The Merger of Korean Mid Front Vowels and the Notation of English Loanwords: An Analysis Through the Social Media*

Sung Hyun, NAM (Chung-Ang University)

1. Introduction

This study aims to observe a patterned confusing among Korean native speakers when they transcribe English words that contain mid front vowels /æ/ and /e/, and to provide an explanation for the phenomenon. As previous studies have argued, mid front vowels conventionally differentiated as 'ㅐ'/ɛ/ and 'ㅔ'/e/ are merged in the vowel system of modern Korean (Kim & Han 2000; 곽충구 2003; 신하영 2013; Shin, Kiaer, & Cha 2013: 99-100 among others). As in <Table 1> the two phones have similar mean values of the first and second formants, and their distribution generally overlaps, indicating the fact that they are not differentiated in Korean speakers' production. Also in perception, Korean speakers cannot distinguish the difference between two sounds synthesized with different formant values (Eychenne & Jang 2015) or between two mid front vowels /e/ and /æ/ produced by a native English speaker (Ingram & Park 1997; Tsukada et al. 2005).

	F1 (Standard distribution)			F2 (Standard distribution)		
	이재강(1998)	Yang (1996)		이재강(1998)	Yang (1996)	
		Male	Female		Male	Female
'ㅐ'/ɛ/	504 (73)	591 (75)	677 (108)	2225 (359)	1849 (106)	2285 (169)
'ㅔ'/e/	485 (64)	490 (105)	650 (113)	2229 (358)	1968 (150)	2377 (77)

<Table 1> F1 and F2 values of Korean mid front vowels

Notwithstanding the fact that these two sounds are not distinguished in production and perception of native speech, the National Institute of the Korean Language suggests a normative guideline where English /æ/ is to be transcribed as 'ㅐ' and /e/ as 'ㅔ,' and the public generally finds no difficulty in following it. This situation seems quite interesting, in that the input form of a foreign word in loanword phonology is assumed to be composed of differentiated sounds units among the source language's surface representation (Silverman 1992; Yip 1993; Kenstowicz 2005). Therefore, we need an explanation for this phenomenon in which Korean speakers transcribe differently two sounds that are not distinguished in their native language. Moreover, for some words like "내비게이션(←navigation)" and "페널티(←penalty)," people more often use the opposite letter to the standard orthography, i.e., "네비게이션" and "패널티," respectively. In those odd forms letter 'ㅔ' is used to transcribe English /æ/, while 'ㅐ' is used for /e/, which shows the opposite pattern of what the norm regulates. From this backgrounds, this study shows that this coexistence of the following of the norm and the showing of the opposite pattern does exist in Twitter, a social media. Then it tries to provide an explanation for it.

2. Analysis of the orthographic pattern through Twitter

This study investigated all Korean tweets between the period of October 20 to 29, 2015. Among them, only the occurrences of English loanwords that contain /æ/ or /e/ were of interest. As it is nearly impossible to directly access the raw data of all tweets, this study utilized R (R Core Team 2015), a statistical programming language, and "TwitteR"

(Gentry 2015), an R package that interacts with Twitter.

For the scope of loanwords, this study followed 김순임(2007), which surveyed 1,948 Korean native speakers and listed up 100 frequently used English loanwords. Among the 100 words, 34 had mid front vowels. <Table 2> shows the scope of the analysis.

#	Word	#	Word	#	Word	#	Word
1	갭	10	매뉴얼	19	아웃렛	27	콘텐츠
2	내비게이션	11	멘트	20	옴부즈맨	28	테마
3	네거티브	12	백업	21	이벤트	29	페널티
4	네트워크	13	벤처기업	22	인센티브	30	프레젠테이션
5	네티즌	14	벤치마킹	23	인터넷뱅킹	31	프로그램
6	네티켓	15	샘플	24	캠페인	32	프로젝트
7	데이터	16	셀카	25	캡처	33	핸디캡
8	데이터베이스	17	스태프	26	콘셉트	34	홈페이지
9	마케팅	18	스팸				

<Table 2> The scope of the analysis

The actual R script that collects the tweets that contain each *keyword* is presented as <Figure 1> below. I first counted the occurrences of the correct transcription and odd transcription*, and then the proportion of odd among all occurrences (that is, $(odd) \div (correct + odd)$)

```
library(twitteR)
keyword <- "(키워드 입력)"
keyword <-enc2utf8(keyword)
    # 인코딩 상의 문제로 인해 enc 인코딩을 utf8로 바꾸어 진행
result<-searchTwitter(keyword, n=1000)
    # 해당 키워드가 들어가 있는 트윗을 1000개 retrieve
result.df<- twListToDF(result)
    # 후처리를 편하게 하기 위해 list 타입으로 된 트윗 검색 결과를 data.frame으로 전환
write.csv(result.df,"twitter.csv")
    # data.frame으로 전환된 트윗 검색 결과를, 엑셀 등에서 읽을 수 있는 .csv 형식으로 출력
```

<Figure 1> An R script that returns the frequency of *keyword* among all tweets

---

* For instance, the second word in <Table 2> which is '내비게이션(←n[æ]vigation)', I first put '내비게이션'(/æ/ transcribed as 'ㅐ') as vector *keyword* to count the frequency (Correct) and then again put '네비게이션'(/æ/ transcribed as 'ㅔ') as vector *keyword* to calculate the frequency (Odd).

<Table 3> presents the proportion of the norm is applied in the opposite way, i.e., /æ/ as '┤' and /e/ as '┤.' Shaded cells in the table represent the cases where /æ/ is transcribed as '┤,' which is against the norm, and hatched cells (#14 and #28) are loanwords that contain both /æ/ and /e/. For "내비게이션(←navigation)" and "페널티(←penalty)," confused notation exceeded 50%. The three top words in <Table 3> are exceptional in that they are more often notated with the opposite letters while /æ/ and /e/ in other loanwords are correctly transcribed as the norm.

	Correct	Odd	Confusion rate (%)		Correct	Odd	Confusion rate (%)
1	페널티	패널티	79.38	19	데이터	대이터	0.13
2	내비게이션	네비게이션	69.68	20	프로젝트	프로잭트	0.11
3	매뉴얼	메뉴얼	27.23	21	콘셉트	콘샙트	0.11
4	벤치마킹	밴치마킹	3.47	22	프로그램	프로그램	0.11
5	인센티브	인샌티브	3.07	23	네거티브	내거티브	0.10
6	스태프	스테프	2.69	24	홈페이지	홈패이지	0.02
7	스팸	스펨	1.95	25	백업	벡업	0.02
8	샘플	셈플	0.83	26	셀카	샐카	0.02
9	멘트	맨트	0.70	27	네티즌	내티즌	0.01
10	핸디캡	헨디캡	0.69	28	캠페인	켐패인	0.01
11	캠페인	캠패인	0.62	29	이벤트	이밴트	0.01
12	테마	태마	0.37	30	마케팅	마캐팅	0.01
13	아웃렛	아웃랫	0.26	31	갭	겝	0.00
14	인터넷뱅킹	인터냇뱅킹	0.25	32	네티켓	내티켓	0
15	캡처	켑처	0.22	33	데이터베이스	대이터베이스	0
16	콘텐츠	콘탠츠	0.19	34	벤처기업	밴처기업	0
17	캠페인	켐페인	0.18	35	옴부즈맨	옴부즈멘	0
18	네트워크	내트워크	0.17	36	프레젠테이션	프래젠테이션	0

<Table 3> The result of Twitter analysis

3. Discussion

As mentioned earlier, English /æ/ and /e/ are not expected to be differentiated in the course of loanword acceptance because mid front vowels are not distinguished in Korean phonemic system. However, /æ/ and /e/ are differently transcribed in the orthography of

English loanwords. This situation indicates that another factor external to phonology plays more role in acceptance: the orthography of the source language. 이진성 (2000) points out that /p, t, k/ are transcribed in Korean as ㅍ, ㅌ, ㅋ, rather than ㅃ, ㄸ, ㄲ which are phonetically more similar to the original language. The transcription of English mid front vowels can be another example where the orthography of the source language impacts the transcription of loanwords.

However, we should take note that in a few exceptions like "페널티(←penalty)," "내비게이션(←navigation)" and "매뉴얼(←manual)," the odd form is more preferred. In accepting those words, loanword-phonological rules are primarily applied before the English orthography. That is to say, Korean speakers may have encountered those words as p[E]nalty, n[E]vigation and m[E]nual, respectively ([E] here indicates a mid front vowel not specifically /æ/ or /e/). After that, [E] is realized as 'ㅐ' or 'ㅔ' in Korean orthography.

Then, which one to choose for this [E]? The mechanism that makes Korean speakers select 'ㅔ' or 'ㅐ' is from the phonotactic possibility and lexical strata. <Table 3> show us another interesting fact. On that table, /æ/ is more often mistakenly transcribed as 'ㅔ' than /e/ is transcribed as 'ㅐ.' Considering that Korean lexicon has three strata of native, sino-Korean and foreign words, people have learned that 'ㅔ' is the more frequent mid front vowel among foreign sub-lexicon and when they are forced to choose between 'ㅔ' and 'ㅐ' they choose 'ㅔ.'

4. Conclusion

This study has shown the pattern of confused transcription of mid front vowels in English loanwords, and tried to present underlying factors beneath the phenomenon. Accepting English words that contain /æ/ or /e/, the source orthography plays more role than loanword-phonological processes. However, some cases have shown that phonotactic possibility and the stratified lexicon play a role in confused notation.

The data analyzed in this study is the most suitable for this kind of research as

people using Twitter feel freer from the normative pressures, and they tend to be more abide by their linguistic intuition. Moreover, this study is qualitatively different from other recent studies that deal with Twitter data, including 손예희·김지연 (2010) and 이정복 (2011, 2013, 2014), in that this study broaden the sphere of analysis from sociolinguistic topics to (loanword) phonology. Also, this study successfully analyzed more than 100 thousand tweets to derive significant implications, while those previous studies only looked at a very few number of tweets.

References

곽충구. 2003. 현대국어의 모음체계와 그 변화의 방향. *국어학* 41. 59-91.

김순임. 2007. *외래어 인지도·이해도·사용도 및 태도 조사*. 국립국어원.

이재강. 1998. 한국어 모음에 대한 한국인과 일본인의 대조 연구. *언어학* 22. 347-369.

이정복. 2011. 트위터 누리꾼들의 호칭어 사용에 대한 사회언어학적 접근. *어문학* 114. 143-174.

_____. 2013. 사회적 소통망(SNS)의 지역 차별 표현. *어문학* 120. 55-83.

_____. 2014. 통신 언어 의성의태어 사용과 성별 차이. *우리말글* 62. 45-74.

이진성. 2000. 외래어 표기와 발음의 실태. *사회언어학* 8(2). 223-245.

손예희 & 김지연. 2010. 소셜 미디어의 소통 구조에 대한 국어교육적 고찰. *국어교육* 133. 207-231.

신하영. 2013. 20세기 초 서울 방언의 고모음화와 모음 합류에 대한 고찰. *어문학* 121. 1-25.

Eychenne, J., & Jang, T. Y. 2015. On the Merger of Korean Mid Front Vowels: Phonetic and Phonological Evidence. *Journal of the Korean Society of Speech Sciences Vol, 7*(2).

Gentry, Jeff. 2015. *twitteR: R Based Twitter Client*. R package version 1.1.9. http://CRAN.R-project.org/package=twitteR

Ingram, J. C., & Park, S. G. 1997. Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of phonetics* 25(3). 343-370.

Kenstowicz, M. 2005. The phonetics and phonology of Korean loanword adaptation. In *Proceedings of the first European conference on Korean linguistics* (Vol. 1, pp. 17-32).

Kim, H., & Han, J. I. 2000. An Optimality-Theoretic Account of the Asymmetry in Korean Vowels. *언어* 25(2). 255-269.

R Core Team. 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Shin, J., Kiaer, J. & Cha, J. 2013. *The Sounds of Korean*. Cambridge: Cambridge University Press.

Silverman, D. 1992. Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology* 9(02). 289-328.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. 2005. A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics* 33(3). 263-290.

Yang, B. 1996. A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics* 24(2). 245-261.

Yip, M. 1993. Cantonese loanword phonology and Optimality Theory. *Journal of East Asian Linguistics* 2(3). 261-291.