



음절구조로 본 서울코퍼스의 글 어절과 말 어절의 음소분포와 음운변동

Phoneme distribution and phonological processes of orthographic and pronounced phrasal words in light of syllable structure in the Seoul Corpus

양 병 곤*

Yang, Byunggon

Abstract

This paper investigated the phoneme distribution and phonological processes of orthographic and pronounced phrasal words in light of syllable structure in the Seoul Corpus in order to provide linguists and phoneticians with a clearer understanding of the Korean language system. To achieve the goal, the phrasal words were extracted from the transcribed label scripts of the Seoul Corpus using Praat. Following this, the onsets, peaks, codas and syllable types of the phrasal words were analyzed using an R script. Results revealed that k0 was most frequently used as an onset in both orthographic and pronounced phrasal words. Also, aa was the most favored vowel in the Korean syllable peak with fewer phonological processes in its pronounced form. The total proportion of all diphthongs according to the frequency of the peaks in the orthographic phrasal words was 8.8%, which was almost double those found in the pronounced phrasal words. For the codas, nn accounted for 34.4% of the total pronounced phrasal words and was the varied form. From syllable type classification of the Corpus, CV appeared to be the most frequent type followed by CVC, V, and VC from the orthographic forms. Overall, the onsets were more prevalent in the pronunciation more than the codas. From the results, this paper concluded that an analysis of phoneme distribution and phonological processes in light of syllable structure can contribute greatly to the understanding of the phonology of spoken Korean.

Keywords: phoneme distribution, syllable structure, phonological processes, orthographic, pronounced, Korean corpus

1. 서론

최근에 서울코퍼스라는 한국어의 자연발화 코퍼스가 발표되었다(Yun *et al.*, 2015). 서울코퍼스는 40명의 서울화자가 다양한 주제에 대한 질문에 즉흥적으로 응답한 것을 녹음하여 전사한 것이다. 이들의 논문에서는 서울코퍼스의 113만개 어절을 이루는 음소 분포를 분석한 결과 자음이 52%, 모음이 48%로 다소 비슷한 분포를 보였다고 한다. 이 수치는 자음이 모음보다 4% 정

도 더 많다. 국어문장은 한글 맞춤법의 띄어쓰기 단위와 대략 일치하는 어절로 구성되어있고, 어절은 한 번에 소리 낼 수 있는 단위인 음절로 구분 된다(이문규, 2015). 음절은 발음상의 소리마디로 각각 초성(Onset), 중성(Peak), 종성(Coda)의 다양한 조합으로 만들어 진다(이호영, 1996). 여기서 한국어 음절구조에서는 초성으로 자음 ‘ㅇ’이 발음될 수 없고, 중성자리에는 ‘ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅇ’의 일곱 자음만이 올 수 있는 음절 구성의 제약(이문규, 2015)이 있기 때문에 단순히 전체의 말에서 쓰인

* 부산대학교, bgyang@pusan.ac.kr

Received 3 August 2016; Revised 16 September 2016; Accepted 20 September 2016

음소의 대략적인 분포보다는 음절구조로 분할하여 자세히 살펴볼 필요가 있다. 국어에서 원순모음은 같은 음절 안의 앞 자음을 원순화 시키기 때문에, 이호영(1996)은 음절이란 단위를 이용해야 음운 현상을 제대로 규칙화할 수 있다고 한다. 이와 같이 신승용(2009:19-20)도 음절을 통해 음운변동 현상의 본질을 파악하는데 도움이 된다고 지적했다. 그는 국어의 ‘ㄴ’ 불규칙 용언어간이 ‘-아/어/X’ 계 어미와 결합할 때 어간이 1음절일 때는 ‘누워, 고와’와 같이 모음조화를 지키지만, 어간이 2음절일 경우에는 모두 ‘-어/X’ 계 어미가 결합하여 ‘더러워, 고마워’와 같이 모음조화가 일어나지 않는다고 한다.

한국어 음소분포에 대한 또 다른 연구로는 배희숙 외(2000)가 1930년대의 한국회곡과 1990년대의 불어소설 번역본의 텍스트를 발음 변환을 시켜서 분석한 연구가 있다. 이들은 운율구단위로 발화음운론적인 차원에서 한글텍스트를 음소 수준으로 변환한 다음 자음과 모음의 분포를 분석했다. 그 결과 두 작품에서 사용된 음소의 빈도분포가 매우 높은 상관관계를 보여서 비슷하게 나타났고, 빈도는 텍스트의 성격에 따라 다르다는 결론을 제시했다. 자음과 모음의 비율은 한국어 회곡과 불어소설 번역본에서 54%:46%로 나타났다고 한다. 한편, 프랑스어의 모음 비율은 42%-44%로 문체나 텍스트의 성격에 따라 약간의 변화를 보인 것으로 조사되었다(Bae, 1997). 이러한 언어 간의 차이는 왜 발생하는지에 대해 설명하려면, 다양한 언어의 음운체계를 자세히 비교 분석할 필요가 있다.

영어단어의 특성을 살펴보기 위해 카네기멜론대학이 온라인으로 공개한 카네기멜론 발화사전(CMUPD)의 116,588개 표제어를 분석한 Yang(2016)의 연구에서는 모든 단어를 음절구조로 나누어 자음과 모음의 분포를 조사했다. 분석해본 결과 그는 CMUPD에서 자음과 모음의 비율이 6:4로 나타나 영어사전 표제어에 자음이 훨씬 더 많이 분포하고 있음을 지적했다. 영어에서 자음은 모음에 비해 약 20% 더 많은 것이 된다. 영어권의 자연발화 코퍼스인 벽아이코퍼스(Yang, 2012)에서는 실제발화의 총 음소 2,638,882개 중 자음은 1,580,547개이고 모음은 1,058,335로 보고되었고, 그 비율은 대략 6:4가 되는데, 이는 CMUPD와 같은 비율이다. Yang(2016)의 연구에서는 가장 많이 쓰인 자음이 N으로 11.7%를 차지했고 이어서 L이 9.7%로 이어졌다. 단어를 이루는 음절의 빈도를 분석해본 결과 2음절이 가장 많이 쓰였고, 이어서 3음절과 1음절의 순서로 나타났다고 한다. Yun *et al.*(2015)에서는 말 어절에서 어절의 크기가 1음절에서 3음절이 82%를 차지하였고, 2음절어절이 1음절과 3음절어절에 비해 약간 더 많다고 보고했다. 국어어절은 “이(가), 은(는)”처럼 주격, “을(를)”처럼 목적격을 나타내는 말마디가 포함되어 있으므로 영어의 사전을 분석한 Yang(2016)의 자료와 직접 비교하기는 어렵지만 대체로 2음절이 많이 쓰였음을 보여준다. 앞으로 단어 품사별 사전파일이 만들어지면 이러한 언어 간의 음절크기 분포에 대해 비교 연구가 가능할 것이다. 서울코퍼스를 음절구조로 구분한 뒤 글 어절과 발화음절로 분리하여 유형별로 분석한 자료에서는 CV(C: consonant, V: vowel)가 가장 자주 쓰인 유형이고 이어서 CVC, V, VC의 순서로 나타났으며

글 어절에서만 소수의 VCC와 CVCC가 나타났다고 한다(Yun *et al.*, 2015).

아직까지 충분한 분량의 자연발화에서 구한 코퍼스에서 음절을 중심으로 음소 분포를 계량적으로 분석한 사례가 드물기 때문에 이 논문에서는 서울코퍼스의 글 어절과 말 어절 자료를 분석하여 한국어의 음절구조에 대해 자세히 살펴봄으로써 한국어를 연구하는 언어학자나 음성학자들에게 구어로 발화된 한국어에 대한 기초자료를 제공하고자 한다. 구체적으로 이 연구에서는 다음과 같은 연구문제를 중심으로 살펴보고자 한다.

1. 서울코퍼스에서 글 어절과 말 어절을 음절구조로 나눈 자음과 모음의 분포는 어떠한가?
2. 서울코퍼스에서 글 어절과 말 어절 사이에 어떤 음운변동의 분포를 보이는가?
3. 서울코퍼스에서 글 어절과 말 어절의 음절유형에서는 어떤 차이를 보이는가?

이러한 연구의 결과는 다른 언어와의 비교에도 활용할 수 있고, 음절구조에 따른 초성과 중성에 분포비율을 참고하여 보다 높은 빈도의 음소로 음성인식에서 판단하는 도구나, 통계적인 분포 규칙에 따라 문자열을 음성으로 합성하는 자료 등으로 활용할 수 있을 것으로 기대된다.

2. 음성 자료

2.1. 서울코퍼스

서울코퍼스는 40명의 서울화자가 인터뷰 형태로 참가자 자신이나 가족, 마을의 이웃사람들이나 정치문제에 대한 의견이나 여가시간 등에 관한 질문에 대해 즉흥적으로 말할 것을 녹음하여 전사한 것이다(Yun *et al.*, 2015). 약 23만개의 어절이 들어가 있고, 정해진 문장을 읽게 하는 통제된 녹음 자료에서는 볼 수 없는 자연스러운 말의 특징을 살펴볼 수 있다. 녹음된 음성은 훈련된 전사자들에 의해 선택된 5,152개의 음소에 대해 98.1%의 일치율을 보일 정도로 일관성 있게 표기되었다. 자동 레이블러를 활용하였지만, 일일이 수작업으로 재확인을 거쳤기 때문에 신뢰할 만한 귀중한 자료로 여겨진다.

서울코퍼스에서 글 어절은 글로 나타낼 때 정서법에 따라 표기된 어절(orthographic syllables)을 말하고 말 어절은 실제 화자가 발성한 것을 소리마디별로 구분하여 표기한 어절(pronounced syllables)이다. 이 코퍼스에서는 말 어절 층위와 최상층의 실제 음소표기에 약간의 차이를 보이고 있다. 예를 들어 첫 번째 화자의 “잘못했다”는 글 어절에서는 c0aall-mmoos0-hhEEss-t0aa로 전사하고, 말 어절에서는 “잘못태따”로 표기하고 c0aall-mmoo-thEE-ttaa로 전사했는데, 최상위층에서는 EE를 ee로 표기하고 있다. 실제 서울 발음에서 남성은 두 개의 모음을 구분하기도 하지만 여성은 구분하지 않고 발음하는 경향을 보이기 때문에 발음을 구분하여 표기했는지는 알기가 어려울 것으로 여겨진다. 이 논문에서는 말 어절로 구분된

층위의 정보를 중심으로 분석하기로 한다. 앞으로 이 부분에 대해 융합적인 기준이나 지각기준을 이용해서 말 어절의 발음을 세분하여 추가적인 연구가 필요하다.

자음표기는 11개의 글 어절에서만 나타나는 중첩자음을 포함하여 30개의 기호가 쓰였는데, 독자들의 편의를 위해 서울코퍼스기호-한글기호-IPA의 순서로 차례로 여기에 적는다(p0-ㅍ-/p/, ph-ㅍ-/pʰ/, pp-ㅍ-ㅍ-/pʰ/, t0-ㄷ-/t/, th-ㅌ-/tʰ/, tt-ㄷ-ㄷ-/tʰ/, k0-ㄱ-/k/, kh-ㅋ-/kʰ/, kk-ㄱ-ㄱ-/kʰ/, s0-ㅅ-/s/, ss-ㅅ-ㅅ-/sʰ/, hh-ㅎ-/h/, c0-ㅈ-/tʃ/, ch-ㅈ-ㅈ-/tʃʰ/, cc-ㅈ-ㅈ-/tʃʰ/, mm-ㅁ-/m/, nn-ㄴ-/n/, ng-ㅇ-/ŋ/, ll-ㄹ-/l/, ks-ㄱㅅ-/ks/, nc-ㄴㅅ-/ntʃ/, nh-ㄴㅎ-/nh/, lk-ㄹㄱ-/lk/, lm-ㄹㅁ-/lm/, lp-ㄹㅍ-/lp/, ls-ㄹㅅ-/ls/, lT-ㄹㅌ-/ltʰ/, lP-ㄹㅍ-/lpʰ/, lh-ㄹㅎ-/lh/, ps-ㅍㅅ-/ps/). 이들 자음 가운데 ng을 제외하고는 초성에 모두 나타나고 중성에는 k0, nn, t0, ll, mm, p0, ng 만 대표자음으로 나타난다. 중성에서 단모음은 8개 (ii-이-/i/, ee-에-/e/, EE-애-/æ/, aa-아-/a/, xx-오-/ɔ/, vv-어-/ə/, uu-우-/u/, oo-오-/o/)로 표시되고 이중모음은 13개 (ye-예-/je/, YE-애-/je/, ya-야-/ja/, yv-여-/jə/, yu-유-/ju/, yo-요-/jo/, wi-위-/wi/, we-웨-/we/, wE-외-/we/, WE-왜-/we/, wa-와-/wa/, wv-워-/wə/, xi-의-/ii/)로 나타내었다. 단모음 가운데 ee, EE는 글어절에서 각각 구분하여 표시했고, 이중모음에서도 je와 YE, wE, WE, we 등으로 구분하여 나타냈는데, 서울화자들이라도 이들 모음을 명확하게 구분하지 않는 화자도 있어서 전사를 담당할 분들이 단모음과 이중모음을 들어서 구별하기에 어려움이 있었을 것으로 여겨진다. 그래서 최상층의 음소표기에서는 ee, je와 we로 표기되어 있다.

2.2. 자료 분석

자료 분석은 Praat(Boersma & Weenink, 2016)에서 글 어절과 말 어절을 먼저 뽑아냈고, 이어서 R(2016)을 이용하여 개별 음절로 구분한 뒤 음절 구성요소와 음절유형별로 나누어 글 어절과 말 어절이 서로 교차된 분할표를 만들어 분석했다. 먼저 Praat에서는 새 스크립트를 열어 아래의 코드를 입력하고 실행하여 outall63.txt이라는 텍스트 파일을 글 어절과 말 어절로 분리하여 하드디스크에 모두 저장했다. 1차 분석에서 llaak0oo가 음절 분리가 안된 채 전사되어 있는 것을 발견하고 llaa-k0oo로 음절표시를 넣어 다시 저장했다.

스크립트에 대해 간략히 설명해보면, 먼저 서울코퍼스의 말을 음성기호로 전사한 label폴더에서 240개의 파일이름 목록을 구한 다음 되풀이작업을 통해 각각의 파일을 열어 실제 말 어절로 표시된 3열과 글 어절로 표시된 6열을 하나씩 열어 시작 시간값과 끝 시간값을 삭제하고, “<”로 시작되는 추가정보 줄은 모두 제외하고, 음소기호로 표시된 음절을 글 어절과 말 어절의 순서로 덧붙여 쓰게 하여 한 개의 최종 텍스트 파일(outall63.txt)로 만들었다(부록 참고).

이어서 R을 이용해서 음절별로 분할한 뒤 음절과 음소 유형별로 분류하였다. R실행코드는 길기 때문에 지면상 전체적인 처리과정만 설명하기로 한다. 먼저 stringr이라는 라이브러리를 설치하고, outall63.txt를 불러와서 전체행의 길이를 구했다. 이 라이브러리에는 하이픈으로 구분된 어절의 요소의 개수를 파

악하거나 모음의 위치를 기준으로 음절의 구성 요소인 문자열을 분리하는데 편리한 다양한 함수가 포함되어 있다. 이런 함수 기능을 활용해서 자음(nn, ll과 같은 단자음과 ks, nc와 같은 중첩자음 목록을 따로 만들어서 처리했음)과 모음 목록을 메모리에 탑재하고, 8개의 열로 된 배열을 만들어 글 어절의 초성, 중성, 종성과 음절유형에 연이어 말 어절의 초성, 중성, 종성과 음절유형으로 문자열을 분할하여 한 줄의 배열로 만든 다음, outall.txt라는 파일에 덧붙여 쓰면서 하드디스크에 자동으로 기록하게 했다. 이 텍스트에는 231,632개의 어절이 포함되어 있다. 구체적으로 파일의 내용을 보면 outall63.txt 파일의 첫 번째 다섯줄은 다음과 같이 되어 있다.

```
Ortho Prono
nnee nnee
c0ee c0ee
ii-llxxmm-xxnn ii-llxx-mmxxnn
ii-k0oo-yo ii-k0uu-yv
```

여기서 첫줄의 Ortho Prono는 첫 번째 열이 글 어절이고 두 번째가 말 어절임을 나타낸다. 이어서 첫 번째 참가자가 “네, 제 이름은 이구요”라는 문장을 4개로 된 어절로 나타냈다. 서울코퍼스에서는 전사할 때 개인정보 보호차원에서 참가자의 이름을 삭제한 것으로 추정 된다(Meyer, 2002). 분류하는 과정은 먼저 철자발음의 각 음절의 문자수를 구한 다음 6개일 경우에는 첫 번째와 두 번째의 자음은 초성으로 세 번째와 네 번째는 중성으로 다섯 번째와 여섯 번째는 종성으로 배열에 넣고, 만약 중성이 중첩자음목록에 들어있으면 CVCC로 분류하여 네 번째 열에 넣었다. 이어서 말 어절에서 동일하게 처리하되 마지막 중첩자음목록은 해당하지 않기 때문에 CVC로 분류하여 네 번째 줄에 넣었다. 예를 들어, 두 번째 줄의 nnee는 모음 ee를 찾아서 모음목록에 들어가 있는지 확인하고 음절의 중성으로 지정하고 그 앞의 자음 nn은 초성으로 지정하고 음절유형은 CV로 분류하였다. 덧붙여, 텍스트 파일을 엑셀에서 불러올 때 열별 구분을 할 수 있도록 각 음소 다음에 콤마를 넣어서 저장했다. 이렇게 처리한 결과로 만들어진 outallsyl.txt 파일의 첫 부분은 다음과 같다.

```
1,OrthoO,OrthoP,OrthoC,OrthoT,PronoO,PronoP,PronoC,PronoT
1,nn,ee,,CV,nn,ee,,CV
1,c0,ee,,CV,c0,ee,,CV
1,,ii,,V,,ii,,V
2,ll,xx,mm,CVC,ll,xx,,CV
3,,xx,nn,VC,mm,xx,nn,CVC
```

첫 번째 숫자는 각 행의 해당 음절 위치를 나타낸다. 마지막으로 이렇게 만든 텍스트파일을 R에 불러와 attach하여 열 이름을 변수로 이용할 수 있게 만든 다음, table(OrthoO, PronoO)과 같이 R의 내장된 함수를 이용해 글 어절과 말 어절의 초성까지 분할표를 만들어 저장하거나 table(OrthoO)과 같이 초성의 유형별

목록과 빈도를 구하여 한 개의 파일로 저장함으로써 음절요소에 따른 각 음소별 분포를 엑셀의 내장된 함수를 이용해서 빈도에 따른 정렬과 비율을 구해 분석했고, 중성과 중성 및 음절유형도 이런 방식을 적용해 모두 살펴보았다.

3. 분석 결과

3.1. 음절요소별 분석

서울코퍼스를 글 어절과 말 어절의 음절 구성요소별로 구분하여 <표 1, 2, 3>과 같이 초성, 중성, 종성별로 분할표를 만들어 나타내었다. 이렇게 글 어절과 말 어절을 서로 교차하여 표로 나타내면 글 어절이 실제 발화에서는 어떻게 변했는지 살펴볼 수 있는 장점이 있다. <표 2>에서 보면 서울코퍼스의 글 어절이나 말 어절에서 중성인 모음이 하나씩 들어간 경우를 모두 합치면 전체음절수는 555,690개이다. 이 수치는 서론에서 살펴본 Yun et al.(2015)이 보고한 글 어절과 말 어절을 중복되게 모두 포함한 113만개의 음소와는 달리 어절별로 분할한 자료이다.

서울코퍼스에서 초성의 음소 분포를 보면 글 어절에서 가장 많이 차지한 음소는 k0로 공백 (Null)을 제외한 순수한 초성의 빈도인 424,847개에서 약 23.4%를 차지하고 있으며 말 어절에서도 k0이 순수한 초성의 빈도인 443,491개에서 19.4%를 차지하고 있다. 이어서 글 어절에서는 t0, c0, nn이 약 10%를 차지하고 s0, ll이 약 8%, hh, mm이 7%, p0은 4%를 차지하고 나머지는 2%이하를 차지한다. 가장 낮은 비율을 보인 초성은 pp로 0.3%가 되며 된소리 ss와 cc도 0.5% 이하를 차지하는데, 이들은 말 어절에서도 0.5%로 매우 낮다. 말 어절에서는 k0다음으로 nn이 12.2%를 차지하고 이어서 ll, t0, c0 등이 약 9%씩 차지하고 mm, s0이 7%, p0, kk, hh가 3%, 나머지 음소들은 2%범위에 해당한다. 참고로 이 논문에서는 앞 절의 outall63.txt에서 보았듯이 nnee와 같은 감탄사에 대한 자세한 분포를 다루지 않았지만, nnee라는 어절이 차지하는 빈도는 4,218개로 나타났고, k0xx가 3,875개, k0vv가 3,093개 등으로 나타났는데 이런 음소들도 빈도에 일부 포함되어 있음을 인지하기 바란다. 이러한 음소분포 결과는 중성을 포함하지는 않았지만, UPSID자료에서 Maddieson(1984)이 조사한 음소의 분포에서 80%이상의 언어가 /p, t, k, m, n, s, j/ 음소를 가지고 있고, /k/를 가장 선호한다는 결과와 일치한다. 글 어절에서 말 어절로 갈 때 가장 많이 줄어든 음소는 hh로 약 16,355개가 줄어들었는데 비율로는 약 4%가 된다. 그 이유는 앞선 음절의 중성자음을 발음하면서 삭제된 경우가 10,208개나 되고 앞선 중성 파열음 등과 결합해서 거센소리로 변환 사례가 많기 때문으로 추정 된다(구체적인 음운변동 사례는 <표 1> 6째 줄 참고). 이 연구결과에서도 보듯이 언어마다 음절구조가 다르기 때문에 한국어와 같이 특징적인 구조를 가지는 경우에는 음절구조로 분리하여 음절요소별로 살펴보아야 전체적인 음소분포에 대해 바르게 논의할 수 있을 것으로 생각한다. 이어서 k0도 약 4%의 비율인 13,218개가 줄어들었는데, 발음이 없어진 사례가 6,383이고 된소리인 kk로 바뀐 경우가 9,449번이며 거센소리인 kh도 2,745개나 된다. 이러한 음운변동

이 일어난 환경과 변화비율 등은 차후에 서울코퍼스의 실제 자료에서 찾아 구체적으로 연구할 필요가 있다.

이번에는 <표 2>를 중심으로 중성으로 분류되는 국어모음의 글 어절과 말 어절의 분포에 대해 살펴보기로 한다. 먼저 글 어절에서 가장 많이 나타난 음소는 aa로 공백을 제외한 중성의 빈도인 441,142개에서 20.6%를 차지했고, 말 어절에서도 순수한 중성의 빈도인 440,610개의 21.1%를 차지했다. 이 모음은 고스란히 말 어절에서도 나타나고 추가로 다른 모음 532개가 aa로 발음될 정도로 안정적이고 선호하는 모음으로 여겨진다. 참고로 백아이크퍼스에서는 글 음소가 실제 말에서는 약 61.8%가 감소했고, 분절음 단위로 보면 모음의 63.0%가 감소했고, 자음의 61.0%가 발음되지 않았다고 한다(Yang, 2012). 감탄사로 된 어절에서는 aa, ee, ye가 각각 1,620개, 1,586개, 1,289개 등으로 나타났는데, 이런 음소들은 중성의 빈도에 미미한 영향을 준 것으로 여겨진다. 이어서 xx, ii가 글 어절과 말 어절에서 모두 14%를 차지했고, vv가 13.1%를, oo가 글 어절에서 9.6%를 차지하고 말 어절에서는 8%로 약간 내려갔다. 글 어절에서 ee와 EE는 각각 7.3%, 4.7%를 차지했으나 말 어절에서는 ee가 7.8%로 미세하게 상승했고, EE는 5.6%로 하락해서 입 벌림을 좀 더 해야 하는 EE발음이 실제 발음에서는 조음동작이 더 적은 ee로 말 어절에서도 대체된 것으로 추정된다. 일반적으로 입 벌림 정도는 F1으로 나타나는데(Pickett, 1987), Yang(1996)은 20명의 한국인 화자가 발음한 국어모음의 포먼트를 측정하여 모음삼각도를 나타내어본 결과 입 벌림 정도를 반영하는 F1값에서 ee와 EE발음의 차이가 남을 지적했다. 앞에서 언급했듯이 최상층의 음소 표기에서는 이들 발음이 모두 ee로 표시되어 있다. uu는 글 어절에서 4.2%인데, 말 어절에서는 6.6%로 12,746개가 증가했다. xx는 철자발음에서는 78,649개인데, 이 가운데 74,583개가 말 어절에서 쓰였고, 다양한 종류의 단모음으로 변형된 모양을 보인다.

이중모음 we는 <표 2>에서 가장 적은 음소로 나타났는데, 글 어절에서는 58회로 0.01%에 해당하며 이마저도 말 어절에서는 39개만 실현되었다. 이중모음은 개별적으로는 대체로 3%이하를 보이는데 글 어절에서 이중모음을 모두 합쳐도 48,210개로 8.8%를 차지한다. 말 어절에서는 이중모음의 빈도가 67,500개로 전체음소에서 15.3%를 차지해서 거의 2배나 증가된 현상을 보인다. 이 가운데 yv가 글 어절 3.2%에서 말 어절에서 4.8%로 8,285개 증가했는데, 그 변화의 중심에는 글 어절의 9,445개를 차지하는 yo가 바뀐 것이다. 앞 절의 outall63.txt 사례에서도 볼 수 있고 s0l1m16f1 화자의 발화에서는 “이구여, 올라가여, 살아여” 등과 같이 30개 글 어절의 “요”가 말 어절에서는 모두 “여”로 교체되었듯이, 아마 젊은 화자들이 친근한 대화를 하면서 서술 어미에서 많이 사용한 것이 증가의 원인으로 여겨지는데, 앞으로 남녀노소 집단별로 분석을 하면 이러한 사회언어학적인 측면의 언어사용 경향도 살펴볼 수 있을 것이다. 음운변동에서 글 어절이 말 어절에서 가장 많이 줄어든 경우는 yo가 11,138개나 바뀌어서 약 2%의 차이를 보였다. w로 시작되는 이중모음군 가운데 단모음으로 변환 경우가 눈에 띄게 많은데, 구체적으로

wa는 4,932개 가운데 40.2%인 1,981개가 aa로 발음되었고, wE에서도 4,981개 중 64.1%인 3,193개가 EE로, wv는 7,257개 가운데 43.5%인 3,157개가 vv로 발음되었고, wi도 1,238개 중 43.5%인 539개가 ii로 발음되었다. 그런데, y로 시작되는 이중모음군은 단모음화의 비율이 상대적으로 낮게 나타난다. 예를 들어, ya 음소 4,766개 가운데 1%인 51개만 aa로 실현되었고, ye와 YE도 각각 e와 E로 각각 36.2%, 15.7%가 바뀌었고, yu는 1,521개 가운데 1.2%인 19개만 uu로 실현되었다. 이러한 이중모음의 단모음화 경향의 차이는 주변음절환경이나 음절핵인 모음의 분포와도 관련이 있고, 단모음과의 구별이 기능적으로 필요한 경우에는 변화가 적은 것으로 추정되는데 서울코퍼스 자료를 중심으로 좀 더 자세한 음운변동 환경과 변화된 음소에 대해 연구할 필요가 있다.

마지막으로 <표 3>에서 중성의 음소분포를 살펴보면, 글 어절에서 가장 많이 나타난 음소는 mn으로 77,298개나 되어 공백을 제외한 순수 중성음소 가운데 35.8%를 차지한다. 이어서 ll이 35,575개로 16.5%를 차지했으며, ng가 19,348개로 11.1%, k0이 19,348개로 8.9%, mm이 16,727개로 7.7%, ss가 14,753개로 6.8%를 차지했고, hh, p0, hn, th 등이 약 2%이고 나머지는 모두 2%이하를 나타낸다. 말 어절에서는 mn이 49,795개로 순수 중성음소 가운데 가장 많은 34.4%를 차지했고, 이어서 ng가 39,541개로 27.3%, ll이 25,399개로 17.5%, mm이 17,830개로 12.3%를 차지했다. k0은 7,436개로 5.1%를, p0은 3,718개로 2.6%, t0은 1,103개로 0.8%를 보였다. 2개로 나타난 ph는 실제 발음에서는 대표음 p0로 표기되어야 하는데 전사상 어려로 여겨진다. 말 어절에서 비음인 mm, ng, nn이 차지하는 비율은 전체에서 74%나 되어서 우리말 중성은 비음이 매우 선호되는 언어라고 할 수 있다. 글 어절을 기준으로 대표음으로 변하는 음운변동 과정을 몇 가지 살펴보면, c0은 930개 가운데 108개는 mn으로 바뀌고 19개는 t0로 변했는데 788개나 발음되지 않는 공백으로 나타났다. hh는 5,460개 가운데 거의 97.8%가 발음되지 않았으며, ll은 35,575개 중 66.2%가 발음되었고, 그 중에 33.5%가 발음되지 않았다. 비

음군에 속하는 mm, ng, nn은 원래의 글 어절음소 가운데 각각 57.3%, 94.6%, 68%나 발음되었다. 왜 이런 비음을 국어의 중성에서 선호하는지는 이런 비음의 분포를 바탕으로 앞으로 발화의 편이성이나 지각 구별의 용이함이라는 변수 등을 중심으로 실험 등을 통해 더 연구할 필요가 있다. 한 가지 더 지적할 것은 비음가운데서도 nn은 비음동화를 통해 다양한 중성자음으로 바뀌어 나타나는데 원래의 음소를 그대로 발음한 것은 44,303개로 57.3%이고, 이어서 ng로 바뀐 것이 13,920개로 18.0%, mm으로는 5.9%인 4,529개로 변했으며, ll로 바뀐 것은 0.7%인 573개나 된다. 비음동화에 의한 발음은 현행 표준 발음법 제5장 18항에서는 많은 중성이 따라오는 ‘ㄴ, ㄹ’ 앞에서 비음으로 발음한다고 적고 있다. 앞서 보았던 ll의 빈도가 높은 것도 원래부터 글에 쓰인 사례도 있지만, 표준 발음법에서 ‘ㄴ’이 ‘ㄹ’앞이나 뒤에서 ‘ㄹ’로 발음되는 국어 유음화현상에 의한 영향도 많을 것으로 여겨진다. 이러한 중성의 분포에 끼치는 음운변동의 영향도 성별 나이별 집단별로 조사해보면 흥미로운 결과가 나올 것이다 (Yang, 2012; Yun *et al.*, 2015). 마지막으로 글 어절에서는 339,497개가 중성이 없는데 이는 전체 철자음절수의 61.1%가 된다. 말 어절에서는 이 숫자가 410,866개로 73.9%로 증가했다. 이는 참여자들이 발음에서 대다수의 중성을 발음하지 않았음을 나타내는데, 다음 절의 음절유형에서 자세히 살펴보기로 한다. 지금까지 음절구조로 나누어 서울코퍼스의 글과 말의 음소 분포를 살펴본 결과 국어에만 볼 수 있는 독특한 음절의 구성요소별 음소의 분포와 음운변동을 관찰할 수 있었다.

표 1. 서울코퍼스 음절초성의 글 음소(OrO)와 말 음소(PrO) 분할표

Table 1. Contingency table of orthographic (OrO) and pronounced phonemes (PrO) of the syllable onsets in the Seoul Corpus

OrO/PrO	Null	c0	cc	ch	hh	k0	kh	kk	ll	mm	nn	p0	ph	pp	s0	ss	t0	th	tt	sum
Null	86556	982	33	588	62	6022	98	365	8291	3963	11589	1475	664	6	767	6017	390	2915	60	130843
c0	941	38419	5207	562	0	71	4	8	6	8	18	2	0	0	19	2	33	2	4	45306
cc	7	20	2115	14	0	0	0	0	0	0	1	0	0	0	0	0	0	1	4	2162
ch	8	13	13	7690	1	1	2	0	0	1	0	0	1	0	3	1	1	3	0	7738
hh	10208	19	2	37	15021	95	1520	60	1250	507	1946	4	352	0	58	7	14	619	7	31726
k0	6383	75	5	2	249	79706	2745	9449	94	14	166	6	0	1	177	16	225	4	40	99357
kh	5	0	0	0	0	5	2092	59	1	0	1	0	2	0	0	0	1	1	1	2168
kk	1310	6	1	0	2	21	6	5646	1	1	7	0	0	0	1	0	1	0	0	7003
ll	1812	58	3	125	4	142	640	560	31885	44	383	2	1	0	496	1	112	13	29	36310
mm	260	0	1	0	3	1	0	1	4	29816	23	45	4	0	40	4	1	0	0	30203
nn	2703	19	0	1	7	33	1	208	467	396	39903	0	0	0	7	3	491	3	5	44247
p0	49	0	0	0	1	8	2	3	1	35	3	15686	11	869	3	0	8	13	1	16693
ph	8	0	0	1	5	1	0	1	2	7	1	10	3516	69	0	1	0	3	0	3625
pp	0	0	0	0	0	0	0	0	0	0	0	2	6	1080	0	0	0	0	0	1088
s0	898	36	5	14	11	8	0	5	11	10	22	6	0	0	32133	4073	18	5	3	37258

ss	3	2	1	0	1	0	0	0	0	0	0	0	1	0	22	1655	0	0	0	1685
t0	975	45	4	1	2	23	0	2	44	4	100	7	0	0	15	4	39283	363	5522	46394
th	70	1	1	8	2	0	2	0	6	0	3	4	0	0	1	1	22	4708	31	4860
tt	3	0	7	0	0	2	0	2	0	0	0	0	0	0	0	0	14	239	6757	7024
sum	112199	39695	7398	9043	15371	86139	7112	16369	42063	34806	54166	17249	4558	2025	33742	11785	40614	8892	12464	555690

표 2. 서울코퍼스 음절중성의 글 음소(OrP)와 말 음소(PrP) 분할표

Table 2. Contingency table of orthographic (OrP) and pronounced phonemes (PrP) of the syllable peaks in the Seoul Corpus

OrP/PrP	Null	aa	ee	EE	ii	oo	uu	vv	wa	we	wE	WE	wi	wv	xi	xx	ya	ye	YE	yo	yu	yv	sum
Null	0	72	53	4	91	35	33	93	1	0	0	0	0	1	3	177	14	0	0	1	2	20	600
aa	1545	111001	102	989	117	46	93	249	83	0	0	0	0	1	0	172	70	1	1	45	0	33	114548
ee	2300	32	37149	534	234	5	4	102	0	0	6	0	3	0	0	91	2	67	0	5	0	73	40607
EE	235	40	571	24317	43	17	15	526	0	1	4	1	0	0	0	114	3	6	2	3	0	12	25910
ii	1404	380	814	27	74164	35	595	78	0	0	1	0	17	1	3	907	16	7	0	7	6	53	78515
oo	522	49	8	8	30	40321	11853	292	2	0	2	0	0	5	0	404	0	0	0	6	3	14	53519
uu	45	13	4	7	37	39	22841	48	2	0	1	0	5	5	0	189	0	0	0	0	22	5	23263
vv	1647	713	765	52	262	229	183	65355	3	0	2	0	1	10	1	3069	13	0	0	40	1	233	72579
wa	5	1981	1	1	4	5	0	21	2907	0	0	0	0	1	0	3	1	0	0	2	0	0	4932
we	0	0	2	2	4	0	0	0	0	35	3	12	0	0	0	0	0	0	0	0	0	0	58
wE	3	4	364	3193	10	1	1	0	0	1	1378	12	0	0	0	13	0	1	0	0	0	0	4981
WE	0	0	43	741	1	1	0	0	0	1	137	473	0	0	0	1	1	0	0	0	0	0	1399
wi	0	0	0	2	539	0	2	1	0	0	0	0	625	0	0	1	0	0	0	0	2	66	1238
wv	9	22	1	3	3	2456	54	3157	8	1	0	0	0	1475	0	51	1	0	0	0	0	16	7257
xi	10	2	622	7	1504	0	0	4	0	0	1	0	2	0	169	76	2	9	1	0	1	1	2411
xx	819	632	790	339	402	133	302	435	1	0	0	0	0	0	4	74583	156	0	0	3	1	49	78649
ya	199	51	14	28	24	0	0	2	2	0	0	0	0	0	0	5	4426	2	0	0	0	13	4766
ye	0	1	1238	84	17	0	0	0	0	0	2	0	0	0	0	0	2	2056	2	2	0	17	3421
YE	0	0	37	214	4	0	0	0	0	0	0	0	0	0	0	1	1	24	1086	0	0	0	1367
yo	412	11	11	2	14	267	9	1039	0	0	1	0	1	1	0	24	14	2	0	5080	7	9445	16340
yu	2	1	1	0	25	0	19	3	0	0	0	0	2	0	0	19	0	0	0	1	1447	1	1521
yv	129	75	17	52	64	9	5	726	1	0	0	0	0	0	0	669	11	0	0	7	1	16043	17809
sum	9286	115080	42607	30606	77593	43599	36009	72131	3010	39	1538	498	656	1500	180	80569	4733	2175	1092	5202	1493	26094	555690

표 3. 서울코퍼스 음절중성의 글 음소(OrC)와 말 음소(PrC) 분할표

Table 3. Contingency table of orthographic (OrC) and pronounced phonemes (PrC) of the syllable codas in the Seoul Corpus

OrC/PrC	Null	k0	ll	mm	ng	nn	p0	ph	t0	sum
Null	335755	93	864	431	855	1069	84	0	346	339497
c0	788	4	0	5	0	108	6	0	19	930
ch	227	4	0	57	0	31	6	0	37	362
hh	5341	16	7	1	21	42	0	0	32	5460
k0	11035	7090	10	8	1187	4	9	0	5	19348
kh	2	0	0	0	0	0	0	0	0	2
kk	314	21	0	0	2	0	0	0	1	338
ks	1	3	0	0	0	0	0	0	0	4
lh	207	0	48	0	0	1	0	0	0	256
lk	9	11	92	0	11	2	0	0	1	126
ll	11913	9	23566	16	11	40	5	0	15	35575
lm	0	0	89	23	19	0	0	0	0	131
lp	1	0	124	0	0	0	0	0	0	125
lT	0	0	1	0	0	0	0	0	0	1
mm	4758	2	9	11379	511	62	5	0	1	16727
nc	0	0	0	0	7	47	0	0	0	54
ng	1192	8	2	54	22614	45	1	0	0	23916
nh	3833	0	1	0	359	441	0	0	0	4634

nn	13948	11	573	4529	13920	44303	3	0	11	77298
p0	2066	60	2	776	3	4	1959	0	0	4870
ph	686	2	0	8	0	0	250	2	0	948
ps	57	20	0	393	0	2	1361	0	0	1833
s0	2735	24	0	136	4	231	26	0	236	3392
ss	11418	42	7	9	17	3080	1	0	179	14753
t0	620	9	4	2	0	133	0	0	155	923
th	3960	7	0	3	0	150	2	0	65	4187
sum	410866	7436	25399	17830	39541	49795	3718	2	1103	555690

3.2. 음절유형별 분석

<표 4>는 서울코퍼스의 글 어절과 말 어절을 음절단위로 구분한 후 음절 유형별로 나누어 분할표를 나타낸 것이다.

표 4. 서울코퍼스의 글 음절유형 (OrT)과 말 음절유형(PrT) 분할표

Table 4. Contingency table of orthographic syllable (OrT) and pronounced syllable types (PrT) in the Seoul Corpus

OrT/PrT	Null	CV	CVC	V	VC	sum
Null	0	243	119	187	51	600
CV	7177	236492	2409	8562	564	255204
CVC	1439	52157	103828	3643	4251	165318
CVCC	0	3559	759	7	0	4325
V	474	28710	203	53910	396	83693
VC	196	4031	10946	9537	19001	43711
VCC	0	5	30	537	2267	2839
sum	9286	325197	118294	76383	26530	555690

앞서 Yun et al.(2015)은 서울코퍼스의 분류에서는 CVC유형이 가장 많은 분포를 보였다고 보고했는데, <표 4>를 보면 글 어절에서는 말 어절에 공백으로 나타난 600개를 제외하고 555,090개에서 CV유형이 46%를 차지하며, 이어서 CVC유형이 29.8%, V유형은 15.1%, 글 어절에서 겹자음이 들어간 CVCC와 VCC유형은 7,164개로 전체에서 약 1.3%밖에 차지하고 있지 않을 정도로 작아서 중첩자음은 우리말에 거의 활용되고 있지 않을 수 있다. 글 어절에 나타난 555,690개의 음절이 말 어절에서는 9,286개가 공백으로 사라졌음을 알 수 있다. 이를 제거하면 546,404개가 되며 이 가운데 말 어절의 CV유형이 차지하는 비율은 59.5%, CVC유형은 21.7%를 차지하고 V는 14.0%, 마지막으로 VC는 4.9%를 차지한다. 이러한 분포를 보면 국어음절의 발화에서 초성이 들어간 CV, CVC유형이 81.2%로 초성이 없는 V, VC유형보다 훨씬 많이 쓰였다고 할 수 있다. 덧붙여, 중성이 없는 음절유형인 CV와 V는 전체음절의 73.5%나 된다. 이러한 결과는 우리말이 중성보다는 초성을 더 많이 활용하여 주변 음절과 구별되는 의미전달을 시도하고 있음을 알 수 있다. 여기서 중성의 자음은 선행하는 초성과 중성으로 된 음절의 소리의 크기가 커서 따라오는 중성을 가리어 지각하기 어렵기 때문에 더 적게 사용되었을 것으로 추정된다. 하지만 중성의 유무는 지각적으로 큰 차이를 보일 수 있고, 울림도가 낮은 중성을

통해 음절의 경계를 인식하게 하거나, 초성과 중성이 서로 다른 음소일 경우에는 다른 음절과 구별하기가 더 쉬운 장점도 있다. 이러한 구별되는 소리의 크기는 영어의 초성에 겹자음이 올 경우 첫 자음보다는 공명도가 높은 자음이 두 번째 따라 오게 되는 경향과도 관련이 있다(Duanmu, 2002). Yang(2016)은 영어 단어의 음절을 분석해본 결과 단어의 자체음절에서 초성과 중성을 93.3% 달리하여 다양한 단어를 구성하였고, 인접한 음절에서도 91.6%가 다른 유형의 모음을 사용하여 서로 구별되는 다양한 단어를 사용하고 있음을 밝혔다. 그의 연구는 2,001개의 CVC단어를 수작업으로 분류하여 보고한 Kessler & Treiman(1997)의 연구에서 제안된 이론을 CMUPD의 116,588개의 표제어에 대해 분석하여 뒷받침한 것이다.

음절유형별 음운변동을 <표 4>에서 살펴보면 CV유형의 음절이 글 어절의 음절에서 공백을 제외한 248,027개에서 말 어절에서는 236,492개로 가면서 다른 유형에 비해 가장 많은 95.3%가 원래의 유형을 유지했다. CVC유형에서는 글 어절 공백을 제외한 163,879개 가운데 63.4%인 103,828개가 말 어절에서 나타났다. V도 64.8%가 유지되었는데, VC는 43.7%만 유지되어 낮은 유지비율을 나타내고 있다. 이러한 결과는 국어의 음운변동에서 해당음절이 중성만 있을 때 앞의 중성과 연결되어 유형이 바뀌거나, 해당음절이 중성이 따라오는 음절의 초성에 연음되어 채음절화 됨으로써 유형이 바뀌었기 때문으로 추정되지만, 구체적으로 음성전사 자료에서 어떤 음운변동에 의해 이런 변화가 일어났는지는 앞으로 더 연구가 필요하다.

4. 논의 및 결론

이 논문에서는 40명의 서울화자가 즉흥적으로 발화한 문장을 전사한 서울코퍼스의 글 어절과 말 어절을 각 음절별로 분리하고 초성·중성·종성의 음절 구성요소별로 나타난 음소들의 분포와 음운변동을 살펴보았다. 연구방법으로는 Praat의 층위표기자료를 읽어 들어 어절단위로 나눈 다음, R의 table 함수를 이용하여 분할표를 통해 자세히 분석해 보았다. 분석결과를 요약하면 다음과 같다.

첫째, 서울코퍼스의 음절을 분석해본 결과 초성의 음소분포에서는 k0이 글 어절에서 23.4%, 말 어절에서 19.4%로 가장 많이 나타났고, 된소리는 낮은 비율을 차지했다. 중성의 음소분포에서는 단모음인 aa가 글 어절에서 20.6%, 말 어절에서 21.1%를 차지했고, 음운변동에서도 가장 안정적으로 유지되었다. 이에

반해 이중모음들은 글 어절에서는 모두 합쳐도 8.8%인데 말 어절에서는 15.3%로 거의 2배나 증가되었다. 이러한 변화의 중심에는 yo가 yu로 바뀐 것이 주된 이유로 관찰되었다. 특히, 글 어절에서 w로 시작되는 이중모음군이 말 어절에서는 단모음으로 바뀌는 현상이 두드러졌다. 말 어절에서 가장 많이 쓰인 종성은 nn으로 34.4%를 차지했고, mm과 ng를 포함하면 전체 종성의 음소빈도에서 74%나 차지할 정도로 선호되었다. 비음인 nn은 글 어절에서 말 어절로 가장 다양하게 변했다.

둘째, 음절유형별로 글 어절과 말 어절을 비교해본 결과 CV 유형이 가장 큰 빈도를 보였고 이어서 CVC, V, VC의 유형 순서로 나타났다. 전체적으로는 초성이 들어간 음절이 종성이 들어간 음절보다 더 많이 나타나 초성이 발화에 더 많이 활용되는 것으로 나타났다. 음운변동에서는 글 어절에서 말 어절로 실현될 때 CV가 가장 변화가 적었고, CVC, V, VC의 순서로 변화가 많이 나타났다.

이러한 결과를 보면 국어음소의 분포를 단순히 전체 말과 글로 표현된 어절에서 나타난 음소분포만으로는 파악하기 힘들고 음절구조로 분리하여 살펴봐야 언어의 쓰임새를 바르게 파악할 수 있다고 결론지을 수 있다.

앞으로 서울코퍼스의 연령과 남녀 집단별 세부 분석을 통해 좀 더 자세히 이런 음운변동을 살펴볼 계획이다.

감사의 글

한국어 자연발화자료를 연구할 수 있도록 서울코퍼스를 만들고 무료로 제공해주신 윤원희를 비롯한 연구자 여러분들께 심심한 감사를 드립니다.

참고문헌

Bae, H.-S. (1997). *Structures lexicales, syntaxiques et phonétiques dans deux pièces de J. Tardieu*. Ph.D. Dissertation, Strasbourg.

Bae, H.-S., Koo, D.-O., Yun, Y.-S., & Oh, Y.-H. (2000). A quantitative study for the distribution of Korean phonemes in the two parts: *The Ox and Waiting for Godot*. *Speech Sciences*, 7(4), 27-40. (배희숙·구동욱·윤영선·오영환 (2000). 한국어 음소분포에 대한 계량언어학적 연구: 『소』와 『고도를 기다리며』를 중심으로. *음성과학*, 7(4), 27-40.)

Boersma, P. & Weenink, D. (2016). Praat: Doing phonetics by computer. Retrieved from <http://www.fon.hum.uva.nl/praat/> on July 1, 2016.

Duanmu, S. (2002). Two theories of onset clusters. *Chinese Phonology*, 11, 97-120.

Kessler, B. & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language*, 37, 295-311.

Lee, M. (2015). *Modern Korean phonology for Korean education*.

Seoul: Hankookmunhwasa. (이문규 (2015). *국어교육을 위한 현대 국어 음운론*. 서울: 한국문화사.)

Meyer, C. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.

Pickett, J. (1987). *The sounds of speech communication: A primer of acoustic phonetics and speech perception*. Austin, Texas: pro-ed.

R. Core Team. (2016). R: A language and environment for statistical computing. Retrieved from <https://www.r-project.org/> [R Foundation for Statistical Computing, Vienna, Austria] on July 1, 2016.

Shin, S. (2009). *Korean syllable phonology*. Seoul: Bakijung. (신승용 (2009). *국어 음절 음운론*. 서울: 박이정.)

Yang, B. (1996). A comparative study of English and Korean monophthongs produced by male and female speakers. *Journal of Phonetics*, 24, 245-261.

Yang, B. (2012). Reduction and frequency analyses of vowels and consonants in the Buckeye Speech Corpus. *Phonetics and Speech Sciences*, 4(3), 75-83.

Yang, B. (2016). Phoneme distribution and syllable structure of entry words in the CMU English Pronouncing Dictionary. *Phonetics and Speech Sciences*, 8(2), 11-16.

Yun, W., Yoon, K., Park, S., Lee, J., Cho, S., Kang, D., Byun, K., Hahn, H., & Kim, J. (2015). The Korean Corpus of Spontaneous Speech. *Phonetics and Speech Sciences*, 7(2), 103-109.

• 양병곤 (Yang, Byunggon)

부산대학교 영어교육과
부산시 금정구 장전동 30
Tel: 051-510-2619
Email: bgyang@pusan.ac.kr
Homepage: <http://fonetiks.info/bgyang>

부록 . 글 어절과 말 어절 추출 Praat 스크립트

```
clearinfo
newname63$="C:\seoulcorpus\out63\outall63.txt"
appendFileLine: newname63$,"Ortho"," ","Prono"
Create Strings as file list... fileList C:\seoulcorpus\label\*. *
numberlist=Get number of strings
for i from 1 to numberlist
select Strings fileList
name$=Get string... i
dotpoint=rindex(name$,".TextGrid")-1
newname$=left$(name$,'dotpoint')
textGridname$="TextGrid "+newname$
Read from file... C:\seoulcorpus\label\name$'
selectObject: textGridname$
Extract one tier: 3
```



```

Down to Table: "no", 1, "no", "no"
norows=Get number of rows
Remove column: "tmax"
Remove column: "tmin"
selectObject: textGridname$
Extract one tier: 6
Down to Table: "no", 1, "no", "no"
Remove column: "tmax"
Remove column: "tmin"
for j from 1 to norows
    selectObject: "Table pWord_prono"
    prono$=Get value: j, "text"
    if left$(prono$,1)=="<"
        prono$="0"
    endif
    selectObject: "Table pWord_ortho"
    ortho$=Get value: j, "text"
    if left$(ortho$,1)=="<"
        ortho$="0"
    endif
    if ((prono$>"0") and(ortho$>"0"))
        appendFileLine: newname63$, ortho$, " ", prono$
    endif
endifor
    selectObject: textGridname$
    Remove
    selectObject: "TextGrid pWord_prono"
    Remove
    selectObject: "TextGrid pWord_ortho"
    Remove
    selectObject: "Table pWord_prono"
    Remove
    selectObject: "Table pWord_ortho"
    Remove
endifor

```