

Analysis of Korean Dialect Obstruents in a Large Corpus Using Speech Recognition Technology*

JOOYOUNG LEE

Seoul National University

SUNHEE KIM

Seoul National University

MINHWA CHUNG

Seoul National University

1 Introduction

Vowels and consonants are fundamental factors in analyzing variations across dialects. In Korean dialectology, vowels have been extensively studied, with numerous comparative analyses available on vowel systems, including Kwak (2003) and Lee (2005). Studies on Korean dialectal vowels often focus on phenomena such as the diphthongization of /y/ and /ø/; the merger of /æ/ and /e/; the contrasts between /ʌ/ and /u/; and the pronunciation of the diphthong /ɥi/ (Kim 2001, Kwak 2003, Ahn 2012, Kang 2018, Park 2022, Park & Yun 2023).

* This work is supported by the Department of Linguistics at Seoul National University under the project *LINGUISTICS SNU 10-10 INITIATIVE* in 2024, and by the Ministry of Science and ICT, Korean Government, under the project *Voice Phishing Information Collection and Processing and Development of a Big Data Investigation Support System* in 2024.

Japanese/Korean Linguistics 31.

Edited by William Giang, Lucien Brown, Shimako Iwasaki, Satoshi Nambu, and Daniel Pieper.

Copyright © 2024, Jooyoung Lee, Sunhee Kim, and Minhwa Chung.

In contrast, research on consonantal patterns across Korean dialects remain fragmented, with many studies documenting individual characteristics limited to specific dialects (e.g. ‘n’ insertion in Gwangju dialect; Jin 2016). Additionally, previous studies on dialectal consonants have relied on small scale datasets. For example, Kim (2021) investigated consonants in coda positions within word stems, while Park (2024) examined the phonology of obstruents and sonorants preceding glottal sounds. Although these studies have contributed to our understanding of dialectal consonants, their analyses rely on small scale datasets, restricting the discovery of broader phonological patterns that may emerge only in large scale corpora.

One notable exception is a large scale study involving 528 speakers (Kim 2006); however, its primary focus was on collecting and providing acoustic features extracted from dialectal speech rather than conducting a comprehensive analysis of consonantal variations across dialects. This gap highlights the need for large scale investigations into consonantal patterns to uncover systematic phonological variations across Korean dialects.

This paper examines consonantal variations across Korean dialects using an extensive dialectal speech corpus, revealing novel patterns that have not been documented in previous research. Specifically, we focus on obstruents, which constitute the majority of Korean consonants—15 out of 19 (Shin 2015: 7). To facilitate large scale comparative analysis, we introduce a novel methodology that autonomously compares standard phoneme sequences with actual pronunciation patterns. This approach integrates grapheme-to-phoneme conversion with automatic phoneme recognition, enabling the extraction of vast phoneme-level data for statistical analysis. Through this method, we systematically identify dialect specific phonological patterns in obstruents.

The significance of this study lies in its pioneering large scale investigation of phonological variation in Korean dialectal obstruents. Moreover, the proposed automated processing framework proves effective in detecting previously unreported patterns, demonstrating its applicability to broader linguistic and phonological research.

2 Methodology

Figure 1 illustrates the overall workflow of our proposed method for the automatic comparison of dialectal obstruents. The core principle underlying our methodology is that dialect specific phonemic patterns emerge from discrepancies between the canonical phonemic representation and the actual pronounced representation. The canonical representation refers to the phoneme sequence as it is pronounced in the standard language, which may differ from actual spoken realizations. In contrast, the pronounced representation corresponds to the sequence of phonemes as they are articulated in speech.

We hypothesize that systematically recurring differences between canonical and pronounced phonemes can be interpreted as distinctive phonemic patterns characteristic of a particular dialect. For instance, in Figure 1, the standard pronunciation of the word *ssawum* ‘a quarrel’ includes the tense sibilant /ʃ/ (i.e. /ʃaum/), whereas in the Gyeongsang dialect, this phoneme is realized as the lax sibilant [s] (i.e. [saum]). If this pattern of detensification—where /ʃ/ is consistently replaced by [s]—occurs with high frequency, it can be inferred that detensification is a phonological feature of the Gyeongsang dialect.

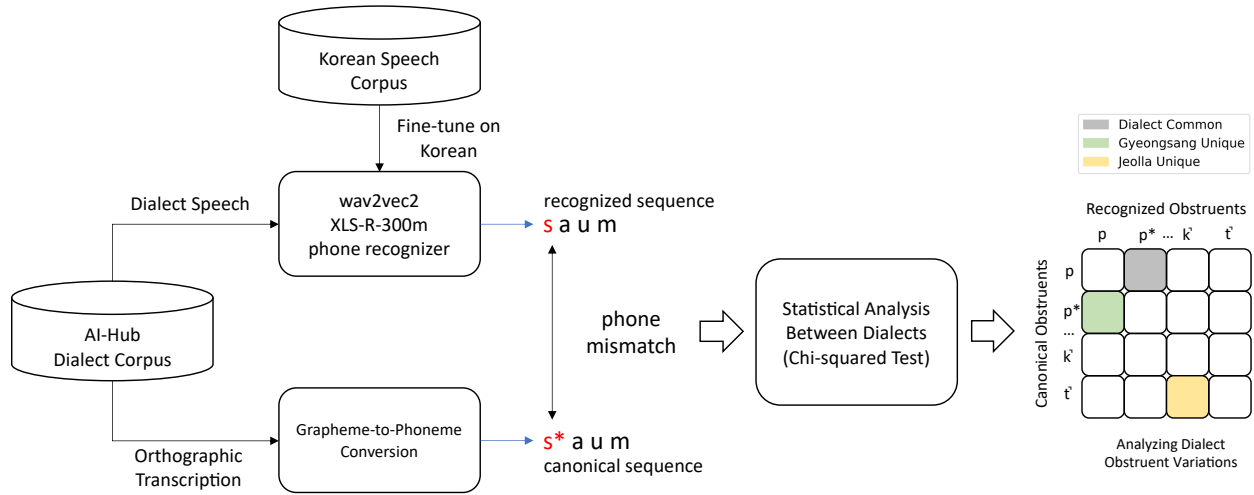


Figure 1: An Overall Framework of the Proposed Automatic Dialectal Obstruent Comparison

2.1 Automatic Phoneme Recognition

Automatic phoneme recognition is the process of transcribing spoken language into a sequence of phonemes, enabling the extraction of phonemic-level representations of a dialect speaker’s actual pronunciation. Unlike word-level speech recognition, which focuses on identifying entire words, phoneme recognition operates at a finer granularity by analyzing the fundamental sound units that constitute spoken language. This process is essential to our study, as it allows us to capture pronunciation variations that may not be evident at the lexical or orthographic level.

2.2 Grapheme-to-Phoneme Conversion

Grapheme-to-phoneme (G2P) conversion is a process of representing the pronunciation of a word using predefined phonemic symbols (phonemes) based on its written form (graphemes) (Bisani & Ney 2008). A given set of graphemes may correspond to different phonemic sequences depending on dialectal variations. The output of G2P conversion represents the expected pronunciation in the standard language rather than the actual spoken form. This step enables us to obtain a standardized phonemic transcription of words spoken by dialect speakers, facilitating direct comparisons between standard and dialectal pronunciations.

2.3 Phonemewise Comparison

Recognized phoneme sequences from automatic phoneme recognition are compared phoneme-by-phoneme with their corresponding canonical phoneme sequences obtained from G2P conversion. However, discrepancies in the number of phonemes between the two sequences frequently occur, requiring the insertion of an empty symbol (denoted as ‘***’ in this study) to align sequences of unequal length.

For example, consider the recognized phoneme sequence ‘KK AA AA JJ I’ (a ‘puppy’ in the Gyeongsang dialect) and its corresponding standard phoneme sequence ‘KK AA NG AA JJ I’. The recognized sequence lacks the phoneme ‘NG’ between the two ‘AA’ phonemes. To ensure alignment, an empty symbol is inserted: ‘KK AA *** AA JJ I’.

Since phoneme sequences in this study are derived from complete sentences, they can be quite long. To optimize alignment, we apply the Needleman-Wunsch algorithm (Needleman & Wunsch

1970), which minimizes the number of inserted empty symbols and mismatched phonemes. This algorithm ensures that phoneme sequence pairs maintain the same length while reducing alignment errors, thereby improving the accuracy of phonemic comparison.

2.4 Statistical Analysis

Once the processed phoneme sequence pairs are obtained, a chi-squared test of independence is conducted to identify significantly frequent mismatches between specific phoneme pairs. Since each phoneme in the recognized sequence is paired with a corresponding phoneme from the canonical sequence, we can systematically quantify occurrences of matched and mismatched phonemes.

The chi-squared test allows us to compare dialectal pronunciations with standard Korean. If a particular mismatch pair appears significantly more frequently in dialectal speech than in the standard language, we interpret the recognized phoneme as a dialect specific variant. Furthermore, if a specific dialect consistently exhibits a phonemic pattern that does not appear in other dialects, we classify that phoneme as characteristic of that particular dialect.

The test results are presented in a confusion matrix, where each cell at row i and column j represents the frequency significance of the canonical phoneme i being recognized as phoneme j . Identical phonemes in both the canonical and recognized transcriptions are excluded from statistical analysis, as they do not indicate dialectal variation.

3 Experiments

3.1 Dialect Corpus

This study examines the AI-Hub Senior Dialect Speech Corpus (AI-Hub 2021), a dataset consisting of spontaneous speech samples from speakers aged 60 and above. We specifically selected the Gyeongsang and Jeolla dialect datasets, which contain 1,203 hours and 1,015 hours of speech data, respectively. The rationale for focusing on senior speakers is that dialectal patterns tend to be more consistently preserved among older generations compared to younger speakers, who are often more influenced by the standard language.

3.2 Experimental Settings

The automatic phoneme recognizer used in this study is built by fine-tuning the pretrained ‘wav2vec 2.0 XLS-R’ model (Babu et al. 2022). XLS-R is a wav2vec 2.0-based model trained for crosslingual speech representation learning, utilizing approximately 500,000 hours of publicly available speech data across 128 languages. We selected this model due to its extensive linguistic diversity, which may help capture the phonetic variability present in different Korean dialects.

For fine-tuning, we used a read speech corpus of standard Korean (in-house dataset), containing 120 hours of speech samples with both orthographic and phonetic transcriptions. Our fine-tuned recognition model achieves a consonant error rate of 3.99%, which is sufficiently low for the purposes of this study.

For grapheme-to-phoneme (G2P) conversion, we employ the Montreal Forced Aligner (MFA) (McAuliffe et al. 2017), version 1.0. The MFA toolkit provides pretrained acoustic models for multiple languages, including Korean. The Korean G2P model used in this study was trained on

48,682 words, incorporating 55 graphemes and 56 phonemes, and achieves a phone error rate of 0.64%.

For statistical analysis, we apply chi-squared tests with a significance threshold of $p = 0.001$ to detect statistically significant differences between dialectal and standard Korean phonemic patterns. Additionally, the experiments are conducted separately for obstruents occurring in onset and coda positions as they have different phonetic values.

4 Results and Discussion

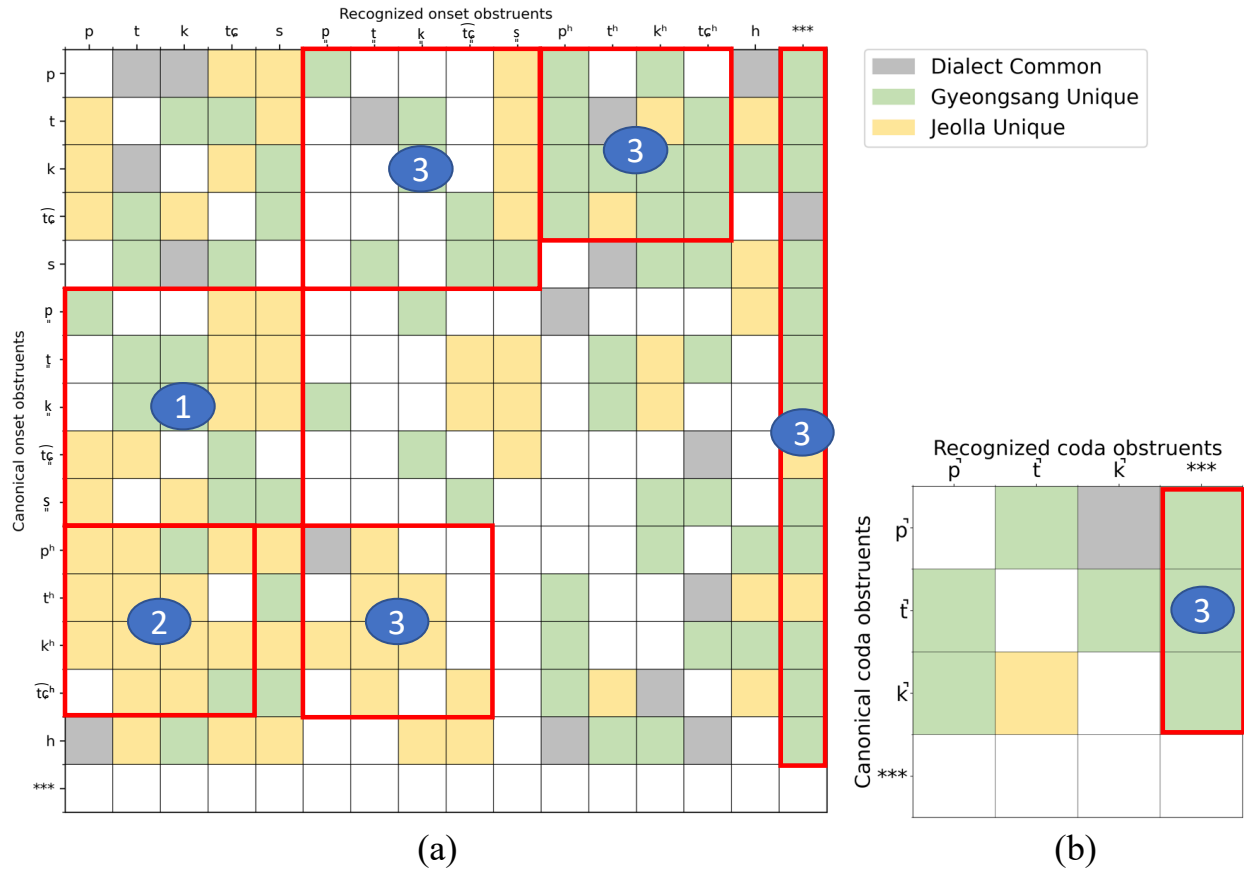


Figure 2: Confusion Matrices of Dialectal Variations in Onset and Coda Obstruents

4.1 Overview of the Results

Figure 2 presents confusion matrices derived from the statistical analysis of obstruents in both onset and coda positions. The rows represent canonical obstruents obtained through grapheme-to-phoneme (G2P) conversion, while the columns indicate the corresponding recognized phonemes from automatic phoneme recognition.

Dialect	Canonical	Recognized	Mismatch Rate (%)
Gyeongsang	h	***	51.26
	$\widehat{t}_{\text{ɸ}}$	$\widehat{t}_{\text{ɸ}}$	37.83
	$t_{\text{ɸ}}$	t	37.52
	$s_{\text{ɸ}}$	s	30.60
	$p_{\text{ɸ}}$	p	29.01
Jeolla	k^h	k	27.31
	t^h	t	17.35
	p^h	p	16.92
	\widehat{k}	k^h	5.33
	$\widehat{t}_{\text{ɸ}}^h$	t^h	5.13

Table 1: Top Five Obstruent Pairs with the Highest Mismatch Rates for Each Dialect

Each cell in the matrices is color-coded to indicate dialectal variation:

- Gray represents obstruents common to both dialects.
- Green denotes Gyeongsang-unique obstruents.
- Yellow signifies Jeolla-unique obstruents.
- Uncolored (blank) cells indicate cases where the obstruents do not significantly differ from the standard language.

To systematically analyze phonemic variation, we categorized obstruents into three groups:

1. Lax obstruents (unmarked)
2. Tense obstruents (marked with a vertical '=' diacritic)
3. Aspirated obstruents (marked with a superscript 'h' diacritic)

Each confusion matrix visualizes how canonical obstruents (rows) correspond to the phonemes recognized in actual pronunciation (columns). The three phoneme categories were analyzed along the right-down diagonal to identify deviations from the expected pronunciation patterns.

4.2 Observed Patterns of Dialectal Variation

Three major patterns are observed in the results:

1. Lenition of tense obstruents to lax obstruents in the Gyeongsang dialect
2. Lenition of aspirated obstruents to lax obstruents in the Jeolla dialect
3. Newly observed phonological processes

Figure 2(a) and Table 1 further highlight that the lenition of tense obstruents in the Gyeongsang dialect extends beyond sibilants, which was reported in a previous study (Lee 2002). While prior research confirmed the lenition of / $\text{ɸ}/$ to [s], this study additionally identifies the lenition of stops and affricates such as / $p_{\text{ɸ}}$, $t_{\text{ɸ}}$, $k_{\text{ɸ}}$, $\widehat{t}_{\text{ɸ}}/$, a phenomenon not previously documented for obstruents. Similarly, in the Jeolla dialect, the lenition of aspirated obstruents through /h/ deletion aligns with finding from previous research (Jang 2019).

4.3 Novel Phonological Shifts

Beyond these expected patterns, this study also uncovers novel phonological shifts in both dialects. In the Gyeongsang dialect, tensification and aspiration of lax obstruents are newly observed, along with obstruent deletion in both onset and coda positions. In the Jeolla dialect, tensification of aspirated obstruents emerges as a previously unreported phenomenon. Further investigation with valid word examples is necessary to determine whether this realization represents a genuine phonological shift or merely a numerical artifact unrelated to dialectal patterns.

4.4 Limitations

While these findings contribute to a deeper understanding of dialectal obstruent variation, the study has three primary limitations:

1. Obstruent position within words: The analysis does not differentiate obstruents by their position within a word (e.g. word-initial vs word-medial). This distinction is critical, as obstruents may be pronounced differently depending on their phonotactic environment. However, due to challenges in distinguishing between word-initial obstruents following an open syllable and word-medial obstruents, this variable was not included in the analysis.
2. Influence of following vowels: The realization of dialectal obstruents may vary depending on the subsequent vowel. While considering this factor could provide deeper insights, it was excluded to avoid excessive complexity in the analysis.
3. Dialectal subgroup variation: The results may not be generalizable to all subregions within the Gyeongsang and Jeolla dialect areas. For instance, the phonemic contrast between /s/ and /ʃ/ differs between Gyeongsang dialects, with southwestern Southern Gyeongsang maintaining the contrast while northeastern Northern Gyeongsang lacking it (Choi 2005: 47). Such variation suggests that future research should conduct analyses at more fine-grained regional levels.

5 Conclusion

This study has identified new findings on the characteristics of obstruents in Korean dialects through the analysis of a large scale speech corpus. The results demonstrate the effectiveness of an automated approach that integrates phoneme recognition, G2P conversion, and statistical analysis for comparing obstruents. This methodology proves to be a valuable tool and has the potential for application in other research areas that require large scale phoneme analysis but have yet to be explored with extensive datasets.

Future studies can further investigate dialectal obstruent realizations by considering their positive effects within a word and the influence of subsequent vowels. This can be achieved by leveraging both phoneme-level and word-level segmentations obtained from MFA outputs, allowing for a more detailed phonetic analysis.

References

- Ahn, M. 2012. The Sociophonetics Study on the Vowel /—/ and /ɿ/ of Daegu Dialect. *Culture and Convergence* 34: 65–83.
- AI-Hub. 2021. AI-Hub senior dialect speech corpus. Accessed September 13, 2024, from <https://www.aihub.or.kr>

- Babu, A., C. Wang, A. Tjandra, K. Lakhota, Q. Xu, N. Goyal, K. Singh, P. von Platen, Y. Saraf, J. Pino, A. Baevski, A. Conneau, and M. Auli. 2022. XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale. *Proceedings of Interspeech 2022*, ed. ISCA, 2278–82. Incheon: ISCA.
- Bisani, M. and H. Ney. 2008. Joint-sequence models for grapheme-to-phoneme conversion. *Speech Communication* 50(5): 434–51.
- Choi, M. 2005. The System of Korean Dialectology. *The Journal of Korean Dialectology* 1: 35–72.
- Jang, S. 2019. A study on the orthography of the Jeollabuk-do Dialect Dictionary. *Journal of Korean Literature* 71: 97–120.
- Jin, J. 2016. A Study on the use of “n”-insertion in Gwangju dialect: focus on the Sino-Korean word. *Journal of The Society of Korean Language and Literature* 65: 55–76.
- Kang, H. 2018. The Diachronic Development and Dialectal Differentiation of ‘eo→eu’ Vowel Raising – Focusing on the Central and the Jeonnam dialect. *The Society for Korean Language & Literary Research* 95: 5–30.
- Kim, H., Y. Choi, and D. Kim. 2006. An Experimental Study of Korean Dialectal Speech. *Phonetics and Speech Sciences* 13(3): 49–65.
- Kim, M. 2001. A Study on the Traits of Korean Diphthong ‘uji’ by Means of Formant Analysis. *The Korean Language and Literature* 72: 1–15.
- Kim, S. 2021. *A Study on the Morphophonological Change of Consonant-ending Stems in Korean*. Doctoral dissertation, Seoul National University.
- Kwak, C. 2003. The vowel system of contemporary Korean and direction of change. *Journal of Korean Linguistics* 41: 59–81.
- Lee, G. 2002. Phonetic study on tensing and voicing of Gyung-Nam dialect -on Fricatives /s/ and /s̺/-. *HAN-GEUL* 255: 7–45.
- Lee, K. 2005. On the Dialectal Variation and the Division of Dialect Areas. *The Journal of Korean Dialectology* 1: 103–23.
- McAuliffe, M., M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger. 2017. Montreal forced aligner: Trainable text-speech alignment using kaldi. *Proceedings of Interspeech 2017*, eds. ISCA, 498–502. Stockholm: ISCA.
- Needleman, S. and C. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* 48(3): 443–95.
- Park, J. 2022. The Phonological Perception and Articulation by Generation in Daegu Dialect /—/ and / ʃ / . *The Journal of Studies in Language* 38(2): 127–42
- Park, J. 2024. *A Phonological Study of Consonant-Glottal Sound Chains in Korean*. Master’s thesis, Seoul National University.
- Park, J. and S. Yun. 2023. South Jeolla Korean speakers’ perception of the /ʌ/ → /i/ vowel raising. *Korean Language Research* 64: 1–18.
- Shin, J. 2015. Vowels and Consonants. *The Handbook of Korean Linguistics*, eds. L. Brown and J. Yeon, 3–21. New Jersey: Wiley.