

한국어 자동 발음열 생성 시스템을 위한 예외 발음 연구

김선희(광운대학교 음성정보처리기술연구센터)

<차 례>

- | | |
|----------------------|----------------------|
| 1. 서론 | 2.3. 자음과 모음이 연속되는 경우 |
| 2. 예외 발음 환경의 설정 | 2.4. 요약 |
| 2.1. 자음과 자음이 연속되는 경우 | 3. 예외 발음 사전 생성 |
| 2.2. 모음과 자음이 연속되는 경우 | 4. 결론 |

<Abstract>

A Study on Exceptional Pronunciations For Automatic Korean Pronunciation Generator

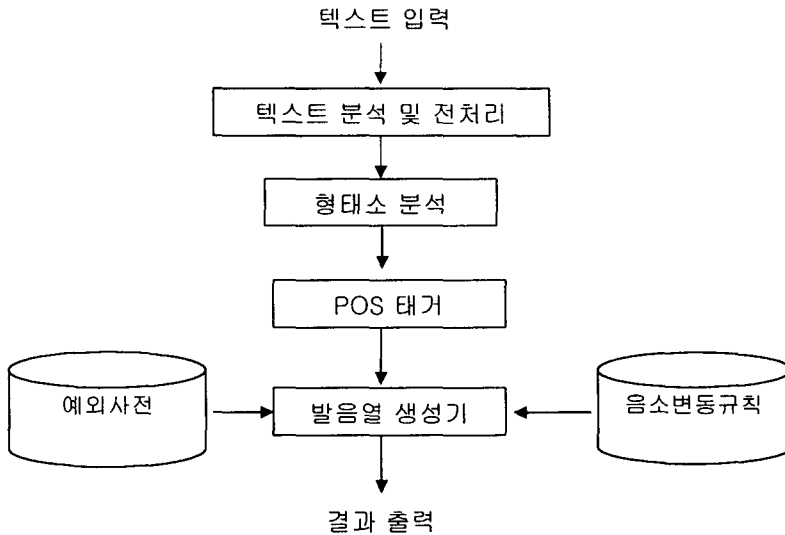
Sunhee Kim(SITRC, Kwangwoon University)

This paper presents a systematic description of exceptional pronunciations for automatic Korean pronunciation generation. An automatic pronunciation generator in Korean is an essential part of a Korean speech recognition system and a TTS (Text-To-Speech) system. It is composed of a set of regular rules and an exceptional pronunciation dictionary. The exceptional pronunciation dictionary is created by extracting the words that have exceptional pronunciations, based on the characteristics of the words of exceptional pronunciation through phonological research and the systematic analysis of the entries of Korean dictionaries. Thus, the method contributes to improve performance of automatic pronunciation generator in Korean as well as the performance of speech recognition system and TTS system in Korean.

* Keywords: Exceptional pronunciations, Automatic pronunciation Generation, Dictionary, Speech recognition, TTS (Text-to-Speech).

1. 서 론

본 논문은 한국어 문서 음성 변환 시스템(Text-to-Speech)이나 음성인식 시스템(Automatic Speech Recognition)의 주요 부분인 자동 발음열 생성 시스템(Automatic Pronunciation Generation System)의 한 부분인 예외 발음 사전을 생성하기 위하여, 사전의 표제어 분석을 기반으로 예외 발음을 보이는 어휘들을 추출하는 방법과 그에 따른 발음 사전 생성 결과를 보이고자 한다. 자동 발음열 생성은 그 언어의 맞춤법 체계를 반영하고 있는 문자열을 음성 체계를 반영하는 발음열로 변환하는 것을 의미한다. 일반적으로 자동 발음열 생성 시스템은 <그림 1>과 같이 텍스트 입력, 텍스트 분석 및 전처리, 형태소 분석, POS (Part of Speech: 품사) 태거, 예외 사전과 음소변동규칙을 바탕으로 한 발음열 생성기를 통해 결과를 출력하는 단계를 포함한다.¹⁾



<그림 1> 자동 발음열 생성 시스템

여기에서 발음열 생성기는 규칙부와 예외 발음 사전의 두 부분으로 구성되는데, 지금까지 규칙부에 관한 연구는 활발하게 진행되었던 반면에 예외 발음 사전에 관해서는 단순히 무작위로 어휘를 수록하는 방법으로 생성하여 왔다. 한국어의 경우 예외 발음의 처리는 TTS 시스템이나 음성인식시스템에서의 선결하여야 할

1) 본 논문에서의 상정하는 자동 발음열 생성기의 경우는 형태소 분석을 전제로 하긴 하지만, 그 기능을 최대한으로 축소시키고[1], 자료의 분석을 통하여 철자법에서 발음으로 바로 변환할 수 있는 최대한의 규칙을 상정하고 있다.

필수적인 과제로 인식되어 왔으나,²⁾ 그러한 예외 발음을 나타내는 어휘들의 특징에 대한 연구의 부재에 따라 예외 발음 사전의 생성 방법도 독립적으로 연구된 적이 없다.

규칙부에 속하게 되는 규칙의 한 예로 ‘경음화 현상’이 있다. 경음화 현상이란 예를 들면, ‘극비’가 [극뵤]로 발음되는 것을 지칭하는 것으로, ‘ㄱ’ 뒤에 오는 ‘ㅂ’은 항상 [뵤]으로 발음되는 것을 규칙화한 것이다. 이러한 경음화 현상은 실제로 ‘ㄱ’ 뒤에 오는 ‘ㅂ’ 뿐만 아니라 ‘ㄷ, ㄱ, ㅅ, ㅈ’도 모두 ‘뵤, ㄱ, ㅅ, ㅈ’으로 각각 발음되는 현상 등을 포함한다. 즉, 단어 내부에서 저해음 ‘ㅂ, ㄷ, ㄱ, ㅅ, ㅈ’이 다른 장애음에 후행하는 경우에 각각 [뵤, ㄷ, ㄱ, ㅅ, ㅈ]로 실현된다. 경음화 현상은 주어진 환경에서는 예외 없이 적용되어 규칙부를 이루게 된다.

이와는 반대로, 같은 음운 환경에서 단어에 따라 일정한 발음 변화 현상이 관찰되는 경우에 이는 개별적인 단어에 의존하므로 규칙화할 수는 없고, 그 단어들을 모아서 예외 발음 사전에 수록하여야 한다. 예를 들면 ‘물고기’와 ‘불고기’의 경우에, 이들은 각각 [물꼬기]와 [불꼬기]로 발음되는데, ‘ㄱ’ 다음에 오는 ‘ㄱ’이 ‘불고기’의 경우는 [ㄱ]으로 발음되는데, ‘물고기’의 경우는 [ㄱ]으로 발음된다. 이와 같이 같은 음운 환경에서 다르게 발음되는 경우의 예들은 자동 발음열 생성 시스템에서 예외 발음 사전에 수록되게 된다.

이에 본 논문은 한국어 음운 현상의 체계적인 연구와 자료의 분석을 통하여 예외 발음이 관찰되는 음운 환경을 규명한 다음, 그 일정한 환경에 해당하는 어휘들을 검토하여 예외 발음 사전을 생성하는 방법을 제시하는 데 그 목적이 있다. 분석 대상이 되는 자료로는 어휘가 중복되지 않으면서도 많은 종류의 어휘가 수록된 [2]를 이용하였다.³⁾

2) 여기에서 제안하는 자동 발음열 생성기는 표준 발음 생성을 위한 것으로, 수의적인 음운 현상들은 제외되었다. 여러 가지 요인에 따라 수의적으로 실현될 수 있는 발음을 생성하기 위한 연구는 음성합성 시스템에 있어서는 화자에 따른 발음 습관을 반영하기 위하여, 음성인식 시스템에 있어서는 성능향상을 위한 다중 발음 사전 개발을 위하여 필수적이다. 본 논문에서는 실제로 음성 인식을 위한 다중 발음 사전 생성을 위한 선행 연구로서 표준 발음 생성을 위한 연구로 제한하였다.

3) 예외적인 발음은 어절 내부만 관찰되는 것으로, 어절과 어절의 경계에서 관찰되는 음운 현상, 즉 동사 어간과 어미의 결합시에 관찰되는 경음화(‘안+다 [안따]’)나, 관형형 어미와 의존명사가 연속되는 경우의 경음화(‘할 수[할쑤] 있다’) 등은, <그림 1>에서 제안하는 자동 발음열 생성 시스템의 형태소 분석의 결과에 따라 형태적 제약에 의한 음운규칙으로 규칙부에서 처리하게 된다[3].

2. 예외 발음 환경 설정

여기에서 예외 발음 환경이란 예외 발음이 관찰되는 음운 환경을 지칭하는 것이다. 한국어에서 예외 발음이 관찰되는 음운 환경을 규명하기 위하여 한국어 음운 현상의 체계적인 연구가 선행되었다[3]. 이를 위하여 사용한 자료로는 어휘의 중복을 최소화하면서 최대한 많은 종류의 어휘를 포함할 수 있도록 하기 위하여 일반 사전인 연세한국어사전에 수록된 49,561개의 표제어를 분석하였다.

일반적으로 음운 현상이 관찰되는 음운 환경으로는 자음과 모음, 자음과 자음, 모음과 모음, 모음과 자음의 네 가지로 분류되는데, 그 가운데 예외 발음이 관찰되는 환경으로는 자음과 자음, 모음과 자음, 그리고 자음과 모음이 결합하는 경우이다[3]. 다음에서는 각각의 환경에서 예외 발음이 관찰되는 음소 결합 관계를 살펴보기로 한다.

2.1. 자음과 자음이 연속되는 경우

자음과 자음이 결합하는 경우에 예외 발음이 관찰되는 음운 환경으로는 선행하는 자음이 유성음(‘ㄱ, ㄴ, ㅇ, ㄹ’)이고 후행하는 자음이 평음인 경우에 후행하는 자음이 경음으로 실현되는 경우이다. 즉, 일반적으로 이러한 환경에서 아무런 음운 현상이 관찰되지 않으나⁴⁾, 단어에 따라 선행하는 유성음 뒤에서 평음이 경음으로 실현되는 것을 볼 수 있다. 이와 같은 현상을 여기에서는 ‘어휘적 경음화’라 부르기로 한다. 예를 들면, /ㄱ/과 /ㄱ/이 연속되는 경우에, ‘음감’에서는 [음감]으로 발음되는 반면, ‘음가’는 [음까]로 발음되는데, 전자의 경우와 대비하여 후자의 경우를 어휘적 경음화를 보이는 예외 발음이라고 한다.

이러한 어휘적 경음화 현상은 한자어에서 많이 관찰되고, 고유어의 경우에는 복합어나 파생어의 경우에 많이 관찰되는데, 각각의 경우를 형태론적, 의미론적 조건에 따라 규칙화하는 것은 용이한 일이 아니다[4]. 각각의 경우에 그 기준을 어떻게 설정하느냐에 따라 다양한 해결책이 가능하고, 한자어이건 고유어이건 이미 단어로 이루어진 예들이 많아서, 경우에 따라서는 하나의 예를 위하여 하나의 규칙을 설정해야 할 필요가 있게 된다. 한자어의 경우는 특히 한 글자가 한 형태소와 같은 행태를 보이므로, 두 글자 이상이 결합하여 한 단어를 형성할 때, 경음화 현상을 수반하는지, 그렇지 않은지는 어휘 의존적(idiosyncratic)이라고 알려져 있다. 따라서 본 논문에서는 한자어와 고유어를 구분하지 않을 뿐만 아니라, 고유어의 경우에도 복합어나 파생어 등 그 단어 형성 관계를 고려하지 않고 모두 단어 내부에서의 음운 현상으로만 처리하기로 한다.⁵⁾

4) 음성학적으로는 유성음화가 관찰됨.

5) [5], [6]에 의하면 자동 발음열 생성을 위하여 일반적인 규칙은 형태소 분석을 전제할 뿐

자음과 자음이 결합하는 경우에 관찰되는 또 다른 예외 발음의 유형이 관찰되는 경우는 /ㄴ/과 /ㄹ/이 연속되는 경우를 들 수 있다. 일반적으로 /ㄴ/과 /ㄹ/이 연속될 때는 유음화라 하여 선행하는 비음 /ㄴ/이 후행하는 유음인 /ㄹ/에 동화되어 /ㄹ/로 실현되는데(예를 들면, /관리/는 [괘리]로 발음됨), 어휘에 따라 유음 /ㄹ/이 선행하는 비음으로 동화되는 예외 현상이 관찰된다. 이러한 현상은 일반적으로 ‘유음의 비음화’라 부른다[7]. 예를 들면, ‘범신론’이 있는데, 이는 [범실론]으로 발음되지 않고 [범신론]으로 발음된다.

다음 <표 1>은 자음과 자음 사이에 관찰되는 예외음운 현상을 관찰되는 환경과 각각의 경우의 예를 들어 나타낸 것이다. 아래 표에서 음영으로 나타낸 부분이 경음화 현상을 포함하는 예외 발음의 예이고, 그렇지 않은 부분에는 이와 대조되도록 동일한 환경에서 일반적으로 경음화 되지 않는 예를 제시하였다.

<표 1> 자음과 자음이 연속되는 경우

	ㅂ	ㄷ	ㅅ	ㅈ	ㅊ	ㄹ
ㄱ	봄비	보름달	숨소리	흙집	꿈길	
	황금비	남달리	감소	금지	금기	
ㄴ	눈병	눈독	눈살	관점	눈가	범신론
	간병	군대	군살	반점	본가	관리
ㅇ	땡별	장독	땅속	동적	땅굴	
	성별	중독	광속	경제적	덩굴	
ㄷ	들보	갈대	결산	결재	들것	
	칠보	홀대	알사탕	줄자	별거	

2.2. 모음과 자음이 연속되는 경우

다음으로 모음과 자음이 연속되는 경우를 살펴보기로 한다. 이때, 모음은 앞 음절의 종성이고, 자음은 뒤에 오는 음절의 초성인 경우를 의미한다. 즉, 뒤에 오는 자음이 선행하는 모음의 종성인 경우는 해당하지 않는다. 이와 같은 예외 발음이 관찰되는 경우에는 모음과 평음 사이로서, 모음 뒤에 평음 /ㅂ, ㄷ, ㅅ, ㅈ, ㅊ/이 올 때, 이 평음이 각각 경음 /ㅃ, ㄸ, ㅆ, ㅉ, ㅊ/으로 발음되는 것이다.

이러한 경음화 현상도 위에서 살펴본 유음에 따르는 평음의 경음화와 마찬가지로 어휘 의존적인 현상으로 어휘적 경음화의 한 부류를 이룬다. 이와 같은 예로는 모음과 /ㅂ/이 연속하는 경우를 들 수 있는데, ‘세법’의 경우는 경음화 현상이

만 아니라, 모든 어휘들을 고유어인지 한자어인지를 구분해 줄 수 있는 기재의 필요성을 언급하였으나, 실제로 제안한 전체 시스템에서는 고유어와 한자어를 구별하고 있지 않다.

적용되어 [세뺨]으로 발음되는데, '사뺨'의 경우는 경음화가 적용되지 않아서 [사뺨]으로 발음된다.

다음 <표 2>는 모음과 자음이 연속하는 경우에 관찰되는 예외 발음을 환경과 각각의 경우의 예를 들어 정리한 것이다.

<표 2> 모음과 자음이 연속되는 경우

	ㅂ	ㄷ	ㅅ	ㅈ	ㄱ
V	육아뺨	대가뺨	역마살	채점	내과
	사뺨	사대	사살	개점	개관

2.3. 자음과 모음이 연속되는 경우

자음과 모음이 연속되는 경우는 철자상에서 자음이 앞 음절의 종성으로 쓰이고, 다음 음절의 초성으로서 철자상 묵음인 /ㅇ/, 즉 모음이 와서 중성인 모음과 연속되는 경우이다. 표기상 2개의 음절이 연속되는 경우에 선행하는 음절의 종성 자음과 후행하는 음절의 중성 모음 사이의 음운 현상은 다음 <표 3>과 같이, 연음, 'ㄴ' 첨가, 구개음화, 그리고 중성중화와 자음군 단음화가 선행하는 연음의 다섯 가지의 음운 현상이 관찰된다. 다음 표는 연세한국어사전의 표제어를 모두 분석하여 주어진 음소 환경이 관찰되는 경우를 음영으로 나타내고, 그 예를 들었다.

<표 3> 자음과 모음이 연속되는 경우

	연음	'ㄴ'첨가	구개음화	중화 (단순화)+연음	'ㅎ'탈락
ㅂ		잡일			
ㅍ		앞일			
ㄷ			말이		
ㅌ		홀이불	같이	겉웃	
ㅅ		나뭇잎		웃웃	
ㅆ					
ㅈ		늦여름		밤낮없이	
ㅊ		낮익다			
ㄱ		색연필			
ㅋ					
ㆁ					쌍이다
ㅇ		담요			
ㄴ		환을			
ㅇ		정열			
ㄹ		서울역			

	연음	'ㄴ'첨가	구개음화	중화 (단순화)+연음	'ㅎ'탈락
ㅂ				값어치	
ㄱ				넋없이	
ㄴ					값이
ㄷ					속았이
ㄹ					
ㄺ					
ㄻ					
ㄼ					
ㄽ					

위 <표 3>에서의 다섯 가지 음운 현상은 다시 규칙적인 경우와 불규칙적인 경우로 나누어진다. 규칙적인 경우로는 연음과 구개음화, 그리고 'ㅎ'탈락이 포함된다. 규칙적인 경우란 위에서 이미 언급한 바와 같이, 일정한 음운론적 혹은 형태론적 환경이 주어졌을 때, 항상 관찰되는 경우이다.

규칙적인 음운 현상 가운데 가장 대표적인 경우는 종성 자음이 모음에 연속되는 경우는 두 음소가 하나의 음절로 발음되는 연음을 들 수 있는데, 예를 들면, '밥이'는 [바비]로 발음된다. 두 번째 규칙적인 예로는 구개음화로서, /ㄷ/과 /ㄷ/의 경우에는 뒤따르는 모음이 /ㅣ/인 경우에는 항상 각각 구개음인 /ㅈ/과 /ㅊ/으로 발음된다. 규칙적인 마지막 경우는 'ㅎ'탈락으로서, /ㅎ/이 모음에 연속될 때 이 /ㅎ/은 항상 동사 어간의 종성으로서 모음으로 시작되는 어미가 오면 예외 없이 탈락된다.

불규칙적인 경우로는 'ㄴ'첨가와 종성 중화나 자음군 단순화가 선행하고 연음이 뒤이어 오는 경우가 있다. <표 1>과 <표 2>에서와 같은 방법으로, 각 예외 발음을 나타내는 어휘와 함께 그렇지 않은 경우(연음이 관찰되는 경우)의 예를 대비하여 보면 다음 <표 4>와 같다.

<표 4> 자음과 모음이 연속되는 경우에 관찰되는 예외 음운 현상과 일반 음운 현상

	'ㄴ'첨가		중화(단순화)+연음	
ㅂ	잡일	법인		
ㄷ	앞일	깊이		
ㄷ	홀이불	같이	겉웃	같이
ㅅ	나뭇잎	깨끗이	웃웃	깨끗이
ㅈ	늦어름	높낮이	밤낮없이	높낮이
ㅊ	낮익다	쫓아가다		
ㄱ	색연필	혹연		
ㄴ	담요	금요일		
ㄷ	환율	군인		
ㄹ	정열	궁인		
ㄺ	서울역	말일		
ㅂ			값어치	가없이
ㄱ			넋없이	(넋이) ¹⁾

‘ㄴ’ 첨가의 예로는 ‘잡일’을 들 수 있는데, 이는 [잡닐]로 발음된다. 이는 /ㅂ/과 모음 사이에 /ㄴ/이 첨가되고 그 결과 /ㄴ/ 앞에 위치하게 되는 장애음 /ㅂ/이 장애음의 비음화로 인하여 [ㅂ]으로 발음된다. ‘앞일’의 경우는 /ㅂ/과 모음 사이에 /ㄴ/이 첨가되고, /ㅍ/이 먼저 중화를 거쳐 /ㅂ/으로 변환된 다음, 다시 장애음의 비음화로 인하여 [ㅂ]으로 발음된다고 설명할 수 있다. 여기에서는 이러한 ‘ㄴ’ 첨가 현상은 [5]6), [6] 등 이전의 자동 발음열 생성에 관한 연구에서 모두 복합명사에 적용되는 규칙적인 현상으로 간주하여 왔었던 것에 반하여, 여기에서는 일정한 환경에서 규칙적으로 관찰되는 현상이 아닌 불규칙한 현상으로 분류한다. 왜냐하면 이러한 ‘ㄴ’ 첨가 현상은 고유어의 복합명사뿐만 아니라, ‘신여성[신녀성]’과 같은 한자어나 ‘늦여름(는녀름)’과 같은 파생어에서도 관찰되는 예가 많기 때문이다.

또 다른 불규칙 현상의 예로는 중성 중화가 선행하고 난 연후에 연음이 관찰되는 경우로, 예를 들면 ‘웃웃’이 [우둔]으로 발음되는데, 이는 중성 /ㅅ/이 [ㄷ]으로 중화된 다음 다시 연음에 의한 것으로 설명할 수 있다.⁷⁾

다시 말하면 중성 자음 다음에 모음이 연속되는 경우에는 총 5가지 음운 현상이 관찰되는데, 그 가운데 연음과 구개음화와 ‘ㅎ’ 탈락은 규칙적인 현상이고, ‘ㄴ’ 첨가와 중화(혹은 단순화)를 거친 연음 현상은 예외적인 현상으로서 예외 발음 사전을 구성하게 된다. 이때 중성으로 오는 자음으로는 단순 자음뿐만 아니라 자음군도 포함된다.

여기에서 한 가지 주의할 점은 예외 발음으로 분류되는 어휘들이 중의적(ambiguous)인 경우가 있을 수 있는데, 이에 관한 해결책을 제시하는 것은 여기에서는 다루지 않는다. 이미 <그림 1>에서 본 바와 같이 발음열이 생성되기 위해서는 형태소 분석이 선행되는데, 형태소에 따라 달리 발음이 되는 경우는 형태소 태깅 정보에 의하여 문제가 해결될 수가 있다. 그러나 동종의 형태소로 태깅되는 경우에 있어서 발음의 중의성 문제는 형태소 태깅이나 이러한 예외 발음 사전으로는 해결할 수가 없게 된다. 예를 들면 ‘화병’의 경우 같은 명사이지만 [화뽕]으로 발음될 때는 ‘花瓶’이고 [화뽕]으로 발음될 때는 ‘火病’이 된다. 이러한 중의성의 문제를 해결하기 위해서는 이러한 현상을 보이는 단어들을 분류한 다음, 그러한 단어들이 출현하는 환경에 대한 통사적, 의미적, 그리고 통계적 연구가 필요할 것이다.

6) [5]에는 “앞 형태소가 자음으로 끝나고 뒤 형태소가 /i, j/로 시작하는 고유어 복합어에는 /ㄴ/이 첨가된다.”라고 되어 있다. 이렇게 정의된 규칙은 형태소 분석을 전제할 뿐만 아니라, 모든 어휘들을 고유어인지 한자어인지를 구분해 줄 수 있는 기재도 필요한데도, 전체 시스템에서 이에 관한 언급은 찾아볼 수 없다.

7) 위 <표 7>에서 보면, /ㅌ, ㅍ, ㄷ/의 경우들은 특히 동일한 자음과 모음의 결합 환경에서 여러 가지 다른 규칙적, 불규칙적인 현상이 관찰되는 것을 볼 수 있다.

2.4. 요약

위에서 살펴본 바와 같이, 예외 발음이 관찰되는 환경은 유음이나 모음이 자음에 연속되는 경우(I), /ㄴ/과 /ㄹ/이 연속하는 경우(II), 그리고, 자음과 모음(철자상 /o/)이 연속되는 경우(III)의 세 가지로 나뉘고, 각각의 경우에 관찰되는 음운 현상으로는 어휘적 경음화, 비음의 유음화, ‘ㄴ’ 첨가, 중화나 단음화에 따르는 연음의 네 가지이다. 이와 같이 각각의 환경과 그 환경에서 관찰되는 음운 현상과의 관계를 다음과 같이 <표 5>로 나타내었다.

<표 5> 예외 발음환경 및 해당 음운 현상 (V: 모든 모음, C: 모든 자음)
 (I) 어휘적 경음화 (II) 비음의 유음화 (III) ‘ㄴ’ 첨가, 중화(단음화) + 연음

	ㅂ	ㄷ	ㅌ	ㅈ	ㅊ	ㄹ	V
ㅁ	(I)						
ㄴ						(II)	
ㅇ							
ㅕ							
V							
C							(III)

이와 같이, 예외 발음이 관찰되는 음운 환경을 한국어 음운 현상과 일반 사전인 [2]의 표제어 분석을 근거로 하여 정리한 결과, 일반적으로 불규칙적이라고 간주되는 예외 발음이 그 유형과 그 관찰되는 환경이 일정하다는 것을 알 수 있었다.

3. 예외 발음 사전의 생성

기본 어휘가 수록된 일반 사전의 표제어의 목록을 바탕으로 예외 음운 환경에 속하는 어휘들을 추출한 다음, 이 가운데 예외 발음이 관찰되는 어휘만을 가려서 예외 발음 사전을 생성한다. 이와 같은 방법으로 연세한국어사전을 이용하여 <표 5>에서와 같은 각각의 환경에 속하는 어휘들과, 그 어휘들을 검토하여 얻어진 예외 발음이 관찰되는 어휘들을 분류하였다. 다음 <표 6>은 그 결과를 나타낸 것이다.

<표 6> 예외 발음 생성 (단위: 어휘수)

	예외 환경 어휘	예외 발음
I	41,619	2,826
II	412	46
III	4,474	143
총계	46,505	3,015

즉, 예외 발음 환경에 해당하는 어휘가 총 46,505 단어이고, 이 가운데 예외 발음에 해당하는 단어는 3,015개로서, 이들이 자동 발음열 생성기에서 예외 발음 사전을 이루게 된다.

4. 결 론

무제한 음성합성의 경우나 대용량 음성인식 시스템의 경우 모두 최대한 많은 예외 발음 어휘를 포함하는 예외 발음 사전을 생성해 내는 것이 필수적이다. 자동 발음열 생성에 필요한 예외 발음 사전을 만들기 위해 먼저 예외 발음이 관찰되는 음운 환경을 한국어 음운 현상과 [2]의 분석을 근거로 정리하여 일반적으로 불규칙적이라고 간주되는 예외 발음의 유형과 그 관찰되는 환경이 일정하다는 것을 보였다. 이러한 연구 결과를 토대로 기본 어휘가 수록된 일반 사전의 표제어의 목록을 바탕으로 예외 발음 사전을 생성하는 방법을 제시하고, 이러한 방법으로 3,015개의 어휘를 포함하는 기본 예외 발음 사전을 생성하였다.

참 고 문 헌

- [1] 김선희, 안주은, 김순협, “텍스트 분석 기반 자동 발음열 생성 시스템”, 제20회 음성통신 신호처리 학술대회 논문집, pp.155-158, 2003.
- [2] 연세대학교 언어정보연구원, 연세한국어사전, 두산동아, 1998.
- [3] 김선희, “한국어 자소열-음소열 변환을 위한 음운 현상 연구”, 한국음향학회 하계학술대회 논문집, pp.105-108, 2003.
- [4] S. Kim, “Recherche sur “sai-sios” en coréen: Relation entre morphologie et phonologie”, Mémoire de D. E. A., E. H. E. S. S., Paris, 1987.
- [5] 전재훈, “형태음운학적 분석에 기반한 한국어 발음열 자동생성”, 서강대학교 전자계산학과 석사학위논문, 1997.
- [6] 이경님, 전재훈, 정민화, “한국어 연속음성인식을 위한 발음열 자동 생성”, 한국음향학회지, 20-2권, pp.35-43, 한국음향학회, 2001.
- [7] 이호영, 국어음성학, 태학사, 1996.

접수일자: 2003년 10월 21일

게재결정: 2003년 11월 6일

▶ 김선희(Sunhee Kim)

주소: 137-909 서울 서초구 잠원동 70 신반포4차 아파트 208동 507호

소속: 광운대학교 음성정보처리기술연구센터(SITRC)

전화: 02) 940-5585

E-mail: sunhkim@daisy.kw.ac.kr