pISSN 2005-8063
eISSN 2586-5854
2018. 03. 31.
Vol.10 No.1
pp. 49-55

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

https://doi.org/10.13064/KSSS.2018.10.1.049



# 지식 기반 프랑스어 발음열 생성 시스템

## A knowledge-based pronunciation generation system for French

김 선 희*

Kim, Sunhee

**Abstract**

This paper aims to describe a knowledge-based pronunciation generation system for French. It has been reported that a rule-based pronunciation generation system outperforms most of the data-driven ones for French; however, only a few related studies are available due to existing language barriers. We provide basic information about the French language from the point of view of the relationship between orthography and pronunciation, and then describe our knowledge-based pronunciation generation system, which consists of morphological analysis, Part-of-Speech (POS) tagging, grapheme-to-phoneme generation, and phone-to-phone generation. The evaluation results show that the word error rate of POS tagging, based on a sample of 1,000 sentences, is 10.70% and that of phoneme generation, using 130,883 entries, is 2.70%. This study is expected to contribute to the development and evaluation of speech synthesis or speech recognition systems for French.

**Keywords:** pronunciation generation, French, speech synthesis, speech recognition, knowledge-based

## 1. 서론

전통적인 방식의 음성합성 시스템, 혹은 텍스트 음성 변환(TTS, text-to-speech) 시스템은 텍스트를 입력하여 음성 기호로 변환하는 언어처리부와, 언어처리부에서 생성된 음성 기호로부터 신호를 생성하는 신호처리부로 구성된다(Allen *et al.*, 1987; Taylor, 2009). 최근 사용되고 있는 음성합성은 신호처리 방식에 따라 Unit Selection 방식과 통계적 파라미터(Statistical Parametric) 방식으로 나뉘고, 통계적 파라미터 방식은 다시 hidden Markov model(HMM) 방식(Tokuda *et al.*, 2013)과 최근의 deep neural network(DNN) 방식에 의한 합성으로 발전되어 왔다. DNN 방식에 의한 합성은 초기에 신호처리부의 음향모델링에 적용되어 음성 합성부의 성능 개선에 기여하였으나(Zen *et al.*, 2013; Zen & Sak, 2015), 최근 end-to-end 방식이라는, 언어처리부 없이 텍스트를 입력하여 바로 신호를 합성해 내는 방식들이 활발하게 연구되고 있고(Oord *et al.*, 2016; Wang *et al.*, 2017; Arik *et al.*, 2017; Shen *et al.*, 2017; Sotelo *et al.*, 2017), 또 상용화되고 있다.

이는 그동안 Unit Selection 방식과 통계적 파라미터 방식에서 공통적으로 음성합성의 한 축으로 간주되던 언어처리부 자체가 실질적으로 사라지고, 문자 단위로 입력된 텍스트로부터 바로 음성을 생성하는 end-to-end 방식의 음성합성 시스템으로 빠르게 진화해 가는 것을 시사한다. 이런 상황에 전통적인 음성합

* ㈜네이버, kim.sunhee@navercorp.com

Received 19 February 2018; Revised 28 March 2018; Accepted 28 March 2018

성의 언어처리부에 대한 연구는 음성합성에서 조만간 더 이상 의미 없는 일이 될 수도 있을 것으로 보인다. 그러나, 현재 상용화 단계에 이른 언어들이 전통적 방법에 의한 오랜 연구에 바탕하고 있고, 고품질의 합성음을 개발하기 위해 아직까지는 기존 방법론에 대한 연구와 새로운 방법론에 대한 연구가 병행적으로 이루어져야 한다.

발음열 생성은 음성합성 및 음성인식 시스템을 구성하는 주요 모듈로, 기본적으로 표제어와 이에 해당하는 발음으로 구성된 발음사전과, 발음 사전에 포함되지 않은 단어들에 대한 입력 텍스트를 그에 상응하는 발음열로 변환하는 자소-음소 변환(Grapheme-to-Phoneme Conversion) 모듈로 구성된다. 음성인식의 경우, 다양한 사람들의 여러 발음을 포섭하기 위해 주어진 단어의 가능한 발음변이를 포함하도록 구성하는데 반해(Hahn *et al.*, 2012), 음성합성의 경우, 일반적으로 한 명의 화자를 기준으로 하여 음성을 생성하기 위한 표준 발음 위주로 구성된다. 자소-음소 변환 연구 역시 크게 규칙 기반 연구(Béchet, 2001)와 데이터 기반 연구로 나누어 볼 수 있는데, 이 중 데이터 기반 방식에서는 결정트리를 이용하거나 유추에 의한 발음 추정 방법, HMM, n-gram 모델링, DNN 모델링 등 다양한 방식이 시도되었다(Black *et al.*, 1998; Taylor, 2005; Marchand & Damper, 2000; Van Den Bosch & Canisius, 2006; Jiampojamarn & Kondrak, 2010; Rao *et al.*, 2015).

한국어나 프랑스어처럼 철자를 기준으로 단어(한국어의 경우는 어절) 내부와 단어와 단어 사이에서 많은 음운현상이 관찰되는 언어의 자소-음소 변환의 문제는 지식 기반 시스템이 좋은 성능을 보이는 것으로 보고되었다(Yoon & Brew, 2006; Lecorvé & Lolive, 2015). 특히 프랑스어의 경우에 오랜 기간의 연구를 기반으로 한 규칙 기반 시스템들을 개발해 왔고, 현재까지 가장 우수한 성능을 보이는 것으로 보고되어 있다(Lecorvé & Lolive, 2015; Yvon *et al.*, 1998; de Mareüil *et al.*, 2005). 이와 같이, 언어지식을 기반으로 한 프랑스어 발음열 생성 시스템이 현재까지 다른 접근법보다 좋은 성능을 보이고 있음에 따라, 음성합성에 있어서 언어처리부를 배제한 모델인 end-to-end 시스템이 주요 연구 과제가 되었고, 자소-음소 변환을 포함한 언어처리부 연구에 있어서도 기계 학습 방법론이 대세인 상황임에도 불구하고, 본 논문은 이와 같은 지식 기반 프랑스어 발음열 생성 시스템에 대하여 기술하는 것을 그 목적으로 한다. 이러한 지식 기반 시스템과 이에 대한 평가 방법 및 결과는 이후 다른 방법론을 적용하여 성능을 개선하는데 있어서 의미 있는 기여를 할 수 있을 것으로 생각된다.

이후 논문 구성은 다음과 같다. 2장에서 프랑스어 자소-음소 변환 시스템을 소개한다. 3장에서는 이에 대한 평가 방법을 제안하고 평가 방법에 따른 결과를 제시하고, 4장의 결론으로 논문을 마무리한다.

## 2. 프랑스어 철자 체계 및 음소 체계

### 2.1. 프랑스어 철자 체계

프랑스어의 철자는 아래 (1)과 같이 기본적으로 알파벳 26글자를 기본으로 대문자와 소문자를 사용하고, 여기에 더해 모음에 4개의 발음 구별 부호(diacritics)를 사용한다: Accent aigu(´), Accent grave( ` ), Accent circonflexe( ^ ), Tréma( ¨ ). 이러한 발음 구별 부호는 모음의 음가를 결정하는 역할을 하며, 각 단어의 강세 위치에서 실현된다. 자음에 나타나는 발음 구별 부호로는 “cédille”가 있다. 일반적으로 음소 /a, o/ 앞에서 /c/는 [k]로 실현되는데, 동일한 환경에서 “c-cédille”는 /s/로 발음되도록 한다. 위 5가지 발음 구별 부호를 갖는 글자는 아래 (2)와 같이 16개이다.

(1) Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm Nn Oo Pp Qq Rr SsTt Uu Vv Ww Xx Yy Zz
(2) Àà Ââ Ää Çç Éé Èè Êê Ëë Îî Ïï Ôô Öö Ùù Ûû Üü Ÿÿ

이 외에도 두 모음이 결합된 형태인 /æ/, /œ/ 가 나타나는데, 이는 종종 두 개의 모음 /ae/, /oe/ 와 혼용된다. 따라서, 대문자, 소문자 각 44개로 총 88개의 글자가 사용된다.

### 2.2. 프랑스어의 음소 체계

프랑스어는 17개의 자음, 3개의 반모음, 12개의 구강 모음과 4개의 비모음으로 구성된다. <표 1>은 프랑스어 음소표를 국제 음성 기호(IPA, international phonetic alphabet)와 X-SAMPA(extended speech assessment methods phonetic alphabet: http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm)로 표기한 것이다.

**표 1.** 프랑스어 음소표의 국제 음성 기호(IPA)와 X-SAMPA 표기
**Table 1.** French Phoneme symbols in X-SAMPA and IPA

| X-SAMP | IPA | Examples |
|---|---|---|
| A | a, ɑ | [mAtin, pAs] |
| A~ | ɑ̃ | [VANtardise, tEMPS] |
| @ | ə | [pEtit] |
| 2 | ø | [crEUser, dEUx] |
| 9 | oe | [malhEUreux, pEUr] |
| E | ɛ | [pERdu, modEle] |
| e | e | [Emu, otE] |
| E~ | ɛ̃, oɛ̃ | [pEINture, IUNdi] |
| i | i | [Idēe, amI] |
| O | ɔ | [Obstacle, cOrps] |
| o | o | [AUditeur, bEAU] |
| O~ | ɔ̃ | [rONdeur, bON] |
| u | u | [cOUpable, IOUp] |
| y | y | [pUnir] |
| w | w | [OUi, Olseau] |
| H | ɥ | [hUlle] |
| j | j | [plētiner, paiLLe] |
| p | p | [Phre, caPe] |
| t | t | [Terre, raTe] |
| k | k | [Cou, saC] |
| b | b | [Bon, roBe] |
| d | d | [Dans, aiDe] |
| g | g | [Gare, baGUe] |
| f | f | [Feu, PHare] |
| s | s | [Sale, taSSe] |
| S | ʃ | [CHanter, maCHine] |
| v | v | [Vous, rjVe] |
| z | z | [Ziro, maiSon] |
| Z | ʒ | [Jardin, manGer] |
| l | l | [Lent, giLet] |
| N | ŋ | [campINg] |

X-SAMPA, extended speech assessment methods phonetic alphabet; IPA, international phonetic alphabet

## 3. 프랑스어 발음열 생성 시스템

영어와 비교하여 프랑스어는 동사 변화가 훨씬 다양하고 복잡하며, 명사 역시 여성과 남성의 성(gender)을 가지며 변화할 뿐만 아니라, 명사와 형용사 및 동사와의 일치 문제가 매우 복잡한 언어라고 할 수 있다. 이러한 특성에 따라 프랑스어의 경우, 발음 변환이 형태소 분석과 좀 더 밀접하게 연결되어 있다고 할 수 있다. 대부분의 프랑스어 발음열 생성 연구는 형태소 분석뿐만 아니라, 텍스트 정규화(text normalization)까지 모두 자소-음소 변화 시스템에 포함되어 있는 경우가 많다(Yvon *et al*., 1998; Béchet, 2001; de Mareüil *et al*., 2005; Rao *et al*., 2015). 본 논문에서도 형태소 분석은 자소-음소 모듈에 포함한다. 대상이 되는 NSW(non-standard word)를 일반 단어로 변환하는 모듈인 텍스트 정규화 모듈은 독립적으로 다루기로 하고, 본 논문에서는 논외로 한다.

<그림 1>은 입력 문장이 “L'équipe comprendra 1,500 anglais.” (그 팀에는 1,500명의 영국인이 포함될 것이다.)일 때, 마지막 단어인 ‘anglais’로부터 음소를 생성하는 과정을 보인다.

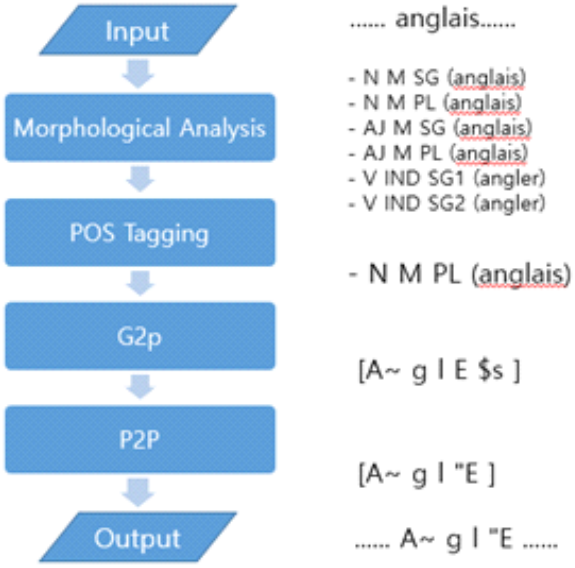


**그림 1.** 프랑스어 발음열 생성 시스템과 예
입력 문장 “L'équipe comprendra 1,500 Anglais.”에서 단어 ‘anglais’로부터 음소열을 추출하는 과정.
**Figure 1.** French pronunciation generation system
Derivation process of the word ‘anglais (English) from the input sentence ““L'équipe comprendra 1,500 Anglais.”

본 논문에서 제안하는 자소-음소 변환 시스템은 위 <그림 1>과 같이 형태소 분석, 품사 태깅, 자소-음소 변환, 음소-음소 변환의 4단계로 구성된다. 아래에서는 형태소 분석, 품사 태깅 단계와 자소-음소 변환 및 음소-음소 변환의 두 단계로 나누어 살펴보도록 하겠다.

### 3.1. 형태소 분석 및 품사 태깅

먼저, 형태론적 분석 단계에서는 문장을 구성하는 각각의 단어를 가능한 형태소로 쪼개고, 각각의 형태소에 따라 여러 가지 다른 의미를 가진 형태론적 자질들을 부여한다(Byrd & Tzoukermann, 1988; Gruaz *et al*., 1996). 본 논문에서는 형태소 분석과 이를 통한 형태소 태그에 대한 기술은 다루지 않기로 한다.

영어나 프랑스어의 경우, 한 단어가 2개 이상의 의미를 갖는 경우들이 종종 있다. 따라서 여러 가지로 다른 형태론적 분석 결과를 보이게 된다. 예를 들면, <그림 1>의 입력 문장에서 단어 ‘anglais’는 2가지 의미(lemma)를 갖는 여섯 가지 형태론적인 분석이 가능하다.

(1) “영국인”을 의미하는 남성 단수 명사
(2) “영국인”을 의미하는 남성 복수 명사
(3) “영국의”를 의미하는 남성 단수 형용사
(4) “영국의”를 의미하는 남성 복수 형용사
(5) “각지게 하다” 라는 의미의 1인칭 단수 (복합)과거 동사
(6) “각지게 하다” 라는 의미의 2인칭 단수 (복합)과거 동사

위와 같이 여섯 가지 형태론적 분석이 가능하다는 것은 그 단어가 중의성(ambiguity)을 갖는다는 것을 의미하며, 단어의 중의성은 다음의 품사 태깅(part-of-speech tagging) 과정에서 여러 규칙을 통해 가장 적합한 결과를 선택하는 중의성 해소(disambiguation) 과정을 거쳐 최종 결정된 품사와 이에 대한 형태론적 정보를 출력한다. 이때, 주어진 단어에 대한 품사와 함께 출력되는 형태론적 정보로는 위의 예에서 살펴본 바와 같이 명사나 형용사의 경우, 남성인지 여성인지를 구분하는 성(gender)과 단수와 복수를 구분하는 수(number) 태그가 함께 출력이 된다. 동사의 경우, 수(number/인칭), 시제(tense)를 같이 출력해 준다. 다음 <표 2>는 본 연구에서 사용한 프랑스어 품사 테그셋 목록이다.

**표 2.** 프랑스어 품사 테그셋 목록
**Table 2.** French POS set

| | | | |
|---|---|---|---|
| AJ | adjective | N | noun |
| AP | adposition | NU | numeral |
| AT | article | PD | pronoun/determiner |
| AV | adverb | PNCT | punctuation mark |
| C | conjunction | U | unique (ne, pas, etc.) |
| I | interjection | V | verb |

POS tagging, part-of-speech tagging

### 3.2. 자소-음소 변화 및 음소-음소 변환

본 연구에서는 형태소 분석을 거쳐 POS 태깅 후 철자 음소 변환 과정을 자소-음소 변환(G2P)과 음소-음소 변환(P2P)의 두 단계로 다시 나누었다. G2P는 입력으로 들어온 단어와 단어의 형태론적 정보로부터 자소에 상응하는 음소로 변환하고, 액센트를 부여하는 단계이다. P2P는 음소열로 변환된 단어들에 단어 내 혹은 단어와 단어 사이에서 적용되는 음운규칙들을 적용하여 최종 음소를 출력하는 단계이다.

프랑스어의 G2P 변환을 어렵게 하는 요인으로 두 가지를 들 수 있다(Yvon *et al.*, 1998). 첫 번째 요인은 프랑스 철자법의 매우 복잡한 규칙성과 규칙성에 상응하는 수준의 불규칙성이다. 예를 들어 동사 변화를 살펴보면 규칙동사들 외에도 많은 불규칙 동사들이 존재하고, 불규칙 동사들의 경우에도 일정한 규칙으로 정의되어 있어 실제로는 매우 복잡한 수많은 규칙이 존재한다고 할 수 있다. 두 번째 요인으로 주어진 단어의 발음이 형태론적 정보 등에 의해 정해지는 많은 동철이음어(Homograph)를 들 수 있다. 대표적인 예로 단어 ‘couvent’을 들 수 있다. 이 단어는 명사인 경우 '수도원'을 의미하고, 발음은 [k u v A~]이 된다. 동사인 경우 '(알을) 품다'라는 의미이며, ‘couver’ 동사 3인칭/복수 현재형으로 발음은 [k u v]가 된다.

G2P는 위와 같이 복잡한 프랑스어 철자법과 동철이음어를 반영하여 각 단어에 대한 음소열을 생성하는 단계로 1,000개 이상의 많은 규칙으로 정의된다. G2P 규칙과 함께 일반적으로 단어의 마지막 음절에 오는 액센트 부여 규칙도 적용된다.

<그림 1>에서와 같이 품사 태깅 과정의 결과물로 생성된 “N M PL (anglais)”은 G2P 과정을 거쳐 액센트(“)가 부여된 음소열 [A~ g l “E $s ]로 출력된다. 여기에서 [$s]는 단어 끝에 위치하는 잠정적인 음소 [s]를 표시하는 것으로, 다음의 P2P 과정에서 컨텍스트에 따라 [s]로 실현될 것인지, 실현되지 않을 것인지가 다시금 규칙에 의하여 결정된다. 이것이 유명한 프랑스어의 연음 Liaison의 문제이다. G2P 단계에서 각 단어의 품사에 맞게 출력된 음소열은 P2P 단계에서 컨텍스트를 고려하여 다시 최종 발음열을 생성하게 되는데, 이때 연음을 포함한 프랑스어의 음운현상들을 반영하는 규칙들이 포함되게 된다.

프랑스어 연음은 연음이 필수적으로 적용되어야 하는 경우(obligatory), 적용되어서는 안 되는 경우(forbidden), 수의적으로 적용될 수 있는 경우(optional)의 세 가지 경우로 나누어진다. 아래 문장(1)은 이 세 가지의 경우를 포함하는 예시이다(Yvon *et al.*, 1998).

(1) les enfants ont écouté (the children have listened)
[l e z A~ f A~ O~ (t) e k u t e]

‘les’와 ‘enfants’사이에 나타나는 /s/는 필수적 연음으로 [l e z A~ f A~] 으로 발음되어야 하고, ‘enfants’과 ‘ont’ 사이의 경우에는 연음이 일어나서는 안 된다. 그리고, ‘ont’과 ‘écouté’사이의 연음은 수의적으로, ‘ont’의 마지막 /t/역시 수의적으로 실현될 수 있다. 위 문장과 같은 경우에는 필수적인 규칙과 수의적인 규칙이 각각 모듈에 포함된다.

프랑스어에서 연음과 함께 발음 생성에서 문제가 되는 것은 ‘묵음-e(mute-e, 혹은 schwa)’라고 하는 중성 모음 /ə/의 탈락 문제로(Larreur & Sorin 1991), 단어 내부에서 혹은 단어와 단어가 연접하는 경우에 필수적으로 혹은 수의적으로 /ə/가 탈락되는 현상을 의미한다.

(1) lorsque vous [l O r s k (ə) v u]
(2) lorsque Helene [l O r s k ə e l E n]
/lorsque/ ‘~할 때’ /vous/ ‘당신’ Helene ‘헬렌’

(1)의 경우 /ə/의 탈락은 수의적이나, (2)와 같이 묵음 /h/ 앞에서는 탈락하지 않아야 한다.

## 4. 시스템 평가

본 논문에서 기술한 프랑스어 발음열 생성 시스템은 형태소 분석을 포함한 품사 태깅 과정과 자소-음소 변환 및 음소-음소 변환의 음소 생성 과정으로 구성된다. 시스템 평가를 위해 언어 전문가들에게 품사 태깅과 음소 변환 결과를 검증할 수 있는 테스트 셋을 작성하도록 의뢰한 후 수작업에 의해 생성된 결과물을 시스템에 따라 생성된 결과물과 비교하는 방식으로 시스템에 대한 평가를 수행하였다.

먼저, 품사 태깅의 평가를 위해 목표한 서비스 도메인인 뉴스 영역에서 다양한 주제를 포괄하는 1,000문장을 선정하도록 하였다. 각 문장은 5단어 이상 60단어 미만으로 구성되었고, 문장

당 평균 단어 수는 25.2개이다.

평가 결과는 아래 <표 3>과 같다. 전체 1,000문장을 구성하는 총 단어 수는 25,182개인데, 이 가운데 잘못 태깅 된 품사 수는 2,701개로 단어 오류율은 10.70%, 이러한 오류 단어를 포함한 문장은 총 797문장으로 문장 오류율은 79.70%이다.

**표 3.** 품사 태깅 평가 결과의 오류 수 및 오류율
**Table 3.** Number of errors and error rate of POS tagging results

| 총 문장 수 | 1,000 | 총 단어 수 | 25,182 |
|---|---|---|---|
| 오류 문장 수 | 797 | 총 POS오류 수 | 2,701 |
| % | 79.70% | % | 10.70% |

POS tagging, part-of-speech tagging

음소 생성 평가는 세 가지로 나누어 수행하였다. 첫 번째 평가는 원어민 언어 전문가 두 명이 동일한 문장 310개에 대해 발음열 정답지를 만들고, 시스템에 의해 산출된 결과와 비교하였다. 평가 결과, 전체 토큰 수 대비 원어민 정답 토큰과의 비율(precision=correct/all)은 각각 0.89와 0.88로서 평균값은 0.89다.

두 번째 평가는 시스템으로부터 생성된 어휘 130,883개 표제어에 대한 평가로 기성 발음 사전과 비교했다. 기성 발음 사전의 결과를 정답으로 하였을 때 기성 발음 사전과 차이를 보이는 5,608개의 표제어를 검출하였다. 이렇게 검출된 표제어는 다시 언어 전문가의 검토를 통해 재확인한 결과, 151개의 표제어 발음이 오류를 보여 2.7% 오류율을 확인할 수 있었다.

위의 두 가지 발음 평가와 더불어 추가적으로 연음에 대한 평가를 수행하기 위해, <표 4>와 같이 프랑스의 아카데미 프랑세즈(http://www.academie-francaise.fr/questions-de-langue#46_strong-em-liaisons-em-strong)에서 규정한 7가지의 통사적 연음 필수 환경에 의거한 총 524건의 테스트 셋을 작성하여 시스템 결과와 비교하였다. 그 결과, 146건의 오류가 나타났고, 28% 오류율이 도출되었다.

**표 4.** 연음에 대한 평가 결과의 오류 문장 수 및 오류율
**Table 4.** Number of errors and error rate of pronunciation generation results for Liaison cases

| 연음 환경 | 해당 문장 수 | 오류 문장 수/% | |
|---|---|---|---|
| 한정사 + 피한정어 | 77 | 1 | 1% |
| 형용사 + 피수식어 | 108 | 21 | 27% |
| 대명사 + 동사 | 105 | 18 | 17% |
| 비인칭구문 혹은 소개사(c'est)구문에서 est+후행어 | 42 | 17 | 40% |
| 부사 + 수식받는 단어 | 70 | 27 | 39% |
| 1음절 전치사 + 후행어 | 67 | 16 | 23% |
| 복합어 및 관용구 내부 | 55 | 46 | 84% |
| 합계 | 524 | 146 | 28% |

## 5. 논의

본 논문은 언어 지식을 기반으로 한 프랑스어 발음열 생성 시스템에 대하여 기술하는 것을 목적으로 프랑스어 음성합성이나 음성인식 시스템을 개발하고자 하는 경우, 사용할 수 있는 프랑스어에 대한 기본적인 지식을 제공하였다. 나아가, 이를 기반으로 한 지식 기반 발음열 생성 시스템에 대하여 기술하였다. 프랑스어는 철자를 기준으로 할 때 단어(한국어의 경우 어절) 내부 그리고 단어와 단어 사이에서 많은 음운현상이 관찰되는 언어로 형태소 분석 결과가 발음열 생성에 지대한 영향을 미치게 된다. 위와 같은 이유로 프랑스어의 발음열 생성 시스템은 형태소 분석과 품사 태깅, 그리고 자소-음소 변환과 음운현상을 반영한 음소-음소 변환 모듈로 구성하였다.

제안한 프랑스어 발음생성 시스템의 평가를 위해 품사 태깅 과정과 발음 생성 과정을 각각 평가하였다. 품사 태깅 정확도 평가는 본 시스템이 적용될 음성합성 시스템의 서비스 도메인인 뉴스 영역에서 다양한 주제를 포함하는 1,000문장을 선정하여 평가하였다. 그 결과, 단어 오류율 10.70%, 문장 오류율 79.70%를 보였다. 위에서 언급한 바와 같이, 프랑스어의 발음열 생성 시스템은 품사 태깅 과정과 텍스트 정규화(text normalization)까지 모두 자소-음소 변환 시스템에 포함되어 있고(Yvon *et al.*, 1998; Béchet, 2001; de Mareüil *et al.*, 2005; Rao *et al.*, 2015), 이러한 연구들에서는 품사 태깅 정확도를 따로 평가하지 않고 있지만, 궁극적으로 자소-음소 변환 결과를 평가하고 있어서 품사태깅 결과에 대한 직접적인 비교는 가능하지 않다.

다만, 본 연구에서 목표 서비스 영역을 뉴스 영역으로 제한하였고, 평균 문장 길이가 25개 정도 되는 긴 문장에 대한 평가가 진행되어 일반적인 평가 셋보다 난이도가 높은 것으로 판단된다. 실제 품사 태깅이 발음에 직접적인 영향을 주는 경우와 동철이음어에 해당하는 경우, 실제 발음에 많은 영향을 미친다고 볼 수 없으므로, 평가 결과 얻은 단어 오류율과 문장 오류율은 다음 단계인 발음 변환 결과와 통합하여 고려해야 할 것이다.

생성된 발음열을 평가를 위해서는 세 가지 방식의 평가를 진행하였다. 먼저, 원어민 언어 전문가 두 명이 동일한 문장 310개에 대하여 발음열 정답지를 만들었고 시스템에 의해 도출된 결과물을 정답지와 비교한 결과, 전체 토큰 수 대비 원어민 정답의 토큰 비율(precision=correct/all)은 평균 0.89였다. 다음으로는, 시스템으로부터 생성된 어휘 130,883개 표제어의 발음열을 기성 발음 사전과 비교한 결과, 151개의 표제어 발음이 오류를 보여 결과적으로 단어 기준 2.7% 오류율이라는 높은 성능을 확인할 수 있었다. 이러한 결과는 현재까지 최근 가장 좋은 성능으로 보고되고 있는 음소 오류률 13.2%와 문장 오류율 57.4%(Béchet, 2001)에 비하여 높은 정확도라고 할 수 있다. 또한, 본 연구에서는 기존 연구들과는 달리 프랑스어의 음소-자소 변환 문제 가운데 성능에 결정적인 영향을 끼치는 연음 환경에 대하여 따로 평가 문장을 만들어 평가를 진행하였다. 평가 결과, 28%의 오류율을 보였는데, 이는 연음 환경만 고려한 문장들을 대상으로 한 평가로서 전체 평가 문장에서 이러한 문장이 차지하는 비율에 따라 그 영향이 다르게 나타날 수 있으며, 그 영향에 대해서는 이후 연구로 미루기로 한다.

앞서 언급한 Yvon *et al.*(1998)은 프랑스어 음성합성을 위한 음소-자소 변환 시스템의 평가 방법에 대한 연구로서, 평가 코

퍼스 구축을 포함한 평가 방법을 제안하고, 당시 지식 기반 시스템으로 프랑스, 스위스, 벨기에, 캐나다 등에서 개발된 8개의 시스템에 대한 평가 결과를 보고하고 있다. 8개 시스템의 경우, 음소 정확도는 97%이고, 문장 정확도는 20%에서 90%의 성능을 보이고 있다. 이 연구는 기존에 개발이 완료된 여러 다른 시스템의 성능을 비교하여 평가하는 것을 목표로 하는 것으로, 프랑스어에 대한 기본 시스템을 목표하는 본 연구와는 그 범위와 목표가 다르다고 할 수 있다. 본 연구는 어느 정도의 기본적인 성능을 확보하기 위한 방법으로, 이후 성능 개선과 함께 Yvon *et al.* (1998)이 제안한 방법에 따른 평가도 이루어져야 할 것으로 본다.

## 6. 결론

본 논문은 프랑스어 발음열 생성 시스템 개발을 위하여 필요한 프랑스어 철자 체계와 음소 체계간의 연관성을 포함한 언어 지식을 제공하고, 이를 토대로 한 프랑스어 발음열 생성 시스템을 개발하는 방법에 대하여 기술하였다. 비록 기본적인 시스템이긴 하나 실질적인 서비스에 적용할 수 있는 테스트 셋을 구성하였고, 품사 태깅과 발음열 생성에 대한 성능 평가를 수행하였다. 평가 결과, 초기 서비스에 적용할 수 있는 정도의 성능은 확보된 것으로는 보이나, 이후 서비스 도메인 확장에 따라 계속 개선할 필요가 있을 것으로 예상한다.

프랑스어의 경우, 현재까지 가장 우수한 성능을 보이는 것으로 보고되는 프랑스어 발음열 생성 시스템에 대한 자료가 그다지 많지 않은 연구 및 개발 환경을 감안할 때, 시스템 개발을 위하여 실질적으로 필요한 자료들과 시스템 구성 및 평가 방법 등을 제안한 본 연구가 이후 프랑스어 음성합성이나 음성인식 시스템을 개발하는데 기여할 수 있기를 기대하는 바이다.

## 참고문헌


Allen, J., Hunnicutt, M., Klatt, D., Armstrong, R., & Pisoni, D. (1987). *From text to speech: The MITalk system.* NY: Cambridge University Press.

Arik, S., Chrzanowski, M., Coates, A., Diamos, G., Gibiansky, A., Kang, Y., Li, X., Miller, J., Ng, A., Raiman, J., Sengupta, S., & Shoeybi, M. (2017). Deep voice: Real-time neural text-to-speech. *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)* (pp. 1234-1252).

Béchet, F. (2001). LIA PHON: Un systeme complet de phonétisation de textes. *Traitement Automatique Des Langues*, 42(1), 47-67.

Black, A., Lenzo, K., & Pagel, V. (1998). Issues in building general letter to sound rules. *3rd ESCA Workshop on Speech Synthesis* (pp. 77-80).

Byrd, R., & Tzoukermann, E. (1988). Adapting an English morphological analyzer for French. *Proceedings of the 26th Annual Meeting on Association for Computational Linguistics* (pp. 1-6). Association for Computational Linguistics.

de Mareüil, P., d'Alessandro, C., Bailly, G., Béchet, F., Garcia, M., Morel, M., Prudon, R., & Véronis, J. (2005). Evaluating the pron unciation of proper names by four French grapheme-to-phoneme converters. *Proceedings of the Interspeech 2005* (pp. 1521-1524). Interspeech.

Gruaz, C., Jacquemin, C., & Tzoukerman, E. (1996). Une approche à deux niveaux de la morphologie dérivationnelle du français. *Actes du séminaire Lexique. Représentations et Outils pour les bases lexicales. Morphologie Robuste*, 107-114.

Hahn, S., Vozila, P., & Bisani, M. (2012). Comparison of grapheme-to-phoneme methods on large pronunciation dictionaries and LVCSR tasks. In *13th Annual Conference of the International Speech Communication Association.*

Jiampojamarn, S., & Kondrak, G. (2010). Phoneme alignment: An exploration. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 780-788). Association for Computational Linguistics.

Larreur, D., & Sorin, C. (1991). Quality evaluation of French text-to-speech synthesis within a task the importance of the mute "e". *Proceedings of the ESCA Workshop on Speech Synthesis*. Lannion. 25-28 September, 1990.

Lecorvé, G., & Lolive, D. (2015). Adaptive statistical utterance phonetization for French. *Proceedings of the Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on* (pp. 4864-4868). IEEE.

Marchand, Y., & Damper, R. (2000). A multistrategy approach to improving pronunciation by analogy. *Computational Linguistics*, 26(2), 195-219.

Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. Retrieved from http://arxiv.org/abs/1609.03499 [Computing Research Repository] on September 19, 2016.

Pérennou, G., & De Calmes, M. (2000). MHATLex: Lexical resources for modelling the french pronunciation. *Proceedings of the LREC 2000*.

Rao, K., Peng, F., Sak, H., & Beaufays, F. (2015). Grapheme-to-pho neme conversion using long short-term memory recurrent neural networks. *Proceedings of the Acoustics, Speech and Signal Proce ssing (ICASSP), 2015 IEEE International Conference on* (pp. 42 25-4229). IEEE.

Shen, J., Pang, R., Weiss, R., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerry-Ryan, R., Saurous, R., Agiomyrgiannakis, Y., & Wu, Y. (2017). Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions. arXiv preprint arXiv:1712.05884. February 16, 2018.

Sotelo, J., Mehri, S., Kumar, K., Santos, J., Kastner, K., Courville, A.,

& Bengio, Y. (2017). Char2Wav: End-to-End Speech Synthesis. *Proceedings of the 5th International Conference on Learning Representations (ICLR 2017) Workshop*. Retrieved from https://openreview.net/forum?id=B1VWyySKx on 18 February, 2017.

Taylor, P. (2005). Hidden Markov models for grapheme to phoneme conversion. *Proceedings of the 9th European Conference on Speech Communication and Technology*.

Taylor, P. (2009). *Text-to-speech synthesis*. NY: Cambridge University Press.

Tokuda, K., Nankaku, U., Toda, T., Zen, H., Yamagishi, J., & Oura, K. (2013). Speech synthesis based on Hidden Markov models. *Proceedings of IEEE* (pp. 1234-1252).

Van Den Bosch, A., & Canisius, S. (2006). Improved morpho-phonological sequence processing with constraint satisfaction inference. *Proceedings of the 8th Meeting of the ACL Special Interest Group on Computational Phonology and Morphology* (pp. 41-49). Association for Computational Linguistics.

Wang, Y., Skerry-Ryan, R., Stanton, D., Wu, Y., Weiss, R., Jaitly, N., Yang, Z., Xiao, Y., Chen, Z., Bengio, S., Le, Q., Agiomyrgiannakis, Y., Clark, R., & Saurous, R. (2017). Tacotron: Towards end-to-end speech synthesis. Retrieved from http://arxiv.org/abs/1703.10135 [Computing Research Repository] on April 6, 2017.

Yoon, K., & Brew, C. (2006). A linguistically motivated approach to grapheme-to-phoneme conversion for Korean. *Computer Speech & Language*, 20(4), 357-381.

Yvon, F., De Mareüil, P., d'Alessandro, C., Auberge, V., Aubergé, V., Bagein, M., Bailly, G., Béchet, F., Foukia, S., Goldman, J., Keller, E., O'Shaughnessy, D., Pagel, V., Sannier, F., Ve´ronis, J., & Zellner, B. (1998). Objective evaluation of grapheme to phoneme conversion for text-to-speech synthesis in French. *Computer Speech & Language*, 12(4), 393-410.

Zen, H., & Sak, H. (2015). Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis. *Proceedings of the* Acoustics, *Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on* (pp. 4470-4474). IEEE.

Zen, H., Senior, A., & Schuster, M. (2013). Statistical parametric speech synthesis using deep neural networks. *Proceedings of the Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 7962-7966). IEEE.

• **김선희 (Kim, Sunhee)**
㈜네이버
경기도 성남시 분당구 불정로 6 NAVER 그린팩토리
Tel: 031-784-3307, Fax: 031-784-1000
Email: kim.sunhee@navercorp.com
관심분야: 음성합성, 음성인식, 자동발음평가, 음성학