

한국어 위치정보 데이터의 발음 분석¹⁾

김 선 희

서울대학교 인문정보연구소

A Pronunciation Analysis on Korean Point-of-Interest Data

Sunhee Kim

Center for Humanities and Information, Seoul National University

E-mail : sunhkim@snu.ac.kr

Abstract

This paper aims to analyze the pronunciation of Korean Point-of-Interest (POI) data, which consist of 224 sound files, from the phonological point of view, adapting the notion of prosodic word within the framework of Intonational Phonology. Each POI word is broken down into prosodic words, which are defined as the minimal sequence of segments which can be produced as one Accentual Phrase (AP). Then the pronunciation of the POI word considering its prosodic words are analyzed. The results show that: in most cases, a prosodic word is realized as one AP; that, in some cases, two prosodic words are pronounced as one AP; and that no cases are found where 3 prosodic words are realized as one AP.

I. 서론

동일한 문장이라도 화자나 발화 속도 등 여러 요인에 의하여 다르게 발화되는데, 이와 같이 여러 다른 요인에 기인한 발음의 다양성은 음성인식 시스템의 성능 저하의 한 원인이 된다. 음성인식 시스템에서 이러

한 다양한 발음 변이 현상을 모델링하는 것은 어휘부(Lexicon)와 음향 모델(Acoustic Model), 그리고 언어 모델(Language Model)의 세 영역에서 가능하다. 어휘부에서의 발음 모델링은 발음사전(lexicon) 기반 모델링이라고도 하는데, 이는 가능한 모든 발음들을 생성하여 어휘부에 추가하는 방법으로 이루어진다[1].

가장 이상적으로는 음성인식 시스템에서 사용하는 음성데이터를 실제로 일일이 듣고 그 발음 변이 현상을 전사하여 사용하는 것이 가장 정확한 어휘부를 구축할 수 있는 방법이 되겠지만, 실제로 대용량의 음성데이터를 이용하는 음성인식 시스템에서 이러한 방법은 불가능하다. 일반적으로 어휘부에 가능한 모든 발음을 추가하는 방법으로는 지식 기반 방식[2, 3]과 데이터 기반 방식[4, 5]이 있다.

본 논문에서 위치 정보 데이터(Point-of-Interest: POI)란 행정구역 및 지명, 인명, 상호명과 같은 위치 관련 어휘로 구성된 데이터로서, 이는 텔레메틱스를 비롯한 다른 지명 정보 관련 분야의 응용 프로그램의 개발에 필수적이다[6]. [6]은 위치 정보 데이터에 대한 발음모델링에 관한 연구로서 운율어를 기본 단위로 상정하여 자동 발음열 생성기를 이용하여 가능한 모든 발음 추출하는 방법을 제안하고 이를 이용한 음성인식 시스템의 성능 향상 여부를 실험하였는데, 실제 음성데이터에 대한 분석은 이루어지지 않았다.

본 논문은 한국어 위치 정보 데이터를 대상으로 한 지식기반 방식에 의한 발음모델링을 위하여 대상 데이터에 대한 음운론적인 연구로서 한국어 위치 정보 데이터를 음절별로 분류하여 그 발음 변이 현상을 분석

1) 이 논문은 2006년도 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (KRF-2006-353-A00063).

한다. 이러한 연구는 한국어 위치 정보 데이터의 발음 모델링의 성능 향상에 직접적으로 기여할 뿐만 아니라, 음운론적으로는 발음 변이 현상에 대한 연구 분야에 새로운 시각을 제시할 수 있을 것이다.

II. 운율어와 발음변이 현상

억양음운론에 의하면은 운율단위가 실제 발화된 문장의 억양 패턴에 의하여 결정되는데, 한국어는 기본적으로 발화의 성조 패턴에 의하여 구분되는 구성 요소인 강세구(Accentual Phrase)와 발화의 억양 패턴에 의하여 구분되는 구성 요소인 억양구(Intonational Phrase)로 이루어진다[7]. [7]은 한국어의 운율 구조가 억양구와 강세구로 이루어지고, 억양구는 다시 하나, 혹은 둘 이상의 운율어로 이루어진다고 하는 운율의 계층적 구조를 상정하지만, 운율어가 실제 운율 구조에서는 아무런 기능을 하고 있지 않다고 지적하였다. 그럼에도 불구하고 [7]에서 운율어는 “강세구로 실현될 수 있는 분절음의 최소 연쇄(the minimal sequence of segments which can be produced as one Accentual Phrase (AP)”라고 정의되었다. 이는 바꾸어 말하면 운율어란 강세구가 될 수 있는 분절음의 최소 연쇄, 즉 일종의 잠재적인 강세구라고 할 수 있다.

실제 발화에서 강세구나 억양구의 실현은 초점, 어휘적 의미, 발화 속도, 음절 수 등 여러 가지의 요소에 의하여 변화될 수 있다. 억양음운론에 의하면 일반적으로 문장의 경우에는 운율어가 어절 경계와 일치하는데, 이러한 운율어가 위에서 언급한 요소들의 영향으로 하나의 강세구로 발화되거나 혹은 다른 어절과 결합하여 하나의 강세구를 형성하는 것을 볼 수 있다. 다음은 발화 속도에 따라 그 강세구 형성이 다른 예를 보인 것이다[7].

- (1) /이건 아주 좋은 그림이다/ ({}: 강세구 경계)
 a. {이건} {아주} {좋은} {그림이다}
 b. {이건} {아주 좋은 그림이다}

[7]은 복합명사구의 운율 구조에 대하여도 구체적으로 논의하였는데, 복합명사구는 어휘부에서 형성되지 않고 통사부에서 형성되는 구 복합어(phrasal compound)로서 일반적으로 어휘부에서 형성되는 복합어(lexical compound)와 구분하였다. 3개의 명사로 구성된 복합명사구의 예들을 가지고, 각각을 구성하고 있는 성분들의 형태론적 혹은 의미론적 구조와 운율 구조가 일치하거나 혹은 그렇지 않을 예들을 보여 주면서, 3개의 명사로 구성된 복합명사구가 보통 2개의

억양구로 실현됨을 보였다. 또한 이러한 복합명사구(2a)와는 달리 어휘적 복합어(2b)는 하나의 억양구로 실현된다고 한다.

- (2) a. /국회의원선거/ → {국회의원} {선거}
 b. /닭싸움/ → {닭싸움}

위치 정보 데이터는 모두 특정 장소를 나타내는 고유명사에 해당하고 새로운 상호명의 출현에 따라 많은 신조어를 포함하는 등, 일반적인 텍스트 데이터와는 다른 특성을 보인다. 구조적으로는 수식어를 포함한 명사구의 형태를 보이는 경우도 있고, 하나 이상의 문장으로 이루어진 경우도 있으나 대부분의 경우는 ‘한국교육문화사’와 같이 명사들이 결합된 복합 명사구로 구성되어 있다. 본 논문에서는 [6]의 제안과 같이 복합명사구의 구성요소를 “강세구로 실현될 수 있는 분절음의 최소 연쇄”인 ‘운율어(prosodic word)’라고 한다.

2개의 운율어가 하나의 강세구로 실현되면, 운율어의 경계에서 음운 규칙이 적용되는데, 이러한 음운규칙의 적용은 화자나 초점, 어휘적 의미, 발화 속도, 음절 수 등 여러 가지의 요소에 기인하는 운율어의 결합에 따라 여러 가지 다양한 발음변이 현상을 야기하게 된다. 따라서 복합명사구의 발음 변이 현상을 모델링하기 위해서는 복합명사구를 그 구성요소인 운율어로 분리하는 것이 필수적이다. 반대로, 복합명사구로 구성된 위치 정보 데이터는 하위 단위인 운율어로 나뉘고, 운율어가 결합하여 하나의 강세구로 발화되는지 아닌지의 여부에 따라 발음변이 현상이 야기되므로, 운율어의 경계에서 음소변동이 관찰되면 인접하는 두 개의 운율어가 하나의 강세구로 실현되었다는 것을 알 수 있다.

본 논문은 실제 음성 데이터 가운데 음소변동 현상이 관찰될 가능성이 높은 데이터들을 모아서, 운율어 경계에서 음소변동 현상이 일어나는 여부에 따라 주어진 데이터의 운율 구조를 파악하고자 한다.

억양음운론은 하나의 발화를 그 계층적 구성요소에 의하여 나타내고 각각의 구성요소는 서로 다른 규칙의 적용 범위가 된다고 한다. 아래 (3)과 (4)는 억양음운론[7, 8]의 범주에서 그 운율 단위와 그 내부에서 적용되는 규칙을 정리한 것이다.

- (3) 강세구를 그 적용 범주로 갖는 규칙
 → 유성음화, 장애음뒤의 경음화, 단모음화, ‘ㄹ’뒤의 경음화, 격음화, 연음규칙, 탄설음화
 (4) 억양구를 그 적용 범주로 갖는 규칙
 → 유음화, 장애음의 비음화, 마찰음화와 /s/의 구개음화

다음절로 구성된 복합명사구는 실제로 여러 개의 강세구로 실현될 수도 있고, 경우에 따라 휴지를 포함하는 억양구로도 실현된다. 따라서 여기에서 대상이 되는 복합명사구의 음소 변동을 고찰하기 위하여 고려해야 하는 음소변동 규칙은 다음과 같이 10개의 규칙이 된다[9]: 1) 종성 중화, 2) 자음군 단음화, 3) 장애음 뒤의 경음화, 4) 격음화, 5) 장애음의 비음화, 6) 유음화, 7) 이중 비음 동화(장애음의 비음화+/르/의 비음화), 8) 연음(재음절화), 9) 구개음화, 10) 단모음화.

III. 분석 결과

3.1 음성데이터

분석에 사용될 한국어 위치 정보 음성 데이터는 한국전자통신연구원에서 2004년 수집 배포한 것으로 총 1049명(남성 519명, 여성 530명)의 화자가 두 가지의 다른 환경에서 발성한 음성으로 구성된다. 첫 번째 환경은 차량을 지하 주차장의 실내에 설치하고 잡음을 추가한 시뮬레이션 환경이며, 두 번째 환경은 도로를 주행하는 차량내부에서 음성을 녹음하는 실제 환경의 조건이었다.

발성목록은 단독 숫자 음과 네비게이션용 POI 지명 단어와 제어 명령어, 지번의 가변 어휘인식을 위한 PRW 등을 포함하는데, 본 논문에서는 이 가운데 네비게이션용 POI 지명 데이터를 그 분석 대상으로 한다. 발음 분석을 위하여 음절별로 분류된 전체 음성 데이터 가운데 음운규칙이 가장 많이 적용될 가능성이 있는 단어를 음절별로 20개씩 선정하였다.

2음절어는 모두 하나의 강세구로 발음될 것으로 보아 분석에서 제외하였다. 3음절어부터 12음절어까지는 각각 20개의 음성 파일이 이용되었고, 13음절어 15개, 14음절어 6개(5), 15음절어 1개(6), 16음절어 1개(7), 18음절어 1개(8)로 총 224개의 음성 파일이 이용되었다.

음소 변동 외에도 성조에 따라 AP나 IP로 나뉘어질 수 있으므로 음향학적 방법에 의한 분석도 생각해 볼 수 있으나, 모든 음성 데이터가 자동화 시뮬레이션 환경이나 실제 환경에서 수집되어 많은 잡음을 포함한 것으로 부득이 발음변화 분석 방법은 청취 실험으로 한정하였다.

3.2 분석 결과

음절별로 운율어 구성 내용과 음성 파일의 분석 결과는 다음과 같다.

음절수 (데이터 수)	구성 운율어와 그 발화
3 (20)	3(20): 3(20)
4 (20)	2+2(18): 2/2(17), 4(1) 3+1(1): 3/1(1) 1+3(1): 1/3(1)
5 (20)	2+3(18): 2/3(16), 5(2) 3+2(2): 3/2(1), 5(1)
6 (20)	2+2+2(20): 2/4(12), 2+2+2(6), 4/2(2)
7 (20)	2+2+3(12): 2/2/3(10), 2/5(2) 3+2+2(5): 3/2/2(4), 3/4(1) 2+2+2+1(1): 2/5(1) 2+3+2(2): 2/3/2(2)
8 (20)	2+2+2+2(15): 2/2/4(9), 4/2/2(2), 2/2/2/2(2), 44(1), 242(1)1 2+4+2(1): 2/4/2(1) 2+3+3(2): 2/3/3(2) 3+2+3(1): 3/2/3(1) 3+3+2(1): 3/5(1)
9 (20)	2+2+2+3(13): 2/2/2/3(10), 2/2/5(2), 4/2/3(1) 2+2+3+2(2): 2/2/3/2(2) 2+3+2+2(2): 2/3/2/2(2) 3+2+2+2(2): 3/2/2/2(1), 3/6(1) 4+2+3(1): 4/2/3(1)
10 (20)	2+2+2+2+2(12): 2/2/2/2/2(4), 2/2/2/4(4), 2/2/4/2(1), 4/2/2/2(1), 4/2/4(1), 2/4/2/2(1) 2+3+2+3(5): 2/3/2/3(4), 5/2/3(1) 3+2+2+3(2): 3/2/2/3(2) 2+2+3+2+1(1): 2/2/3/2/1(1)
11 (20)	2+3+2+2+2(9): 5/2/4(3), 2/3/2/2/2(2), 5/6(1), 2/3/4/2(1), 2/3/6(1), 2/3/2/4(1) 2+2+2+2+3(6): 2/4/2/3(4), 2/2/4/3(1), 2/2/2/5(1) 3+2+2+2+2(2): 3/2/2/2/2(1), 3/2/4/2(1) 2+2+2+3+2(1): 2/4/3/2(1) 2+2+3+2+2(1): 2/2/3/4(1)
12 (20)	2+2+2+2+2+2(14): 2/2/2/2/2/2(6), 2/2/2/2/4(2), 2/4/2/4(2), 4/4/4(2), 2/2/4/4(1), 4/4/2/2(1) 2+3+3+2+2(2): 2/3/3/4(1), 5/3/2/2(1) 2+2+2+3+3(2): 2/4/3/3(2)

	2+3+2+2+3(1): 2/3/2/2/3(1) 2+2+3+2+3(1): 2/2/3/2/3(1)
13 (15)	2+2+2+2+3+2(6): 2/2/4/3/2(3), 2/2/2/2/3/2(2), 2/2/2/2/5(1) 2+3+2+2+2+2(2): 2/3/2/2/4(1), 5/4/4(1) 2+3+2+3+3(1): 2/5/3/3(1) 2+3+3+2+3(1): 2/5/2/3(1) 2+2+2+3+2+2(2): 2/2/2/3/2/2(1), 4/5/4(1) 2+2+2+2+2+3(2): 2/4/2/2/3(1), 4/4/2/3(1) 3+2+2+2+2+2(1): 3/4/2/2/2(1)
14 (6)	2+2+2+2+2+2+2(5): 2/2/2/2/2/2/2(4), 2/4/4/4(1) 3+2+2+2+2+3(1): 3/2/4/2/3(1)
15 (1)	2+3+2+2+2+2+2(1): 2/3/2/2/2/2/2(1)
16 (1)	2+2+2+3+2+3+2(1): 2/2/2/3/2/3/2(1)
18 (1)	2+2+2+3+2+2+3(1): 2/2/2/3/2/2/3(1)

표 1. 음절수별로 분류한 POI 데이터의 운율어 구성 및 발화 현황

총 224개의 데이터는 대부분 2음절어, 3음절어의 운율어로 구성되어 있다. 3음절어의 경우는 모두 하나의 운율어로 구성되고 하나의 AP로 발화하였다. 음절수에 따라 차이는 있지만 음성데이터를 분석한 결과 많은 경우는 하나의 운율어가 하나의 AP를 구성하였고, 2개의 운율어가 하나의 AP를 이루는 경우들이 있었다. 하지만 3개 이상의 운율어가 결합하여 하나의 AP를 형성하는 경우는 없었다.

IV. 토의 및 결론

본 논문은 한국어 위치 정보 데이터를 대상으로 한 지식기반 방식에 의한 발음모델링을 위하여 총 224개의 음성데이터에 대한 음운론적인 연구로서 한국어 위치 정보 데이터를 음절별로 분류하여 그 발음 변음 현상을 분석하였다. 총 224개의 데이터는 대부분 2음절어, 3음절어의 운율어로 구성되어 있다. 음절수에 따라 차이는 있지만 음성데이터를 분석한 결과 많은 경우는 하나의 운율어가 하나의 AP를 구성하였고, 2개의 운율어가 하나의 AP를 이루는 경우들이 있었다. 하지만 3개 이상의 운율어가 결합하여 하나의 AP를 형성하는 경우는 없었다.

이러한 연구는 한국어 위치 정보 데이터의 발음 모델링의 성능 향상에 직접적으로 기여할 뿐만 아니라, 음운론적으로는 발음 변이 현상에 대한 연구 분야에

새로운 시각을 제시할 수 있을 것이다.

참고문헌

- [1] H. Strik and C. Cucchiaroni, "Modeling Pronunciation Variation for ASR: A Survey of the Literature," *Speech Communication* 29, pp. 225-246, 1999.
- [2] J. M. Kessens, M. Wester, H. Strik, "Improving the performance of Dutch CSR by modeling within word and cross-word pronunciation variation," *Speech Communication* 29, pp193-207, 1999.
- [3] J. H. Jeon, S. Wee, M. Chung, "Generating Pronunciation Dictionary by Analyzing Phonological Variations Frequently Found in Spoken Korean", *Proc. of International Conference on Speech Processing*, pp. 519-523, 1997.
- [4] E. Fosler-Lussier, "Multi-level decision trees for static and dynamic pronunciation models," *Proc. Eurospeech 1999*, 1999.
- [5] M. Riley, W. Byrne, M. Finke, S. Khudanpur, A. Ljolje, J. McDonough, H. Nock, M. Saraclar, C. Wooters, G. Zavaliagos, "Stochastic pronunciation modeling from hand-labeled phonetic corpora," *Speech Communication* 29, pp. 209-224, 1999.
- [6] 김선희, 박전규, 나민수, 전재훈, 정민화, "운율 정보를 이용한 한국어 위치 정보 데이터의 발음 모델링," *정보과학회논문지: 소프트웨어 및 응용*, 제34권 제2호, pp104-111, 2007.
- [7] Jun, S.-A. (1996), *The Phonetics and Phonology of Korean Prosody: intonational phonology and prosodic structure*, Garland Publishing Inc., New York : NY.
- [8] 문양수, 이호영, 하세경(2001), "유음화 규칙의 적용 영역", *언어학* 31, 101-125.
- [9] Kim, S. (2003), "Korean phonology for letter-to-sound rules," *Proc. Internoise 2003*.