# Pronunciation Modeling of Loanwords for Korean ASR Using Phonological Knowledge and Syllable-based Segmentation

Hyuksu Ryu[1], Minsu Na[2], and Minhwa Chung[1,2]
[1] Department of Linguistics, Seoul National University, Seoul, REPUBLIC OF KOREA
E-mail: {oster01, mchung}@snu.ac.kr
[2] Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, REPUBLIC OF KOREA
E-mail: dix39@snu.ac.kr

***Abstract*—This paper aims to improve the performance of automatic pronunciation generation of foreign loanwords in Korean by using phonological knowledge and syllable-based segmentation. The loanword text corpus used for our experiment consists of 16.6K words extracted from the frequently used words in set-top box, music, and POI domains. At first, pronunciations of loanwords in Korean are obtained by manual transcriptions, which are used as target pronunciations. A syllable-based segmentation method considering phonological differences is proposed for loanword pronunciation modeling. Performance of the baseline and the proposed method are measured using PER/WER and F-score at various context spans. The result shows that the proposed method outperforms the baseline. We also observe performance decrease when training and test sets come from different domains, which implies that loanword pronunciations are influenced by data domains. It is noteworthy that pronunciation modeling for loanwords in Korean is enhanced by reflecting phonological knowledge. The loanword pronunciation modeling in Korean proposed in this paper can be used for (1) ASR of application interface such as navigation and set-top box and (2) computer-assisted pronunciation training for Korean learners of English.**

## I. Introduction

Loanwords are borrowed words from L2 (foreign language), which are incorporated into L1 (native language) phonetic system [1] and are made to conform with phonological rules of L1 [2]. In Korean, it is reported that loanwords appear in 63% of the titles of TV programs [3], while only 4.7% of entry words of the Standard Korean Dictionary are loanwords [1][4]. Recently, automatic speech recognition (ASR) technology is commonly used for application interfaces of navigation and TV set-top box. Foreign loanwords are frequently used as point-of-interest (POI) entries for navigation interface and as parts of titles of TV programs for TV set-top box interface. Furthermore, various foreign proper nouns such as the names of singers and films are frequently used in TV set-top box applications. Therefore, to improve ASR performance for such application interfaces, loanword pronunciation should be modeled and reflected in a pronunciation dictionary of ASR systems.

Loanwords used in POI or multimedia are usually neologisms. Thus, it is difficult to manually establish loanword pronunciation dictionary, since it consumes massive amount of time, manpower, and cost to continuously update neologisms to the dictionary. Therefore, grapheme-to-phoneme (G2P) converter is necessary for pronunciation modeling of loanwords in Korean.

Loanwords spoken in Korean have separate phonological/phonetic system from Korean native vocabulary [1][5], although the loanwords are part of Korean. For instance, as can be seen in Fig. 1, an English word "SECRET" has three loanword pronunciation variants in Korean. Vowel epenthesis of /ɯ/ appearing in a consonant cluster or word-final consonant is obligatory according to phonological rules in native Korean [6]. However, tensification of alveolar fricative /s/ as /s͈/ occurs only in loanwords of Korean, not in native Korean vocabulary [7][8]. Therefore, besides a normal Korean G2P [9], an additional G2P for loanwords in Korean is required in order to deal with such pronunciation variations.

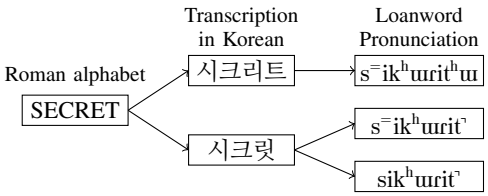


Fig. 1. Pronunciation variations of loanwords in Korean

There are several previous studies regarding pronunciation modeling of foreign loanwords [10][11]. Reference [10] dealt with establishing pronunciation dictionary of loanwords in Chinese Mandarin for ASR. They extended the pronunciation dictionary by mapping phonemes using differences in phonemic system and syllabic structure between English and Mandarin. By the results of the experiment, the pronunciation dictionary that considers loanwords improved the performance of ASR. Reference [11] studied loanword pronunciation modeling in Sepedi, which is one of the public languages in South Africa. They extended pronunciation of vocabularies which come from English or other public languages in South Africa. They predicted pronunciation variants by using (1) foreign-

to-Sepedi phoneme mapping and (2) Letter-to-sound rule of Sepedi regardless of source languages. The result showed that letter-to-sound rules outperformed phoneme mapping. These previous studies presented that the performance of ASR was enhanced by considering pronunciation modeling of loanwords.

However, there have been no previous studies concerning pronunciation modeling of loanwords in Korean. Only a few quantitative studies dealt with pronunciation variations of loanwords [1][12]. Pronunciation modeling of loanwords in Korean is necessary for improving ASR performance for navigation and set-top box applications.

The goal of this paper is to improve pronunciation modeling of loanwords in Korean by using phonological knowledge and syllable-based segmentation. The remaining part of this paper is organized as follows. Section II describes phonological and phonetic characteristics of loanwords in Korean. In Section III, a segmentation scheme and experimental setup are proposed. Experimental results using the proposed method are presented in Section IV, which is followed by conclusion in Section V.

## II. Loanwords in Korean

### A. Phonological difference

In this section, phonological differences between English and Korean are described, considering that many loanwords in Korean originate from English [13]. Description regarding phonological differences is necessary since many characteristics of loanword pronunciation come from the fact that pronunciations of the source language (English) are not fully covered by that of the target language (Korean). Thus, pronunciations of loanwords are realized in different manner depending on phonological system of the target language.

There are two kinds of phonological differences between English and Korean: one is difference of phonemic system and the other is difference of syllabic structure. First of all, regarding phonemic systems of English and Korean, there are some English phonemes which do not exist in Korean. These phonemes are mapped by regulations concerning Hangeul transcription of loanwords [14]. The phonemes and the corresponding mapping in Korean are listed below.

TABLE I
List of phonemes which exist only in English

| Category | English | mapping in Korean |
|---|---|---|
| Fricative | f, v, θ, ð, ʃ, ʒ | p, b, s, d, si, ʤi |
| Affricate | ʦ, ʣ | ʧi, zi |
| Vowel | ɪ, ʊ, ɔ, ə | i, u, o, ʌ |

Secondly, syllabic structure in Korean and English shows difference in dealing with consonant clusters and codas. Consonant clusters are not allowed in Korean, while at most three consonants at onset position and four at coda position are allowed in English, e.g. STRENGTHS /stɹɛŋkθs/. In addition, only seven consonants of /kˀ, n, tˀ, l, m, pˀ, ŋ/ can appear at coda position in Korean. On the contrary, English allows all consonants except /h/ at coda. Korean speakers tend to insert a vowel /ɯ/ as a strategy to adapt to the differences in consonant clusters and codas [15].

### B. Pronunciation of loanwords in Korean

As mentioned previously, loanword phonology in Korean have different characteristics from Korean native vocabulary phonology [5]. This leads to the unique pronunciation rules that are realized only in loanwords in Korean [1][7][12], such as onset tensification, fricative /s/ tensification, affrication, and vowel variation, etc. Each rule and the corresponding examples are shown in Table II.

The pronunciation rules listed above are difficult to be used to standardize pronunciation for loanwords because of the following reasons: (1) there is no standard pronunciation of loanwords in Korean [16] and (2) they are optional phonological rules, not obligatory [1]. Therefore, in this study, pronunciation modeling is performed by using data-driven approach, not rule-based.

TABLE II
Loanword pronunciation rules and examples

| Pronunciation rules | Examples | |
|---|---|---|
| Onset tensification | GAME /k͡=ɛim/ | BOX /p͡=ak's͡=ɯ/ |
| Fricative /s/ tensification | ACE /ɛis͡=ɯ/ | SIGN /s͡=ain/ |
| Affrication | BASIC /bˌɛiʤ͡ikˀ/ | BEARS /bˌɛʌʤ͡ɯ/ |
| Vowel variation | LIGHTER /laitʰa͡/ | COLOR /kʰa͡lla͡/ |

## III. Method

### A. Corpus and Transcription

We use the loanword corpus provided by SK Telecom. It is a text corpus written in Roman alphabet. The corpus consists of 16.6K words extracted from the frequently used words in set-top box, music, and POI domains. Details of the loanword corpus are described in Table III.

TABLE III
Details of the loanword corpus

| Domain | # of words | Proportion(%) |
|---|---|---|
| Set-top box | 2,991 | 18.00 |
| Music | 2,217 | 13.35 |
| POI | 11,405 | 68.65 |
| Total | 16,613 | 100.00 |

At first, pronunciation of loanwords in Korean is obtained by manual transcription. Five groups participate in loanword transcription. Each group is composed of six annotators: two graduate and two undergraduate students majoring in linguistics, and two non-linguistic majors. Non-linguistic majors are included to reflect the pronunciation difference according to linguistic knowledge. Each group annotates approximately 3K to 3.5K words regardless of domains.

The groups annotate the loanwords based on their own pronunciation. They are asked to transcribe as many pronunciations as possible in the environment that they speak in Korean, not in source language such as English.

To prevent from allowing too many pronunciation variants, the results transcribed by more than three annotators are chosen as target pronunciations. As a result, 19.4K pronunciation variants are selected. Thus, each loanword has 1.17 variants per word in average.

### *B. Syllable-based segmentation*

Loanwords in Korean may have multiple pronunciation variants, although the grapheme forms are the same. For example, a loanword "EMMA" can be realized as /ɛma/ or /ɛmma/. Unfortunately, it is difficult to predict such completely different pronunciations using only the grapheme sequences "EMMA".

Considering phonological differences such as syllable structures described in Section II, the loanword presented above that has two different pronunciation variants can be segmented in different ways. As shown in Fig. 2, the two pronunciation variants of the word "EMMA" have different syllable structures in the first syllable. The first syllable /ɛ/ of the pronunciation sequence /ɛma/ has only a nucleus. On the contrary, in the second pronunciation /ɛmma/, the first syllable /ɛm/ has a nucleus and a coda. In accordance with this difference, loanwords "EMMA" is segmented differently in grapheme level as well. In the case of "EMMA" /ɛma/, "MM" of grapheme "EMMA" belongs to the second segment, since there is no coda in the first grapheme segment. On the other hand, "MM" is split into each grapheme segment when the pronunciation /ɛmma/ is realized.

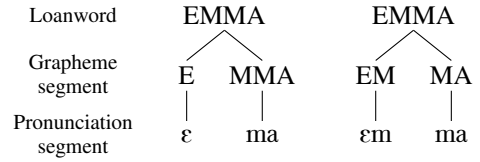


Fig. 2. Segmentation example

As described above, an identical grapheme sequence may have different grapheme segmentations for the corresponding pronunciations according to the phonological syllable structures. If segmentation information of loanwords at grapheme level is provided additionally besides loanwords and their pronunciations, loanword pronunciations can be predicted more effectively. Therefore, considering phonological differences, we propose a syllable-based segmentation scheme for modeling loanword pronunciations in Korean. Segmentation rules we propose are as follows. Basically, loanwords are segmented based on syllables in graphemes. However, since there is difference of syllable structure between Korean and source languages, vowel epenthesis in consonant clusters and word-final consonants should be considered for segmentation. Thus, inserted vowels are also segmented. For example, as shown in Fig. 3, while "SECRET" is composed of two syllables in English pronunciation, loanword pronunciations in Korean are realized as three or four syllables according to vowel insertion.

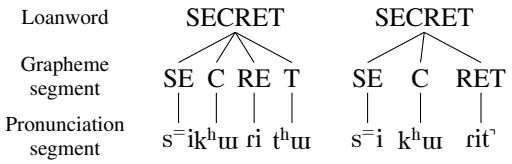


Fig. 3. Segmentation considering vowel epenthesis

### *C. Experimental setup*

We adopt the Sequitur G2P toolkit [17] for our loanword pronunciation modeling task. The Sequitur toolkit relies on a joint sequence n-gram model, which computes its probability considering sequences of graphemes and the corresponding phoneme sequence. The Sequitur G2P toolkit is one of the state-of-art approach for grapheme-to-phoneme conversion [18–20].

Pronunciation generation experiments for loanwords are performed to compare the proposed method with the baseline. For the baseline, following the grapheme-phoneme joint multigram in [17], G2P model is trained using pairs of loanword and pronunciation. The fundamental idea of the method is to generate the most likely pronunciation considering a given grapheme sequence. The method is formalized as (1) [17].

$$\varphi(g) = \arg\max_{\varphi \in \Phi} p(g, \varphi) \tag{1}$$

In (1), $g$ and $\varphi$ denote a sequence of grapheme and pronunciation, respectively. The set of pronunciation is denoted as $\Phi$.

For the proposed method, we use two dictionaries for training: one is a segment dictionary and the other is a pronunciation dictionary. As shown in Table IV, the segment dictionary contains loanword and the corresponding syllable-based grapheme segmentation, while the pronunciation dictionary contains loanword and the corresponding pronunciation. G2P models for the proposed method are trained by mapping syllable-based segments and pronunciation in two dictionaries.

TABLE IV
SEGMENT AND PRONUNCIATION DICTIONARIES

| Dictionary | Loanword | Segment/Pronunciation |
|---|---|---|
| Segment dictionary | SECRET | SE — C — RE — T |
| Pronunciation dictionary | | s= i kʰ ɯ ɾ i tʰ ɯ |

The training process is performed in incremental way. The initial training creates a very simple model, and each training depends on the previously trained model. In this experiment, training is performed seven times. In addition, 5% of training data is used as development set for optimizing parameters of the models. When each iteration is completed, test data is applied to each trained model.

Considering the purpose of ASR, G2P should generate multiple pronunciation variants [18]. For this reason, we extract a number of pronunciation variants until the sum of the probabilities of the pronunciation variants exceed 0.5.

However, if the probability of each variant is too small, too many garbage variants are generated to exceed the criteria. Thus, we restrict the total number of pronunciations to ten variants at most.

The Sequitur G2P has an option to adjust context spans to calculate joint sequence model. We vary context spans from tri-gram, i.e. one previous, one current, and one following context, to nine-gram, i.e. four previous, one current, and four following contexts.

We use two kinds of measures to assess the quality of the G2P models: (1) phone error rate (PER)/word error rate (WER), and (2) precision/recall/F-score. PER and WER measures are usually used for evaluation of pronunciation modeling [19–21]. On the other hand, the precision, recall and F-score measures are also good indicators for quality of pronunciation variations [18][22]. Especially in ASR context, not only the ratio of generated pronunciations which are correct pronunciation (precision), but the ratio of correct pronunciation variants which are actually predicted (recall) is also important. F-score is the harmonic mean of precision and recall.

TABLE V
STATISTICS OF THE DATA FOR THE EXPERIMENT 1

| | # of words | Proportion(%) |
|---|---|---|
| Training | 13,300 | 80.06 |
| Development | 700 | 4.21 |
| Test | 2,613 | 15.73 |
| Total | 16,613 | 100.00 |

Using experimental setup above, we perform two experiments: (1) performance comparison between the baseline and the proposed method, and (2) comparison among different domains of the data. In the experiment 1, we divide training, development, and test data regardless of data domains. The statistics of the data in our experiments is shown in Table V.

The aim of the experiment 2 is to observe the influence of loanword domains on the performance of pronunciation modeling. We train the pronunciation model using POI data only, since the data from other domains are too small to be used for training as shown in Table III. Three kinds of test data are prepared: set-top box, music, and POI. The detailed statistics for experiment 2 is as follows in Table VI.

TABLE VI
STATISTICS OF THE DATA FOR THE EXPERIMENT 2

| | Domain | # of words | Proportion(%) |
|---|---|---|---|
| Training | POI | 9,500 | 66.90 |
| Development | POI | 500 | 3.52 |
| Test | Set-top | 1,400 | 9.86 |
| | Music | 1,400 | 9.86 |
| | POI | 1,400 | 9.86 |
| Total | | 14,200 | 100.00 |

## IV. EXPERIMENTAL RESULTS

### A. Experiment 1

Fig. 4 and Fig. 5 show the experimental results of PER/WER comparison between the baseline and the proposed method. The proposed method outperforms the baseline in all context spans from 3-gram to 9-gram, except 3-gram in WER. The proposed method shows 6.95% and 29.19% in PER and WER, respectively, while the baseline - 7.74% and 29.81% at best. Thus, the proposed method presents 10.2% and 2.07% of relative improvement in PER and WER, respectively. Regarding the context spans, 5-gram shows the lowest PER and WER. In addition, it is observed that the higher the n-gram is, the earlier the performance is saturated.

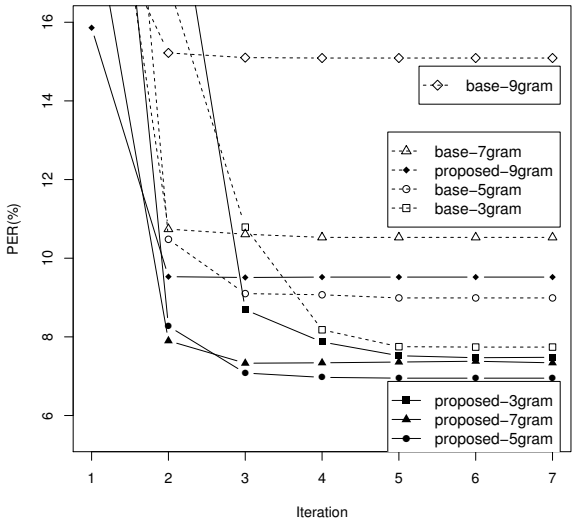


Fig. 4. Performance comparison in PER

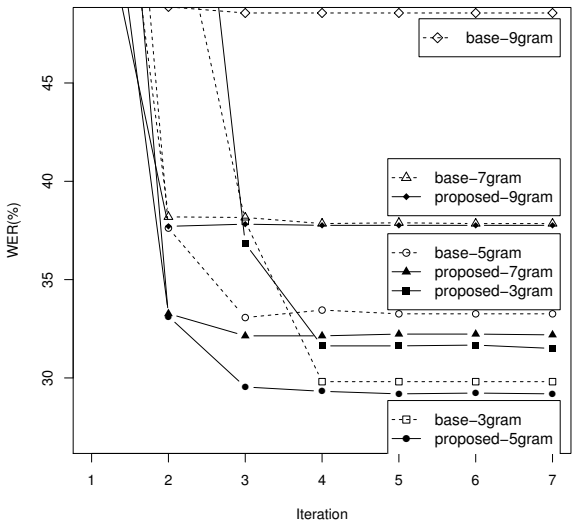


Fig. 5. Performance comparison in WER

The second measure we use for evaluation is F-score by

considering both precision and recall. By the result, the graph of F-score is shown in Fig. 6. Like the preceding results of PER/WER, the proposed method shows better performance than the baseline in F-score as well. The best performances in the proposed method and the baseline are 0.539 and 0.509 at 5-gram, respectively, which means 5.89% in relative improvement.

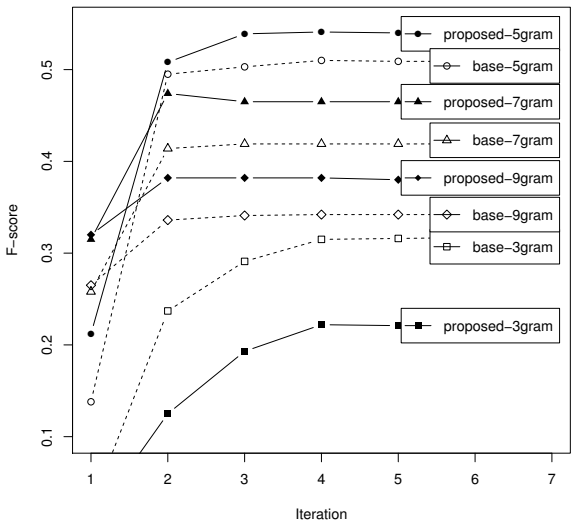


Fig. 6. Performance comparison in F-score

### B. Experiment 2

In the experiment 2, we compare the performance in terms of domains of the data to observe the influence of loanword domains on the G2P performance. The model is trained using the proposed method. We measure the performance applying 5-gram of context span at $6^{th}$ iteration, where the best performance is shown. The experimental results are presented in Table VII.

TABLE VII
EXPERIMENTAL RESULTS IN TERMS OF DOMAINS

| Training set | Test set | PER(%) | WER(%) | F-score |
|---|---|---|---|---|
| POI | POI | 6.66 | 30.20 | 0.530 |
| | Music | 10.47 | 35.44 | 0.417 |
| | Set-top box | 11.49 | 41.16 | 0.340 |

Table VII presents that loanword pronunciation modeling performance degrades when training and test sets come from different domains. Compared with the decreases in PER and WER, reduction of F-score is relatively larger. The performances of POI vs. music and POI vs. set-top box pairs are relatively similar. The results imply that loanword pronunciation modeling is influenced by data domains.

## V. CONCLUSION

This paper provides pronunciation modeling of loanwords in Korean by using phonological knowledge and syllable-based segmentation. We propose a syllable-based segmentation method considering phonological knowledge of loanwords in Korean. Performance of the baseline and the proposed method are measured using PER/WER and F-score at various context spans. Experimental results show that the proposed method outperforms the baseline in every measure. Especially, the best result is shown at 5-gram, which contains two previous, one current, and two following contexts. It is noteworthy that pronunciation modeling for loanwords in Korean is enhanced by reflecting phonological knowledge.

This research is studied in the context of ASR for navigation and set-top box applications. In our future research, therefore, we need to observe how pronunciation modeling using the proposed method influences the performance of ASR. Furthermore, in computer-assisted pronunciation training (CAPT) for Korean learners of English, error pronunciations by learners should be predicted for automatic pronunciation error detection and feedback [22]. Korean learners, especially in beginner level, are likely to pronounce English words in their own loanword pronunciations. Therefore, it is expected that there would be positive effect on CAPT for Korean learners of English, when results of loanword G2P results are included into the mispronunciation sequences of the learners.

## ACKNOWLEDGMENT

This work was supported by SK Telecom for modeling loan word pronunciations in Korean.

## REFERENCES

[1] J.-E. Cha, "A study on the standard pronunciation of the words of foreign origin and related matters," *Korean Linguistics*, vol. 35, pp. 363–390, 2007.
[2] C. Paradis and D. Lacharité, "Preservation and minimality in loanword adaptation," *Journal of Linguistics*, vol. 33, no. 2, pp. 379–430, 1997.
[3] E. Lee, "Usage of loanwords in multimedia," *Saegugeosaenghwal*, vol. 8, pp. 41–59, 1998.
[4] The National Institute of Korean Language, "The Korean standard dictionary," 1999.
[5] H. Kang, "English loanword in Korean," *Studies in Phonetics, Phonology, and Morphology*, vol. 2, pp. 21–47, 1996.
[6] S. Davis and S.-H. Shin, "The syllable contact constraint in Korean: An optimality-theoretic analysis," *Journal of East Asian Linguistics*, vol. 8, no. 4, pp. 285–312, 1999.
[7] S. Davis and M.-H. Cho, "Phonetics versus phonology: English word final /s/ in Korean loanword phonology," *Lingua*, vol. 116, no. 7, pp. 1008–1023, 2006.
[8] H. Kim, "Korean adaptation of English affricates and fricatives in a feature-driven model of loanword adaptation," in *Loan phonology* (A. Calabrese and W. L. Wetzels, eds.), vol. 307, pp. 155–180, Amsterdam; Philadelphia: John Benjamins Publishing, 2009.
[9] K.-N. Lee and M. Chung, "Morpheme-based modeling of pronunciation variation for large vocabulary continuous speech recognition in Korean," *IEICE Transactions on Information and Systems*, vol. E90-D, no. 7, pp. 1063–1072, 2007.
[10] L. Wang and R. Tong, "Pronunciation modeling of foreign words for mandarin ASR by considering the effect of language transfer," in *INTERSPEECH 2014*, (Singapore), pp. 1443–1447, 2014.
[11] T. Modipa and M. H. Davel, "Pronunciation modelling of foreign words for sepedi ASR," in *21st Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)*, (Stellenbosch, South Africa), pp. 185–189, 2010.
[12] H. Ryu and M. Chung, "Pronunciation variations of loan words produced by Korean speakers," in *Korean Society of Speech Sciences (KSSS) 2014 Fall Conference*, (Seoul, Korea), pp. 173–174.

[13] Y. Kang, M. Kenstowicz, and C. Ito, "Hybrid loans: a study of English loanwords transmitted to Korean via Japanese," *Journal of East Asian Linguistics*, vol. 17, no. 4, pp. 299–316, 2008.

[14] The National Institute of the Korean Language, "Regulation concerning Hangeul transcription of loanwords," 28 February, 2015 2015.

[15] H. Hong, J. Kim, and M. Chung, "Effects of Korean learners' consonant cluster reduction strategies on English speech recognition performance," in *INTERSPEECH 2010*, (Makuhari, Japan), pp. 1858–1861, 2010.

[16] G. D. Yurn, "In justification of "jjajangmyeon"," *Korean Linguistics*, vol. 30, pp. 181–205, 2006.

[17] M. Bisani and H. Ney, "Joint-sequence models for grapheme-to-phoneme conversion," *Speech Communication*, vol. 50, no. 5, pp. 434–451, 2008.

[18] D. Jouvet, D. Fohr, and I. Illina, "Evaluating grapheme-to-phoneme converters in automatic speech recognition context," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2012*, (Kyoto, Japan), pp. 4821–4824, 2012.

[19] T. Schlippe, W. Quaschningk, and T. Schultz, "Combining grapheme-to-phoneme converter outputs for enhanced pronunciation generation in low-resource scenarios," in *4th Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU '14)*, (St.Petersburg, Russia), pp. 139–145, 2014.

[20] T. Schlippe, S. Ochs, and T. Schultz, "Grapheme-to-phoneme model generation for Indo-European languages," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2012*, (Kyoto, Japan), pp. 4801–4804, 2012.

[21] S. Hahn, P. Vozila, and M. Bisani, "Comparison of grapheme-to-phoneme methods on large pronunciation dictionaries and LVCSR tasks," in *Interspeech 2012*, pp. 2538–2541, 2012.

[22] J. Bang, J. Lee, G. G. Lee, and M. Chung, "Pronunciation variants prediction method to detect mispronunciations by Korean learners of English," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 13, no. 4, pp. 1–21, 2014.