

# 성인 자유 발화 자료 분석을 바탕으로 한 한국어의 음소 전이 빈도

신지영<sup>§</sup>

(고려대학교 국어국문학과)

---

신지영. 성인 자유 발화 자료 분석을 바탕으로 한 한국어의 음소 전이 빈도. 『언어청각장애연구』, 2008, 제13권, 제3호, 477-502. **배경 및 목적:** 이 연구의 목적은 한국어를 정상적으로 구사하는 성인 화자들의 자유 발화를 수집·전사하여 한국어의 자유 발화에서 관찰되는 음소 전이 빈도(두음소 빈도) 정보를 살펴보고, 체계적으로 기술하여 한국어의 말소리 관련 기초 자료를 제공하려는 것이다. 음소 전이 빈도에 대한 이러한 지식은 한국어 화자들이 가지고 있는 한국어의 기본적인 지식이기 때문에 이러한 지식을 체계화시키는 작업은 기초 연구로서 반드시 필요하다. **방법:** 서울 방언을 구사하는 성인 57(남 28, 여 29)명의 자유 발화를 수집하고 이를 음성 전사한 후에 운율 단위를 고려하여 음소 전이 빈도(두음소 빈도)를 추출하였다. 본 연구의 분석 대상 자료는 총 35,439 발화, 403,572음절, 858,479음소였다. **결과:** 발화 자료를 분석한 결과, 모두 980,639개의 두음소가 관찰되었다. 논리적으로 연결 가능한 두음소 1,225종 중에서 1,009종의 두음소가 관찰되었고, 이중 5.5%에 해당하는 고빈도 두음소 55종의 출현 빈도 합이 50%를 넘는 것으로 나타났다. 가장 높은 출현 빈도를 보인 두음소는 #+ㄱ> ㄱ+#> ㄷ+#> ㄷ+ㄹ> ㄴ+#> ㄴ+ㄹ> ㄱ+ㄷ 등이었다. 고빈도 두음소들은 대체로 고빈도 문법 형태소의 음절형과 관련한 것으로 보였다. **논의 및 결론:** 음소 전이 빈도를 면밀히 살펴본 결과, 한 음소와 다음 음소의 전이 역시 고빈도 유형이 발화에서 반복적으로 사용됨을 알 수 있다. 음소 전이 빈도에 대한 이러한 연구 결과는 국어학은 물론, 언어병리학, 한국어 교육, 음성 공학 분야에서도 요긴하게 활용될 수 있을 것으로 기대한다.

---

**핵심어:** 음소 전이 빈도, 두음소, 두음소 빈도, 자유 발화, 음성 전사

## I. 서론

이 연구의 목적은 한국어를 정상적으로 구사하는 성인 화자들의 자유 발화를 수집·전사하여 한국어의 자유 발화에서 관찰되는 음소 전이 빈도와 관련된 정보를 다각적으로 살펴보고 체계적으로 기술하여 한국어의 말소리 관련 기초 자료를 제공하려는 데 있다. 한 음소가 다음 음소와 연결되어 만들어 내는 두 음소의 연결체를 두음소(bi-phoneme)라고 한다면, 음소 전이 빈도란 두음소의 빈도를 의미하게 될 것이다. 본 연구에서는 한국어의 음소 연결 양상의 특징을 두음소의 빈도를 상세히

---

게재 신청일: 2008년 6월 15일; 최종 수정일: 2008년 8월 15일; 게재 확정일: 2008년 8월 20일

§ 교신저자: 신지영, 고려대학교 국어국문학과 교수, 서울시 성북구 안암동 5-1, e-mail: shinjy@korea.ac.kr, tel.: 02-3290-1973

© 2008 한국언어청각임상학회 <http://www.kasa1986.or.kr>

살펴봄으로써 알아보고자 한다.

두음소 빈도에 대한 지식은, 한국어 화자들이 음향 신호를 언어의 유의미한 단위로 분절하여 이해하는 데 사용하는 가장 기본적인 지식 중 하나일 것이다. 그럼에도 불구하고 한국어의 음소 전이 빈도에 대한 선행 연구를 다양하게 찾기는 매우 어려운 형편이다. 그동안 말뭉치 언어학이나 전산 언어학적 방법론이 한국어 연구에 적극적으로 활용되면서 한국어의 여러 문법 단위들, 예를 들어 어절, 어휘, 형태소 등에 대한 연구 결과가 다양하게 보고되고 있지만, 한국어의 말소리와 관련된 빈도 정보는 양적으로나 질적으로 풍성하다고 하기 어렵다. 그 이유는 신지영(2005a)에서 논의하였듯이 말소리와 관련된 빈도를 알아보기 위해서는 음성 자료를 수집한 후에 음성 전사를 수행해야만 연구를 진행할 수 있고, 이 음성 전사는 훈련된 전문가의 엄청난 시간과 노력을 요하는 일이기 때문이다.

한국어의 음소 전이 빈도와 관련된 연구를 수행했던 선행 연구로는 김민정·배소영·고도홍(2001), 신지영(2005b), 박동근·이석재(2005) 등 3편의 논문을 찾을 수 있다. 이 중에서 김민정·배소영·고도홍(2001)과 신지영(2005b)은 아동의 자유 발화를 분석 대상 자료로 하였으며, 성인 자유 발화를 대상으로 연구를 진행한 것은 박동근·이석재(2005)가 유일한 것으로 보인다.

#### <표 - 1> 음소 전이 빈도에 대한 선행 연구

| 분석 대상 자료 | 해당 선행 연구                      |
|----------|-------------------------------|
| 아동 자유 발화 | 김민정·배소영·고도홍(2001), 신지영(2005b) |
| 성인 자유 발화 | 박동근·이석재(2005)                 |

김민정·배소영·고도홍(2001)에서는 2세에서 5세 아동, 총 64명(각 연령당 16명)의 자유 발화를 수집하여 아동당 50발화씩, 총 21,897음절을 대상으로 음소의 출현 빈도와 함께 전이 빈도를 간략히 논의하였다. 이 연구에서 다루고 있는 전이 빈도는 초성과 중성의 전이 빈도와 종성과 다음 음절의 초성의 전이 빈도뿐이다. 그 이외의 전이 빈도, 예를 들어 중성과 중성의 전이 빈도에 대한 정보 등은 제시하고 있지 않다. 이 연구는 전이 빈도에 대한 전면적인 정보를 제공하지 못한다는 문제와 함께, 연구 대상 자료가 아동의 자유 발화를 직접 음성 전사한 것이 아니라 철자 전사를 수행한 후에 어절 단위로 발음 추정형을 만들어 빈도 정보를 추출했다는 문제를 안고 있다.

신지영(2005b)은 3세에서 8세 아동, 총 49명이 발화한 음성 자료를 운율 단위를 고려하여 음성 전사 한 후에 전이 빈도(두음소 빈도)를 살펴보았다. 자료의 크기는 5,851발화, 51,554음절이었고, 여기서 관찰된 두음소의 총수는 124,588개였다. 이 연구에서는 음소의 전이 빈도를 운율 단위를 고려하여 음절 내 두음소 빈도, 음절 경계를 사이에 둔 두음소 빈도, 강세구 경계에 놓인 두음소 빈도 등을 다양하게 살펴보았다.

박동근·이석재(2005)는 대학생 총 154(남 65, 여 89)명이 30개의 대화 장면에서 발화한, 총 93,844어절, 218,683음절을 대상으로 음소 빈도와 함께 음소 배열 빈도를 살펴보았다. 이 연구에서 살

펴본 음소 배열 빈도는 초성과 중성, 중성과 종성, 종성과 초성의 배열 빈도였다. 하지만 이 연구는 몇 가지 문제점이 발견된다. 첫째, 발화자들이 어떤 방언을 구사하는지에 대한 정보가 전혀 없다는 점이다. 연구자들은 154명의 발화자 수는 중복해서 대화 장면에서 참여한 사람들을 고려하지 않고 계산된 것이라고 밝히고 있지만, 이들이 어떠한 방언을 구사하는지, 발화자의 방언이 통제되어 녹음되었는지 등에 대한 언급이 전혀 없다.

둘째, 전사 결과로 예를 든 자료에 의하면, 운율 단위에 대한 고려 없이 어절 단위로 전사를 수행하였고, 또한 실제 발화형이 아니라 약간은 이상적인 발화형을 전사 결과로 삼았다는 점이다. 예를 들어 전사의 예로 제시되어 있는 ‘있구’의 발화형 /일꾸/의 경우는 아주 예외적인 경우를 제외하고는 일상적인 자유 발화에서 관찰되지 않는 발화형으로 보인다. 본 연구의 자료에서는 거의 대부분의 발화자가 이 경우 /이꾸/로 발음하였다.

셋째, 운율 단위에 대한 고려 없이 음소 배열 제약을 논의하고 있어서 특히 종성과 초성의 배열 빈도(즉, 전이 빈도)의 경우 제시된 통계가 정확히 어떠한 환경에서의 배열을 의미하는지 명확하지 않다는 점이다. 한 음절의 종성과 다음 음절의 초성 사이에는 음절 경계만 놓일 수도 있고, 강세구나 억양구 경계가 나타날 수도 있다. 각각의 경우마다 음소 배열 제약이 달리 존재하기 때문에 운율 단위를 무시하고 음소 배열 제약을 논의하는 것은 정밀한 결과로 보기 어렵다. 연구에 명시적으로 밝히고 있지는 않지만, 음성 전사를 어절 단위로 한 것으로 보이기 때문에 대체로 어절 내의 종성과 초성의 전이빈도를 의미하는 것 같기는 하다. 하지만, ‘어절’이라는 단위는 운율적인 단위가 아니기 때문에 어절의 경계와 음소의 배열은 직접적인 관련이 없으므로 정밀한 기술이라고 하기 어렵다.

아울러 모든 가능한 음소 배열 제약을 논의하고 있지 않다는 점도 문제점으로 지적할 수 있다. 예를 들어 선행 음절의 종성이 /ㅇ/이고 후행하는 음절의 초성이 비어 있는 경우, 즉 /ㅇ/으로 끝난 음절이 모음으로 시작하는 음절과 이어질 때 생기는 음소 배열 빈도는 굳이 이름을 붙이자면, ‘종성과 중성의 배열 빈도’가 될 텐데, 박동근·이석재(2005)에는 이에 대한 통계가 제공되어 있지 않다.

이상에서 살펴본 바와 같이 한국어 음소의 전이 빈도와 관련된 연구는 우선 양적으로 매우 빈약하다. 특히 성인의 자유 발화를 대상으로 한 연구는 단 1편을 찾을 수 있었고, 이 1편의 연구는 앞서 살펴본 바와 같이 자료의 수집 단계, 전사 단계, 분석 단계에서 여러 가지 문제점을 안고 있는 것으로 보였다. 이에 본 연구에서는 한국어 발화에서 관찰되는 음소의 전이 빈도를 운율 단위를 고려하여 정밀하게 고찰해 보고자 한다. 이를 위하여 본 연구에서는 서울 방언을 구사하는 성인 57(남 28, 여 29)명의 자유 발화를 수집하고, 음소 수준의 음성 전사를 수행한 후에 음소와 음소의 연쇄 빈도를 정밀하게 살펴보았다.

## II. 연구 방법

### 1. 연구 자료

본 연구의 대상 자료는 신지영(2008)에 사용한 것과 동일하였다. 이를 간략히 요약하면, 피험 자료는 자신이 표준어에 가까운 방언을 구사한다고 생각하는 화자 총 57(남 28, 여 29)명이 참여하였고, 이들은 거의 대부분이 대학생들이었다. 최대한 자연스럽게 자유 발화를 유도하기 위하여 친밀한 관계에 있는 화자들을 3명씩 녹음실로 불러 녹음을 진행하였다. 친밀한 관계에 있는 3명의 화자들은 통제자의 개입 없이도 자발적으로 대화를 이어가는 데 거의 문제가 없었다.

녹음이 끝난 후에는 정해진 전사 원칙을 바탕으로 훈련을 받은 음성학/음운론 전공 석사과정 이상의 연구보조원들(석사과정 3인, 박사과정 2인)이 1차 음성 전사를 수행하였다. 이후 4차에 걸친 검토 작업(연구자)과 3차에 걸친 수정 작업(음성학/음운론 전공 박사과정 이상의 훈련받은 연구보조원 4인)을 통하여 분석 대상 자료가 완성되었다. 이렇게 완성된 말뭉치의 크기는 <표-2>에 정리된 바와 같다. 본 연구의 분석 대상 말뭉치의 크기는 선행 연구인 박동근·이석재(2005)의 약 2배 정도였다.

<표-2> 발화 자료 통계

|        |              |               |
|--------|--------------|---------------|
| 말뭉치 크기 | 발화자 수(남, 여)  | 57(28, 29)    |
|        | 발화 수(화자당 평균) | 35,439(662.5) |
|        | 억양구 수        | 55,927        |
|        | 강세구 수        | 122,919       |
|        | 음절 수         | 403,572       |
|        | 음소 수         | 858,479       |

본 연구에서 관찰한 발화에서 화자들은 19개의 자음, 7개의 단모음, 11개의 복모음을 구사하였다. 단모음 /e/와 /æ/가 중화된 모음은 /ɛ/로, 복모음 /we/와 /wæ/가 중화된 모음은 /ɛ/로, /je/와 /jæ/가 중화된 모음은 /ɛ/로 표기하였고, 삼중모음으로 실현된 /wɔ/는 /u/로 표기하였다.

### 2. 분석 방법

음성 전사를 마친 자료를 대상으로 음소 전이 빈도와 관련된 정보를 다각적으로 추출하였다. 음소의 전이 빈도란, 한 음소가 다음 음소로 연결되는 빈도를 의미하며, 이는 두음소(bi-phoneme)의 빈도로 표현할 수 있을 것이다. 두음소의 빈도를 정확하게 산출하기 위해서는 운율 단위에 대한 고려

가 필요하다. 음소의 연결이 같더라도 중간에 음절 경계가 오느냐, 강세구 경계가 오느냐에 따라 음소 사이의 배열 제약이 달라질 수 있기 때문이다. 한국어의 음절화는 강세구를 영역으로 하여 그 안에서 일어나기 때문에 음소와 음소 사이에 어떠한 운율 단위가 놓이느냐에 대한 정보는 매우 중요하다. 이를 위하여 신지영(2005b)과 같은 방법으로 강세구의 앞뒤에 강세구 경계를 의미하는 휴지 음소인 ‘#’을 삽입하였다. 이를 통하여 강세구의 시작이나 끝에 위치한 음소 P의 두음소는 각각 ‘#+음소’와 ‘음소+#’으로 표현되었다.

음소 전이 빈도 등을 얻기 위해서 SynKD1.5.2(일명 깜짝새)가 분석 도구로 사용되었는데, 깜짝새만으로는 다양한 전이 빈도 정보를 얻기 어렵기 때문에 깜짝새를 최대한 활용하였으나, 대부분의 작업은 직접 수행하였다.

### Ⅲ. 연구 결과

#### 1. 한 음절 내의 두음소 빈도

한 음절 내에서 출현 가능한 두음소의 종류에는 두 가지가 있는데, 초성과 중성(C+V), 중성과 종성(V+C)으로 만들어지는 두음소가 그것이다. <표-3>은 한 음절 내 두음소의 종류 수와 출현 총수, 누적 빈도 50% 이상을 보이는 상위 두음소의 종류 수와 비율을 보인 것이다. 표에서 보듯이, C+V 두음소가 V+C 두음소보다 출현 빈도가 약 3배 정도 높은 것으로 나타났다. 그 이유는 초성이 있는 음절의 빈도가 높고, 중성이 없는 음절의 빈도가 높은 것과 유관하다.

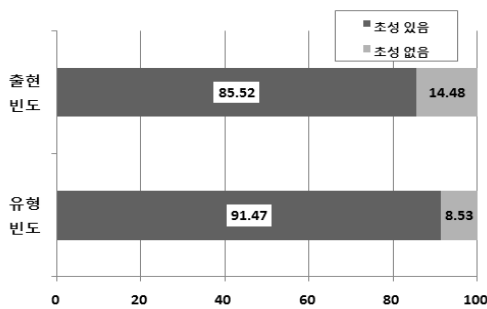
<표-3> 한 음절 내 두음소의 유형 빈도와 출현 빈도

|                 | C+V              | V+C        |
|-----------------|------------------|------------|
| 가능 종류 수         | 306(=(18×18)-18) | 126(=18×7) |
| 출현 종류 수(비율)     | 232(75.8%)       | 98(77.8%)  |
| 상위 50% 종류 수(비율) | 21(9.1%)         | 8(8.2%)    |
| 출현 총수           | 345,128          | 109,779    |

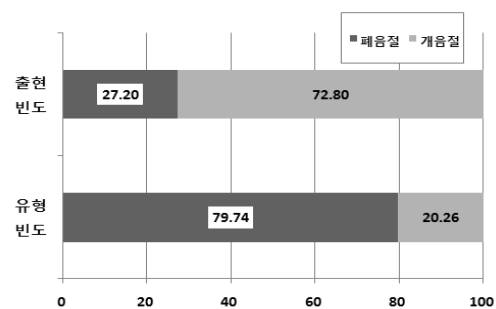
한 음절 내 두음소 빈도의 특성을 살펴보기 위해, 음절 빈도에 대한 정보를 먼저 정리해 보았다. <그림-1>과 <그림-2>는 자료에서 관찰된 초성이 있는 음절과 없는 음절, 중성이 있는 음절(폐음절)과 중성이 없는 음절(개음절)의 유형 빈도와 출현 빈도를 정리한 것이다<sup>1)</sup>. 그림에서 보듯이

1) <그림-1>과 <그림-2>는 신지영(2008)의 음절 빈도 기술에서 제시했던 것과 같다.

초성이 있는 음절은 유형 빈도와 출현 빈도가 모두 높았고, 종성이 있는 음절(폐음절)은 유형 빈도는 높았지만, 출현 빈도가 낮았다. 따라서 초성이 있는 음절에 의해 만들어지는 C+V 두음소의 출현 빈도가 종성이 있는 음절에 의해 만들어지는 V+C 두음소의 출현 빈도보다 훨씬 높게 나타나는 것은 당연하다.



<그림 - 1> 초성 유무에 따른 음절의 유형 빈도와 출현 빈도



<그림 - 2> 종성 유무에 따른 음절의 유형 빈도와 출현 빈도

### 가. 초성과 종성(C+V)의 두음소 빈도

한 음절 내에 놓인 초성과 종성의 전이 빈도를 알아보았다. <표 - 3>에 보였듯이 음절 내에서 만날 수 있는 가능한 초성과 종성의 결합 종류 수는 총 306종(=18×18 - 18<sup>2</sup>)이었다. 하지만 실제 발화 자료에서 관찰된 출현 종류 수는 모두 232종이었다. 빈도순으로 누적 빈도 50% 이상을 보인 C+V 두음소는 불과 21종이었다. 결국, 전체 가능한 306종 중 21종, 즉 6.9%에 불과한 특정 초성과 종성의 결합이 발화에서 매우 반복적으로 사용되고 있음을 알 수 있다. <표 - 4>는 누적 빈도 50% 이상을 차지하는 고빈도 C+V 두음소의 목록과 출현 빈도를, <표 - 5>는 C+V 두음소 전체의 출현 빈도를 보인 것이다.

2) 초성과 종성의 논리적으로 가능한 두음소의 종류 수는 '초성 가능 자음의 수×종성 가능 모음의 수', 즉 '18×18'에서 한국어의 표면 음성 제약(surface phonetic constraint)에 의하여 출현이 불가능한 [ʒ, ʒʒ, ʒʒ]와 /j/-계 이중모음의 연쇄 총수 '18'을 빼는 것으로 산출된다. 신지영(2005b)에 제시된 C+V 두음소의 가능 종류 수에는 표면 음성 제약으로 인해 출현할 수 없는 15개의 두음소가 제외되어 있지 않아 문제가 있다. 따라서 신지영(2005b) <표 - 2>에 제시된 가능 종류 수는 '306'이 아니라 '291(306-15)'이 되어야 한다. 신지영(2005b)은 아동 발화를 연구한 것인데, 아동 발화에서는 삼중모음이 관찰되지 않았기 때문에 현행 연구보다 모음의 수가 하나 적었다.



<표 - 5> C+V 두음소 전체 빈도(%) (\*: 출현 불가능 두음소, -: 미출현 두음소)

|   | ㄱ    | ㅋ   | ㆁ    | ㄷ   | ㄱ   | ㄲ   | ㄷ   | ㅌ   | ㄴ    | ㄷ    | ㄹ    | ㅂ   | ㅃ   | ㅅ   | ㅆ   | ㅇ   | ㅈ   | ㅊ   | ㅊ   | ㅊ     | 계 |
|---|------|-----|------|-----|-----|-----|-----|-----|------|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-------|---|
| ㄱ | 4.0  | 0.0 | 2.5  | 0.4 | 1.7 | 0.1 | 1.5 | 0.0 | 5.1  | 1.7  | 1.8  | -   | 0.1 | 0.2 | 0.0 | 0.0 | -   | 0.0 | 0.0 | 19.2  |   |
| ㅋ | 3.4  | 0.6 | 0.7  | 0.4 | 0.4 | 0.0 | 0.2 | 0.0 | 3.4  | 2.6  | 1.5  | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | -   | -   | -   | 13.4  |   |
| ㆁ | 1.8  | -   | 0.3  | 0.0 | 1.1 | -   | 0.6 | 0.0 | 1.5  | 0.2  | 2.8  | -   | 0.5 | -   | 0.0 | 0.0 | -   | -   | -   | 8.8   |   |
| ㄷ | 1.9  | 0.0 | 2.4  | 0.4 | 0.7 | 0.0 | 0.2 | 0.0 | 1.2  | 1.6  | 1.3  | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | -   | -   | -   | 9.8   |   |
| ㄱ | 2.5  | 0.0 | 0.6  | 1.2 | 1.0 | 0.0 | 0.8 | 0.0 | 0.2  | 0.7  | 0.4  | -   | -   | 0.0 | 0.2 | 0.0 | -   | -   | -   | 7.8   |   |
| ㄷ | 0.9  | -   | 0.4  | 0.2 | 0.8 | 0.0 | 0.6 | 0.0 | 0.1  | 0.3  | 0.3  | -   | 0.0 | 0.1 | -   | 0.0 | -   | -   | -   | 3.7   |   |
| ㄴ | 1.3  | 0.0 | 1.9  | 0.1 | 0.3 | 0.0 | 0.4 | 0.0 | 0.4  | 1.4  | 0.6  | -   | 0.0 | -   | -   | 0.0 | -   | -   | 0.0 | 6.6   |   |
| ㄷ | 1.7  | *   | 1.4  | *   | 0.6 | *   | 0.7 | *   | 0.1  | 2.4  | 0.8  | *   | 0.0 | 0.0 | 0.0 | 0.0 | -   | -   | *   | 7.8   |   |
| ㄷ | 0.3  | *   | 0.4  | *   | 0.1 | *   | 0.1 | *   | 0.0  | 0.7  | 0.1  | *   | 0.0 | 0.0 | 0.0 | 0.0 | -   | -   | *   | 1.8   |   |
| ㄷ | 0.3  | 0.0 | 0.1  | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.1  | 0.2  | 0.7  | -   | 0.0 | 0.0 | -   | 0.0 | -   | -   | -   | 1.6   |   |
| ㄷ | 0.4  | -   | 0.2  | 0.0 | 0.1 | -   | 0.0 | 0.0 | 0.4  | 0.1  | 0.5  | -   | 0.0 | -   | -   | 0.0 | -   | -   | -   | 1.8   |   |
| ㄷ | 0.2  | -   | 0.1  | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.2  | 0.2  | 0.1  | -   | -   | 0.0 | -   | -   | -   | -   | -   | 1.0   |   |
| ㄷ | 2.7  | 0.0 | 0.1  | 0.2 | 0.2 | 0.0 | 0.1 | 0.0 | 0.0  | 0.2  | 1.0  | -   | 0.1 | 0.2 | 0.0 | 0.0 | -   | -   | -   | 4.8   |   |
| ㄷ | 1.1  | -   | 0.4  | 0.1 | 0.3 | 0.2 | 0.3 | 0.0 | 0.2  | 0.2  | 0.3  | -   | 0.0 | 0.1 | 0.0 | 0.0 | -   | 0.0 | 0.0 | 3.3   |   |
| ㄷ | 0.9  | -   | 0.3  | 0.0 | 0.4 | -   | 0.1 | -   | 0.2  | 0.0  | 0.7  | -   | 0.0 | -   | -   | 0.0 | -   | -   | 0.0 | 2.6   |   |
| ㅃ | 0.3  | -   | 0.1  | 0.0 | 0.1 | 0.0 | 0.1 | -   | 0.0  | 0.0  | 0.1  | -   | -   | -   | -   | -   | -   | -   | -   | 0.7   |   |
| ㅅ | 0.3  | 0.0 | 1.4  | 0.0 | 0.1 | 0.0 | 0.2 | 0.0 | 0.6  | 0.4  | 0.1  | -   | 0.0 | 0.0 | -   | 0.0 | -   | -   | -   | 3.1   |   |
| ㅅ | 0.9  | *   | 0.2  | *   | 0.4 | *   | 0.1 | *   | 0.0  | 0.5  | 0.1  | *   | 0.0 | 0.0 | -   | -   | -   | -   | *   | 2.2   |   |
| 계 | 24.9 | 0.6 | 13.5 | 3.1 | 8.4 | 0.5 | 6.2 | 0.1 | 14.0 | 13.4 | 13.3 | 0.1 | 0.8 | 0.6 | 0.3 | 0.2 | 0.0 | 0.0 | 0.0 | 100.0 |   |



<표 - 6> 고빈도 음절의 출현 빈도 및 누적 빈도(%)

| 등위 | 음절 | 출현 빈도 | 누적 빈도 | 등위 | 음절 | 출현 빈도 | 누적 빈도 |
|----|----|-------|-------|----|----|-------|-------|
| 1  | 그  | 3.4   | 3.4   | 19 | 고  | 1.2   | 34.6  |
| 2  | 가  | 3.1   | 6.5   | 20 | 구  | 1.1   | 35.8  |
| 3  | 이  | 2.7   | 9.2   | 21 | 썩  | 1.1   | 36.9  |
| 4  | 나  | 2.4   | 11.6  | 22 | 라  | 1.1   | 38.0  |
| 5  | 대  | 2.3   | 13.9  | 23 | 래  | 1.1   | 39.1  |
| 6  | 는  | 2.1   | 16.0  | 24 | 야  | 1.1   | 40.1  |
| 7  | 니  | 2.0   | 18.0  | 25 | 마  | 1.0   | 41.2  |
| 8  | 거  | 1.8   | 19.8  | 26 | 자  | 1.0   | 42.2  |
| 9  | 아  | 1.6   | 21.4  | 27 | 애  | 0.9   | 43.1  |
| 10 | 지  | 1.5   | 22.9  | 28 | 여  | 0.9   | 44.0  |
| 11 | 개  | 1.4   | 24.4  | 29 | 어  | 0.8   | 44.8  |
| 12 | 하  | 1.4   | 25.7  | 30 | 시  | 0.8   | 45.6  |
| 13 | 러  | 1.4   | 27.1  | 31 | 까  | 0.8   | 46.4  |
| 14 | 서  | 1.3   | 28.4  | 32 | 해  | 0.7   | 47.2  |
| 15 | 다  | 1.3   | 29.7  | 33 | 드  | 0.7   | 47.9  |
| 16 | 기  | 1.3   | 31.0  | 34 | 도  | 0.7   | 48.6  |
| 17 | 리  | 1.2   | 32.2  | 35 | 면  | 0.7   | 49.3  |
| 18 | 내  | 1.2   | 33.4  | 36 | 사  | 0.7   | 50.0  |

#### 나. 중성과 종성(V+C)의 두음소 빈도

다음은 한 음절 내에서 관찰되는 중성과 종성의 전이 빈도를 알아보았다. <표 - 3>에서 보듯이 논리적으로 가능한 두음소 총 126종(=18×7) 중에서 관찰된 두음소의 종류는 모두 98종이었다. 하지만 관찰된 98종의 두음소 중에서 6.3%에 불과한 8종의 두음소가 전체 발화 자료의 50% 이상을 차지하는 것으로 나타났다. <표 - 7>은 누적 빈도 50% 이상을 차지하는 고빈도 V+C 두음소의 목록과 출현 빈도를, <표 - 8>은 V+C 두음소 전체의 출현 빈도를 보인 것이다.

&lt;표 - 7&gt; 고빈도 V+C 두음소의 출현 빈도 및 누적 빈도(%)

| 등위 | V+C | 출현 빈도 | 누적 빈도 | 등위 | V+C | 출현 빈도 | 누적 빈도 |
|----|-----|-------|-------|----|-----|-------|-------|
| 1  | ㄴ+ㄴ | 13.3  | 13.3  | 5  | ㄹ+ㄴ | 4.9   | 38.0  |
| 2  | ㄷ+ㄴ | 9.0   | 22.3  | 6  | ㄷ+ㅇ | 4.8   | 42.7  |
| 3  | ㄴ+ㄹ | 5.6   | 27.9  | 7  | ㄷ+ㄹ | 4.3   | 47.1  |
| 4  | ㄱ+ㄴ | 5.2   | 33.0  | 8  | ㄱ+ㄴ | 3.9   | 50.9  |

&lt;표 - 8&gt; V+C 두음소 전체 빈도(%) (- : 미출현 두음소)

|   | ㄱ   | ㄴ    | ㄷ   | ㄹ    | ㅁ   | ㅂ   | ㅇ    | 계     |
|---|-----|------|-----|------|-----|-----|------|-------|
| ㄷ | 2.9 | 9.0  | 0.1 | 4.3  | 2.7 | 0.4 | 4.8  | 24.1  |
| ㄷ | 0.0 | 0.0  | 0.0 | 0.0  | 0.0 | 0.0 | 1.0  | 1.1   |
| ㄱ | 0.4 | 5.1  | 0.1 | 1.7  | 1.5 | 1.4 | 2.9  | 13.2  |
| ㄱ | 0.2 | 3.9  | 0.0 | 0.8  | 0.1 | 0.1 | 2.4  | 7.4   |
| ㄱ | 0.8 | 1.2  | 0.1 | 1.4  | 1.0 | 0.1 | 2.4  | 7.1   |
| ㄱ | 0.0 | 0.0  | -   | 0.0  | -   | -   | 0.3  | 0.3   |
| ㄱ | 0.3 | 1.7  | 0.0 | 1.2  | 0.2 | 0.0 | 1.1  | 4.4   |
| ㄱ | 0.1 | 0.1  | -   | 0.0  | 0.0 | 0.0 | 0.1  | 0.2   |
| ㄱ | 0.1 | 13.3 | 0.0 | 5.6  | 1.6 | 0.1 | 3.5  | 24.2  |
| ㄱ | 0.5 | 4.9  | 0.0 | 2.5  | 1.3 | 0.6 | 1.1  | 11.0  |
| ㄱ | 0.3 | 1.6  | 0.0 | 0.3  | 0.1 | 0.0 | 1.9  | 4.3   |
| ㄱ | -   | 0.2  | -   | 0.0  | -   | -   | -    | 0.2   |
| ㄱ | 0.0 | 0.4  | 0.0 | 0.1  | 0.0 | 0.0 | 0.1  | 0.5   |
| ㄱ | 0.1 | 0.5  | 0.0 | 0.1  | -   | -   | 0.3  | 1.0   |
| ㄱ | -   | 0.4  | -   | 0.5  | 0.0 | -   | 0.0  | 0.8   |
| ㄱ | -   | 0.0  | 0.0 | 0.0  | 0.0 | 0.0 | 0.0  | 0.1   |
| ㄱ | -   | -    | -   | -    | -   | -   | -    | 0.0   |
| ㄱ | -   | 0.0  | -   | -    | -   | -   | -    | 0.0   |
| 계 | 5.9 | 42.3 | 0.3 | 18.5 | 8.5 | 2.8 | 21.7 | 100.0 |

최고빈도를 보인 두음소 'ㄴ+ㄴ'은 최고빈도 중성과 최고빈도 중성의 결합이 아니었다. 최고 빈도 중성과 최고빈도 중성의 결합인 'ㄷ+ㄴ'은 두 번째로 빈번하게 관찰된 두음소였다. 최고빈도 초성과 최고빈도 중성의 결합이 만들어 낸 두음소보다 'ㄴ+ㄴ'이 최고빈도 두음소로 관찰된 이유는 이

음소의 결합이 문법 형태소인 조사 ‘는’, 어미 ‘-는’, ‘-은’에 존재하기 때문인 것으로 보인다. <표 - 6>에 보였듯이, 폐음절 유형 중에서 가장 높은 빈도를 보인 음절형은 ‘는’이었고, 다음은 ‘면’이었는데, ‘-+ㄴ’는 최고빈도를, ‘ㄷ+ㄴ’은 빈도 순위 8위를 보였다.

## 2. 음절 경계를 사이에 두고 나타나는 두음소의 빈도

다음은 음절 경계를 사이에 두고 나타나는 두음소의 빈도를 알아보았다. 이 위치에서 가능한 두음소의 유형으로는 C\$C, C\$V, V\$C, V\$V 등이 있다. <표 - 9>에 정리되어 있듯이, 이중 유형 빈도가 가장 높은 것은 V\$C형이었고, 가장 유형 빈도가 낮은 것은 C\$V형이었다. 반면에 가장 출현 빈도가 높은 것은 V\$C형이었고, 가장 출현 빈도가 낮은 것은 C\$V형이었다. V\$C형의 출현 빈도가 가장 높고, V\$C형의 출현 빈도가 가장 낮은 이유는 앞에서 살펴본 음절형의 출현 빈도와 유관하다.

<표 - 9> 음절 경계를 사이에 두고 나타나는 두음소의 유형 빈도와 출현 빈도

|                 | C\$C             | C\$V        | V\$C          | V\$V           |
|-----------------|------------------|-------------|---------------|----------------|
| 가능 종류 수         | 86 (= (7×18)-40) | 17 (= 1×17) | 324 (= 18×18) | 306 (= (18×17) |
| 출현 종류 수(비율)     | 86 (100%)        | 16 (94.1%)  | 292 (90.1%)   | 225 (73.5%)    |
| 상위 50% 종류 수(비율) | 9 (10.5%)        | 2 (12.5%)   | 21 (7.2%)     | 16 (5.2%)      |
| 출현 총수           | 71,264           | 4,576       | 179,834       | 24,220         |

즉, 앞에서 살펴본 바와 같이, 폐음절보다는 개음절이, 초성이 없는 음절보다는 있는 음절이 훨씬 출현 빈도가 높게 나타났기 때문에, 음절을 경계에 두고 앞 음절은 종성이 없어 모음으로 끝나고, 뒤 음절은 초성이 있어 자음으로 시작하는 음절형의 연쇄가 높은 빈도를 보일 것이고, 따라서 음절 경계를 사이에 두고 가장 높은 출현 빈도를 보일 것으로 예상되는 두음소 유형은 당연히 V\$C형이 될 것이다. 또한, C\$V형의 유형 빈도와 출현 빈도가 낮은 이유는 강세구 내에서 자음과 모음의 연쇄는 초성이 될 수 없는 /o/을 제외하고는 모두 초성으로 음절화가 되기 때문으로 설명이 가능하다. 한편, V\$V형이 유형 빈도가 높은 편에 속하지만, 출현 빈도가 낮은 이유는 히아투스(hiatus) 회피 현상과 관련된 것으로 설명할 수 있다.

### 가. 종성과 초성(C\$C)의 두음소 빈도

음절의 경계를 사이에 두고 나타나는 자음과 자음의 전이 빈도, 즉 음절 경계를 사이에 두고 나타난 앞 음절의 종성과 다음 음절의 초성이 연결되는 C\$C 두음소의 빈도를 알아보았다. <표 - 9>

에서 보듯이 논리적으로 가능한 두음소의 종류는 모두 86종(=(7×18) - 40<sup>5</sup>)이었고, 86종 모두 발화 자료에서 관찰되었다. 하지만 그중에서 10.5% 에 불과한 9종의 두음소가 전체 발화 자료의 50% 이상을 차지하는 것으로 나타났다. <표 - 10>은 누적 빈도 50% 이상을 차지하는 고빈도 C\$C 두음소의 목록과 출현 빈도를, <표 - 11>는 C\$C 두음소 전체의 출현 빈도를 정리한 것이다.

<표 - 10> 고빈도 C\$C 두음소의 출현 빈도 및 누적 빈도(%)

| 등위 | C\$C | 출현 빈도 | 누적 빈도 | 등위 | C\$C | 출현 빈도 | 누적 빈도 |
|----|------|-------|-------|----|------|-------|-------|
| 1  | ㄴ\$ㄷ | 13.1  | 13.1  | 6  | ㄴ\$ㅈ | 4.0   | 43.7  |
| 2  | ㅇ\$ㄱ | 9.0   | 22.1  | 7  | ㄴ\$ㅅ | 3.5   | 47.3  |
| 3  | ㄴ\$ㄴ | 7.6   | 29.7  | 8  | ㅇ\$ㅎ | 2.4   | 49.7  |
| 4  | ㄹ\$ㄹ | 5.5   | 35.2  | 9  | ㄴ\$ㅊ | 2.3   | 52.0  |
| 5  | ㄴ\$ㄱ | 4.6   | 39.7  |    |      |       |       |

<표 - 11> C\$C 두음소 전체 빈도(%) (\*: 출현 불가능 두음소, -: 미출현 두음소)

|   | ㄱ    | ㄲ   | ㅋ   | ㆁ   | ㆅ   | 표   | ㄷ    | ㅌ   | ㅍ   | ㅊ   | ㅌ   | ㅈ   | ㅊ   | ㅊ   | ㅈ   | ㅈ   | ㅈ    | ㅈ   | ㅈ     | 계   |
|---|------|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-------|-----|
| ㄱ | *    | *   | *   | *   | 0.4 | 0.1 | *    | 0.7 | 0.1 | *   | 1.0 | 0.1 | *   | 1.4 | *   | *   | *    | *   | *     | 3.9 |
| ㄴ | 4.6  | 0.5 | 0.3 | 1.1 | 0.1 | 0.2 | 13.1 | 0.4 | 1.3 | 4.0 | 2.3 | 1.0 | 3.5 | 0.4 | 1.4 | 1.4 | 7.6  | *   | 42.9  |     |
| ㄷ | *    | 0.1 | 0.0 | *   | 0.0 | 0.0 | *    | *   | *   | *   | *   | *   | *   | *   | *   | *   | *    | *   | *     | 0.1 |
| ㄹ | 1.9  | 1.8 | 0.1 | 0.8 | 0.1 | 0.2 | 1.0  | 1.4 | 0.3 | 0.5 | 0.9 | 0.3 | 0.3 | 1.6 | 1.3 | 1.4 | *    | 5.5 | 19.5  |     |
| ㅁ | 0.4  | 0.2 | 0.0 | 1.0 | 0.1 | 0.2 | 1.3  | 0.1 | 0.1 | 1.0 | 0.2 | 0.2 | 0.8 | 0.1 | 0.3 | 1.2 | 1.1  | *   | 8.3   |     |
| ㅂ | *    | 0.7 | 0.0 | *   | *   | *   | *    | 0.7 | 0.0 | *   | 0.7 | 0.1 | *   | 1.3 | *   | *   | *    | *   | *     | 3.5 |
| ㅇ | 9.0  | 0.7 | 0.2 | 0.8 | 0.0 | 0.1 | 1.3  | 0.1 | 0.2 | 1.0 | 0.1 | 0.4 | 1.6 | 0.2 | 2.4 | 1.5 | 2.2  | *   | 21.8  |     |
| 계 | 15.9 | 3.9 | 0.7 | 3.7 | 0.8 | 0.8 | 16.7 | 3.4 | 1.9 | 6.5 | 5.3 | 2.1 | 6.2 | 5.0 | 5.4 | 5.5 | 10.8 | 5.5 | 100.0 |     |

특징적인 것은, 고빈도를 보인 C\$C 두음소 유형 중에는 동기관적(homorganic)인 두음소가 대부분을 차지하고 있다는 점이다. 고빈도 9개의 유형 중에서 'ㄴ\$ㄷ', 'ㅇ\$ㄱ', 'ㄴ\$ㄴ', 'ㄹ\$ㄹ', 'ㄴ\$ㅈ', 'ㄴ\$ㅅ', 'ㄴ\$ㅊ' 등 7개의 유형이 같은 조음 위치에서 산출되는 동기관적 자음의 연쇄였다. 고빈도 9개 유형 중에서 동기관적이지 않은 유형은 'ㄴ+ㄱ'(5위)와 'ㅇ+ㅎ'(8위)의 두 개 유형이었다.

5) 한국어에서 올 수 없는 자음 결합의 수를 뺀 것이다. 구체적으로 어떤 연쇄가 불가능한지는 <표 - 9>에 표시되어 있다.

### 나. 중성과 중성(C\$V)의 두음소 빈도

음절 경계를 두고 나타난 자음과 모음의 연쇄 빈도를 알아보았다. <표-9>에 정리된 바와 같이 출현 가능한 두음소 종류는 모두 17(=1×17<sup>6</sup>)종이었고, 한 종류를 제외한 16종의 두음소가 자료에서 관찰되었다. 관찰된 16종의 두음소 중에서 단 2종의 두음소, ‘ㅇ+ㅣ’, ‘ㅇ+ㅡ’가 전체 자료의 50% 이상을 차지하였다. C\$V 두음소의 전체 출현 빈도와 누적 빈도에 대한 자세한 내용은 <표-12>에 정리되어 있다. 앞에서 논의하였듯이, C\$V형의 유형 빈도와 출현 빈도가 낮은 이유는 강세구 내에서 자음과 모음의 연쇄는 초성이 될 수 없는 /ㅇ/을 제외하고는 모두 초성으로 음절화가 되기 때문이다.

<표-12> 고빈도 C\$V 두음소의 전체 출현 빈도 및 누적 빈도(%) (- : 미출현 두음소)

| 등위 | CV  | 출현 빈도 | 누적 빈도 | 등위 | CV  | 출현 빈도 | 누적 빈도 |
|----|-----|-------|-------|----|-----|-------|-------|
| 1  | ㅇ+ㅣ | 37.8  | 37.8  | 10 | ㅇ+ㄱ | 1.6   | 97.2  |
| 2  | ㅇ+ㅡ | 16.2  | 54.0  | 11 | ㅇ+ㄴ | 1.5   | 98.7  |
| 3  | ㅇ+ㅐ | 14.3  | 68.3  | 12 | ㅇ+ㅓ | 0.6   | 99.3  |
| 4  | ㅇ+ㅑ | 10.6  | 78.9  | 13 | ㅇ+ㅕ | 0.3   | 99.5  |
| 5  | ㅇ+ㅓ | 4.1   | 83.0  | 14 | ㅇ+ㅗ | 0.2   | 99.7  |
| 6  | ㅇ+ㅕ | 3.9   | 86.9  | 15 | ㅇ+ㅛ | 0.2   | 99.9  |
| 7  | ㅇ+ㅗ | 3.4   | 90.4  | 16 | ㅇ+ㅜ | 0.1   | 100.0 |
| 8  | ㅇ+ㅛ | 2.8   | 93.2  | 17 | ㅇ+ㅠ | -     | 100.0 |
| 9  | ㅇ+ㅜ | 2.4   | 95.6  |    |     |       |       |

<표-12>에서 한 가지 특기할 것은, 고빈도 두음소 유형을 보이는 예가 대부분 문법 형태소와 관련이 있다는 점이다. 1, 2, 3위를 차지한, 누적 빈도 68.3%를 보인, ‘ㅇ+ㅣ’, ‘ㅇ+ㅡ’, ‘ㅇ+ㅐ’는 ‘이’, ‘이다’, ‘은’, ‘을’, ‘으로’, ‘에’, ‘에서’, ‘에게’ 등의 조사가 연결되는 환경과 관련한 것으로 보인다.

### 다. 중성과 초성(V\$C)의 두음소 빈도

다음은 음절의 경계를 사이에 두고 나타나는 선행 음절의 중성과 후행 음절의 초성으로 구성된 두음소의 빈도를 알아보았다. 논리적으로 가능한 두음소의 종류는 324종(=18×18)이었는데, <표-9>에서 보듯이 자료에서 관찰된 V\$C 두음소의 종류는 292종(90.1%)이었다. 그중 8.2%에 불과한 20종의 두음소가 전체 발화 자료의 50% 이상을 차지하는 것으로 나타났다. 결국, 논리적으로 가능한 두음소 중에서 발화에서 관찰된 두음소의 유형 비율은 90.1%로 높았지만, 그중에서 8.2%에 불과한 특정 두음

6) 하나의 강세구 안에서 자음이 후행 모음의 초성이 되지 못하는 경우는 /ㅇ/이 모음에 선행하는 경우이다. 따라서 자음의 종류 수는 1이고, 후행하는 모음의 종류 수는 삼중모음 이외에는 모두 올 수 있기 때문에 17이다. 따라서 출현 가능 두음소의 총 종류 수는 1×17이 된다.

소가 발화에서 반복적으로 나타난다는 것을 알 수 있다. <표 - 13>은 누적 빈도 50% 이상을 차지하는 고빈도 두음소 목록을 보인 것이고, <표 - 14>는 두음소 전체의 출현 빈도를 정리한 것이다.

<표 - 13> 고빈도 V\$C 두음소의 출현 빈도 및 누적 빈도(%)

| 등위 | V\$C | 출현 빈도 | 누적 빈도 | 등위 | V\$C | 출현 빈도 | 누적 빈도 |
|----|------|-------|-------|----|------|-------|-------|
| 1  | ㅏ\$ㄴ | 6.7   | 6.7   | 12 | ㅣ\$ㄴ | 1.7   | 38.0  |
| 2  | ㅡ\$ㄹ | 6.2   | 12.9  | 13 | ㅏ\$ㅇ | 1.6   | 39.6  |
| 3  | ㅏ\$ㄱ | 4.2   | 17.1  | 14 | ㅡ\$ㄱ | 1.4   | 41.1  |
| 4  | ㅣ\$ㄱ | 3.2   | 20.3  | 15 | ㅓ\$ㄹ | 1.4   | 42.4  |
| 5  | ㅣ\$ㄹ | 2.9   | 23.2  | 16 | ㅏ\$ㅇ | 1.4   | 43.8  |
| 6  | ㅏ\$ㄹ | 2.8   | 26.0  | 17 | ㅣ\$ㅇ | 1.3   | 45.2  |
| 7  | ㅍ\$ㄱ | 2.3   | 28.4  | 18 | ㅍ\$ㄷ | 1.3   | 46.5  |
| 8  | ㅏ\$ㅈ | 2.3   | 30.7  | 19 | ㅏ\$ㄷ | 1.3   | 47.8  |
| 9  | ㅓ\$ㄱ | 2.0   | 32.7  | 20 | ㅡ\$ㄴ | 1.3   | 49.1  |
| 10 | ㅓ\$ㄴ | 1.9   | 34.6  | 21 | ㅣ\$ㅇ | 1.2   | 50.3  |
| 11 | ㅍ\$ㅅ | 1.8   | 36.3  |    |      |       |       |

<표 - 14> V\$C 두음소 전체 빈도(%) (- : 미출현 두음소)

|   | ㄱ    | ㅋ   | ㆁ   | ㄷ   | ㅌ   | ㅍ   | ㅊ   | ㅌ   | ㅍ   | ㅈ   | ㅊ   | ㅊ   | ㅊ   | ㅊ   | ㅊ   | ㅊ   | ㅊ    | ㅊ    | ㅊ     | 계 |
|---|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|-------|---|
| ㅏ | 4.2  | 1.4 | 0.3 | 0.6 | 0.2 | 0.2 | 1.3 | 0.4 | 0.8 | 2.3 | 0.2 | 0.5 | 1.1 | 0.5 | 1.6 | 6.7 | 2.8  | 25.7 |       |   |
| ㅣ | 3.2  | 1.3 | 0.4 | 0.7 | 0.1 | 0.2 | 0.9 | 0.3 | 0.2 | 1.1 | 0.3 | 0.2 | 0.7 | 1.0 | 0.9 | 1.2 | 1.7  | 2.9  | 17.5  |   |
| ㅓ | 2.0  | 0.3 | 0.9 | 0.5 | 0.1 | 0.1 | 0.8 | 0.7 | 0.2 | 0.7 | 0.3 | 0.2 | 0.4 | 0.5 | 0.3 | 1.1 | 1.9  | 1.1  | 11.9  |   |
| ㅍ | 2.3  | 0.2 | 0.1 | 0.4 | 0.0 | 0.1 | 1.3 | 0.6 | 0.1 | 0.6 | 0.3 | 0.1 | 1.8 | 0.7 | 0.5 | 0.8 | 1.0  | 0.6  | 11.6  |   |
| ㅡ | 1.4  | 0.3 | 0.3 | 0.1 | 0.0 | 0.1 | 0.2 | 0.3 | 0.3 | 0.3 | 0.1 | 0.1 | 0.3 | 0.0 | 0.0 | 1.0 | 1.3  | 6.2  | 12.1  |   |
| ㅓ | 1.1  | 0.2 | 0.3 | 0.2 | 0.2 | 0.1 | 0.6 | 0.1 | 0.4 | 0.3 | 0.0 | 0.1 | 0.5 | 0.1 | 0.3 | 0.6 | 0.9  | 0.9  | 6.8   |   |
| ㅓ | 1.0  | 0.2 | 0.0 | 0.2 | 0.0 | 0.0 | 0.4 | 0.0 | 0.2 | 0.3 | 0.0 | 0.1 | 0.6 | 0.0 | 0.3 | 0.3 | 1.1  | 1.4  | 6.3   |   |
| ㅊ | 0.1  | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.2 | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0  | 0.1  | 0.7   |   |
| ㅋ | 0.4  | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.0 | 0.4 | 0.1 | 0.0 | 0.1 | 0.3 | 0.1 | 0.0 | 0.8  | 0.2  | 2.7   |   |
| ㅌ | 0.2  | 0.1 | 0.0 | 0.0 | -   | 0.0 | 0.3 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0  | 0.0  | 0.8   |   |
| ㅍ | 0.1  | 0.0 | 0.0 | 0.0 | -   | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | -   | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1  | 0.1  | 0.4   |   |
| ㅍ | 0.5  | 0.0 | -   | 0.0 | 0.0 | 0.0 | 0.0 | -   | -   | 0.0 | -   | -   | 0.0 | 0.0 | 0.0 | 0.0 | 0.0  | 0.0  | 0.7   |   |
| ㄱ | 0.0  | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0  | 0.0  | 0.3   |   |
| ㅊ | 0.5  | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.2  | 0.0  | 1.2   |   |
| ㅊ | 0.1  | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.1  | 0.1  | 0.9   |   |
| ㅊ | 0.0  | 0.0 | 0.0 | 0.0 | 0.0 | -   | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | -   | 0.0 | 0.0 | 0.0 | 0.0 | 0.2  | 0.1  | 0.4   |   |
| ㅊ | 0.0  | -   | -   | -   | -   | -   | 0.0 | -   | -   | 0.0 | -   | -   | 0.0 | -   | 0.0 | 0.0 | 0.0  | -    | 0.0   |   |
| ㅊ | 0.0  | 0.0 | -   | 0.0 | -   | -   | 0.0 | 0.0 | -   | -   | 0.0 | -   | -   | 0.0 | -   | -   | -    | -    | 0.0   |   |
| 계 | 17.3 | 4.2 | 2.4 | 2.8 | 0.7 | 1.0 | 6.0 | 2.6 | 2.2 | 6.6 | 1.4 | 1.4 | 5.8 | 3.4 | 3.1 | 6.8 | 16.0 | 16.4 | 100.0 |   |

<표 - 13>에서 주목할 점은, 고빈도 V\$C 유형에서 관찰되는 선행 모음의 환경은 대체로 고빈도 모음형이었고, 후행하는 자음의 환경은 대체로 고빈도 자음형이었다. 선행하는 모음이 고빈도형 상위 4개 유형인 /ㅏ, ㅣ, ㅓ, ㅡ/인 경우가 전체의 67.2%, 후행하는 자음(초성)이 고빈도 상위 3개 유형인 /ㄱ, ㄴ, ㄷ/인 경우가 전체의 49.8%를 차지하는 것으로 나타났다. 한편, 두음소 고빈도 유형에 저빈도 초성인 /ㄱ/ 연결형이 존재하는 이유는 표면 음성형에서 경음화 후 동일 조음 위치 폐쇄음 탈락이 일어난 유형, 혹은 수의적인 자음 위치 동화가 일어난 후 동일 조음 위치 폐쇄음 탈락이 일어난 유형이 다수 존재하는 것과 관련한 것으로 보인다<sup>7)</sup>.

### 라. 중성과 중성(V\$V)의 전이 빈도

다음은 음절 경계를 사이에 두고 나타나는 모음과 모음의 전이 빈도를 알아보았다. <표 - 9>에서 보듯이 논리적으로 가능한 V\$V 두음소의 종류는 모두 306종(=18×17<sup>8)</sup>)이었으나, 이 중에서 자료에서 관찰된 두음소의 종류는 총 225종으로 가능한 두음소의 종류 중에서 73.5%가 자료에서 관찰되었다. 그리고 관찰된 225종의 두음소 가운데 5.2%에 불과한 16종의 두음소가 전체 발화 자료의 50% 이상을 차지하는 것으로 나타났다. <표 - 15>은 누적 빈도 50% 이상을 차지하는 고빈도 두음소 목록을 보인 것이고, <표 - 16>은 두음소 전체의 출현 빈도를 정리한 것이다.

<표 - 15> 고빈도 V\$V 두음소의 출현 빈도 및 누적 빈도(%)

| 등위 | V\$V | 출현 빈도 | 누적 빈도 | 등위 | V\$V | 출현 빈도 | 누적 빈도 |
|----|------|-------|-------|----|------|-------|-------|
| 1  | ㅏ\$ㅏ | 5.8   | 5.8   | 9  | ㅓ\$ㅏ | 2.7   | 34.4  |
| 2  | ㅏ\$ㅣ | 4.9   | 10.7  | 10 | ㅓ\$ㅣ | 2.6   | 37.1  |
| 3  | ㅏ\$ㅓ | 4.4   | 15.1  | 11 | ㅓ\$ㅓ | 2.4   | 39.4  |
| 4  | ㅣ\$ㅏ | 4.4   | 19.5  | 12 | ㅓ\$ㄱ | 2.3   | 41.8  |
| 5  | ㅏ\$ㄱ | 3.5   | 23.0  | 13 | ㅓ\$ㄱ | 2.3   | 44.1  |
| 6  | ㅣ\$ㅏ | 3.0   | 26.0  | 14 | ㅓ\$ㄴ | 2.2   | 46.4  |
| 7  | ㅣ\$ㅓ | 2.9   | 28.9  | 15 | ㅏ\$ㅏ | 2.1   | 48.4  |
| 8  | ㅣ\$ㅓ | 2.9   | 31.8  | 16 | ㅓ\$ㅣ | 2.0   | 50.4  |

7) 전자의 예로는 /그래 갖꾸/ → [그래가꾸], 후자의 예로는 /찾거나/→[차꺼나] 등을 들 수 있다. 특히, ‘ㅏ+ㅓ’형은 담화표지로 출현 빈도가 높은 ‘갖꾸[가꾸]’와도 관련한 것으로 보인다.

8) 논리적으로 가능한 두음소의 종류 수는 ‘중성 가능 모음의 수×중성 가능 모음의 수’, 즉 ‘18×17’이 된다. 여기서 후행 중성 가능 모음의 수가 17인 이유는 이 위치에서 후행하는 모음이 삼중모음인 경우를 생각할 수 없기 때문이다.

&lt;표 - 16&gt; V\$V 두음소 전체 빈도(%) (\* : 출현 불가능 두음소, - : 미출현 두음소)

|   | ㅏ    | ㅑ    | ㅓ   | ㅕ   | ㅗ   | ㅛ   | ㅜ   | ㅠ    | ㅡ   | ㅣ    | ㅞ   | ㅝ   | ㅚ   | ㅜ   | ㅟ   | ㅠ   | ㅡ   | ㅢ | ㅣ | ㅤ    | 계 |
|---|------|------|-----|-----|-----|-----|-----|------|-----|------|-----|-----|-----|-----|-----|-----|-----|---|---|------|---|
| ㅏ | 1.4  | 4.9  | 0.5 | 1.7 | 1.6 | 2.3 | 1.0 | 2.2  | 0.5 | 1.7  | 1.8 | 0.4 | 0.1 | 0.4 | 0.3 | 0.0 | -   | * |   | 20.7 |   |
| ㅑ | 2.9  | 1.4  | 3.0 | 2.9 | 0.1 | 0.5 | 0.3 | 1.1  | 0.3 | 4.4  | 0.2 | 0.6 | 0.1 | 0.2 | 0.2 | 0.1 | -   | * |   | 18.2 |   |
| ㅓ | 0.9  | 2.0  | 0.4 | 1.0 | 0.8 | 0.6 | 0.9 | 2.3  | 0.6 | 5.8  | 0.5 | 0.7 | 0.0 | 0.2 | 0.0 | 0.0 | -   | * |   | 16.8 |   |
| ㅕ | 1.2  | 4.4  | 1.0 | 0.9 | 0.0 | 0.2 | 0.6 | 3.5  | 0.9 | 2.1  | 0.2 | 0.5 | 0.2 | 0.1 | 0.1 | 0.0 | 0.0 | * |   | 16.0 |   |
| ㅗ | 0.1  | 0.4  | 0.0 | 0.5 | 0.0 | 0.1 | 0.0 | 0.2  | 0.0 | 0.1  | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 | * |   | 1.9  |   |
| ㅛ | 2.7  | 2.6  | 0.4 | 0.7 | 1.0 | 0.2 | 0.2 | 0.7  | 0.2 | 0.6  | 0.3 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | -   | * |   | 9.9  |   |
| ㅜ | 0.5  | 2.4  | 1.5 | 0.9 | 0.0 | 0.1 | 0.2 | 1.3  | 0.2 | 0.3  | 0.2 | 0.3 | -   | 0.0 | 0.0 | 0.1 | 0.0 | * |   | 7.9  |   |
| ㅠ | 0.0  | 0.2  | 0.0 | 0.0 | -   | 0.1 | 0.4 | 0.1  | 0.0 | 0.3  | 0.1 | 0.2 | -   | -   | 0.0 | -   | -   | * |   | 1.4  |   |
| ㅡ | 0.1  | 0.5  | 0.0 | 0.4 | 0.0 | 0.0 | 0.0 | 0.0  | 0.0 | 0.2  | 0.0 | 0.0 | 0.1 | 0.0 | 0.3 | -   | -   | * |   | 1.7  |   |
| ㅣ | 0.1  | 0.0  | 0.0 | 0.0 | -   | -   | 0.0 | 0.0  | 0.0 | 0.0  | -   | 0.1 | 0.0 | 0.0 | -   | -   | -   | * |   | 0.4  |   |
| ㅞ | 0.1  | 0.2  | 0.0 | 0.3 | -   | 0.0 | 0.1 | 0.2  | 0.1 | 0.3  | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | -   | * |   | 1.4  |   |
| ㅝ | 0.0  | 0.1  | -   | 0.0 | -   | 0.0 | -   | 0.1  | 0.0 | 0.4  | 0.0 | 0.0 | -   | 0.0 | -   | -   | -   | * |   | 0.6  |   |
| ㅚ | 0.1  | 0.4  | 0.1 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1  | 0.1 | 0.0  | 0.0 | 0.1 | 0.0 | 0.0 | -   | -   | -   | * |   | 1.2  |   |
| ㅜ | 0.0  | 0.1  | 0.0 | -   | -   | 0.0 | -   | 0.2  | 0.0 | 0.0  | -   | -   | 0.0 | 0.0 | -   | -   | -   | * |   | 0.3  |   |
| ㅠ | 0.0  | 0.1  | 0.0 | 0.1 | -   | 0.1 | -   | 0.0  | 0.0 | 0.0  | 0.0 | -   | 0.0 | -   | -   | -   | -   | * |   | 0.4  |   |
| ㅡ | 0.1  | 0.0  | 0.3 | 0.4 | -   | -   | 0.1 | 0.1  | -   | 0.0  | -   | 0.0 | -   | -   | -   | 0.0 | -   | * |   | 1.1  |   |
| ㅣ | 0.0  | -    | -   | -   | -   | -   | -   | 0.0  | -   | -    | 0.0 | 0.0 | -   | -   | -   | -   | -   | * |   | 0.0  |   |
| ㅞ | -    | -    | -   | -   | -   | -   | -   | -    | -   | 0.0  | -   | -   | -   | -   | -   | -   | -   | * |   | 0.0  |   |
| 계 | 10.2 | 19.5 | 7.4 | 9.9 | 3.6 | 4.3 | 3.8 | 12.2 | 3.0 | 16.3 | 3.3 | 3.1 | 0.8 | 1.0 | 1.1 | 0.4 | 0.0 | * |   | 100  |   |

<표 15>에서 보듯이, 가장 높은 빈도를 보인 두음소는 ‘ㄱ\$ㅏ’였으며, 그 다음으로는 ‘ㅏ\$ㅣ’, ‘ㅞ\$ㅣ’, ‘ㅣ\$ㅏ’, ‘ㅞ\$ㅏ’, ‘ㅣ\$ㅏ’의 순이었다. 특징적인 것은 고빈도 결합형 중에 ‘고빈도 중성 + 저빈도 중성’의 연결이 많다는 것이다. 물론, 그 이유는 모음끼리의 연쇄를 회피하는 히아투스(hiatus) 회피 현상과 관련된 것으로 보이며, 이러한 이유 때문에 후행 음절의 중성이 저빈도이기는 하지만 이중모음인 경우가 많았다. 특히 이중모음 중에서도 /j/계 이중모음이 둘째 음절의 중성인 경우가 높은 빈도를 보였다.

### 3. 강세구 경계에 나타나는 두음소의 빈도

마지막으로 강세구 경계에 나타나는 두음소의 유형 빈도와 출현 빈도를 알아보았다. 이 위치에서 가능한 두음소의 유형에는 #+음소, 음소+#의 두 형태가 있다. 이 두음소 빈도는 강세구 초와 말에 나타나는 음소의 빈도와 같다. <표 - 17>은 강세구 초와 말에 나타나는 음소, 즉 휴지와 연결되는 두음소의 유형 빈도와 출현 빈도를 정리한 것이다.



<표 - 17> 강세구 경계에 나타나는 두음소의 유형 빈도와 출현 빈도

|                 | #+음소     | 음소+#     |
|-----------------|----------|----------|
| 가능 종류 수         | 35       | 25       |
| 출현 종류 수(비율)     | 35(100%) | 25(100%) |
| 상위 50% 종류 수(비율) | 5(14.3%) | 4(16%)   |
| 출현 총수           | 122,919  | 122,919  |

### 가. 강세구 초 음소(두음소, ‘#+음소’)의 빈도

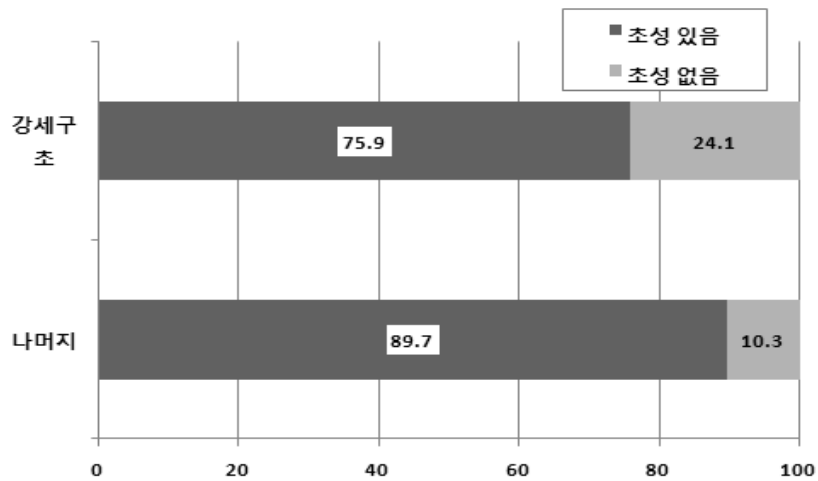
우선 강세구 초의 음소 빈도, 즉 두음소 ‘#+음소’의 빈도는 <표 - 17>에 보인 바와 같다. 논리적으로 가능한 종류는 모두 35종(= 18 +17<sup>9)</sup>)이었으며, 35종 모두 발화에서 관찰되었다. 하지만 이 중에서 5종(14.3%)이 전체 자료의 50.6%를 차지하는 것으로 나타났다. <표 - 18>은 강세구 초에 나타난 두음소를 빈도순으로 보인 것이다.

<표 - 18> #+음소 두음소의 전체 출현 빈도 및 누적 빈도(%) (음영: 누적 빈도 50%선)

| 등위 | #+음소 | 출현 빈도 | 누적 빈도 | 등위 | #+음소 | 출현 빈도 | 누적 빈도 |
|----|------|-------|-------|----|------|-------|-------|
| 1  | #+ㄱ  | 19.2  | 19.2  | 19 | #+ㅍ  | 0.9   | 91.6  |
| 2  | #+ㅋ  | 8.7   | 27.9  | 20 | #+ㅑ  | 0.9   | 92.5  |
| 3  | #+ㆁ  | 8.5   | 36.4  | 21 | #+ㅓ  | 0.9   | 93.3  |
| 4  | #+ㄴ  | 7.8   | 44.2  | 22 | #+ㅕ  | 0.8   | 94.1  |
| 5  | #+ㄷ  | 6.5   | 50.6  | 23 | #+ㅗ  | 0.8   | 94.9  |
| 6  | #+ㅌ  | 6.3   | 56.9  | 24 | #+ㅛ  | 0.8   | 95.7  |
| 7  | #+ㄷ  | 6.1   | 63.0  | 25 | #+ㅜ  | 0.7   | 96.4  |
| 8  | #+ㅎ  | 5.7   | 68.8  | 26 | #+ㅠ  | 0.6   | 97.0  |
| 9  | #+ㅈ  | 5.6   | 74.3  | 27 | #+ㅡ  | 0.6   | 97.6  |
| 10 | #+ㅊ  | 4.1   | 78.5  | 28 | #+ㅣ  | 0.5   | 98.0  |
| 11 | #+ㄹ  | 3.1   | 81.6  | 29 | #+ㅍ  | 0.4   | 98.5  |
| 12 | #+ㅍ  | 1.9   | 83.5  | 30 | #+ㅑ  | 0.4   | 98.8  |
| 13 | #+ㅋ  | 1.6   | 85.2  | 31 | #+ㅓ  | 0.3   | 99.2  |
| 14 | #+ㄴ  | 1.4   | 86.6  | 32 | #+ㅕ  | 0.3   | 99.5  |
| 15 | #+ㄷ  | 1.2   | 87.8  | 33 | #+ㅗ  | 0.3   | 99.8  |
| 16 | #+ㅌ  | 1.0   | 88.8  | 34 | #+ㅛ  | 0.1   | 99.9  |
| 17 | #+ㄷ  | 1.0   | 89.8  | 35 | #+ㅜ  | 0.1   | 100.0 |
| 18 | #+ㅎ  | 0.9   | 90.7  |    |      |       |       |

9) 초성에 출현 가능한 자음의 수(18)와 자음 없이 나타날 수 있는 중성의 수(17)를 합한 것이다. 자음 중에서는 /ㅇ/이, 모음 중에서는 /ㅣ/가 제외된다.

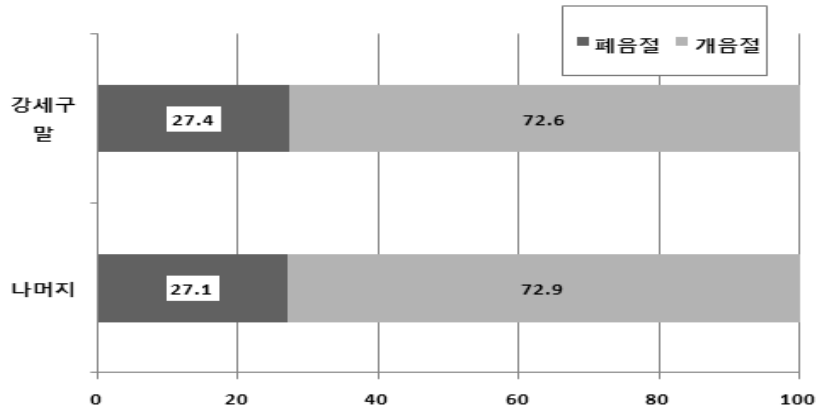
위의 두음소 빈도 중 ‘#+자음’의 빈도는 자음으로 시작하는 강세구를, ‘#+모음’은 모음으로 시작하는 강세구를 의미한다. 강세구의 첫 음절이 초성을 갖는 경우는 그렇지 않은 경우에 비하여 많았다. <그림 - 3>에 보였듯이 강세구 초가 자음으로 시작하는, 즉 강세구의 첫 음절이 초성을 갖는 경우는 75.9%, 강세구 초가 모음으로 시작하는, 즉 강세구의 첫 음절이 초성을 갖지 않는 경우는 24.1%로 나타났다. 하지만 강세구 초는 나머지 위치와 비교할 때, 초성이 있는 음절의 비율이 낮은 것으로 나타났다. 나머지 위치의 경우는 초성이 없는 음절의 비율이 89.7%, 초성이 있는 음절의 비율이 10.3%로 나타나서, 강세구 초의 위치에서 초성이 없는 음절이 나타나는 비율이 나머지 위치에 비하여 약 13.8% 정도 높다는 것을 알 수 있다. 즉, 강세구 초가 다른 운율적 위치에 비하여 초성을 갖지 않는 음절의 비율이 약간 높게 나타난다는 사실을 관찰할 수 있다.



<그림 - 3> 강세구 초와 나머지 위치에서의 초성 자음 유무 음절의 출현 빈도 비교

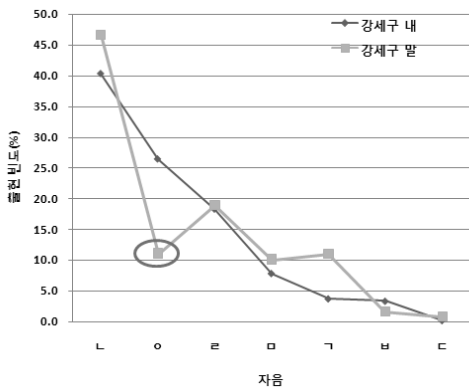
한편, 강세구 초에 나타나는 음소의 빈도와 강세구 내에 나타난 음소의 빈도를 자음과 모음으로 나누어 비교하면 <그림 - 4>와 <그림 - 5>에 나타난 것과 같다. 그림에 보였듯이, 강세구 초에는 강세구 내 초성의 자음과는 달리 /ㄱ, ㅈ, ㅋ, ㅎ/의 빈도가 높았고, /ㄹ, ㄷ, ㅌ/의 빈도가 낮았다. 강세구 초에 위치한 모음의 빈도를 강세구 내에 위치한 모음의 빈도와 비교하면, 강세구 초에는 /ㅣ, ㅝ/의 빈도가 높았고, /ㅡ, ㅜ/의 빈도가 낮았다.



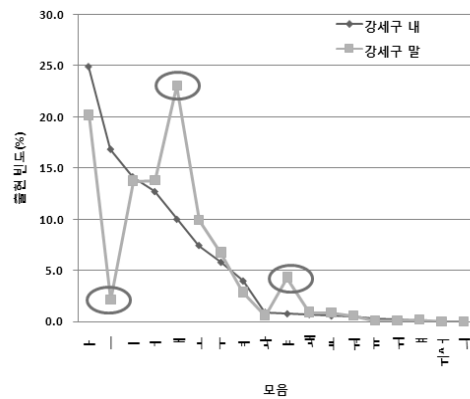


<그림 - 6> 강세구 말과 나머지 위치에서의 종성 자음 유무 음절의 출현 빈도 비교

위의 두음소 빈도 중 ‘자음+#’의 빈도는 자음으로 끝나는 강세구를, ‘모음+#’은 모음으로 끝나는 강세구를 의미한다. 강세구의 끝 음절이 종성을 갖는 경우는 그렇지 않은 경우에 비하여 현저히 낮은 비율을 보였다. <그림 - 6>에 보였듯이, 강세구 말이 개음절인 경우는 72.6%인 데 비하여 강세구 말이 폐음절인 경우는 27.4%에 불과하였다. 하지만 이 비율은 나머지 위치 비교할 때 크게 다르지 않았다. 나머지 위치, 즉 비강세구 말 위치에 나타난 음절의 경우 또한 개음절이 72.9%, 폐음절이 27.1%의 출현 빈도를 보였다.



<그림 - 7> 강세구 초와 강세구 내 초성 자음의 빈도 비교



<그림 - 8> 강세구 초와 강세구 내 모음의 빈도 비교

강세구 말에 나타나는 음소의 빈도와 강세구 내에 나타난 음소의 빈도를 자음과 모음으로 나누어 비교하면 <그림 - 7>과 <그림 - 8>에 나타난 것과 같다. 그림에 보였듯이, 연구개음인 /ㄱ, ㅇ/을 제외하고는 대체로 비슷한 출현 빈도가 관찰되었다. 강세구 말에는 강세구 내 종성의 자음과는 달

리 /ㄱ/의 빈도가 높았고, /ㅇ/의 빈도가 낮았다<sup>10)</sup>. 또한, 강세구 초에 위치한 모음의 빈도를 강세구 내에 위치한 모음의 빈도와 비교하면, 강세구 말에는 /h, ʃ/의 빈도가 높았고, /-/의 빈도가 낮았다. 강세구 말 위치에서 /h, ʃ/의 빈도가 높은 이유는 강세구의 끝이 이러한 소리를 가진 문법 형태소로 끝나는 것과 관련한 것으로 보인다. 반면에 /-/의 빈도가 낮은 이유는 문법 형태소의 끝이 해당 모음으로 끝나는 경우가 없는 것과 관련한 것으로 보인다.

#### 4. 두음소(bi-phoneme) 전체 빈도

끝으로 지금까지의 논의를 종합하여 가능한 모든 두음소(bi-phoneme)의 빈도를 알아보았다. 가능한 두음소의 종류는 <표 - 20>에 보인 바와 같이 1,225종이지만, 그중에서 발화에서 확인된 두음소는 828종이었다. 그러나 불과 5.5%에 해당되는 55종의 두음소가 전체 발화의 50% 이상을 차지하고 있는 것으로 나타났다.

<표 - 20> 초성과 중성의 두음소 종류 수와 출현 총수

|                 |              |
|-----------------|--------------|
| 가능 종류 수         | 1,225        |
| 출현 종류 수(비율)     | 1,009(82.4%) |
| 상위 50% 종류 수(비율) | 55(5.5%)     |
| 출현 총수           | 980,639      |

<표 - 21>은 빈도순으로 누적 빈도 50% 이상을 보이는 두음소의 목록을 보인 것이다. 표에서 볼 수 있듯이 강세구 말이나 초에 놓은 두음소의 비율이 상대적으로 높다는 것을 알 수 있다. 누적 빈도 상위 50%를 보이는 55종 중에서 강세구 경계와 관련된 두음소는 모두 18종이었으며, 이들의 출현 빈도의 합은 20.0%에 달하는 것으로 나타났다. 하지만 이 두음소들이 발화에서 반복적으로 관찰되는 이유는 본 연구의 분석 방법과 유관하다. 본 연구에서는 강세구의 앞뒤에 강세구 경계를 의미하는 휴지를 음소로 설정하는 방법을 이용하고 있기 때문에 강세구 경계에 놓인 음소의 두음소를 '#+음소' 혹은 '음소+#'으로 보았다. 따라서 강세구 경계의 유형 빈도가 낮아져서 한 유형당 출현 빈도가 높을 수밖에 없다.

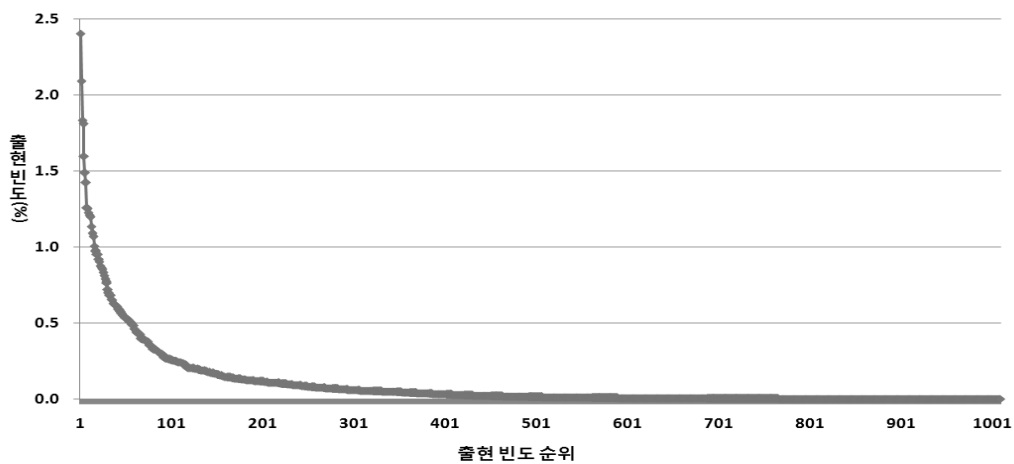
한편, 전체 두음소 중에서 강세구 경계에 놓인 음소들을 제외하고 가장 높은 출현 빈도를 보인 두음소는 'ㄱ+-'였고, 다음으로 '-+ㄴ', 'ㄱ+ㄴ', 'ㄴ+ㄴ', 'ㄴ+-'의 순이었다. 이 두음소들은 앞서도 논의하였듯이 대체로 문법 형태소에 사용되는 것들이었다. <그림 - 9>는 두음소의 전체 빈도 분

10) 왜 연구개음에서만 현저한 차이를 나타내는지에 대해서는 앞으로 좀 더 살펴볼 필요가 있어 보인다. 하지만 아마도 기저 음소가 /ㄱ/로 끝나는 억양구 말 강세구는 후행 강세구 초의 공명 자음에 의해 장애음의 비음화를 겪지 않지는 반면에 나머지 위치에서는 같은 음운론적 환경에서 장애음의 비음화를 겪는 것과 관련한 것이 아닌가 생각된다.

포를 보인 것이다. 그림에서 보듯이 소수의 두음소가 전체 자료의 대부분을 차지하는 것으로 나타났다. 다시 말해서 발화에서는 소수의 두음소가 발화에서 매우 반복적으로 사용되는 것으로 관찰되었다.

<표 - 21> 고빈도 두음소의 출현 빈도 및 누적 빈도(%)

| 등위 | 두음소 | 출현 빈도 | 누적 빈도 | 등위 | 두음소 | 출현 빈도 | 누적 빈도 | 등위 | 두음소 | 출현 빈도 | 누적 빈도 |
|----|-----|-------|-------|----|-----|-------|-------|----|-----|-------|-------|
| 1  | #+ㄱ | 2.4   | 2.4   | 20 | ㅎ+ㅏ | 1     | 26.9  | 39 | ㄷ+ㅏ | 0.6   | 41.3  |
| 2  | ㅏ+# | 2.1   | 4.5   | 21 | ㄴ+ㅣ | 0.9   | 27.9  | 40 | ㅓ+# | 0.6   | 41.9  |
| 3  | ㅏ+# | 1.8   | 6.3   | 22 | ㄴ+# | 0.9   | 28.8  | 41 | ㄱ+ㅣ | 0.6   | 42.5  |
| 4  | ㄱ+ㅡ | 1.8   | 8.1   | 23 | ㅓ+ㅏ | 0.9   | 29.7  | 42 | ㅣ+ㄱ | 0.6   | 43.1  |
| 5  | ㄴ+# | 1.6   | 9.7   | 24 | ㄱ+ㄱ | 0.9   | 30.5  | 43 | ㅈ+ㅏ | 0.6   | 43.7  |
| 6  | ㅡ+ㄴ | 1.5   | 11.2  | 25 | ㅈ+ㅣ | 0.9   | 31.4  | 44 | ㄱ+ㄴ | 0.6   | 44.3  |
| 7  | ㄱ+ㅏ | 1.4   | 12.7  | 26 | ㄹ+ㄱ | 0.8   | 32.2  | 45 | ㅊ+ㄴ | 0.6   | 44.9  |
| 8  | ㅊ+# | 1.3   | 13.9  | 27 | #+ㅓ | 0.8   | 33    | 46 | ㄹ+ㅣ | 0.6   | 45.4  |
| 9  | ㅣ+# | 1.3   | 15.2  | 28 | #+ㅣ | 0.8   | 33.8  | 47 | ㅣ+ㄴ | 0.6   | 46    |
| 10 | ㅏ+ㄴ | 1.2   | 16.4  | 29 | ㅏ+ㄱ | 0.8   | 34.6  | 48 | ㄴ+ㄴ | 0.6   | 46.5  |
| 11 | ㄴ+ㅡ | 1.2   | 17.6  | 30 | #+ㄷ | 0.8   | 35.3  | 49 | ㄱ+ㅓ | 0.5   | 47.1  |
| 12 | ㄴ+ㅏ | 1.2   | 18.8  | 31 | #+ㅎ | 0.7   | 36.1  | 50 | ㅏ+ㅇ | 0.5   | 47.6  |
| 13 | ㅡ+ㄹ | 1.1   | 19.9  | 32 | #+ㅏ | 0.7   | 36.8  | 51 | ㅣ+ㄹ | 0.5   | 48.1  |
| 14 | #+ㅓ | 1.1   | 21    | 33 | ㅓ+ㄱ | 0.7   | 37.4  | 52 | ㄴ+ㅏ | 0.5   | 48.7  |
| 15 | #+ㅈ | 1.1   | 22.1  | 34 | ㄹ+ㅏ | 0.7   | 38.1  | 53 | ㄷ+ㅡ | 0.5   | 49.2  |
| 16 | ㅏ+ㄴ | 1     | 23.1  | 35 | ㅇ+ㄱ | 0.7   | 38.8  | 54 | #+ㅓ | 0.5   | 49.7  |
| 17 | ㄷ+ㅏ | 1     | 24.1  | 36 | ㄹ+# | 0.7   | 39.4  | 55 | ㅏ+ㄹ | 0.5   | 50.2  |
| 18 | #+ㄴ | 1     | 25    | 37 | ㅡ+ㄹ | 0.6   | 40.1  |    |     |       |       |
| 19 | ㄴ+ㄷ | 1     | 26    | 38 | ㄱ+ㅏ | 0.6   | 40.7  |    |     |       |       |



<그림 - 9> 두음소 전체 빈도

## IV. 논의 및 결론

지금까지 성인 자유 발화 자료를 분석하여 한국어의 음소 전이 빈도(두음소 빈도)를 운율적 환경에 따라서 알아보았다. 분석 대상 자료의 규모는, 57명의 성인 화자가 발화한 총 35,439발화, 403,572음절, 858,4793음소였다. 발화 자료를 분석한 결과, 모두 1,009종의 두음소 유형이 발화 자료에서 관찰되었는데, 이 중에서 5.5%에 해당하는 55종의 두음소가 전체 자료의 50% 이상을 차지하는 것으로 나타났다. <표-22>에 보인 바와 같이 이러한 경향은 본 연구에서 살펴본 모든 운율적 환경에서 공통적으로 나타났다.

<표-22> 운율 환경에 따른 두음소의 유형 빈도 및 출현 빈도

| 두음소 종류 | 가능 종류 수 | 출현 총수   | 출현 종류 수(비율) | 상위 50% 종류 수(비율) |
|--------|---------|---------|-------------|-----------------|
| C+V    | 306     | 345,128 | 232(75.8%)  | 21(6.9%)        |
| V+C    | 126     | 109,779 | 98(77.8%)   | 8(6.3%)         |
| C\$C   | 86      | 71,264  | 86(100%)    | 9(10.5%)        |
| C\$V   | 17      | 4,576   | 16(94.1%)   | 2(12.5%)        |
| V\$C   | 324     | 179,834 | 292(90.1%)  | 21(7.2%)        |
| V\$V   | 306     | 24,220  | 225(73.5%)  | 16(5.21%)       |
| #+음소   | 35      | 122,919 | 35(100%)    | 5(14.3%)        |
| 음소+#   | 25      | 122,919 | 25(100%)    | 4(16%)          |

발화에서 관찰되는 이러한 불균형성은 발화에서 소수의 특정한 음소 연쇄가 대다수의 음소 연쇄보다 훨씬 반복적으로 사용되는 것을 알 수 있다. 이러한 사실은 발화에서 일부 특정 음소의 연쇄가 대부분의 음소 연쇄에 비하여 훨씬 높은 기능 부담량을 가진다는 사실을 말해준다.

전체 두음소 중에서 강세구 말이나 초에 놓은 두음소의 비율이 상대적으로 높았는데, 누적 빈도 상위 50%를 보이는 55종 중에서 강세구 경계와 관련된 두음소는 모두 18종이었다. 이들의 출현 빈도 합은 20.0%에 달했는데, 이는 본 연구에서 강세구 경계를 하나의 음소로 인정하는 연구 방법론을 사용하고 있기 때문이었다. 한편, 강세구 경계에 위치한 두음소 이외의 두음소 중에는 ‘ㄱ+ㄹ’, ‘ㄹ+ㄱ’, ‘ㄱ+ㄷ’, ‘ㄷ+ㄱ’, ‘ㄷ+ㄹ’, ‘ㄹ+ㄷ’와 같이 문법 형태소에 사용되는 두음소들이 고빈도를 차지하는 것으로 나타났다.

본 연구는 한국어의 말소리의 빈도 관련 정보를 체계적이고 종합적으로 기술하기 위해 수행된 여러 가지 연구 과제 중의 하나라고 할 수 있다. 따라서 한국어 말소리의 빈도 관련 정보를 체계적이고 종합적으로 기술하기 위해서는 다양한 종류의 후행 연구 과제의 수행이 필수적이라고 할 수 있다. 앞으로 수행되어야 할 연구 과제는 크게 두 가지 방향으로 정리될 수 있을 것이다. 한 가지 방향은 지금까지의 연구 결과를 비교하여 그 시사점을 찾아내는 것이고, 다른 한 가지 방향은 필요하지만 아

직 연구가 진행되지 않은 부분의 연구 결과를 기술하고 정리하는 것이다. 전자의 예로는 본 연구와 동일한 연구 방법을 이용하여 아동의 자유 발화를 연구했던 신지영(2005a, b)의 연구와 비교함으로써, 언어 발달의 각 단계에서 관찰되는 말소리 관련 빈도의 변화를 살펴보고 그 시사점을 논의하는 것을 들 수 있다. 또한, 후자의 예로는 본 연구와 동일한 연구 방법을 활용하여, 사전에 등재된 단어를 대상으로 연구를 진행하여 결과를 기술하고 정리하는 것을 들 수 있다<sup>11)</sup>.

한국어 말소리의 빈도 관련 정보에 대한 지식은 한국어 화자들이 음향 신호를 언어의 유의미한 단위로 분절하여 이해하는 데 사용하는 가장 기본적인 지식 중 하나일 것이다. 따라서 이와 관련된 연구 결과는 한국어 사용자들의 음성 언어 이해 과정과 음성 언어 산출 과정을 이해하기 위한 기초 자료로 활용될 수 있을 것이다. 아울러 국어학의 응용 분야인, 언어 병리학, 한국어 교육, 음성 공학에서도 기초 자료로 활용될 수 있을 것이다.

## 참 고 문 헌

- 김민정·배소영·고도홍(2001). 2~5세 아동의 자발적 발화에 나타난 한국어 음절 및 음운 빈도. 『음성 과학』, 8, 99-107.
- 박동근·이석재(2005). 대학생 구어의 음운 실현 연구. 서상규·구현정(편). 『한국어 구어 연구(2)-대학생 말뭉치를 중심으로-』. 서울: 한국문화사.
- 신지영(2005a). 3세~8세 아동의 자유 발화 분석을 바탕으로 한 한국어 말소리의 빈도 관련 정보. 『한국어학』, 27, 163-200.
- 신지영(2005b). 한국어 음소의 전이 빈도: 3세~8세 아동의 자유 발화 자료를 바탕으로. 『한국어학』, 28, 81-110.
- 신지영(2008). 성인 자유 발화 자료의 분석을 바탕으로 한 한국어의 음소 및 음절 관련 빈도 『언어 청각장애연구』, 13, 193-215.

11) 자유 발화의 경우는 문법 형태소에서 사용되는 음소가 고빈도를 차지하기 때문에 한국어 단어가 가지고 있는 말소리의 특징을 기술하기 위해서는 자유 발화 자료뿐만 아니라 사전에 등재된 단어들의 말소리 관련 정보를 정리하는 것이 반드시 필요하기 때문이다. 이를 위하여 앞으로의 연구에서는 연구의 대상을 사전에 등재된 단어들로 확장하여, 이들을 문법 형태소와 어휘 형태소로 나누고, 어휘 형태소들을 품사별, 원정보별로 나누어 한국어 말소리의 빈도 관련 정보를 총괄적으로 정리해 보고자 한다.

- 이 논문은 2007년도 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2007-231-A0107).



ABSTRACT

## Bi-phoneme Frequency of Korean Based on the Analysis of Spontaneous Speech Data<sup>1)</sup>

Jiyoung Shin<sup>§</sup>

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

---

**Background & Objectives:** The aim of the present study was to provide some information on frequencies of occurrence for units of Korean bi-phoneme based on the careful analysis of spontaneous speech spoken by adult speakers of Korean. Information on bi-phoneme frequency as well as phoneme and syllable frequencies is one of the most fundamental and basic knowledges on Korean language and is used in the segmentation of speech signal into meaningful utterance units.

**Methods:** 57 adult (28 male and 29 female) speakers of standard Korean were employed as subjects for the present study. Spontaneous speech spoken by the subjects were recorded and phonemically transcribed. Bi-phoneme frequency was calculated from the database. The data size of the present study was as follows: 35,439 utterances, 403,572 syllables, and 858,479 phonemes. **Results:** 980,668 classified into 8 types (C+V, V+C, C\$C, C\$V, V\$C, V\$V, #+P, P+#) according to prosodic boundaries between phonemes, were examined. Logically possible numbers of bi-phoneme types were 1,225 types and 1,009 types of bi-phonemes were found in the data. However, only 5.5% (55 types) of bi-phoneme types covered over 50% of the speech data. **Discussion & Conclusion:** Information on bi-phoneme frequency of Korean is one of the most fundamental and basic data for speech pathology, Korean language teaching, and speech engineering as well as Korean linguistics. (*Korean Journal of Communication Disorders* 2008;13;477-502)

**Key Words:** phoneme transition frequency, bi-phoneme, bi-phoneme frequency, spontaneous speech, phonemic transcription

---

---

1) This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2007-231-A0107).

Received June 15, 2008; final revision received August 15, 2008; accepted August 20, 2008.

<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

© 2008 The Korean Academy of Speech-Language Pathology and Audiology  
<http://www.kasa1986.or.kr>

## References

- Kim, M. J., Pae, S., & Ko, D. (2001). Syllable and phoneme frequencies in the spontaneous speech of 2-5 year-old Korean children. *Speech Sciences, 8*(4), 99-107.
- Park, D., & Lee, S. (2005). A study on phonological realization of spoken Korean. In S. Seo & H. Koo (Eds.), *A study on spoken Korean*. Seoul: Hankookmunhwasa.
- Shin, J. (2005a). Phoneme frequency of 3 to 8-year-old Korean Children. *Korean Linguistics, 27*, 163-200.
- Shin, J. (2005b). Bi-phoneme frequency of 3 to 8-year-old Korean children. *Korean Linguistics, 28*, 81-110.
- Shin, J. (2008). Phoneme and syllable frequencies of Korean based on the analysis of spontaneous speech data. *Korean Journal of Communication Disorders, 13*(2), 193-215.