

인공지능 학습을 위한 중노년층 방언 음성 데이터 구축 전략 및 분석*

옥성수** · 김수연***

(세종대학교)

Ok, Sungsoo & Kim, Soo-Yeon. (2024). Strategies and analysis for constructing middle-aged and elderly dialect speech data for artificial intelligence training. *The Linguistic Association of Korea Journal*, 32(1), 1-19. This paper presents a comprehensive strategy and analysis for constructing dialect speech data for middle-aged and elderly populations aimed at enhancing artificial intelligence (AI) training. Recognizing the critical role of high-quality, diverse speech datasets in improving AI's real-world performance, especially in speech recognition, this study focuses on the underrepresented dialects of older demographics. It outlines the methodologies employed in collecting, processing, and labeling the speech data, ensuring the inclusion of various dialectal nuances, intents, and emotional states. Additionally, the paper discusses the project's challenges, including ensuring data diversity and the technical aspects of data processing. By addressing these areas, the research contributes to the development of AI systems better attuned to the linguistic diversity and needs of older users, potentially improving AI accessibility and user experience across different applications.

주제어(Key Words): 인공지능 학습용 데이터(AI training data), 노인 음성(elderly speech), 방언 데이터(dialect data), 데이터 라벨링(data labeling), 발화 의도(speech intent), 감성(emotion)

* 이 논문은 2023년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2023S1A5A2A03087587).

** 제1저자

*** 교신저자

1. 서론

인공지능(이하, “AI”) 학습용 데이터의 양과 질은 AI 모델이 실세계에서 다양한 상황과 사용자를 얼마나 잘 이해하고 대응할 수 있는지를 결정하는 핵심 요소이다. 대규모 데이터셋은 모델이 더 다양한 시나리오와 상황을 학습할 수 있게 해주는데, 이는 특히 음성 인식과 같이 다양한 발음, 억양, 방언을 처리해야 하는 경우에 중요하다. 데이터가 많을수록, 모델은 더 넓은 범위의 음성 패턴과 언어적 특성을 포괄할 수 있으며, 이는 실세계의 복잡한 상황에서 더 나은 성능을 발휘하는 데 기여한다. 데이터 질의 측면에서는, 고품질의 데이터는 모델이 더 정확하고 신뢰할 수 있는 예측을 할 수 있도록 한다. 데이터의 질은 녹음의 명료성, 배경 소음의 최소화, 정확한 전사 및 발화 의도나 감성 등을 포함한 라벨링과 같은 요소들에 의해 결정된다. 높은 질의 데이터는 모델이 잘못된 정보를 학습하는 것을 방지하고, 더 정확한 음성 인식 및 자연어 처리 능력을 개발하는 데 필수적이다.

다양한 음성 데이터 중에서도 노년층 음성 데이터 구축의 중요성은 사회적 배경과도 맞닿아 있다. 2023년 통계청 자료에 따르면, 대한민국의 65세 이상 노년층 인구 비율은 전체의 18%이며, 2025년에는 20%를 돌파하여 초고령사회(Post-Aged Society)로 진입할 것으로 예측된다. 이러한 상황에서 음성인식 기반의 스마트홈 기기 제어, 통신, 가상 비서 등의 서비스는 노년층의 더 안전하고 편리한 생활에 기여할 수 있다. 하지만, 일반적으로 노인 음성의 인식률은 젊은 성인에 비해 낮은 경향이 있다. 이는 노화로 인한 음성의 변화, 발음의 명확성 감소, 말의 속도 변화 외에도 문장 길이 및 복잡성 감소, 구문적 다양성 감소와 같은 구조적인 변화 등 여러 요인 때문인데(Linville & Rens, 2001; Kemper et al., 2003; Harnsberger et al., 2008; Horton & Spieler, 2010; Beese et al., 2019), Fukuda et al.(2020)는 음성인식 시스템이 이러한 변화를 정확하게 인식하고 해석하는 데 어려움을 겪을 수 있으며, 이것으로 인하여 인식률이 저하될 수 있으므로, 노인 음성의 독특한 특성을 반영한 모델 개발을 통해 인식률을 개선할 필요성을 제기했다. 노년층 언어의 또 하나의 특징으로는 높은 지역 방언 사용 비율을 들 수 있다. 세대별 언어사용과 의식의 변화를 탐구한 사회언어학적 연구들은 노인 세대에서 방언 사용이 더욱 두드러질 수 있음을 시사한다(이기갑, 2008; 류성기, 2016; 2017). 방언 음성 데이터를 AI 학습용 데이터로 구축할 경우, 방언의 지리적, 성별, 연령별 다양성을 포함하는 광범위한 데이터를 수집하는 것이 중요하고, 그 데이터가 해당 지역의 방언으로서의 대표성을 가지는지를 고려하여야 한다. 또한, 방언의 특성을 정확히 반영하는 텍스트 및 억양, 모음의 장단 등 음운정보를 포함한 라벨링 역시 학습용 데이터의 질을 높이는 데 중요하다.

본 논문에서는 2022년 과학기술정보통신부가 주관하고 한국지능정보사회진흥원(NIA)이 지원하여 수행된 인공지능 학습용 데이터 구축 사업 중 “2-18 중노년층 방언 데이터” 과제(이하, “NIA 2-18 과제”)의 데이터 설계, 수집, 가공의 과정을 소개하고, 구축된 데이터의

분석을 진행하고자 한다. 해당 과제는 증가하는 중노년층의 AI 음성 인식 서비스 의존도와 그에 반해 상대적으로 부족한 AI 학습용 중노년층 음성 및 그들의 방언 데이터 간의 괴리를 좁히고자 기획되었다. NIA 2-18 과제는 단순한 음성과 텍스트의 페어링을 넘어, 음성의 억양 및 화자의 발화 의도, 감성, 문장 유형 등의 라벨링 정보가 포함된 최적화된 AI 학습용 데이터 구축을 목표로 수행되었다.

2. 데이터 설계 및 수집

2.1. 데이터 설계

NIA 2-18 과제는 한국어 5개 방언 권역별로 50대부터 90대까지의 중노년층 남녀의 음성 데이터를 수집하여 총 4,000시간 이상의 라벨링된 원천데이터 구축을 목표로 하였다. 총시간 중 2,000시간은 강원, 경상 방언 발화, 나머지 2,000시간은 충청, 전라, 제주 방언 발화 데이터로 구성되었다. 구축 목표량은 각 지역의 인구수를 고려하여 아래와 같이 할당되었다.

표 1. 방언 지역별 데이터 구축 목표량

지역	강원	경상	충청	전라	제주
시간	800시간	1,200시간	800시간	1,000시간	200시간

중노년층의 음성 특징과 지역별 방언 특징을 최대한 포착하여 다각적으로 활용 가능한 데이터를 구축하기 위해서는 원시 데이터 수집 단계에서부터 목적에 특화된 발화 유도들의 설계가 필수적이다. 이에, 해당 과제에서는 중노년층에 특화된 발화 도메인의 설정과 함께, 스크립트를 기반으로 한 정형 발화와 특정 주제를 바탕으로 한 비정형 발화(자유 발화)를 모두 획득할 수 있는 유도들을 개발하였다. 발화 도메인을 설정하는 것은 수집된 음성 데이터의 다양성과 품질을 보장하고, 특정 응용 프로그램이나 연구 목적에 맞게 데이터를 최적화하는 데 도움이 된다. 도메인을 특정하면, 발화 내용이 특정 상황이나 주제에 초점을 맞추게 되어 AI 모델이 더 정확하고 효과적으로 해당 도메인에 대한 언어 패턴을 학습할 수 있는데, 심우창 외(2023)는 모델에게 적합한 지시를 내리는 것이 중요하다는 점을 강조하며, 특정 도메인에 초점을 맞춘 언어모델의 학습 효율성과 정확도를 높일 수 있는 방법을 제시하였다. NIA 2-18 과제에서는 아래와 같이 총 9가지의 발화 도메인을 설정하였다.

표 2. 발화 도메인 및 세부 주제

도메인	세부 주제
농경	경작/타작/농기구 등
가족·생활	가문, 가족 소개/결혼/자녀 등
의	복식과 장식 등
식	주식과 부식/농촌의 식생활 등
주	가옥 현대화 (과거 가옥과 비교 등)
자연	산과 들/강과 바다/날씨와 방향 등
건강	각종 질병과 의료 환경 등
풍속	민속놀이/전통 등
응급상황	화재/사고 등

발화 유형 중 정형 발화의 경우, 방언 출현 빈도를 높이기 위해 방언이 포함된 도별 스크립트를 제작하여 각 도당 약 300개의 데이터 유형, 총 1,500개의 데이터 유형을 만들었다. 정형 발화가 가져올 어색한 발화를 지양하기 위하여 도별 방언 화자가 미리 자연스러운 억양으로 녹음한 음성 파일을 스크립트와 함께 수집 앱에 탑재하였다. 이를 통해, 수집 시 화자에게 해당 음성 파일을 들려준 후 따라말하기를 유도하도록 정형 발화 유도틀을 설계하였다.

표 3. 정형 발화 스크립트 예시(제주)

번호	도메인	방언 표기	표준어 표기
1	가족·생활	죽은사원 우리 딸신디 더할 나우 엇인 사름입주.	작은사위는 우리 딸에게 더할 나위 없는 사람이지요.
2	건강	자이 풀에 독술 막난 거 보난 언 생이우다.	재 팔에 닭살이 많이 난 거 보니까 추운 모양입니다.
3	풍속	글피 모듬벌초 헌덴 허난 아방 산이랑 널 헤 볼게.	글피 모듬벌초 한다고 하니 아버지 되는 내일 해 버리자.
4	식	할망 얼굴이 발강혜신디 약주 혼 잔 드셨수과?	할머니 얼굴이 발강데 약주 한 잔 드셨습니까?
5	농경	조나 보린 몰래와디 농서가 잘 안돼.	조나 보리는 모래밭에 농사가 잘 안돼.




비정형 발화는, 최대한 자연스러운 발화를 유도하고 끊이지 않는 지속적인 발화가 가능할 수 있도록 노인들의 관심 주제를 선정, 주제별로 두 가지 유형의 발화 유도 형태를 설정하였다. 첫 번째는, 질문에 답하기 유형으로, 수집자가 먼저 주제별 질문을 제시한 후 화자로 하여금 주제에 맞는 본인의 답을 자연스럽게 발화하도록 유도하였다. 질문은 화자가 많은 생각의 준비나 머뭇거림, 어려움 없이 직관적으로 대답할 수 있도록 구성되었다.

표 4. 비정형 발화(질문에 답하기) 질문 예시

번호	도메인	질문
1	가족·생활	(저는 1남 1녀인데 누나는 지금 서울에 살아요.) 어르신께서는 형제 관계가 어떻게 되세요? 다들 어디에 사세요?
2	주	(저는 어릴 때부터 같은 집에서 계속 살고 있는데요.) 지금 집에서는 얼마나 사셨어요? 이사하신 적 있으신가요?
3	농경	(저희 아버지께서는 귀농하셔서 농사도 조금 지으시는데요.) 농사지어 보신 적 있으세요? 무슨 작물을 재배해 보셨나요? 어떤 작물이 재배하기 좋아요?
4	식	(얼마 전에 친구 어머니께서 부추를 엄청 보내 주셨는데요.) 부추로 무슨 요리를 해 먹으면 맛있을까요? 좋아하는 식재료 있으세요?
5	건강	(저는 아침에 일어나자마자 물 한잔을 마시는데요.) 아침에 일어나면 제일 먼저 뭐 하세요? 건강을 위해서 특별히 하시는 거 있으세요?
6	의	(저는 주로 옷을 인터넷에서 사는 편인데요.) 지금 입고 계신 옷은 언제 어디에서 사셨어요? 마음에 드세요? 어떤 옷을 좋아하세요?
7	자연	(저는 봄마다 꽃구경을 가는데요.) 봄에 친구분들이랑 꽃구경 가신 적이 있으세요? 언제 어디로 가셨어요? 어떠셨어요?
8	풍속	(저는 명절 음식을 좋아해서 추석이 기다려지는데요.) 명절에 어떤 음식을 준비하시나요? 명절에 꼭 하시는 음식이 있으세요? 그 음식은 어떻게 만드시나요?
9	응급상황	(지난달에 저희 동네에 불이 났는데, 다행히 빨리 꺾어요.) 만약 불이 나면 제일 먼저 무엇을 챙기시겠어요? 그 이유가 뭐예요?

두 번째는, 2인 대화 유형으로, 두 명의 화자가 대화를 이끌어갈 수 있도록 찬반을 유도하는 질문을 설정하였다. 2인 대화 유형은 턴 테이킹 방식의 자유 발화로, 대화를 통해 화자의 발화에 다양한 의도와 감성이 반영되는 것을 목표로 하였다. 질문에 답하기와 마찬가지로 실생활과 관련된 주제 등을 선정하여 자연스러운 대화를 유도하였다. 또한, 화자가 고령자임을 고려하여 질문 내용을 삽화로 제작한 후 수집 시 제시하여 대화 주제의 이해를 돕도록 하였다.

표 5. 비정형 발화(2인 대화) 질문 예시

번호	도메인	삽화	질문	보조질문
1	가족 · 생활	 <p>대가족 핵가족</p>	<p>(대가족과 핵가족 그림입니다. 예전에는 3대 이상 함께 사는 대가족이 많았지만, 요즘은 찾아보기가 어려운데요. 다음 1, 2번 중 하나씩 선택하셔서 서로 대화해 주세요.) 1. 대가족이 더 좋다. 2. 핵가족이 더 좋다.</p>	<p>지금은 누구와 함께 사세요? 어렸을 때는 몇 대가 함께 사셨어요? 요즘 1인 가구가 점점 늘고 있는데, 혼자 살아서 좋은 점과 나쁜 점은 각각 뭐라고 생각하세요?</p>
2	풍속	 <p>옷놀이 화투</p>	<p>(옷놀이 하는 그림과 화투 치는 그림입니다. 명절에 옷놀이 하는 집도 있는 반면에 화투 놀이를 즐기는 집도 있죠. 다음 1, 2번 중 하나씩 선택하셔서 서로 대화해 주세요.) 1. 명절에는 전통 놀이를 한다. 2. 명절에는 화투나 다른 놀이를 한다.</p>	<p>어릴 때 즐겨 하였던 놀이는 어떤 것이 있으셨어요? 어떻게 하는 놀이인지 그 방법을 말씀해 주세요. 이 동네에서 사람들이 모여서 하던 놀이가 있으면 말씀해 주세요.</p>
3	건강	 <p>운동 음식</p>	<p>(운동하는 그림과 삼계탕 먹고 있는 그림인데요. 다음 1, 2번 중 하나씩 선택하셔서 서로 대화해 주세요.) 1. 건강을 위해 운동을 하는 것이 더 중요하다. 2. 건강을 위해 몸에 좋은 음식을 잘 챙겨 먹는 것이 더 중요하다.</p>	<p>복날이면 삼계탕을 꼭 챙겨 드세요? 선호하시는 건강보양식이나 건강을 위해 꼭 챙겨 드시는 보양식이 있으세요? 꾸준히 하고 계시는 운동이 있으신가요?</p>

2.2. 데이터 수집

음성 수집은 NIA 2-18 과제를 위해 제작된 음성 수집용 모바일 앱을 통해 진행되었다. 수집자들은 수집 앱이 설치된 모바일 기기를 휴대하고 다녔으며, 앱을 통해 발화자의 성별, 출생년도, 거주지역, 거주기간, 직업, 학력, 그리고 치아손실 등의 건강상태와 같은 메타 데이터 정보를 입력한 후 녹음을 진행하였다. 거주기간의 경우 해당 지역에서 출생하여 최소 15년 이상 거주한 사람을 음성 수집 대상으로 하였다.



그림 1. 음성 수집 앱 화면 예시

중노년층 방언 화자와 직접 대면하여 수집하는 상황에서 발화자가 자연스러운 분위기에서 발화할 수 있도록 수집자는 모두 해당 수집 지역의 방언 화자로 지정하였다. 녹음 중에는 수집자의 음성을 포함하여 불필요한 소리가 녹음되지 않도록 소음을 최대한 통제하도록 하였으며, 보조배터리 등을 휴대하여 모바일 장비의 전원을 충분히 확보하도록 하였다. 특히, 중노년층 화자를 배려하여 수집 앱에 탑재된 수집 스크립트와 그림을 크게 출력하여 소지한 후, 필요시 발화자에게 보조도구로 제시하도록 하였다. 수집한 발화 유형은 1시간 음성 기준으로 정형 발화 따라말하기 18분(30%), 비정형 발화 질문에 답하기 30분(50%), 그리고 비정형 발화 2인 대화 12분(20%)의 비율로 수집하도록 하였으며, 발화자에게는 녹음량 기준 소정의 참여비가 지급되었다.

3. 데이터 가공

3.1. 정제 및 전사

AI 학습을 위해 수집된 음성 데이터에 텍스트 전사(transcription)를 하는 것은 정확한 데이터 학습을 위해 필수적이다. AI 시스템, 특히 음성 인식 시스템은 다양한 음성 데이터로부터 학습하여 정확한 텍스트 변환을 수행할 수 있어야 한다. 음성 데이터에 대한 정확한 텍스트 전사는 이러한 시스템이 실제 발화와 그에 해당하는 정확한 텍스트 매핑을 학습하는 데 필수적이다. 또한, 다양한 발화 스타일, 방언 등을 포함한 음성 데이터의 텍스트 전사는 AI가 더 넓은 범위의 음성 변형을 인식하고 이해하는 데 도움을 줄 수 있다. 음성 데이

터의 텍스트 전사는 음성 인식뿐만 아니라 자연어 처리(NLP) 작업에도 중요한데, 전사된 텍스트를 통해 감정 분석, 의도 인식, 기계 번역 등과 같은 고급 자연어 처리 작업을 수행할 수 있다. 한편, 음성과 텍스트 데이터의 결합은 멀티모달 학습을 가능하게 한다. 이는 AI가 더 풍부한 정보를 학습하고, 음성 데이터만으로는 파악하기 어려운 맥락이나 의미를 이해하는 데 유용하다.

NIA 2-18 과제에서는, 수집된 음성 데이터의 전사를 진행하기 전에 정제 작업이 선행되었다. 정제는 정형 발화의 경우 문장 단위로, 비정형 발화의 경우 최대 2분이 넘지 않도록 재가공 되었으며¹⁾, 각각의 발화가 하나의 웨이브(WAV) 파일로 전사 작업자들에게 전달되었다. 정제 단계에서는 또한, 배경 소음 제거와 같은 오디오 품질 개선이 이루어졌으며, 불필요한 부분(잡음, 긴 침묵 구간²⁾ 등)을 제거하고, 발화의 시작과 끝이 명확하도록 조정하였다. 마지막으로 자동 신호 처리(automatic signal processing)을 통해 음성의 기계적 분석이 이루어졌으며, 분석된 음성의 음높이 변화를 시각적으로 표현한 피치 그래프(pitch graph)를 전사 및 라벨링 도구에 탑재하여 작업자들의 라벨링 작업을 보조하도록 하였다.

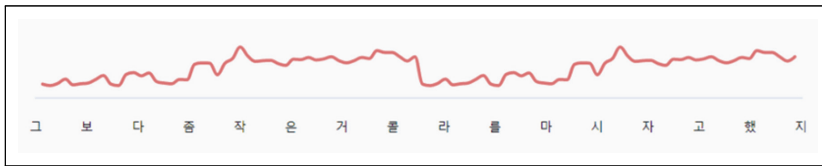


그림 2. 피치 그래프 예시

전사 단계에서는 음성 데이터 전사 전문 작업자들이 투입되었다. NIA 2-18 과제의 전사는 해당 과제를 위해 개발된 전사 도구를 통해 이루어졌다. 정제된 음성 데이터는 네이버에서 제공하는 CLOVA Speech Recognition(CSR)을 통해 텍스트로 변환(Speech-to-Text, STT)

- 1) 비정형 발화 데이터를 최대 2분으로 제한한 결정은 데이터 라벨링 작업자의 작업 효율성과 인공지능 학습용 데이터의 품질 관리라는 두 가지 중요한 목표에 기반을 두고 있다. 첫째, 라벨링 작업의 효율성 측면에서, 긴 발화 데이터는 작업자의 주의 분산과 피로도를 증가시킬 수 있다. 이는 오류 발생률을 높이고, 따라서 전체 데이터셋의 품질에 부정적인 영향을 미칠 수 있다. 이러한 양상은 특히, 본 과제와 같이 발화 의도, 감정, 문장 유형 등 복잡한 라벨링 정보를 요구하는 작업에서 더 두드러지며, 실제 작업에서도 긴 음성 파일에 대한 작업자들의 피로도가 보고되기도 했다. 둘째, AI 학습용 데이터로서의 품질을 고려할 때, 비정형 발화의 길이를 적절하게 관리하는 것은 중요하다. 너무 긴 발화로 인해 특히 초기 학습 단계에서는 모델이 중요한 언어적 특성과 패턴을 파악하는 데 어려움을 겪을 수 있다.
- 2) 발화 중의 침묵은 화자의 감정이나 의도를 파악하는 데 중요한 단서가 될 수 있다. 이러한 침묵은 감정적 긴장, 망설임, 생각하는 시간 등 다양한 의미를 내포할 수 있으므로, 이러한 침묵 구간도 데이터에 포함시키는 것이 추후 관련 연구를 위해서 필요할 수 있다. 따라서, 본 과제에서는 발화 중 나타난 침묵은 유지하되, 발화 전후에는 0.5초에서 1초간의 침묵 구간만 유지하는 방식으로 데이터 정제를 실시하였다.

되어 전사 도구 상에 제시되었으며, 전사 작업자들은 정제된 음성과 거기에 대응되는 텍스트를 참고하여 어절 단위로 전사를 진행하였다.



그림 3. 전사 도구 화면 예시

해당 과제에서는 방언 데이터의 특징을 감안하여 3중 전사가 이루어졌는데, 표 6의 예시와 같이 표준어 및 방언 표기법에 맞는 ‘철자 전사’가 1차로 진행되었고, 방언에 해당하는 어절에 한해 2차로 발음 표기법을 바탕으로 소리값에 최대한 가까운 ‘방언 발음 전사’, 그리고 3차로 해당 방언 어절에 대응하는 ‘표준어’가 제시되었다.

표 6. 3중 전사 예시

1차	철자 전사	아침에	새복시장에	가서	쌈	한가마를	사부렀당께
2차	방언 발음 전사		새복씨장예				사부러당께
3차	표준어		새벽시장에				사버렀다니까

그 외에, 데이터의 일관성과 정확성을 보장하기 위해 TN(Text Normalization)을 적용해 숫자, 외국어를 포함한 모든 발화는 한글로 전사되었다. TN은 비정형화된 텍스트를 읽을 수 있는 형태로 변환하는 과정으로, 예를 들어 숫자 “2024”를 “이천이십사”로 변환한다. 전사 과정에서 문장부호 및 특수문자는 쓰지 않았으며, 간투사, 웃음소리, 반복되는 발화 등도 모두 전사되었다. 해당 과제에서는 라벨링이 문장 단위로 이루어졌기 때문에, 전사 단계에서 데이터를 문장 단위로 나누는 작업도 실시되었다. 1차 작업자의 작업물은 두 차례의 검수를 거쳐 반려되지 않은 데이터에 한해 라벨링 작업자들에게 전달되었다.

3.2. 라벨링

AI 학습용 데이터로서 수집된 음성 데이터에 라벨링을 하는 것은 매우 중요한 과정이다. 라벨링은 데이터에 태그나 분류 정보를 추가하는 작업으로, AI 모델이 데이터를 이해하고 학습하는 데 필수적인 기반을 제공한다(장영우 외, 2023; Khatun & Sarkar, 2024 등). AI 모델은 제공된 데이터로부터 패턴을 학습하므로, 라벨링된 음성 데이터는 모델에게 어떤 음성 특성이 중요한지, 특정 발화가 어떤 의미를 가지는지 알려준다. 이를 통해 모델은 정확하게 학습하여 음성 인식, 감정 분석, 언어 이해 등의 작업을 더 잘 수행할 수 있다. 또한, 효과적인 라벨링은 학습 과정을 최적화하여 모델이 더 빠르게 그리고 더 적은 데이터로도 학습할 수 있도록 돕는데, 이는 개발 비용과 시간을 절약할 수 있게 한다. 이처럼 음성 데이터에 대한 정확하고 세밀한 라벨링은 AI 모델의 학습 효율성을 높이고, 모델의 성능을 최적화하며, 특정 응용 프로그램에 대한 맞춤형 솔루션을 개발하는 데 핵심적인 역할을 한다.

NIA 2-18 과제에서도 수집된 데이터에 대한 발화 의도, 감정, 문장 유형 등 다양한 라벨링 작업이 실시되었다. 라벨링 작업자들은 국어국문학 또는 한국어 교육 전공자들이 중심이 되어 구성되었으며, 그들을 대상으로 라벨링 기준, 라벨링 도구 사용법 등의 내용으로 대면 및 비대면 교육이 이루어졌다. 전사와 마찬가지로 라벨링 역시 해당 과제를 위해 개발된 특정 작업 도구들을 사용해 작업이 이루어졌으며, 인터-레이터 신뢰도(inter-rater reliability)를 높이기 위해 이중 독립 검토(dual independent review) 후 붙일치 데이터에 대해 제 3자 검증을 거쳐, 최소한 2인 동의를 거친 라벨링이 되도록 작업 결과를 통제하였다.

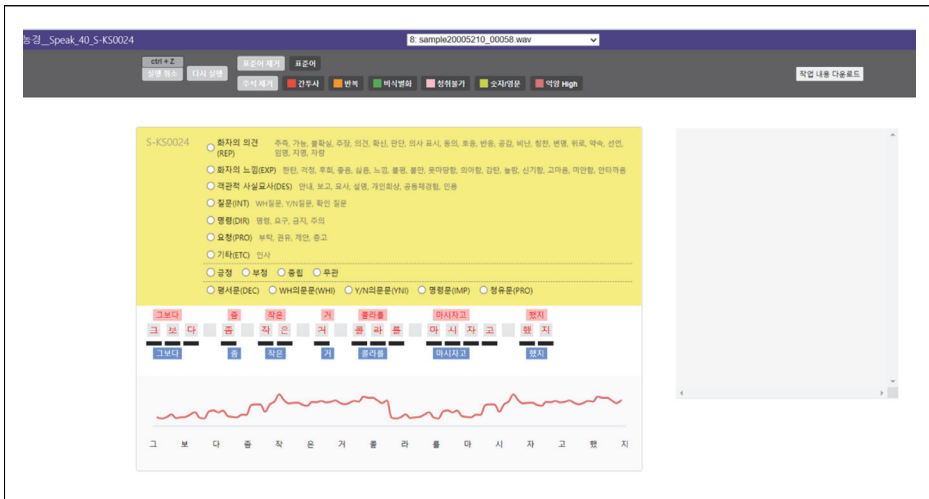


그림 4. 라벨링 도구 화면 예시

3.2.1. 발화 의도

음성 데이터에 발화 의도를 명확하게 라벨링함으로써, AI 시스템의 이해력 향상, 상호 작용 개선, 맥락 이해, 정확도 및 효율성 향상에 도움을 줄 수 있다. NIA 2-18 과제에서 실시된 데이터 라벨링 중의 하나인 발화 의도는 Jacobson(1960)의 전통적인 언어 6기능(six functions of language)과 Searle(1969)의 화행 분류에 기초하여, '표 7과 같이 재구성되었다. 라벨링 기준을 바탕으로 '기타(ETC)'를 포함, 총 7개의 발화 의도 라벨링 값이 설정되었으며, 각각의 라벨링 값은 모두 합쳐 56개의 세부 발화 의도를 포함하였다.

표 7. 발화 의도 라벨링 값 및 기준

라벨링 기준	라벨링 값	세부 발화 의도
화자	화자의 생각, 판단, 행위 약속 (REPresentative)	추측, 가능, 불확실, 주장, 의견, 확신, 판단, 의사표시, 동의, 호응, 반응, 공감, 비난, 칭찬, 변명, 위로, 약속, 선언, 임명, 지명, 주관적경험, 자랑
	화자의 심리 상태 표현 ('~를 느낀다'에 해당) (EXPressive)	한탄, 걱정, 후회, 좋음, 싫음, 느낌, 불평, 불만, 못마땅함, 의아함, 감탄, 놀람, 신기함, 고마움, 미안함, 안타까움
중립	객관적 사실 정보 제공 (DEScriptive)	안내, 보고, 묘사, 설명, 인용, 객관적경험
	청자의 예상 답 기준 (INterrogative)	WH질문, Y/N질문, 확인질문
청자	청자 의무 있음 / 강제성(의무) 있음 (DIRective)	명령, 요구, 금지, 주의
	강제성(의무) 없음 / 상대의 행동 기대 (PROpositive)	부탁, 권유, 제안
기타	기타 (ETC)	인사, 맞장구

1차 라벨링 작업자들은 작업자 교육 및 '표 7'의 내용을 숙지하고 라벨링 도구 상에서 해당 문장이 가지는 가장 우세한 발화 의도를 그 문장의 라벨링 값으로 판정하였는데, 도구 상에서 세부 발화 의도를 선택하면 라벨링 값도 자동으로 라벨링되는 방식이었다. 문장의 발화 의도가 중의성을 띠는 경우에도, 인공지능 학습의 효율을 높이고, 빅데이터의 의도 분산 효과를 고려하여, 제 1의미(primary interpretation) 기반 라벨링을 진행하였고, 동일 해석에 대해 다른 검수자의 동의 여부가 최종 라벨링 값을 결정하였다. 구체적으로, 2차 작업자들

은 1차 작업자들의 발화 의도 판정에 동의하면 판정을 유지하고, 동의하지 못할 때는 본인의 판정을 메모하여 3차 작업자에게 전달하였다. 마지막으로 박사급 관리자들로 구성된 3차 작업자들이 최종적으로 발화 의도 판정을 검수 및 확정하였다.

3.2.2. 감성

감성 라벨링을 통해 개발된 AI 모델은 텍스트나 음성에서 사람의 감정을 인식하고 분석할 수 있다. 이는 고객 의견 분석이나 트렌드 분석 및 예측과 같은 산업 분야뿐만 아니라, 인간의 정서 상태를 모니터링하고, 정신 건강 문제를 조기에 발견하는 데에도 사용될 수 있다. NIA 2-18 과제에서는 '표 8'과 같이 총 4가지 척도로 감성 라벨링 값을 설정하였다.

표 8. 감성 라벨링 값 및 기준

라벨링 값	긍정	중립	부정	무관
라벨링 기준	명확한 긍정일 때	중립 감성일 때 또는 긍정과 부정이 양립할 때	명확한 부정일 때	기본값

작업자들은 감성 라벨링 작업 시, '긍정' 값은 발화가 기쁨, 만족, 흥미 등과 같은 긍정적인 감정, 태도, 또는 평가를 나타낼 때 선택하였고, '부정' 값은 발화가 슬픔, 불만족, 분노 등과 같은 부정적인 감정, 태도, 또는 평가를 나타낼 때 선택하였다. 발화에 감정은 존재하지만, 긍정적이기도 부정적이기도 않은 상태를 나타낼 때 또는 긍정과 부정이 혼합되어 어느 한 쪽으로 기울기 어려운 경우에는 '중립'으로 판정하였는데, 예를 들어, "그것은 괜찮은 아이디어 이긴 한데, 실현하기 어려울 것 같아."와 같은 발화는 중립적으로 해석될 수 있다. 마지막으로 발화가 감정이나 태도와 관련 없는 사실적인 정보, 질문, 또는 일반적인 서술 등 주어진 감성 분류 척도와 무관한 정보를 담고 있을 때에는 '무관'으로 판정하였다. 감성 라벨링 역시 발화 의도 라벨링과 마찬가지로의 기준으로 3차에 걸쳐 진행되었는데, 특히 감성 라벨링은 작업자의 주관적인 요소가 강하게 작용할 수 있기 때문에 여러 평가자가 동일한 음성 데이터에 대해 라벨링을 수행하고 그 결과를 비교함으로써 라벨링의 일관성과 신뢰성을 높일 수 있다.

3.2.3. 문장 유형

통사적 형태에 기반한 문장 유형 라벨링은 언어 분석의 정밀도를 높여, 언어 모델이 음성이나 텍스트 데이터의 복잡한 구조와 의미를 더 깊이 이해할 수 있게 할 수 있다. 대화 시스템이나 챗봇에서는 사용자의 발화 의도와 감성을 이해하는 것뿐만 아니라, 발화의 문장 유형을 파악하는 것도 중요한데, 예를 들어, 명령문이나 청유문을 정확히 인식하면 시스템이 적절한 동작을 수행하거나 반응을 제공하는 데 도움이 된다. 해당 과제에서 작업자들은

‘표 9’와 같이 데이터의 통사 정보에 기반하여 5가지 문장 유형 중 하나를 라벨링 값으로 판정하였다.

표 9. 문장 유형 라벨링 값 및 기준

라벨링 값	평서문 (DECLarative)	명령문 (IMPerative)	WH의문문 (WHInterrogative)	Y/N의문문 (YNInterrogative)	칭유문 (PROpositive)
라벨링 기준	-다, -구면, -을게요, -구나, -군(요), -습디다, -는가 봐(요), -더라, -습니다, -라고(요), -잖아(요) 등	-아/어라, -으시오, -아/어 주세요, -아/어 보세요, -어 주소, -으십시오, -지 마(라), -지 말아요 등	의문사가 존재하는 형태	“네/아니요”로 대답 가능한 형태	-읍시다, -자

3.2.4. 기타 라벨링

발화 의도, 감성, 문장 유형 이외에도 작업자들은 해당되는 경우, 라벨링 도구 상에서 ‘간투사, 반복, 비식별화, 청취불가, 숫자/영어, 억양 High’의 라벨링을 실시하였다. "음...", "그게...", "저..." 등과 같이 문장 내에서 문법적 기능을 가지지 않으면서 대화의 흐름을 조절하기 위해 사용되는 표현에는 ‘간투사’ 라벨링이 이루어졌고, 비의도적으로 반복된 표현이라고 판단되는 경우에는 ‘반복’ 라벨링이 표시되었다. 개인 정보 보호를 위해 특정한 이름, 주소, 전화번호, 이메일주소와 같이 개인을 식별할 수 있는 표현이나 심한 욕설과 같은 표현에는 라벨링하여 ‘비식별화’ 처리되도록 하였다. 그리고 전사는 되어있으나, 의미 파악이 힘들고 음성으로도 확인이 되지 않을 때는 ‘청취불가’ 처리를 하였고, 2인 대화에서 주로 일어난 말걸침 구간에 대해서도 ‘청취불가’ 라벨링을 하였다. 전사 단계에서 TN처리된 숫자나 영어 표현은 라벨링 단계에서 다시 ‘숫자/영어’ 표시를 하여 ITN(Inverse Text Normalization) 처리되도록 하였다. 마지막으로, 특히 경상 방언과 같이 높은 억양으로 의도나 감정을 표현하는 경우 해당 부분에 대해 ‘억양 High’ 라벨링이 실시되었다.

4. 데이터 분석

4.1. 기초 데이터 분석

최종 구축되어 과학기술정보통신부의 AI 통합 플랫폼인 AI-Hub(www.aihub.or.kr)에 공개된 데이터는 웨이브 파일 형식의 원천 데이터와 그에 대응하는 제이슨(json) 파일 형식의 라벨

링 데이터로 구성되어 있고, 그중 본 논문에 인용될 발화 의도 및 감성 라벨링 데이터의 구성 내용은 '표 10'과 같다. 인용 대상 데이터의 총시간은 3,303여 시간의 분량으로, 이 데이터의 구축에 참여한 발화자의 수는 총 13,422명이다.

표 10. 연령, 성별 및 지역별 데이터 시간

		강원도	경상도	전라도	제주도	충청도	합계
50대	남	21.1	4.6	41.8	14	42.6	124.1
	여	130.2	29.8	76	75.2	118.7	429.9
60대	남	84.6	103.5	95.7	14.9	40	338.7
	여	282.1	255.9	216.6	42.5	172.8	969.9
70대	남	50.7	130.8	93.1	1.5	54.8	330.9
	여	103.7	383.3	196.8	5.5	155.2	844.5
80대	남	12.1	39.4	20	0	17.1	88.6
	여	21.3	76.2	44.7	0.4	29.6	172.2
90대	남	1.2	0	0.2	0	0	1.4
	여	0	0	2.7	0	0.2	2.9
합계		707	1023.5	787.6	154	631	3303.1

발화자 분포는 아래 '그림 5'와 같이 연령별로 50대 2,414명(18%), 60대 4,924명(37%), 70대 이상이 6,084명(45%)이고, 성별 비율에서는 여성이 남성보다 3배 높은 참여도를 보였다.

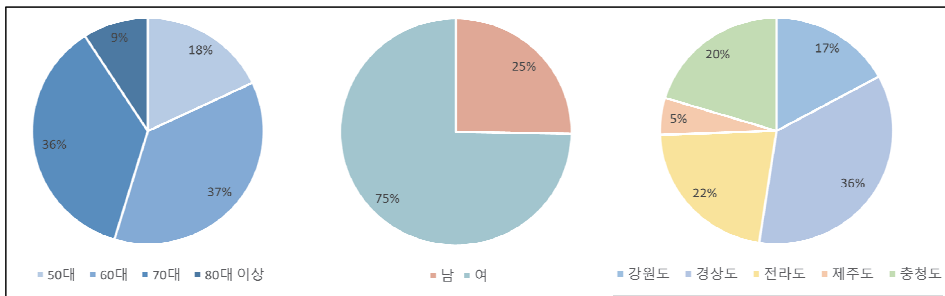


그림 5. 연령별, 성별, 지역별 발화자 분포

4.2. 집단별 발화 의도 및 감성 분포

본 절에서는 대한민국 중노년층 화자의 발화에 나타나는 전반적인 감성 및 의도 특징을 알아보기 위해, 중노년층 화자의 발화 의도 및 감성이 가장 자연스럽게 드러나도록 기획 및 구축된 비정형 발화의 2인 대화 데이터에 나타난 집단별 발화 의도 및 감성 분포 통계 결과를 보고한다. 우선 발화 의도에서는 ‘그림 6’의 결과로 알 수 있듯이 남성과 여성이 유사한 분포를 보여 주는데, 양 성별에서 모두 ‘화자의 의견(REP)’이 50% 이상의 비율을 차지하며 가장 높은 빈도를 보여 주었고, ‘객관적 사실묘사(DES)’가 남녀 모두 30% 이상을 차지하며 두 번째로 높은 비율을 차지하였다. 그다음으로는 ‘화자의 느낌(EXP)’이 10% 이상의 비율로 뒤를 이었으며, 그다음이 ‘질문(INT)’, 마지막은 ‘요청(PRO)’과 ‘명령(DIR)’으로 두 항목을 합쳐도 1% 미만의 비율을 보여주었다(‘그림 6-7’에서는 ‘PRO’ 항목으로 통합).

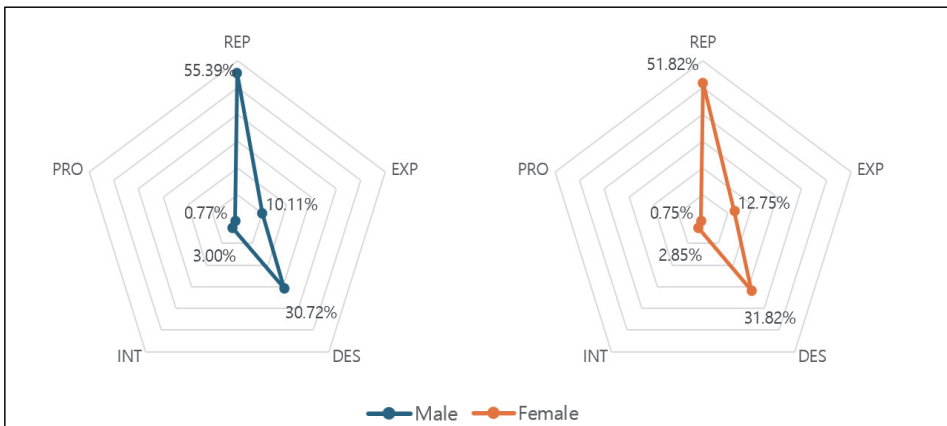


그림 6. 성별 발화 의도 분포(2인 대화)

발화 의도 분포를 ‘그림 7’과 같이 50대(Middle Aged, MA), 60대(Middle Elderly, ME), 70대 이상(Old Elderly, OE)의 연령대별로 살펴보면, 대화에서 노년층으로 갈수록 주관적 의견의 비율이 줄고, 객관적 발화가 늘어나는 것을 확인할 수 있었다. 질문이나 명령, 요청 등의 의도를 가지는 발화는 50대에서 가장 높은 비율을 보여 주었고, 노년층으로 갈수록 그 비율이 줄어들었다.

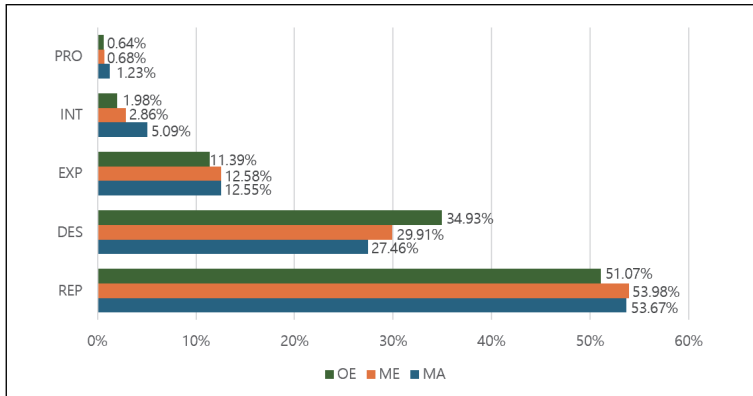


그림 7. 연령대별 발화 의도 분포 (2인 대화)

감성에서는 ‘무관(Irrelevant)’을 제외하고 ‘부정(Negative)’, ‘중립(Neutral)’, 그리고 ‘긍정(Positive)’의 라벨링 비율을 살펴보았는데, 우선 성별 비교에서는 ‘그림 8’에서와 같이 부정과 긍정의 감성 모두 여성이 남성보다 높은 비율을 보였고, 이는 중노년층 여성의 감성 표현이 남성보다 더 풍부하다는 것을 암시한다.

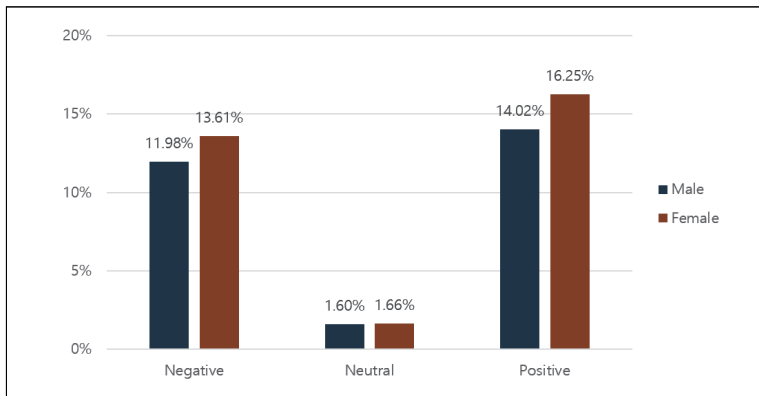


그림 8. 성별 감성 분포 (2인 대화)

마지막으로 연령대별로는 ‘그림 9’에서 확인할 수 있듯이, 연령대가 높아지면서 근소하게 부정 감성이 줄고 긍정 감성이 증가하는 경향을 보여 주었다.

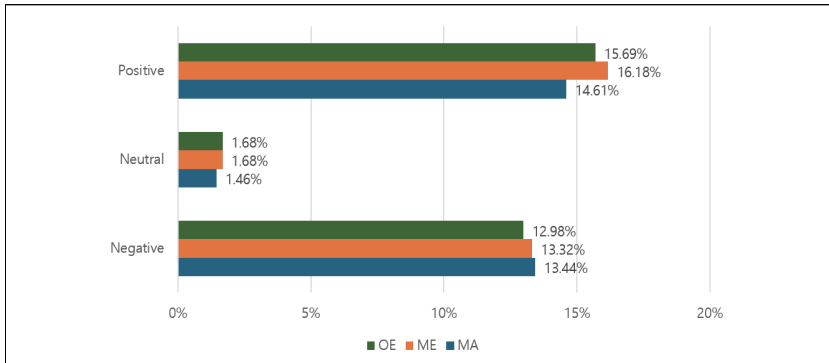


그림 9. 연령대별 감성 분포(2인 대화)

5. 결론 및 제언

본 논문에서는 2022년 진행된 AI 학습용 데이터 구축 사업 중 중노년층 방언 데이터 과제의 기획, 수집, 그리고 가공의 구축 과정을 소개하고, 구축된 데이터의 분석을 진행하였다. 50대부터 90대까지의 중노년층 방언 화자가 수집 대상이었던 해당 과제의 결과로 대한민국 5개 방언권의 중노년층 음성 데이터가 원천 데이터로 구축되었으며, 3중 텍스트 전사 및 억양, 발화 의도, 감성, 문장 유형, 그 외 다양한 라벨링 정보가 메타 정보와 함께 포함된 라벨링 데이터가 원천 데이터에 매칭되어 함께 구축되었다.

해당 과제에서 구축된 데이터는 이용 분야에 따라 다방면의 효과가 기대된다. 우선, 학문적으로는 중노년층의 방언 데이터를 기반으로 한 음운, 구문 특성, 발화 의도, 감성 유형 등에 관한 이론적 연구의 발전을 도모할 수 있고, 이는 언어학의 다양한 분야에 대한 새로운 지식을 창출하는 기반이 될 수 있다. 또한, 인공지능 학습 데이터의 구축 방향과 데이터 스키마 형식에 대한 기준을 설정함으로써 후속 연구에 대한 방향 제시가 가능하다. 사회적으로는, 노령화 사회에서 점점 중요해지는 노년층의 언어 분석과 인공지능 연구 결과물을 통해 독거노인의 응급상황 대처 및 사회 문제 해결에 기여하고, 세대 간 소통을 지원하여 노년층의 디지털 소외를 방지하는 데 기여할 수 있다. 그리고 기술적으로는 데이터를 이용한 기술 실용화가 이루어질 수 있는데, 특히 인공지능 음성 인식 프로그램 개발에 있어 노년층의 한국어 데이터를 통해 성능을 보완하고 개선할 수 있을 것으로 기대된다.

물론, 향후 추가적인 라벨링 정보를 포함한 데이터의 복합적 분석이 필요하고, 방언, 억양, 문장 유형 등 더 다양한 데이터 특성을 고려하여, 인공지능 학습 데이터의 품질을 더욱 높이는 연구도 병행되어야 할 것이다. 또한, 데이터 고도화를 통해 다양한 응용 분야에서의 활용 가능성을 탐색하는 것도 남은 과제라고 할 수 있겠다.

참고문헌

- 류성기. (2016). 경남 하동 방언의 문법 형태와 세대 및 의식 변화에 따른 문법 형태 사용 변화 연구. *국제언어문학*, 33, 1-37.
- 류성기. (2017). 남원 방언 문법 형태와 사용 양상 변화 연구. *국제언어문학*, 37, 57-94.
- 심우창, 진혜빈, 김세진, 김선동. (2023). ARC 문제 해결을 위한 프롬프트 엔지니어링의 가능성. *한국정보과학회 학술발표논문집*, 397-399.
- 이기갑. (2008). 농촌 지역의 이주 외국인 여성들을 위한 방언 교육. *한글*, 280, 165-202.
- 장영우, 서하린, 서영균. (2023). 주어진 주제에 대한 소셜미디어 데이터셋 생성을 위한 파이프라인 설계 및 구현. *한국정보과학회 학술발표논문집*, 438-440.
- Beese, C., Vassileiou, B., Friederici, A. D., & Meyer, L. (2019). Age differences in encoding-related alpha power reflect sentence comprehension difficulties. *Frontiers in Aging Neuroscience*, 11. <https://doi.org/10.3389/fnagi.2019.00183>
- Fukuda, M., Nishizaki, H., Iribe, Y., Nishimura, R., & Kitaoka, N. (2020). Improving speech recognition for the elderly: a new corpus of elderly Japanese speech and investigation of acoustic modeling for speech recognition. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 6578-6585, Marseille, France. European Language Resources Association.
- Harnsberger, J. D., Shrivastav, R., Brown, W. S., Rothman, H., & Hollien, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice*, 22(1), 58-69.
- Horton, W. S., Spieler, D. H., & Shriberg E. (2010). A corpus analysis of patterns of age-related change in conversational speech. *Psychology and Aging*, 25(3), 708-713.
- Jakobson, R. (1960). Closing statements: linguistics and poetics. In T. A. Sebeok (Ed.), *Style in Language* (pp. 350-377). MIT Press.
- Kemper, S., Herman, R., & Lian, C. (2003). Age differences in sentence production. *The Journals of Gerontology: Series B*, 58(5), 260-268.
- Khatun, R. & Sarkar, A. (2024). Deep-keyword net: automated English keyword extraction in documents using deep keyword network based ranking. *Multimed Tools Applications*. <https://doi.org/10.1007/s11042-024-18110-5>
- Linville, S. E. & Rens, J. (2001). Vocal tract resonance analysis of aging voice using long-term average spectra. *Journal of Voice*, 15(3), 323-330.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

옥성수

05006 서울시 광진구 능동로 209
세종대학교 대양휴머니티칼리지 초빙교수
이메일: okss@sejong.ac.kr

김수연

05006 서울시 광진구 능동로 209
세종대학교 인문대학 영어영문학전공 교수
전화: 02-3408-3634
이메일: kimsy@sejong.ac.kr

Received on March 4, 2024

Revised version received on March 24, 2024

Accepted on March 31, 2024