Analyzing the Disfluency of Reading Tasks of Persons Who Stutter Based on Deep Learning and Word Embedding

Sanghoun Song^a, HeeCheong Chon^b, Soo Bok Lee^c

^aDepartment of Linguistics, Korea University, Seoul, Korea

^bDepartment of Speech-Language Pathology, Chosun University, Gwangju, Korea

^cDepartment of Speech-Language Therapy & Aural Rehabilitation, Woosong University, Daejeon, Korea

Correspondence: Soo Bok Lee, PhD

Department of Speech-Language Therapy & Aural Rehabilitation, Woosong University, 171 Dongdaejon-ro, Dong-gu, Daejeon 34606, Korea
Tel: +82-42-630-9222
Fax: +82-42-630-9229
E-mail: sblee@wsu.ac.kr

Received: July 20, 2020

Revised: August 29, 2020

Accepted: September 7, 2020

Objectives: Recent natural language processing systems employ embedding techniques, which convert linguistic expressions into numerical vectors in order to measure the geometric distance between expressions. Using skills and focusing on the reading tasks, the present study aims to reveal the distributional properties of disfluencies. **Methods:** The current work segmented the reading data of 110 adolescents and adults who stutter, transformed the data into a vector space, and then conducted the embedding calculation. Utilizing Word2Vec, the cosine similarity was measured so as to look at how the types of disfluencies were co-related to each other. **Results:** The eight ND (Normal disfluencies) and AD (Abnormal disfluencies) types, excluding the R2 (Repetition 2) and DP (Disrhythmic Phonation) types, were close to each other with respect to the cosine similarity (>.9). In particular, the AD types such as Ha (Abnormal hesitation), Ia (Abnormal interjection), URa (Abnormal unfinished/revision word), and R1a (Abnormal Repetition1) largely overlapped with each other. R2 and DP showed different distributional properties from other types of disfluencies. The results also indicated that each ND and AD pair seldom differed in their distributional properties. Finally, this study it found that several consonants tended to appear more often when the speakers produced disfluencies. **Conclusion:** This study draws the distributional patterns of fluency disorders in an automatic way using deep learning skills. The findings are of use for the diagnosis and treatment of the fluency disorders.

Keywords: Deep learning, Word embedding, Cosine similarity, Stuttering reading task, Normal disfluencies, Abnormal disfluencies

인공지능 및 딥러닝은 현재 학계를 비롯하여 우리가 살아가는 실 세계에서 빠지지 않는 주요 키워드이다. 딥러닝에 기반한 인공지능 시대가 본격화되면서 언어병리학 분야도 딥러닝 기술을 활용하여 새로운 방향을 개척하고자 하는 융합형 연구 모형이 제시되고 있다. Park, Lim과 Sung (2019)은 신경언어장애에 영향을 끼치는 것으로 드러난 한국어의 통사적 특질을 워드임베딩 기법으로 확인하는 연구를 진행하였고, Kourkounakis, Hajavi와 Etamad (2020)는 순환신경망의 일종인 LSTM 모델을 활용하여 영어 비유창성을 판별하는 시스템을 개발하였다. 이러한 연구가 주목할 만한 선행연구 이기는 하나, 언어병리학 데이터 자체가 내재한 패턴을 찾아내어 그

안에 내재된 함의를 도출하는 데까지는 논의가 미치지 못하였다.

본 연구는 본격적으로 딥러닝 기반의 자연어처리 기술과 한국어인 언어병리학 데이터를 접목하여 실제 말더듬과 관련된 주요 의제를 살피고자 한다. 이를 통해, 그 결과가 기존 유창성장애 연구 전반에서 분석된 바와 크게 배치되지 않으면서 동시에 새로운 연구문제를 양산할 수 있음을 입증하고자 한다. 적용하는 기술 체계는 크게 Mikolov, Chen, Corrado와 Dean (2013)에서 제안된 Word2Vec을 준용하였으며, gensim 라이브러리를 이용한 시각화 및 코사인 유사도 측정 분석법을 활용하였다(Rehurek & Sojka, 2010).

딥러닝의 가장 큰 장점은 한정적인 데이터를 활용하여 큰 그림을

제시할 수 있다는 것이다. 컴퓨터를 활용하기 때문에 데이터가 많아질수록 더 진가를 발휘하게 될 것이다. 언어병리학 분야도 데이터를 축적해 가는 과정에 있기 때문에 컴퓨터를 활용한 분석이 필요하며, 그 준비의 차원에서 언어병리학 데이터에 대한 딥러닝 분석의 경험 역시 축적되어야 할 것이다. 물론 언어병리학 연구와 관련하여 딥러닝을 사용하는 것의 한계가 없는 것은 아니다. 일반적인 언어병리학 연구에서는 추론 통계에 따라 논지를 기술하는 반면, 딥러닝은 유의성(p value) 검증을 사용하지 않는 까닭에 그 결과로 어떠한 연구가설을 직접적으로 확인할 수 없다. 즉, 연구결과를 바탕으로 어떠한 연구문제에 대하여 “예/아니오” 형식의 대답을 곧바로 내놓을 수는 없다는 한계를 가진다. 비슷한 맥락에서 딥러닝은 단일 모델을 가정하는 까닭에 연구에 있어서 대조군이 명확하지 않다는 특성이 있다. 대부분의 과학적 연구는 실험군과 대조군을 나누어 비교를 통한 통계분석을 통해 입증을 취하는 데 비해, 딥러닝은 본질적으로 과학적 검증을 상정하지 않기 때문이다. 또한, 딥러닝의 작동방식 자체가 가지는 문제점도 존재한다. 전통적인 과학 연구는 재현 가능성을 중요시하는데 비해 딥러닝 모델은 학습을 달리할 때마다 결과치가 갱신되곤 한다는 점에서 완전히 재현 가능하다고 볼 수는 없다. 정리하자면, 딥러닝 분석은 전통적인 언어병리학의 과학적 방법론과는 차이점을 보인다. 그럼에도 언어병리학의 새로운 연구 주제와 방법론을 탐색할 수 있다는 점에서 본 연구가 가지는 함의와 가치는 명확히 존재한다. 유의성 검증을 가정하지 않아 p value 등의 지표를 도출할 수 없는 한계는 있으나, 그 결과를 보면 대개 어떠한 경향성을 내포한다. 즉, 전통적인 방식을 택하지 않았지만, 딥러닝이 어떠한 현상의 일반적 속성을 파악하는 능력이 탁월하다는 것은 이미 여러 사례를 통해 입증된 사실이다(Goodfellow, Bengio, & Courville, 2016). 또한, 실험군-대조군으로 구성된 분석틀을 따르지 않은 측면에서 과학적 설명력이 다소 취약할 수 있으나, 딥러닝 분석과 구현이 가지는 실효성과 발전 가능성은 이미 알파고를 비롯한 최근의 수많은 실례에서 확인할 수 있다. 연구 결과를 확장하여 실제 애플리케이션을 개발하거나 말더듬 평가 시스템을 구현할 때를 가정하면 딥러닝 환경이 가지는 장점을 결코 무시할 수 없을 것이다. 또한, 기존의 이론을 딥러닝-워드 임베딩이라는 다른 방법론을 통해 재분석하고 재검토하는 것만으로도 분명한 가치가 있다고 판단한다. 지금까지 구성된 언어병리학의 이론적 토대와 경험적 발견이 딥러닝 분석에서도 그대로 일치하는지 혹은 그렇지 않은지를 확인하는 과정은 시의적절한 시도가 될 것이다.

유창한 말의 흐름을 방해하는 비유창성(disfluencies)은 언어병리학에서 오랫동안 다루어 온 주제이다. 특히 비유창성은 말더듬

(stuttering) 분야에서 장애의 본질과 특성을 파악하는 데 도움을 줄 수 있을 뿐 아니라 임상현장에서 대상자의 말더듬 장애 유무를 정확하게 진단하고 치료 효과를 검증하는 데 영향을 미치기 때문에 꾸준한 연구가 이루어졌다. 또한 비유창성은 말을 더듬는 사람에게만 나타나는 것이 아니므로 말더듬는 사람과 일반인의 비유창성의 차이를 찾아내려는 시도가 꾸준히 이루어졌다(Lee & Sim, 2003). 말더듬는 사람과 일반인의 비유창성을 어떻게 구별할 것인가에 대해서는 연구자들 사이에서 상당한 논쟁이 있어 왔고 이런 시각 차이는 비유창성 범주 및 유형 분류에 영향을 미쳤다(Kolk & Postma, 1997; Wingate, 1988; Zebrowski, 1994). 말더듬는 사람과 일반인의 유창성 문제는 상호배타적이라고 주장하기도 하지만(Hamre, 1992), 대부분의 연구자는 유창성 문제가 동일한 연속선상에 있는 것으로 간주하며(Bloodstein, 1992; Starkweather, 1992), 말더듬는 사람과 일반인 모두 동일한 말산출 체계를 사용하고 말산출 본질에서는 동일한 점이 있을 것이라는 가정을 한다(Manning, 2010). 그러나 말더듬는 사람과 일반인이 보이는 비유창성은 표면적으로 양적 차이뿐만 아니라 질적인 차이를 보이기 때문에(Ambrose & Yairi, 1999; Johnson & Associates, 1959; Yaruss, LaSalle, & Conture, 1998) 이러한 차이를 고려하여 대부분의 선행연구는 비유창성을 크게 일반인과 말더듬는 사람 모두에게서 관찰되는 비유창성 범주와 말더듬는 사람에게서 주로 관찰되는 비유창성 범주로 나누고(Campbell & Hill, 1987; Conture, 1990; Meyers, 1986; Yairi, Ambrose, Paden, & Throneburg, 1996), 말더듬는 사람들에게서 주로 관찰되는 비유창성의 특성을 보고하였다.

먼저, 언어학적 측면에서 비유창성 분류를 살펴보면 다음과 같다. Conture (1990)는 ‘날말’이라는 언어학적 단위를 기준으로 하여 말더듬는 사람이 주로 보이는 비유창성을 날말 내 비유창성(within-word disfluencies) 범주로, 장애 유무와 상관없이 모든 사람에게서 나타나는 비유창성을 날말 간 비유창성(between-word disfluencies) 범주로 분류하였다. 날말 내 비유창성에는 일음절날말전체반복, 음절/말소리반복, 연장, 막힘을 포함하였고, 날말 간 비유창성에는 다음절날말전체반복, 구반복, 수정, 간투사를 포함하였다. Zebrowski (1994)는 비전형적 비유창성(atypical disfluency)과 전형적 비유창성(typical disfluency)으로 나누고, 비전형적 비유창성에는 음절반복, 말소리반복, 연장, 막힘을 포함하였으며, 전형적 비유창성은 주저, 간투사, 미완성구, 수정, 구반복, 단어반복을 포함하였다. Yairi 등(1996)은 말더듬는 사람만이 보이는 특정한 비유창성을 규정짓기보다는 진성비유창성(stuttering-like disfluencies, SLD)과 가성비유창성(other disfluencies, OD)이라는 용어를 사용하여 날말부분반복(다음절날말일부반복, 음절반복, 음절부

분반복, 말소리반복), 일음절날말반복, 비운율적 발성(disrhythmic phonation [DP], 연장, 막힘, 깨진날말을 포함한 용어임)을 SLD로, 다음절날말전체반복, 구반복, 간투사, 미완성/수정 유형을 OD로 분류하였다. 이처럼 선행연구는 말더듬는 사람에게서 주로 관찰되는 비유창성 유형이 일음절날말반복, 날말부분반복, 연장, 막힘이라고 일관되게 보고하고 있으며, 말더듬는 사람과 일반인 모두에게서 관찰되는 비유창성 범주의 유형은 연구마다 세부 유형에 조금씩의 차이가 있긴 하지만 주저, 간투사, 미완성/수정, 구반복, 다음절날말전체반복이라고 보고하고 있다. 장애 유무와 상관없이 모든 사람에게서 관찰되는 비유창성 범주는 산출하고자 하는 언어를 계획하고 처리하는 과정에서 발생하기 때문에(Manning & Shirley, 1981; Navarro-Ruiz & Rallo-Fabra, 2001; Yoon & Chon, 2019) 이 범주를 언어(학)적 비유창성(linguistic disfluency) 혹은 미로(maze)이라고 하였다(Manning, 2010; Navarro-Ruiz & Rallo-Fabra, 2001).

또한, 연구자들은 비유창성에 동반되는 특성을 반영하여 말더듬과 일반을 구분할 수 있도록 비유창성 유형을 구분하기도 하였다. 예를 들어, Campbell과 Hill (1987)은 다른 선행연구들과 동일하게 말더듬는 사람에게서 관찰되는 비유창성을 덜 전형적인(less-typical) 범주로, 장애 유무와 상관없이 관찰되는 비유창성을 더 전형적인(more-typical) 범주로 나누었다. 하지만 일반인에게서도 일음절날말반복과 날말부분반복이 종종 관찰되는 것을 고려하여 추가적인 기준을 비유창성 범주의 구분에 포함하였다. 즉, 선행연구와 동일하게 말소리반복, 연장, 막힘을 덜 전형적인 범주에, 주저, 간투사, 수정, 구반복을 더 전형적 범주에 포함시켰으나, 일음절날말반복과 부분날말반복의 경우 2회 이하로 반복되고 긴장이 동반되지 않은 경우는 덜 전형적인 범주에 포함하였고 3회 이상 반복된 경우 더 전형적인 비유창성 범주에 포함하였다. Van Riper (1982)는 음절반복의 경우 날말당 빈도, 100날말당 빈도, 속도 및 규칙성, 중성모음의 출현 여부 등 다양한 측면을 고려하여 이 유형이 정상적인 비유창성 범주에 포함되는지 비정상적인 비유창성에 포함되는지를 구분하게 하였다.

국내의 경우, Sim, Shin과 Lee (2010)가 비유창성을 정상적 비유창성(normal disfluencies, ND)과 비정상적 비유창성(abnormal disfluencies, AD)의 두 범주로 분류하였다. ND 범주는 주저(Hesitation, H), 간투사(Interjection, I), 미완성/수정(Unfinished/Revision word, UR), 반복1(Repetition1, R1: 다음절날말전체반복, 구/어절반복) 등을 포함하였고, AD 범주는 비정상적 주저(Abnormal hesitation, Ha), 비정상적 간투사(Abnormal interjection, Ia), 비정상적 미완성/수정(Abnormal unfinished/revision word, URa), 반

복2 (Repetition2, R2: 말소리/음절/음절부분/날말부분반복, 일음절날말반복), 비운율적 발성(Disrhythmic Phonation, DP) 등이다. 이 분류 방법은 기존의 선행연구들과 다른 양상을 보인다. 즉, 선행연구에서 말더듬 유무와 상관없이 관찰된다고 보고한 언어적 비유창성이라고 해도 3회 이상 반복되거나 시각적 긴장과 같은 질적 양상이 동반되면 말더듬는 사람에게 주로 관찰되는 AD 범주에 포함시켰다. 이는 정도의 차이는 있으나 AD와 ND에 속하는 비유창성 유형이 말더듬는 사람뿐만 아니라 일반인 모두에게서 관찰될 수 있으며, 유형 자체만으로 정상과 비정상 여부를 판단할 경우 말더듬 문제를 과소평가할 수 있기 때문이라고 하였다(Ha & Sim, 2008; Sim et al., 2010). 비유창성 분류와 관련된 선행연구들을 기반으로 하여 비유창성 유형들 사이의 분포적 유사성 혹은 차이점을 찾는 것은 현재 비유창성 유형들을 두 범주로 분류하는 방법의 타당성을 검증해볼 수 있을 뿐 아니라 비유창성 범주를 나누는 또 다른 기준을 제시해 줄 수 있을 것으로 보인다.

비유창성의 특성 중 하나는 비유창성이 2개 이상으로 군집(cluster)화 되어 나타나는 경향을 보인다는 것이다(Hubbard & Yairi, 1988; LaSalle & Conture, 1995; Penttilä, Korpiaakko-Huuhka, & Kent 2019; Silverman, 1973). 군집화는 인접해 있는 낱말 간에 혹은 하나의 낱말 내에서 비유창성이 연속적으로 발생하거나 연관 있는 비유창성 유형 간에 근접하여 앞 또는 뒤에 나타나는 것을 말한다. 비유창성 유형 간의 근접성을 확인하기 위한 Jeon과 Jeon (2015)의 연구에서 말더듬 성인은 혼합유형(OD와 SLD), OD유형(OD와 OD), SLD유형(SLD와 SLD) 순으로 비유창성 군집을 산출하였으며, OD유형에서는 간투사가 가장 높았고, SLD유형에서는 단어부분반복과 비운율적 발성이 가장 높은 비율로 나타났다. 비유창성 군집이 발생하는 원인은 다양할 수 있다. 말 산출을 위한 말운동 기제의 문제 때문일 수 있고(Hubbard & Yairi, 1988), 언어처리의 어려움으로 인해서 발생할 수도 있다(Wexler & Mysak, 1982). Bóna (2019)는 메시지 생성과 언어학적 계획의 두 관점을 사용하여 비유창성 군집을 설명하기도 하였다. 즉, 메시지 생성이 어려운 경우(무엇을 말해야 할지 모를 때), 언어학적 계획이 어려운 경우(하고 싶은 말은 있으나 언어적으로 그 내용을 구성하는 것이 어려울 때), 혹은 메시지 생성과 언어학적 계획이 둘 다 잘 안될 때 비유창성 군집이 발생한다고 하였다. 비유창성 군집은 말을 더듬지 않는 사람에게서도 관찰되기는 하지만 말을 더듬는 사람이 말을 더듬지 않는 사람보다 비유창성 군집의 빈도가 유의하게 높았으며, 비유창성 군집 유형의 특성도 다른 것으로 나타났다(Hubbard & Yairi, 1988; Jeon & Jeon, 2015; Sawyer & Yairi, 2010). 또한 Yairi와 Seery (2015)는 낱말 내에서 관찰되는 비유창성 군집은 외현적인

말더듬 중증도와 정적인 상관성이 있으며, 만성적 말더듬과 상관성이 있다고 주장하였으며, Hubbard와 Yairi (1988)는 비유창성 군집을 이루는 비유창성 유형들을 살펴보는 것은 ‘비유창성이 다른 비유창성을 유발하는 이유’에 대한 통찰력을 제공할 수 있다고 하였다. 따라서 말더듬은 사람의 발화를 사용하여 비유창성 군집의 특성을 파악하고 비유창성 유형 간의 근접성 또는 유사성을 분석하는 것은 말더듬의 진단 목적뿐만 아니라 각 비유창성 유형의 발생 원인을 파악하는 데 도움을 줄 수 있다(Yairi & Ambrose, 2005). 비유창성 군집들을 토대로 산출한 비유창성 밀도와 이러한 군집을 형성하는 비유창성의 유형은 임상적이고 이론적인 시사점을 준다. 예를 들어, 비유창성의 군집 밀도가 높을수록, 즉, 서로 근접하여 발생하는 비유창성의 수가 많을수록 더 심각한 말더듬을 반영한다고 예측할 수 있다.

언어, 인지 및 운동 기술과 같은 여러 요인이 복합적으로 작용하여 발생하는 말더듬은(Smith & Weber, 2017) 전통적으로 낱말의 길이 및 복잡성 등 말 산출에 영향을 주는 여러 언어학적인 변인들의 영향을 받는 것으로 보고되었다(Hubbard & Prins, 1994; Ronson, 1976). 다수의 선행연구는 말더듬이 첫 음소나 긴 단어 또는 익숙하지 않은 단어와 같은 언어 정보의 하중이 높거나 음운적인 복잡성이 높아 발음하기 어려운 경우에 자주 나타난다고 하였다(Bloodstein, 2002; Han, Lee, & Sim, 2005; Nippold, 2002). 또한 말더듬은 사람들은 어휘, 구문, 음운 및 음성 정보 처리에서 비전형적인 특성이 보고되었다(Byrd, Conture, & Ohde, 2007; Howell & Bernstein Ratner, 2018; Newman & Bernstein Ratner, 2007). 이런 현상은 구어의 실행뿐만 아니라 산출 전에 언어단위의 계획이나 조합부터 어려움이 반영되었음을 시사한다. 선행연구에서 음성학적으로 복잡한 단위산출과 말더듬 간의 관계를 규명하려는 시도가 많았다. 조음복잡성과 말더듬과의 관련성에 관한 연구에서는 조음복잡성이 증가된 음절 및 어절에서 말더듬이 증가되어 나타난다는 보고가 지배적이지만 모든 연구에서 동일한 결과를 보고하고 있지 않다(Kwon, 1991; Lee, Han, & Sim, 2004; Shin 1996). 일찍 발달하는 말소리나 조음적으로 덜 복잡하게 구성된 말소리에서 비유창성 발생빈도가 낮을 것으로 예상되나, 예상과 달리 비유창성 빈도가 더 높게 나타나거나 말더듬 대상자 별로 어려운 음소들이 각기 다른 특징을 보여주었다.

이러한 말더듬 현상은 내적수정가설(covert repair hypothesis [CRH]; Lee & Sim, 2003; Postma & Kolk, 1993)로 말더듬은 사람들이 말 산출을 위한 처리과정에서 의미적, 음성학적, 통사적 계획의 오류를 수정하면서 말더듬이 나타나는 것으로 설명하였다. 나아가 EXPLAN 모델(Howell & Au-Yeung, 2002; Howell, 2004)은

말더듬과 구문 프로세싱과의 관련을 주장하며, 언어계획(plan-ning, PLAN)과 운동실행(execution, EX)을 독립시켜 내적수정가설을 정교화하여 말더듬 현상을 설명하였다. 말산출은 통사적, 의미적, 음성학적 특징을 외부에 보이지 않게 순차적으로 계획하는(PLAN) 내적 인지-언어체계에 의해 시작되고, 운동과정을 통해 말산출이 조직화되고 실행된다. 유창한 발화는 운동체계가 언어학적 배열을 순서에 따라 실행할 때 산출되는 것으로 보았으며 비유창한 발화는 언어계획과 운동실행 사이의 일시적 비동기화로 나타난다고 보았다. 말더듬 아동이 주저 또는 반복같은 비유창성을 보이는 것은 언어계획을 위한 시간을 확보하기 위한 전략으로 보았다. Brocklehurst (2008)는 말더듬 아동과 일반 아동의 음운부호화 속도는 비슷하지만 말더듬 아동은 일반 아동과 다른 방식으로 어휘 정보를 조직화하면서 말을 더듬게 되며, 음운부호화의 과정을 모니터링하고 오류를 수정하면서 지속해서 말을 더듬는 것으로 보았다.

Chon (2010)은 비유창성에 대한 연구문제를 크게 비유창성의 원인, 비유창성의 유형, 그리고 비유창성의 위치로 삼분하였다. 딥러닝 기반 자연어처리 기술에 입각하는 본 연구에서는 이 가운데 유형과 위치에 집중하고자 한다. 이는 딥러닝이 일반적인 설명력이 우수하지만 어떠한 현상에 대한 원인을 직접 도출하지 못한다는 나름의 한계를 가지기 때문이다. 물론 각 연구자의 역량에 따라 딥러닝 분석 결과에서 후행 연구를 도출할 수 있으므로 딥러닝이 비유창성의 원인 분석에 전혀 이바지하는 바가 없다고는 볼 수 없다. 다만, 본 연구는 워드임베딩의 일차적 결과를 바탕으로 유형 및 위치에 대한 분석에 집중하고자 한다.

비유창성의 유형과 관련하여, 본 연구는 파라다이스-유창성검사 II (Paradise-Fluency Assessment II; Sim et al., 2010)의 기준에 따라 유창성장애의 정상적 비유창성(ND) 및 비정상적 비유창성(AD) 유형 10개가 가지는 분포적 유사성을 측정하고자 한다. 비유창성이 발생하는 데 있어서 어떠한 잠재적 원인이 존재한다고 가정한다면 서로 다른 유형의 비유창성이 발생하는 분포에도 어떠한 패턴이 발견될 것이다. 이는 비유창성의 유형을 자료 기반으로 재조직하고 분류하면서 동시에 앞서 말한 바와 같이 비유창성의 기저 원인을 짐작할 수 있도록 하는 장치가 된다. 더 나아가 현재 흔히 이루어지고 있는 ND와 AD의 구분이 분포적으로도 유의미한지 검토할 기회가 마련될 수 있을 것이다.

비유창성의 위치와 관련하여, Chon (2010)은 한국어의 말더듬 발생 위치를 밝히는 연구가 이루어진다면 말더듬의 발생을 예측할 수 있어 치료에 관한 중요 정보를 얻을 수 있다고 주장하였다. 문제는 단순 빈도에만 의존하는 기존 방식으로는 말더듬의 발생 위치를 예측하는 모델이 충분한 기능을 수행할 수 없다는 것이다. 단순

히 출현 빈도에 의존하는 방식이 가지는 한계에 관해서는 다른 연구에서도 공통으로 지적되는 사항이다. Ha와 Sim (2008)은 간투사의 사용과 관련한 분포적 특성에서 상당한 개인차가 발생하고 있음을 지적하고 이 경우 각 개인차를 하나의 정보로 결합하는 방식인 단순 빈도계수는 자칫 비유창성에 대한 과소평가 혹은 과대평가로 귀결될 위험성을 내포한다고 하였다. 이러한 측면에서 보면, 답리닝-워드임베딩은 기존 선행연구에서 수행해온 언어병리데이터에 대한 일차원적인 분석을 넘어선 종합적인 분석을 할 수 있게 하는 대안적 장치라고 할 수 있다.

본 연구의 연구문제를 정리하면 다음과 같다. 첫째, 비유창성의 유형과 관련하여, 비유창성 항목들 사이의 분포적 유사성이 존재하는가? 둘째, 비유창성의 유형과 관련하여, 정상적 비유창성과 비정상적 비유창성 항목들 사이의 분포적 차이가 존재하는가? 셋째, 비유창성의 위치와 관련하여, 음소 환경에 따른 비유창성의 분포 유형이 존재하는가?

연구방법

워드임베딩 개괄

컴퓨터를 이용한 언어 분석의 핵심 개념은 단어를 비롯한 언어 표현의 의미와 자질을 수치화하여 ‘계산’할 수 있도록 하는 것이다 (Boleda, 2020). 이는 구조주의 언어학자인 Harris (1954) 및 Firth (1957)에 의해 제창된 ‘분포가설’에 입각한 것이고, 그러한 전통은 ‘분포의미론’(distributional semantics)으로 이어지고 있다. 워드임베딩은 그 가운데 가장 잘 알려진 기법으로, 사람의 언어를 기계가 이해할 수 있도록 수치 ‘벡터’로 변환하여 공간상에서 연산을 도모하는 과정과 그 결과를 칭한다.

임베딩이라는 표현은 단어나 문장 각각을 벡터로 변환해 공간(vector space)에 ‘끼워넣는다(embed)’는 것이지만, 이는 공학적 해설일 뿐 일반 언어병리학 연구자들에게는 다소 낯선 개념이다. 임베딩은 한자어로는 ‘사상(寫像)’, 영어로는 ‘mapping’에 해당한다. 이를 집합론으로 설명하면 어떤 집합의 임의의 원소가 다른 집합의 하나의 원소에 대응할 때 그 두 집합 간의 대응 관계를 설정하는 것을 임베딩이라고 표현할 수 있다. 다시 말해, ‘워드임베딩’의 요체는 어떠한 단어가 가진 속성을 다른 공간에 펼쳐서 대응시킨다는 것이다. 이때 두 집합의 척도가 달라지는데, 본래 단어는 언어적 속성(예컨대, 의미)을 문자적으로 가지지만, 이 속성이 사상된 집합 공간은 수치 척도로 구성된다. 간단히 요약하면, 범주적 속성을 수치적 속성으로 변환하는 과정을 임베딩이라 할 수 있다.

워드임베딩을 통해 말더듬 데이터를 다룰 때 얻을 수 있는 이점

에 대해 논의하기 위해서는, 이른바 ‘분포표상(distributional representation)’에 대한 이해가 필요하다. 분포의미론은 ‘어떠한 단어의 의미는 그 주변 환경에 의해서 결정된다’는 가정에 입각한다. 예컨대, 한국어의 ‘다리’는 전형적인 동음이의어로서 영어의 ‘leg’ 또는 ‘bridge’에 대응할 것이다. 그러나 ‘다리가 아프다’에서 ‘다리’와 ‘다리가 지어졌다’에서 ‘다리’가 다른 의미라는 것을 쉽게 알 수 있는 것처럼, 실제 쓰임에 있어서 그 의미 구분은 그 주변 단어(‘아프다’ 혹은 ‘지어졌다’)들에 의해 선택된다. 이러한 가설이 최근의 답리닝 환경과 결합하여 나타난 것이 분포표상인데, 이는 단어의 의미와 구조를 ‘자질화’하여 n -차원의 ‘벡터(vector)’로 표상하는 것을 말한다. 벡터는 ‘숫자로 표현된 원소들로 구성된 리스트’라고 할 수 있는데, 좀 더 기술적으로 설명하면 $[0.5, 0.1, 0.05, 0.001, \dots]$ 과 같은 형식으로 구성된 ‘1차원 행렬(matrix)’을 말한다. 즉, 워드임베딩은 각 단어가 가진 언어적 자질을 수치로 표현된 리스트로 변환하는 과정을 의미한다.

여기에서 두 가지 질문이 가능한데 하나는 ‘언어적 정보를 어떻게 수치 행렬로 변환할 수 있는가’이며, 다른 하나는 ‘수치화하였을 때 어떠한 장점이 있는가’일 것이다. 먼저 수치 정보로의 변환 방식에 대하여 살펴보면 다음과 같다. 문자 혹은 음성으로 표현되는 언어를 수치화한다는 것은 코퍼스언어학에서 가장 기본이 되는 ‘빈도’를 대표적인 예로 들 수 있다. 이는 ‘어떤 단어가 많이 쓰였는가?’에 주목하는 것으로 흔히 백오브워즈(Bag of Words) 기법이라고 칭한다. 단순히 빈도를 측정하는 방식을 넘어 단어 연쇄의 n -gram을 측정하는 것도 통계적 패턴을 언어 이해에 활용하는 방법의 일종이며, 이는 ‘단어가 어떤 순서로 쓰였는가?’에 주목하여 출현 순서에 확률을 부여하는 것이다(Jurafsky & Martin, 2019, Chapter 3). 이들 방법론과 달리 워드임베딩은 ‘어떤 단어가 함께 쓰였는가?’에 주목한다. 공기(共起)관계에 초점을 두는 가장 큰 이유는 공간상에서 수치로 표현된 정보들을 통해 상호 비교를 도모할 수 있는 연산체계를 구성하기 위해서이다.

Figure 1을 참조하여 공기관계에 기초한 분포적 유사성의 의미를 이해할 수 있다.

첫째, Figure 1의 첫 번째 행은 A가 나타난 위치 근처에 B가 동시에 나타나는 경향으로 이해할 수 있다. 데이터 상에서 양자가 서로 밀접한 위치에서 함께 나타나는 것으로, 예컨대 자연어에서 ‘여자’와 ‘남자’가 가까운 위치에서 함께 나올 가능성이 큰 것과 같다. 둘째, Figure 1의 두 번째와 세 번째 행에서는 A와 B가 주변 문맥 w, x 및 y, z 를 공유하고 있는데, 데이터상에서 서로 분포적 환경이 비슷하여 “교환가능”하다는 뜻이 된다. 예컨대 자연어에서 ‘책방’을 ‘서점’으로 바꾸어도 대부분은 전달하는 의미에서 차이가 없는 것과

Figure 1. Distributional similarity.

Figure 2. Cosine similarity.

같다. 이때, A와 B는 분포적으로 유사한 방향성을 가진다. 여기서 이 두 가지 형태가 반드시 배타적인 것은 아니다. 가까운 위치에 함께 출현하는 요소들끼리 비슷한 맥락을 가질 확률이 높다는 것이 분포가설의 기본 가정이기 때문이다. 이를 나타낸 것이 Figure 1의 네 번째 행인데, A와 B가 동시에 출현하면서 주변 문맥을 공유하고 있다.

다음으로 이러한 관계가 어떻게 수치로 '측정'될 수 있는지를 기계학습 연구에서 광범위하게 쓰이는 코사인(cosine) 유사도 측정 방식을 통해 살펴보면 다음과 같다. 세 단어 car, desk, chair가 어떠한 공간에 대응되었을 때, 그들 사이의 거리는 이들 각각이 서로 얼마나 관련성을 가지고 있는지의 문제로 볼 수 있다. 이렇게 세 단어가 공간상에 배치되면 그들 사이의 거리는 동심원을 기준으로 한 코사인 각도를 계산하여 구할 수 있다. 아래 Figure 2에서 car와 desk의 사이각은 45°이며 desk와 chair의 사이각은 30°이다. $\cos 45^\circ \approx 0.5253$, $\cos 30^\circ \approx 0.1543$ 이므로 desk는 car보다 chair와 더 의미적으로 가까운 관계라고 정리할 수 있다.

언어 정보를 벡터 공간상에 표상하면 두 요소가 분포적으로 얼마나 동질적인가를 그 방향으로부터 확인할 수 있다. 화살표를 통해 이를 잘 보여주는 그림은 아래와 같다(Cited in <https://wikidocs.net/book/2155>). 크기만을 나타내는 이른바 스칼라(scalar) 척도와는 달리, 벡터 척도는 크기와 동시에 방향을 갖는 물리량이다. 따라서 그 방향성 정보(directional information)는 집체적인 결과 해석을 가능하게 한다. 아래 Figure 3의 첫 번째 그림에서는 두 벡터의

Figure 3. Directional vector information.

방향이 완전히 같다. 이럴 때 코사인 유사도는 1로 수렴된다. 두 번째 그림에서는 두 벡터의 방향이 반대 방향으로 코사인 유사도는 -1로 수렴된다. 이는 통계 일반에서 사용되는 상관계수(correlation coefficient)와 사실상 같은 개념인데, 두 그림을 두 분포가 서로 정적(正的) 상관관계 혹은 부적(負的) 상관관계를 가지고 있음을 말한다. 이와 달리 두 분포가 서로 직교하는 세 번째 그림의 경우에는 유사도가 0으로 측정되는데, 두 분포상의 유사성이 전혀 포착되지 않음을 말한다.

다시 말해, 두 분포가 가지는 상호 방향성에 따라 그들이 상관성을 지니고 있는지 아닌지, 상관성을 지닌다면 배타적 상관성인지 아닌지를 계산식을 통해 확인할 수 있다. 이러한 방식을 통하여 음소, 형태소, 단어 및 구 등의 각 언어단위를 그들이 가지는 의미-구조적 제약을 기준으로 하여 공간상에 분포하는 개체로 표상하게 되면 그들 사이의 유사성 및 상호작용을 '선형' 관계화하여 '연산'을 수행할 수 있다. 이러한 방식들로 자연언어의 이면에 숨겨진 여러 관계와 제약을 실질적으로 계산할 수 있다.

연구대상

본 연구의 연구대상자는 말더듬으로 진단받은 중학생 이상의 청소년 및 성인 110명(남성 92명, 여성 18명)으로 평균 연령은 26.21세(± 8.62)였다. 연구대상자는 (1) 만 13세 이상의 청소년과 성인이며, (2) 모국어가 한국어이며, (3) 자신이 말더듬는 사람이라고 인식하고 있으며, (4) 자발화 분석 결과 100음절당 3회 이상의 말더듬 행동이 관찰되었고(Yairi & Ambrose, 2005), (5) 2급 이상의 언어재활사가 P-FA-II (Sim et al., 2010)를 실시한 결과 말더듬 중증도가 '약함' 이상으로 진단되었고, (5) 말더듬 외 언어, 조음음운, 읽기, 시청각 및 신경학적인 문제를 가지고 있지 않는 경우 대상으로 선정하였다. P-FA-II 평가 결과, 총점수를 기준으로 20명은 '약함', 1명은 '약함-중간', 46명은 '중간', 43명은 '심함' 중증도의 말더듬으로 진단되었고 필수과제의 ND, AD 평균점수는 다음과 같다(Table 1).

말자료(speech samples) 수집 및 분석

연구대상자 선정 및 말자료 수집을 위하여 검사자는 소음이 없는 조용한 공간에서 각 연구대상자와 일대일로 검사를 시행하였

Table 1. Participants' information

	N	Mean age	Results of P-FA-II		
			Required task' mean ND rate	Required task' mean AD rate	Stuttering severity
Male	92	25.50 (8.25)	9.11 (5.57)	14.33 (13.14)	13 Mild, 1 Mild-Moderate, 39 Moderate, 39 Severe
Female	18	29.82 (9.77)	8.12 (5.81)	10.17 (11.02)	7 Mild, 7 Moderate, 4 Severe
Total	110	26.21 (8.62)	8.95 (5.59)	13.65 (12.87)	20 Mild, 1 Mild-Moderate, 46 Moderate, 43 Severe

Values are presented as mean (SD).

P-FA-II=Paradise-Fluency Assessment-II (Sim et al., 2010); ND=Normal disfluencies; AD= Abnormal disfluencies.

다. 모든 연구대상자는 P-FA-II를 사용하여 말더듬 평가를 받았으며, 평가의 전 과정은 말더듬 평가 및 진단, 자료 수집 및 분석의 정확성을 위해 디지털 캡코더를 사용하여 녹화하였다. 대상자들의 P-FA-II 읽기과제 자료는 3년 이상의 유창성장애 치료 경력을 지닌 언어재활사 4명이 전사한 후 P-FA-II의 비유창성 유형 분석지침(Sim et al., 2010)을 따라 비유창성 분석을 하였다.

임베딩 구현

데이터 전처리

파라다이스-유창성검사-II를 통해 기술된 데이터는 수작업 계수를 염두에 두고 만든 데이터이기 때문에 전산처리에 적합한 형태로 일부 수정을 하였다. 예컨대, 띄어쓰기나 음절 분리 등이 군데군데 틀렸거나 평가자가 전사를 하면서 각 파일마다 조금씩 달리 표기했다고 하더라도 연구자가 눈으로 보고 확인하여 연구할 때는 문제가 되지 않을 것이지만, 컴퓨터 처리를 목적으로 할 때는 이 부분들에 대한 수정이 불가피하다. 또한, 딥러닝 처리는 대부분 데이터의 선형성을 기준으로 분절을 하는 까닭에 일부 선형 처리에 적합하지 않은 데이터는 수정 처리를 하였다. 예컨대, 'Ur{R2(소)소} R1남자'와 같이 괄호를 여러 겹 동반하여 전사된 경우에는 비유창성의 구조적 특질을 보여준다는 점에서는 특기할 사항이지만, 컴퓨터상의 직렬처리(sequential processing)에는 부적합하다. 이러한 사례가 발견되면 최대한 순차적인 배열로 데이터를 일부 조정하였다. 데이터 정제 작업이 끝난 이후에는 각 파일의 내용을 전산처리를 위해 텍스트 파일로 변환하였다. 이 과정에서 문단 단위로 구성된 데이터는 모두 문장 단위로 변환하여 텍스트 파일의 한 행에 한 문장이 출현하는 형태로 데이터를 재구성하였다. 결과로 구성된 학습 데이터는 총 112,996개 음절, 총 236,956개 음소를 포함하였다.

본 연구에서는 음소 환경을 고려하기 때문에 각 한글 문자를 자모단위로 분리하는 과정이 전처리 단계에서 이루어졌다(파이썬 hgtk라이브러리 사용). 각 음절을 초성, 중성, 종성으로 분리한 이후에는 초성으로 나타난 자음에는 'C'를 붙여 표시하였으며, 중성의 경우에는 7종성법에 따라 종성을 7개로 구분하여 변환한 뒤 'J'

• 원문: 안녕하세요 선생님

• 'narrow' 버전 변환: ['ㄸ', 'ㅇC', 'ㅏ', 'ㄴJ', 'ㄴC', 'ㅣ', 'ㅇJ', 'ㅎC', 'ㅏ', 'ㄸC', 'ㅣ', 'ㅇC', 'ㅓ', '@', 'ㄸC', 'ㅣ', 'ㄴJ', 'ㄸC', 'ㅣ', 'ㅇJ', 'ㄴC', 'ㅣ', 'ㅇJ', '@']

• 'wide' 버전 변환: ['ㅇC', 'ㅏ', 'ㄴJ', 'ㄴC', 'ㅣ', 'ㅇJ', 'ㅎC', 'ㅏ', 'ㄸC', 'ㅣ', 'ㅇC', 'ㅓ', 'ㄸC', 'ㅣ', 'ㄴJ', 'ㄸC', 'ㅣ', 'ㅇJ', 'ㄴC', 'ㅣ', 'ㅇJ']

Figure 4. Example of the narrow and wide versions.

를 붙였다. 이는 글자 환경이 아니라 소리 환경을 조금이라도 더 결과에 반영하고 이중자음(예컨대 'ㅃ' 등)이 과도하게 해석되는 경향을 배제하기 위한 목적에 따른 것이다. 예컨대, '짓'과 같은 글자라면 차례로 'ㄸC', 'ㅣ', 'ㄴJ'와 같은 형태로 변환하였다. 단, 변환 과정은 파이썬 스크립트를 이용하여 기계적으로 이루어지기 때문에 각 글자 단위로 처리되었다. 이에 덧붙여, 말더듬이 빈번하게 발생하는 위치가 문장시작 혹은 낱말초성인 것을 고려하여(Bloodstein, 2002) 각 문장의 시작점과 어절 경계에는 각기 'ㄸ'와 '@'를 임의로 삽입하였다. 단, 문장 및 어절 경계에 해당하는 이 표지는 하단에서 설명할 'narrow' 버전 모델에만 음소 환경의 특수성을 더 세밀히 살피기 위한 용도로 사용하였다. 변환과정을 예로 들면 Figure 4와 같다.

위와 같은 형태로 원자료를 변환한 이후, 각 음소가 해당 읽기발화에서만 비정상적으로 많이 출현한 것은 아닌지를 확인하기 위한 용도로 세종 구어 코퍼스(<https://ithub.korean.go.kr>, 약 80만 어절)를 마찬가지로 방식으로 처리하였다. P-FA-II에 속한 2개의 읽기발화는 음소 별로 고르게 나타나도록 설계된 자료이지만, 실제 한국어 일반의 음소 출현과 혹 동떨어진 것은 없는지를 추가로 확인할 필요가 있었다. 세종 구어 코퍼스와 읽기발화 110명 데이터의 초성, 중성, 종성(7종성 변환) 자모를 따로 계산하여 비교하였으며, 출현 백분율로 보아 양자의 데이터에서 크게 분기되는 요소는 없는 것으로 확인되었다.

모델 학습

학습을 시키기 위한 데이터 처리가 끝난 이후에는 워드임베딩을 연산하여 모델을 구성하였다. 잘 알려진 워드임베딩 도구에는

Word2Vec (Mikolov et al., 2013), fastText (Bojanowski, Grave, Joulin, & Mikolov, 2017), 그리고 GloVe (Pennington, Socher, & Manning, 2014) 등이 존재하는데 본 연구에서는 그 가운데 가장 일반적인 Word2Vec을 사용하였다. FastText는 이른바 하위단어(subword)에 강점을 보이는 것으로 알려져 있는데(Ganegedara, 2018), 본 연구에서는 음소와 비유창성 유형의 분포를 임베딩하는 것이기 때문에 fastText의 강점이 부각될 수 없다. 다음으로 GloVe는 코퍼스 전체의 통계량을 Word2Vec보다 더 잘 반영하는 것으로 알려져 있는데(Ganegedara, 2018), 본 연구의 대상이 되는 약 24만개 음소는 딥러닝의 관점에서는 작은 데이터에 속하여 GloVe의 강점을 충분히 발휘할 수 없다고 판단하였다. 이상의 과정을 통합적으로 운용하기 위해서 본 연구에서는 파이썬 gensim 라이브러리를 사용하였다(Rehurek & Sojka, 2010).

다음으로 문맥의 범위를 설정하기 위한 방식(CBOW 또는 skip-gram)과 크기를 결정하여야 한다. 지난 몇 년간 축적된 연구결과에 따르면 대부분은 skip-gram이 더 좋은 성능을 내는 것으로 보고되므로(Ganegedara, 2018), 경험적 준칙에 따라 본 연구에서는 skip-gram을 사용하였다. 다음으로 적절한 문맥의 범위를 결정하여야 한다. 딥러닝에서 문맥이란 특정 범위 내에서 나타난 다른 언어단위들을 말하며, 그 특정 범위를 통상 윈도우(window)라 칭한다. 한국어 일반에서 적용 가능한 윈도우 크기에 대해 몇 편의 선행 연구가 존재한다. Park, Byun, Baek, Cho와 Oh (2018)는 음소 단위 및 음절 단위로 한국어 워드임베딩을 시도하면서 윈도우 크기를 5로 고정하였으나 윈도우 크기에 따른 별도의 실험 결과는 제시하지 않았다. Choi, Kim, Seol과 Lee (2017)는 한국어 단어의 대부분(95% 이상)의 음절 수를 고려하였을 때 한국어 음절 단위 임베딩에서 윈도우 크기는 1에서 4까지가 적당하다고 주장하였다. 이러한 점을 고려하여, 본 연구에서는 윈도우 크기에 따른 복수의 모델을 구성하여 비교하기로 하였다. 첫 번째는 윈도우 크기를 2로 하여(즉, 앞 뒤 총 4개의 출현 단위) 임베딩을 실시하고, 두 번째는 윈도우 크기를 4로 하여(즉, 앞 뒤 총 8개의 출현 단위) 임베딩을 실시하였다. 이어지는 논의에서 편의상 전자는 ‘narrow’버전 후자를 ‘wide’버전이라고 부르기로 한다.

물론 이상과 같은 방식으로 구현된 모델이 완벽한 것은 아니다. 몇 가지 남아있는 문제점이 존재한다. 첫째, 앞서 말한 괄호 묶기의 문제가 존재한다. ‘Ur{R2(소)소} R1남자’ 등의 예에서 볼 수 있듯이 원본 데이터에서 괄호는 반복 등의 단위를 표시하는데, 본 연구에서는 실제 발화된 내용의 선후 관계만을 보기 때문에 이러한 괄호의 쓰임이 무시된다. 둘째, 원본 데이터도 어디까지나 전사(transcription)된 내용이기 때문에, 실제 음성적 환경을 온전히 반영하

지 못한다는 한계가 존재한다. 예컨대, 어두 자음과 어중 자음의 유무성 구분이라든지, 이중자음에서 7종성법의 예외가 존재한다던지(예: 밝아>발가, 닭이>다익), 경음화(예: 특강>특깁)가 나타나 는 경우 현재의 모델이 그 세밀한 음성적 특성까지는 반영할 수 없다. 또한, 24만개 음소의 데이터 항목은 임베딩 구성의 측면에서 보았을 때 비교적 작은 데이터이기 때문에, 크기에 따른 왜곡이 발생할 가능성도 일부 존재한다.

임베딩 모델 시각화

일반적인 워드임베딩 연구는 결과의 시각화를 포함한다. 본 임베딩의 결과도 마찬가지로, 청소년 및 말더듬 성인의 읽기자료를 토대로 임베딩한 결과를 t-SNE 그림으로 시각화하였다. 앞서 설명한 임베딩 구현에서는 워드임베딩의 통상적 방식에 따라 300차원의 공간으로 음소 및 비유창성 유형의 분포를 구성하였다. 문제는 우리는 3차원 공간에 살고 있어 이러한 고차원의 정보를 직접 살필 수가 없다. 이 문제를 해결하기 위해 모델의 차원을 축소하여 2차원 혹은 3차원으로 변환하는 기법이 개발되어 있으며, t-SNE 시각화는 결과물을 2차원 평면상으로 옮기는 과정이다. 이를 통해 우리는 전체 분포가 공간상에 위치하는 바를 직관적으로 조망할 수 있다. 일반적으로 t-SNE 그림을 분석할 때는 주로 세 가지 측면에 주목한다. 첫째 군집의 형성, 둘째 군집의 위치, 셋째 군집의 중심부이다. 군집 형성 여부는 해당 데이터 안에서 어떠한 항목들이 서로 가까운 분포 관계를 맺는가를 나타낸다. 군집의 위치는 해당 항목들이 얼마나 균질하게 출현하는가를 말하는데, 군집이 그림의 중심부에 있을수록 해당 데이터에서 고루 분포하고 있다는 의미이다. 군집의 중심은 해당 군집과 나아가 전체 데이터를 대표할 수 있는 항목이 무엇인가를 나타낸다.

신뢰도 측정

신뢰도 측정을 위하여 전체 자료의 10%인 11명의 자료를 무작위로 선정하였다. 비유창성 유형(4개 ND 유형, 5개 AD 유형)에 대한 평가자 간 신뢰도는 2명의 유창성장애 전문가에 의해 독립적으로 이루어졌다. 이 평가자들은 언어재활사 1, 2급 국가자격증을 소지하고 있으며 언어병리학과 석사과정을 전공하고 유창성장애 임상 경력이 5년 이상인 전문가이다. ND 유형과 AD 유형은 일치율(%) 분석을 통하여 신뢰도를 측정하였다.

비유창성 유형을 분류하는 평가자 간 일치율은 읽기 과제에서 일치한 ND(또는 AD)유형의 수를 읽기과제에 나타난 ND(또는 AD)유형의 총 수로 나눈 후 100을 곱하여 평가자 간 신뢰도를 측정하였다. 그 결과, ND 유형에서 89%, AD 유형에서 96%의 일치율을 보였다.

연구결과

임베딩 결과의 전체 특성

앞선 과정에서 구축된 두 개의 모형, 즉 ‘narrow’버전과 ‘wide’버전을 각기 시각화하면 차례로 Figure 5 및 Figure 6과 같다.

위 시각화를 바탕으로 얻을 수 있는 결과에 대한 개괄적 이해는 다음과 같다. 첫째, Figure 5와 Figure 6에서는 모두 비유창성의 주요 유형들이 일정한 군집을 형성하고 있다. 이것이 의미하는 바는 상이한 비유창성 유형들이 동시에 출현하거나, 혹은 각 비유창성 유형들이 출현하는 분포가 서로 유사하다는 것이다. 전자를 예를 들면, 간투사 사용 뒤에 곧바로 특정 음절 반복을 하는 경우를 들 수 있다. 후자를 예를 들면, Ia가 출현하는 환경에서 Ia를 URa로 바꾸어도 현재 데이터에서는 별로 이상한 것이 없다는 의미이다. 물론, 앞서 Figure 1에서 설명한 바와 같이 이 양자가 배타적인 것은 아니나, 둘 가운데 어느 것이 더 주된 특성인가에 대해서는 주어진 결과만으로 판단하기는 어렵다. 둘째, 두 그림 모두에서 비유창성의 유형들로 구성된 군집은 모두 그림의 정중앙에 위치한다. 이것이 의미하는 바는 두 가지로 볼 수 있다. 첫째로 위 비유창성 유형 표지들이 해당 데이터의 특성을 가장 잘 대변하고 있는 항목들이라는 것이다. 이는 말더듬 대상자들의 데이터를 기준으로 구성된 임베딩이 실제로 말더듬 대상자들의 대표 특성을 충분히 잘 반영하고 있음을 나타낸다. 즉, 구성된 모델이 상식적인 차원에서 자연스럽게 구성되었음을 보여준다. 둘째로 비유창성 유형들이 각 음소 단위와 서로 비슷한 거리 관계를 가지고 있음을 참작하면, 비유창성은 특정 음소에 크게 영향을 받지 않는 듯하다. 셋째, 두 그림에서 군집의 중심에 놓이는 요소는 ‘URa’이며 ‘Ur’, ‘I’, ‘Ia’ 등이 그에 근접한 요소들이다. 이 중심은 위 네 항목이 이른바 비유창성 읽기 발화의 특성을 우선적으로 대변할 수 있다는 점을 시사한다.

대부분의 과학적 연구는 큰 그림을 중심으로 설명을 도모하는 거시적 분석과 각 부분의 특성을 보다 세밀하게 살피는 미시적 분

석으로 구성된다. 워드임베딩을 통한 자료 분석도 마찬가지이다. 상기의 시각화를 통한 결과분석은 거시적으로 자료의 윤곽을 확인하기 위한 절차에 해당한다. 이에 비하여 이후 제시되는 각 유형 별 분석은 워드임베딩에서 연산된 수치를 바탕으로 세부적 특성을 도출하기 위한 미시 분석에 해당한다.

비유창성 유형의 분포적 특징

각 비유창성 유형이 서로 가지는 관련성을 유사도(similarity)의 측면에서 살펴보았다. Table 2는 각 비유창성 유형 표지를 키워드로 하여 그 표지와 관련성이 가장 높은 항목(음소 및 비유창성 유형 모두 포함) 상위 10개를 뽑은 결과이다(most_similar 함수).

비유창성 유형들은 모두 다른 비유창성 유형과 높은 관련성을 보였다. 특히 ‘DP’를 제외한 9개의 비유창성 유형들은 예외 없이 상위 10개 항목에 자신을 제외한 나머지 9개를 포함하고 있었으며, 대부분의 비유창성 간에 .87 이상의 유사도를 보여 분포가 매우 가깝다는 점을 알 수 있었다. 특히 ‘URa’, ‘Ia’, ‘Ha’, ‘R1a’는 상호간의 분포적 특징이 거의 같다고 분석되었다. ‘URa’, ‘Ia’, ‘Ha’, ‘R1a’ 모두 유사도 순위 3위 안에 다른 요소들이 포함되어 있었으며 그 유사도 수치 또한 .95를 상회하였다. 이 정도 수치의 유사도가 나타날 경우 통상 두 요소가 유의미하게 동질적인 분포를 공유한다고 해석할 수 있다. 즉, 네 가지 가운데 어느 한 유형이 나타났을 때 그 유형을 다른 세 유형 중의 하나로 바꾸어도 현재 분포적 연산에서는 전혀 이상하지 않다는 것이다. 결론적으로, ‘URa’, ‘Ia’, ‘Ha’, ‘R1a’는 발생 요인이 거의 중첩된다고 볼 수 있었다.

‘R2’와 ‘DP’는 다른 비유창성 유형과 분포적 양상이 매우 다르게 나타났다. 특히 ‘DP’는 나머지 비유창성 유형들과 눈에 띄는 차이를 보였다. Table 2에서 DP는 다른 비유창성 유형들과의 유사도가 9위 혹은 10위에 있는 것으로 나타났으며 유사도 점수 또한 평균 .6631로 낮았다. 이는 ‘DP’는 다른 비유창성 유형과 분포적 특징이 다름을 나타내는 것으로 해석이 가능하다. 즉, ‘DP’는 나머지 비유

Figure 5. t-SNE visualization. (narrow version, windows=2).

Figure 6. t-SNE visualization. (wide version, windows=4).

Table 2. Top 10 similar items of 10 disfluency types

H		Ha		I		Ia		Ur		URa		R1		R1a		R2		DP	
T	S	T	S	T	S	T	S	T	S	T	S	T	S	T	S	T	S	T	S
R1	.97	URa	.98	R1a	.96	R1a	.98	URa	.93	Ha	.98	R1a	.97	Ia	.98	R1	.92	R2	.78
R1a	.95	Ia	.98	Ia	.96	URa	.98	Ha	.92	Ia	.98	H	.97	Ha	.98	Ha	.91	R1	.71
Ia	.95	R1a	.98	Ha	.96	Ha	.98	Ia	.93	R1a	.97	Ia	.95	URa	.97	R1a	.91	H	.69
Ha	.92	I	.96	URa	.95	I	.96	H	.91	I	.95	Ha	.92	R1	.97	Ur	.89	I	.68
Ur	.91	Ur	.93	R1	.92	H	.95	R1a	.90	Ur	.93	I	.92	I	.96	H	.89	R1a	.68
URa	.90	R1	.92	Ur	.89	R1	.95	R2	.89	H	.90	R2	.92	H	.95	URa	.89	Ha	.66
R2	.89	H	.92	R2	.88	Ur	.93	I	.89	R1	.90	URa	.90	R2	.91	I	.88	URa	.65
I	.87	R2	.91	H	.87	R2	.88	R1	.88	R2	.89	Ur	.88	Ur	.90	Ia	.88	Ur	.65
DP	.69	DP	.66	DP	.68	DP	.64	□	.69	ㅍ	.61	DP	.71	DP	.68	DP	.78	Ia	.64
□	.66	□	.57	ㄷ	.52	□	.58	DP	.65	DP	.59	ㄴ	.62	ㅈ	.56	ㄴ	.67	H	.63

Values are presented as median (range).

T = Type; S = Similarity; H = Hesitation, I = Interjection, Ur = Unfinished/Revision word, R1 = Repetition1; Ha = Abnormal hesitation, Ia = Abnormal interjection, URa = Abnormal unfinished/revision word, R2 = Repetition2, DP = Disrhythmic Phonation.

창성 유형과는 다른 요인에 의한 영향을 받고 있다는 점이 임베딩 결과에서 드러난 것이다. 한 가지 더 주목할 점은 ‘DP’와 가장 분포적 유사성이 높은 요소가 말운동통제 능력과 관련 있는, 즉, 비유창성 범주를 분류할 때 일관되게 비정상적인 범주에 포함되는 ‘R2’라는 점이다. 이는 본 임베딩 결과가 지금까지 말더듬 분야에서 얻어온 결과와 배치되지 않을 뿐만 아니라, 선행연구에서 포착된 내용이 딥러닝 임베딩에서도 재현됨을 의미한다. 즉, 워드임베딩 결과가 적어도 말더듬 대상자들의 일반적인 특질을 충분히 포착하고 있음을 알 수 있었다. ‘R2’의 경우 DP를 제외한 다른 비유창성 유형들과 .85 이상의 유사도를 보였다. ‘R2’는 발화 분석 시 비유창성 유형들과 동반되어 나타나는 경우가 빈번하였는데 이를 학습데이터를 구축하는 과정에서 데이터를 순차적인 정보로만 변환한 것에 영향을 받은 것으로 보인다(‘Ur{R2(소)소} R1납자’).

정상적 비유창성과 비정상적 비유창성의 분포적 차이

다음으로 정상적 비유창성(ND) 범주와 비정상적 비유창성(AD) 범주 사이에 분포적 차이가 실재하는지를 확인하였다. 여기서는 ‘R2’와 ‘DP’를 제외하고 ND 및 AD에서 각 대응쌍(예컨대, ‘Ur’ vs. ‘URa’)이 존재하는 8개 유형으로 분석을 제한하여 진행한 결과 양자는 분포적 차이가 거의 없었다. 구체적인 증거는 Table 2에서 제시된 유사도 값으로 다시 확인할 수 있다. ‘H’와 ‘Ha’는 다소 거리가 먼 편(4위, .9237)이기는 하나, ‘Ur’과 ‘R1’은 자신과 가장 분포적 특질이 가까운 요소로 각기 ‘URa’(1위, .9314)와 ‘R1a’(1위, .9657)가 있었다. ‘I’의 경우에도 ‘Ia’가 2위이며 그 유사도는 .9562로 측정되었다. 즉, 세 정상적 비유창성(ND) 유형의 경우 현재 태깅된 각 유형을 모두 ‘Ia’, ‘URa’, ‘R1a’로 바꾸어도 전체적인 양상에서는

현재의 분포와 대동소이하다는 뜻이다. 그 뿐만 아니라 위 표에서 각 ND 및 AD 쌍의 순위에서 상위 다섯 개의 분포를 보아도 이 또한 서로 상당히 유사함을 알 수 있었다. 이는 각 ND 및 AD 항목이 임베딩된 공간 안에서 서로 가까운 위치에 놓여있음을 말하는데, 앞서 제시된 t-SNE 그림에서도 어느 정도 드러나는 부분이다. 위 그림에서 보면 ‘R1’과 ‘R1a’가 약간 거리가 있을 뿐, 실제로 ‘H’, ‘I’, ‘Ur’과 가장 가까운 위치에 놓인 다른 요소는 ‘Ha’, ‘Ia’, ‘URa’이었다. 이를 바탕으로 동일한 비유창성 유형의 정상적·비정상적 구분은 적어도 임베딩 공간에서는 명시적이지 않다고 정리할 수 있었다.

음소에 따른 비유창성의 분포적 환경

첫 번째 연구문제와 관련하여 어떠한 특정 음소환경에 따라 말더듬의 출현이 영향을 받는가를 확인하고자 하였다. 두 모델 ‘narrow’ 버전과 ‘wide’ 버전에서 각 초성 자음이 ‘H’, ‘Ha’, ‘I’, ‘Ia’, ‘Ur’, ‘URa’, ‘R1’, ‘R1a’ 8개 유형과 가지는 유사도를 계산한 이후 그 평균과 표준편차를 구한 결과는 Table 3과 같다. 양쪽 모델에서 모두 상대적으로 높은 유사도를 가지는 것으로 분석되는 음소는 굵은 글씨로 표시된 ‘ㄱ’, ‘ㄷ’, ‘ㅈ’ 음소로 나타났다.

이 초성 자음들이 P-FA-II의 읽기발화에 많이 등장하였기 때문에 이러한 결과치가 나온 것일 수 있으므로 앞서 구한 세종 구어 코퍼스 비교표를 참조하였다. Table 4를 보면 ‘ㄱ’, ‘ㄷ’, ‘ㅈ’ 등의 음소가 절대적으로나 상대적으로 더 많이 출현했다고 볼 수 없다고 판단되었다. 절대적으로 보자면 위 세 음소보다 ‘ㄴ’ 등이 더 큰 비율로 사용되었는데 오히려 위 표에서 보듯이 ‘ㄴ’은 말더듬 음소 환경과 관련도가 떨어지는 편이었다. 상대적으로 보아도 위 세 음소가 해당 읽기발화에서 실제 사용량보다 큰 비중을 차지한다고 보기도

Table 3. Phonemes with high frequency of stuttering

Narrow			Wide		
Phoneme	Mean	SD	Phoneme	Mean	SD
ㅅ	0.4674	0.0766	ㅁ	0.5754	0.0862
ㅋ	0.4593	0.0926	ㄷ	0.5188	0.0639
ㅈ	0.4579	0.0745	ㅊ	0.4785	0.0786
ㅊ	0.4556	0.0801	ㅂ	0.4717	0.0796
ㄷ	0.4473	0.0710	ㅅ	0.4454	0.0830
ㄹ	0.4455	0.0819	ㅇ	0.4399	0.0619
ㅃ	0.4370	0.0666	ㄱ	0.4382	0.0967
ㅇ	0.4266	0.0653	ㅍ	0.4336	0.0899
ㄴ	0.4024	0.0919	ㄴ	0.3827	0.1011
ㅎ	0.3980	0.0907	ㄱ	0.3776	0.1041

어려웠다. 즉, 각 음소의 출현 빈도가 위 결과에 영향을 주었을 가능성이 별로 없음을 확인할 수 있었다.

다음으로 각 음소가 어두에서 사용되었을 가능성도 판단해 보아야 한다. 본 읽기발화 2개의 지문을 계수한 결과 ‘ㄷ’은 16회, ‘ㄱ’은 21회, ‘ㅈ’은 19회 어두에서 사용되었다. 이에 비해 ‘ㄱ’은 44회 사용되었는데 마찬가지로 비유창성과의 관련도가 비교적 떨어졌다. ‘ㅎ’은 16회 어두에서 사용되어 위 ‘ㄱ’, ‘ㄷ’, ‘ㅈ’의 어두 사용빈도와 유사하였지만 위 관련도 분석에서는 후순위에 놓였다. 즉, 어두에서 사용된 빈도가 결과에 직접적 영향을 준 것은 아니라는 것을 확인할 수 있었다. 결과적으로 적어도 본 연구에서 사용된 110개 읽기 발화 데이터 안에서는 ‘ㄱ’, ‘ㄷ’, ‘ㅈ’ 음소가 8개의 비유창성 발생과 어떤 형식으로든 관련이 있다고 추론할 수 있다.

논의 및 결론

본 연구는 110명 말더듬 대상자의 읽기발화에 나타난 비유창성 위치와 특성을 워드임베딩을 통해 규명하고자 하였다. 이를 위해 t-SNE 그림으로 모델을 시각화하여 제시하였고 음소 환경을 분석하고자 P-FA-II의 읽기발화와 세종 코퍼스 데이터를 비교 분석하였다.

연구결과를 요약하면, 1) 비유창성 유형들 사이의 분포적 유사성을 분석한 결과, 양적/질적 자질 유무에 따라 ND 및 AD에 각 대응쌍이 존재하는 8개 비유창성 유형은 .9 이상의 높은 유사도로 서로 근접하여 출현하는 것으로 나타났다. 또한, ‘la’, ‘Ha’, ‘URa’, ‘Rla’는 .95 이상의 유사도로 발생 위치가 거의 중첩되는 것으로 나타났고, AD 범주에만 속하는 ‘R2’와 ‘DP’는 다른 비유창성 유형들과 상이한 분포양상을 보이는 것으로 나타났다. 2) ND 범주와 AD 범주에 모두 포함될 수 있는 8개의 비유창성 유형 간의 분포적 특성을 대응 별로 분석한 결과, 차이가 크지 않은 것으로 나타났다. 3)

Table 4. Frequency table of onset consonants (Sejong vs. P-FA-II)

Phoneme	# (Sejong)	% (Sejong)	# (Reading)	% (Reading)
ㄱ	209,704	7.50%	11,029	5.40%
ㅋ	13,144	0.47%	809	0.40%
ㄴ	120,587	4.31%	10,084	4.93%
ㄷ	92,908	3.32%	7,568	3.70%
ㄹ	13,894	0.50%	832	0.41%
ㄱ	88,684	3.17%	6,657	3.26%
ㄴ	65,535	2.34%	4,429	2.17%
ㅂ	39,932	1.43%	3,075	1.50%
ㅃ	3,587	0.13%	207	0.10%
ㅅ	81,321	2.91%	7,874	3.85%
ㅆ	5,327	0.19%	619	0.30%
ㅇ	285,614	10.22%	20,064	9.82%
ㅈ	90,563	3.24%	4,897	2.40%
ㅊ	5,160	0.18%	403	0.20%
ㅌ	16,438	0.59%	2,377	1.16%
ㅋ	4,591	0.16%	410	0.20%
ㅍ	11,104	0.40%	418	0.20%
ㅑ	9,370	0.34%	713	0.35%
ㅎ	65,410	2.34%	4,009	1.96%
ㄱ	209,704	7.50%	11,029	5.40%

비유창성 위치와 관련하여 음소환경에 따른 8개 비유창성 출현 빈도를 분석한 결과, ‘ㄱ’, ‘ㄷ’, ‘ㅈ’ 음소에서 비유창성의 출현 가능성이 비교적 높게 나타났다. 연구문제 별로 자세히 살펴보면 다음과 같다.

첫 번째, 비유창성 유형들 사이의 분포적 유사성을 분석한 결과, ND 범주에 속하는 유형 ‘UR’, ‘I’, ‘H’, ‘R1’과 AD 범주에 속하는 유형인 ‘URa’, ‘la’, ‘Ha’, ‘Rla’ 들은 .9 이상의 높은 유사도로 서로 근접하여 출현하는 것으로 나타났다. 그리고 ‘URa’, ‘la’, ‘Ha’, ‘Rla’ 등의 유형들은 서로 .95 이상의 유사도를 가지는 것으로 나타났는데, 이는 발생 위치가 거의 중첩된다는 것을 함의한다. 또한, 비유창성 군집의 중심이 ‘URa’라는 결과는 ‘미완성수정’이 말더듬의 핵심적인 특질이라고 한 Shin과 Ha (2015)의 견해와 상통한다. 말더듬 대상자의 데이터에서 가장 중요한 특징이 되는 요인은 ‘미완성수정’ 유형이라고 정리할 수 있고, 그 유형을 중심으로 여러 다른 유형과 여러 음소가 배치된 형태로 볼 수 있다. 다시 말해, 이 비유창성 유형들은 상호 근접하여 비유창성 군집으로 나타날 수도 있고, AD 범주에 속한 유형이 출현하는 자리에 그에 상응하는 ND 범주의 유형이 출현할 수도 있다는 것이다. Sim, Shin과 Lee (2010)는 P-FA-II 유형 분류 시 ND에 속한 유형이라도 3회 이상 반복되거나 질적 양상이 동반되면 AD로 분류하였다. 즉, ‘H’, ‘I’, ‘Ur’, ‘R1’ 유형이라도 비유창성의 지속시간이나 반복횟수, 동반되는 질적 양상(청각적/

시각적 긴장, 음의 고저나 강도의 변화, 말속도의 변화, 들리는 들숨이나 날숨, 모음의 중모음화 등)을 고려하여 AD에 포함시키도록 하였다('Ha', 'Ia', 'URa', 'RIa'). 이렇게 분류한 이유는 나타난 유형 자체만으로 단순히 정상과 비정상을 판단하였을 경우 말더듬 문제를 과소평가할 수 있는 위험을 최소화하기 위함이었는데, 이 유형들은 ND 범주나 AD 범주에 포함되더라도 유사한 발생위치 속성을 가진 것으로 워드임베딩 분석 결과 나타난 것이다.

AD에 속한 유형 중 'R2'와 'DP'는 'Ur', 'I', 'H', 'RI'와 다른 분포 양상을 보이는 것으로 나타났다. 'R2'의 경우 다른 비유창성 유형과 .85의 유사도를 보이나 'DP'는 .6 정도의 낮은 유사도를 보여 질적으로 다른 유형이라는 것을 보여주었다. 'DP'는 막힘, 연장, 깨진 낱말이 포함된 전형적인 AD 유형이다(Yairi & Ambrose, 2005). 본 연구의 임베딩 결과는 'DP'는 언어적 비유창성이라고도 불리는 ND 범주와 분포적 특질이 상이함을 나타낸다. 이런 결과는 말더듬는 사람이 주로 보이는 AD의 특성은 일반인도 보이는 ND의 특성과는 질적으로 다르다는 것을 보여준다.

두 번째, ND 범주와 AD 범주에 모두 속하는 비유창성 유형 8개의 분포적 차이가 실제하는지를 확인한 결과, 그 차이가 크지 않은 것으로 나타났다. 말더듬 집단과 일반 집단이 보이는 비유창성은 양적인 차이뿐 아니라 질적인 차이를 보이며 연구자마다 조금씩 다른 비유창성 분류체계를 사용한다. 연구자들은 비유창성의 유형 중 말더듬는 사람의 말 행동을 적절하게 나타낼 수 있는 유형이 무엇인지에 대하여 연구하였다. Conture (1990)는 비유창성 유형을 크게 낱말 내 비유창성(within-word disfluencies)과 낱말 간 비유창성(between-word disfluencies)으로 제시하였다. Conture (1990)는 일음절낱말반복, 소리 및 음절반복, 연장, 막힘을 말더듬는 사람의 특징인 낱말 내 비유창성 유형에 포함시켰고, 일반인이 보일 수 있는 비유창성인 낱말 간 비유창성에는 구반복, 다음절낱말반복, 간투사, 수정을 포함시켰다. 그러나 이 유형 분류는 모든 낱말 내 비유창성 유형이 반드시 말더듬는 사람만에게만 나타나는 것은 아니라는 점과(Cordes & Ingham, 1994), 일음절낱말반복이 모호하고 비일관적으로 처리된다는 점에서(Conture, 1990; Yairi et al. 1996) 제한점을 갖는다. 게다가 말더듬 아동만 보이고 일반 아동은 산출하지 않는 비유창성은 없는 것으로 나타났기 때문에(Yairi & Lewis, 1984; Yairi et al., 1996) Yairi 등(1996)은 말더듬는 사람만이 보이는 특정한 비유창성을 규정짓기보다는 진성비유창성(SLD)과 가성비유창성(OD)이라는 용어를 사용하여 낱말부분반복, 일음절낱말반복, 비운율적 발생(DP), 긴장성 멈춤을 진성비유창성으로, 다음절낱말반복, 구반복, 간투사, 수정-불완전 구문을 가성비유창성으로 분류하였다. 또한, Yairi와 동료들(1996)은 진성비유창성이 초

기 말더듬을 가려낼 수 있는 도구가 될 수 있음을 보여주었다. Sim과 동료들(2010)에서는 'UR', 'I', 'H', 'RI'를 ND로, 그리고 질적양상이 동반된 ND 유형들을 AD로 구분하였는데, 이는 다른 연구자들이 이 네 가지 유형을 정상적인 비유창성으로만 분류했던 방식과 다른 체계이다. 전술하였듯이 이는 비유창성 유형만으로 판단하기보다 동반되는 양적, 질적인 양상을 고려하여 비유창성 범주를 구분하기 위함이었다. 하지만 본 연구에서 'UR', 'I', 'H', 'RI'의 ND 범주 유형과 그에 대응하는 AD 범주 유형은 유사한 발생위치 속성을 가진 것으로 드러나 실제 말더듬는 사람에게서 나타나는 양적, 질적인 양상을 고려하지 않는다면 이 유형들을 ND와 AD로 구분하는 것이 매우 어려움을 제시한 것이다.

끝으로, 비유창성 발생위치와 관련하여 음소환경에 따른 비유창성 출현 빈도를 분석한 결과, 'ㄱ', 'ㄷ', 'ㅈ' 음소에서 R2와 DP를 제외한 8개 비유창성 빈도가 높게 나타났다. 두 모델 'narrow' 버전(원도우 크기 2, 앞뒤 총 4개 음소단위)과 'wide' 버전(원도우 크기 4, 앞뒤 총 8개 음소단위)에서 비유창성은 특정 음소에 크게 영향을 받지는 않고 초성 'ㄱ', 'ㄷ', 'ㅈ' 자음에서 많이 나타났다. P-FA-II의 읽기발화에 나타난 음소 출현 빈도와 무관하게 t-SNE 시각화 자료와 세종 구어 코퍼스과 비교한 자료에서도 이와 근사한 결과를 보였다. 선행연구에서 음성학적으로 더욱 복잡한 단위산출시 말더듬이 나타나는지를 규명하려는 시도가 많았다. Kwon (1991)의 연구에서는 말더듬 성인을 대상으로 9가지 과제상황에서 말더듬이 빈번하게 나타난 자음의 특징을 조사하였다. 그 결과, 조음위치에서는 경구개음, 연구개음, 치조음, 성문음 순으로, 조음방법에서는 파열음, 마찰음, 파찰음, 비음, 유음 등의 순으로 말더듬이 빈번하게 나타나 늦게 발달하는 말소리일수록 말더듬이 더욱 빈번하게 발생되었다고 보고하였다. Lee, Han과 Sim (2004)의 연구에서는 말더듬 성인 10명을 대상으로 읽기과제를 실시한 결과 조음복잡성이 증가된 어절일수록 비유창성 빈도도 함께 증가하는 결과를 보여주었다. 그러나 모든 연구결과가 일치된 것은 아니었다. Shin (1996)은 말더듬 진단을 받은 아동과 성인으로 구성된 78명을 대상으로 읽기과제 및 자발화 수집을 통하여 자료를 분석한 결과 조음위치에 따라 양순음, 성문음, 연구개음, 치조음, 경구개음 순으로, 조음방법에 따라 비음, 폐쇄음, 마찰음, 유음 순으로 말더듬이 빈번하게 발생하였다고 보고하였다. 일찍 발달하는 말소리나 조음적으로 덜 복잡하게 구성된 말소리에서 비유창성 발생빈도가 더 낮게 나타나지는 않았다고 보고하여 선행연구들과는 다른 결과를 보여주었다. 이처럼 조음복잡성과 말더듬과의 관련성에 관한 연구에서는 조음복잡성이 증가된 음절 및 어절에서 말더듬이 증가되어 나타난다는 보고가 지배적이지만 모든 연구에서 동일한 결과를 보고하고 있지

않다. 그리고 말더듬 성인의 경우 개인별로 어려운 음소들이 다르기 때문에 어려운 음소를 일반화하기는 한계가 있으므로, 더욱 정확한 정보를 얻기 위하여 대상자와 과제를 변화시킨 다양한 연구가 계속되어야 할 것이다. 현재까지 얻어진 내용을 일반화할 수는 없다는 한계는 분명히 존재하지만, 그 결과가 무의미한 것은 아니다. 본 결과는 후행연구와 분석에 있어서 시발점이 된다. 딥러닝을 이용한 현상 분석은 많은 경우 직접적인 답을 주지는 않으나, ‘어느 부분에서 무엇인가 이상한 점이 포착되었으니 보다 면밀히 관찰하기를 바란다’와 같은 알림 메시지를 주는 기능은 탁월하다. 다시 말해, ‘ㄱ’, ‘ㄷ’, ‘ㅈ’의 음소를 말더듬의 관점에서 더 세밀히 살펴야 할 필요성은 제기되었다고 할 수 있다. 말더듬이라는 것이 복합적인 이유에 의한 통합적인 결과라는 점을 가정하면 이러한 단계적 접근법이 타당성을 지닌다고 할 수 있다. 아울러, 딥러닝 그리고 워드임베딩은 단순한 빈도 계산보다 설명력이 더 좋다는 것은 주지의 사실이다. 따라서, 후행연구를 위한 참조 사항이라는 측면에서는 더 신뢰성을 가지는 결과일 것이다.

본 연구 결과가 선행 및 후행연구에 가지는 함의를 정리하면 다음과 같다. 첫째, 말더듬의 각 유형은 군집을 이루어 동시다발적으로 출현하는 경향성이 있음이 포착되었다. 즉, 하나의 말더듬 유형이 발생하는 환경이 되면 연쇄적으로 다른 말더듬도 함께 발생할 가능성이 커진다는 것인데, 이는 선행연구의 군집화 가설과 합치하는 결과이다. 둘째, 정상적 비유창성과 비정상적 비유창성 사이의 분포적 특징, 특히 양적 양상과 질적 양상에 따라 비유창성 범주가 구분되는 비유창성 유형 간의 분포적 특징이 크게 두드러지지 않는다는 점을 확인하였다. 비유창성 유형에 동반되는 양적, 질적 양상은 말더듬은 사람의 말 특성을 파악하는데 중요한 역할을 한다. 하지만 이러한 양상을 고려하지 않고 비유창성 유형 자체만을 본다면 말더듬의 두 범주로 나누기에 이 유형들의 분포적 특징이 유사함을 확인할 수 있는 결과라 할 수 있다. 셋째, 내적수정가설과 관련된 말더듬이란 현상의 중심에는 간투사 사용 전략 혹은 미완성수정 사용 전략이 놓인다고 판단된다. 이는 전체 t-SNE 그림에서 ‘I’, ‘Ia’, ‘Ur’, ‘URa’ 등의 유형이 중심에 놓인다는 점에서, 말더듬 대상자 데이터의 대표적인 속성을 위 4개 유형으로 정리할 수 있기 때문이다. 넷째, EXPLAN 모델에서 언어계획과 운동실행을 나누어 본 관점을 적용하였을 때, ‘R2’ 그리고 ‘DP’ 유형이 나머지 비유창성 유형과 분포적으로 동떨어져 있다는 점은 시사하는 바가 크다. 특히 ‘DP’는 여타의 비유창성과 발생요인 등에서 차이가 있을 것으로 예측되는 바, 추가적인 연구가 필요하다.

본 연구의 제한점은 비교대조군이 제시되지 않아 결과가 가지는 유의미성에 대한 해석이 다소 취약하다는 점이다. 이러한 관점에서

본 연구의 결과에 바탕을 두어 추후 연구에서 다룰 사항은 다음과 같다. 첫째, 읽기발화 결과와 언어계획이 더 요구되는 그림묘사나 자유발화의 결과와 비교할 수 있어야 한다. 읽기발화와 자유발화가 사용환경 등이 동일하다고 가정할 수 없으므로 양자의 비교를 통해서만 말더듬과 언어계획과 관련된 분포적 특성에 대한 보다 종합적 결론을 내릴 수 있을 것이다. 둘째, 본 연구는 중학생 이상의 (준)성인의 발화로만 연구 영역을 한정하였는데, 후행연구에서는 아동 말더듬 데이터와의 비교가 요구된다. 셋째, 본 연구에서 정상적 비유창성과 비정상적 비유창성의 표지 구분이 크게 기능을 하지 않는 것으로 결과를 제시하였으나, 이를 더 확실하게 검증하기 위해서 일반인 데이터와도 비교해야 한다. 말더듬을 대상으로 이미 실행한 연구에서 일정 정도 의미 있는 결과가 도출된 만큼, 더 방대하고 구분이 지어진 데이터를 가지고 진행하는 후행연구 또한 유의미한 결과를 보일 것으로 예상된다. 넷째, 음소 별 비유창성 빈도 외에 음운자질에 따른 비유창성 출현 연구도 필요하다.

본 연구의 의의는 딥러닝 방법론을 언어병리학과 융합하여 기존의 연구에서는 포착되지 못했던 연구 주제를 발견하거나 혹은 논쟁의 소지가 있는 주제에 대하여 또 다른 입증 결과를 제시하였다는 것이다. 향후 언어병리학 분야에서 딥러닝 방법론에 기대할 수 있는 바도 이와 마찬가지로 될 것이다. 언어병리학은 궁극적으로 대상자들의 치료 가능성을 지향하기 때문에 본질적으로 실용적인 목적을 염두에 두며, 이를 성취하기 위해 새로운 기술과 방법을 받아들이는 것에 포용적인 성격을 지닌다. 즉 이런 연구결과는 말더듬 평가 및 치료 애플리케이션, 플랫폼 개발에 적용될 수 있다. 또한 언어병리학은 이미 언어학 및 보건의료 계열 등과 융합을 이룬 분야로써 학제간 연구에 낯설어 하지 않는다는 특성을 보인다. 이러한 측면에서 보면 딥러닝이 중요한 연구의 매개로 자리를 잡아가고 있는 현재, 언어병리학에 딥러닝 연구 방법론을 접목하여 연구의 도구와 관점을 넓히는 일은 충분한 타당성을 지닌다.

REFERENCES

- Ambrose, N. G., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language, and Hearing Research*, 42(4), 895-909.
- Bloodstein, O. (1992). Response to Hamre: Part I. *Journal of Fluency Disorders*, 17(1-2), 29-32.
- Bloodstein, O. (2002). Early stuttering as a type of language difficulty. *Journal of Fluency Disorders*, 27(2), 163-166.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word

- vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135-146.
- Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, 6, 213-234.
- Bóna, J. (2019). Clustering of disfluencies in typical, fast and cluttered speech. *Clinical Linguistics & Phonetics*, 33(5), 393-405.
- Brocklehurst, P. (2008). Phonological encoding in children who stutter (CWS): is it impaired?. *Contemporary Issues in Communication Science and Disorders*, 35, 25-43.
- Byrd, C. T., Conture, E. G., & Ohde, R. N. (2007). Phonological priming in young children who stutter: holistic versus incremental processing. *American Journal of Speech-Language Pathology*, 16(1), 43-53.
- Campbell, J., & Hill, D. (1987). Systematic disfluency analysis: accountability for differential evaluation and treatment. *Miniseminar presented to the Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA*.
- Chon, H. C. (2010). Stuttering-like disfluencies in phonological words in preschool children who stutter. *Korean Journal of Communication & Disorders*, 15(3), 422-432.
- Choi, S., Kim, T., Seol, J., & Lee, S. G. (2017). A syllable-based technique for word embeddings of Korean words. *Proceedings of the First Workshop on Subword and Character Level Models in NLP*, 36-40.
- Conture, E. G. (1990). *Stuttering* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Cordes, A. K., & Ingham, R. J. (1994). The reliability of observational data: II. issues in the identification and measurement of stuttering events. *Journal of Speech and Hearing Research*, 37(2), 279-294.
- Firth, J. R. (1957). *A synopsis of linguistic theory 1930-55*. Reprinted in E.R. Palmer (Ed.), (1968), *Selected papers of J.R. Firth 1952-59* (pp. 1-32). London: Longman
- Ganegedara, T. (2018). *Natural language processing with TensorFlow: teach language to machines using Python's deep learning library*. Packt Publishing Ltd.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge: The MIT Press.
- Ha, J. W., & Sim, H. S. (2008). A comparison study of interjectional characteristics between people who stutter and people who do not stutter. *Korean Journal of communication disorders*, 13(3), 438-453.
- Hamre, C. (1992). Stuttering prevention I: primacy of identification. *Journal of Fluency Disorders*, 17(1-2), 3-23.
- Han, J. S., Lee, E. J., & Sim, H. S. (2005). The phonemic characteristics of disfluencies in children and adults who stutter. *Korean Journal of Speech Sciences*, 12(3), 59-77.
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3), 146-162.
- Howell, P. (2004). Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary Issues in Communication Science and Disorders*, 31, 123-140.
- Howell, P., & Au-Yeung, J. (2002). *The EXPLAN theory of fluency control and the diagnosis of stuttering*. In E. Fave (Ed.), *Current issues in linguistic theory series: pathology and therapy of speech disorders* (pp. 74-94). Amsterdam: John Benjamins.
- Howell, T. A., & Bernstein Ratner, N. (2018). Use of a phoneme monitoring task to examine lexical access in adults who do and do not stutter. *Journal of Fluency Disorders*, 57, 65-73.
- Hubbard, C., & Prins, D. (1994). Word familiarity, syllabic stress pattern, and stuttering. *Journal of Speech and Hearing Research*, 37(3), 564-571.
- Hubbard, C. P., & Yairi, E. (1988). Clustering of disfluencies in the speech of stuttering and nonstuttering preschool children. *Journal of Speech and Hearing Research*, 31(2), 228-233.
- Jeon, H. S., & Jeon, H. E. (2015). Characteristics of disfluency clusters in adults who stutter. *Journal of Speech-Language & Hearing Disorders*, 24(1), 135-144.
- Johnson, W., & Associates. (1959). *The onset of stuttering*. Minneapolis, MN: University of Minnesota Press.
- Jurafsky, D., & Martin, J. H. (2019). *Speech and Language Processing* (3rd ed. draft). Retrieved from <https://web.stanford.edu/~jurafsky/slp3/>.
- Kolk, H., & Postma, A. (1997). *Stuttering as covert repair phenomenon*. In R. F. Curlee & G. M. Siegel (Eds.), *Nature and treatment of stuttering*. Boston: Allyn & Bacon.
- Kourkounakis, T., Hajavi, A., & Etemad, A. (2020). Detecting multiple speech disfluencies using a deep residual network with bidirectional long short-term memory. *Proceedings of the 2020 IEEE international conference on Acoustics, Speech and Signal Processing*, 6089-6093.
- Kwon, D. H. (1991). The onset of stuttering behaviors (2): in the phonological aspects of language contexts. *Journal of Speech-language & Hearing Disorders*, 1, 147-165.
- LaSalle, L. R., & Conture, E. G. (1995). Disfluency clusters of children who stutter: Relation of stutterings to self repairs. *Journal of Speech, Language, and Hearing Research*, 38(5), 965-977.
- Lee, E. J., Han, J. S., & Sim, H. S. (2004). The effects of the phonetic complex-

- ity on the disfluencies and the articulation errors of people who stutter. *Korean Journal of Communication & Disorders*, 9(3), 139-156.
- Lee, E., & Sim, H. S. (2003). A preliminary study for the test of covert repair hypothesis by the analysis of disfluencies during spontaneous utterance. *Korean Journal of Communication & Disorders*, 8(1), 201-216.
- Manning, W. H. (2010). *Clinical decision making in fluency disorders*. (3rd ed.). Clifton Park, NY: Delmar Cengage Learning.
- Manning, W. H., & Shirkey, E. A. (1981). *Fluency and the aging process*. In D. S. Beasley & G. A. Davis (Eds.), *Aging: Communication Processes and Disorders* (pp. 175-189). New York, NY: Grune & Stratton.
- Meyers, S. C. (1986). Qualitative and quantitative differences and patterns of variability in disfluencies emitted by preschool stutterers and nonstutterers during dyadic conversations. *Journal of Fluency Disorders*, 11(4), 293-306.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *Proceedings of the 2013 International Conference on Learning Representations*, 1301-3781.
- Navarro-Ruiz, M. I., & Rallo-Fabra, L. (2001). Characteristics of mazes produced by SLI children. *Clinical Linguistics and Phonetics*, 15(1-2), 63-66.
- Newman, R. S., & Bernstein Ratner, N. (2007). The role of selected lexical factors on confrontation naming accuracy, speed, and fluency in adults who do and do not stutter. *Journal of Speech and Hearing Research*, 50(1), 196-213.
- Nippold, M. A. (2002). Stuttering and phonology: is there an interaction?. *American Journal of Speech-Language Pathology*, 11(2), 99-110.
- Park, S., Byun, J., Baek, S., Cho, Y., & Oh, A. (2018). *Subword-level word vector representations for Korean*. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2429-2438.
- Park, C. W., Lim, Y. S., & Sung, J. E. (2019). Machine-learning and corpus-based analyses of Korean nouns from subjects and objects: stimuli development of verb treatment for neurogenic communication disorders, *Communication Sciences & Disorders*, 24(4), 968-985.
- Pennington, J., Socher, R., & Manning, C. D. (2014). *Glove: global vectors for word representation*. In Proceedings of the 2014 conference on empirical methods in natural language processing, 1532-1543.
- Penttilä, N., Korpiaakko-Huuhka, A. M., & Kent, R. D. (2019). Disfluency clusters in speakers with and without neurogenic stuttering following traumatic brain injury. *Journal of Fluency Disorders*, 59, 33-51.
- Postma, A., & Kolk, H. (1993). The covert repair hypothesis: prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, 36(3), 472-487.
- Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45-50.
- Ronson, I. (1976). Word frequency and stuttering: the relationship to sentence structure. *Journal of Speech and Hearing Research*, 19(4), 813-819.
- Sawyer, J., & Yairi, E. (2010). Characteristics of disfluency clusters over time in preschool children who stutter. *Journal of Speech and Hearing Research*, 53(5), 1191-1205.
- Silverman, E. (1973). Clustering: a characteristic of preschoolers' speech disfluency. *Journal of Speech and Hearing Research*, 16(4), 578-583.
- Shin, M. J. (1996). Characteristics of fluency disorders. *Speech-Language Disorder Research*, 1, 82-103.
- Shin, G. E., & Ha, J. W. (2015). The effects of phonetic complexity on the disfluency and articulation errors of children with speech sound disorders. *Journal of Speech-language & Hearing Disorders*, 24(1), 91-102.
- Sim, H. S., Shin, M. J., & Lee, E. J. (2010). *Paradise Fluency Assessment-II*. Seoul: Paradise Welfare Foundation.
- Smith, A., & Weber, C. (2017). How stuttering develops: the multifactorial dynamic pathways theory. *Journal of Speech, Language, and Hearing Research*, 60(9), 2438-2505.
- Starkweather, C. W. (1992). Response and reaction to Hamre, "Stuttering prevention I." *Journal of Fluency Disorders*, 17(1-2), 43-55.
- Van Riper, C. (1982). *The nature of stuttering* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Wexler, K. B., & Mysak, E. D. (1982). Disfluency characteristics of 2-, 4-, and 6-yr-old males. *Journal of Fluency Disorders*, 7(1), 37-46.
- Wingate, M. E. (1988). *The structure of stuttering*. New York, NY: Springer Verlag.
- Yairi, E., & Ambrose, N. G. (2005). *Early childhood stuttering: for clinicians by clinicians*. Austin, TX: Pro-Ed.
- Yairi, E., Ambrose, N., Paden, E., & Throneburg, R. (1996). Predictive factors of persistence and recovery: pathways of childhood stuttering. *Journal of Communication Disorders*, 29(1), 51-77.
- Yairi, E., & Lewis, B. (1984). Disfluencies at the onset of stuttering. *Journal of Speech and Hearing Research*, 27(1), 154-159.
- Yairi, E., & Seery, C. H. (2015). *Stuttering: foundations and clinical applications* (2nd ed.). Boston: Pearson Education, Inc.
- Yaruss, J. S., LaSalle, L. R., & Conture, E. G. (1998). Evaluating stuttering in young children: diagnostic data. *American Journal of Speech Language Pathology*, 7(4), 62-76.

- Yoon, H., & Chon, H. (2019). Linguistic disfluency in retelling of cause-effect and compare-contrast expository discourses for third to fourth graders from multicultural families. *Audiology Speech Research*, 15(4), 283-292.
- Zebrowski, P. M. (1994). *Stuttering*. In J. B. Tomblin, H. L. Morris & D. C. Spriestersbach (Eds.), *Diagnosis in speech-language pathology*. San Diego: Singular Publishing Group.

국문초록

딥러닝-워드임베딩을 기반으로 한 말더듬 대상자의 읽기과제 비유창성 분석

송상현¹ · 전희정² · 이수복³¹고려대학교 언어학과, ²조선대학교 언어치료학과, ³우송대학교 언어치료청각재활학과

배경 및 목적: 최근의 딥러닝 자연어 처리는 언어단위를 수치 벡터로 변환하여 공간상에서 연산을 도모하는 임베딩 기술을 활용한다. 본 연구는 이 기법을 언어병리학 유창성장애 데이터에 적용하여 비유창성의 위치와 분포 특성을 파악하고자 하였다. **방법:** 110명의 중학생 이상 청소년 및 말더듬 성인의 읽기발화(800음절) 데이터를 음소 단위로 분절한 뒤 수치 벡터로 변환하여 거리 연산을 수행하였다. Word2Vec을 활용하여 코사인 유사도를 측정하여 각 비유창성 유형 별 유사성을 도출하고, 또한 전체 데이터를 t-SNE 그림으로 모델을 시각화하여 제시하였다. 또한, 음소 환경을 분석하고자 파라다이스-유창성검사-II의 읽기발화와 세종 코퍼스 데이터를 비교 분석하였다. **결과:** 첫 번째, 총 8개의 ND 유형('UR', 'I', 'H', 'RI')과 AD 유형('URa', 'Ia', 'Ha', 'RIa')은 .9 이상의 유사도로 근접하여 출현하였다. 두 번째, ND 유형과 AD 유형 간의 분포적 차이를 분석한 결과, 분포적 차이가 크지 않은 것으로 나타났다. 또한, AD 유형은 상당히 높은 유사도로 AD 유형 간에 발생 위치가 중첩되는 것으로 나타났고 'R2'와 'DP'는 다른 비유창성 유형과 다른 양상을 보였다. 마지막으로, 음소 환경에 따른 비유창성 출현 빈도는 'ㄹ', 'ㄷ', 'ㅈ' 음소에서 비유창성 빈도가 비교적 높게 나타났다. **논의 및 결론:** 본 연구는 딥러닝 기법을 활용하여 비유창성 유형들 간의 군집화 특성, 정상적 비유창성과 비정상적 비유창성 사이의 분포적 차이가 크지 않다는 것, 그리고 비유창성 출현 음소 등을 확인하였고 이러한 결과는 유창성장애 진단과 치료에 활용될 수 있을 것이다.

핵심어: 딥러닝, 워드임베딩, 코사인 유사도, 말더듬 읽기과제, 정상적 비유창성, 비정상적 비유창성

참고문헌

- 권도하(1991). 말더듬 발생(2): 음운론적 측면에서. *언어치료연구*, 1, 147-165.
- 박채원, 임윤섭, 성지은(2019). 신경언어장애군의 동사 중재 자극 개발을 위한 기계학습 및 빅데이터 기반 한국어의 주격 및 목적격 명사 분석. *Communication Sciences & Disorders*, 24(4), 968-985.
- 신가은, 하지완(2015). 말소리장애 아동의 조음복잡성에 따른 비유창성 및 조음오류 특성. *언어치료연구*, 24(1), 91-102.
- 신문자(1996). 유창성장애의 특성 연구. *말-언어장애연구*, 1, 82-103.
- 심현섭, 신문자, 이은주(2010). *파라다이스 유창성 검사-II*. 서울: 파라다이스 복지재단.
- 이은주, 심현섭(2003). 내적수정가설 검증을 위한 기초 연구: 말더듬 및 정상 성인의 자발화에서 나타난 비유창성을 중심으로. *언어청각장애연구*, 8(1), 201-216.
- 이은주, 한진순, 심현섭(2004). 조음복잡성이 비유창성과 조음오류에 미치는 영향. *언어청각장애연구*, 9(3), 139-156.
- 윤호진, 전희정(2019). 초등 3-4학년 다문화가정 아동의 원인-결과와 비교-대조 설명담화 다시 말하기에 나타난 언어학적 비유창성 특성. *Audiology and Speech Research*, 15(4), 283-292.
- 전희숙, 전효은(2015). 말더듬성인의 비유창성군 특성. *언어치료연구*, 24(1), 135-144.
- 전희정(2010). 음운적 단어에 나타난 학령전기 말더듬 아동의 비유창성 발생 위치. *언어청각장애연구*, 15(3), 422-432.
- 하지완, 심현섭(2008). 유창성장애 집단과 정상 집단의 간투사 비교연구. *언어청각장애연구*, 13(3), 438-453.
- 한진순, 이은주, 심현섭(2005). 말더듬 아동과 성인에게서 나타난 비유창성의 음운특성. *음성과학*, 12(3), 59-77.

ORCID

송상현(제1저자, 교수 <https://orcid.org/0000-0002-4234-232X>); 전희정(공동저자, 교수 <https://orcid.org/0000-0002-9715-7164>); 이수복(교신저자, 교수 <https://orcid.org/0000-0001-6942-4635>)