



Frequency of grammar items for Korean substitution of /u/ for /o/ in the word-final position

Eunkyung Yoon*

Department of Korean Language and Multi-Cultural Studies, Daegu Cyber University, Gyungbuk, Korea

Abstract

This study identified the substitution of /u/ for /o/ (e.g., py θ llo [py θ llu]) in Korean based on the speech corpus as a function of grammar items. Korean /o/ and /u/ share the vowel feature [+rounded], but are distinguished in terms of tongue height. However, researchers have reported that the merger of Korean /o/ and /u/ is in progress, making them indistinguishable. Thus, in this study, the frequency of the phonetic manifestation /u/ of the underlying form of /o/ for each grammar item was calculated in The Korean Corpus of Spontaneous Speech (Seoul Corpus 2015) which is a large corpus from a total of 40 speakers from Seoul or Gyeonggi-do. It was then confirmed that linking endings, particles, and adverbs ending with /o/ in the word-final position were substituted for /u/ approximately 50% of the stimuli, whereas, in nominal items, they were replaced at a frequency of less than 5%. The high rates of substitution were the special particle “-do[du]” (59.6%) and the linking ending “-go[gu]” (43.5%) among high-frequency items. Observing Korean pronunciation in real life provides deep insight into its theoretical implications in terms of speech recognition.

Keywords: Korean vowels, /o/-rising, substitution, frequency of grammatical items, Seoul Corpus 2015, vowel shift

1. 서론

현재 안드로이드 OS 기반의 text-to-speech(TTS)에서는 [-구어]로 발음하면 자동으로 ‘-고요’의 표준발음형태로 변환된다. 이처럼 컴퓨터의 텍스트에 문장을 입력하면 사람의 음성으로 변화해 주는 음성합성 기술인 TTS는 사물 인터넷 시대를 맞아 더욱 활발하게 사용되고 있다. 최근에는 각종 인터넷 뉴스에서 기사를 읽어주는 서비스부터 SNS 또는 네비게이션에서도 빈번하게 활용되고 있다. 언중들의 발화 양식을 TTS 기술에 접목하여 보다 자연스러운 음성을 산출 및 처리하는 알고리즘을 개발

하기 위해서는 표준발음 이외에 현실발음에 대한 조사가 요구된다. 한국어 표준발음은 문서로 규정화되어 있지만, 일상 언어 생활에서 개인 변이형은 매우 다양하게 나타난다. 이에 본 논문에서는 음성인식 시스템의 성능을 개선시키기 위한 기초 연구의 일환으로 구어 말뭉치 분석을 통해 현재 진행 중인 한국어 /ㅊ/의 고모음화 현상을 문법 형태별로 출현빈도와 비율을 집중적으로 살펴보는 데 연구 목적이 있다.

한국어의 음운론적 10모음 체계에서 /ㅊ/와 /ㅊ/는 이중모음으로 간주하는 경향이 있고, /ㅊ/와 /ㅊ/가 합류되어 최근에는 음성학적으로 /i, E, a, ʌ, o, u, ɨ/ 7모음 체계를 이룬다는 것이 통설

* heavynub@dcu.ac.kr, Corresponding author

Received 31 January 2020; Revised 14 March 2020; Accepted 14 March 2020

© Copyright 2020 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

로 자리 잡았다. 이 중 한국어 /ㄴ/와 /ㄷ/는 ‘우리:오리’와 같이 원순성 자질을 공유하지만 혀높이의 차이로 변별된다. 그러나 최근에는 /ㄴ/와 /ㄷ/의 합류에 대한 논의도 활발히 진행되고 있다.¹

선행연구(Chae, 1999; Han & Kang, 2013; Ha & Oh, 2017; Kang & Han, 2013; Kang & Kong, 2016; Lee et al., 2016; Lee et al., 2017; Seong, 2004; Yang, 2018; Yoon & Kim, 2015)에 따르면, 주로 서울·경기 방언에서 혀높이의 자질이 변별력을 상실해 비어두음절 위치에서 /ㄴ/의 고모음화가 진행되고 있다. 이러한 현상은 일상생활에서 상당히 빈번하게 드러난다. 예를 들어, 명사 ‘삼촌’을 [삼춘], ‘부사격조사 -로’를 [루], 부사 ‘별로’를 [별루], 연결어미 ‘-고’를 [구]로, 보조사가 결합된 형태인 ‘-고요’를 [구요] 또는 [구어] 등으로 사용되는 것을 자주 목격하게 된다.

Chae(1999)에 의하면, 이러한 비어두 모음 /ㄴ/의 고모음화는 음운론적 환경에 조건 짓기 어렵고, 사회언어학적인 변이나 단어의 출현빈도에 따라 다르게 나타난다고 하였다. 이 연구를 기반으로 실험음성학적으로 /ㄴ/와 /ㄷ/의 모음공간 근접성을 연구한 Han & Kang(2013)은 /ㄴ/와 /ㄷ/의 포먼트를 측정하여 유클리디언 거리(Euclidean’s distance)의 차이를 살펴보았다. 20대, 30대, 40·50대의 3세대로 구분한 피험자 36명의 산출상 /ㄴ/와 /ㄷ/의 거리는 여성이 남성보다 더 가깝고, 연령대가 낮을수록 모음 공간이 더 가깝게 위치한 것이 확인되었다.

위의 Han & Kang(2013)의 연구 결과는 Chae(1999)의 연구를 발전시켜 구어체의 음성 실현의 양상과 사회언어학적인 요인을 성별·연령별로 살펴봄으로써 언어 변화를 포착했다는 데 학문적 의의가 있다. 그러나 위의 연구는 구어체 발화를 분석하기는 했으나, ‘천천히 _____ 하세요’의 문장 틀을 사용하여 낭독체 발화를 연구 대상으로 삼았다. 표준발음형이 아닌 음성 변이형을 추적 관찰하기 위해서는 낭독체가 아닌 자유발화 형태의 음성 자료가 필요하다.² 시료들이 통제된 조건에서 진행된 낭독체 발화라면 피험자에게 제공되는 녹음용 자료 목록의 활자에 의해 자극이 강화되어 자료를 표준발음인 /ㄴ/로 구사한 피험자들도 있었을 것으로 예측된다. 자유발화를 통해 보다 일상생활에 가까운 언어 변이형을 관찰하는 것이 더욱 타당할 것으로 사료된다. 이러한 측면에서 본고는 한국어 자연발화 음성 코퍼스(The Korean Corpus of Spontaneous Speech, 이하 Seoul Corpus)를³ 기반으로 어말 음절에서 한국어 모음 /ㄴ/가 /ㄷ/로 대체되는 /ㄴ/의 고모음화 현상에 대해 탐구하는 것을 연구 목적으로 삼는다.

2. 연구 방법

2.1. 연구 대상

Seoul Corpus는 Yun et al.(2015)의 연구에서 구축된 것으로 약 220,000어절로 구성된 대규모 음성 코퍼스이다. 10대, 20대, 30대, 40대 세대로 남녀 각 5명씩 선발하여 총 40명의 피험자들의 음성 자료가 파일별로 60분 분량의 240개 파일이 녹음되어 있다. 10대 자료는 평균 8,841어절로 남성이 4,571어절, 여성이 4,270어절 규모이고, 20대 자료는 평균 10,749어절로 남성이 6,021어절이고, 여성이 4,728어절, 30대 자료의 평균은 12,675어절이고 남성은 7,246어절, 여성은 5,429어절이고, 40대 자료는 평균 12,292어절이고 남성은 6,277어절, 여성은 6,015어절의 발화량을 보였다. 주로 학생, 직장인, 주부 등 피험자들의 신변 이야기부터 직업, 문화, 정치까지 다양한 주제를 인터뷰 진행자가 질문을 하고 피험자가 대답을 하는 형식으로 진행된다. 녹음은 TASCAM HD-P2 레코더와 AKG C420 마이크로폰을 사용하였고, 44 kHz의 표본추출율과 16-bit로 양자화하여 녹음되어 있다.

2.2. 자료 추출

Seoul Corpus에서 제공하는 모든 음성 파일에 딸린 TextGrid는 그림 1에서 보는 바와 같이 크게 기저형 발음과 표면형 발음을 어절 또는 발화 단위로 한글 또는 로마자로 전사한 자료이다.⁴

Seoul Corpus에 구축된 240개의 음성 파일(wav)과 텍스트 파일(Textgrid)을 분석하기 위해 음성 말뭉치 분석 프로그램인 Phonometrica(Eychenne & Courdès-Murphy, 2019)를 사용하였다. 우선 tier 6에서 기저형이 '오'로 끝나는 인터벌을 모두 추출한 후, tier 3에서 표면형이 [oo]-‘오’ 또는 [uu]-‘우’로 나타난 자료를 모두 추출하고 이 두 tier의 한글 전사형을 확인하기 위하여 tier 2와 5도 함께 추출하여 비교하였다.

1 /ㄴ, ㄷ/와 관련된 최근 한국어 모음 추이 상세한 논의는 Kang & Kong(2016) 참고.

2 Kim & Yoon(2015a, 2015b)의 연구에서도 낭독체 발화와 자유발화에서의 포먼트 분석 결과의 차이에 대해 언급하였다.

3 이 말뭉치는 낭독체가 아니라 성별, 연령대별 자유발화를 녹음한 대규모 말뭉치라는 점에서 굉장히 가치가 있고, 귀중한 자료이다. 자세한 내용은 Yun et al.(2015) 참고. 그러나 위의 자유발화 역시 실험실 환경에서 인터뷰 진행자가 피험자를 인터뷰한 것으로서 일상생활에서 사용되는 진정성이 있는 자유발화와는 어느 정도 차이가 있을 것으로 예상된다.

4 Textgrid의 층렬 1은 ‘phoneme’ 단위, 층렬 2는 ‘pWord.prono’로 어절 단위로 끊어 발음 나는 대로 한글로 전사 층렬 3은 ‘utt.prono’로 표면형을 어절 단위를 내부 규약대로 로마자로 표기, 층렬 4는 ‘utt.prono’로 표면형을 발화 단위로 끊어 한글 전사, 층렬 5는 ‘pWord.ortho’로 기저형을 어절 단위로 한글로 철자 전사, 층렬 6은 ‘pWord.ortho’로 기저형을 어절 단위로 로마자로 표기, 층렬 7은 ‘utt.ortho’로 기저형을 발화 단위로 한글 전사함.

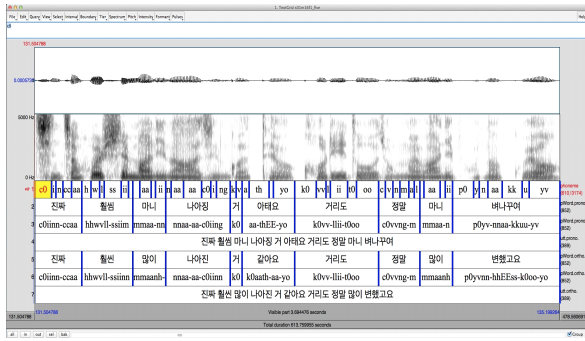


그림 1. Seoul Corpus의 음성파일과 텍스트 전사 층위 샘플 화면
Figure 1. Screenshot of wav and Textgrid files in the Seoul Corpus

2.3. 자료 분석

위의 과정을 통해 추출한 연구 자료는 총 24,805개이다. 이 중 기저형 /ㅏ/의 표면형으로 /ㅑ/ 이외에 다른 모음들도 추출되어 제거하고, 5 본 연구의 초점인 /ㅑ/ 대체형만을 선정하였다. 그리고 화자들이 말을 반복하며 교정했을 경우 이전 형태의 음성형은 제거하였다. 그 결과 본 실험에서 선정한 자료는 2,103개를 제외한 22,702개이다. 그중에서 음성표면형이 [o]인 토큰은 12,310개, [u]인 토큰은 10,392개로 나타났다.

최종적으로 선택된 자료들은 형태소 분석기 McCab(메카브)를 이용하여 품사를 1차적으로 분류하고, 품사 코딩이 잘못된 것들이 발견되어 수작업으로 형태소를 재분류하였다. 예컨대, ‘하고’의 경우 3 가지의 문법 형태로 분류가 가능하여 (1) 부사격조사, (2) 접속조사, (3) 접미사 ‘-하다’에 연결어미 ‘-고’가 결합된 형식이 있다. 해당 문법 형태가 텍스트에서는 어떤 형태로 나타나는지 확인하기 어려워 발화 단위의 층렬(tier)을 검색해서 하나씩 확인하는 과정을 거쳐 분석하였다.

어말에서 /ㅏ/로 끝나는 문법 항목은 연결어미, 조사, 부사, 체언, 감탄사로 5개의 범주로 나타났으나, 감탄사는 ‘어보’의 예한 개만 출현하였고 [o]의 형태로 실현되었기 때문에 더 이상 논의를 진행하지 않았다. 나머지 문법 항목은 대분류로 연결어미, 조사, 부사, 체언으로 구분하였다. 그리고 이들을 중분류하여 조사는 보조사, 부사격조사, 접속조사로 재분류하였다. 그리고 부사는 다시 문장부사와 성분부사로, 체언은 보통명사, 고유명사, 수사, 외래어, 의존명사로 재분류하였다.

3. 분석 결과

3.1. 대분류: 연결어미, 조사, 부사, 체언

기저형의 /o/가 표면형에서 [o] 또는 [u]로 실현되는 22,702개의 음성을 문법 항목별로 출현빈도와 출현비율을 조사한 결과가 표 1에 제시되어 있다. Seoul Corpus에서 [o]로 끝나는 문법

항목들 중 가장 높은 빈도로 출현한 항목은 연결어미(10,790개)>조사(8,684개)>부사(2,129개)>체언(1,099개) 순으로 나타났다. 그리고 아래의 표에서 보는 바와 같이 [o]의 [u] 대체율을 종합적으로 분석해 볼 때 체언은 다른 문법 항목들과는 달리 굉장히 낮은 비율인 4.6%의 대체율을 보였고, 그 외의 문법 항목은 조사(52.1%)>부사(51.3%)>연결어미(43.8%) 순으로 높게 나타났고, 평균 45.8%의 대체율을 보였다.

표 1. 문법 항목별 출현빈도 및 비율 (%)

Table 1. A summary of substitution frequency of linking endings, particles, adverbs, and nominal items

대분류	중분류	출현 빈도	출현 비율	[u] 대체율
연결어미 (n=10,790)	연결어미	10,790	43.8	43.8
조사 (n=8,684)	보조사	5,007	59.4	52.1
	부사격조사	3,431	42.4	
	접속조사	246	39.8	
부사 (n=2,129)	문장부사	1,079	67.5	51.3
	성분부사	1,050	34.8	
	접미사	-	-	
체언 (n=1,099)	보통명사	726	5.9	4.6
	고유명사	126	3.2	
	수사	193	0.5	
	(명/의.명) 외래어	54	5.6	
총합		22,702		38.0

아래의 그림 2에서 문법 항목별 출현비율을 보다 간략하게 밀도 플롯으로 제공하였다. 문법 항목 중 ‘체언’에서만 대체율이 매우 크게 차이가 나는 것을 볼 수 있다.

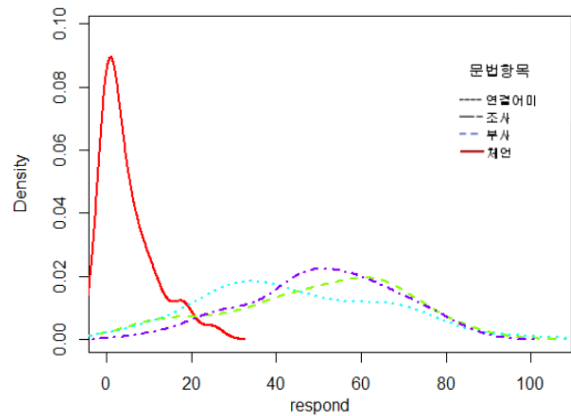


그림 2. 문법 항목별 [o]→[u] 대체율(%)의 밀도 플롯
Figure 2. Density plot of substitution rate for each grammar item

문법 항목별 [u]의 대체율만을 기술통계량으로 정리한 것이 표 2에 있다. 연결어미, 조사, 부사의 최솟값은 최소 0.0%에서

5 /ㅑ/ 이외에도 /ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅟ, ㅠ, ㅡ, ㅢ/와 같은 모음이 추출되었다.

6 논문 내에서 표기의 일관성을 갖추기 위해 [ㅓ]와 [ㅑ]로 표기하는 것이 옳지만, 분석 결과를 표로 삽입하여 두 모음만을 일렬로 표기했을 때 상당히 유사해 보여 시각적으로 혼동이 된다는 점이 확인되었다. 이에 3장의 분석 결과에서는 부득이 두 모음을 IPA로 다르게 표기하였다.

최대 11.0%의 대체율을 보였다.

표 2. 화자별 각 문법 항목의 기술통계량
Table 2. Descriptive statistics for each grammar item

문법 항목	n	최솟값	최댓값	평균	표준편차
연결어미	40	6.7	95.8	44.2	20.7
조사	40	11.0	81.9	50.9	16.9
부사	40	5.3	82.9	49.4	20.2
체언	40	0.0	25.0	4.8	6.2

이러한 문법 항목별 음성 변이형의 출현이 개인마다 차이가 있는지를 살펴보기 위해 아래의 그림 3과 같이 [u] 대체율에 대한 밀도 분포를 확인해 보았다. 그 결과 위의 표에 나타난 연결어미, 조사, 부사의 최솟값(6.7%, 11.0%, 5.3%)은 남성 화자 1번의 것으로 나머지 화자들의 대체율은 대체로 유사했다.

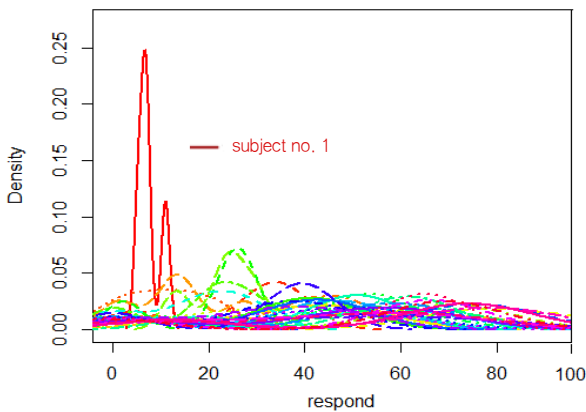


그림 3. 개인별 [o]→[u] 대체율(%)의 밀도 플롯
Figure 3. Density plot of substitution rate for each subject

위에서 언급한 문법 항목 간 [u] 대체율에 차이가 있는지를 검증하기 위해 R 프로그램(ver.1.2.5033)의 lme4 패키지를 사용하여 선형혼합효과 모델(Linear mixed-effects model) 분석을 실시하였다. 고정 효과로는 ‘문법 항목’, 반응변수로는 ‘응답([u] 대체율)’, 랜덤 효과로는 ‘화자’를 입력하여 변수를 통제한 후 다른 항목과 가장 큰 차이를 보이는 ‘체언’을 참조(reference)로 코딩하여 분석하였다. 그 결과는 아래의 표 3과 같다. 체언은 연결어미, 부사, 조사의 응답에서 통계적으로 유의미한 차이가 있었으나($p < .001$), Tukey 방식으로 사후 검정을 실시한 결과 연결어미, 부사, 조사 간에는 유의미한 차이가 나타나지 않았다($p > .05$).

표 3. 문법 항목별 대체율에 대한 선형혼합효과(Lmer) 모델 통계량
Table 3. Results of Linear mixed-effects model for grammar items

Fixed effects:	Estimate	Std Error	df	t-value	Pr (> t)
(Intercept)	4.88	4.98	2.48	0.98	0.413
연결어미	39.383	2.599	120	15.15	<.000***
조사	46.115	2.599	120	17.74	<.000***
부사	44.548	2.599	120	17.14	<.000***

*** $p < .05$, 코딩: 0=체언, 1=연결어미, 2=조사, 3=부사.

다음 절에서는 출현빈도가 높게 나타난 순서대로 연결어미, 조사, 부사, 체언을 좀 더 중분류로 세분화하여 구체적인 용례와 함께 대체 양상을 살펴보기로 한다.

3.2. 중분류: 형태소별 연결어미

연결어미는 형태소 간 출현빈도에 편차가 있어 20개를 기준으로 형태소를 구분하여 출현빈도, 출현비율, 대체율을 다음의 표 4와 같이 제시하였고 이해를 돕기 위해 용례를 추가하였다. 연결어미는 총 10,790개 중에서 [o]가 6,067개(56.2%), [u] 4,723개(43.8%)가 출현하였다. 그림 4에서도 알 수 있는 바와 같이 [u]의 대체율은 50% 미만이지만, [u]의 출현빈도가 4천 개 이상으로 매우 높으므로 청자들은 일상생활에서 연결어미의 음성 변이형을 고빈도로 지각하게 될 것이다.

보다 구체적으로 연결어미의 형태소별 차이를 살펴보면 다음과 같다. Seoul Corpus에는 /o/로 끝나는 연결어미가 11개가 출현하였다. 가장 높은 출현빈도를 보인 것은 {-고}였는데, 9,948개로 놀랍게도 전체의 92.2%를 차지했다. 이 중 [o]가 5,620개(56.5%), [u]가 4,328개(43.5%)로 나타났다. 출현빈도 이외에 [o]의 [u] 대체율이 가장 높은 연결어미는 {-어/아도}로 [o]가 192개(47.4%), [u]가 213개(52.6%)로 나타났다. 표준형보다 변이형이 더 높은 비율로 사용되고 있었다. 그리고 20개 미만으로 나타난 연결어미는 표의 아랫부분에 따로 제시하였는데, {-랍시고, -느라고, -답시고, -(으)르지라도}가 있었다. 이 연결어미들의 [u] 대체율은 50% 이상으로 높게 나타났다.

표 4. 연결어미에서 [o]와 [u]의 출현빈도 및 비율(%)

Table 4. Frequency and substitution rates of [o] and [u] in linking endings

형태소	n	음성형	출현 빈도	출현 비율	[u] 대체율
-어/아도	405	[o]	192	47.4	50% 이상
		[u]	213	52.6	
-(으)려고	190	[o]	106	55.8	50% 미만 ~ 30% 이상
		[u]	84	44.2	
-고	9,948	[o]	5,620	56.5	50% 미만 ~ 30% 이상
		[u]	4,328	43.5	
-(이)라도	62	[o]	36	58.1	50% 미만 ~ 30% 이상
		[u]	26	41.9	
-다고	91	[o]	57	62.6	50% 미만 ~ 30% 이상
		[u]	34	37.4	
-더라도	38	[o]	24	63.2	50% 미만 ~ 30% 이상
		[u]	14	36.8	
-ㄴ/는다고	34	[o]	22	64.7	50% 미만 ~ 30% 이상
		[u]	12	35.3	
<20개 미만 자료>					
-랍시고	1	[o]	0	0	50% 이상
		[u]	1	100	
-느라고	15	[o]	7	46.7	50% 이상
		[u]	8	53.3	
-답시고	4	[o]	2	50.0	50% 이상
		[u]	2	50.0	
-(으)르지라도	2	[o]	1	50.0	50% 이상
		[u]	1	50.0	
종합	10,790	[o]	6,067	56.2	50% 이상
		[u]	4,723	43.8	

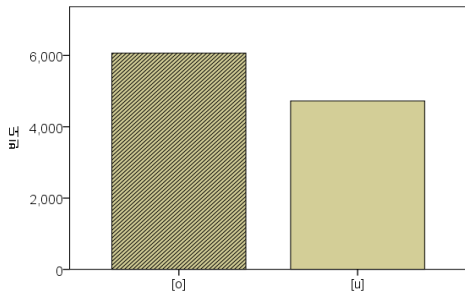


그림 4. 연결어미에서의 [o]와 [u] 출현빈도
Figure 4. Frequency of [o] and [u] in linking ending

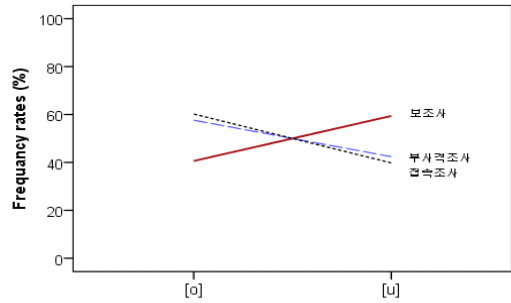


그림 6. 조사에서의 [o]와 [u] 출현비율(%)
Figure 6. Frequency rates(%) of [o] and [u] in particles

3.3. 중분류: 조사 (보조사, 부사격조사, 접속조사)

조사를 보조사, 부사격조사, 접속조사로 세분화하여 [o]와 [u]의 출현빈도, 출현비율, 대체율을 제시한 것이 아래의 표 5에 제시되어 있다. 조사는 총 8,684개 중에서 [o]가 4,159개(47.9%), [u] 4,525개(52.1%) 출현하였다. 표준발음보다는 [u]로 대체된 변이형태가 더 많이 출현한 것을 알 수 있다. 그리고 조사별 출현빈도를 나타내는 그림 5에서 알 수 있는 바와 같이 출현빈도는 보조사(5,007개)>부사격조사(3,431개)>접속조사(246개) 순으로 나타났고, 조사별 [u]의 대체율은 그림 6에서 보이는 바와 같이 특히 보조사의 대체율이 59.4%로 가장 높게 나타났고, 그 다음으로 부사격조사(42.4%), 접속조사(39.8%) 순으로 나타났다.

표 5. 조사에서 [o]와 [u]의 출현빈도 및 비율(%)
Table 5. Frequency and substitution rate of [o] and [u] in particles

대분류	중분류	음성형	출현 빈도	출현 비율	[u] 대체율
조사	보조사 (n=5,007)	[o]	2,035	40.6	50% 이상
		[u]	2,972	59.4	
	부사격조사 (n=3,431)	[o]	1,976	57.6	50% 미만
		[u]	1,455	42.4	
	접속조사 (n=246)	[o]	148	60.2	30% 이상
		[u]	98	39.8	
	종합 (n=8,684)	[o]	4,159	47.9	
		[u]	4,525	52.1	

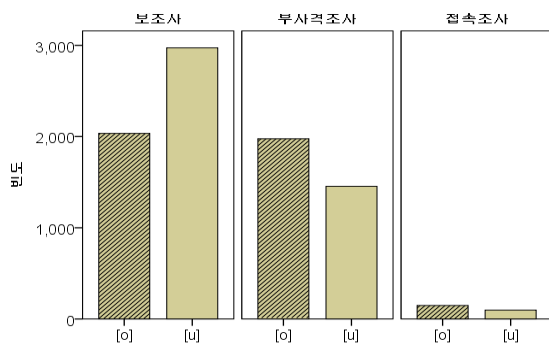


그림 5. 조사에서의 [o]와 [u] 출현빈도
Figure 5. Frequency of [o] and [u] in particles

3.3.1. 보조사

아래의 표 6과 같이 Seoul Corpus에는 조사 중에서 /o/로 끝나는 보조사는 5,007개이고, 형태소로는 {-도, -대로} 2개가 출현하였는데, {-도}는 4,956개로 전체 문법 항목 중에서 가장 높은 출현빈도(전체의 99.0%)를 보였다. {-도}는 [o]가 2,002개(40.4%), [u]가 2,954개(59.6%)로 나타났다. {-대로}는 35.3%의 [u] 대체율을 보였다.

표 6. 보조사에서 [o]와 [u]의 출현빈도 및 비율(%)
Table 6. Frequency and substitution rates of [o] and [u] in special particles

형태소	n	음성형	출현 빈도	출현 비율	[u] 대체율
-도	4,956	[o]	2,002	40.4	50% 이상
		[u]	2,954	59.6	
-대로	51	[o]	33	64.7	50% 미만
		[u]	18	35.3	
종합	5,007	[o]	2,035	40.6	
		[u]	2,972	59.4	

3.3.2. 부사격조사

부사격조사는 {-보고, -에로, -(으)로, -하고} 4개가 출현하였는데, 표 7과 같이 가장 높은 출현빈도를 보인 것은 {-(으)로}로 1,935개(전체의 56.4%)이고, 이 중 [o]가 1,096개(56.6%), [u]가 839개(43.4%)로 나타났다. 그 다음은 {-하고}는 [o]가 45개(67.2%), [u]가 22개(32.8%)로 출현하였다. 나머지 {-보고, -에로}는 단지 4개 미만으로 출현하였고, 이 중 50%는 [u]로 대체되었다.

부사격조사가 포함된 인용표현으로는 {-더라고, -자고, -ㄴ/는다고, -다고, -냐고, -(이)라고, -느냐고}의 7개의 형태가 나타났는데, 이 중 {-더라고, -자고}는 50% 이상의 [u] 대체율을 보였고, {-ㄴ/는다고, -다고, -냐고, -(이)라고}는 30~50% 정도의 [u] 대체율을, {-느냐고}는 4개의 출현빈도를 보이고, 25%의 [u] 대체율을 나타냈다.

표 7. 부사격조사에서 [o]와 [u]의 출현빈도 및 비율 (%)
Table 7. Frequency and substitution rates of [o] and [u] in case particles

형태소	n	음성형	출현 빈도	출현 비율	[u] 대체율
-보고	4	[o]	2	50.0	50% 이상
		[u]	2	50.0	
-에로	2	[o]	1	50.0	50% 미만
		[u]	1	50.0	
-(으)로	1,935	[o]	1,096	56.6	50% 미만
		[u]	839	43.4	
-하고	67	[o]	45	67.2	50% 미만
		[u]	22	32.8	
<인용표현>					
-더라고	49	[o]	7	14.3	50% 이상
		[u]	42	85.7	
-자고	24	[o]	11	45.8	50% 미만
		[u]	13	54.2	
-니/는다고	219	[o]	125	57.1	50% 미만
		[u]	94	42.9	
-다고	440	[o]	260	59.1	50% 미만
		[u]	180	40.9	
-냐고	69	[o]	40	58.0	50% 미만
		[u]	29	42.0	
-(이)라고	617	[o]	386	62.6	30% 미만
		[u]	231	37.4	
-느냐고	4	[o]	3	75.0	30% 미만
		[u]	1	25.0	
종합	3,431	[o]	1,976	57.6	30% 미만
		[u]	1,455	42.4	

3.3.3. 접속조사

아래의 표 8과 같이 접속조사는 {-하고}의 형태만 246개가 출현하였다. 이 중 [o]가 148개(60.2%), [u]가 98개(39.8%)로 나타났다. 앞서 본 부사격조사 {-하고}보다는 약간 높은 대체율을 보였으나, 큰 차이는 없었다(32.8% vs. 39.8%).

표 8. 접속조사에서 [o]와 [u]의 출현빈도 및 비율 (%)
Table 8. Frequency and substitution rates of [o] and [u] in conjunctive particles

형태소	n	음성형	출현 빈도	출현 비율	[u] 대체율
-하고	246	[o]	148	60.2	40% 미만
		[u]	98	39.8	

3.4. 중분류: 부사(문장부사, 성분부사)

부사는 표 9와 같이 총 2,129개 중에서 [o]가 1,036개(48.7%), [u]가 1,093개(51.3%) 출현하여, 표준발음형보다는 [u] 대체형이 더 많이 출현했다. 그리고 부사는 문장부사와 성분부사로 구분했다. 문장부사는 총 1,079개 출현했고, [o]가 351개(32.5%), [u]가 728개(67.5%)로 나타났다. 성분부사는 [o]가 685개(65.2%), [u]가 365개(34.8%)로 나타났다.

표 9. 부사에서 [o]와 [u]의 출현빈도 및 비율 (%)
Table 9. Frequency and substitution rates of [o] and [u] in adverbs

대분류	중분류	음성형	출현 빈도	출현 비율	[u] 대체율
부사	문장부사 (n=1,079)	[o]	351	32.5	50%
		[u]	728	67.5	이상
	성분부사 (n=1,050)	[o]	685	65.2	30%
		[u]	365	34.8	이상
	종합 (n=2,129)	[o]	1,036	48.7	
		[u]	1,093	51.3	

아래의 그림 7은 위의 표 9에서 출현빈도를, 그림 8은 출현비율을 그림으로 표현한 것이다. 보다 상세한 문장부사와 성분부사의 형태소별 비교는 다음 절에서 보고하겠다.

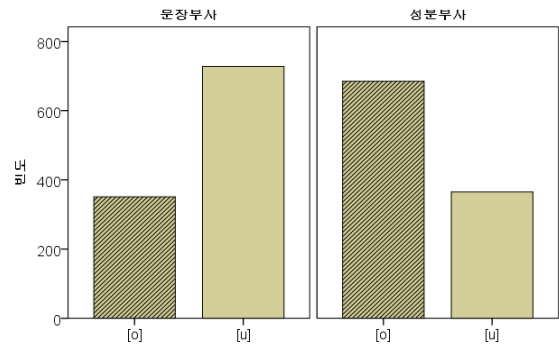


그림 7. 부사에서의 [o]와 [u] 출현빈도
Figure 7. Frequency of [o] and [u] in adverbs

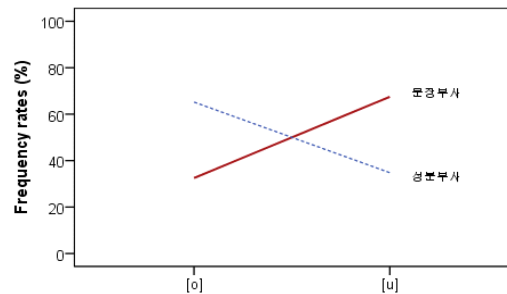


그림 8. 부사에서의 [o]와 [u] 출현비율(%)
Figure 8. Frequency rates(%) of [o] and [u] in adverbs

3.4.1. 문장부사

문장부사는 표 10과 같이 {그리고, 그래도, 아무래도} 3개가 출현하였는데, 가장 높은 출현빈도를 보인 것은 {그리고}로 575개 중 [o]가 184개(32.0%), [u]가 391개(68.0%)로 나타났다. 그 다음은 {그래도}는 290개 중 [o]가 101개(34.8%), [u]가 189개(65.2%)로 출현하였다. 마지막으로 {아무래도}는 214개 중 [o]가 66개(30.8%), [u]가 148개(69.2%)로 나타났다. 이 세 개의 문장부사들은 모두 60% 이상의 높은 대체율을 보였다.

표 10. 문장부사에서 [o]와 [u]의 출현빈도 및 비율(%)

Table 10. Frequency and substitution rates of [o] and [u] in sentential adverbs

형태소	n	음성형	출현 빈도	출현 비율	[u] 대체율
그리고	575	[o]	184	32.0	60% 이상
		[u]	391	68.0	
그래도	290	[o]	101	34.8	
		[u]	189	65.2	
아무래도	214	[o]	66	30.8	
		[u]	148	69.2	
종합	1,079	[o]	351	32.5	
		[u]	728	67.5	

3.4.2. 성분부사

성분부사로 추출된 1,050개는 아래의 표 11과 같이 33개의 형태소로 나타났는데, 이 중에서 20개 이상 추출된 형태소를 출현 빈도별로 나열하면 {별로, 서로, 바로, 주로, 따로, 제대로, 새로, 그대로, 실제로, 하도, 스스로, 정말로, 억지로}의 13개 형태가 있었다.

성분부사 중에서 가장 높은 출현빈도를 보인 {별로}는 374개 중에서 [o]가 200개(53.5%), [u]가 174개(46.5%)로 나타나 높은 대체율을 보였다. 대체율이 가장 높은 형태소는 {억지로}로 12개 중에서 11개(91.7%)가 [u]로 대체한 것으로 나타났다.

표 11. 성분부사에서 [o]와 [u]의 출현빈도 및 비율(%)

Table 11. Frequency and substitution rates of [o] and [u] in componential adverbs

형태소	n	음성형	출현 빈도	출현 비율	[u] 대체율
억지로	12	[o]	1	8.3	50% 이상
		[u]	11	91.7	
하도	21	[o]	5	23.8	
		[u]	16	76.2	
그대로	27	[o]	11	40.7	
		[u]	16	59.3	
제대로	33	[o]	17	51.5	50% 미만 ~ 30% 이상
		[u]	16	48.5	
별로	374	[o]	200	53.5	
		[u]	174	46.5	
실제로	21	[o]	13	61.9	
		[u]	8	38.1	
따로	83	[o]	58	69.9	
		[u]	25	30.1	
새로	32	[o]	23	71.9	30% 미만
		[u]	9	28.1	
바로	113	[o]	82	72.6	
		[u]	31	27.4	
서로	135	[o]	102	75.6	
		[u]	33	24.4	
스스로	20	[o]	16	80.0	
		[u]	4	20.0	
정말로	20	[o]	17	85.0	
		[u]	3	15.0	
주로	107	[o]	104	97.2	10% 미만
		[u]	3	2.8	
종합	1,050	[o]	685	65.2	
		[u]	365	34.8	

3.5. 중분류: 체언(보통·고유명사, 수사, 외래어)

체언은 아래의 표 12와 같이 총 1,099개 중에서 [o]가 1,048개(95.4%), [u]가 51개(4.6%) 출현하여, 표준발음형인 [o]가 [u] 대체형보다 압도적으로 더 많이 출현했다. 그리고 체언은 보통명사, 고유명사, 수사, 외래어로 중분류했을 때 가장 높은 비율로 대체된 것은 보통명사(5.9%)>외래어(5.6%)>고유명사(3.2%)>수사(0.5%) 순으로 나타났다. 앞서 통계 분석에서도 살펴본 바와 같이 체언은 다른 문법 항목과 다르게 [u]의 대체율이 낮게 나타났다.

표 12. 체언에서 [o]와 [u]의 출현빈도 및 비율(%)

Table 12. Frequency and substitution rates of [o] and [u] in nominal items

중분류	음성형	출현 빈도	출현 비율	형태소
보통명사 (n=726)	[o]	683	94.1	정도, 사고, 피아노, 부모, 정보, 외고 등
	[u]	43	5.9	정도(16), 사고(9), 목공소(3), 우도(3) 등
고유명사 (n=126)	[o]	122	96.8	경기도, 카카오, 송도, 제주도, 강원도 등
	[u]	4	3.2	인도(2), 구로, 시라노
수사 (n=193)	[o]	192	99.5	5, 15, 25, 35 등
	[u]	1	0.5	5
외래어 (명/의·명) (n=54)	[o]	51	94.4	프로, 킬로, 비포, 레고, 코스트코, 디아블로 등
	[u]	3	5.6	프로(2), 킬로(1)
종합 (n=1,099)	[o]	1,048	95.4	
	[u]	51	4.6	

위의 표를 보다 간략하게 출현빈도로 나타낸 것이 아래의 그림 9, 출현비율로 나타낸 것이 그림 10에 나타나 있다.

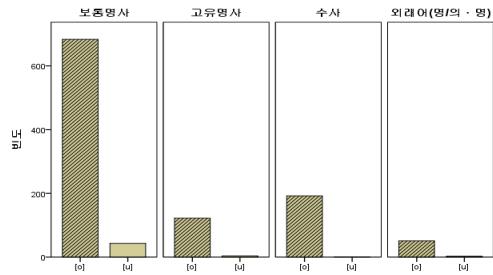


그림 9. 체언에서의 [o]와 [u]의 출현빈도
Figure 9. Frequency of [o] and [u] in nominal items

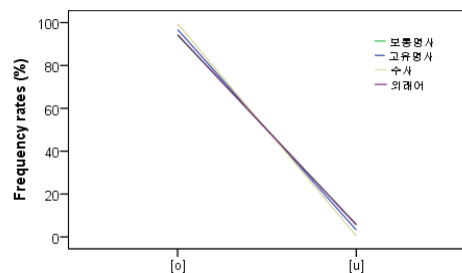


그림 10. 체언에서의 [o]와 [u]의 출현비율(%)
Figure 10. Frequency rates(%) of [o] and [u] in nominal items

4. 결론

본 논문의 연구 목적은 최근 활발하게 논의되고 있는 한국어 /ㄴ/의 고모음화 현상에 대해 한국어 자연발화 음성코퍼스 (Seoul Corpus)를 기반으로 문법 항목별 음성 변이형인 /ㄷ/의 출현빈도와 출현비율을 밝히는 데 있었다. 그 결과 다음과 같은 사실을 확인할 수 있었다.

구어 말뭉치에서 /ㄴ/로 끝나는 문법 항목은 연결어미, 조사, 부사, 체언, 감탄사 5가지의 형태로 나타났다. 감탄사는 ‘어보’ 단 한 개만 출현하였고 변이형이 나타나지 않아 더 이상 논의를 진행하지 않았다. 연결어미, 조사, 부사, 체언의 형태를 대부분류했을 때 체언을 제외하고는 나머지 문법 형태의 /ㄷ/ 대체율은 50% 정도로 상당히 높게 나타났다. 반면 체언은 대체율이 5% 미만으로 나타났다. 즉, /ㄴ/의 /ㄷ/ 대체 현상은 체언을 제외하고 실질형태소(부사)와 형식형태소(연결어미, 조사)의 구분 없이 상당히 높은 비율로 나타났다.

해당 문법 형태들을 보다 세분화한 중분류로 볼 때 조사는 보조사(59%)>부사격조사(42%)>접속조사(40%)의 순으로 /ㄷ/의 대체율이 높게 나타났다. 보조사 {도}는 전체 보조사의 99%를 차지할 만큼 출현비율이 높았는데, 이 항목을 제외하면 {대로}는 35%로 다른 조사들과 큰 차이가 없었다.

한편 용언이나 다른 말을 수식하는 역할을 하는 부사는 크게 문장부사와 성분부사로 나뉜다. 본고의 실험에서는 문장부사는 성분부사보다 /ㄷ/ 대체형의 출현이 높은 것으로 나타났다 (68% vs. 35%). 문장부사는 문장 전체의 문법적 의미에 관여한다. 반면 성분부사는 뒷말만을 수식하는 경우가 많고, 어떤 상태나 속성을 나타내는 형용사를 수식할 수 있다(Seo, 2005). 문장부사보다는 성분부사가 어휘적 의미에 깊게 관여한다. 이러한 속성이 음성 대체형에도 투영되어 성분부사보다는 문장부사에서 음성 변이형이 더 많이 나온 것으로 추측할 수 있다.

마지막으로 다른 항목과는 달리 유독 체언은 다른 문법 항목과 다르게 /ㄷ/의 대체율이 낮게 나타났다. 체언은 문법적인 특성보다는 어휘적인 특성이 강하게 드러나므로 음성 변이형의 출현이 어려운 것으로 보인다. 특히 명사는 사물의 개념을 표현하고, 수사는 사물의 수량이나 순서를 나타내므로 음소가 다른 음으로 대체된다면 사물의 의미가 변질되므로 변이형이 나타나기 어려울 것으로 판단된다. 그럼에도 불구하고 5% 정도의 대체율이 나온 것에 대해 다음과 같은 점을 고려해 볼 수 있다. /ㄷ/ 대체형이 나타난 체언의 음성 파일을 들어 보면 웅얼거리거나 목소리가 작아 제대로 목표음을 지각하기 어려웠고, 텍스트 파일에는 /ㄷ/로 전사되어 있으나 /ㄴ/와 /ㄷ/의 구분이 명확하지 않은 것들이 있었다. 체언 이외의 다른 문법 항목의 대체형들은, 예컨대 ‘별로 [별루]는 분명하게 /ㄷ/로 지각되는 반면 체언 대체형들은 /ㄷ/로 그다지 뚜렷하게 지각되지 않는 경우가 많았다.

그러나 본 실험의 분석은 필자의 주관적 판단에 따르기보다는 원자료의 전사 자료를 기반으로 분석했다. Seoul Corpus의 음성 및 전사 자료는 연구원들이 많은 노력을 들인 소중한 자료이다. 1차로 연구원들이 한글로 음소, 어절, 발화 단위로 자료를

전사하고, 2차로 자동음성인식기(Automatic Speech Recognition)를 통해 음소 경계를 자동으로 정렬한 후, 마지막으로 9명의 수작업으로 오류를 교정한 것이다. 차후 체언의 대체형에 대해서는 보다 정밀한 분석이 필요할 것으로 보인다.

한국인들의 발음형이 실제 언어생활에서 어떠한 방식으로 실현되는지를 관찰하는 것은 매우 중요하다. 그간의 선행연구는 주로 음향분석을 통해 /ㄴ/의 포먼트 값을 측정함으로써 고모음화 현상을 사회언어학적으로 살펴보려는 노력이 주를 이루었다. 본 논문은 음성 인식 연구에 실증적인 자료를 제공하고 자 문법 항목의 출현빈도와 대체율의 측면에서 논의를 진행했다. 구어 말뭉치를 기반으로 분석한 본 연구를 통해 한국어 후설원순 중모음 /ㄴ/의 표준발음과 실제발음형 /ㄷ/의 출현빈도 및 출현비율을 추출해 봄으로써 음성 인식과 훈련을 위한 실재성이 있는 자료를 제공했다는 점에서 연구 의의가 있다.

감사의 글

실험음성학 강독회 여름 워크숍을 통해 Seoul Corpus 음성 시료와 스크립트를 제공해 주신 윤원희, 윤규철 교수님께 진심으로 감사드립니다. 또한 phonometrica를 활용하게 해 주신 Julien Eychenne 교수님께도 감사의 말씀을 드립니다.

References

- Chae, S. Y. (1999). The core-periphery structure in the Korean lexicon reflected in a phonological variation and change. *Studies in Phonetics, Phonology and Morphology*, 5(1), 217-236.
- Eychenne, J., & Courdès-Murphy, L. (2019). Phonometrica: An open platform for the analysis of speech corpora. *Proceedings of the Seoul International Conference on Speech Sciences 2019* (pp. 107-108). Seoul, Korea.
- Ha, Y., & Oh, J. (2017). A study on vowel systems and aspects focusing on announcers vowel pronunciation. *Studies in Phonetics, Phonology and Morphology*, 23(1), 55-94
- Han, J. I., & Kang, H. (2013). Cross-generational change of /o/ and /u/ in Seoul Korean I: Proximity in vowel space. *Phonetics and Speech Sciences*, 5(2), 25-31.
- Kang, H., & Han, J. I. (2013). Cross-generational change of /o/ and /u/ in Seoul Korean II: Spectral interactions in normalized vowel space. *Phonetics and Speech Sciences*, 5(2), 33-41.
- Kang, J., & Kong, E. (2016). Static and dynamic spectral properties of the monophthong vowels in Seoul Korean: Implication on sound change. *Phonetics and Speech Sciences*, 8(4), 25-31.
- Kim, S., & Yoon, K. (2015a). A comparative study on the effects of age on the vowel formants of the Korean corpus of spontaneous speech. *Phonetics and Speech Sciences*, 7(3), 65-72.
- Kim, S., & Yoon, K. (2015b). A study on the male vowel formants of the Korean corpus of spontaneous speech. *Phonetics and Speech*

Sciences, 7(2), 95-102.

- Lee, H., Shin, W., & Shin, J. (2017). A sociophonetic study on high/mid back vowels in Korean. *Phonetics and Speech Sciences*, 9(2), 39-51.
- Lee, J., Yoon, K., & Byun, K. (2016). A study of vowel shift in Seoul Korean. *The Journal of Studies in Language*, 31(4), 979-998.
- Seo, S. (2005). Adverbs and adjectives. *Understanding of Korean Language*, 15(1), 179-192.
- Seong, C. (2004). An acoustic analysis on the Korean 8 monophthongs: With respect to the acoustic variables on the F1/F2 vowel space. *The Journal of the Acoustical Society of Korea*, 23(6), 454-461.
- Yang, S. I. (2018). A critical review on the studies of the Korean monophthongs: At the contact point of Korean educational linguistics. *Journal of Korean Linguistics*, 85, 429-462.
- Yoon, K., & Kim, S. (2015). A comparative study on the male and female vowel formants of the Korean corpus of spontaneous speech. *Phonetics and Speech Sciences*, 7(2), 131-138.
- Yun, W., Yoon, K., Park, S., Lee, J., Cho, S., Kang, D., Byun, K., ... Kim, J. (2015). The Korean corpus of spontaneous speech. *Phonetics and Speech Sciences*, 7(2), 103-109.

- **윤은경 (Eunkyung Yoon)** 교신저자
대구사이버대학교 한국어다문화학과 부교수
경북 경산시 진량읍 대구대로 201
Tel: 053-859-7428
Email: heavynub@dcu.ac.kr
관심분야: 음성학, 음운론

어말 위치 /ㄴ/의 /ㄷ/ 대체 현상에 대한 문법 항목별 출현빈도 연구

윤은경

대구사이버대학교 한국어다문화학과

국문초록

본 논문은 구어 말뭉치를 기반으로 한국어 /ㄴ/가 /ㄷ/로 고모음화되는 현상(예, ‘별로’ [별루])에 대해 문법 항목별로 차이를 살펴보는 데 연구 목적이 있다. 한국어의 /ㄴ/와 /ㄷ/는 [+원순성] 자질을 공유하지만, 혀 높이 차이로 변별된다. 그러나 최근 /ㄴ/와 /ㄷ/의 두 모음의 음성적 구분이 모호해지는 병합 현상이 진행 중이라고 여러 논문에서 보고되었다. 본 연구에서는 어말 위치의 /ㄴ/가 한국어 자연언어 구어 말뭉치(The Korean Corpus of Spontaneous Speech)에서 음성적으로 [o] 또는 [u]로 실현되는 현상을 연결어미, 조사, 부사, 체언의 문법 항목별로 출현빈도 및 출현비율에 대해 살펴보았다. 실험 결과 연결어미, 조사, 부사에서 /ㄴ/는 약 50%의 비율로 /ㄷ/로 대체되는 것을 확인했고, 체언에서만 상당히 낮은 비율인 5% 미만으로 대체가 되는 것을 알 수 있었다. 고빈도 형태소 중에서 가장 높은 /ㄷ/ 대체율을 보인 형태소는 '-도 [두]' (59.6%)였고, 연결어미에서는 '-고 [구]' (43.5%)로 나타났다. 구어 말뭉치를 통하여 실제 발음형과 표준발음의 차이를 살펴보았다는 데 연구 의의가 있다.

핵심어: 한국어 모음, /ㄴ/ 고모음화, 문법 항목 출현빈도, 모음 추이, 한국어 자연발화 음성코퍼스

참고문헌

- 강지은, 공은정 (2016). 서울 방언 단모음의 소리 변화와 음향 단서 연구: 단일지점 포먼트와 궤적 양상. *말소리와 음성과학*, 8(4), 39-47.
- 김순옥, 윤규철 (2015a). 한국어 자연발화 음성코퍼스의 연령별 모음 포먼트 비교 연구. *말소리와 음성과학*, 7(3), 65-72.
- 김순옥, 윤규철 (2015b). 한국어 자연발화 음성코퍼스의 남성 모음 포먼트 연구. *말소리와 음성과학*, 7(2), 95-102.
- 서상규 (2005). 부사와 관형사. *새국어생활*, 15(1), 179-192.
- 성철재 (2004). 한국어 단모음 8개에 대한 음향분석. *한국음향학회지*, 23(6), 454-461.
- 양순임 (2018). 단모음 연구에 대한 비판적 고찰-한국어교육학과 음성·음운론의 접점에서. *국어학*, 85, 429-462.
- 윤규철, 김순옥 (2015). 한국어 자연발화 음성코퍼스의 남녀 모음 포먼트 비교 연구. *말소리와 음성과학*, 7(2), 131-138.
- 이주희, 윤규철, 변군혁 (2016). 서울 방언의 단모음 변화 연구: 중세부터 현대까지. *한국현대언어학회지*, 31(4), 979-998.
- 이향원, 신우봉, 신지영 (2017). 한국어 후설 고중모음에 대한 사회음성학적 연구. *말소리와 음성과학*, 9(2), 39-51.
- 채서영 (1999). 음운 변화에 나타난 한국어 어휘의 층위구조. *음성·음운·형태론 연구*, 5(1), 217-236.
- 하영우, 오재혁 (2017). 아나운서의 단모음 실현 양상과 특징. *음성·음운·형태론 연구*, 23(1), 55-94.