이 연구의 결과가 뉴스 자체의 반영임에는 논리적 의의가 없겠지만, 보다 다양한 핵심어의 추출을 위해서 뉴스 본문에 대한 연구의 필요성이 제기될 수도 있다. 이를 위해서는 핵심어의 빈도추출을 넘어서 텍스트마이닝의 형태소 분석 알고리즘의 방법을 통해 완전한 기계적 방법으로 기사 전체를 분석하거나 나아가 댓글 분석을 통해 이용자들의 시각을 확인하는 것 역시 향후 연구의 과제가 될 것이다. 또한 본 연구는 빅데이터 연구에 가까운 양적 방법을 채택하여 기사가 주는 정성적 효과를 간과했을 가능성이 있다. 인간과 기계의 대결과 같은 이벤트성 기사가 양적으로는 극소수에 불과하지만, 그것이 기계번역에 대한 인간의 인지에 어느 정도 깊이 있는 영향을 주었을지는 이 연구 방법으로는 파악하기 힘들다. 따라서 해당 기사에 대한 실험 연구 혹은 인터뷰와 같은 질적인 방법을 통해서 그 영향력을 파악할 후속 연구의 필요성도 고려해야 한다. 하지만 1,306개의 헤드라인을 통해서 나타난 것은

인간과 기계의 대결이 아닌 기술의 확장성에 보다 집중된 관심과 플랫폼, 시장, 서비스로서의 가치에 초점이 맞춰져 있었음은 분명하다.

참고문헌

- 강명구 (2000) 「정치뉴스에 나타난 한국 정치권력구조의 네트워크 분석」, 『언론정보연구』 37: 93-129.
- 강수정 (2020) 「전문번역사들의 NMT에 대한 인식과 수용에 관한 연구 - 심층인터뷰를 중심으로」, 『번역학연구』 21(3): 9-35.
- 강수정 (2021) 「통번역대학원생들의 기계번역 사용현황과 인식에 관한 연구」, 『번역학연구』 22: 41-63.
- 김만재, 전방욱 (2012) 「언어네트워크 분석 기법을 활용한 인간배아복제 신문보도 분석」, 『생명윤리』 13(2): 19-34.
- 김문수, 김순양 (2018) 「사회연결망분석을 적용한 영남권신공항 건설 논쟁과정의 동태성 분석: 이명박 정부 하의 제2단계 논쟁을 중심으로」, 『국정관리연구』 13: 159-189.
- 김순미 (2017) 「신경망번역기(NMT) 활용 학부 번역교육의 가능성 연구」, 『통번역교육연구』 15(3): 5-87.
- 김순미 (2018) 「AI시대 인간번역과 기계(NMT)번역의 공존-경영학 ‘확장(Augmentation)전략’ 중심」, 『통역과 번역』 20(2): 1-32.
- 김용학 (2016) 『사회 연결망 분석』(4판), 서울: 박영사.
- 김용학, 윤정로, 조혜선, 김영진 (2007) 「과학기술 공동연구의 연결망 구조」, 『한국사회학』 41(4): 68-103.
- 김유호 (2012) 「의료민영화 논의에 따른 이슈용어의 연결 중심성 분석」, 『한국콘텐츠학회논문지』 12(8): 207-214.
- 김준석 (2017) 「일상생활 속으로 들어온 기계번역」, 『새국어생활』 27(4): 63-79.
- 남경덕 (2021) 「미디어와 국민청원 사이트에서 공중의제를 다루는 방식을 통해 제시하는 3단계 역의제 설정 이론 : 의미연결망분석(SNA) 및 연결망군 집분석(CONCOR)을 통해」, 한국방송학회 학술대회 논문집.

- 마승혜 (2018) 「문학작품 기계번역의 한계에 대한 상세 고찰」, 『통번역학연구』 22(3): 65-88.
- 문안나 (2016) 「한국 사회의 IR(Investor Relations) 개념: 1994~2014년 신문기사 내용분석 및 의미연결망 분석」, 『홍보학연구』 20(1): 50-78.
- 박지영, 김태호, 박한우 (2013) 「의미연결망 분석을 통한 셀러브리티의 SNS 메시지 탐구」, 『방송통신연구』 봄호: 36-74.
- 박한우와 Loet Leydesdorff (2004) 「한국어의 내용분석을 위한 KrKwic 프로그램의 이해와 적용: Daum.net에서 제공된 지역혁신에 관한 뉴스를 대상으로」, 『Journal of The Korean Data Analysis Society』 6: 1377-1387.
- 반현과 Maxwell E. McCombs (2007) 「의제설정 이론의 재고찰 - 5단계 진화 모델을 중심으로」, 『커뮤니케이션 이론』 3(2): 7-53.
- 송상현 (2020) 「기계 번역에 대한 몇 가지 오해」, 『번역·언어·기술』 1: 5-23.
- 신지선 (2017) 「테크놀로지 패러다임에서의 번역능력 재조명」, 『통번역학연구』 21(4): 51-71.
- 오창우 (2017) 「한국에서의 사회갈등 논의의 의미연결망 분석: 주요 포털에서의 핵심어간 네트워크를 중심으로」, 『정치커뮤니케이션 연구』 45: 37-67.
- 이수상 (2014) 「언어 네트워크 분석 방법을 활용한 학술논문의 내용분석」, 『정보관리학회지』 31(4): 49-68.
- 이승일 (2018) 「AI 번역의 속성 및 번역주체에 대한 논의」, 『통번역학연구』 22(4): 183-209.
- 이완수, 최명일 (2019) 「뉴스 언어의 선정성과 언론의 상업성: ‘김학의 성접대 사건’ 보도에 대한 의미연결망 분석을 통해」, 『미디어, 젠더 & 문화』 34(3): 105-156.
- 이완수, 최명일 (2020) 「범죄뉴스 헤드라인 언어와 의미구성: 포털사이트 ‘네이버’에 보도된 범죄 유형별 의미연결망 분석을 통해」, 『한국언론정보학보』 101: 367-398.
- 이준웅 (2005) 「갈등적 사안에 대한 여론 변화를 설명하기 위한 프레임링 모형 검증 연구」, 『한국언론학보』 49(1): 133-161.
- 이준호 (2015) 「헤드라인 저널리즘」, 『지역과 커뮤니케이션』 19: 249-280.
- 이창수 (2019) 「문학번역에서의 기계번역과 인간번역 문체에 대한 전산문체학

- 적 비교 연구, 『번역학연구』 20(2): 111-130.
- 이하자, 안동미 (2011) 「한·일 기계번역시스템의 문제점 고찰 - 번역가 번역문과의 대조분석을 통해」, 『일본어교육』 58: 151-165.
- 이형민, 박진우 (2020) 「언택트 시대의 넷플릭스: 넷플릭스 보도에 대한 언어네트워크 분석」, 한국방송학회 학술대회 논문집.
- 임종수, 신민주, 문훈복, 윤주미, 정태영, 이연주, 유승현 (2017) 「AI 로봇 의인화 연구」, 『한국언론학보』 61: 113-143.
- 장애리 (2017) 「국내 기계 통번역의 발전 현황 분석: 한중언어쌍을 중심으로」, 『번역학연구』 18(2): 171-206.
- 장한수, 조성경, 홍정석 (2020) 「미래에너지의 사회적 수용성에 대한 행동경제학적 접근: 프레임 효과에 대한 네트워크분석 사례」, 한국기술혁신학회 학술대회 논문집.
- 정수영, 황경호 (2015) 「한·일 주요 일간지의 한류 관련 뉴스 프레임과 국가 이미지: 기사 헤드라인에 대한 의미연결망 분석을 중심으로」, 『한국언론학보』 59(3): 300-334.
- 차현주, 전재민, 권상희 (2020) 「미세먼지 이슈에 대한 미디어 간 네트워크 의제설정 연구: 언론사 보도와 네이버 블로그 간 네트워크 의제설정 중심으로」, 『사이버커뮤니케이션학보』 37(4): 5-52.
- 천종성 (2020) 「전문번역사들의 기계번역 수용에 관한 연구」, 『한국융합학회논문지』 11(6): 281-288.
- 최문선 (2019) 「국내 번역학 기계번역 연구 동향 : 내용 분석과 키워드 분석을 중심으로」, 『언어학연구』 24(1): 275-297.
- 최운정, 권상희 (2014) 「'빅데이터' 관련 신문기사의 의미연결망 분석」, 『사이버커뮤니케이션학보』 31(1): 241-285.
- 한관중 (2003) 「사회과학 방법론으로서의 연결망 분석기법 적용의 의의와 연구 과제: 의미와 연결망 분석(semantic network analysis)을 중심으로」, 『사회과학교육연구』 10: 219-235.
- Borgatti, Stephen P., Martin G. Everett and Jeffrey C. Johnson (2018) *Analyzing Social Networks*, London: Sage.
- Conboy, Martin (2007) *The Language of the News*, London: Routledge.

- Entman, Robert M. (1993) 'Framing: Toward Clarification of a Fractured Paradigm', *Journal of Communication* 43(4): 51-58.
- Lee, Eun-Ju and Edson C. Tandoc, Jr. (2017) 'When News Meets the Audience: How Audience Feedback Online Affects News Production and Consumption', *Human Communication Research* 43(4): 436-449.
- Luo, Qiuju and Dixi Zhong (2015) 'Using Social Network Analysis to Explain Communication Characteristics of Travel-related Electronic Word-of-mouth on Social Networking Sites', *Tourism Management* 46: 274-282.
- Newell, Jay, Ulrike Genschel and Ni Zhang (2015) 'Media Discontinuance: Modeling the Diffusion "S" Curve to Declines in Media Use', *Journal of Media Business Studies* 11(4): 27-50.
- Rogers, Everett M. (2010) *Diffusion of Innovations*, Hohenheim: Simon and Schuster.
- van Dijk, Teun A. (2013) *News as Discourse*, London: Routledge.

[Abstract]

**A Study on the Semantic Structure of News Articles on
‘Neural Network Machine Translation’:
Focusing on the Semantic Network Analysis for Headlines**

Jong-Sung Chun* & Su-Jung Kang**
(Hanyang University* & Sookmyung Women’s University**)

This study delves into how artificial neural network machine translation is making its way fast into our language life. The study analyzes the meaning of the news headlines, focusing on how the media helps the rapid spread of AI translation. 1,306 news cases reported from September 2016 to June 2022 were collected for the research and were later refined through a news big data providing site named BIGKinds. The refined data were then put through morphological analysis to generate a co-occurrence matrix before conducting semantic network analysis. As a result of the analyses, the keywords in those news articles turned out to be linked to value-neutral words, consisting of sub-concepts that can be divided into markets, technologies, and services. It is noteworthy that general news articles are restructured or framed by the media and gatekeepers, whereas news related to machine translation remain value-neutral. In conclusion, this study recommends that translation researchers and practitioners have a market and industry-oriented cooperation model.

Keywords: AI translation, neural network machine translation, semantic network analysis, language network, semantic network, news analysis

주제어: AI번역, NMT, 의미네트워크분석, 언어연결망, 의미연결망, 뉴스분석

천종성(1저자, <https://orcid.org/0000-0003-0988-0035>)

한양대학교 정보사회미디어학과 겸임교수

cjs1006@hanyang.ac.kr

관심분야: 미디어산업, 플랫폼커뮤니케이션, 네트워크 효과, 네트워크분석

강수정(교신저자)

숙명여자대학교 중어중문학부 부교수

xiaojiang@hanmail.net

관심분야: 미디어번역, 번역교육, 학부통번역교육, 기계번역

논문 투고: 2022년 9월 3일

1차 심사 완료: 2022년 9월 27일

2차 심사 완료: 2022년 10월 11일

게재 확정: 2022년 10월 21일