

# 한국어 사전 표제어 발음의 음소 및 음절 빈도

신 지 영<sup>§</sup>

고려대학교 국어국문학과

**배경 및 목적:** 이 연구의 목적은 한국어 사전에 등재된 표제어의 발음형을 대상으로 말소리 관련 빈도를 알아보는 것이다. 사전 표제어의 발음형이 보여주는 말소리 관련 빈도 정보, 즉 음소 및 음절 관련 빈도 정보는 한국어에 대한 가장 기본적인 정보로서, 한국어 화자들의 한국어에 대한 기본적인 언어 지식이라고 할 수 있다. **방법:** 표제어 선정에 사용 빈도의 개념이 반영된 『연세한국어사전』의 전체 표제어 49,553개 중 잘못된 형, 방언형, 자모, 접사(접두사와 접미사) 등과 『표준국어대사전』의 표제어에서 찾을 수 없는 어휘 항목을 제외한 47,401개 표제어를 연구의 대상으로 삼았다. 해당 표제어들을 대상으로 필자가 정한 원칙에 따라 사전의 발음형을 일일이 확인하고 입력하는 방법으로 자료를 완성하였다. 완성된 자료에서 음소 및 음절 빈도를 얻기 위해 SynKDP 1.5.2(일명 껌짜새)가 사용되었다. **결과:** 분석 대상이 된 47,401개 표제어에서 관찰된 음절은 총 122,761개였고, 음소는 총 290,131개였다. 음소 중에서 모음은 122,761(42.3%)개, 자음은 167,370(57.7%)개로, 자음의 수가 약간 많았다. 자음 중에는 초성이 110,122(65.8%)개, 종성이 57,248(34.2%)개로 초성의 빈도가 높았다. 한국어에 존재하는 36개의 음소 중에서 가장 높은 빈도를 보인 음소는 /a/였고, 가장 낮은 빈도를 보인 음소는 /wi/였다. 가장 높은 빈도를 보인 자음은 /k/였고, 이어서 /n/, /ŋ/, /l/, /tɕ/의 순서를 보였다. 가장 높은 출현 빈도를 보인 모음은 /a/였고, 이어서 /i, ʌ, o, u, ε, w, jʌ, wε, wa, jo, ju, ja, wi, wʌ, jε, wi/의 순서를 보였다. 사전의 발음형 전체에서 관찰된 음절의 유형 수는 총 1,283이었다. 이 가운데 가장 출현 빈도가 높았던 음절은 5,413회를 보인 /ha/였다. 이어서 /li, tɕi, i, ki, sa, tɕak/ 등이 그 뒤를 따랐다. /ha/가 고빈도를 보인 이유는 동사 파생 접사 (-하)의 생산성과 유관한 것으로 보인다. 1,283개의 유형 중에서 고빈도 82개 유형, 즉 6.4%에 불과한 음절의 유형이 전체 자료의 50.25%를 차지하는 것으로 나타났다. **논의 및 결론:** 한국어 사전 표제어의 음소 및 음절 관련 빈도 정보는 한국어와 관련된 기초 자료로서 음성학, 음운론은 물론, 언어 병리학, 한국어 교육, 음성 공학 분야에서도 꼭 필요한 기초 자료로 활용될 것이다. 『언어청각장애연구』, 2010;15:94-106.

**핵심어:** 사전 표제어, 음소 빈도, 자음 빈도, 모음 빈도, 음절 빈도

<sup>§</sup> 교신저자

신지영  
고려대학교 국어국문학과 교수  
서울특별시 성북구 안암동  
e-mail: shinjy@korea.ac.kr  
tel.: 02-3290-1973

## I. 서론

이 연구의 목적은 한국어 사전에 등재된 표제어의 발음형을 대상으로 말소리 관련 빈도를 알아보는 것이다. 사전 표제어의 발음형이 보여주는 말소리 관련 빈도 정보, 즉 음소 및 음절 관련 빈도 정보는 한국어에 대한 가장 기본적인 정보로서, 한국어 화자들의 한국어에 대한 기본적인 언어 지식을 정리하는 일이 될 것이다.

한국어 단어에 대한 말소리 관련 지식은 한국어 화자들이 음향 신호로부터 발화 단위를 분절하는 데 사용되는 매우 기본적인 지식으로서, 이와 관련된 지식을 체계화하는 것은 반드시 필요한 기초 연구라고 생각된다.

하지만 선행 연구에서 한국어 사전 표제어 전체를 대상으로 한국어의 말소리 관련 빈도를 전면적이고 다각적으로 알아본 연구는 찾아보기 어렵다. 사전의 어휘 항목을 대상으로 음소 빈도를 연구한 선행 연구로

■ 게재 신청일: 2010년 1월 19일 ■ 최종 수정일: 2010년 2월 22일 ■ 게재 확정일: 2010년 2월 25일

© 2010 한국언어청각임상학회 <http://www.kasa1986.or.kr>

는 김경일(1985), 이상억(1990), 한국방송공사(1993), 한성우(2006) 등을 들 수 있는데, 이 가운데 한국방송공사(1993)을 제외한 나머지 연구들은 모두 사전 표제어의 일부를 대상으로 하고 있다. 또한, 전체 표제어를 대상으로 한 한국방송공사(1993)은 연구서가 아니기 때문에 음소 빈도에 대한 평면적이고 초보적인 나열에 머문 한계가 있다고 할 수 있다.

김경일(1985)는 1979년 한글학회에서 간행한 『새 한글 사전』에 수록된 어휘 중에서 명사(대명사 수사 포함), 단음절 동사, 형용사, 접두사 등 37,000여 개의 고유어 어휘를 형태소 분석하여 동음이의인 형태소를 하나로 처리하는 등의 과정을 거쳐 1음절 형태소 599개와 2음절 형태소 2,230개의 발음형, 총 5,059음절을 대상으로 말소리의 출현 빈도를 계산하였다.

이상억(1990) 역시 사전의 표제어를 대상으로 연구를 진행하였다. 이희승 편 『국어대사전』을 대상으로 하였고, 어휘의 선정 방법은 어휘의 원어정보, 품사 등을 고려한 김경일(1985)와는 달리 사전의 각 면의 왼쪽 맨 아래에 위치한 표제어를 무작위로 추출하는 방법이 사용되었다. 이러한 방법으로 선정된 4,370개의 단어, 총 14,053음절을 발음 나는 대로 전사한 후에 빈도를 조사하였다.

그리고 한성우(2006)은 『표준국어대사전』의 고유단일어 6,364개 단어만을 대상으로 연구를 진행하였다. 이 논문은 품사별 음소 빈도를 살펴본 데에 있어서 기존의 연구와 약간의 차별성을 인정할 수 있으나, 분석 대상을 고유어 단일어로만 한정된 것, 그리고 품사 또한 명사, 동사, 형용사로만 그 대상을 한정된 것 등을 한계라고 할 수 있겠다.

한편, 한국방송공사(1993)은 발음 사전을 만드는 과정에서 산출된 음소 및 음절 빈도와 관련된 내용을 부록에 실고 있다. 표제어 65,973개의 70,113 발음 항목이 보여 주는 음소 및 음절 빈도가 정리되어 있는데, 발음 항목이 표제어 수보다 많은 이유는 복수 발음을 인정한 단어가 4,139개 포함되어 있기 때문이다. 장단 음소를 포함한 총 42개 음소에 대한 통계가 제시되어 있으며, 이와는 별도로 어말과 어두의 음소

빈도가 함께 제공되어 있다. 또한 전체 자료에서 관찰된 1,453개 음절 종류에 대한 빈도를 제시하였다<sup>1)</sup>. 하지만 이 자료는 발음 사전의 부록에 실린 것이어서, 음소 빈도에 대한 평면적이고 초보적인 나열에 그친 것이 한계라고 할 수 있다.

지금까지 살펴본 것과 같이 선행 연구가 몇 개 존재하기는 하지만, 한국어 단어의 말소리 관련 빈도에 대한 전면적이고 치밀하며 체계적인 기초 연구가 필요한 실정이다. 이에 본 연구에서는 사전의 표제어 선정에 사용 빈도의 개념이 반영된 『연세한국어사전』의 표제어를 대상으로 한국어 말소리의 빈도와 관련된 정보를 체계적으로 추출해 보고자 한다. 이와 같이 한국어의 말소리 관련 빈도 정보를 체계적으로 추출해 정리하는 것은 한국어의 특성을 밝히는 데 기여할 수 있을 뿐만 아니라, 한국어 응용 분야에 기초 자료를 제공한다는 점에서 의의를 가질 수 있을 것이다.

## II. 연구 방법

### 1. 자료

분석 대상이 된 사전은 『연세한국어사전』이었다. 이 사전을 분석의 대상으로 삼은 이유는, 이 사전의 표제어 선정에 사용 빈도의 개념이 반영되었기 때문이다. 이 사전의 표제어들은 기본적으로 일정 정도 이상의 사용 빈도를 보이는 어휘 항목들이기 때문에 실제로 사용된 예를 찾을 수 있는 단어들이<sup>2)</sup> 『연세한국어사전』의 표제어 수는 총 49,553개였는데, 이 가운데 잘못된 형, 방언형, 자모, 접사(접두사와 접미사) 등과 『표준국어대사전』의 표제어에서 찾을 수 없는 어휘 항목들은 분석의 대상에서 제외하는 것을 원칙으로 하였다<sup>3)</sup>. 이렇게 선정된 분석 대상 자료의 수는 총 47,401개였다.

<표 - 1>은 분석 대상 자료를 내용어와 기능어<sup>4)</sup>로 나눈 후, 내용어를 원어정보별로 나누어 정리한

1) 부록에 기술된 바에 따르면, 관찰된 음절의 총 종류 수가 1,453개로 되어 있으나, 전체 음절 누적빈도표에는 1,454개 음절에 대한 빈도가 제시되어 있다.

2) 『연세한국어사전』은 기본적으로 4300만 어절의 말뭉치에서 빈도 14 이상을 보이는 항목을 표제어로 삼는 것을 원칙으로 하였다.

3) 단, '-적, -하, -되-'에 의해 파생된 파생어의 경우는 표준 사전의 등재 여부와는 무관하게 모두 분석의 대상에 포함하였다.

4) 한국어에서 기능어는 조사의 경우는 단어이지만 어미의 경우는 단어가 아니다. 따라서 필자가 말하는 '기능어'란 '어'는 '단

것이다<sup>5)</sup>. 내용어는 총 46,529개, 기능어는 총 872개였다. 또한, 내용어를 원어정보별로 나누어 보면, 고유어가 11,540개, 한자어가 24,716개, 외래어가 1,302개, 그리고 기타 복합형이 8,971개를 나타내, 각각 24.34%, 52.14%, 2.75%, 18.92%의 비율을 보였다. 한자어가 고유어의 2배 이상의 빈도를 보였고, 외래어의 비율이 매우 낮았다. 기능어는 모두 고유어였으며, 기능어 중에는 조사가 135개, 어미가 603개, 그리고 기타 축약형<sup>6)</sup>이 134개였다.

<표 - 1> 분석 대상 단어의 원어정보별 구성

		원어정보	빈도	(%)
내용어		고유어	11,540	(24.34)
		한자어	24,716	(52.14)
		외래어	1,302	(2.75)
		복합	8,971	(18.92)
		<b>소계</b>	<b>46,529</b>	<b>(98.16)</b>
기능어	고유어	조사	135	(0.28)
		어미	603	(1.27)
		축약형	134	(0.28)
		<b>소계</b>	<b>872</b>	<b>(1.84)</b>
<b>총계</b>			<b>47,401</b>	<b>(100.00)</b>

한편, 분석 대상이 된 내용어의 품사별 구성은 <표 - 2>에 제시된 것과 같다. 가장 높은 비율을 차지한 품사는 66.53%를 보인 명사였고, 가장 낮은 비율을 차지한 품사는 0.16%를 보인 수사였다. 동사가 형용사에 비해 약 3배 이상의 빈도를 보인다는 것과 관형사의 목록이 매우 빈약한 것이 특징적이다.

<표 - 2>는 또한 각 품사에 속한 단어의 원어정보별 구성을 보이고 있다. 표에서 보듯이 고유어는 모든 품사의 단어를 구성하는 반면, 한자어는 동사와 형용사, 즉 용언류를 제외한 모든 품사의 단어를, 외래어는 명사만을 구성하는 특징을 보였다. 그리고 복합형의 대부분은 명사와 동사 및 형용사, 그리고 부사를 구성하는 것으로 나타났다. 복합형 동사와 형용

사는 한자어와 외래어 명사의 어근에 동사 파생 접사, -하-, -되-, 시키-, -내-, -롭-, -답-, 등의 접사가 붙어서 파생된 것이 대부분이었고, 부사 또한 한자어에 부사파생 접사 '-히, -이' 등이 붙어서 파생된 것이 대부분이었다.

<표 - 2> 내용어의 품사별, 원어정보별 구성

	고유어	한자어	외래어	복합	계	
명사	4,571	23,655	1,302	1,428	30,956	66.53%
대명사	73	19	-	4	96	0.21%
수사	42	34	-	-	76	0.16%
동사	3,501	-	-	5,702	9,203	19.78%
형용사	1,390	-	-	1,459	2,849	6.12%
관형사	59	858	-	1	918	1.97%
부사	1,766	148	-	373	2,287	4.92%
감탄사	138	2	-	4	144	0.31%
계	11,540	24,716	1,302	8,971	46,529	100.00%

## 2. 분석 방법

음소 빈도를 살펴볼 때 이중모음을 두 음소의 결합으로 보고 분석하는 것보다는 하나의 단위로 보고 분석하는 것이 여러 가지 이점을 준다. 특히 두 개의 음소로 쪼개어 보는 경우에는 이중모음들 사이에 어떠한 빈도 차이를 보이는가를 알기 어렵다는 단점이 있다. 이중모음을 한 단위로 보는 경우에는 이중모음들 사이의 빈도 차이는 물론, 결합된 요소들을 쉽게 쪼개어 사후 분석을 할 수 있기 때문에 장점이 더 많다고 할 수 있다. 따라서 이 논문에서는 이중모음을 하나의 단위로 분석하여 결과를 제시하였다. 따라서 음소 빈도의 논의에 등장하는 전체 음소의 개수는 자음 19개, 단모음 7개, 이중모음 10개를 포함한 총 36개가 될 것이다.

한편, 표제어의 발음형은 다음 (1)에 보인 기준을 가지고 설정하였다.

어가 아니라 사전의 표제어를 지칭하는 것이다.

5) 품사 설정과 원어정보 관련 내용이 「표준국어대사전」과 일치하지 않는 경우에는 「표준국어대사전」을 따르는 것으로 하였다.

6) 기타 축약형이란 「연세한국어사전」에 '준꼴'로 표시된 부류를 의미하는 것으로, 예를 들어 '냐고 한, 려고 하다가'의 축약형인 '냐, 려다가' 등이 이에 해당한다.

(1) 발음형 설정 기준

1. /h/ 탈락<sup>7)</sup>은 인정하지 않았다.  
예) 기하학[기하학], 심화[심화]
2. 조음 위치 동화 현상<sup>8)</sup>은 반영하지 않았다.  
예) 신문[신문], 신발[신발]
3. 동일 위치 장애음 탈락<sup>9)</sup>은 적용한다.  
예) 학교[하교]
4. 장애음 뒤 /je/의 /j/ 탈락은 적용한다.  
예) 계단[kjetan]
5. 자립 형태소 2음절 이하의 /-니/는 [i]를, 조사  
의 /-니/는 [ɛ]를 발음형으로 한다.  
예) 주의[주이], 의[에]
6. 용어 어간의 발음형을 분석 대상으로 삼는다.  
예) 놓다[논], 예쁘다[예쁘], 아름답다[아름답]
7. 어미의 음소 시작형은 분석 대상에 포함시키되, 종  
성으로 취급하고, 음절 수와 음절형의 비율을 셀  
때는 음소형을 제외한다.  
예) ㄴ지: 음소 수는 3개, 자음 2(초성 1, 종성 1),  
모음 1. 음절 수 1음절
8. 이 이외의 것은 「표준국어대사전」의 발음형을 따  
른다.

이러한 분석 기준을 가지고 사전의 발음형을 일일이 확인하고 입력하는 방법으로 자료를 완성하였다. 완성된 자료에서 음소 및 음절 빈도를 얻기 위해 SynKDP 1.5.2<sup>10)</sup>(일명 깜짝새)가 사용되었다.

- 7) 비격식 발화, 빠른 발화 등에서 공명음 사이에 위치한 /h/는 자주 탈락한다.
- 8) 조음 위치 동화 현상이란 치경음이 후행하는 연구개음이나 양순음의 영향으로 조음 위치가 변화하는 현상을 말한다. 치경음의 변화보다는 적게 나타나지만, 양순음이 후행하는 연구개음에 의하여 조음 위치가 연구개음으로 변화하는 경우도 있다.
- 9) 장애음과 장애음의 연쇄에서 두 장애음의 조음 위치가 같은 경우, 선행하는 장애음이 탈락하는 현상을 말한다. /학+교/는 장애음 뒤에 위치한 평장애음이 경음으로 변하여 ‘학교’가 된 다음, 두 장애음의 조음 위치가 동일하기 때문에 선행하는 장애음이 탈락하여 [하교]가 된 것으로 분석할 것이다. 표준발음법에는 이 둘이 수의적인 실현을 보인다고 하였지만, 실제로 한국어 화자들에게 [하교]와 [하교]의 대립은 무의미하며, 특히, 마찰음이 후행하는 경우 동일 위치 장애음 탈락이 필수적으로 적용된다는 사실은 이 규칙의 확대 가능성을 보여 준다고 할 수 있다. /웃+소/는 ‘운소 → 운쏘 → 우쏘’가 되는데, 이 과정에서 [운쏘]는 한국어 화자들에게서 관찰 불가능한 표면형이다. 하지만 표준 발음법에는 /웃+소/의 표면형으로 [운쏘]와 [우쏘]가 모두 가능한 것으로 설명하고 있다. 이에 본 연구에서는 동일 위치 장애음 탈락이 적용된 형을 발음형으로 삼았다.
- 10) SynKDP 1.5.2는 공개된 프로그램으로, 한글로 입력된 자료를 대상으로 어절 빈도, 음절 빈도, 자소 빈도 등을 계산해 준다. 본 연구에서는 발음형을 한글로 입력한 후에 프로그램을 돌렸기 때문에 자소 빈도가 곧 음소 빈도였다.
- 11) 표에 제시된 음소 기호는 신지영(2000)을 따랐다.

### Ⅲ. 연구 결과 및 논의

분석 대상이 된 47,401개 표제어에서 관찰된 음절은 총 122,761개였고, 음소는 총 290,131개였다. 음소 중에서 모음은 122,761(42.3%)개, 자음은 167,370(57.7%)개로, 자음의 수가 약간 많았다. 자음 중에는 초성이 110,122(65.8%)개, 종성이 57,248(34.2%)개로 초성의 빈도가 높았다.

<표-3> 사전 전체 자료의 음절 및 음소 수

표제어 수		47,401	
음절 수		122,761	
음소 수	모음	122,761(42.3%)	
	자음	초성	110,122(65.8%)
		종성	57,248(34.2%)
		소계	167,370(57.7%)
총계		290,131	

#### 1. 음소 빈도

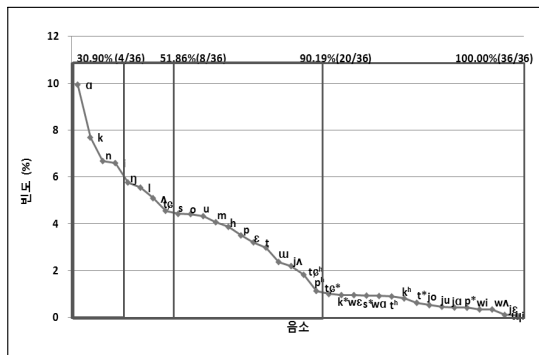
##### 가. 음소 전체 빈도

<표-4>와 <그림-1>은 음소 전체 출현 빈도를 보인 것이다<sup>11)</sup>. 한국어에 존재하는 36개의 음소 중에서 가장 높은 빈도를 보인 음소는 /a/였고, 가장 낮은 빈도를 보인 음소는 /wi/였다. 표와 그림에서 보듯이 어휘 항목의 발음형에 나타난 음소 빈도는 매우 불균형한 것으로 나타났다. 이러한 불균형성은 <그림-1>을 통해 훨씬 쉽게 확인할 수 있다. 고빈도 상위 4개 음소의 출현 빈도 합은 30.90%나 되었

고, 고빈도 상위 8개의 음소의 출현 빈도 합이 50% 이상, 더 정확히는 51.86%를 차지하는 것으로 나타났다. 또한 고빈도 상위 20개 음소의 출현 빈도 합이 90.19%를 보이는 것으로 나타났다. 따라서 나머지 하위 16개의 음소의 출현 빈도 합은 9.81%에 불과하여, 10%에도 미치지 못하는 것으로 나타났다.

<표 - 4> 음소 전체 빈도

순위	음소	출현 빈도	출현 빈도 (%)	누적 빈도 (%)	순위	음소	출현 빈도	출현 빈도 (%)	누적 빈도 (%)
1	a	28846	9.94	9.94	19	te <sup>h</sup>	5291	1.82	89.06
2	k	22311	7.69	17.63	20	p <sup>h</sup>	3289	1.13	90.19
3	n	19384	6.68	24.31	21	te <sup>*</sup>	2909	1.00	91.19
4	i	19134	6.59	30.90	22	k <sup>*</sup>	2778	0.96	92.15
5	ŋ	16727	5.76	36.67	23	wε	2772	0.96	93.11
6	l	16092	5.55	42.21	24	s <sup>*</sup>	2714	0.94	94.04
7	ʌ	14784	5.09	47.31	25	wa	2662	0.92	94.96
8	te	13218	4.56	51.86	26	t <sup>h</sup>	2608	0.90	95.86
9	s	12831	4.42	56.29	27	k <sup>h</sup>	2358	0.81	96.67
10	o	12817	4.42	60.70	28	t <sup>*</sup>	1819	0.63	97.30
11	u	12558	4.33	65.03	29	jo	1529	0.53	97.82
12	m	11804	4.07	69.10	30	ju	1329	0.46	98.28
13	h	11228	3.87	72.97	31	ja	1232	0.42	98.71
14	p	10169	3.50	76.47	32	p <sup>*</sup>	1213	0.42	99.12
15	ε	9309	3.21	79.68	33	wi	998	0.34	99.47
16	t	8660	2.98	82.67	34	wʌ	997	0.34	99.81
17	w	6882	2.37	85.04	35	je	314	0.11	99.92
18	ja	6373	2.20	87.23	36	wj	230	0.08	100.00



<그림 - 1> 음소 전체 빈도(그림 위의 % 표시는 누적 빈도를, 괄호 안의 숫자는 36개 음소 중 해당 누적 빈도를 보인 음소의 수를 나타낸 것이다)

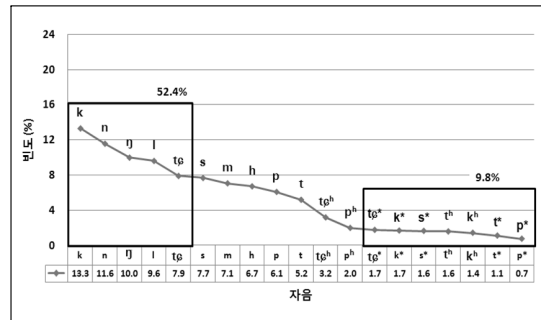
## 나. 자음 빈도

36개 음소에 대한 전체적인 빈도 경향을 살펴보는 것도 좋지만, 자음과 모음의 빈도를 나누어 살펴보는 것도 필요할 것이다. 또한 자음 빈도는 자음의 전체 빈도는 물론, 자음을 음절 내 위치에 따라 초성과 종성으로 나누어 각각의 위치에서 보이는 빈도도 알아 보았다. 그럼 먼저 자음의 전체 빈도를 알아보도록 하자.

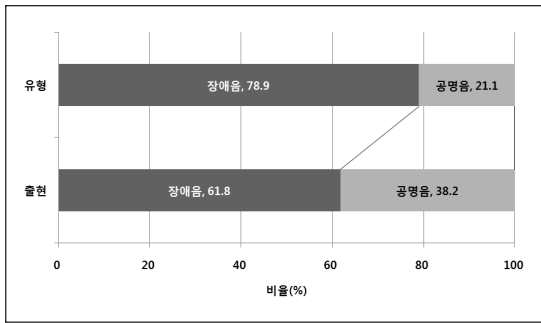
### (1) 전체 빈도

<그림 - 2>는 자음의 전체 빈도를 고빈도순으로 보인 것이다. 가장 높은 빈도를 보인 자음은 /k/였고, 이어서 /n/, /ŋ/, /l/, /te/의 순서를 보였다. 그림 안에 상자로 표시한 것과 같이 고빈도 5개 자음의 출현 빈도 합은 52.4%를 보인 반면, 저빈도 7개 자음의 출현 빈도 합은 9.8%를 보였다. 자음 전체 빈도에서 연구개 폐쇄음 /k/ 과 연구개 비음 /ŋ/이 높은 비율을 나타낸다는 점은 주목할 만하다. /k/와 /ŋ/의 빈도는 각각 13.3%와 10.0%를 보였다. 특히, 연구개 비음이 음절의 종성에만 나타날 수 있다는 점을 고려한다면, 자음 전체 빈도에서 3위를 보인 것은 주목할 만하다.

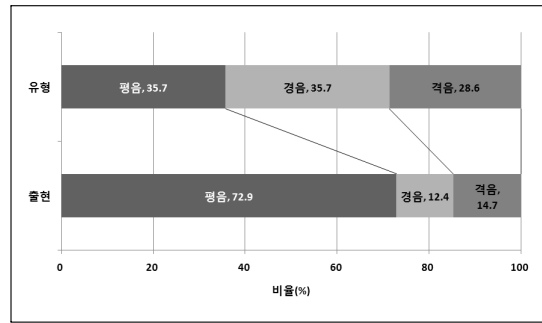
다음은 자음의 하위 분류별 출현 빈도 특성을 알아보았다. 유형 빈도를 고려해 볼 때, 공명음이 장애음에 비하여 높은 빈도를 보였고(<그림 - 3>). 또한, 장애음 중에는 평음이 압도적인 빈도를 보였다(<그림 - 4>). 조음 방법별로 보면, <그림 - 5>에 보인 바와 같이 폐쇄음의 출현 빈도가 33.0%를 보여 가장 높은 빈도를 보이기는 했지만, 유형 빈도를 고려할 때 가장 높은 출현 빈도를 보인 것은 비음이었다. 조음 위치별로는 유형 빈도를 고려할 때, <그림 - 6>에 보인 것과 같이 양순음의 출현 빈도가 낮고, 연구개음의 출현 빈도가 높았다.



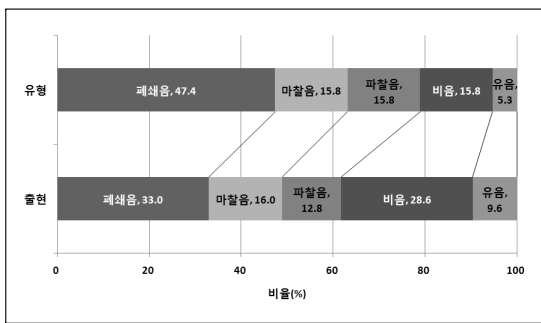
<그림 - 2> 자음 전체 빈도



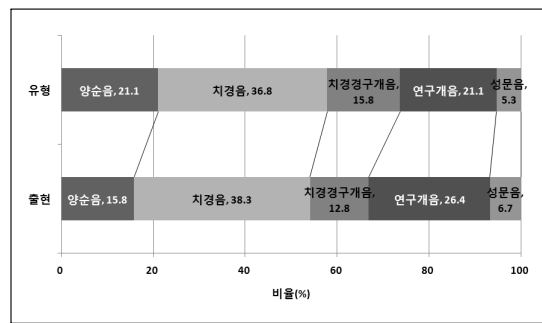
<그림 -3> 장애음 대 공명음의 출현 빈도와 유형 빈도 (자음 전체)



<그림 -4> 발성 유형별 출현 빈도와 유형 빈도 (자음 전체)



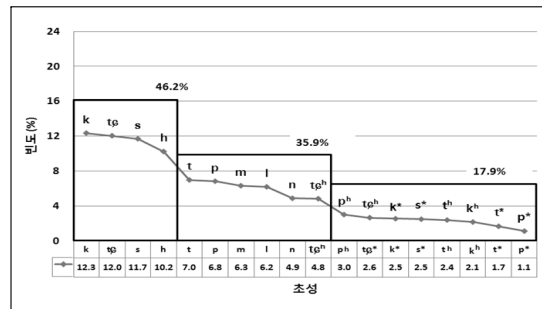
<그림 -5> 조음 방법별 출현 빈도와 유형 빈도 (자음 전체)



<그림 -6> 조음 위치별 출현 빈도와 유형 빈도 (자음 전체)

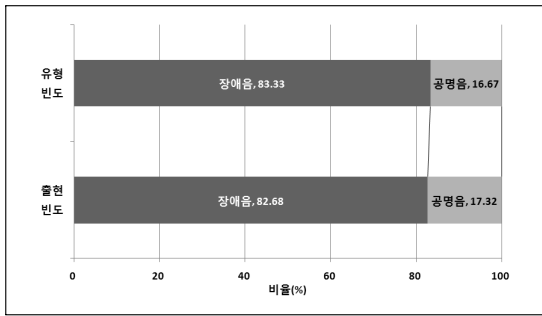
(2) 초성 빈도

다음은 자음의 음절 내 위치에 따른 출현 빈도의 경향을 살펴보기 위하여, 우선 음절의 초성에서 관찰되는 18개 자음의 출현 빈도를 알아보았다. <그림 -7>은 초성에 쓰인 자음의 출현 빈도를 보인 것이다. 초성에서 가장 높은 출현 빈도를 보인 자음은 /k/였고, 가장 낮은 출현 빈도를 보인 자음은 /p\*/였다. 출현 빈도가 급격히 차이를 보이는 지점을 주목하여 고빈도, 중빈도, 저빈도를 보이는 초성의 유형을 나누어 보았다. 고빈도 유형에 속하는 초성 자음은 /k, t̥, s, h/였는데, 최고 빈도를 보인 /k/를 제외하면 마찰음과 파찰음이 고빈도군을 이루었다. 중빈도 유형에 속하는 초성 자음은 /t, p, m, l, n, t̥ʰ/로, /t̥ʰ/를 제외하고는 모두 평 폐쇄음과 공명음이었다. /pʰ, t̥ʰ, k\*, s\*, tʰ, kʰ, t\*, p\*/는 저빈도 유형에 속했는데, 이 자음들은 공통적으로 격음 혹은 경음이었다.

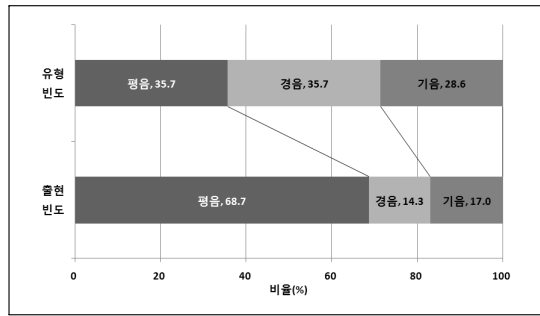


<그림 -7> 초성 빈도

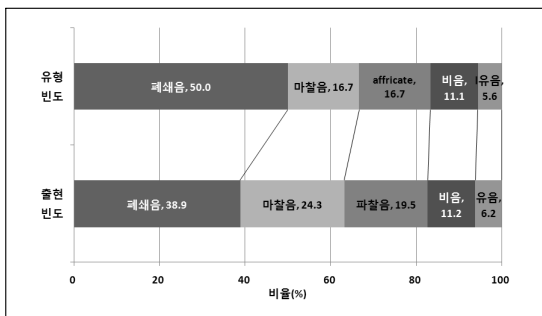
초성의 하위 분류별 출현 빈도 특성도 알아보았다. <그림 -8>에 보인 것과 같이 초성에는 공명음보다는 장애음이 압도적인 비율로 사용되었다. 초성에 쓰인 장애음의 비율은 82.7%로, 공명음의 비율 17.3%에 비하여 훨씬 높게 나타났다. 자음의 전체 빈도에서 장애음의 출현 비율이 61.8%였다는 것을 고려해 볼 때, 대체로 초성에서 장애음의 비율이 높고, 중성에서는 공명음의 비율이 높게 나타난다는 것을 알 수 있다. 한편, 유형 빈도와 출현 빈도가 크게 다르지 않은 것도 특징적이다. 발성 유형 중에는 평음이 역시



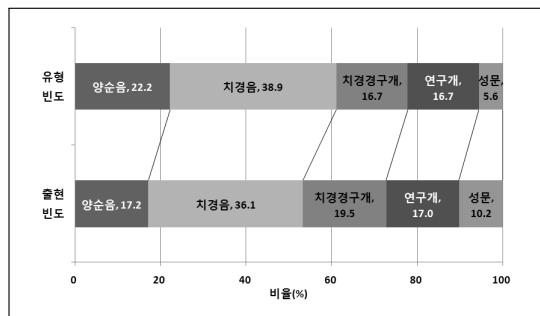
<그림 - 8> 장애음 대 공명음의 출현 빈도와 유형 빈도 (초성)



<그림 - 9> 발생 유형별 출현 빈도와 유형 빈도 (초성)



<그림 - 10> 조음 방법별 출현 빈도와 유형 빈도(초성)



<그림 - 11> 조음 위치별 출현 빈도와 유형 빈도(초성)

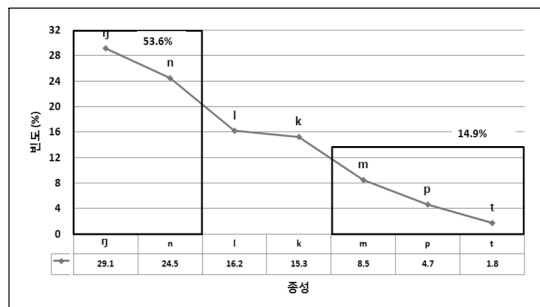
압도적인 비율을 보였고, 격음의 비율이 경음에 비하여 약간 높게 나타났다(<그림 - 9>).

조음 방법에 있어서는 폐쇄음이 38.9%로 가장 높은 비율을 보였지만, 유형 빈도가 50%인 점을 고려할 때, 높은 출현 빈도를 보인다고 하기는 어렵다. 반면에 유형 빈도가 16.7%씩인 마찰음의 출현 빈도가 각각 24.3%를 보여서 상대적으로 높은 출현 빈도를 보였다. 비음과 유음 등 공명음의 출현 빈도는 초성에서는 높지 않았다(<그림 - 10>). 조음 위치에 있어서 가장 높은 출현 빈도를 보인 것은 치경음이었지만 유형 빈도를 고려할 때 상대적으로 높은 출현 빈도를 보인 것은 치경경구개음과 후두음이었다. 한편, 유형 빈도에 비하여 낮은 출현 빈도를 보인 것은 양순음이었다(<그림 - 11>).

(3) 종성 빈도

다음은 음절의 종성 위치에서 관찰된 자음의 출현 빈도를 살펴보았다. 한국어의 종성에는 /m, n, ŋ, l, p, t, k/ 등 7개의 자음만이 출현할 수 있다. <그림 - 12>에 보인 바와 같이, 종성에서 가장 높은 빈도를 보인 자음은 29.1%의 출현 빈도를 보인 연구개 비음

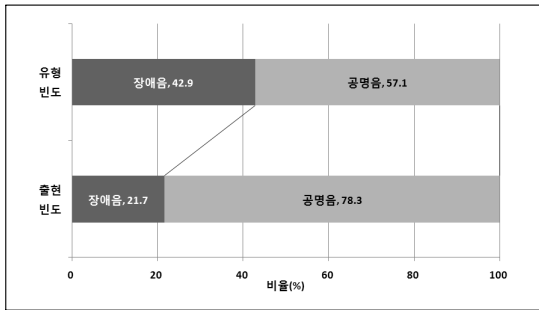
/ŋ/이었고, 가장 낮은 빈도를 보인 자음은 1.8%의 출현 빈도를 보인 치경 평 폐쇄음 /t/였다. 종성에서 높은 출현 빈도를 보인 자음은 /ŋ/과 /n/였는데, 이 두 자음의 출현 빈도 합은 53.6%로 과반을 차지했다. 반면에 가장 적은 빈도를 보인 세 자음 /m, p, t/의 출현 빈도 합은 14.9%에 불과한 것으로 나타났다.



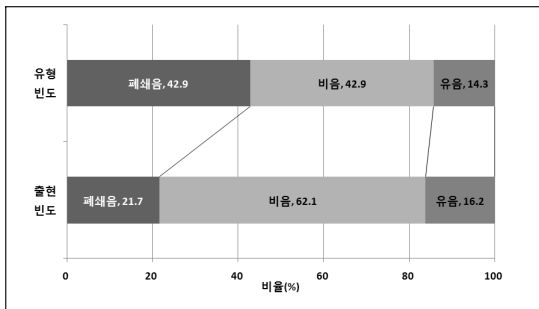
<그림 - 12> 종성 빈도

다음은 자음의 유형별로 출현 빈도를 알아보았다. <그림 - 13> 보인 바와 같이 종성에서 공명음의 선호도가 높다는 것을 알 수 있다. 하지만 /k/와 /m/는 이러한 일반적인 경향에서 벗어나 있음을 알 수 있

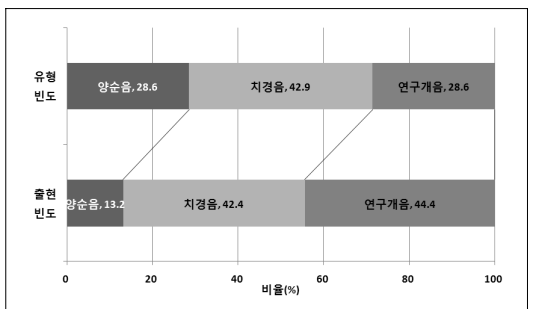
다. 장애음 중에서 /k/만은 15.3%의 출현 빈도를 보여서 유일하게 저빈도 유형에 속하지 않았다. 또한, 공명음 중에서 /m/만은 8.5%의 출현 빈도를 보여서 역시 유일하게 저빈도 유형에 속했다. 조음 방법에 있어서는 비음이 높은 비율을 보였고, 상대적으로 폐쇄음의 비율이 높지 않았다(<그림 - 14>). 조음 위치에 있어서는 유형 빈도가 28.6%인 연구개음이 44.4%의 출현 빈도를 보여서, 유형 빈도가 42.9%인 치경음의 출현 빈도 42.4%와 거의 비슷했다(<그림 - 15>). 반면에, 28.6%의 유형 빈도를 보인 양순음의 출현 빈도는 13.2%에 불과한 것으로 나타나서, 양순음은 초성과 종성 모두에서 낮은 빈도가 관찰되었다.



<그림 - 13> 장애음 대 공명음의 출현 빈도와 유형 빈도 (중성)



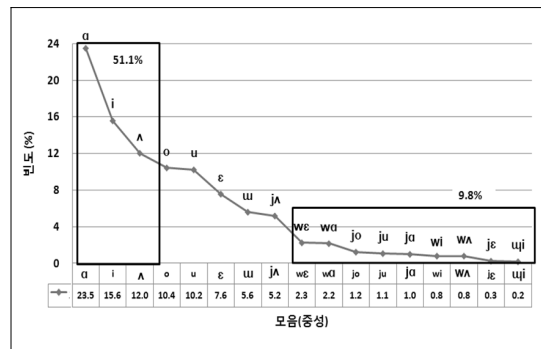
<그림 - 14> 조음 방법별 출현 빈도와 유형 빈도(중성)



<그림 - 15> 조음 위치별 출현 빈도와 유형 빈도(중성)

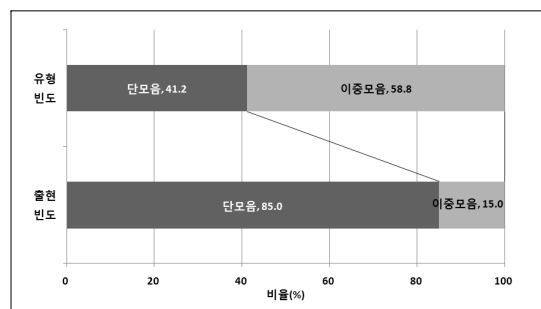
### 다. 모음 빈도

자음에 이어 모음의 출현 빈도를 알아보았다. <그림 - 16>은 모음의 출현 빈도를 고빈도 순서로 보인 것이다. 그림에서 보듯이 가장 높은 출현 빈도를 보인 모음은 23.5%를 보인 /a/였고, 이어서 /i, ʌ, o, u, ε, w, jʌ, wε, wa, jo, ju, ja, wi, wʌ, jε, wji/의 순서를 보였다. 흥미로운 것은 단모음 7개가 상위 빈도 7위까지를 차지하고 있다는 점이다. <그림 - 16>에 표시한 것처럼, 고빈도 상위 3개 모음의 출현 빈도 합이 51.1%로 과반을 차지했지만, 저빈도 하위 9개 모음의 출현 빈도 합은 9.8%를 보여 10%에도 미치지 못하는 것으로 나타났다.



<그림 - 16> 모음(중성) 빈도

<그림 - 17>에 보인 바와 같이, 단모음 7개의 출현 빈도 합은 85.0%로 10개의 이중모음의 출현 빈도 합인 15.0%에 비하여 훨씬 높았다. 한 가지 특기할 것은, 자음에 비하여 모음의 음소 간 출현 빈도 차이가 더 크게 나타난다는 것이다. 단모음 중에는 /w/의 비율이 가장 낮았고, 이중모음 중에는 /jʌ/의 비율이 가장 높았다.



<그림 - 17> 단모음과 이중모음의 출현 빈도와 유형 빈도



다음은 단모음과 이중모음의 유형별 빈도를 알아보았다. <그림 - 18>~<그림 - 21>은 그 결과를 정리한 것이다. 유형 빈도가 낮은 원순 모음과 전설 모음은 출현 빈도도 낮았지만, 유형 빈도가 낮은 저모음은 출현 빈도에서는 높은 비율을 나타냈다. 한편, 이중모음의 활음 유형별 빈도는 <그림 - 21>에 보인 것과 같이 /j/계 이중모음이 58.5%의 출현 빈도를 나타내서, 40.3%의 출현 빈도를 보인 /w/계 이중모음보다 높았다. 그리고 /w/계 이중모음의 출현 빈도는 1.2%에 불과했다. 유형 빈도를 고려할 때, /j/계 이중모음의 출현 빈도가 /w/계에 비하여 약간 높았다.

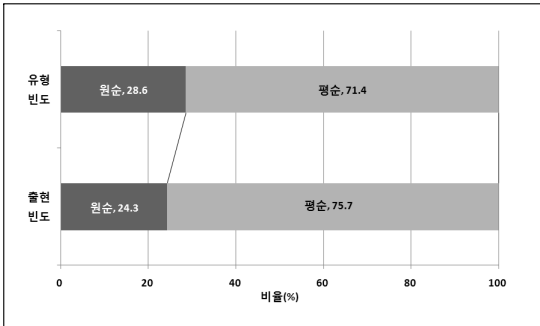
2. 음절 빈도

다음은 음절 빈도를 알아보았다. 사전의 발음형 전체에서 관찰된 음절의 유형 수는 총 1,283이었다. 이 가운데 가장 자주 사용된 음절은 출현 빈도 5,413 회를 보인 /ha/였다. 이어서 /li, tɕi, i, ki, sa, tɕak/ 등이 그 뒤를 따랐다. /ha/가 고빈도를 보인 이유는

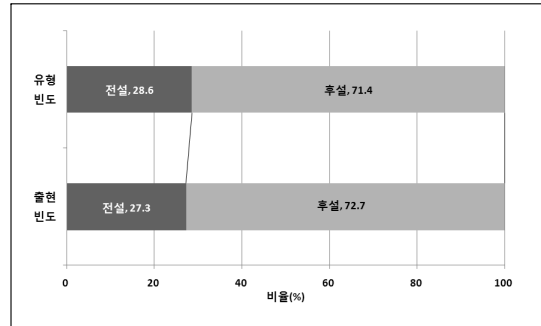
동사 파생 접사 {-하}의 생산성과 관련한 것으로 보인다. <표 - 5>는 누적 빈도 50% 이상을 보인 상위 82개 음절형의 출현 빈도와 누적 빈도를 출현 빈도 순서로 보인 것이다. 1,283개의 유형 중에서 고빈도 82개 유형, 즉 6.4%에 불과한 음절의 유형이 전체 자료의 50.25%를 차지하는 것으로 나타났다.

고빈도 음절형들은 대체로 V형과 CV형이었다. V형은 관찰된 7개 유형 중에서 5개가, CV형은 관찰된 127개 유형 중에서 45개가 고빈도 유형에 속했다. 한편, 가장 많은 유형 빈도가 관찰된 CVC형은 관찰된 778개 유형 중에서 단지 21개 유형만이 고빈도 유형에 속하는 것으로 나타났다.

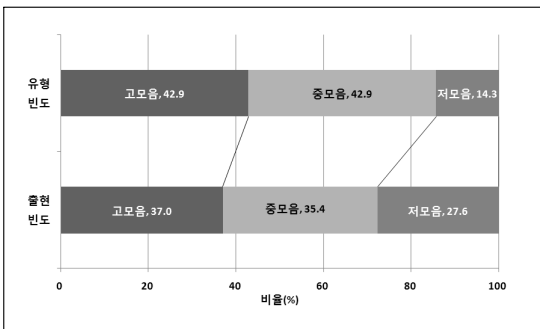
다음은 음절 구조에 따른 분포를 살펴보았다. <표 - 6>과 <그림 - 22>에 보인 것과 같이, 한국어에 존재하는 8가지 음절 유형 중에서 가장 높은 출현 빈도를 보인 것은 CV형으로 42.8%의 출현 빈도를 보였다. 반면에 가장 낮은 출현 빈도를 보인 음절형은 GV형으로, 2.0%의 출현 빈도를 보였다. CV형에 이어 높은 출현 빈도를 보인 음절형은 CVC형이



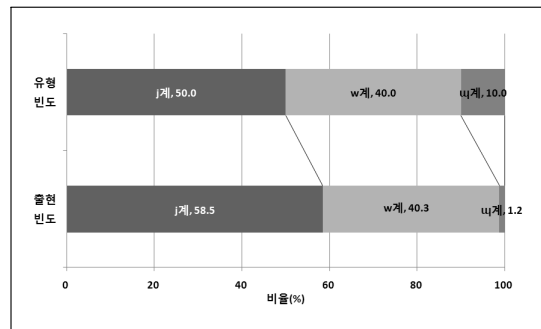
<그림 - 18> 원순 대 평균 모음의 출현 빈도와 유형 빈도(%)



<그림 - 19> 전설 대 후설 모음의 출현 빈도와 유형 빈도(%)



<그림 - 20> 혀의 높이별 모음의 출현 빈도와 유형 빈도(%)



<그림 - 21> 활음의 유형별 이중모음의 출현 빈도와 유형 빈도(%)

<표-5> 누적 빈도 50% 이상을 차지하는 고빈도 음절형

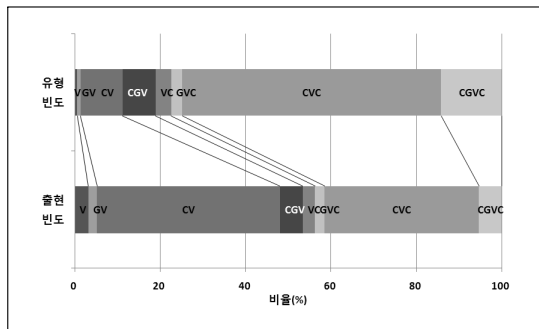
순위	음절	출현 빈도	출현 빈도 (%)	누적 빈도 (%)	순위	음절	출현 빈도	출현 빈도 (%)	누적 빈도 (%)
1	ha	5413	4.41	4.41	42	koŋ	550	0.45	35.51
2	li	1805	1.47	5.88	43	mi	545	0.44	35.96
3	tei	1643	1.34	7.22	44	su	545	0.44	36.40
4	i	1636	1.33	8.55	45	po	542	0.44	36.84
5	ki	1626	1.32	9.87	46	ne	541	0.44	37.28
6	sa	1389	1.13	11.01	47	ma	531	0.43	37.71
7	tɛʌk	1373	1.12	12.12	48	sin	519	0.42	38.14
8	twe	1210	0.99	13.11	49	sɛ	517	0.42	38.56
9	tɛa	1197	0.98	14.09	50	ʌ	513	0.42	38.98
10	si	1176	0.96	15.04	51	mul	500	0.41	39.38
11	su	1154	0.94	15.98	52	a	495	0.40	39.79
12	tɛ	1095	0.89	16.88	53	la	486	0.40	40.18
13	ka	1064	0.87	17.74	54	mun	483	0.39	40.58
14	tɛʌŋ	1011	0.82	18.57	55	kʰi	480	0.39	40.97
15	pu	1006	0.82	19.38	56	ni	479	0.39	41.36
16	tɛɛ	952	0.78	20.16	57	sʌn	467	0.38	41.74
17	kʰa	943	0.77	20.93	58	mo	464	0.38	42.12
18	ko	904	0.74	21.66	59	tɛʰɛ	461	0.38	42.49
19	saŋ	882	0.72	22.38	60	il	459	0.37	42.87
20	kɛ	879	0.72	23.10	61	wʌn	450	0.37	43.23
21	tɛaŋ	861	0.70	23.80	62	sʰʌŋ	449	0.37	43.60
22	hwa	850	0.69	24.49	63	kjʌŋ	446	0.36	43.96
23	tɛu	847	0.69	25.18	64	mjʌŋ	438	0.36	44.32
24	to	836	0.68	25.86	65	ku	437	0.36	44.67
25	mu	824	0.67	26.54	66	u	434	0.35	45.03
26	tɛʌn	815	0.66	27.20	67	sil	433	0.35	45.38
27	pi	784	0.64	27.84	68	jʌn	428	0.35	45.73
28	ku	780	0.64	28.47	69	tɛʌ	419	0.34	46.07
29	so	750	0.61	29.08	70	tɛuŋ	419	0.34	46.41
30	tɛʰi	711	0.58	29.66	71	pal	415	0.34	46.75
31	ju	657	0.54	30.20	72	paŋ	414	0.34	47.09
32	kʌ	654	0.53	30.73	73	ʌ	407	0.33	47.42
33	tɛo	636	0.52	31.25	74	le	406	0.33	47.75
34	toŋ	634	0.52	31.77	75	pul	399	0.33	48.07
35	tɛʰk	591	0.48	32.25	76	tɛʰa	395	0.32	48.40
36	pʰa	587	0.48	32.72	77	kjo	392	0.32	48.71
37	sʌ	586	0.48	33.20	78	kwan	389	0.32	49.03
38	na	580	0.47	33.67	79	tan	387	0.32	49.35
39	sʌŋ	580	0.47	34.15	80	tʰa	387	0.32	49.66
40	ta	563	0.46	34.61	81	san	386	0.31	49.98
41	in	563	0.46	35.06	82	o	381	0.31	50.29

었다. CVC형이 출현 빈도는 36.1%였다. 유형 빈도가 가장 높은 음절형은 CVC형이었는데, 유형 빈도가 60.6%나 되었다. 하지만 출현 빈도를 36.1%를 보여서 유형 빈도가 9.9%에 불과한 CV형의 출현 빈도 42.8%보다 낮았다. 또한, 활음을 포함한 음절, 즉 이중모음을 포함한 음절은 유형 빈도와 출현 빈도가 모두 낮았다. 단모음의 유형 빈도와 출현 빈도는 각각 74.4%와 85%인 데 반하여, 이중모음의 유형 빈도와 출현 빈도는 각각 25.6%와 15.0%로 나타났다.

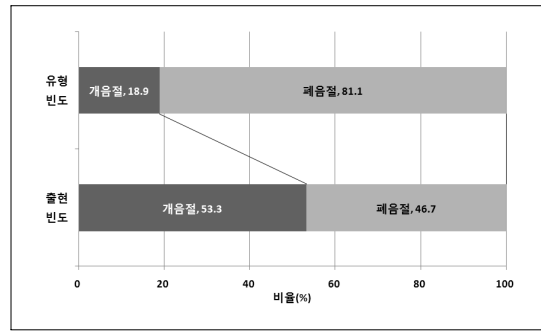
음절 유형별 빈도를 초성이 있는 음절과 없는 음절, 그리고 종성이 있는 음절(폐음절)과 종성이 없는 음절(개음절)로 나누어 살펴보았다. <그림 - 23>에 보인 것과 같이, 초성이 있는 음절은 유형 빈도와 출현 빈도가 모두 높았다. 반면에 <그림 - 24>에 보인 것과 같이 종성이 없는 개음절은 낮은 유형 빈도(18.9%)에도 불구하고 높은 출현 빈도(53.3%)를 보이는 것으로 나타났다.

<표 - 6> 음절의 유형별 출현 빈도와 유형 빈도

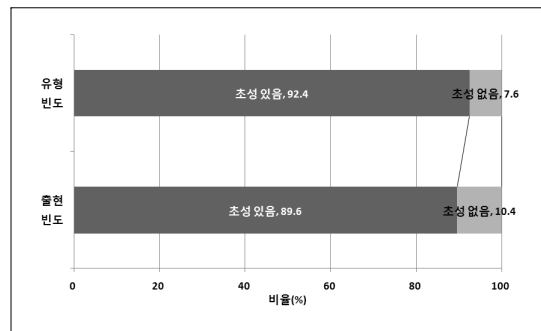
	출현 빈도	(%)	유형 빈도	(%)
V	3,897	3.17	7	0.55
GV	2,511	2.05	10	0.78
CV	52,551	42.81	127	9.9
CGV	6,554	5.34	99	7.72
VC	3,547	2.89	47	3.66
GVC	2,791	2.27	33	2.57
CVC	44,330	36.11	778	60.64
CGVC	6580	5.36	182	14.19
계	122,761	100.00	1,283	100.00



<그림 - 22> 음절의 유형별 출현 빈도와 유형 빈도



<그림 - 23> 개음절과 폐음절의 출현 빈도와 유형 빈도(%)



<그림 - 24> 초성 유무에 따른 음절 유형의 출현 빈도와 유형 빈도(%)

#### IV. 논의 및 결론

지금까지 『연세한국어사전』의 47,401개 표제어의 발음형에서 관찰된 음소 및 음절 관련 빈도를 상세히 알아보았다. 분석 대상이 된 47,401개 표제어에서 관찰된 음절은 총 122,761개였고, 음소는 총 290,131개였다.

음소 중에서 모음은 42.3%, 자음은 57.7%로 나타나, 자음의 비율이 약간 높았다. 자음 중에는 초성이 65.8%, 종성이 34.2%로, 초성의 비율이 높았다. 자음의 전체 빈도 순서는 /k/ > /n/ > /ŋ/ > /l/ > /tɕ/ > /s/ > /m/ > /h/ > /p/ > /t/ > /tɕʰ/ > /pʰ/ > /tɕʰ/ > /kʰ/ > /sʰ/ > /tʰ/ > /kʰ/ > /tʰ/ > /pʰ/와 같았다. 음절 내 위치에 따른 출현 빈도로 다시 나누어 분석하면 초성에서 높은 빈도를 보인 자음은 순서대로 /k/ > /tɕ/ > /s/ > /h/ > /t/ > /p/ > /m/ > /l/ > /n/ > /tɕʰ/ > /pʰ/ > /tɕʰ/ > /kʰ/ > /sʰ/ > /tʰ/ > /kʰ/ > /tʰ/ > /pʰ/와 같았다. 한편, 종성에서는 /ŋ/ > /n/ > /l/ > /k/ > /m/ > /p/ > /t/의

순으로 나타났다.

자음 중에서는 대체로 공명음이 장애음보다, 평음이 여타의 발성 유형보다, 그리고 높은 비율을 보였다. 조음 위치에서는 연구개음의 출현 빈도가 높게 나타나는 것이 특징적이었다. 조음 방법별로는 초성에서는 폐쇄음이, 종성에서는 비음이 높은 출현 빈도를 나타내는 것으로 관찰되었다.

사전의 발음형 전체에서 관찰된 음절의 유형 수는 총 1,283이었다. 이 가운데 가장 높은 출현 빈도를 보인 음절은 /ha/였다. 이어서 /li, tɕi, i, ki, sa, tɕak/ 등이 그 뒤를 따랐다. 그리고 1,283개의 유형 중에서 고빈도 82개 유형, 즉 6.4%에 불과한 음절의 유형이 전체 자료의 50.25%를 차지하는 것으로 나타났다. 고빈도 음절형들은 대부분 V형과 CV형이었다.

이 연구는 한국어의 기초 통계가 풍부하지 않은 점을 고려할 때, 기존의 연구에서 제시하지 못한 사

전 표제어 전체를 대상으로 음소 빈도를 다면적이고 체계적으로 정리함으로써, 앞으로 응용 분야에서 활용될 수 있는 기초 자료로서의 가치를 인정받을 수 있을 것이라고 생각된다. 한국어 말소리의 빈도 관련 정보를 좀 더 치밀하게 알아보기 위하여 앞으로의 연구에서는 표제어의 품사별, 원어정보별 음소 빈도의 차이에 대해 논의할 필요가 있을 것으로 생각된다. 한국어 말소리의 빈도 관련 정보를 체계적이고 종합적으로 정리하는 일은 한국어의 말소리 특징을 밝히는 국어학적 의의뿐만 아니라, 조음 장애나 청각 장애의 진단과 치료에 방향을 제시할 수 있다는 점에서 언어 병리학적으로도 큰 의미를 가질 수 있을 것이다. 또한, 음성 공학 분야에서도 음성 합성기와 인식기의 성능 향상에도 도움을 줄 수 있는 기초 자료로 활용될 수 있을 것이다.

<표-7> 음절의 구조별 유형 빈도와 출현 빈도 총정리

대분류	V, GV	CV, CGV	CVC, CGVC	VC, GV
유형 빈도(%)	1.3	17.6	74.8	6.2
출현 빈도(%)	5.2	48.1	41.5	5.2
종성 유무	종성 없음(개음절)		종성 있음(폐음절)	
유형 빈도(%)	18.9		81.1	
출현 빈도(%)	53.4		46.6	
초성 유무	초성 없음	초성 있음		초성 없음
유형 빈도(%)		92.4		
출현 빈도(%)		89.6		
유형 빈도(%)	7.6		10.4	
출현 빈도(%)	10.4		7.6	
음절 종류수	1,283			
상위 50%(비율)	82(6.4)			
고빈도 음절형	/ha>/li>/tɕi>/i>/ki>/sa>/tɕak>/twe>/tɕa>/si>/su>/tɕ/>/ka/			
관찰 음절 총수	122,761			

### 참 고 문 헌

김경일(1985). 『한국어 음절구조에 관한 통계분석』. 서울대학교 석사학위논문.

신지영(2000). 『말소리의 이해』. 서울: 한국문화사.

이상억(1990). 국어 어휘부의 계량언어학적 연구. 『강신함 교수회갑기념 국어학논문집』, 331-346.

한국방송공사(1993). 『표준한국어발음대사전』. 서울: 어문각.

한성우(2006). 국어 단어의 음소 분포 『어문학』, 91, 163-191.

ABSTRACT

## Phoneme and Syllable Frequencies Based on the Analysis of Entries in the Korean Dictionary

Jiyoung Shin<sup>§</sup>

Department of Korean Language and Literature, Korea University, Seoul, Korea

**Background & Objectives:** The aim of the present study is to provide information on the occurrence frequencies of Korean phoneme and syllable units based on the careful analysis of entries in the Korean Dictionary. Information on phoneme and syllable frequencies of the entries is a fundamental knowledge component of the Korean language. **Methods:** Using 47,401 entries from the Yonsei Korean Dictionary, the occurrence frequencies for units of Korean phonemes and syllables were analyzed. Pronunciations of all the entries were manually checked, and SynKDP 1.5.2 was used to calculate the phoneme and syllable frequencies. **Results:** A total of 122,761 syllables and 290,131 phonemes were observed from 47,401 entries, with 65.8% of consonants occupying the onset position in the syllable structure. Among 36 phonemes, /a/ showed the highest frequency, and /w/ showed the lowest frequency. Among 19 consonants, /k/ occupied the highest ranking, and /n/, /ŋ, /l/, and /t/ were the next rankings, in that order. Syllable frequency was also investigated. A total of 1,283 syllable types were observed. However, the most frequent 82 syllable types, i.e., 6.4% of all observed syllable types, accounted for over 50% of the data. **Discussion & Conclusion:** Information on the phoneme and syllable frequencies of entries in the Korean dictionary is a fundamental knowledge component of the Korean language. Therefore, this type of information is useful not only for Korean linguistics, but also speech pathology, Korean language teaching, and speech engineering. (*Korean Journal of Communication Disorders* 2010;15:94-106)

<sup>§</sup> Correspondence to

Prof. Jiyoung Shin PhD,  
Department of Korean  
Language and Literature,  
Korea University,  
5-1 Anam-dong,  
Sungbook-Gu, Seoul, Korea  
e-mail: shinjy@korea.ac.kr  
tel.: +82 2 3290 1973

**Key Words:** dictionary entries, phoneme frequency, vowel frequency, consonant frequency, syllable frequency

### REFERENCES

- Han, S. (2006). A study on phonemic distribution of native single word in Korean. *Eomunhak*, 91, 163-191.
- KBS (1993). *Pronunciation Dictionary of Standard Korean*. Eomunkak: Seoul.
- Kim, K. (1985). *A statistical analysis of the syllable structure in Korean*. Unpublished masters's thesis. Seoul National University, Seoul.
- Lee, S. (1990). A quantitative study on Korean lexicon. *Papers on Korean Linguistics in honor of professor Kang, Shin-Hang*, 331-346.
- Shin, J. (2000) *Understanding Speech Sounds*. Hankukmonhwas: Seoul.

■ Received, January 19, 2010 ■ Final revision received, February 22, 2010 ■ Accepted, February 25, 2010.

© 2010 The Korean Academy of Speech-Language Pathology and Audiology <http://www.kasa1986.or.kr>