Phonetic integration in Heritage Korean language-mixing

Nicoline Butler

University of Hawai'i at Mānoa
butlern@hawaii.edu

ABSTRACT

Phonetic integration, the application of native phonological features to non-native utterances, is frequently operationalized to differentiate between loanwords and other categories of language mixing. Amongst bilingual communities however, recent studies have reported variability of phonetic integration across all types of language mixing, arguing that factors such as phonological segment serve as stronger predictors of integration.

This study analyzes social and linguistic factors which motivate phonetic integration amongst Heritage Korean-English bilinguals and explores the interaction between phonetic integration and speaker identity. Counter to previous studies, a linear regression model shows language mixing type to have the greatest effect on integration, with social factors such as speaker generation, language preference and cultural identity also surfacing as significant. These findings suggest that rather than merely linguistic diagnostic, phonetic integration can also function as an indicator of the unique identity and culture of bilingual speakers.

Keywords: phonetic integration, language contact code-switching, Heritage Korean, corpus linguistics

1. INTRODUCTION

Language mixing, the alternation of two languages within a single discourse, sentence, or constituent [1], is a common language practice among bilinguals, and has received attention not only for how it is used to signal identity, but also for its potential to drive language change. Within bilingual research, the distinction between established loanwords and other spontaneous language-mixing has been particularly polarizing, and phonetic integration, the application of recipient language phonology to donor language words, has frequently been operationalized as a key diagnostic.

In stark contrast to these long-held assumptions, recent studies have demonstrated that phonetic integration is not a reliable metric for delineation. They report that, in fact, that amongst bilingual French-English speakers of Quebec, phoneme itself is the most significant predictor of phonetic integration [2].

These findings lead to interesting questions about the role of phonetic integration in language mixing strategies:

i. Does the primacy of phoneme in integration apply to other bilingual language-mixing communities or is it exclusive to French-English speakers of Quebec?

ii. are there other linguistic and social features not considered in previous studies that factor into integration? If so, how significant is their contribution to this phenomena?

Poplack and colleagues themselves stress that patterns of phonetic integration should not be assumed, but rather "established segment by segment, word by word, speaker by speaker, [and] community by community" [2, p. 152]. While integration may, in some cases, be used to signal loanword status, it is likely that other factors, both linguistic and social, condition phonetic integration amongst bilingual speech communities.

This study endeavors to pursue these answers by investigating the phonetic integration of bilingual Heritage Korean speakers of the Greater Toronto Area. Applying methods of quantifying integration from previous research to sociolinguistic interviews with bilingual speakers of Heritage Korean, this study explores what social and/or linguistic factors may condition phonetic integration across English language-mixing types.

We begin by providing an overview of phonetic integration and then outline language-mixing habits amongst heritage bilingual communities reported in previous work. We then present the data, design, and results of the current study. In discussing the methodology used to address the above questions, this paper also highlights challenges inherent to investigating heritage languages as a distinct variety, rather than simply as an extension of their homeland counterpart.

2. BACKGROUND

2.1. Language Mixing and Phonetic Integration

Language mixing strategies of bilingual speakers have long-held the attention of linguists. While qualitative discussions of language mixing behaviors and their influence on language variation are numerous, quantitative analyses supporting these findings, particularly of the phonetic variety, are more difficult to come by.

Previous investigations have demonstrated that phonetic integration in language-mixing exhibits

significant variability [3, 2], and can function independently from morphosyntax of the same phenomena [1, 5, 2]. Of most significance to the current study are the findings of Poplack *et al.*, who assert that the assessment of integration at any level "depends of the successful identification of conflict sites - elements that can be unambiguously be characterized as belonging to [the donor] or [recipient language]" [2, p. 133].

In their work on French-English language mixing Québécois bilinguals, they report variability across all factors of analysis. They summarize that "intra-individual variation overwhelms inter-individual variation", and that conflict phonemes themselves are the most significant factor, each exhibiting different amenability to integration.

2.2. Heritage Korean and language mixing

Discussions on the contact between English and Korean have been widely represented in the literature. As one of the first languages analyzed in loanword phonology, Korean has been a popular choice for investigating the phonological processes involved in adapting donor tokens into the recipient language [6, 7, 8]. Additionally, there have been numerous studies on Korean-English code-switching, with particular emphasis on its use in constructing identity and positioning oneself vis-à-vis their interlocutor [see 9, 10, and 11]. Previous studies on linguistic integration and language mixing have also found that a speaker's cultural identity and generational heritage were the most significant predictors of language mixing [12].

An investigation into phonetic integration in Heritage Korean-English language mixing provides an opportunity to examine the findings of previous research within the context of a bilingual community consisting of languages with different societal status in the country. Previous studies have focused on bilingual speakers of French and English, both official languages of Canada which enjoy relatively similar prestige. It could be argued that speakers who display a working proficiency of both English and French enjoy higher prestige than their monolingual counterparts, due to the socioeconomic benefits afforded to bilinguals. In contrast, Heritage Korean speakers do not experience the same level of social status when speaking Korean as they do when speaking English, and many interviews coded in this study included accounts of racial discrimination experienced when speaking Korean. Comparing the integration patterns of Heritage Korean speakers and French-English bilinguals could offer valuable insights into the relationship between phonetic integration and extralinguistic social factors.

3. MATERIALS AND METHODS

This study makes use of sociolinguistic interviews from the Heritage Language Documentation and Change Project, hereon HLVC [13]. Typically 30-40 minutes in length, these interviews were conducted in Korean between 2009 and 2010. During interviews, participants were asked questions pertaining to their cultural identity, social lives, and experiences based on questionnaires adapted from [14] and [15].

Interviews were transcribed in Hangul by various members of the HLVC project, with instances of language mixing indicated by a shift to English transcription. (see [12] for more details on the interview procedures and transcription procedures).

Twenty-eight interviews were included in the study, and instances of language mixing were segmented and coded in ELAN (see Table 1 for demographics).

Gen.	Korean	Kor-Can	Canadian	*total*
1st	6	8	1	*15*
2nd	2	7	4	*13*
total	*8*	*15*	*5*	*28*

Table 1: Speakers included in study by generation of immigration and cultural orientation. Ages range from 39-85 (avg. 60) in first generation speakers and 18 to 42 (avg. 27) in second.

3.1. Quantifying Integration

Phonetic integration was quantified using conflict phonemes, sounds that do not occur in the recipient language and are easily identifiable as belonging to the donor language [16]. The fricatives /z v f θ, ð, and s/ which occur in English are not native to Korean phonology, and the application of Korean phonological processes, when adapted, is predictable (illustrated in Table 2).

PHONEME	ENGLISH DONOR		KOREAN ADAPTATION	
/z/ → [ts]	/zig.zag/	zigzag	[tsi.kɨ.tsæ.kɨ]	지그재그
/v/ → [p]	/vik.to.ri/	victory	[pik.to.ɹi]	빅토리
/f/ → [p^h]	/al.fa.bet/	alphabet	[ɑɭ.p^hɑ.pɛt]	알파벳
/θ/ → [t^h]	/kaθ.lik/	Catholic	[k^hɑ.t^hoʟ.ʟik]	가톨릭
/ð/ → [t]	/ɹi.ðum/	rhythm	[ɹi.tɨm]	리듬
/s/ → [ʃ]	/bei.sik/	basic	[pɛ.i.ʃik]	베이식

Table 2: Phonological processes by which conflict phonemes are adapted into Korean.

This predictability allows a clear coding system for instances of language mixing, where instances containing [f, v, θ, ð, z and s] are assigned a zero (*unintegrated*), and instances in which they are not are assigned a one (*integrated*).

3.2. Categorizing Language Mixing

Poplack *et al.* [2] identifies three types of language-mixing: established loanwords, nonce borrowings, and codeswitches. In order to better suit an investigation of Heritage Korean, delineations for each have been adapted as follows:

(1) *loanwords*: words adopted into the language.

(2) *nonce borrowings*: single speaker borrowings.

(3) *codeswitching* : multi-word switches.

In comparison to speech communities in previous studies, identifying established loanwords in Heritage Korean is less straightforward. The use of existing Korean dictionaries as the diagnostic for determining the status of Heritage Korean loanwords is reductive, possibly perpetuating the false narrative that homeland Korean is 'proper' Korean, and limits our understanding of Heritage speakers. For this reason, loanwords in this study are considered "adopted into the language" when they have been uttered more than once by more than one speaker across all interviews (1).

Additionally, nonce borrowings (2) are extended from "single-word utterance used once by a single speaker", to any single word used by a single speaker, regardless of frequency. This adjustment is made to account for the style of interview utilized in the HLVC, which frequently results in the repetition of certain nonce-borrowings. The definition of code-switching given in (3) remains unchanged.

3.3. Coding and factors

Approximately 14.5 hours of transcribed data was coded, resulting in a total of 2642 unique instances of language mixing. To produce a multi-use dataset, each of these instances were coded for language mixing-type and grammatical category, and well for heritage generation, language preference, and cultural orientation. 943 language-mixing utterances containing the fricatives /f, v, θ, ð, z and s/ were selected extracted for analysis. These tokens were also coded to include integration, conflict phoneme, and segment position.

4. DATA ANALYSIS

Of the 943 coded tokens, 295 phonemes were phonetically integrated to Korean, resulting in a mean of 33 tokens per speaker and a total integration average of 30.7%.

While working with naturalistic spontaneous speech has many benefits, it also poses difficulties for creating balanced data sets. Amongst the interviews selected for this study, not all speakers produced every target phoneme, /θ/ and /ð/ being the most difficult to come by (13 and 18 interviews respectively). Speakers do, however, all exhibit the ability to produce both integrated and unintegrated conflict phonemes.

TYPE	θ	ð	v	z	s	f	TOTAL
loanwords	0	101	109	18	61	42	*331*
nonce borrowing	5	124	81	38	86	38	*372*
codeswitches	41	59	50	21	61	8	*240*
TOTAL	*46*	*77*	*88*	*208*	*240*	*284*	*943*

Table 3: Distribution of data by language-mixing type and phoneme.

Across tokens, utterances containing /f/ occurred most, while /θ/ occurred least. When looking at integration, /s/ is integrated most and /θ/ is integrated least. The distribution of each conflict phoneme is outlined in Table 3.

4.2. Linguistic factors

Language-mixing type: Also presented in Table 3 is the total number of tokens observed in each language mixing type, as well as their occurrence with each conflict phoneme. Most frequent was nonce borrowings, representing approximately 39% of all data, while codeswitching occurred the least, representing only 25% approximately. Phonemes occurring in loanwords were integrated most often (50.4% of all realizations), while phonemes contained in codeswitches integrated least often (only approx. 13.7%).

Position of phoneme: Amongst the phonemes observed in this dataset, most occurred word-medially (406 phonemes), some occurred word-initially (312 phonemes), and phonemes occurring word-finally were the most rare (225 phonemes). Word-medial phonemes also integrated the most (36.2% of instances), while word-initial phonemes integrated the least (22.8% of instances).

Grammatical category: Grammatical category is by far the most disproportionate of all factors. Conflict phonemes occurring in nouns represent approximately 64% of all data, while sentences represent only around 25%. All other categories including adjectives and verbs each represent 6% or less of the remaining data. Phonemes occurring in nouns also integrated the most (37.4%), while phonemes occurring in sentences integrated the least (14.7%).

4.3 Social factors

Generation: Despite the higher number first generation speakers, second generation speakers made up a larger proportion of the data (55.5% of all instances) and exhibited a higher number of tokens per speaker. Conversely, despite lower total instances of language mixing (44.5%; avg. of 28 tokens per speaker), first generation speakers integrate much

more frequently than second (47.4% and 14.9% respectively).

Cultural orientation: While the largest number of speakers identified as Korean-Canadian, speakers who identified as Canadian and Korean both display a higher average number of tokens per speaker (39 and 37 tokens respectively). In addition, Korean speakers integrated the most frequently (47.4% of the time), while Canadian speakers integrated the least (14.9% of the time).

Language Preference: Although most speakers indicate a preference for speaking English or Korean, speakers who have no preference (both) tended to integrate the most (45.6% of the time).

5. REGRESSION ANALYSIS

Logistic regression modeling [17] was used to test the comparative influence of each factor on phonetic integration. Anova and *phia* [18] were used to conduct post-hoc analyses. The results of these tests are summarized in Table 4.

Mixing Type (p < .000)	[DF]	[chi.sq.]	[Pr(>|z|)	
nonce borrowing	1	15.56	< .000	***
codeswitch	1	12.72	< .001	***
loanword	1	0.00	0.949	
Language Preference (p < .000)				
English	1	20.38	< .000	***
Korean	1	30.03	< .000	***
both	1	**3.30**	0.069	
Speaker Generation (p < .000)				
1st	1	2.86	0.091	
2nd	1	39.55	< .000	***
Conflict Phoneme (p < .000)				
f	1	31.23	< .000	***
ð	1	10.58	0.004	**
θ	1	9.63	0.005	**
v	1	5.87	0.016	*
s	1	6.98	0.016	*
z	1	16.37	< .000	***
Cultural Identity (p = .02)				
Korean-Canadian	1	22.20	< .000	***
Canadian	1	22.33	< .000	***
Korean	1	8.86	0.003	**

Table 4: results of fixed effects of logistic regression model. glm (integration ~ 1 + mixing type + generation + phoneme + language preference + cultural identity + segment position + grammatical category, family = "binomial")

Based on this model, mixing-type and language preference appear to be particularly strong predictors of integration, with second generation speakers and those who identify as Canadian also showing higher levels of integration. No significant effects were found for grammatical category or segment position. Overall, these results suggest that the motivations behind phonetic integration amounts Heritage Korean speakers are complex and multifaceted.

6. DISCUSSION

The results of this study confirm that phonetic integration is not realized uniformly across language-mixing types. However, extra-linguistic factors such as language preference and speaker generation demonstrate a more pronounced effect on integration. The significance of these social factors challenges the assertions of previous studies, which suggested that "extralinguistic characteristics do not clarify the motivations behind phonetic realization" [2, p.147].

The study also raises important questions about the complex relationship between language, culture, and identity. While phonetic integration may be a matter of choice for many bilingual speakers, for Heritage speakers this choice may hold more consequences beyond mere ease of speech.

Pronunciation has long been held as the strongest hallmark of proficiency, and for Heritage speakers, the decision to integrate may be strongly motivated by how their proficiency will be perceived by interlocutors. This presents an interesting but complex conundrum: Do speakers measure their own proficiency to decide on language preference and/or cultural identity? What other criteria are considered when making this decision? Heritage language and culture in the diaspora is an incredibly complex and nuanced topic that undoubtedly goes beyond the scope of this paper, but highlights a need for future research on the nature of language contact and change in multilingual communities

This study is not without its limitations. Phonetic integration in Korean-English language mixing cannot be fully captured by diagnostics of this study. To more accurately account for phonetic integration as a whole, future studies could explore a more complex diagnostic that considers vowels, syllable structure, and suprasegmental features. Additionally, future studies could also consider speaker proficiency and the possibility of L1 intrusions in 1st get speakers [19]. Finally, another area of interest wound be to observe the effect of phonetic convergence and the cultural identity of the interlocutor on an individual's propensity to integrate.

Overall, this study provides a preliminary account of phonetic integration amongst speakers of Heritage Korean in the Greater Toronto area. Results speak to an interaction between social factors and language-mixing behaviors, and reinforce the long-held understanding that language does not exist in isolation, but rather is intricately interwoven with culture and identity.

7. REFERENCES

[1] Poplack, S. 1988. Contrasting patterns of code-switching in two communities. Codeswitching: anthropological and sociolinguistic perspectives 48, 215-244.

[2] Poplack, S., Robillard, S., Dion, N., Paolillo, J. 2020. Revisiting phonetic integration in bilingual borrowing. Language 96(1), 126-159.

[3] Fought, C. 2010. Ethnic Identity and Linguistic Contact. In The Handbook of Language Contact, Wiley-Blackwell, pp. 282–298.

[4] Bullock, B. E. 2009. Phonetic reflexes of code-switching. In B. E. Bullock, & A. J. Toribio (Eds.), The Cambridge handbook of linguistic code-switching (pp. 163–181). Cambridge: Cambridge University Press.

[5] Muldner, K., Hoiting, L., Sanger, L., Blumenfeld, L., and Toivonen, I. 2007. The phonetics of code-switched vowels. International Journal of Bilingualism, vol. 23, no. 1, pp. 37–52, Jun. 2017.

[6] Kenstowicz, M. 2005. The phonetics and phonology of Korean loanword adaptation. In: Rhee, S.-J. (ed), Proceedings of the First European Conference on Korean Linguistics. Seoul: Hankook Publishing Co., 17-32.

[7] Kang, Y. 2010. The emergence of phonological adaptation from phonetic adaptation: English loan-words in Korean. Phonology 27(1), 225-253.[8] Kent, D. B. 1999. Speaking in tongues: Chinglish, Japlish and Konglish. *KOTESOL Proceedings of PAC2.* Seoul, 197-209.

[8] Lee, D., Kim, D., Yun, S., Kim, S. 2021. Phonetic variation modeling and a language model adaptation for Korean English code-switching speech recognition. Applied Science 11.

[9] Kent, D. B. 1999. Speaking in tongues: Chinglish, Japlish and Konglish. KOTESOL Proceedings of PAC2. Seoul, 197-209.

[10] Kang, A. 2003. Negotiating conflict within the constraints of social hierarchies in Korean American discourse. Journal of Sociolinguistics 7(3), 299–320.

[11] Song, J. 2009. Bilingual creativity and self negotiation: Korean–English bilingual children's language socialization into address terms. In: Reyes, A., Lo, A. (eds), Beyond Yellow English: Towards Linguistic Anthropology of Asian Pacific America. New York: Oxford University Press, 213–232.

[12] Chung, S. 2010. Code-switching as a marker of cultural identity among Koreans in Toronto. LIN 497. University of Toronto: unpublished assignment.

[13] Nagy, N. 2009. Heritage language and variation change in Toronto. http://individual.utoronto.ca/ngn/research/heritage_lgs.htm#objectives

[14] Keefe, S., Padilla, A. 1987. Chicano ethnicity. Albuquerque, NM: University of New Mexico Press.

[15] Labov, W. 1984. Field methods of the project on linguistic change and variation. University of Pennsylvania.

[16] Poplack, S., Meehan, M. 1998. How languages fit together in codemixing. International Journal of Bilingualism 2, 127-138.

[17] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

[18] De Rosario-Martinez, H. (2015). phia: Post-Hoc Interaction Analysis. R package version 0.2-1. https://CRAN.R-project.org/package=phia.

[19] Kirkham, S. 2017. "Ethnicity and phonetic variation in Sheffield English liquids" in Journal of the International Phonetic Association, 47(1): 17—35