

Machine-Learning and Corpus-Based Analyses of Korean Nouns from Subjects and Objects: Stimuli Development of Verb Treatment for Neurogenic Communication Disorders

Chaewon Park^a, Yoonseob Lim^b, Jee Eun Sung^a

^aDepartment of Communication Disorders, Ewha Womans University, Seoul, Korea

^bCenter for Intelligent and Interactive Robotics, Korea Institute of Science and Technology, Seoul, Korea

Correspondence: Jee Eun Sung, PhD
Department of Communication Disorders,
Ewha Womans University, 52 Ewhayeodae-gil,
Seodamun-gu, Seoul 03760, Korea
Tel: +82-2-3277-2208
Fax: +82-2-3277-2122
E-mail: jeesung@ewha.ac.kr

Received: October 5, 2019
Revised: November 11, 2019
Accepted: November 11, 2019

This work was supported by the Technology Innovation Program-Industrialized Technology Innovation Project funded By the Ministry of Trade, Industry & Energy (MOTIE) of Korea (No. 10077553, Development of Social Robot Intelligence for Social Human-Robot Interaction of Service Robots).

This article is based on the master's thesis of the first author.

Objectives: Subjects and objects associated with treatment verbs are common stimuli in verb treatment using the argument structure for neurogenic patients. This study investigated Korean nouns from subjects and objects associated with the target verbs in the corpus to suggest the characteristics of subjects and objects for developing Korean-specific verb treatment stimuli. In addition, we shared raw data through cloud so that anyone can use the data for clinical or academic purposes. **Methods:** We used Korean textbook corpus to investigate the differences between subjects and objects in terms of frequency, number of type, and machine-learning based semantic distance to the target verbs. We also examined how machine-learning based semantic distance is correlated with behavioral rating semantic distance. **Results:** Subjects significantly showed less frequency, less number of types, and farther machine-learning based distance to the target verbs than objects did. Machine-learning based semantic distance was strongly correlated with behavioral rating. **Conclusion:** The results demonstrated strong evidence of the subject ellipsis phenomenon in Korean. The weak semantic relation, as proven by the machine-learning based semantic distance, indicated that subjects as verb stimuli for Korean-speaking neurogenic patients need to be reconsidered. We confirmed the possibility of machine-learning based semantic distance as a criterion in selecting the treatment stimuli, but more detailed verification is required for future studies.

Keywords: Corpus analysis, Machine learning, Subject ellipsis, Semantic distance, Verb treatment

신경언어장애군은 전반적으로 어휘 인출(lexical retrieval)에 어려움을 겪는다(Gillam, Marquardt, & Martin, 2011). 어휘 인출과 관련하여 다수의 실어증 환자군 연구에서는 실어증 유형에 따라 동사와 명사 인출의 해리(dissociation) 현상이 나타남을 보고하였다(Chen & Bates, 1998; Miceli, Silveri, Villa, & Caramazza, 1984; Zingeser & Berndt, 1990). 즉, 유창성 환자군은 동사보다 명사 인출을, 비유창성 실어증 환자군은 명사보다 동사 인출을 더 어려워한다는 것이다. 그러나 유창성 실어증 환자군 중 일부 베르니케실어증 환자들뿐만 아니라 정상 노년층도 동사 인출에 어려움을 겪는

다는 결과가 보고되면서 최근에는 명사보다 동사의 인출이 더 어렵다는 주장에 무게가 실리고 있다(Basso, Razzano, Fraglioni, & Zanolio, 1990; De Bleser & Kauschke, 2003; Luzzatti et al., 2002; Nicholas, Obler, Albert, & Goodglass, 1985).

동사 인출이 어려운 원인 중 하나로 동사가 명사 인출에 비해 더 많은 구문론적(syntactic) 처리를 요한다는 점을 들 수 있다(Kim & Thompson, 2000). 명사와 달리 동사는 특정 문장 성분(예: 주어, 목적어, 관형어, 부사어 등)과 함께 쓰여야지만 의미가 제대로 전달된다(Nam, 2004; Nam & Ko, 1993). 이때 동사가 필요로 하는 문장

성분을 동사의 ‘논항(argument)’이라고 하는데, 몇 개의 혹은 어떤 유형의 논항이 필요한지는 동사마다 다르다(Thompson, Lange, Schneider, & Shapiro, 1997). 동사의 인출에는 이와 같은 논항정보가 함께 인출되어야 하기 때문에 신경언어장애군의 동사 인출 중재법은 동사의 논항구조와 연관되어 다양하게 연구되어 왔다(Thompson & Shapiro, 2005, 2007; Thompson, Shapiro, & Roberts, 1993; Thompson, Riley, den Ouden, Meltzer-Asscher, & Lukic, 2013). 이 중에서도 특히 동사의 논항과 의미역(thematic role)을 중재 자극으로 활용하는 동사의미역강화중재(verb network strengthening)는 치료의 일반화 효과가 큰 것으로 알려져있다(Edmonds & Babb, 2011; Edmonds, Mammino, & Ojeda, 2014; Edmonds, Nadeau, & Kiran, 2009). 동사의 의미역이란 논항과 동사가 맺는 의미적 관계(semantic relation)를 일컫는다(Nam, 2007). 예컨대 ‘나는 밥을 먹는다’라는 문장에서 주어 ‘나’는 밥을 먹는 행동을 하는 ‘행동주(agent)’ 의미역이며, 목적어 ‘밥’은 행동주인 내가 먹는 ‘대상(theme)’ 의미역이다. 동사의미역강화중재에서는 주어와 목적어 두 개의 논항을 필요로 하는 2항 동사(2-place verb)를 사용한다. 그리고 그 동사의 주어 위치에 올 수 있는 주격 의미역 중 행동주와 목적어 위치에 올 수 있는 목적격 의미역 중 대상을 훈련하여 동사의 논항구조에 대한 인출을 촉진한다.

동사의미역강화중재는 한국어를 사용하는 소수의 실어증 환자군에게도 실행되어 중재 및 일반화 효과가 나타난 바 있다(Kwag, Sung, Kim, & Cheon, 2014). 한국어에서 동사는 대부분 문장의 맨 마지막에 위치하여 전체 문장의 의미를 결정하는 데 중요한 영향을 미치므로 중재의 필요성이 큰 품사이다(Son, 2001). 따라서 한국어 사용 실어증 환자군을 대상으로 한 동사의미역강화중재에서 중재 및 일반화 효과가 나타난 점은 매우 고무적이다. 그러나 동사 중재의 효과를 극대화하기 위해서는 한국어의 언어적 특수성이 고려되어야 하는데 이와 관련된 기초연구는 매우 부족한 실정이다.

본 연구에서는 신경언어장애군의 동사 중재에 쓰일 수 있는 의미역 명사를 한국어 특징과 관련하여 분석하고자 한다. 이를 위해 우선 한국어의 특징인 주어생략현상(subject ellipsis)이 빅데이터 기반에서 확인되는지 살펴보고자 한다. 또한 동사와 의미역 간의 의미거리를 기계학습(machine learning) 기반으로 산출하고 그에 관한 기초자료를 클라우드에 공유하여 언어치료에 활용할 수 있도록 원자료를 배포하는 것에 목적이 있다.

본 연구의 첫 번째 목적은 한국어의 주어생략현상을 빅데이터 기반으로 살펴보는 것이다. 한국어의 특징 중 특히 주어생략현상이 동사의 논항구조에 미치는 영향에 대해 주시할 필요가 있다. 영어권 비유창성 실어증 환자군의 동사 산출은 동사의 논항수(num-

ber of arguments)와 관련되어 있는 것으로 나타났다. 즉, 영어권 비유창성 실어증 환자군은 논항수를 더 많이 필요로 하는 동사일수록 산출의 어려움을 보였다(De Bleser & Kauschke, 2003; Jonkers & Bastianse, 1996; Kemmerer & Tranel, 2000; Kim & Thompson, 2000). 반면 한국어 사용 경도인지장애군이나 실어증 환자군은 1항 동사와 3항 동사 간 산출에서만 유의한 차이를 보였다(Choi, Sung, Jeong, & Kwag, 2013; Sung, 2016). Sung (2016)은 그 원인 중 하나로 한국어 화자들이 2항 혹은 3항 문장에서 주어를 빈번히 생략한다는 점을 거론했다. 즉, 한국어에서는 2항 동사가 주어 생략으로 인해 1항 동사처럼 기능하게 되어 1항 동사와 2항 동사의 산출 차이가 유의하지 않을 수 있다는 것이다. 한국어의 주어생략현상은 구어(spoken language) 및 문어(written language) 자료 분석에서 모두 밝혀진 바 있다(Kim, 2016; Kim & Choi, 2013; Lee, 2014; Park, 2012). 구어 연구의 경우 일상 대화를 분석하였다(Lee, 2014). 그러나 형식이 엄격하지 않은 구어의 특성상 주어생략현상이 빈번하게 일어났을 가능성이 있다. 문어 연구의 경우 소설, 사설, 영화 시나리오와 같이 제한된 장르에서만 주어생략현상이 밝혀졌다(Kim, 2016; Kim & Choi, 2013; Park, 2012). 이에 본 연구는 기존 연구와 비교하였을 때 매우 방대한 양일뿐만 아니라 다양한 글 종류를 포함하며, 아주 정제된 표현을 담고 있는 텍스트인 교과서에서도 주어탈락현상이 발생하는지 확인하고자 한다.

본 연구의 두 번째 목표는 교과서 말뭉치에서 동사와 주격 명사 간 의미거리 및 목적격 명사 간 의미거리를 기계학습으로 산출하고 그 특성에 대해 확인하는 것이다. 더불어 분석에 사용된 약 천 여개의 주격 명사와 목적격 명사의 목록은 언어치료사들이 중재자극 선정 과정에서 활용할 수 있도록 클라우드 기반(구글드라이브)으로 원자료를 배포하고자 한다. 기계학습이란 기계가 새로운 지식과 기술을 습득해나가는 과정을 통칭한다(International Organization for Standardization, 2015). 최근에는 기계학습을 통해 어휘를 숫자열 벡터(vector)로 나타내는 워드 임베딩(word embedding) 기술의 약진으로 어휘 간 의미 관계에 대한 정량적 파악이 가능해졌다(Mikolov, Sutskever, Chen, Corrado, & Dean, 2013a; Mikolov, Chen, Corrado, & Dean, 2013b). 이에 본 연구는 동사로부터 주격 및 목적격 명사까지의 의미 관계를 수치로 정량화(quantification)하여 의미거리 분포의 차이를 분석하고자 한다. 기계학습에 기반한 의미거리는 최근 가장 주목을 받고 있는 단어 임베딩 기술 중 하나인 Word2vec에 기반하여 구하고자 한다. Word2vec으로 임베딩한 단어 벡터들이 단어의 의미를 대변하고 있는지에 대한 검증은 여러 연구에서 시도되었다(Mikolov, Le, & Sutskever, 2013; Mikolov et al., 2013a; Mikolov, Yih, & Zweig, 2013d; Ororbia, Mikolov,

& Reitter, 2017). 예를 들어, Mikolov 등(2013b)은 동일한 의미적 관계(semantic relation)를 가지는 여러 단어쌍들(예: ‘남자-여자’라는 의미적 관계를 가지는 ‘아들-딸’ 혹은 ‘손자-손녀’)을 얼마나 정확하게 유추할 수 있는지를 측정하였다. 이러한 연구들은 사람이 정의한 단어 간 의미 관계를 Word2vec이 학습한다는 점을 밝혔는데 의의가 있다. 그러나 의미거리를 치료에 활용하기 위해서는 한 단계 더 나아가 행동적(behavioral) 증거가 반드시 필요하나, 이와 관련된 연구는 찾아보기 어렵다. 따라서 본 연구에서는 기계학습으로 산출한 의미거리가 실제 한국어 화자들이 인식하고 있는 의미거리와 어떠한 상관관계를 보이는지 또한 검증하고자 한다.

요약하면, 본 연구의 목적은 한국어를 사용하는 신경언어장애군을 위한 동사 중재에 적용할 수 있도록 한국어 특성에 대한 기초 연구를 시행하는 것이다. 보다 구체적으로는 치료 목표 동사 선정 후 이와 관련된 목표 동사의 주격 명사와 목적격 명사의 특징을 밝히고, 이에 대한 기초자료를 제시하는 것이다. 본 연구의 분석대상은 2항 동사의 주격 명사 중 의미역이 행동주인 명사, 그리고 목적격 명사 중 의미역이 대상인 명사이다. 단, 주격 조사 ‘이/가’나 목적격 조사 ‘을/를’이 명시적으로 붙은 명사로만 한정하였다. 본 연구의 연구질문은 아래와 같다.

첫째, 교과서 말뭉치에서 목표 동사와 결합하는 주격 및 목적격 명사의 평균 출현빈도수(frequency) 및 유형수(type)의 차이는 어떠한가?

둘째, 교과서 말뭉치에서 목표 동사로부터 주격 및 목적격 명사까지의 평균 의미거리(semantic distance)를 기계학습으로 산출하였을 때 그 분포의 차이는 어떠한가?

셋째, 기계학습으로 산출한 의미거리와 행동데이터 기반 의미거리의 상관관계는 어떠한가?

연구방법

실험 자극

목표 동사 선정

목표 동사는 한국어를 사용하는 신경언어장애군을 대상으로 한 동사 관련 선행연구 기반으로 선정하였다(선행연구 목록 및 선행연구 동사 목록은 Appendix 1 참고). 선행연구는 과제 유형에 따라 이해, 산출, 중재, 규준 연구로 분류되었으며, 언어적 수준에 따라 형태소, 단어, 문장, 이야기 수준으로 분류되었다. 동사 취합 시에는 과제 유형이나 언어적 수준에 제한을 두지 않고 단 한 번이라도 실험 동사로 쓰였을 경우 목표 동사에 포함하였다. 단, 기계학습으로 의미거리를 구하는 과정에서 오류를 야기할 수 있는 경우, 즉

(1) 동형이의어(homographic)인 형용사가 존재하는 동사(예: ‘쓰다’), (2) 합성 동사(예: 붙잡다), (3) ‘-하다’류 동사(예: 참가하다), (4) 동사 활용에 의해 의미 오류가 발생할 수 있는 동사(예: ‘들다’와 ‘들다’)는 목표 동사에서 제외하였다. 최종적으로 22개의 목표 동사가 선정되었다.

말뭉치 구성

본 연구는 두 개의 말뭉치를 구축하였다. 첫 번째 말뭉치는 ‘교과서 말뭉치’이다. 교과서는 비문(非文)이 적고 표현과 내용이 검증된 텍스트이므로 정제된 말뭉치에서 주격 및 목적격 명사의 특징을 밝히고자 한 본 연구의 목적에 부합하는 자료이다. 이에 교과서 말뭉치를 구축하여 교과서 말뭉치에서 목표 동사와 결합하는 주격 및 목적격 명사의 평균 출현빈도수 및 유형수의 차이를 구하였다. 두 번째 말뭉치는 ‘통합 말뭉치’이다. 통합 말뭉치는 교과서 말뭉치를 확장한 말뭉치이다. 교과서 말뭉치, 세종문어말뭉치, 그리고 한국어 위키피디아(Wikipedia) 말뭉치를 하나로 통합하였다. 세종문어말뭉치 및 한국어 위키피디아 말뭉치는 교과서 말뭉치보다 덜 정제된 성격의 문어 말뭉치이다. 그럼에도 불구하고 통합 말뭉치를 구축한 이유는 기계학습의 정확도를 높이기 위해서이다. 기계학습에서는 여러 용례를 학습할수록 어휘 간 의미관계를 좀 더 정확하게 파악하기 때문이다. 또한 통합 말뭉치는 기계학습으로 산출한 의미거리와 행동 데이터 간 상관관계를 분석하는 실험에서 설문 문항을 선정하는 데에도 사용하였다.

교과서 말뭉치

교과서 말뭉치는 중학교 국어교과서를 취합하여 구축하였다. 중학교 국어 수준으로 교과서 말뭉치를 구축한 이유는 중학교 교육 수준과 언어치료 대상자에게 기대하는 언어구사 수준이 근접하기 때문이다. 교육부에 따르면 중학교 교육의 목표는 ‘일상 생활과 학습에 필요한 기본 능력을 기르는 것’이다(교육부 고시 2015개정교육과정 제2015-74호 [별책3]에 의거함). 반면 초등학교 국어교과서의 경우 학년 간 텍스트의 수준 차이가 커서 선택하는 데 부적절하였고, 고등학교 국어교과서의 경우 일상 생활의 필요성과 수준을 넘어서는 문법 및 고전문학의 비중이 컸기 때문에 제외하였다. 교과서 말뭉치는 2009학년도 검인정 중학교 국어 교과서 7종—창작과 비평, 교학사(남미영), 미래엔(윤여탁), 신사고(민현식), 신사고(우한용), 비상교육(김태철), 비상교육(한철우)—을 취합하였으며, 교과서 간 중복된 텍스트는 하나만 남기고 삭제하였다. 중복 텍스트의 목록은 Appendix 2에 제시하였다.

Table 1. Size of corpus (Total = 81,278,738 eojeols)

	Number of eojeols (%)
Textbook	1,416,433 (1.74)
Sejong corpus (written)	35,149,824 (43.25)
Wikipedia-Korean	44,712,481 (55.01)

통합 말뭉치

세종말뭉치는 CD로 배포된 세종말뭉치(2010년 12월 수정판)의 문어 자료를, 한국어 위키피디아 말뭉치는 2016년 7월까지의 데이터를 모았다. 교과서, 세종, 위키피디아 각 말뭉치의 어절 단위(즉, 띄어쓰기 단위) 규모를 실험실에서 자체적으로 작성한 프로그램(Python 3.5.4)으로 구하여 Table 1에 제시하였다.

목표 동사의 주격 및 목적격 명사 추출

교과서 말뭉치에서 목표 동사와 결합하는 주격 및 목적격 명사를 추출하기 위해 ‘꼬꼬마(KKMA) 구문 분석기 버전 2.1’를 사용하였다. 이때 명사의 품사는 일반명사로 제한하였다. 최종 선정된 통사구조는 참고자료로 활용할 수 있도록 본 연구에서 분석한 출현 빈도수, 기계학습으로 산출한 의미거리 정보와 함께 구글드라이브(https://docs.google.com/spreadsheets/d/1Me7_LJ7r9gqSX-HwTmIqHL1ZhG5dMFFWwmkXQ4DpNKs/edit?usp=sharing)에 게시하였다.

실험 절차 및 분석

목표 동사와 결합하는 주격 명사와 목적격 명사의 평균 출현 빈도수 및 유형수 차이

평균 출현빈도수란 교과서 말뭉치에서 각 목표 동사와 결합한 주격 혹은 목적격 명사가 평균 몇 회 출현하였는지를 의미한다. 평균 유형수란 각 목표 동사와 결합한 주격 혹은 목적격 명사가 평균 몇 종이었는지를 의미한다. 추출한 주격 명사 총 258개, 목적격 명사 총 4,722개에 대하여 평균 출현빈도수 및 유형수 차이를 분석하였으며, 이때 평균 출현빈도수 및 유형수는 Python으로 계산하였다.

목표 동사로부터 주격 명사와 목적격 명사까지의 의미거리 분포의 차이

교과서 말뭉치에서 목표 동사와 결합한 서로 다른 주격 및 목적격 명사에 대하여 각 목표 동사별 평균 기계학습 기반 의미거리를 구하였다. 주격 명사 총 139개, 목적격 명사 총 1,002개가 분석 대상이 되었다. 기계학습 기반 의미거리 산출 과정은 다음과 같다.

Table 2. Parameters setup for Word2vec

	Value
Vector size	300
Min_count	5
Sample	10 ⁵
Window size	5

See Goldberg & Levy (2014) for more detailed explanation of the parameters.

워드 임베딩(Word Embedding)

어휘를 숫자열로 벡터화(vectorize)하는 워드 임베딩 방식은 여러가지가 있는데, 이 중 학습 속도를 현저히 개선한 Word2vec이 널리 쓰인다(Mikolov et al., 2013a). 이에 본 연구는 Word2vec으로 워드 임베딩을 수행하였다. Word2vec은 띄어쓰기를 기준으로 어휘를 구분하여 벡터를 산출한다. 예를 들어 ‘거짓말이 거짓말을 낳았다’라는 문장을 학습하면 띄어쓰기 단위로 세 개의 어휘(‘거짓말이’, ‘거짓말을’, ‘낳았다’)에 대한 벡터가 각각 산출된다. 본 연구는 먼저 통합 말뭉치를 꼬꼬마 구문분석기로 형태소 태깅(tagging)하여 형태소 단위로 띄어쓰기를 하였다. 즉, ‘거짓말이 거짓말을 낳았다’라는 문장은 형태소 태깅 후 ‘거짓말/NNG 이/JKS 거짓말/NNG 을/JKO 낳/VV 았/EPT 다/EFN’가 된다. 형태소 태깅의 목적은 목표 동사의 벡터를 어근 기준으로 산출하기 위함이다. 그리하여 동사의 활용형(예: ‘낳다’의 경우 ‘낳았다’ 등)마다 각각의 벡터가 산출되는 것을 방지하였다. 형태소 태깅 후에는 체언(보통명사, 고유명사, 대명사, 의존명사, 수사, 일반명사)과 조사(격조사, 보조사, 접속조사)를 붙였다. 즉, 앞의 예시 문장은 최종적으로 ‘거짓말이/NNG/JKS 거짓말을/NNG/JKO 낳/VV 았/EPT 다/EFN’가 된다. 체언과 조사를 붙인 이유는 명사가 주격 명사로 쓰였을 때와 목적격 명사로 쓰였을 때를 구분하여 벡터를 산출하기 위함이다. 또한 Word2vec은 매개변수(parameters) 조정을 통해 학습기준을 달리 할 수 있다. 이에 Gensim3.2.0으로 매개 변수를 바꾸어가며 말뭉치를 학습해 보았다. 결과적으로 단어 간 관계가 가장 잘 학습되었던 매개변수 값을 택하였으며, 그 값은 Table 2에 제시하였다. 각 매개 변수가 의미하는 바는 Goldberg와 Levy (2014)의 연구에 자세히 설명되어 있다.

Word2vec으로 임베딩한 단어들의 벡터를 다차원의 벡터 공간에 투영(project)하면 Figure 1처럼 비슷한 의미적 특성을 지닌 단어끼리 군집하여 분포하게 된다(Mikolov et al., 2013d). 따라서 임베딩 후 단어 벡터 간의 차이는 두 단어 간의 의미적 특성을 반영하는 ‘의미거리’라고 정의할 수 있다.

어포던스 벡터(Affordance vector)

워드 임베딩 후 다차원의 벡터 공간에서 명사는 자신과 결합이 가능한 동사들과 Figure 2A처럼 일정한 거리를 유지하고 있다. 이때, ‘명사-결합 가능한 동사 집단’ 사이의 평균 델타(delta)를 ‘어포던스’라고 한다. 따라서 Figure 2B처럼 다차원의 벡터 공간에 위치한 명사의 벡터에 결합 가능한 동사들과의 어포던스 벡터만큼을 더해주면, 명사는 결합 가능한 동사들이 군집해 있는 방향으로 이동(translation)한다(Fulda, Ricks, Murdoch, & Wingate, 2017). 본 연구에서도 어포던스 벡터를 사용하였다. 단, 본 연구에서는 명사가 아닌 목표 동사의 벡터를 이동시켰다. 즉, Word2vec으로 구해진

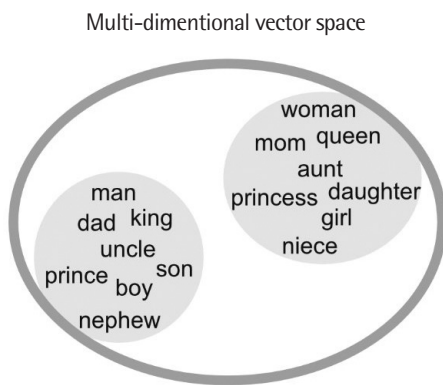


Figure 1. Example of the distribution of words in the multi-dimensional vector space after producing word embeddings by Word2vec. Words with similar semantic features are clustered as a result of Word2vec.

목표 동사의 벡터에 주격 명사와의 평균 어포던스나 목적격 명사와의 평균 어포던스를 빼주어 목표 동사를 각 명사가 군집해 있는 방향으로 이동시켰다.

코사인 유사도(Cosine similarity)

목표 동사를 결합 가능한 주격 혹은 목적격 명사들의 군집으로 이동시킨 후, Gensim으로 목표 동사와 명사들 간 코사인 유사도를 계산하였다. 코사인 유사도는 다차원의 벡터 공간에서 두 벡터 간의 거리를 계산하는 여러 방법 중의 하나이다. 두 벡터의 내적(dot product)으로 두 벡터가 이루는 각(angle)을 측정하며, 각이 클수록 두 단어의 거리가 먼 것으로 계산된다. 코사인 유사도값은 -1과 1 사이로, -1에 가까울수록 의미거리가 멀고 1에 가까울수록 의미거리가 가깝다는 것을 의미한다.

기계학습 기반 의미거리와 행동데이터 기반 의미거리 간 상관관계

연구 대상

기계학습으로 산출한 의미거리와 사람이 실제 인지하는 의미거리와의 상관관계를 알아보기 위해 20-30대 정상청년층 123명을 대상으로 구글 온라인설문을 실시하였다. 연구참여자의 평균 연령은 29.0세(SD = 4.4; range, 21-37세), 평균 교육년수는 15.4년(SD = 1.1; range, 9-16년), 남녀의 성비는 0.86:1이었다. 이때 발달성 장애력이 있는 경우, 모국어가 한국어가 아닌 경우는 설문 대상에서 제외하

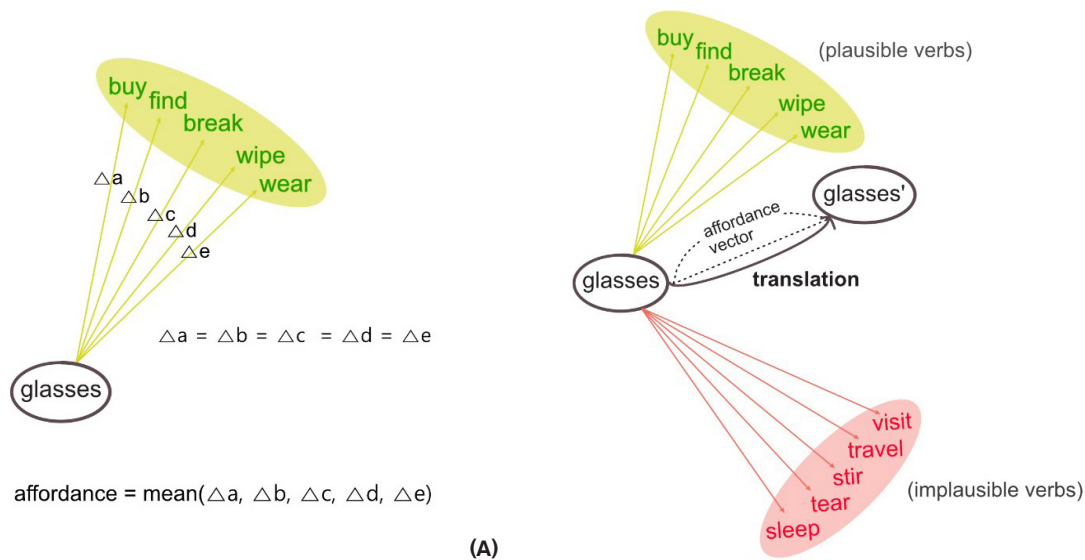


Figure 2. Example of affordance between a noun and verbs (A) and translation (B). (A) The mean delta (Δ) between a noun and its plausible verbs is called *affordance*. (B) Semantically implausible verbs with noun *glasses* are clustered at the bottom (red), whereas semantically plausible verbs are clustered on the top (green). After vector *glasses* are translated with the amount of affordance vector, the *glasses* move toward the vector space where their plausible verbs are clustered.

였다. 더불어 본 연구는 중학교 교과서 말뭉치에 기반하였으므로 피험자는 반드시 한국에서 중학교를 재학해야 했으며, 초등학교를 4년 이상 외국에서 다닌 경우도 제외하였다. 본 연구는 이화여자대학교 생명윤리위원회의 승인을 받아 진행하였다(No. 130-15).

설문 문항 구성

설문 문항은 통합 말뭉치에서 목표 동사와 기계학습에 기반한 의미거리가 최소:최대인 주격 명사를 1개씩, 목적격 명사를 1개씩 선정하였다. 의미거리가 최대인 명사는 목표 동사와 의미적으로 가장 먼 명사이므로 실제 언어 생활에서 쓰이지 않는 통사구조가 설문 문항이 되기도 하였다(예: '폭풍을 가르치다'). 명사 선정 과정에서 (1) 다의어, (2) 고유어로 대체할 수 없는 외래어(예: '벨브'), (3) 문맥없이 이해가 어려운 명사(예: '방도'), (4) 불쾌감을 주는 명사(예: '시신'), (5) 표준국어대사전에 등재되지 않은 명사, (6) 통합 말뭉치에서 5회 미만 등장하여 벡터가 산출되지 않은 명사는 제외하였다. 최종적으로 18개 동사에 대해 네 가지 조건('주격 명사-목표 동사 의미거리 최소', '주격 명사-목표 동사 의미거리 최대', '목적격 명사-목표 동사 의미거리 최소', '목적격 명사-목표 동사 의미거리 최대')에서 각 한 문항씩, 총 72문항으로 설문 문항을 구성하였다. 전체 설문 문항은 Appendix 3에 제시하였다.

실험참여자는 제시된 명사와 동사가 의미적으로 얼마나 관련이 있는지를 Likert 5점 척도로 평정하였다. 설문은 네 개의 섹션(section)으로 나누어졌으며, 각 섹션마다 4-5개의 동사에 대한 설문이 포함되었다. 홀수 섹션에는 주격 명사가, 짝수 섹션에서는 목적격 명사가 먼저 제시되었다. 본 연구의 목표 동사는 주격 및 목적격 명사가 모두 필요로 하는 2항 동사였다. 따라서 '주격 명사-목표 동사'

혹은 '목적격 명사-목표 동사'로만 설문 문항을 제시할 경우 설문 문항이 비문(非文)이 되어 어색하였다. 이에 주격 명사 설문에서는 목적어 위치에 '(~를)'을, 목적격 명사 설문에서는 주어 위치에 '(~가)'를 추가하여 설문 문항을 제시하였다. 설문의 예시는 Figure 3에 제시하였다.

자료의 통계적 처리

목표 동사와 결합하는 주격 명사와 목적격 명사의 평균 출현빈도수 및 유형수의 차이, 그리고 목표 동사로부터 주격 및 목적격 명사까지의 평균 기계학습 기반 의미거리 차이를 알아보기 위해 IBM SPSS Statistics version 20으로 대응표본 t-검정(paired t-test)을 실시하였다. 이때 평균에서 3 SD 이상 떨어진 이상치(outlier)는 제거 후 분석하였다. 기계학습으로 산출한 의미거리와 행동데이터 기반 의미거리의 상관관계 분석은 각 의미거리를 표준화 점수(z-score)로 변환한 후 SPSS로 Pearson 상관분석을 실시하였다.

연구결과

목표 동사와 결합하는 주격 및 목적격 명사의 평균 출현빈도수와 유형수의 차이

출현빈도수

이상치(>3 SD)를 제외한 결과 교과서 말뭉치에서 총 19개의 목표 동사가 분석에 포함되었다. 목표 동사와 결합하는 주격 명사는 총 121개, 목적격 명사는 총 1,237개가 추출되었다. 각 목표 동사별 주격 및 목적격 명사의 평균 출현빈도수를 알아보기 위하여 대응표본 t-검정을 실시하였다. 그 결과, 목표 동사별 주격 명사의 평균

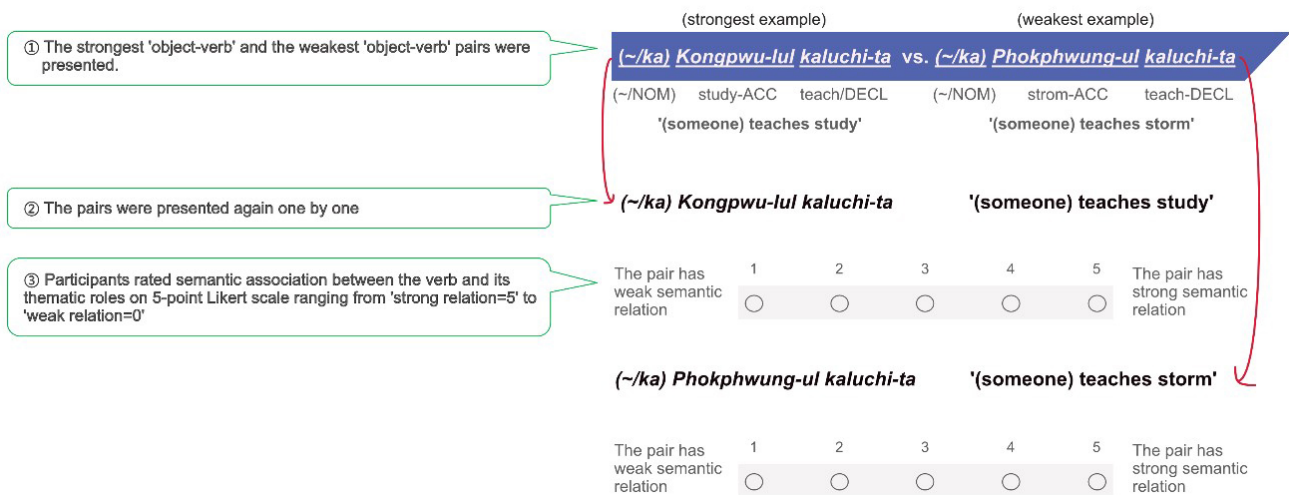


Figure 3. Example of a Google survey. Yale Romanization of Korean is used in transcribing the Korean alphabet (Martin, Yi, & Chang, 1967).

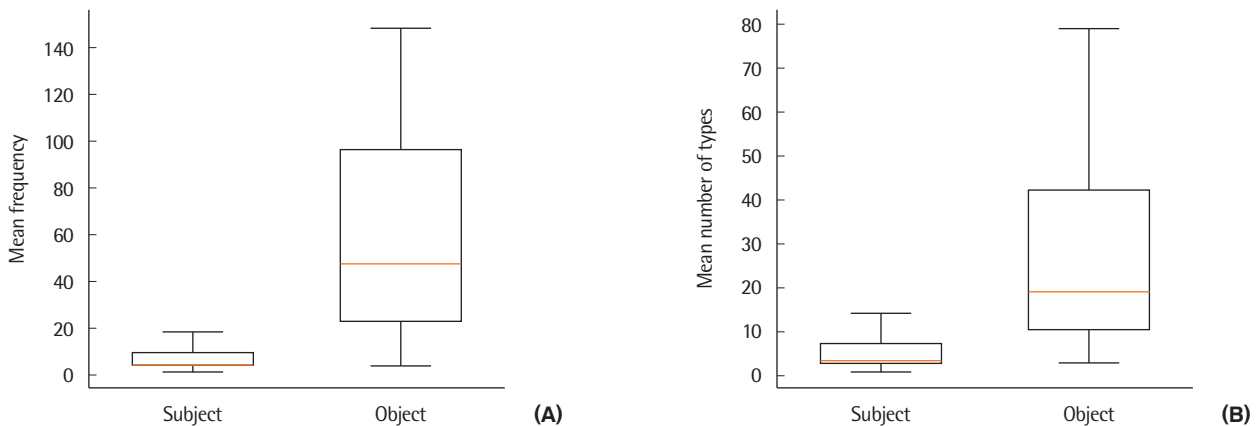


Figure 4. Boxplots of the mean frequency (A) and the mean number of types (B).

출현빈도수(6.37 ± 1.50)가 목적격 명사의 평균 출현빈도수(65.11 ± 66.35)보다 유의하게 낮았다($t_{18} = -4.171, p < .005$). 주격 및 목적격 명사의 평균 출현빈도수에 대한 상자도표(boxplot)는 Figure 4A에 제시하였다.

유형수

이상치를 제외한 결과 총 20개의 목표 동사가 분석에 포함되었다. 목표 동사와 결합하는 주격 및 목적격 명사 중 서로 다른 유형수의 주격 명사 총 105개, 목적격 명사 총 708개가 추출되었다. 각 목표 동사별 주격 및 목적격 명사의 평균 유형수를 알아보기 위하여 대응표본 *t*-검정을 실시하였다. 그 결과, 주격 명사의 평균 유형수(5.25 ± 4.40)가 목적격 명사의 평균 유형수(35.40 ± 39.37)보다 유의하게 낮았다($t_{19} = -3.766, p < .005$). 주격 및 목적격 명사의 유형수에 대한 상자도표는 Figure 4B에 제시하였다.

‘목표 동사-주격 명사’와 ‘목표 동사-목적격 명사’의 기계학습 기반 의미거리 차이

이상치를 제외하고 교과서 말뭉치에서 21개의 목표 동사와 결합한 주격 명사(N=124)와 목적격 명사(N=814)를 대상으로 의미거리를 산출하였다. 각 목표 동사에서 주격 및 목적격 명사까지의 평균 기계학습 기반 의미거리의 분포 차이를 알아보기 위하여 대응표본 *t*-검정을 실시하였다. 기계학습으로 산출한 의미거리는 -1부터 1 사이의 값을 가지며, 1에 가까울수록 의미거리가 가까운 것으로 해석한다. 분석 결과 목표 동사와 주격 명사의 평균 의미거리(0.33 ± 0.08)가 목표 동사와 목적격 명사의 평균 의미거리(0.38 ± 0.06)보다 유의하게 먼 것으로 나타났다($t_{20} = -3.325, p < .005$). 기계학습으로 산출한 의미거리에 대한 상자도표는 Figure 5에 제시하였다.

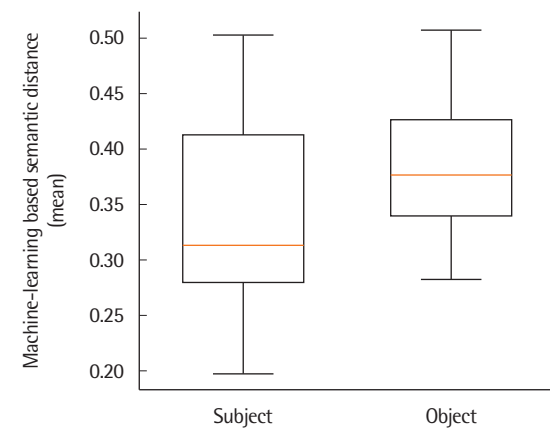


Figure 5. Boxplots of the mean machine learning-based semantic distance to the target verbs.

기계학습으로 산출한 의미거리와 행동데이터 기반 의미거리의 상관관계

기계학습으로 산출한 의미거리와 행동데이터 기반 의미거리 간 상관관계 분석을 위해 Pearson product-moment를 사용하였다. 그 결과, 기계학습으로 산출된 의미거리와 행동데이터 기반 의미거리 간 상관관계는 주격과 목적격 모두에서 매우 강한 정적(positive) 관계를 보이는 것으로 나타났다(주격명사: $r = .806, p < .001$, 목적격명사: $r = .930, p < .001$). 기계학습으로 산출한 의미거리와 행동데이터 기반 의미거리의 산점도(scatter plot)는 Figure 6에 제시하였다.

논의 및 결론

본 연구는 한국어를 사용하는 신경언어장애군의 동사 중재에서 중재 자극으로 사용될 수 있는 한국어 주격 및 목적격 명사에 대한 기초자료를 빅데이터 기반으로 분석하고 원자료를 제공하고자 하

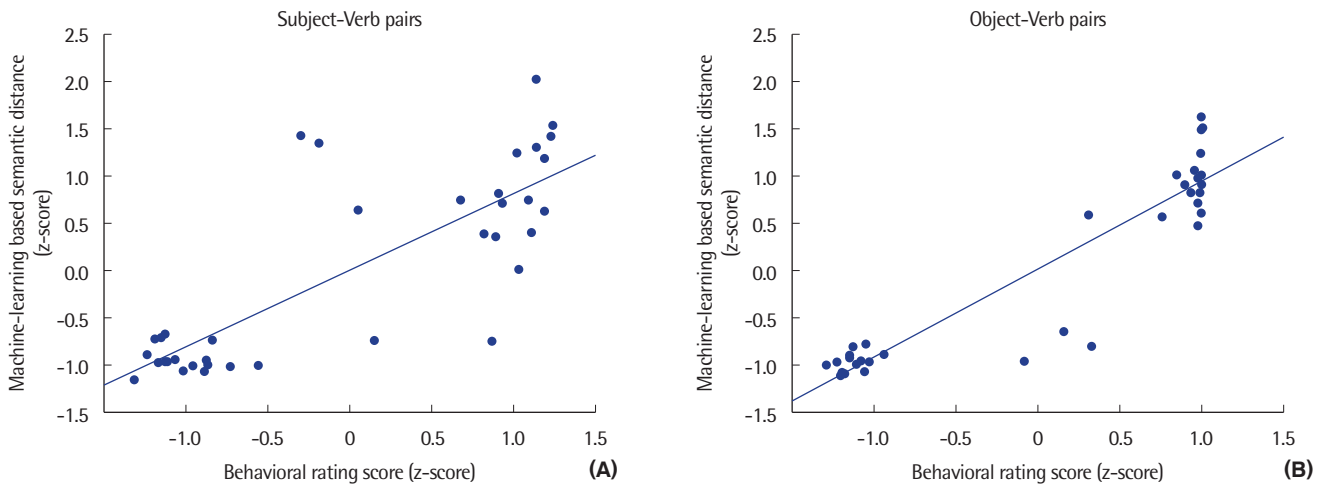


Figure 6. Scatter plots of the correlation between machine learning-based semantic distance and the behavioral distance rating score: subject-verb pairs (A) and object-verb pairs (B).

었다. 이를 위해 교과서 말뭉치를 분석하여 목표 동사와 결합하는 주격 및 목적격 명사의 평균 출현빈도수, 그리고 유형수 차이를 규명하였다. 또한 목표 동사로부터 주격 및 목적격 명사까지의 평균 의미거리를 기계학습으로 산출하여 의미거리의 분포 차이를 분석하였다. 더불어 기계학습으로 산출한 의미거리와 행동데이터 기반 의미거리의 상관관계를 밝혀 방법론에 대한 검증은 수행하였다.

분석 결과 주격 명사는 목적격 명사보다 평균 출현빈도수 및 유형수가 모두 낮은 것으로 드러났다. 대규모 텍스트 자료에서, 특히나 정제된 표현이 담긴 교과서 말뭉치에서 주어생략현상이 일관되게 나타난다는 점은 주목할 필요가 있다. 교과서보다 표현의 제약이 엄격하지 않은 여타의 문어 말뭉치나 구어 말뭉치에서는 주어의 생략이 더욱 빈번할 것임을 예상할 수 있기 때문이다. 앞서 한국어 사용 경도인지장애군 및 실어증 환자를 대상으로 동사 논항수에 따른 산출 과제를 수행하였을 때 1-3항 동사 간의 산출만 유의하게 어려움이 증가하는 현상이 나타난다고 하였다(Choi et al., 2013; Sung, 2016). 그리고 Sung (2016)은 그 원인을 한국어 2항 동사에서 의 갖은 주어생략현상으로 해석하였다. 의미역 선택 과제에서도 주어의 빈번한 생략이 정반응률에 영향을 미치는 것으로 논의된 바 있다. Jeong과 Sung (2018)은 한국어 사용 실어증 환자군과 정상 집단에게 여러 유형의 명사를 제시한 후 그 안에서 동사의 주격 의미역(행위자)과 목적격 의미역(대상)을 선택하는 과제를 수행하게 하였다. 그 결과 두 집단은 모두 동사의 행위자를 선정하는데 더 어려움을 겪었으며, 이는 구어에서 주어의 생략 비율이 높은 한국어의 특징에 기인한 것으로 보았다(Jeong & Sung, 2018). 이처럼 지금까지 주어생략현상은 한국어를 사용하는 신경언어장애군의 동사 관련 연구에서 가설로만 제기되었다. 그러나 본 연구는 주어생략이

빅데이터에서 확인할 수 있는 현상임을 밝혔다는 점에서 의의가 있다. 즉, 이러한 결과는 임상적 현상을 빅데이터 기반으로 설명할 수 있는 기초자료를 제공하였다는 점에서 의미가 있다. 앞으로도 본 연구 결과는 동사 논항이나 의미역과 관련하여 한국어를 사용하는 신경언어장애군이 보이는 한국어 특정적(Korean-specific) 현상에 대한 근거자료로 활용될 수 있을 것이다.

기계학습으로 산출한 의미거리 분석에서는 주격 명사가 목적격 명사보다 동사와의 의미거리가 먼 것으로 드러났다. 이는 앞서 밝힌 주어생략현상이 반영된 결과로 보인다. 즉, 주어가 자주 생략되는 한국어 특성상 주어가 동사와 결합하는 빈도가 낮아지고 의미거리도 멀어지는 것으로 해석된다. 따라서 한국어를 사용하는 신경언어장애군을 대상으로 하는 동사 중재에서 주어 활용 방안에 대한 고찰이 필요하다. 예를 들어, 영어권 실어증 환자군을 대상으로 한 동사의미역강화중재에서는 주격 및 목적격 의미역을 동등한 빈도로 치료를 진행한다(Edmonds & Babb, 2011). 한국어 사용 실어증 환자군을 대상으로 한 동사의미역강화중재연구에서도 영어권 환자들과 동일한 중재법을 적용하였다(Kwag et al., 2014). 하지만 주어 탈락이 빈번하고 주격 명사-동사의 의미거리가 먼 것을 빅데이터 기반으로 밝힌 본 논문의 결과를 감안할 때, 다음과 같은 임상적 시사점이 있다. 한국어에서는 주격 명사를 중재자극으로 목적격 명사와 자극 빈도를 동일하게 치료하는 것은 동사 중재 효율성의 문제를 야기할 수 있다. 세종 구어 말뭉치 자료에서 주어생략현상을 분석한 Lee (2014)에 따르면 일반적인 대화 상황에서 한국어 화자는 1.2인칭 주어를 쓰지 않으며, 화용적(pragmatics) 의도가 있을 때만 주어를 명시한다고 하였다. 본 연구에서도 주격 명사는 빈번히 생략되며 동사와의 의미거리가 멀었다. 따라서 한국어를 사용

하는 신경언어장애군을 치료할 때 주격 이외의 논항으로 동사의 논항구조를 활성화시켜 중재 및 일반화 효과를 극대화하는 한국어 특성을 반영한 중재 프로토콜을 개발할 필요가 있다. 예를 들어, 여격(-에/에게), 탈격(-에서/에게서), 방향격(-로/으로)을 논항으로 취하는 3항 동사의 활용 방안을 고려할 수 있다.

기계학습으로 산출한 의미거리와 행동데이터 기반 의미거리의 상관관계는 매우 강한 정적 상관관계를 보였다. 즉, 기계학습 기반 의미거리가 실제 한국어 화자가 느끼는 의미거리와 관련성이 높다는 것을 입증하였다. 최근 국외에서는 신경언어장애군을 대상으로 한 기계학습 연구가 다양하게 이루어지고 있으며, 본 연구에서처럼 기계학습 결과를 행동데이터로 증명하고 있다(Kiran & Thompson, 2003; Le, Licata, Mercado, Persad, & Provost, 2014). 기계학습 데이터로는 컴퓨터 단층촬영(computed tomography, CT) 결과 및 신경심리검사 결과, 자발화 등의 자료를 이용한다. 그리하여 기계학습 결과를 바탕으로 신경언어장애군의 진단 시 활용할 수 있는 분류기준(classifier)을 찾고, 자동으로 환자의 임상적 진단을 내리기 위한 시스템을 개발하려는 시도가 주를 이루고 있다(Bentley et al., 2014; Fraser et al., 2014; Garrard, Rentoumi, Gesierich, Miller, & Gorno-Tempini, 2014; Järvelin & Juhola, 2011; Orimaye, Wong, Golden, Wong, & Soyiri, 2017). 기계학습 결과를 치료에까지 적용한 연구도 있다(Kiran & Thompson, 2003). Plaut (1996)은 컴퓨터 모의 실험으로 단어의 전형성(typicality)이 치료 일반화에 미치는 영향을 분석한 바 있다. Plaut (1996)은 먼저 컴퓨터 네트워크에 단어의 의미(semantic) 정보를 학습시킨 후, 학습 내용을 다시 의도적으로 손상시키고 재학습시켰다. 그 결과, 비전형적(atypical) 단어들(예: ‘새’라는 범주에서 펭귄)을 먼저 재학습시키면 전형적(typical) 단어들(예: ‘비둘기’)을 인식하는 데 향상이 나타났다. 반면 전형적 단어들을 먼저 재학습시키면 훈련한 단어에서만 향상이 일어났다. 이후 Kiran과 Thompson (2003)은 이 모의실험 결과를 유창성 실어증 환자군의 이름대기 중재에 적용하여 비전형적인 단어를 중재하는 것이 일반화 효과가 크다는 것을 증명하였다.

국내에서는 So, Hooshyar, Park과 Lim (2017)이 기계학습을 통해 치매를 조기판별할 수 있는 신경심리검사를 찾아내고자 한 바 있으나 기계학습 결과를 행동데이터로 검증한 연구는 찾아보기 힘들다. 본 연구에서는 기계학습으로 의미거리를 측정하고 그것을 행동데이터로 직접 증명하는 새로운 시도를 하였다. 특히 본 연구에서 시도한 기계학습 기반 의미거리는 ‘동사-명사’ 개별 자극에 대한 의미거리를 하나씩 정량화하여 나타낼 수 있다는 강점이 있어 중재 자극 선정 과정에서 임상적으로 유용하게 활용될 수 있다. 예를 들어, 신경언어장애군의 동사 중재 이론 중 복잡성 이론(Complex-

ity Account of Treatment Efficacy)에 따르면 복잡한 구조로부터 중재를 시작할 경우 단순한 구조 산출로까지 일반화가 일어나지만, 단순한 구조로부터 중재를 시작하면 복잡한 구조의 산출로까지 일반화가 일어나지 않는다고 하였다(Thompson, Shapiro, Kiran, & Sobecks, 2003). 복잡성 이론에 기초한 중재 연구들은 중재 자극의 난이도를 조절하는 방법으로 내포문의 활용, 논항수의 증가, 명사의 전형성 및 친숙도 조절 등을 활용하여 중재 효과를 증명하였다(Kiran & Thompson, 2003; Schneider & Thompson, 2003; Thompson et al., 2003; Thompson & Shapiro, 2005). 본 연구에서 제시한 의미거리도 중재 난이도 조절하는 데 있어 또 하나의 기준으로 활용할 수 있다. 예를 들어, 의미거리가 먼 명사부터 중재하는 것이 한국어에서도 일반화 효과를 극대화하는 방법이 될 수 있는지에 대한 검증에 본 자료를 활용할 수 있을 것이다. 이에 본 연구는 목표 동사와 주격 및 목적격 명사의 빈도수 및 기계학습 기반 의미거리를 클라우드 기반으로 공개하여 언어치료사들이 적극적으로 활용할 수 있게 하였다. 특히 생략이 잦아 중재 자극으로써의 효율성이 낮을 가능성이 큰 주격 명사보다는 목적격 명사 위주의 활용이 적절할 것으로 사료된다. 다만 본 연구는 기계학습으로 산출한 의미거리와 행동데이터 간 상관관계를 알아본 첫 시도였기 때문에 목표 동사와 최대·최소 거리에 위치한 명사만을 검증하였다는 실험적 한계를 가진다. 따라서 기계학습과 행동데이터 간 검증방법을 보다 다양화하는 후속연구가 필요할 것으로 보인다. 예를 들어 명사를 기계학습에 기반한 의미거리에 따라 범주화하고, 각 범주에서 모두 한국어 화자의 의미거리 평정과 정적 상관관계를 나타내는 지 검증하는 방법도 대안으로 생각해 볼 수 있다.

본 연구에서 빅데이터를 기반으로 제시한 한국어 동사의 주격 명사 및 목적격 명사의 특징은 임상적 가설을 뒷받침하는 기초자료로써 유용하게 사용될 수 있다. 특히 본 연구는 누구나 활용할 수 있도록 클라우드 기반으로 원자료를 공개하여 한국어를 사용하는 신경언어장애군에게 적합한 동사 중재 자극을 선정하는 데 기여를 하고자 하였다. 차후에는 정제된 교과서 말뭉치와 성격을 달리하는 구어 말뭉치와의 비교를 통해 한국어 특징을 밝혀 최적의 동사 중재 자극을 선정하기 위한 연구가 지속될 필요가 있다. 아울러 선정한 중재 자극을 치료에 적용하여 효과성에 대한 부분도 검증되어야 할 것이다.

REFERENCES

- Ahn, T. S. (2002). *Production of grammatical inflections of adults with Broca's aphasia* (Master's thesis). Yonsei University, Seoul, Korea.

- Basso, A., Razzano, C., Faglioni, P., & Zanobio, M. E. (1990). Confrontation naming, picture description and action naming in aphasic patients. *Aphasiology*, 4(2), 185-195.
- Bentley, P., Ganesalingam, J., Jones, A. L. C., Mahady, K., Epton, S., Rinne, P., ... & Rueckert, D. (2014). Prediction of stroke thrombolysis outcome using CT brain machine learning. *NeuroImage: Clinical*, 4, 635-640.
- Chen, S., & Bates, E. (1998). The dissociation between nouns and verbs in Broca's and Wernicke's aphasia: findings from Chinese. *Aphasiology*, 12(1), 5-36.
- Choi, E., Sung, J. E., Jeong, J. H., & Kwag, E. (2013). Noun-verb dissociation in a confrontation naming task for persons with mild cognitive impairment. *Dementia and Neurocognitive Disorders*, 12(2), 41-46.
- Choi, S. J., & Sung, J. E. (2014). Task-specific and argument structure effects on verb production in normal elderly adults: animation vs. picture comparisons. *Journal of Rehabilitation Research*, 18(4), 279-293.
- Choi, S. J., Sung, J. E., Jeong, J. H., & Kim, G. H. (2016). Effects of working memory treatment using verb-carrier phrases on language processing in persons with mild cognitive impairment. *Communication Sciences & Disorders*, 21(3), 524-537.
- De Bleser, R., & Kauschke, C. (2003). Acquisition and loss of nouns and verbs: parallel or divergent patterns? *Journal of Neurolinguistics*, 16(2-3), 213-229.
- Edmonds, L. A., & Babb, M. (2011). Effect of verb network strengthening treatment in moderate-to-severe aphasia. *American Journal of Speech-Language Pathology*, 20(2), 131-145.
- Edmonds, L. A., Mammino, K., & Ojeda, J. (2014). Effect of verb network strengthening treatment (VNeST) in persons with aphasia: extension and replication of previous findings. *American Journal of Speech-Language Pathology*, 23(2), S312-S329.
- Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, 23(3), 402-424.
- Eom, B. R., & Sung, J. E. (2015). Effects of syntactic complexity-based working memory treatment on sentence comprehension abilities in persons with aphasia. *Proceedings of the 2015 Conference of Korean Speech-Language & Hearing Association*, 204-205.
- Fraser, K. C., Meltzer, J. A., Graham, N. L., Leonard, C., Hirst, G., Black, S. E., & Rochon, E. (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex*, 55, 43-60.
- Fulda, N., Ricks, D., Murdoch, B., & Wingate, D. (2017). What can you do with a rock? Affordance extraction via word embeddings. Retrieved from <https://arxiv.org/abs/1703.03429>
- Garrard, P., Rentoumi, V., Gesierich, B., Miller, B., & Gorno-Tempini, M. L. (2014). Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. *Cortex*, 55, 122-129.
- Gillam, R. B., Marquardt, T. P., & Martin, F. N. (2011). *Communication sciences and disorders: from science to clinical practice*. Sudbury, MA: Jones & Bartlett Learning.
- Goldberg, Y., & Levy, O. (2014). Word2vec explained: deriving Mikolov et al's negative-sampling word-embedding method. Retrieved from <https://arxiv.org/abs/1402.3722>
- Hwang, S. H., & Choi, J. N. (2013). A case study of verb production during action naming in Korean patients with Alzheimer's disease. *Journal of Language Sciences*, 20(3), 143-173.
- Hwang, S. H. (2014). Alzheimer's disease patients' and normal children's use of Korean [SEPARATION] verbs in an action naming task. *Journal of Language Sciences*, 21(1), 235-262.
- Hyun, J. M., Kim, H. H., Shin, J. C., & Seo, S. G. (2003). Retrieval of nouns and verbs in Broca's and Wernicke's aphasia. *Korean Journal of Communication & Disorders*, 8(3), 171-187.
- International Organization for Standardization. (2015). *Information technology - Vocabulary (ISO/IEC 2382:2015)*. Retrieved from <https://www.iso.org/standard/63598.html>
- Järvelin, A., & Juhola, M. (2011). Comparison of machine learning methods for classifying aphasic and non-aphasic speakers. *Computer Methods and Programs in Biomedicine*, 104(3), 349-357.
- Jeong, K. H., & Sung, J. E. (2018). Verbs and their thematic role processing abilities for people with aphasia. *Communication Sciences & Disorders*, 23(2), 337-346.
- Jonkers, R., & Bastiaanse, R. (1996). The influence of instrumentality and transitivity on action naming in Broca's and anomic aphasia. *Brain and Language*, 55(1), 37-39.
- Jung, H. S. (2009). *Production of grammatical morphemes by Korean: speaking adults with Broca's aphasia* (Master's thesis). Dankook University, Yongin, Korea.
- Kemmerer, D., & Tranel, D. (2000). Verb retrieval in brain-damaged subjects. 1. Analysis of stimulus, lexical, and conceptual factors. *Brain and Language*, 73(3), 347-392.
- Kim, H. L., & Sung, J. E. (2016). Effects of syntactic complexity on sentence comprehension in the discourse of persons with aphasia. *Communication*

- Sciences & Disorders*, 21(4), 668-682.
- Kim, H. N., & Sung, J. E. (2014). Age-related changes in story retelling procedures and their relation to working memory capacity. *Special Education Research*, 13(13), 7-24.
- Kim, J. H. (2008). *A study on the verb comprehension depending on severity and types in aphasic patient* (Master's thesis). Myongji University, Seoul, Korea.
- Kim, J., Sung, J. E., Lee, S. E., & Sim, H. S. (2018). ERP components associated with locality dependency of time references between young and elderly adults. *Communication Sciences & Disorders*, 23(1), 129-145.
- Kim, K. Y. (2006). *Verb production and argument structures in aphasics* (Master's thesis). Yonsei University, Seoul, Korea.
- Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs in agrammatism: implications for lexical organization. *Brain and Language*, 74(1), 1-25.
- Kim, S. J. (2016). A study on the form of subjects in Korean written text by genre. *Hanminjok Emunhak*, 72, 25-62.
- Kim, S. J., & Choi, D. J. (2013). A study on the form of subjects in Korean written text. *Hanminjok Emunhak*, 64, 37-69.
- Kim, S. R., Park, C. I., Kim, D. Y., & Hwang, M. A. (2004). Production of nouns and verbs in adults with Broca's aphasia: comparison between naming and narration. *Korean Journal of Communication & Disorders*, 9(2), 1-18.
- Kiran, S., & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits. *Journal of Speech, Language, and Hearing Research*, 46(4), 771-787.
- Kwag, E. J., Sung, J. E., Kim, Y. H., & Cheon, H. J. (2014). Effects of verb network strengthening treatment on retrieval of verbs and nouns in persons with aphasia. *Communication Sciences & Disorders*, 19(1), 89-98.
- Lee, D., Licata, K., Mercado, E., Persad, C., & Provost, E. M. (2014). Automatic analysis of speech quality for aphasia treatment. *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4853-4857.
- Lee, J. H. (2017). *The effects of constraint induced language therapy focusing on verb production with tense in Korean individuals with aphasia* (Master's thesis). Ewha Womans University, Seoul, Korea.
- Lee, J., Kwon, M., Na, H. R., Bastiaanse, R., & Thompson, C. K. (2013). Production and comprehension of time reference in Korean nonfluent aphasia. *Communication Sciences & Disorders*, 18(2), 139-151.
- Lee, N. R. (2014). A reconsideration of the omission of first and second person subjects in modern spoken Korean: focusing on the pragmatic meanings of the overt subject expressions. *Discourse and Cognition*, 21(3), 145-163.
- Lee, O. B., & Lee, S. H. (2007). The effects of the story-retelling technique on the word retrieval and production of sentences of an aphasic subject: a case study. *Journal of Speech & Hearing Disorders*, 16(2), 147-159.
- Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A., & Pinna, G. D. (2002). Verb-noun double dissociation in aphasic lexical impairments: the role of word frequency and imageability. *Brain and Language*, 81(1-3), 432-444.
- Miceli, G., Silveri, M. C., Villa, G., & Caramazza, A. (1984). On the basis for the agrammatic's difficulty in producing main verbs. *Cortex*, 20(2), 207-220.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013b). Efficient estimation of word representations in vector space. Retrieved from <https://arxiv.org/abs/1301.3781>
- Mikolov, T., Le, Q. V., & Sutskever, I. (2013c). Exploiting similarities among languages for machine translation. Retrieved from <https://arxiv.org/abs/1309.4168>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013a). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26, 3111-3119.
- Mikolov, T., Yih, W. T., & Zweig, G. (2013d). Linguistic regularities in continuous space word representations. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746-751.
- Minjungseogwan. (1967). *New Korean-English Dictionary*. Seoul: Minjungseogwan.
- Min, O. Y., Kang, J. S., Min, K. O., & Hwang, Y. J. (2011). A study on sentence comprehension characteristics according to the types of aphasia in sentence correction judgement task: focused on phonologic, semantic and syntactic errors. *Journal of the Institute of Special Education & Rehabilitation Science*, 50(2), 369-386.
- Nam, K. S. (2004). *Modern Korean syntax*. Seoul: Taehaksa.
- Nam, K. S., & Ko, Y. G. (1993). *The standard Korean grammar*. Seoul: Tap Publishing.
- Nam, S. H. (2007). *Event structure and argument structure of Korean predicates*. Seoul: Seoul National University Press.
- Nicholas, M., Obler, L., Albert, M., & Goodglass, H. (1985). Lexical retrieval in healthy aging. *Cortex*, 21(4), 595-606.

- Orimaye, S. O., Wong, J. S., Golden, K. J., Wong, C. P., & Soyiri, I. N. (2017). Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. *BMC Bioinformatics*, 18(1), 34.
- Ororbia II, A. G., Mikolov, T., & Reitter, D. (2017). Learning simpler language models with the differential state framework. *Neural Computation*, 29(12), 3327-3352.
- Park, C. H. (2012). Statistical approach about ellipsis of Korean and English. *Journal of the Society of Korean Language and Literature*, 66, 171-191.
- Park, K. A., Kim, H., Park, E. S., & Shin, J. C. (2006). Verbal morpheme retrieval in aphasia. *Korean Journal of Communication & Disorders*, 11(1), 51-63.
- Plaut, D. C. (1996). Relearning after damage in connectionist networks: toward a theory of rehabilitation. *Brain and Language*, 52(1), 25-82.
- Schneider, S., & Thompson, C. (2003). Verb production in agrammatic aphasia: the influence of semantic class and argument structure properties on generalisation. *Aphasiology*, 17(3), 213-241.
- Shin, S. E. (2008). *Changes in verb production and comprehension ability depending on severity in Alzheimer's disease* (Master's thesis). Ewha Womans University, Seoul, Korea.
- So, A., Hooshyar, D., Park, K., & Lim, H. (2017). Early diagnosis of dementia from clinical data by machine learning techniques. *Applied Sciences*, 7, 651.
- Son, B. R. (2001). The influence of thought and culture on language. *Modern Studies in English Language and Literature*, 45(2), 131-152.
- Sung, J. E. (2016). The effects of verb argument complexity on verb production in persons with aphasia: evidence from a subject-object-verb language. *Journal of Psycholinguistic Research*, 45(2), 287-305.
- Sung, J. E., & Kwag, E. J. (2012). Age-related verb naming abilities depending on the argument structures. *Korean Journal of Communication & Disorders*, 17(4), 550-564.
- Thompson, C. K., & Shapiro, L. P. (2005). Treating agrammatic aphasia within a linguistic framework: treatment of underlying forms. *Aphasiology*, 19(10-11), 1021-1036.
- Thompson, C. K., & Shapiro, L. P. (2007). Complexity in treatment of syntactic deficits. *American Journal of Speech-Language Pathology*, 16(1), 30-42.
- Thompson, C. K., Lange, K. L., Schneider, S. L., & Shapiro, L. P. (1997). Agrammatic and non-brain-damaged subjects' verb and verb argument structure production. *Aphasiology*, 11(4-5), 473-490.
- Thompson, C. K., Riley, E. A., den Ouden, D. B., Meltzer-Asscher, A., & Lukic, S. (2013). Training verb argument structure production in agrammatic aphasia: behavioral and neural recovery patterns. *Cortex*, 49(9), 2358-2376.
- Thompson, C. K., Shapiro, L. P., & Roberts, M. M. (1993). Treatment of sentence production deficits in aphasia: a linguistic-specific approach to wh-interrogative training and generalization. *Aphasiology*, 7(1), 111-133.
- Thompson, C. K., Shapiro, L. P., Kiran, S., & Sobecks, J. (2003). The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia. *Journal of Speech, Language, and Hearing Research*, 46(3), 591-607.
- Yoon, J., & Kim, Y. T. (2002). The category specificity in Korean Broca's aphasic patients: the impairment of verb comprehension. *Korean Journal of Communication & Disorders*, 7(3), 39-54.
- Zingeser, L. B., & Berndt, R. S. (1990). Retrieval of nouns and verbs in agrammatism and anomia. *Brain and Language*, 39(1), 14-32.

Appendix 1. 신경언어장애군 동사 관련 선행연구 및 동사 목록

선행연구 목록 (저자 가나다순)

연번	저자(년도)	동사 개수
1	Kwag, Sung, Kim, & Cheon (2014)	36
2	Kim (석사, 2006)	64
3	Kim, Park, Kim, & Hwang (2004)	20
4	Kim (석사, 2008)	120
5	Kim, Sung, Lee, & Sim (2018)	7
6	Kim & Sung (2014)	34
7	Kim & Sung (2016)	48
8	Min, Kang, Min, & Hwang (2011)	60
9	Park, Kim, Park, & Shin (2006)	50
10	Sung & Kwag (2012)	40
11	Shin (석사, 2008)	160
12	Ahn (석사, 2002)	5
13	Eom & Sung (2015)	18
14	Yoon & Kim (2002)	75
15	Lee & Lee (2007)	16
16	Lee (석사, 2017)	24
17	Jeong & Sung (2018)	15
18	Jung (석사, 2009)	40
19	Choi & Sung (2014)	8
20	Choi, Sung, Jeong, & Kim (2016)	7
21	Choi, Sung, Jeong, & Kwag (2013)	30
22	Hyun, Kim, Shin, & Seo (2003)	25
23	Hwang (2014)	13
24	Hwang & Choi (2013)	13
25	Lee, Kwon, Na, Bastiaanse, & Thompson (2013)	18

선행연구 동사 목록 (총 296개, 가나다순)

	동사 목록
ㄱ	가꾸다* , 가다, 가두다, 가르치다* , 가리키다, 감기다, 감다* , 갖다주다, 걱정하다, 건설하다, 건지다, 겹다, 걸다, 걸리다, 결석하다, 결혼하다, 계산하다, 고르다, 고치다* , 공부하다, 구걸하다, 구르다, 굶다, 권하다, 그만두다, 그치다, 기대하다, 기도하다, 기뻐하다, 깎다* , 깔다, 깨다* , 깨뜨리다, 깨우다, 깨지다, 꺼내다, 꺾이다, 꼬집다* , 꼬집히다, 꽃다, 꽃히다, 꿇다, 꿈꾸다, 꺾다, 꿇다, 끝나다, 끼다
ㄴ	나다, 나누다, 나오다, 낙서하다, 낚시하다, 날다, 날리다, 날아가다, 남다, 내다, 내려가다, 내려놓다, 내리다, 내쫓다, 널다, 넘다, 넘어지다, 넘치다, 널다, 노래하다, 노력하다, 녹다, 녹이다, 놀다, 놀리다, 놓다, 놓아두다, 놓치다, 누르다, 눅히다
ㄷ	다가가다, 다가오다, 다치다, 달리다, 담다, 당황하다, 던지다, 데려오다, 도착하다, 돌다, 돌려보내다, 돌리다, 돌아가다, 돌아보다, 돕다* , 되다, 두드리다* , 들다, 들리다, 들어가다, 들어오다, 듣다, 따라가다, 따르다, 때리다* , 떠올리다, 떨어뜨리다, 떼다, 뚫다, 뛰다
ㄹ	막다* , 막히다, 말하다, 망가뜨리다, 망가지다, 망설이다, 맞다, 맞이하다, 맞추다* , 매달다, 맷다, 맷히다, 먹다* , 먹이다, 먹히다, 멈추다, 메다, 모이다, 목욕하다, 무치다, 묵다, 묵이다, 묻다, 물리다, 미끄러지다, 미안해하다, 밀리다
ㅂ	바꾸다* , 바르다, 바치다, 박다, 받다, 발견하다, 발생하다, 밝히다, 밝히다, 배우다, 버리다, 벗기다, 베다, 변하다, 보이다, 부딪히다, 부러뜨리다, 부러지다, 부촉하다, 부치다, 부화하다, 붓다, 붙이다, 붙잡다, 비우다, 빌다, 빗다, 빛나다, 빼뜨리다, 빠지다, 빼다, 뽑다, 뽑히다, 뿌리다
ㅅ	사과하다, 사랑하다, 사망하다, 사용하다, 살다, 상상하다, 생각하다, 서다, 섞다, 설명하다, 세우다, 소개하다, 소리치다, 숨기다, 숨다, 시들다, 시키다, 신기다, 신다, 심다, 싸다, 싸우다, 쌓이다, 썩다, 썰다, 쏘다, 쏘이다, 쓰다, 쓰러지다, 씹우다, 씹다, 씻다, 씻기다
ㅇ	안기다, 안다, 앉다, 알리다, 양보하다, 얻다, 열다, 업다, 업히다, 없다, 여기다, 열리다, 오다, 오르다, 오리다, 올라가다, 올리다, 옮기다, 요리하다, 운전하다, 울다, 울리다, 움직이다, 웃기다, 웃다, 원하다, 위하다, 이야기하다, 인사하다, 일어나다, 일어서다, 일으키다, 일하다, 읽다* , 입히다, 있다
ㅈ	자다, 자라다, 잠그다* , 잠기다, 잡다* , 잡히다, 재우다, 저촉하다, 전하다, 전화하다, 좇다, 주다, 죽다, 지나다, 지르다* , 지적하다, 지치다, 진찰받다, 짓다, 짜다, 쫓아가다, 쫓기다, 쫓아내다, 찍다* , 찢리다, 찢기다, 찢다*
ㅊ	차다, 차이다, 차지하다, 참가하다, 찾다* , 찾아가다, 채우다, 축하다, 춤추다, 취하다, 치다, 치료하다, 칠하다
ㅋ	캐다, 켜다, 키우다
ㅌ	타다, 터뜨리다
ㅍ	파티하다, 팔다, 팔리다, 퍼지다, 포기하다, 포장하다, 풀다* , 피다, 피우다
ㅎ	하다, 하품하다, 할다, 흐르다, 흔들리다

*목표 동사(22개).

Appendix 2. 교과서 간 중복 텍스트 목록 (제재명 가나다 순)

연번	저자	제재명	출판사
1	김상헌	가노라 삼각산아	미래엔, 비상(김)
2	최재천	개미 제국의 발견	비상(한), 신사고(민)
3	송재경 외	거침없이 하이킥	미래엔, 신사고(민)
4	장영희	관찰아	미래엔, 신사고(우), 창비
5	고은	그 꽃	미래엔, 비상(한)
6	심훈	그날이 오면	미래엔, 비상(한)
7	곽재구	그림 엮서	미래엔, 비상(한)
8	금난새	금난새의 클래식 여행	비상(한), 신사고(민)
9	유치환	깃발	신사고(민), 신사고(우)
10	이직	까마귀 검다 하고	신사고(우), 창비
11	영천 이씨	까마귀 싸우는 골에	미래엔, 창비
12	박완서	꽃씨에게 보내는 갈채	미래엔, 신사고(우)
13	류강하 외	꽃 피는 봄이 오면	미래엔, 신사고(민)
14	한용운	나룻배와 행인	교학사, 미래엔, 비상(김)
15	현덕	나비를 잡는 아버지	신사고(민), 창비
16	조세희	난쟁이가 쏘아 올린 작은 공	교학사, 미래엔
17	최일남	노새 두 마리	미래엔, 비상(한), 신사고(민), 신사고(우)
18	민예지 외	달리는 차은	미래엔, 비상(김), 비상(한)
19	박경화	도시의 밤은 너무 눈부시다	비상(김), 신사고(우)
20	윤석중	돌과 물	미래엔, 비상(한)
21	김영랑	들담에 속삭이는 햇발	비상(김), 신사고(민), 신사고(우), 창비
22	김유정	동백꽃	교학사, 미래엔, 비상(김), 신사고(민), 신사고(우)
23	황진이	동짓달 기나긴 밤을	비상(김), 비상(한)
24	미상	두꺼비 파리를 물고	교학사, 비상(김), 신사고(우)
25	법정	먹어서 죽는다	교학사, 비상(김)
26	김소월	먼 후일	미래엔, 비상(한), 신사고(우)
27	이효석	메밀꽃 필 무렵	교학사, 미래엔
28	백석	멧새 소리	비상(김), 비상(한)
29	기 드 모파상	목걸이	신사고(민), 창비
30	조항미	못난 사과	교학사, 비상(김)
31	정희성	민지의 꽃	비상(김), 신사고(우)
32	미상	박씨전	미래엔, 비상(김), 비상(한)
33	윤동주	반딧불	비상(김), 비상(한)
34	윤오영	방망이 깎던 노인	비상(한), 신사고(우)
35	나희덕	배추의 마음	신사고(우), 창비
36*	김구	백범일지	신사고(민)
37	한훤구	보리	미래엔, 신사고(민), 신사고(우)
38	정호승	봄 길	비상(한), 신사고(민)
39	앨빈 토플러, 하이디 토플러	부의 미래	교학사, 미래엔
40	박제가	북학의	교학사, 비상(김), 신사고(우)
41	황인숙	비	비상(김), 창비
42	주요섭	사랑손님과 어머니	교학사, 미래엔, 비상(김), 창비
43	홍량	산버들 골라 꺾어	비상(한), 창비
44	김지하	새봄	비상(김), 비상(한)
45	윤동주	서시	교학사, 비상(한), 신사고(우), 창비
46	이시영	성장	미래엔, 비상(김), 신사고(민)
47	김종길	성탄제	신사고(민), 신사고(우)
48	세종대왕	세종어제훈민정음	교학사, 미래엔, 비상(김), 신사고(민), 신사고(우)
49	황순원	소나기	교학사, 비상(한)
50	전성태	소를 죽다	미래엔, 비상(김)
51	오경희	소음 공해	신사고(민), 신사고(우)
52	하근찬	수난시대	교학사, 비상(김), 신사고(민), 창비
53	미상	시집살이 노래	교학사, 비상(김), 신사고(우)
54	유병록	식구	미래엔, 신사고(우), 창비

(Continued to the next page)

Appendix 2. Continued

연번	저자	제재명	출판사
55	최은숙	심재현, 사랑해	비상(김), 창비
56	미상	아기장수 우투리	교학사, 비상(한), 신사고(우)
57	한국방송광고공사	안과 밖이 다른 가족	비상(김), 신사고(민)
58	박지원	양반전	교학사, 신사고(민)
59	생텍쥐페리	어린 왕자	비상(김), 비상(한), 창비
60	이순원	어머니는 왜 숲 속의 이슬을 떨었을까	미래엔, 신사고(민)
61	기형도	엄마 걱정	교학사, 미래엔, 비상(김), 비상(한), 신사고(우), 창비
62	김소월	엄마야 누나야	교학사, 신사고(민)
63	이문구	열보다 큰 아홉	신사고(민), 창비
64	오주석	옛 그림 읽기의 즐거움	교학사, 비상(김), 비상(한), 신사고(우)
65	윤선도	오우가	미래엔, 신사고(민)
66	김려령	완득이	미래엔, 신사고(우), 창비
67	양귀자	우리 동네 예술가 두 사람	교학사, 미래엔
68	안도현	우리가 눈발이라면	미래엔, 비상(김)
69	이문열	우리들의 일그러진 영웅	비상(김), 신사고(우)
70	현진건	운수 좋은 날	교학사, 미래엔, 비상(김), 신사고(민)
71	경민	울림이 있는 말	미래엔, 비상(김)
72	양귀자	원미동 사람들	비상(김), 비상(한), 신사고(우), 창비
73	정약용	유배지에서 보낸 편지	비상(김), 신사고(우)
74	이규보	이옥설	교학사, 신사고(민)
75	정병욱	잊지 못할 운동주	교학사, 신사고(우)
76	박완서	자전거 도둑	교학사, 비상(한), 신사고(민), 창비
77	김광섭	저녁에	비상(한), 신사고(민)
78	정성춘	'저출산 고령화가 왜 경제성장을 방해할까요?	교학사, 신사고(민)
79	정인지	정인지 서문	비상(한), 신사고(민), 신사고(우)
80	김무곤	종이 책 읽기의 즐거움	교학사, 창비
81	황동규	즐거운 편지	신사고(민), 신사고(우)
82	김소월	진달래꽃	비상(김), 신사고(민), 창비
83	이정향	집으로	신사고(민), 신사고(우)
84	페터 빅셀	책상은 책상이다	미래엔, 신사고(민), 신사고(우), 창비
85	민태원	청춘 예찬	미래엔, 신사고(우)
86	이육사	청포도	비상(한), 신사고(민), 창비
87	최만리 외	최만리 등의 상소문	교학사, 신사고(우)
88	미상	춘향전	교학사, 비상(김), 신사고(민)
89	윤대성	출세기	비상(김), 비상(한), 창비
90	서거정	태평화화골계전	교학사, 비상(김), 창비
91	미상	토끼전	교학사, 비상(김), 신사고(우)
92	박상천	통사론	교학사, 비상(한)
93	국립국어원	표준 언어 예절	비상(한), 신사고(민)
94	국립국어원	표준어 규정	비상(김), 비상(한), 신사고(민), 신사고(우)
95	앤드루 클레먼츠	프린들 주세요	미래엔, 비상(김), 신사고(우)
96	현덕	하늘은 맑건만	미래엔, 신사고(우)
97	김형배	한글, 모든 언어의 꿈	비상(김), 신사고(우)
98	오승희	할머니를 따라간 메주	미래엔, 비상(김), 신사고(우)
99	정호승	항아리	교학사, 신사고(민)
100	박두진	해	교학사, 비상(한), 신사고(우)
101	정지용	해바라기 씨	미래엔, 신사고(민)
102	이규태	헛기침으로 백 마디 말을 한다	비상(한), 신사고(민)
103	헬렌 켈러	헬렌 켈러 자서전	미래엔, 신사고(우)
104	허균	홍길동전	교학사, 미래엔, 비상(한), 신사고(민), 신사고(우), 창비
105	정철	훈민가	미래엔, 비상(한), 신사고(민)
106	미상	홍부전	교학사, 비상(한)
107	공선옥	힘들다, 힘들어	신사고(민), 창비

*교과서 내 중복.

Appendix 3. 설문 문항: 동사 가나다 순, 의미거리 최대-최소 순

연번	동사	'주격 행동주-동사' 통사구조		'목적격 피동주-동사' 통사구조	
		설문 문항	의미거리	설문 문항	의미거리
1	가르치다	어머니가 (~을) 가르치다	0.504	(~가) 공부를 가르치다	0.580
		뚜껑이 (~을) 가르치다	-0.034	(~가) 폭풍을 가르치다	0.033
2	감다	숨소리가 (~을) 감다	0.505	(~가) 털실을 감다	0.557
		문서가 (~을) 감다	-0.002	(~가) 확장을 감다	-0.008
3	깎다	지하수가 (~을) 깎다	0.341	(~가) 암벽을 깎다	0.447
		통신이 (~을) 깎다	0.010	(~가) 성공을 깎다	0.004
4	깨다	노인이 (~을) 깨다	0.283	(~가) 예상을 깨다	0.487
		종류가 (~을) 깨다	-0.001	(~가) 죄인을 깨다	0.027
5	꼬집다	며느리가 (~을) 꼬집다	0.468	(~가) 엉덩이를 꼬집다	0.458
		이용이 (~을) 꼬집다	0.009	(~가) 임무를 꼬집다	0.005
6	돕다	아내가 (~을) 돕다	0.338	(~가) 이해를 돕다	0.422
		값이 (~을) 돕다	-0.003	(~가) 흔적을 돕다	0.074
7	두드리다	스님이 (~을) 두드리다	0.292	(~가) 목탁을 두드리다	0.540
		견해가 (~을) 두드리다	-0.015	(~가) 후손을 두드리다	0.002
8	때리다	축구공이 (~을) 때리다	0.363	(~가) 뺨을 때리다	0.632
		관련이 (~을) 때리다	0.003	(~가) 확장을 때리다	-0.001
9	막다	장사꾼이 (~을) 막다	0.378	(~가) 역류를 막다	0.565
		명칭이 (~을) 막다	-0.003	(~가) 생활을 막다	0.034
10	먹다	가축이 (~을) 먹다	0.630	(~가) 찌개를 먹다	0.735
		작성이 (~을) 먹다	0.057	(~가) 요청을 먹다	0.049
11	바꾸다	채식이 (~을) 바꾸다	0.288	(~가) 이름을 바꾸다	0.517
		신뢰가 (~을) 바꾸다	0.052	(~가) 약초를 바꾸다	0.076
12	읽다	중고생이 (~을) 읽다	0.481	(~가) 책을 읽다	0.698
		진출이 (~을) 읽다	0.059	(~가) 확보를 읽다	0.043
13	잡다	사냥꾼이 (~을) 잡다	0.528	(~가) 살바를 잡다	0.627
		작성이 (~을) 잡다	0.068	(~가) 불편을 잡다	0.084
14	지르다	아주머니가 (~을) 지르다	0.359	(~가) 고함을 지르다	0.565
		함량이 (~을) 지르다	0.013	(~가) 혼동을 지르다	0.032
15	찍다	누나가 (~을) 찍다	0.289	(~가) 사진을 찍다	0.703
		명령이 (~을) 찍다	0.023	(~가) 명령을 찍다	0.024
16	찢다	발소리가 (~을) 찢다	0.489	(~가) 껍질을 찢다	0.454
		확보가 (~을) 찢다	-0.013	(~가) 수요를 찢다	0.033
17	찾다	주인이 (~을) 찾다	0.456	(~가) 기사문을 찾다	0.540
		전체가 (~을) 찾다	0.053	(~가) 강세를 찾다	0.118
18	풀다	친구가 (~을) 풀다	0.363	(~가) 결박을 풀다	0.516
		비중이 (~을) 풀다	0.005	(~가) 면허를 풀다	0.053

국문초록

신경언어장애군의 동사 중재 자극 개발을 위한 기계학습 및 빅데이터 기반 한국어의 주격 및 목적격 명사 분석

박채원¹ · 임윤섭² · 성지은¹

¹이화여자대학교 언어병리학과, ²한국과학기술연구원 지능로봇연구단

배경 및 목적: 본 연구는 신경언어장애군의 동사 중재 자극으로 쓰이는 동사의 주격 및 목적격 명사의 한국어 특징을 빅데이터 기반으로 분석하였다. 또한 분석한 원자료는 향후 동사 중재 관련 연구에 활용할 수 있도록 클라우드에 배포하여 공유하였다. **방법:** 교과서 말뭉치에서 목표 동사와 결합한 주격 및 목적격 명사 간 출현빈도수 및 유형수 차이를 분석하여 주어생략현상을 확인하였다. 또한 목표 동사로부터 주격 및 목적격 명사까지의 의미거리를 기계학습으로 산출하여 의미거리 분포 차이를 분석하였다. 더불어 기계학습으로 산출한 의미거리와 한국어 화자들이 인식하고 있는 의미거리 간 상관관계를 살펴보았다. **결과:** 주격 명사는 목적격 명사보다 유의하게 출현빈도수 및 유형수가 낮았으며, 동사와의 기계학습 기반 의미거리가 더 멀었다. 기계학습으로 산출한 의미거리는 행동데이터와 강한 정적 상관관계를 보였다. **논의 및 결론:** 본 연구에서 밝힌 한국어의 주어생략현상은 한국어를 사용하는 신경언어장애군의 특징을 설명하는 근거자료로 활용될 수 있다. 주격 명사와 동사 간 기계학습 기반 의미거리가 멀다는 결과는 앞으로 한국어를 사용하는 신경언어장애군을 위한 동사 중재에서 주격 명사의 활용 방안이 재고될 필요가 있음을 시사하였다. 더불어 기계학습으로 산출한 의미거리를 중재 자극 선정에서 하나의 기준으로 사용할 수 있는 가능성을 확인하였으나, 차후 정밀한 추가 검증이 필요하다.

핵심어: 말뭉치 분석, 기계학습, 주어 생략, 의미거리, 동사 중재

본 연구는 2019년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(No. 10077553).

본 논문은 제1저자의 석사학위논문을 발췌 및 수정한 것이다.

참고문헌

- 곽은정, 성지은, 김연희, 전희정(2014). 동사의미역강화중재가 실어증 환자의 동사 및 명사 이름대기에 미치는 효과. *Communication Sciences & Disorders*, 19(1), 89-98.
- 김가영(2006). 실어증 환자의 동사 산출과 논항구조 간의 관계. 연세대학교 대학원 석사학위논문.
- 김수련, 박창일, 김덕용, 황민아(2004). 브로카 실어증 환자의 과제 간 명사와 동사의 산출 비교. *언어청각장애연구*, 9(2), 1-18.
- 김수정(2016). 문어 텍스트의 장르별 주어의 실현 양상. *한민족어문학*, 72, 25-62.
- 김수정, 최동주(2013). 소설 텍스트에서의 주어의 실현 양상: 박완서의 소설 '그 가을의 사흘 동안'을 중심으로. *한민족어문학*, 64, 37-69.
- 김주희(2008). 실어증 환자의 중증도와 유형에 따른 동사 이해에 관한 연구. 명지대학교 대학원 석사학위논문.
- 김지인, 성지은, 이수은, 심현섭(2018). 문장 성분 간 거리에 따른 시제 표지 처리 과정에 대한 청년층 및 노년층의 ERP 성분 비교. *Communication Sciences & Disorders*, 23(1), 129-145.
- 김하나, 성지은(2014). 노화에 따른 이야기 다시 말하기 수행력 및 작업기억과의 상관관계 연구. *특수교육*, 13(3), 7-24.
- 김혜림, 성지은(2016). 실어증 환자의 통사적 복잡성에 따른 답화속 문장이해 능력 비교. *Communication Sciences & Disorders*, 21(4), 668-682.
- 남기심(2004). *현대 국어 통사론*. 서울: 태학사.
- 남기심, 고영근(1993). *표준국어문법론*. 서울: 탑출판사.
- 남승호(2007). *한국어 술어의 사건 구조와 논항 구조*. 서울: 서울대학교출판부.
- 민오영, 강정숙, 민경옥, 황영진(2011). 문장정오판단과제로 살펴본 실어증 유형에 따른 문장이해 특성연구. *특수교육재활과학연구*, 50(2), 369-386.
- 민중서관(1967). *한영사전*. 서울: 민중서관.

- 박경아, 김향희, 박은숙, 신지철(2006). 실어증 환자의 문법형태소 산출 특성. *언어청각장애연구*, 11(1), 51-63.
- 박정희(2012). 한국어와 영어의 생략 현상에 대한 통계적 접근. *어문논집*, 66, 171-191.
- 성지은, 곽은정(2012). 연령 및 동사 논항 구조에 따른 애니메이션을 활용한 동사 이름대기 과제 수행력 차이. *언어청각장애연구*, 17(4), 550-564.
- 손병룡(2001). 사고와 문화가 언어에 미치는 영향. *현대영어영문학*, 45(2), 131-152.
- 신상은(2008). 알츠하이머성 치매환자의 중증도에 따른 동사 표현 및 이해 능력에 관한 연구. 이화여자대학교 대학원 석사학위논문.
- 안태선(2002). 브로카 실어증 환자의 문법 형태소 산출. 연세대학교 대학원 석사학위논문.
- 엄보라, 성지은(2015). 구문복잡성을 활용한 작업기억 중재가 실어증 환자의 문장이해 능력에 미치는 효과. *한국언어치료학회 학술발표대회 논문집*, 204-205.
- 윤정미, 김영태(2002). 브로카실어증 환자들의 동사이해결함의 범주특정적 특색. *언어청각장애연구*, 7(3), 39-54.
- 이나라(2017). 한국어의 1·2인칭 주어 생략 현상에 대한 재고: 명시적 주어 표현의 화용적 의미를 중심으로. *담화와 인지*, 21(2), 145-163.
- 이옥분, 이상희(2007). 이야기 되말하기 방법이 실어증자의 어휘 산출력과 문장 표현력에 미치는 영향: 사례 연구. *언어치료연구*, 16(2), 147-159.
- 이지연, 권미선, 나혜리, 바스티안세로엘린, 탐슨신디아(2013). 비유창성 실어증 환자의 시제표시 이해와 산출 능력. *Communication Sciences & Disorders*, 18(2), 139-151.
- 이지혜(2017). 시제 중심 동사강화 CILT 중재가 실어증 환자의 동사산출에 미치는 치료 및 일반화 효과. 이화여자대학교 대학원 석사학위논문.
- 정규현, 성지은(2018). 의미역 유형에 따른 실어증 환자의 동사 및 의미역 처리 능력과 중증도 간의 관계. *Communication Sciences & Disorders*, 23(2), 337-346.
- 정현선(2009). 브로카 실어증 환자의 문법형태소 산출: 조사, 피동·사동 접사를 중심으로. 단국대학교 대학원 석사학위논문.
- 최수진, 성지은(2014). 과제 제시 유형 및 논항 수에 따른 노년층의 동사 이름대기 수행력 차이. *재활복지*, 18(4), 279-293.
- 최수진, 성지은, 정지향, 김건하(2016). 동사 운반구를 활용한 작업기억중재가 경도인지장애군의 언어처리과정에 미치는 효과. *Communication Sciences & Disorders*, 21(3), 524-537.
- 최은정, 성지은, 정지향, 곽은정(2013). 경도인지장애군의 명사 및 동사 하위유형에 따른 대면이름대기 산출능력. *대한치매학회지*, 12(2), 41-46.
- 현정문, 김향희, 신지철, 서상규(2003). 베르니케실어증과 브로카실어증 환자들의 명사와 동사 인출 비교. *언어청각장애연구*, 8(3), 171-187.
- 황순희(2014). 동작명명하기 과제에 나타난 알츠하이머형 치매환자와 일반 아동의 분리성 동사 사용. *언어과학*, 21(1), 235-262.
- 황순희, 최진남(2013). 알츠하이머형 치매환자의 동사 사용 양상 분석. *언어과학*, 20(3), 143-173.

ORCID

박채원(제1저자, <https://orcid.org/0000-0003-0912-5590>); 임윤섭(공동저자, <https://orcid.org/0000-0003-4754-6038>); 성지은(교신저자, <https://orcid.org/0000-0002-1734-0058>)