

한국어 음성인식 시스템 향상을 위한 동음이철 단위의 중의성 유형 분류*

Ambiguity Types of the Homonymic & Heterographic Units for Improving Korean Voice Recognition System - a Preliminary Research

윤 애 선** · 강 미 영***
Aesun Yoon · Mi-young Kang

ABSTRACT

The accuracy rate of P2G (Phoneme-to-Grapheme) is one of the important factors determining the quality of unlimited voice recognition (VR) systems. Few studies were, however, conducted to reduce ambiguities of a phoneme string which can be segmented into a variety of different linguistic units (i.e. morphemes, words, eo-jeols), thus be transformed into more than one grapheme string. This paper is a preliminary research for building a large knowledge base of those homonymic & heterographic units(HHUs), which will provide unlimited Korean VR systems with more accurate P2G information. This paper analyzes 2 main factors generating HHUs: (1) boundary determination of the prosodic unit; (2) its segmentation into linguistic units. In this paper, linguistic characteristics determining variable boundaries of a prosodic unit are investigated, and the ambiguity types of HHUs are classified in accordance with their morphological and syntactic structures as well as with the phonological rules governing them.

Keywords: Korean voice recognition, homonymic & heterographic unit, P2G(Phoneme-to-Grapheme), ambiguity type

1. 들어가기

최근 이동통신 기기의 소형화와 인간친화형 로봇 실용화에 따라, 인간과 기계 간에 인간언어와 음성을 기반으로 하는 지능적 서비스의 필요성이 다시 대두하고 있다. 인간언어는 음성으로 발화되는 것이 가장 자연스러운 방법이기도 하지만, 현실적으로는 사람과 기계 간 대화 환경(Man Machine Interface, 이하 MMI)에서 자판과 화면을 쓸 수 없거나 손과 눈을 다른 일로 사용하는 경우 음성 인터페이스의 필요성이 더욱 커진다.

하지만, 인간언어의 생성과 분석 전 과정에 걸쳐 나타나는 중의성(ambiguity)이 해결되지 않으면 어떤 인터페이스를 사용하든 메시지의 정확한 전달이 불가능하다. 자판과 화면을 이용한 문자에 비해

* 이 논문은 2007 년 정부(교육과학기술부)의 지원(과학재단 R01-2007-000-20517-0)으로 이루어짐.

** 부산대학교 한국어 정보처리 연구실

*** 국립국어원

음성을 인터페이스로 사용하는 경우, 입출력 과정에서 새로운 특성이 있는 중의성 단계가 증가한다. 예를 들어, 음성합성(Text-To-Speech, 이하 TTS)¹⁾에서는 문자 기호열(grapheme string)로 표현된 텍스트를 ① 적절한 음성 기호열(phoneme string)로 바꾸고, ② 자연스럽게 끊어 읽는 운율 단위를 찾아, ③ 음성 및 운율 데이터베이스에서 가장 적합한 단위를 찾아 음성(speech)으로 출력하는 과정 중 첫 번째 단계(Grapheme-to-Phoneme, 이하 G2P)가 이에 해당한다. 반면에 음성인식(Voice Recognition, 이하 VR)에서는 ① 입력 음성의 음성적 특질을 분석하여, ② 음성 데이터베이스와 운율 데이터베이스를 이용하여 음성 기호열로 전사하고, ③ 음성 기호열에서 의미 단위를 이루는 어휘, 구, 문장 형태의 문자화된 텍스트로 출력하는 과정 중 세 번째 단계(Phoneme-to-Grapheme, 이하 P2G)가 이에 해당한다. 바로 음성공학(speech engineering)과 언어 공학(language engineering)이 만나는 지점이다.

- (ex1) (a) 출생 신고를 하러 가자.
 (a') [출생 신고를 하러 가자]
 (b) 신발을 신고 기다렸다.
 (b') [신발을 신꼬 기다렸따]
 (ex2) (a) [산나무를]
 (b) 산나물을 잘 먹는구나.
 (c) 내가 오늘 산 나물을 무쳐라
 (d) 산 나무를 땀 적은 한 번도 없다.

TTS의 G2P 과정에서 나타나는 (ex1)의 문자 기호열 '신고'는 품사 및 구성 형태소의 성분에 따라 음성 기호열 [신고]와 [신꼬]로 변환되어야 하며, VR의 (ex2) 과정에서 전사된 음성 기호열 [산 나무를]은 문맥에 따라 '산나물을', '산 나물을', '산 나무를'로 출력되어야 한다.

띄어쓰기나 철자를 교정하여 정규화할 수 있는 TTS의 문자 기호열을 대상으로 G2P를 수행하는 것보다, VR의 P2G는 문제의 복잡성이 훨씬 크다. 사용 어휘가 통제되지 않는 무제한 VR의 경우, P2G의 음성 기호열을 어휘나 어절 단위로 구분해야 하는 단계가 추가되기 때문이다.

- (ex3) (a) [어제뽀아훙씨삼십뽀네 || 큰따리 | 처다이를나안따는소시글든꼬 | 밤열씨오십뽀받기차를타고 | 경주에내려간따. || 도차카자마자 | 그관마으메 | 신바를신꼬 | 마루에올라간따.||]
 (a') [어제뽀[∇]아훙씨삼십뽀네 | 큰따리 | 처다이를[∇]나:안따는[∇]소시글든꼬 | 밤[∇]열씨오십뽀[∇]받기차를타고 | 경주에[∇]내려간따. || 도차카자마자 | 그관마으메 | 신바를신꼬 | 마루에[∇]올라간따.||]
 (b) 어젯밤 아훙 시 삼십 분에 큰딸이 첫아이를 낳았다는 소식을 들고 밤 열 시 오십 분 발 기차를 타고 경주에 내려갔다. 도착하자마자 급한 마음에 신발을 신고 마루에 올라갔다.

1) 이 논문은 음성합성의 여러 기술 중에서 '단위 선택기반 음성 합성(unit selection-based TTS) 기술'을 대상으로 한다.

VR에서 음성 기호열은 (ex3-a) 또는 (ex3-a')처럼 ‘끊어 읽기 단위(prosodic unit)’로 전사된다.²⁾ 이를 (ex3-b)와 같이 정확한 문자 기호열로 변환하기 위해서는 끊어 읽기 단위에서 나타나는 음성 기호열 중의성을 해결해야 한다. 동음이의어(homonym)와 같은 중의적 현상이나, 동음이의어가 어휘 또는 어절 단위로 제한되는 것에 비해 끊어 읽기 단위는 (ex3-a) 또는 (ex3-a')처럼 발화 상황, 화자의 의도 및 언어습관 등에 따라 편차가 심하여, 동음이철 단위(homonymic & heterographic unit, 이하 HHU)의 중의성을 해결하는 것이 P2G의 정확도를 크게 향상시킬 수 있다.

음성자료나 음성 기호열로 전사된 결과로부터 언어정보를 얻어 문자 기호열로 출력을 하면, 끊어 읽기 단위별로 연속된 음성 기호열로 묶이므로, 지금까지의 연구에서와는 다른 중의성 단위가 발생한다. ① 가변적인 끊어 읽기 경계에 따른 중의성과 ② 음운규칙 적용의 결과로 파생되는 음성 형태들에서 발견되는 중의성이다(4.2절 참조). 따라서 이들 중의성 발생 범위는 형태소에서부터 초어절에까지 확장된다.

이 논문은 한국어 VR 시스템에서 P2G의 정확도를 개선하기 위해, 대규모 HHU 말뭉치를 구축하기 위한 기초 연구로, 2 절에서는 중의성을 낳은 다양한 언어 단위의 특성 분석과 추정에 관한 선행연구를 살펴보고,³⁾ 3 절에서는 한국어 끊어 읽기 단위를 추정하기 위한 언어 단위의 특성을 분석하며, 이를 바탕으로, 4 절에서는 끊어 읽기 단위인 HHU의 중의성 유형을 분류한다. 5 절에서는 향후 연구방향을 알아보겠다.

2. 선행연구

국내에서 TTS는 전화번호 안내, 신문 기사 읽기 서비스, 장애우를 위한 공문서 읽기 서비스, 박물관 유물 안내와 같은 공공 서비스에 이용될 뿐 아니라 최근에는 보이싱피싱(Voice Phishing)과 같은 사기행각에도 사용될 정도로, 음성 품질이 높아 활용 범위가 급증하고 있다. 이에 비해, VR은 아직 특정 분야에서 제한된 사용자에 한해 통제된 어휘를 대상으로 하는 서비스만이 실용화되어 있다. 따라서 중의성 단위의 특성 분석과 추정에 관한 선행연구도 P2G보다는 G2P 영역에서 더 활발하게 이루어져 왔다.

2) ‘prosodic unit’은 ‘운율 단위’로 번역되나, 운율 단위 경계에서 짧은 길든 끊어 읽기가 이루어지고, 본 논문이 이러한 끊어 읽기 단위의 중의성을 다루고 있으므로 이를 강조하기 위해 운율 단위와 동일한 의미로 ‘끊어 읽기 단위’라는 용어를 사용한다. 이는 운율 단위 형성에 영향을 미치는 요소가 다양하고 말하는 사람의 자의성과 언어습관이 중요한 영향을 미치는 단위임을 표현하고자 하는 이유도 있다.

또한, 음성 기호열을 국제음성부호와 한글 자모를 이용하여 표시하는 방법이 있는데, 한국어 TTS와 VR 시스템에서는 후자를 사용하므로 이 논문에서도 후자 방식을 채택한다. 다음은 음성 기호열과 관련하여 이 논문에서 사용하는 기호다.

[]: 음성 기호열, ||: 강한 끊어 읽기, |: 약한 끊어 읽기, ^v: 수의적 끊어 읽기

3) P2G와 G2P에서 중의성 해결 문제는 어느 언어에나 해당되는 문제이고 언어에 따라 다른 양상으로 나타나므로, 논문의 초점을 모으기 위해 한국어를 대상으로 한 선행연구로 제한하겠다.

2.1 중의성 단위

음성 인터페이스에서 나타나는 중의성 단위는 형태소, 어휘, 어절 끊어 읽기 단위인데, 언어학 분야의 선행연구는 주로 어휘 단계의 양상 기술에 치중되어 있으며, 그 결과가 대개 사전의 형식을 빌어 나타난다. 원영섭(1994)은 2,000 여 개의 음과 의미는 같으나, 한자어가 다른 어휘를 제시한다. 하지만, 한글을 기준으로 하면 동음동철동의어 사전이다. 김병균(2000)은 약 96,000 여 개의 남북한 일반용어, 전문용어, 고유명사 중 품사가 같은 동철이의어를 수록하고 이를 해설했다. 동철이의어의 사용 양상을 연구한 것으로는, 윤미경(1975)이 동철이의어를 분류하고 품사별 통계를 제시했다.⁴⁾ 한국어 고유어 및 한자어에 2 음절짜리 동철이의어가 가장 많으며, 그중 81%가 한자어다. 품사별로는 명사(95%)가 가장 많이 나타났으며, 동사(2.5%), 부사(0.8%) 순서대로 나타났다. 모음의 장단을 변별적인 자질로 보고 길이의 대립([말], [말:])에 따른 동철이의어와 음소의 변동([눈:바람], [눈:빠람])에 의한 동철이의어를 분석했으며 문맥 이용, 접사나 설명요소 첨가, 한자 토시 사용 등과 같은 방법으로 한국어 화자들이 중의성을 없애는 양상에 대해 관찰했다. 강범모(2005)는 550 만 어절 규모의 세종 의미분석 말뭉치를 대상으로 동철이의어⁵⁾의 사용 양상을 조사했다. 동철이의어 대부분이 2 음절짜리 한자어이자 명사이고 동철이의어 중 대부분(90%)이 한 가지 의미로 쓰인다는 사실을 밝혀냈으며, 빈도 1 짜리의 동철이의어의 비율이 높지 않은 어휘형들도 담화 맥락 속에서는 대부분 어느 한 가지의 의미로 사용된다는 사실을 관찰했다.

이론 언어학 분야에서는 동철이의 단위의 연구 대상이 형태소나 어휘였으며 그 경계를 넘어가는 동철이의적 단위에 대한 연구는 이루어지지 않았다. 이는 기존의 언어 이론적 연구에서 형태소나 어휘가 언어학에서 확고한 문법적 지위를 가졌던 반면에, 그 이상의 단위들은 언어분석 단위로서의 문법적 지위를 갖지 못했기 때문이다. 하지만, 인간언어처리 분야에서는 띄어쓰기 단위인 ‘어절’이 언어분석 단위로 부상되면서부터 어절 경계 내에서 발생하는 형태·범주적 중의성에 대한 논의가 활성화되었다. 황화상 외(2003)는 어절의 형태소 분리와 해석의 양상에 따라, 크게 형태소 분리와 해석의 양상에 따른 유형과, 문법 범주의 차이에 따른 유형으로 나누었다. 형태론적 중의성의 해결을 위한 첫 번째 방법으로 선·후행 요소와의 의존관계가 뚜렷한 의존명사, 수사, 관형형 어미, 연결 어미 등의 문법 범주를 이용한 방안을 제시했고, 선·후행 요소와의 범주 결합관계와 어휘 결합관계를 토대로 한 어절 독립적 해소 방법을 제시했다. 임희석 외(1993)에서는 한국어 어절에서 발생하는 중의성이 어절을 이루고 있는 형태소의 분리 위치가 다르거나, 형태소의 분리 위치가 같은 형태소 결합이라 할지라도 형태소들의 품사가 서로 다르기 때문에 발생하는 것으로 보고, 어절의 중의성 유형을 4 가지로 분류하고, 인간언어처리를 위해서 어절 단위를 넘어선 문장 단위의 분석이 필요하다는 것을 강조하고 있다. 유혜원 외(2006)는 세종말뭉치로부터 형태적 중의성 (morphological ambiguity)이 있는 어절을 추출하여, 중의적 어절 유형 정보, 분석 후보 정보, 공기

4) 김병균(2000)과 윤미경(1975)에서는 동철이의어를 ‘동음어’라고 칭했으나 철자와 음을 구분하지 않아 사용한 용어다. 후자 논문의 분석 대상을 엄밀히 규정하면 동철이의어의 두 유형인 동철동음이의어와 동철이음이의어를 모두 포함했다.

5) 해당 논문에서는 동철이의어를 ‘동음이의어’라고 칭했다. 세종 의미분석 말뭉치가 텍스트 형태이므로 이 논문에서는 동일한 음이 아니라 동일한 철자를 사용하는 어휘를 대상으로 다의어 사용 양상을 분석한 것이다.

어 정보, 빈도 정보, 용례 정보 등 관련 정보를 포함하는 ‘한국어 형태론적 중의 어절 사전’ 개발을 계획했다.

위 대부분의 연구는 G2P에 부분적으로 유용한 동철이의 단위를 다루고 있고, 이러한 언어학적 기초연구를 바탕으로 구축된 지식베이스는 TTS의 G2P 단계에서 동철이음 단위 처리에 어느 정도의 우수한 성능을 보인다. 이경님 외(2001)가 동철이음어와 관련한 문제를 처리하기 위해 연결 제약 조건과 어절 내 구성 형태소의 종류와 연결 형태 2 가지를 고려하여 연음법칙, 구개음화, 경음화 등의 규칙으로 분류하고, 다시 그 규칙들을 적용양상에 따라서 세부규칙으로 나누어 처리했으며, Lee et al.(2002)은 발음변동 규칙을 단순화해서 표준발음규칙이 적용되는 경계점의 앞뒤 형태소가 실질 형태소인지 아닌지만 구분하기도 했다. 그리고 KIM et al.(2002)은 음성 패턴 사전과 형태 음소 연결 표, 그리고 CCV(Consonant Consonant Vowel)규칙과 LTS(Letter to Sound) 규칙들을 이용하여 음성 기호열을 생성했다.

하지만, 이론 언어학적 틀에서 P2G 단계의 중의적 단위를 해결할 수 있는 선행 연구는 거의 없다고 해도 과언이 아니다. 다만, 리형태 외(1994)가 어절 단위의 동음어와 HHU를 분류하여 수록하고 있을 따름이다. 약 1,000 여 개의 언어생활에서 주로 많이 쓰이는 어휘를 기본으로 하여 동의어, 반의어, 동음어를 수록한 사전을 편찬하면서, 동음어를 다음과 같은 세 가지 유형으로 분류하여 수록하고 있다. ① 기본형에서 음과 철자가 같은 동음 어휘인 동음동철(이의)어, ② 기본형에서 철자는 다르지만 음이 같은 동음 어휘인 동음이절(이의) 어휘, ③ 특정한 활용 형태에서 음이 동일한 동음이절(이의) 어절이다. 하지만 이 연구에서 다루는 범위는 정확도가 높은 P2G 시스템 구현에 사용하기에는 규모가 지나치게 작고, 분류가 매우 단순하다.

2.2 끊어 읽기 단위

음성 기호열 및 음성으로 출력된 결과물의 중의성을 해소하기 위해서는 음성이나 발음을 입력할 때의 여러 요인을 고려해야 한다. 사람들이 발화할 때는 단숨에 말할 수 없어 중간마다 숨을 쉬어 주며, 숨이 차지 않더라도 명확한 의미 전달을 위해 특정한 부분을 끊어 읽게 된다. 따라서 G2P에서 끊어 읽기 단위의 추정은 자연스러운 운율을 찾는 데 필요하고, P2G에서는 중의성 대부분이 끊어 읽기 단위 내부에 존재하므로 매우 중요하다.

따라서 끊어 읽기 단위를 추정하는 방법에 대해 음성학적, 전산언어학적, 공학적 연구가 이루어졌다. 운율음운론(Prosodic Phonology)에서 운율 계층의 한 단위로 제시되는 ‘운율구(phonological phrase)’는 통사구조에 바탕을 두어 도출된 운율 단위로서 형태·음운론적 발음규칙이 적용되는 최대 단위로 규정하며, 이보다 하위 계층으로 ‘운율어(phonological word)’가 있으며 상위 단위로는 ‘억양구(intonational phrase)’로 구분한다(강옥미, 1994).⁶⁾ 음성처리 분야의 운율구 경계 예측 모델링에서의 운율구 경계는 다양한 이름으로 구분되기는 하지만 보통 ① 경계 없음, ② 약한 경계, ③ 강한 경계 등의 세 단계로 나누는 것이 일반적이다(김상훈 외, 1977; 김선미, 2002; 김선희 외, 2008;

6) 여기서 ‘운율어’와 ‘운율구’는 학자에 따라서 각각 ‘음운 어휘’와 ‘음운구’로도 번역하고 있다(이찬도, 1997; 신지영, 2000 등). 한편, 음성학적인 연구에서는 소리의 음 높이 형태에 기초하여, 운율구가 아니라 강세구가 존재한다고 보는 학자도 있다(Jun, S.A. 1993. *The Phonetics and Phonology of Korean Prosody*. PhD dissertation. Ohio State University), 신지영(2000)에서 재인용함).

김성규, 1999; 김승원 외, 1999; 이호영, 1997).

이찬도(1997)는 ‘운율어’⁷⁾를 음운 단위로 보고, 이 범주에 ① 어휘에 조사와 어미가 결합된 형태 ② 내부에 아무런 경계가 없는 복합어 ③ 수식어와 의존 명사구를 포함했다. 또한, 인공신경망 중 단순회귀신경망(simple recurrent network)을 기초로 한 학습 알고리즘을 이용하여 경계 추출 모델을 개발했다.

김상훈 외(1997)는 운율구 경계를 자동으로 추출하고자 학습용 데이터⁸⁾에서 강한 경계로 구분된 운율구 내에 나타나는 품사나 문장 성분의 열을 추출하고 운율구 형태를 325 개로 나눈 후, 품사 및 문장 성분의 Bi-gram과 Tri-gram을 이용해서 운율구의 경계 강도를 측정했다. 경계 강도의 임계치(임의 값 0.5)를 정해서 경계 강도를 나누어 구 경계를 추출했다. 운율구 경계유형은 크게 ① 경계가 없는 강도, ② 약한 경계 강도, ③ 강한 경계 강도처럼 3 단계로 정하는 것이 통계적 의의가 있다고 보고 있다.

운율구 경계를 2 단계와 4 단계로 구분하여 기계학습을 시도한 연구로는 권오일 외(2002)와 전진욱 외(2002)가 있다. 전자에서는 ‘경계, 비경계’로 이분하여, 학습을 통해 끊어 읽는 부분의 존재 여부만을 예측하는 모델을 제안했다. 하지만, 이러한 연구 결과는 운율구 경계가 지나치게 단순화되어, 실제 TTS나 VR 시스템에 적용하기가 미흡하다. 반대로 전진욱 외(2002)는 4 단계로 끊어 읽기 단위를 구분하고 학습을 통해 그 경계를 추정하는 연구를 수행했다. 운율구를 세분화하면 좀 더 자연스런 끊어 읽기 단위를 생성할 수 있으나, 학습 데이터를 만드는 과정에서 인간이 이렇게 세분화된 단계의 차이를 듣고 구별하기는 어렵다.

끝으로 김선희 외(2008)는 음성모델과 문법모델을 결합한 기계학습모델을 제안했다. 이 연구는 대부분의 연구처럼 운율구 경계를 세 단계로 나누고 이를 추정하기 위해서는 운율(prosody) 정보가 중요한 역할을 한다고 보고, 기본주파수(F0), 강도, 지속 시간 등을 포함하는 14개의 자질을 이용한 음성모델을 구성하여, 띄어쓰기, 음절 수, 형태소 수, 단어의 품사 등을 포함하는 10개의 자질을 이용한 문법모델과 결합하고 있다.

3. 한국어 끊어 읽기 단위 예측

인간 언어에서 끊어 읽기 위치는 여러 요소를 고려하여야 정확한 예측을 할 수 있다. 가장 직접적인 요소로는 형태·통사적인 특성을 들 수 있으며, 한국어도 예외가 아니다(권재일 외, 1997; 김선미, 2002; 이찬도, 1997; Taylor & Black, 1997). 한국어에서 음운 현상은 ① 필수적이고도 보편적인 음운 규칙이 발생하는 기초단위인 ‘형태소 및 어휘’, ② 형태소와 단어를 기반으로 하는 체언과 조사, 용언과 어미, 체언 간 결합 등 다양한 형태론적 경계에서 발생할 수 있는 음운 현상이 발생할 수 있는 단위로 철자법상의 띄어쓰기와 일치하는 단위인 ‘어절’, ③ 어절의 경계를 넘어서 하나의 발음 단위로 인식되는 통사적 연쇄의 단위인 ‘초어절’ 단위에서 일어나며, 3 가지 단위 간에 형태·

7) 해당 논문에서는 ‘운율어’ 대신 ‘음운 어휘’라는 용어를 사용했다.

8) 이 연구에서는 강한 경계가 표시된 운율 데이터베이스의 156 개 문장과 KBS 뉴스 문장 5,300 여 개 문장을 사용했다.

통사적 결합이 긴밀하게 나타나서, 초어절 운율 단위를 형성한다. 이들의 품사 결합 패턴은 다음과 같다.⁹⁾

- <관형사/관형형어미> # <체언>(+조사)
예) 이 가지 잘라야 해/그런 아이는
- <체언> + <관형격조사> # <체언>
예) 학교의 서고
- <부사>¹⁰⁾ # <용언>:
예) 고기를 다 다지지 않고는
- 복합어: <<용언> # <보조용언>>^{복합용언}/<<명사> # <명사>(조사) # (명사...)>^{복합명사}
예) 네가 잡지 않아도 남을 것이다./수학 문제를 풀다.

<표 1>처럼 띄어쓰기로 구분될 수 있는 어절 단위와 수의적인 띄어쓰기가 허용되는 복합어가 짧은 끊어 읽기의 대상이 되며, 수식어과 피수식어의 결합이 초어절 단위의 끊어 읽기 단위를 형성하는데, 수식어와 피수식어가 단일어인지 아닌지에 따라 중간 끊어 읽기와 강한 끊어 읽기 단위로 구분된다.

표 1. 형태·통사적 결합 기준 끊어 읽기 단위

짧은 끊어 읽기 (운율어)	어절 단위, 복합명사나 복합용언(수의적 띄어 읽기)
중간 끊어 읽기 (운율구)	통사적 구 단위(수식어와 체언이나 용언의 묶음, 단 수식어를 뒤따라오는 피수식어가 단일어가 아니면 제외)
강한 끊어 읽기 (억양구)	통사적 구 단위(수식어 뒤에 오는 체언이나 용언이 단일어인지 아닌지 구분하지 않고 묶는다는 점만 중간 끊어 읽기와 차이가 남), 통사적 절 단위

그러나 2.2절에서 보았듯이, 끊어 읽기는 통사구(syntactic phrase)와 항상 일치하는 것은 아니며, 실제 발화에서는 화자의 의도에 따라 초어절 단위의 운율 경계는 가변적이어서 다양한 중의적 운율 단위가 발생할 수 있다. 가변적인 끊어 읽기 단위에 대한 정확한 예측을 위해서는 끊어 읽기 표지를 포함한 발음 전사 말뭉치로부터 추출한 통계 자료를 기반으로 한 더욱 체계적인 연구가 필요하다. 그러나 이런 말뭉치를 확보할 수 없는 상황에서 형태분석 말뭉치로부터 추출한 통계정보를 간접적으로 살펴볼 수 있다. 형태소, 형태 범주 패턴이 자주 연속하여 나타난다면 화자가 하나의 연속 단위로 인식하고 있다고 볼 수 있다. 다음 목록은 500 만 어절 말뭉치¹¹⁾로부터 형태소와 형태소 범주 두 가지 요소를 조합한 패턴을 조사한 결과 고빈도로 출현하는 패턴을 정리한 것이다.¹²⁾

9) <>: 형태소 범주/품사, /: 선택 항목, #: 띄어쓰기/어절 경계, +: 형태소 경계, (): 수의적 요소, ...: 가능한 공기 정보

10) 문장 수식 부사는 제외한다.

11) 부산대학교 한국어 정보처리연구실 형태 분석 말뭉치이며 약 500 만 어절 규모이다.

- <체언>+(와/과/은/는) > # {같다/관련하다/다르다/동일하다/비교하다/유사하다/틀리다}
 - 예) ‘한국+과 # 달리’
- <체언>+(에) # {관하다/대하다/비하다/비교하다/따르다}
 - 예) ‘논문+에 # 대해’
- <체언>+(와/과)# {같은} # <명사>
 - 예) ‘나무+와 # 같은’
- <수 관형사> # <명사/수 단위 의존명사>
 - 예) ‘세 # 참가자’
- <수사> # <명사>
 - 예) ‘팔팔 # 올림픽’, ‘이천이 # 동계유니버시아드’
- <수사+수사> # {기념식/광복/사태/운동/의거/전쟁/혁명}
 - 예) ‘12.12 # 사태’
- <수사># {분/점}(+<조사>) # <수 관형사/수사>
 - 예) ‘사분+의 # 일’, ‘십이 # 점 # 일’
- <수사> # {대} # <수사>
 - 예) ‘삼 # 대 # 이’
- <수 관형사> # {년/시/할} # <수 관형사> # {분/월/푼} # <수 관형사> # {리/일/초}
 - 예) ‘이천오 # 년 # 이 # 월 # 이 # 일’
- {강력/회계} # <수 관형사>+(반/부/팀)
 - 예) ‘강력 # 일반’, ‘회계 # 삼팀’
- <수 관형사> # <수 단위 의존명사>(+<조사>) # {내/들/등/만/말/밖/선/새/외/중/치/뽕/후}
 - 예) ‘두 # 시간 # 만+에’, ‘세 # 명 # 외’, ‘삼 # 일 # 중’
- {만} # <수 관형사> # <수 단위 의존명사>
 - 예) ‘만 # 열일곱 # 살’, ‘만 # 삼 # 개월’

이와 같이 형태소와 범주 패턴이 빈번하게 자주 출현한다는 것은 화자가 이러한 형태 범주 조합 패턴을 하나의 단위로 인식하고 있다는 것이므로 이들을 끊어 읽기 단위로 사용할 확률이 높다고 볼 수 있다. 즉 형태·통사적 단위를 기반으로 한 일반적 발화에서는 경계가 생기는 연결이지만 초어절의 운율 단위에 포함될 수 있는 후보이다. 통계에 기반을 둔 이런 초어절 패턴을 추출하여 휴리스틱을 확보할 수 있다면 전체 문장 분석(full sentence parsing)이 불가능한 부분 분석(partial parsing) 상황에서도 효율적인 끊어 읽기 예측 시스템을 구축할 수 있을 것이다.

또한, 화자가 의도에 따라 끊어 읽기가 가능하다 하더라도 언어적인 직관을 벗어나지 않도록 의미 전달에 방해가 가지 않는 선에서 끊어 읽기를 한다. 이런 맥락에서 ① 관용어구, ② 인용구 등은 여러 어절로 구성되어 있지만 하나의 의미 단위로 묶이므로 끊어 읽기 단위라고 볼 수 있다.¹³⁾

12) 형태소, 용언은 기본 형태를 { }에 밝혔다.

13) 관용어구와 인용구 경계 안이나 밖에서는 가변적 끊어 읽기가 가능하나, 설명의 편의상 관용어구와 인용구와 관련한 끊어 읽기만 표현한다.

- 관용어구 포함 문장

예) 그 사람은 기고만장해 | / || 하늘 높은 줄 모르고 날뛰더라.

- 인용구 포함 문장

예) 뮤직비디오에도 출연해 | / || “소년들이여 꿈을 가져라.”라고 외치는 연기를 소화했다.

끝으로, 끊어 읽기는 물리적인 요소에 의해 가변적이다. 사람이 일정한 간격으로 호흡을 해주어야 한다. 이러한 호흡 활동이 발화와 바로 연관되므로, 끊어 읽기는 ① 어절 내의 음절 개수, ② 한 문장 (구/절 포함) 내의 어절의 개수, ③ 한 문장 내에서의 어절의 위치, ④ 말하기 속도 등과 같은 물리적 요소에 영향을 받는다. 예를 들어, 문장 내에 어절이 많다면 아무리 하나의 통사구나 의미 전달의 단위로 묶어야 하더라도 두 개의 끊어 읽기 단위로 분리될 것이다.

4. HHU의 중의성 유형

P2G에서 기본적으로 어휘를 중심으로 발생할 수 있는 중의성은 상이한 철자로 구현되는 유형이 가장 뚜렷한 것이며 리형태(1994)는 바로 어휘 단위의 동음이절어 사전이다. 그러나 VR 시스템이 작동하는 발화 상황에서는 음운 변동이 다양하게 나타나므로 어휘 항목 이외에도 중의성을 이루는 단위가 많다. 현존하는 음성 발음치에 나타난 결과만을 가지고 그 유형을 살피는 것은 자료의 정확성과 균질성이 보장되지 않으며 중요한 유형이 빠져서 불완전한 유형을 도출할 가능성이 크다. 그러므로 음성 발음치를 분석하기에 앞서 이론적으로 가능한 한국어의 음운 현상을 확인하여 예상되는 도출형을 파악하고 이를 유형화의 기본 작업으로 삼아야 한다. 4.1 절에서는 HHU를 생성하는 한국어의 일반적인 음운현상을 살펴보고, 이를 바탕으로 4.2 절에서는 ‘끊어 읽기 단위’ 내부의 구성 성분별, 상호 단위 간 HHU의 유형을 분류한다.

4.1 음운 현상

한국어에서 음성 기호열 전사 단계에서 중의성을 발생시키는 음운 현상으로는 연음, 교체, 탈락, 축약을 들 수 있다. 음운 교체는 선행하거나 후행하는 자음의 자질에 따라 후행하거나 선행하는 자음이 교체되는 것으로 미파화, 자음동화(비음화, 유음화, 구개음화), 경음화, 유기음화를 들 수 있다. 이중 자음동화는 선행하거나 후행하는 자음의 자질에 의해 후행하거나 선행하는 자음이 동화되는 것으로 자음을 구성하는 어떤 자질의 값이 인접한 자음을 구성하는 자질의 영향으로 그와 같은 자질의 값을 가지게 된다. 이러한 자음동화에는 필수적 변동인 비음화, 유음화, 구개음화가 있고 수의적 변동인 연구개음화, 양순음화가 있는데, 수의적 변동은 표준발음으로 인정하지 않으므로 대상에서 제외한다.

- 연음화

앞 음절의 종성이 뒤 음절의 초성으로 발음 나는 현상.

① 어휘 내부: 예) ‘곡우’ [고구]

② <어간(체언, 용언)> + <어미/조사>: 예) ‘내각+이’ [내가기]

- 미과화

음절 말 장애음 종성이 어절과 같은 끊어 읽기 경계 마지막이나 이웃하는 음절의 장애음 초성 앞에서 파열되지 않고 [ㅂ, ㄷ, ㄱ]로 교체되는 것.

- ① 어휘 내부: 예) '짚' [집]
- ② <어간(체언, 용언)> + <어미/조사>: 예) '짚과' [집과]

- 비음화

평음 또는 미과화된 음절 말 [ㅂ, ㄷ, ㄱ]가 이웃하는 비음 초성 앞에서 동일 조음 위치의 비음으로 변동하는 현상.

- ① 어휘 내부: 예) '국문' [궁문]
- ② <어간(체언, 용언)> + <어미/조사>: 예) '죽+만' [중만]

- 유음화

음절말 유음 초성이나 종성과 이웃하는 치경음이나 후음이 [ㄴ]나 [ㄹ]로 바뀌는 것.

- ① 어휘 내부: 예) '산림' [살림]
- ② <어간(체언, 용언)> + <어미/조사>: 예) '활+는다' [할른다]

- 구개음화

음절 말 [ㄷ, ㅌ]가 모음 [ㅣ]로 이어질 때 구개음인 [ㅈ, ㅊ]으로 바뀌는 현상.¹⁴⁾

- ① <체언> + <조사>: 예) '술+이', '숯+이' [수치]
- ② <용언 어간> + <접사>: 예) '받히다' [바치다]
- ③ <어근> + <접사>¹⁵⁾: 예) '곧-이' [고지], 말-이[마지]

- 경음화

평음 [ㅂ, ㄷ, ㄱ, ㅈ]가 경음 [ㅃ, ㄸ, ㄱ, ㅊ]로 바뀌는 현상.

- ① 음운론적 제약을 받지 않은 어두 경음화: 예) '과사' [파사], '달리다' [딸리다]¹⁶⁾
- ② 어휘 내부: 예) '독기' [도끼/독끼]
- ③ 형태소 경계에서의 음운론적 경음화
 - <장애음으로 끝나는 체언 어간> + <조사>: 예) '낮+도' [나또/날또]
 - <장애음이나 비음으로 끝나는 용언 어간> + <어미>: 예) '삼+지만' [삼찌만]
- ④ 형태소 경계에서의 형태·음운론적 경음화: 예) '잠+자리' [잠짜리]

- 유기음화

평음 [ㅂ, ㄷ, ㄱ, ㅈ]가 'ㅎ'을 만나 유기음 [ㅃ, ㄸ, ㄱ, ㅊ]로 교체되는 현상.

- <어간(체언, 용언)> + <어미/조사>: 예) '놓+지' [노치], '죽+하고' [조카고]

14) 한국어에서는 [ㄸ]로 끝나는 명사나 용언은 없으므로 [ㄸ]이 [ㅊ]으로 바뀌는 예는 없다.

15) 구개음화는 실질 형태소와 형식 형태소가 만나는 경계에서만 일어나며, 단일어 내부나 합성에서는 발생하지 않는다.

16) 수의적으로 일어나는 발음이나 요즘 빈도가 높게 일어나는 현상이다. 예에서 '과사'는 '학과 사무실'의 준말이고 '달리다'는 '모자르다'라는 의미로 사용되는 경우다.

- 단모음화¹⁷⁾
 - 구개음 뒤의 반모음 탈락으로 ([저, 처, 쯤]가 [저, 처, 쯤]로 바뀌는 현상: 예) ‘처서’ [처서]
 - [-i]가 [캐, ㅣ]로 바뀌는 현상: 예) ‘무니’ [무니]
- 후음 [ㅎ]의 탈락
 - 후음 [ㅎ]이 유성음과 이웃해 있을 때 탈락하는 현상.
 - ① 어휘 내부: 예) ‘만희(曼姬)’ [마니]
 - ② <용언 어간> + <어미>¹⁸⁾: 예) ‘싫+은’ [시른]

4.2 끊어 읽기 단위별 중의성 유형

P2G 과정에서 동음이철 현상이 나타날 수 있는 최대 언어 단위는 ‘끊어 읽기 단위’이며 그 구성 성분의 음운적·형태적 지위, 품사 등에 따라 동음이철을 발생하는 음운현상이 다르다. 먼저 한국어에서 음운 현상이 일어날 수 있는 단위로 ① 음운 현상이 어휘 내부의 음소 연쇄 과정에서 생기는 필수적이고도 보편적인 음운규칙이 발생하는 기초단위인 형태소나 어휘, ② 체언과 조사, 용언의 활용, 체언의 결합 등 형태론적 경계에서 발생할 수 있는 음운 현상의 단위인 어절, ③ 어절의 경계를 넘어서 하나의 발음 단위로 인식되는 통사적 연쇄의 단위인 초어절을 들 수 있다. 4.1 절에서 살펴본 것과 같은 한국어에서 나타나는 음운 현상을 중심으로 형태소·어휘 내부, 어절 내부, 어절 경계를 넘어서 초어절 단위를 대상으로 적용하여 결과형을 검토하면 중의성의 유형을 <표 2>~<표 4>처럼 분류할 수 있다.¹⁹⁾

표 2. 어휘 내부의 동음이철 유형

음운현상	동음 음성 기호열	이철	
		음운변동	음운현상 미적용
연음화	[고구]	곡우	고구(考究)
미파화	[집]	짚	집
비음화	[궁문]	국문	궁문(宮門)
유음화	[살림]	산림	살림
경음화	[도끼]	독기	도끼
단모음화	[자미]	잠의	자미(姿媚)
ㅎ탈락	[마니]	만희(曼姬)	마니(摩尼)

17) 한국어에서 발생하는 음운 탈락 현상은 다양하다. 그러나 이 중에 먼저 구개음 다음에 반모음 [j]가 탈락하는 현상이나 일부의 ‘ㅎ’의 탈락 현상을 제외하고는 이미 활용으로 그 형태가 고정되므로 동음이철형을 발생시키지 않는다.

18) 어미 ‘-어도’나 ‘-은’이 용언의 어간에 통합하는 경우에 ‘ㅎ’탈락의 보상적 장음을 고려하지 않았을 때 HHU가 발생한다. 하지만, 최근의 음성합성 관련 연구들은 장음과 단음의 구분에 대해 의견을 달리하는 경우가 많지만 대체로 세대에 따라 다른 양상을 보인다는 데에는 의견의 일치를 보고 있다.

19) <표 2>~<표 4>는 부산대학교 이근열 선생님과 논의 결과를 포함하였다.

표 3. 어절 내의 형태소 경계 포함 동음이철 유형

음운현상	동음 음성 기호열	이철	
		음운현상 적용	음운현상 미적용
연음화	[내가기]	내각+이	내가+기
미과화/경음화	[집과]	집+과	집+과
비음화	[중만]	죽+만	중+만
구개음화	[바치고]	받+히고, 받+이고	바치+고
	[바지]	받+이	바지
경음화	[찌끼/찍기]	찍+기	찌끼
유기음화	[조카고]	죽+하고	조카+고
단모음화	[처서]	치+어서=처서	처서(處暑)
ㅎ탈락/연음화	[시름]	싫+은, 실+은	-
	[시름]	싫+음	시름

표 4. 초어절 단위의 동음이철 유형

어절/형태소 경계	음운현상	동음 음성 기호열	이철	
			패턴 1	패턴 2
용언+관형형 어미 # 명사	연음화	[그러나이는]	그런 아이는	그러나 이는
	유음화	[과는]	내가 <u>관</u> 은목걸이	내가 <u>과</u> 는 목걸이
(수)관형사 # 명사	연음화	[이가지]	그곳은 종로3가가 아니고 <u>2</u> 가지.	<u>이</u> 가지 잘라야 해.
	경음화	[전체육교사용]	전 체육 교사용	전체 육교 사용
명사+관형격 조사 # 명사	단모음화	[하교에서고]	학교의 서고(書庫)	학교에서고
부사 # 용언	연음화	[다다지지]	돼지고기는 <u>다</u> 다지지 않고 넣는 것이 더 맛있기도 해.	문이 제대로 닫아지지 않고 반쯤 열려 있던걸.
	유기음화	[이러케도와]	비가 오는데 <u>이렇게도</u> 와 주신 분이	어려운 사정임에도 <u>이렇게</u> <u>도와</u> 주신 분이
본용언 (#) 보조용언	경음화, [ㅎ]탈락 연음화	[잡찌어나도]	내가 <u>잡지</u> <u>않아도</u> 남을 것이다.	가벼워 보이는 <u>잡지</u> [雜誌] <u>안아도</u> 무게는 마찬가지다.
논항(주어/보어) # 용언	연음화	[저니가]	<u>전</u> 이가 빠졌어요.	<u>전이</u> (轉移)가 됐어요.
	-	[나물]	나 물 좀 먹자.	나물 좀 먹자.
		[감사는]	<u>감</u> 사는 저 사람에게 물어보세요.	<u>감</u> 사는 저 사람에게 하세요.
		[나가]	나 가?	집에서 나가!

<표 2>와 <표 3>에서 볼 수 있듯이 어휘 내부와 어절 내부의 HHU가 형태소 경계에 민감한 음운변동의 결과로 생성된 것이라면, 어절과 어절이 긴밀하게 연쇄되는 경우는 발음의 편의를 위한 음운 동화가 필수적으로 일어나며 화자의 발음 의도에 따라 다양한 경계의 변화가 이루어질 수 있다. 이러한 경계의 변화와 음운 현상은 <표 4>에서와 같이 다양한 초어절 HHU를 생성한다.

초어절 단위의 HHU는 형태소나 띄어쓰기 경계에 민감하다. 지금까지 언어학에서는 어휘 내부의 HHU만 연구 대상으로 삼았으나, P2G에서 발생하는 중의성 해결에는 도움이 되지 못한다. 따라서 어절 내부의 중의성 패턴과 초어절 단위의 중의성 패턴 분석까지 연구를 확장하게 된다면 그 결과를 언어 이해 과정에서 발생할 수 있는 중의성 해소를 위한 지식베이스 구축에 응용할 수 있다.

5. 이어 가기

이 논문은 한국어 VR 시스템에서 P2G의 정확도를 개선하고자 대규모 HHU 말뭉치를 구축하고 언어 이해 과정의 초어절 중의성 해소 지식베이스를 구축하기 위한 기초 연구로서 한국어 끊어 읽기 단위 생성에 관여하는 언어 단위의 특성을 분석하고 끊어 읽기 단위인 HHU의 중의성 유형을 분류했다.

음성자료나 음성 기호열로 전사된 결과로부터 언어정보를 언어 문자 기호열로 출력을 하면 중의성 발생 범위는 형태소에서부터 초어절에까지 확장된다. 이러한 언어 이해 단계에서 발생할 수 있는 운율 단위 중의성을 해소하기 위해서는 초어절 중의성 해소 지식베이스가 필요하다. 이 지식베이스는 본 논문의 분석 결과를 바탕으로 하여 추출될 수 있는 ① 중의성 단위(어절, 초어절)의 형태·통사적 결합 정보, ② 인용구, 관용구 인식 정보, ③ 중의성 단위(어절, 초어절)의 맥락 정보, ④ 중의성 단위와 관련된 통계 정보(중의성 단위의 빈도, 중의성 단위가 포함된 문장의 어절 수, 문장 내에서의 중의성 단위의 위치)를 포함할 수 있다. 또한, 기존 연구의 결과로 얻어진 ⑤ 중의성 단위의 기본주파수, 중의성 단위 내에 포함된 음절의 길이, 발화 속도 등의 음성적 정보를 포함한다면 중의성 해소 성능이 높아질 것이다.

초어절 중의성 해소 지식베이스는 음성인식 기술의 연구뿐만 아니라 기계 번역, 정보 검색, 텍스트 마이닝 등 다양한 언어처리 기반기술 연구를 위한 자원으로 활용되어, 언어처리 기술 전반에 걸친 연구를 활성화할 것이다. 특히, 언어 이해에 기반을 둔 우리말 의미 분석 기술을 발전시킬 수 있다. 지능형 로봇, 디지털 콘텐츠, 홈네트워크, 텔레매틱스, 차세대 PC 등은 인간의 인지적 능력을 기계가 이해하여 인간 중심의 지능적인 서비스를 강조한다. 지능적인 서비스라 함은 인간의 의미 개념을 기계가 이해하고 인간과 기계 간의 정보교환이 인간의 언어로 이루어지는 인터페이스의 실현을 의미한다. 인간의 언어는 철자가 같은 어휘가 의미가 다르거나 철자가 다르더라도 발음생성 과정에서 발음이 같아져 의미 혼동을 가져오는 경우가 많다. 따라서 초어절 중의성 해소 지식 베이스는 음성언어처리과정에서 같은 발음 때문에 의미 해석이 달라질 수 있는 경우에 대한 초어절 중의성 유형 분석 결과를 음성인식 또는 음성합성에서의 의미 중의성 해소 기술 개발에 직접 이용할 수 있다.

참 고 문 헌

- 강범모. 2005. “동음이의어의 사용 양상.” *언어연구* 41(1), 1-29.
- 강옥미. 1994. “한국어의 음절화.” *어학연구* 30(3), 535-567.
- 권오일, 홍문기, 강선미, 신지영. 2002. “코퍼스 방식 음성합성에서의 개선된 운율구 경계예측.” *음성과학* 9(3), 25-34.
- 권재일, 김운한, 문양수, 남승호, 전종호. 1997. “통사 구조와 운율 구조의 상관성 연구: 중의성 해소 양상을 중심으로.” *언어학* 20, 103-109.
- 김병균. 2000. *한국어 동음어사전*. 서울: 태학사.
- 김상훈, 성철재, 이정철. 1997. “운율구 경계현상 분석 및 텍스트에서의 운율구 추출.” *한국음향학회지* 16(1), 24-32.
- 김선미. 2002. *한국어의 리듬 단위와 문법 구조: 음성 합성에서 리듬 구현의 자연성 향상을 위한 음성·언어학적 연구*, 서울대학교 박사학위 청구논문.
- 김선희, 전재훈, 홍혜진, 정민화. 2008. “음성정보와 문법정보를 이용한 한국어 운율 경계의 자동 추정.” *말소리* 66, 117-130.
- 김성규. 1999. “잠재적 휴지의 실현.” *선정어문* 27, 831-860.
- 김승원, 김병창, 정민우, 이근배. 2005. “CRF를 이용한 한국어 운율 경계 추정.” *제17회 한글 및 한국어 정보처리 학술대회 논문 발표집*, 134-138.
- 리형태, 류은중. 1994. *동의어, 반의어, 동음어 사전*. 서울: 한국문화사.
- 신지영. 2000. *말소리의 이해*, 서울: 한국문화사.
- 원영섭. 1994. *동음동의이차어*. 서울: 세창출판사.
- 유혜원, 남경완, 홍종선. 2006. “한국어의 형태론적 중의 어절 사전 구축과 표제어 선정.” *한국어학* 31, 279-313.
- 윤미경. 1975. “국어 동음어에 대한 연구.” *군자어문학* 2, 21-33.
- 이경민, 전배훈, 정민화. 2001. “한국어 연속음성인식을 위한 발음열 자동 생성.” *한국음향학회지*, 20(2), 35-43.
- 이찬도. 1997. “음성인식·합성을 위한 한국어 운율 단위 음운론의 계산적 연구: 음운단위에 따른 경계의 발견.” *한국정보처리학회 논문지* 4(1), 280-287.
- 이호영. 1996. *국어 음성학*. 서울: 태학사.
- 임희석, 이호, 임해창. 1993. “형태소분석 단계에서 발생하는 어절의 중의성 분석 방안.” *제20회 한국정보과학회 봄 학술발표논문집* 20(1), 773-776.
- 전진욱, 김한우, 김동건, 이양희. 2002. “한국어 TTS 시스템을 위한 운율구 경계 예측.” *2002년도 한국음향학회 학술발표대회 논문집* 21(1), 77-82.
- 황화상, 최정혜. 2003. “연구논문: 한국어 어절의 형태론적 중의성 연구.” *한국어학회* 20, 287-311.
- Kim, B. C., Lee G. B. & Lee. J. H. 2002. “Morpheme-Based Grapheme to Phoneme Conversion Using Phonetic Patterns and Morphophonemic Connectivity Information.” *ACM TALIP* 1(1), 65-82.
- Lee, D. G. & Kwon. H. C. 2002. “Automatic String Generator based on Standard Korean Pronunciation.” *APIS II*, Jakarta, Indonesia, 47-51.
- Taylor, P. & Black. A. W. 1997. “Assigning Phrase Breaks from Part-of-Speech Sequences.” *In Proceedings of Eurospeech 1997*. Rhodes. 995-998.

접수일자: 2008. 10. 30

수정일자: 2008. 11. 28

게재결정: 2008. 12. 8

▲ 윤애선

부산광역시 금정구 장전2동 산 30 (우: 609-735)

부산대학교 불어불문학과 교수

Tel: +82-51-510-2049 Fax: +82-51-583-0996

E-mail: asyoon@pusan.ac.kr

▲ 강미영(교신저자)

서울특별시 강서구 방화3동 금남화길 148 (우: 157-857)

국립국어원 학예사

Tel: +82-2-2669-9754 Fax: +82-2-2669-9787

E-mail: mykang@mcst.go.kr