

3세~8세 아동의 자유 발화 분석을 바탕으로 한 한국어 말소리의 빈도 관련 정보

신지영*

* 고려대학교 국어국문학과

Phoneme Frequency of 3 to 8-year-old Korean Children

Department of Korean Language and Literature, Korea Univ.

shinjin@korea.ac.kr

Abstract

The aim of this study is to provide some information on frequencies of occurrence for units of Korean phonemes and syllables analysing spontaneous speech spoken by 3 to 8-year-old Korean children. 49 Korean Children(7~10 children for each age) were employed as subjects for this study. Speech data were recorded and phonemically transcribed 120 utterances for each child were selected for analysis except one child whose data were only 91 utterances. The data size of the present study were 5,971 utterances, 5,1554 syllables, and 105491 phonemes. Among 19 consonants, /n/ showed highest frequency rate and /k, l, m/ were occupied the next rankings. The frequency rate of these four consonants were over 50% for all age groups. Among 18 vowels, /a/ was the most frequent one and /u/ and /ʌ/ were the second and third respectively. The frequency rate of these four consonants were over 50% for all age groups. Frequently occurring syllable types were a part of grammatical word in most cases. Only 5~6% of syllable types covered 50% of speech.

I. 서론

이 연구의 목적은 3세에서 8세 사이 아동의 자유 발화를 수집, 전사하여 한국어 말소리의 빈도와 관련된 정보를 추출해 보고자 하는 데 있다. 자유 발화를 수집, 전사하여 말소리 관련 정보를 추출하는 작업은 많은 시간과 노력을 요한다. 더욱이 아동의 발화 자료를 수집하고 전사하여 말소리의 빈도를 조사하는 작업은 성인의 발화를 대상으로 하는 것보다 더 많은 시간과 노력이 필요한 것이 사실이다. 하지만 아동 발화를 대상으로 말소리의 빈도를 살펴보는 일은 그간의 국어 연구에서 소외되었던 아동 발화를 대상으로 국어학적 정보를 제공한다는 의미 외에도 성인 발화와 비교를 통하여 언어 습득과 관련된 문제들을 살펴볼 수 있다는 점과 장애 아동의 진단과 치료에 활용될 수 있다는 점에서 아주 중요한 의미를 지닌다고 할 수 있다.

말소리 빈도와 관련된 연구를 수행하기 위해서는 우선 문자 언어 자료가 아니라 음성 언어 자료를 수집해야 하며, 이를 대상으로 '음성 전사'를 수행해야 한다. 형태 위주의 표기를 원칙으로 하고 있는 현재의 정서법 체계 하에서는 표기와 발음이 일치하지 않는 경우가 많기 때문에 문자 언어 자료를 가지고 바로 말소리 빈도와 관련된 연구를 수행할 수 없기 때문이다. 결국, 음성 언어 자료를 수집, 전사하여 연구를 진행하는 것은 문자 언어를 대상으로 연구를 진행하는 것보다 많은 시간과 노력과 비용을 요한다[1].

이러한 이유 때문에 지금까지 한국어의 말소리 빈도 관련 연구는 어휘, 어절, 형태소 등 한국어의 다른 문법 단위나 한글 자소 빈도 관련 연구에 비하여 양적으로나 질적으로 매우 빈약하다.

한국어 말소리 빈도와 관련된 선행 연구는 자료의 성격에 따라서 크게 두 가지로 나눌 수 있다. 하나는 문자 언어 자료를 추정되는 발음형으로 연구자가 바꾸어 연구의 대상으로 삼은 것이고, 다른 하나는 음성 언어 자료를 직접 전사하여 연구의 대상으로 삼은 것이다. [2], [3], [4], [5] 등은 전자의 방법을 이용한 연구이고, [6], [7], [8]은 후자의 방법을 이용한 연구이다.

문자 언어 자료를, 추정되는 발음형으로 바꾼 후 빈도를 조사한 연구는 다시 사전의 표제어를 대상으로 한 연구와 문장을 대상으로 한 연구로 나눌 수 있다. [2], [3]은 사전을 대상으로 한 연구이고, [4], [5]는 문장을 대상으로 한 연구이다.

음성 언어 자료를 수집한 후에 이를 직접 전사한 자료를 바탕으로 말소리 빈도를 조사한 연구는 발화자의 연령에 따라서 아동 발화를 연구의 대상으로 하였는지, 성인 발화를 대상으로 하였는지에 따라 나누어 볼 수 있다. [6]과 [8]은 아동 발화를 대상으로 한 연구이고, [7]은 성인 발화를 대상으로 한 연구이다.

II. 연구 방법

2.1 연구 대상

부모가 서울 방언 화자이고, 서울에서 태어나 현재 서울에 살고 있는 만 3세에서 만 8세 사이의 정상 발달을 보이는 아동을 연구의 대상으로 삼았다. 각 연령별로 10명씩, 모두 60명을 연구의 대상으로 삼아 자료를 수집하였다. 그러나 철자 전사와 1차 발음 전사를 마치고 발음 전사를 수정하는 과정에서 11명분의 원자료가 유실되었다. 그 결과 연령별로 7명에서 10명의 아동이 연구의 대상이 되었으며, 연구의 대상이 된 아동의 수는 모두 49명(여아: 31명, 남아: 18명)이었다). 대상 아동의 연령별, 성별 구성은 <표 1>과 같다.

<표 1> 대상 아동 정보

	3세	4세	5세	6세	7세	8세	계
수	8명	10명	9명	7명	8명	7명	49명
남여	35	46	45	34	26	25	1831

2.2. 자료 수집

1) 원자료는 남아와 여아의 비율이 55(3세, 6세), 46(나머지 연령)이었다 그런데 공개롭게도 남아의 자료가 대거 유실되는 바람에 7세와 8세의 경우는 성비에서 심한 불균형을 보이게 되었다 하지만 앞에서 언급하였듯이 선행 연구 결과 성인[7]과 아동[8] 모두에서 성비에 따른 음소 빈도의 차이는 통계적으로 유의미하지 않은 것으로 나타났으므로 성비의 불균형으로 인한 문제는 크게 우려하지 않아도 좋을 것으로 판단하였다.

자연스러운 환경에서 대상 아동들의 자유 발화를 수집하기 위하여 학령기 이전 아동, 즉 만 3세~6세 아동의 경우는 1) 사진보고 말하기, 2) 소꿉놀이, 3) 그림책 보고 말하기 등의 방법이, 학령기 아동, 즉 만 7세 이상의 아동의 경우는 1) 사진보고 말하기, 2) 그림책 보고 말하기, 3) 학교생활이나 가정생활 등 일상생활에 관한 말하기 등의 방법이 사용되었다.

자유 발화의 수집은 주로 아동의 집(주로 아동의 방)에서 언어병리학을 전공하는 대학원생과 학부생들에 의해 이루어졌다. 발화 상황에 대한 이해를 필요로 하는 경우를 대비하여 모든 발화 자료는 캠코더를 이용하여 수집하였고, 다음의 원칙에 따라 전사되었다

(1) 음성 전사 원칙

- 1 발화의 구분 기준은 Owens(1992)와 김영대(2002)를 따른다
- 2 발음 나는 대로 전사하도록 한다
- 3 음성 전사 자료의 띄어쓰기 단위는 경세구(accentual phrase) 단위로 한다
- 4 잘 들리지 않는 음절은 음절의 수만큼 X자로 표시한다

이러한 원칙에 의하여 전사된 자료 중에서 자동 발화, 반복 발화, 감탄사만 있는 발화, 불명확한 발음이 포함된 발화 등을 제거한 후에 한 아동 당 무작위로 120발화씩을 선정하여 분석하였다. 단, 한 아동(3세 여아)의 경우는 발화 선정 원칙에 따라 발화를 선정한 결과, 총 발화량이 충분치 않아서 선정 발화의 수가 91발화인 경우가 있었다. 따라서 분석의 대상이 된 자료는 49명이 발화한 5,851발화, 51,554음절이었다. 발화 자료와 관련된 상세한 정보는 <표 2>와 같다.

표에서 볼 수 있듯이 대체로 연령이 높아지면서 발화 당 음절 수, 발화 당 강세구 수, 발화 당 음소 수가 모두 증가 하는 것으로 나타났다. 결국, 고연령으로 갈수록 발화의 길이가 길어진다는 사실을 보여주고 있다. 한 가지 흥미로운 것은 대체로 3세와 4세, 5세와 6세, 7세와 8세가 한 묶음으로 묶이는 경향을 볼 수 있다는 것이다. 다시 말해서 두 연령 사이의 차이보다는 다른 연령 묶음의 차이가 훨씬 큰 것으로 나타났다.

<표 2> 발화 자료 정보

	3세	4세	5세	6세	7세	8세	계 평균
발화 수	1,051	1,200	1,080	840	960	840	5,971
음절 수 (=모음 수)	4,743	7,294	9,703	7,073	11,348	11,193	51,554
발화당 음절 수	4.5	6.1	9.0	8.4	12.0	13.3	8.6
강세구 수	1,883	2,873	3,516	2,703	4,138	3,918	19,071
발화당 강세구 수	2.0	2.4	3.3	3.2	4.3	4.6	3.2
음소 수	9,250	14,524	19,627	14,532	24,116	23,442	105,491
발화당 음소 수	8.8	12.1	18.2	17.3	25.1	27.9	17.7

2.3 분석 방법

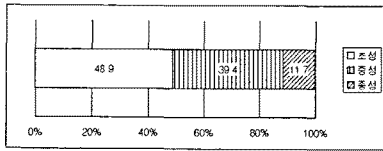
음성 전사를 마친 자료를 대상으로 음소 관련 정보, 음절 관련 정보 등 말소리 관련 정보를 추출하였다. 음소 빈도와 음절 빈도를 얻기 위해서 SynKD1.5.2(일명 캄썩새)가 분석 도구로 사용되었다.

본 연구에서는 말소리 관련 정보로 크게 음소 관련 정보와 음절 관련 정보를 살펴보았다. 음소 관련 정보로는 자음과 모음의 빈도를 알아보았으며 음절 관련 정보로는 고빈도 음절 유형에 대한 정보와 세부 음절 유형에 대한 정보를 알아보았다.

III. 연구 결과

2.1 음소 관련 정보

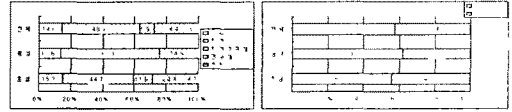
자료에 나타난 음소의 수는 모두 105,491개였다. 이 중에서 자음이 53,937개, 모음이 51,534개로 비슷한 수를 보였다. 결국, 한 음절이 모음 하나와 자음 하나의 비율로 구성되어 있음을 알 수 있다. 좀더 구체적으로 알아보기 위하여 초성과 중성 중성에 각각 쓰인 음소의 비율을 살펴보았다. <그림 1>에 보인 바와 같이 모음이 음소의 반 정도인 48.9%를, 초성이 39.4%를, 중성이 11.7%를 차지하였다.



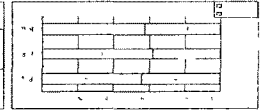
<그림 1> 음소의 구성 비율

아동 발화에 나타난 자음을 조음 위치별로 살펴보면, <그림 2>에서 보듯이 전반적으로 치경음이 여타 조음 위치보다 높은 비율을 나타냈다. 특히, 중성의 경우 치경음이 압도적인 비율을 보였다. 자음을 설정성에 따라 분류해 보면, <그림 3>에서 보듯이 설정음이 비설정음에 비하여 약간 높은 비율을 보였다. 설정음은 초성에서보다는 중성에서 조금 더 높은 비율을 보였다

자음의 조음 방법별 비율은 <그림 4>에서 볼 수 있듯이 전체적으로는 폐쇄음과 비음의 비율에 큰 차이를 보이지 않았지만, 초성에서는 폐쇄음이, 중성에서는 비음이 압도적인 비율을 보이고 있다.

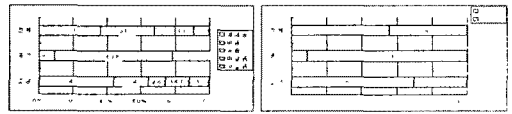


<그림 2> 자음의 조음 위치별 비율

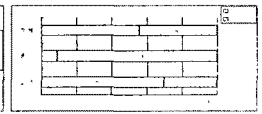


<그림 3> 설정음과 비설정음의 비율

다음은 자음의 음절 내 위치에 따른 공명음과 장애음의 비율을 살펴보았다. 자음 전체적으로 볼 때는 거의 비슷한 비율을 보였지만, <그림 5>에 보였듯이 초성에서는 장애음이 약 70% 정도의 비율을 보인 반면에 중성에서는 공명음이 90%를 상회하는 높은 비율을 보였다.



<그림 4> 자음의 조음 방법별 비율

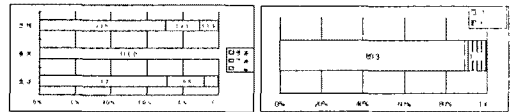


<그림 5> 공명음과 장애음의 비율

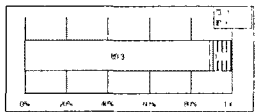
장애음의 발생 유형별 출현 빈도에도 큰 차이가 있었다. 평음이 압도적인 비율을 차지하고 있으며, 경음과 기음의 순으로 나타났다. 물론, 중성에는 평음만이 나타났지만, 중성에서 폐쇄음의 비율이 매우 낮았기 때문에 전체 비율과 초성 비율 사이에는 큰 차이가 없었다.

모음의 경우는 <그림 7>에서 볼 수 있듯이 단모음이 이중모음에 비하여 높은 비율을 나타냈다. 전체 자료에서 단모음은 89.3%를 보였다. 이중모음의 유형별 빈도를 보면, <그림 8>에 보였듯이 /i/개가 약 80%를 차지하였으며, /-/-개는 0.4%에 불과했다.

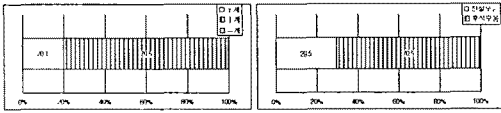
단모음의 경우는 <그림 9>와 <그림 10>에서 볼 수 있듯이 전설 모음보다는 후설 모음이, 원순 모음보다는 평순 모음이, 높은 비율을 차지하는 것으로 나타났다. 또한 <그림 11>에서 보듯이 모음의 개구도에 따라서는 중모음이 고모음이나 저모음에 비하여 약간 높은 비율을 나타냈다. 고모음과 중모음이 각각 3개씩이고, 저모음이 1개에 불과하다는 점을 고려한다면, /-/의 빈도가 매우 높다는 것을 알 수 있다.



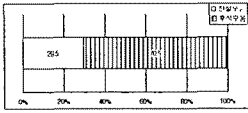
<그림 6> 자음의 발생 유형별 비율



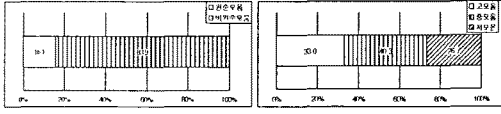
<그림 7> 단모음과 이중모음의 비율



<그림 8> 이중모음의 구성 비율



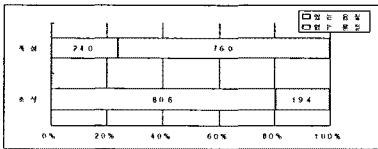
<그림 9> 전설과 후설 모음의 비율



<그림 10> 원순과 비원순 모음의 비율 <그림 11> 모음의 개구도에 따른 비율

전체적으로 852종의 음절형이 자료에 나타났는데, 그 중에서 단지 38%를 차지하는 32개의 음절형이 전체 자료의 50% 이상을 차지하는 것으로 나타났다. 고빈도 음절형 중에는 폐음절형이 /ㄴ/ 하나밖에 없었고, 대부분은 CV형이었다. 고빈도 음절형의 대부분은 조사나 어미 등 문법 형태소를 구성하는 음절형인 경우가 많았다. 특히, 연령이 높아지면서 /ㄴ/의 빈도가 높아지는 것은 동사의 습득 순서, 문장 유형의 습득 순서나 조사의 발달과 관련한 것으로 생각되어 흥미롭다.

전체 자료를 대상으로 개음절과 폐음절의 비율, 그리고 초성이 있는 음절과 없는 음절의 비율을 알아보았다. <그림 12>에서 보듯이 초성이 있는 음절이, 그리고 초성이 없는 음절이 큰 비율을 차지하였다.



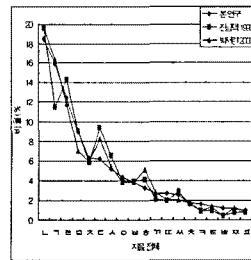
<그림 12> 초성, 종성 유무에 따른 비율

IV. 논의 및 결론

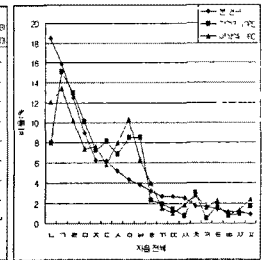
본 연구 결과와 선행 연구의 결과를 비교해 보았다. 자음의 전체 빈도에 대한 비교부터 실시하였다²⁾. <그림 13>은 성인 자료와의 비교이며³⁾, <그림 14>는 사전 자료와의 비교이다. 그림에서 보듯이 본 연구의 결

과는 성인의 자유 발화를 분석한 [7]의 자료와 유사한 결과를 보였다. 약간의 차이를 보이는 것은 /ㄷ/와 /ㅎ/의 비율이 [7]에서 높게 나타났다는 점이다.

사전 자료와의 비교 결과, <그림 14>에서 보듯이 많은 차이가 관찰되었다. /ㄴ/의 비율이 사전 자료에서 낮게 나타나는 점, /ㄴ, ㅁ, ㅂ/가 사전 자료에서 높은 비율을 차지하고 있다는 점, 그리고 경음의 비율이 사전 자료에서 전반적으로 낮게 나타나고 있다는 점을 들 수 있다. 발화 자료에서 /ㄴ/의 비율이 높은 이유는 /ㄴ/를 포함하는 문법 형태소가 많기 때문인 것으로 보인다. 또한, 경음의 비율이 발화 자료에서 상대적으로 높게 나타난 것은 경음화와 관련한 것으로 생각된다.



<그림 13> 자음 전체 비율에 대한 선행 연구와의 비교(1)



<그림 14> 자음 전체 비율에 대한 선행 연구와의 비교(2)

다음은 모음의 출현 빈도를 선행 연구와 비교해 보았다. <그림 15>는 발화 자료 혹은 문어 언어 자료에 대한 발음 추정형 자료들과의 비교이고, <그림 16>은 사전 자료와의 비교이다

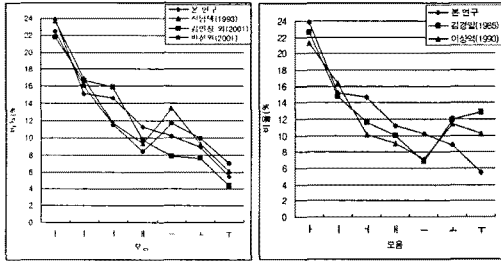
모음이 빈도 비교는 압도적인 비율을 보였던 단모음만을 비교하였으며, 본 연구와의 비교를 위하여 선행 연구의 결과를 7모음 체계로 변환하였다⁴⁾. <그림 15>에서 보듯이 본 연구와 [8]은 거의 유사한 빈도순을 보였으나, [4], [5]과는 약간의 차이를 보였다. [4], [5]에서는 /-/의 빈도가 상당히 높게 나타났지만, 상대적으로 /ㅁ/와 /ㄴ/의 비율이 낮은 것으로 나타났다. 이러한 차이가 문어 자료와 구어 자료의 차이를 반영하는 것인지, 성인 발화와 아동 발화의 차이를 반영하는 것인지는 성인의 자유 발화에 대한 선행 연구가 없는 현재로서는 알기 어렵다.

한편, <그림 16>에서 보듯이 단모음의 경우는 사전 자료와 발화 자료는 많은 차이를 보이는 것으로 나타났다. 발화 자료에 비하여 /-/의 비율이 낮고, /ㄱ/와 /ㄷ/의 비율이 높았다. 발화 자료에서 /-/의 비율이

2) 뒤에 제시된 그래프에 비교 대상이 된 선행 연구의 수가 다른 것은 선행 연구에서 자료를 제시한 방법이 서로 달랐기 때문이다. 전체 자료에 대한 정확한 수치가 제시된 경우에만 본 연구의 결과와 비교하였다.
3) [8]에는 자음 전체 자료에 대한 정보가 나와 있지 않다. 자음의 음절 내 위치에 따라 초성과 종성의 비율만을 구체적으로 밝히고 있어서 그림에 함께 제시하지 못했다

4) 즉, /ㅁ/와 /ㅂ/를 통합하고, /ㄴ/와 /ㄹ/를 이중모음으로 처리하였다

현저히 높은 것은 문법 형태소에 /-/가 많이 포함되어 있기 때문인 것으로 보인다 물론, [2]와 [3]이 모두 사전의 모든 표제어를 대상으로 빈도를 낸 것이 아니므로 사전의 결과는 전체 사전의 표제어를 대상으로 한 경우에 달라질 수 있을 것이다.



<그림 15> 모음 비율에 대한 선행 연구와의 비교(1)

<그림 16> 모음 비율에 대한 선행 연구와의 비교(2)

앞으로의 연구에서는 성인의 자유 발화에 나타난 음소 빈도를 알아보는 일과, 국어 사전의 전체 표제어에 대한 음소 빈도 조사가 이루어져야 할 것이다. 이러한 자료들을 바탕으로 궁극적으로는 문자 언어 자료를 발화한 자료와 성인과 아동의 자유 발화 자료, 그리고 사전 표제어의 발음 자료를 비교하여 각 자료의 말소리 특징을 면밀히 비교해야 할 것이다. 이러한 연구는 한국어의 특징을 밝힌다는 국어학적 의미 외에도 조음 장애, 청각 장애의 진단과 치료에 방향을 제시한다는 점에서 언어병리학적으로도 중요한 의미를 지니며, 음성 공학적으로도 긴요한 정보가 될 것이다.

참고문헌

[1] 신지영, “음성 코퍼스를 활용한 국어 연구,” *한국어학* 23, pp 23-47, 2004.
 [2] 김경일, *한국어 음절구조에 관한 통계분석*, 서울대학교 석사학위논문, 1985
 [3] 이상여, “國語 語彙部の 計量言語學的 研究”, *姜信弘教授回甲紀念 國語學論文集*, pp. 331-346, 1990.
 [4] 진남택, “한국어 음소의 기능부담량: 계량 언어학적 연구” *말소리* 25. pp. 65-91, 1993.
 [5] 변성완, “한국어의 발음 음소별 빈도로 본 한국어 PB Word의 타당성,” *대한이비인후과학회지* 44(5), pp. 485-489, 2001.
 [6] 양인홍, 권정순, 이규식, “幼兒들의 會話에서 나타난 音素 頻度調査.” *難聽과 言語障碍研究* 9(1), pp

35-39, 1986.

[7] 박서란, *성인의 일상적인 대화에서 나타나는 말소리 출현빈도 연구*, 이화여자대학교 석사학위논문, 2000.
 [8] 김민정, 배소영, 고도홍. “2-5 세 아동의 자발적 발화에 나타난 한국어 음절 및 음운 빈도,” *음성과학* 8(4), pp 99-107, 2001.