

# INITIAL SIMPLIFICATION IN HO NE: A CASE OF ETHNOGRAPHY-INFORMED VARIATIONIST STUDY IN AN ENDANGERED LANGUAGE COMMUNITY

Jie Cui

Nanjing University of Science and Technology  
jiēcui@njust.edu.cn

## ABSTRACT

This paper presents a sociophonetic study of initial cluster simplification (k(h)w-/k(h)-/) in Ho Ne language, a critically endangered Hmong-Mien language spoken in southern China that has been undergoing a dramatic restructuring in its phonological system. All community members are at least bilinguals in Hakka and Ho Ne. Based on ethnographic fieldwork, this paper draws data from speakers of four generations and of both ethnic Ho Ne and women who married into the community (n=42). The findings suggest that the onglide /w/ in /k(h)w-/ almost disappeared among the youngest generation. The predominant Hakka-speaking in-married women (who are also the primary caretakers of their children) not only are initiators of this change but also played an essential role in passing down the innovative variant to their offspring. However, among these in-married women, those who are more culturally identified with Ho Ne seem to be resisting this trend.

**Keywords:** (Peripheral) multilingualism; Variation in endangered languages; SLSCs; Ho Ne (She), China

## 1. RESEARCHING VARIATION IN SMALL LANGUAGES AND SMALL COMMUNITIES

Endangered languages are usually underdocumented, and the lack of solid baseline studies has hindered outsider researchers from studying variation in these languages. However, doing variationist study in tandem with grammar-orientated documentation in these small languages and small communities (SLSCs) [1][6][9][11][13] is a necessary task, since we might miss the best timing for conducting variationist study if we keep waiting for a solid baseline grammar. Moreover, the implementation of the variationist paradigm requires surveying more than just NORM (non-mobile, older, rural male) speakers [2]; this may help expand our sampling of the language under study to a fuller spectrum and bring our documentation closer to the linguistic reality. Meanwhile, it is also a challenging job to locate the most locally relevant social factors.

Taking Ho Ne as a case study and focusing on the initial simplification variable, this paper was set to demonstrate how an ethnography-informed variationist study could help outsider researchers to approach variation in SLSCs. The empirical ramifications of the

variationist research paradigm in SLSCs may lend sights into both contact linguistics and sociolinguistics.

## 2. BACKGROUND ON HO NE

### 2.1. The Ho Ne Communities in Guangdong

Ho Ne (ISO 639-3 shx) is a little-studied critically endangered language Hmongic [12] language spoken in rural areas of Guangdong Province in the People's Republic of China with an estimated 1,000 speakers. [5] is the first and to date the only sketch grammar of the Ho Ne language. The four counties in which Ho Ne speaker reside are Zengcheng (增城) of Guangzhou Municipality (广州市) and Boluo (博罗县), Huidong (惠东县) in Huizhou Municipality as well as Haifeng (海丰县) in Shanwei Municipality (汕尾市). In all four counties, Ho Ne speakers are a minority and in contact with Hakka, Cantonese, Hokkien, and Mandarin as well as some other varieties of local *patois*, albeit to different extents. Situated the Reform and Opening-up frontier (i.e., the Pearl Delta Economic Zone 珠江三角洲), the Ho Ne communities had been integrated to the national economy in the past 40 years, which directly led to a shift from the traditional swiddening to cash crops; many peasants were transformed to migrant workers. Meanwhile, intermarriage with Han Chinese women are now the new marital norm replacing the traditional pattern of endogamy.

### 2.2. The Linguistic Inventory of Boluo Ho Ne

Data for this paper was gathered in Boluo, Huizhou. Based on my own field research, the last generation of speakers who grew up speaking Ho Ne monolingually before their adolescence was born in the 1970s. Today, Boluo Ho Ne are almost all bilingual in Ho Ne and Hakka. Some middle-aged and young Ho Ne picked up Cantonese as a work-related language. Younger speakers in their twenties or early thirties are almost fluent Mandarin speakers. Most of the school-age Ho Ne have shifted to monolingual speakers of Mandarin or bilingual speakers of Hakka and Mandarin.

### 2.3. "It does pay to speak Ho Ne."

Endogamy was common among the older generations (speakers born before the 1970s). The youngest couple of endogamy in Boluo Ho Ne commune was in their mid-forties; in contrast, the majority of their cohorts married Hakka-speaking women from neighbouring

villages. Currently, most first (oldest) generation in-marrying women who were born in the 1930s to 1940s are both fluent speakers and active users of Ho Ne. The same could be said to the second or the grandmother-generation in-marrying women. Language choice began diversifying drastically among third generation mothers(wives). Some reported having picked up Ho Ne as soon as they married into the village and were also the ones who taught their children the Ho Ne language. Some learned enough Ho Ne to communicate or understand but were reluctant to use it anywhere. More than half of the 3rd-generation speakers never learned Ho Ne because they believed that Ho Ne is either "useless" or "too hard" to learn.

### 3. UNCOVERING THE VARIABLE

The current study focused on the simplification of the Ho Ne initial /k(h)w-/. To my best knowledge, this variable was not mentioned in any published materials on the Ho Ne language; rather, it is a variable uncovered gradually during the author's ethnographic fieldwork. I noticed earlier on during my fieldwork at the local village school that the labiovelar /-w-/ in initial clusters /kw-/ and /khw-/ was omitted in most school-age children's speech. And it is not consistent with what was documented in the sketch grammar [10] published in the 1980s.

Also worth mentioning was that this change only occurs in a restricted phonological environment: between the velar stops ([k-] or [kh-]) and rimes prefaced by low vowel [-a]. Tokens for this variable were extracted from the following syllables: [kwa], [kwan], [kwaŋ], [khwa], [khwan], and [kwaŋ].

Two questions that I explored in this study are: (1) how do different members of different social categories in the Ho Ne community produce this consonant cluster? (2) who are the leaders of this change?

### 4. DATA AND PROCEDURES

#### 4.1. The Ho Ne Corpus

Data in this study come from a larger research project on language documentation and variation in Ho Ne language (the Western dialect). The entire corpus contains recordings of over 200 hours of naturally occurring conversations and 50 hours of field elicitation and interview sessions targeting at specific phonetic features. These recordings were made between 2014-2016 during the author extensive ethnographic fieldwork in Boluo. All elicitation sessions were recorded using a Zoom4n recorder with a Shure overhead microphone at a sampling rate of 44,000 Hz. Each interview lasts from forty minutes to three hours. Tokens were extracted from naturally occurred narratives or conversations for those older speakers who suffered some degree of hearing loss. Most of the tokens were collected from picture-prompted elicitation. All

elicitation sessions were carried out by the author in Ho Ne.

#### 4.2. Data and coding

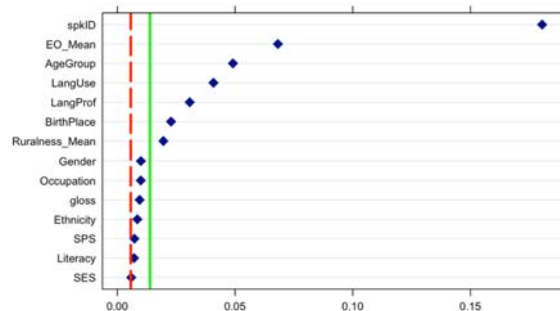
All the recordings used in analysis were made either at speakers' household or individuals' workplace. 42 speakers (22 females, 18 males) of four generations generated altogether 1080 tokens for the current study and eight of them are not native Ho Ne speakers but women who married into the village for at least fifteen years by the time of the recording. Each token containing the targeted variable was given a binary coding, indicating whether /-w-/ was deleted or present. Thirteen independent variables included were 1). Speaker ID; 2). age group (1-4); 3). gender; 4). ethnic orientation index; 5). language use (i.e., how often Ho Ne is used); 6). Ho Ne language proficiency; 7). birthplace (i.e., village in which the speaker was born and raised until adolescence); 8). ruralness index; 9). occupation; 10). ethnicity; 11). sociopolitical impact (i.e., whether one belong to a well-established clan; 12). literary(education); 13). socioeconomic stratification.

### 5. ANALYSES AND RESULTS

#### 5.1. Statistical analyses

The random forest model (mtry=5) [15] was applied to sort the relative importance of all the independent variables. The number of trees in the random forest was set to 1000 to guarantee highly stable results (Figure 1). Then the important predictors will again be used to draw a conditional inference tree [16].

Figure 1: Variable importance ranking for [-w-] deletion variable



A red dashed line was added on the plot at the absolute value of the lowest ranking predictor as well as a solid green threshold line (the median point of variable importance). Only the variables listed to the right of the green line were then included in further analysis.

#### 5.2. Results and interpretations

The two-step statistical analyses indicate that among the tested social factors, ruralness, birthplace, ethnic

orientation, age, language use, and language proficiency made into the second round of analysis. These factors were later included to make conditional inference tree (Figure 2).

### 5.1.1. Age

The youngest generation shows a stronger tendency of favouring cluster simplification [-w-] ( $p < .001$ ). Some exceptions among this fourth generation were also observed: those who were born and raised in the innermost village and are more oriented to Ho Ne culture tend to preserve the cluster.

### 5.1.2. Birthplace

The birthplace variable was coded in five levels, namely the 2k village, the 3K village, 5K village, and the Old Ho Ne habitat, and outside of the village. 5K is the innermost village where the densest social network could be found, while 2K being the outermost village. The highest percentage of endogamy is also found in 5K. Among the three older age groups (generation 1-3), birthplace was found to be relevant. The results suggest that those who were born and raised outside of the Ho Ne village tend to simplify the cluster ( $p < .001$ ).

The birthplace factor also interacts with the age factor. Those second-generation speakers who were born and raised outside of the village (i.e., in-marrying women) tend to delete [-w-]. One caveat here is about how this birthplace should be interpreted. According to the current analysis, the most significant division based on the birthplace variable was between the “outsider”

versus the “insider.” Given that all ethnic Ho Ne were born and raised inside one of the sub-villages, the “outsider” speakers in this dataset are exactly the eight in-marrying women. The detailed discussion regarding the interaction among the gender, ethnicity, and age factors will be given in section 6.

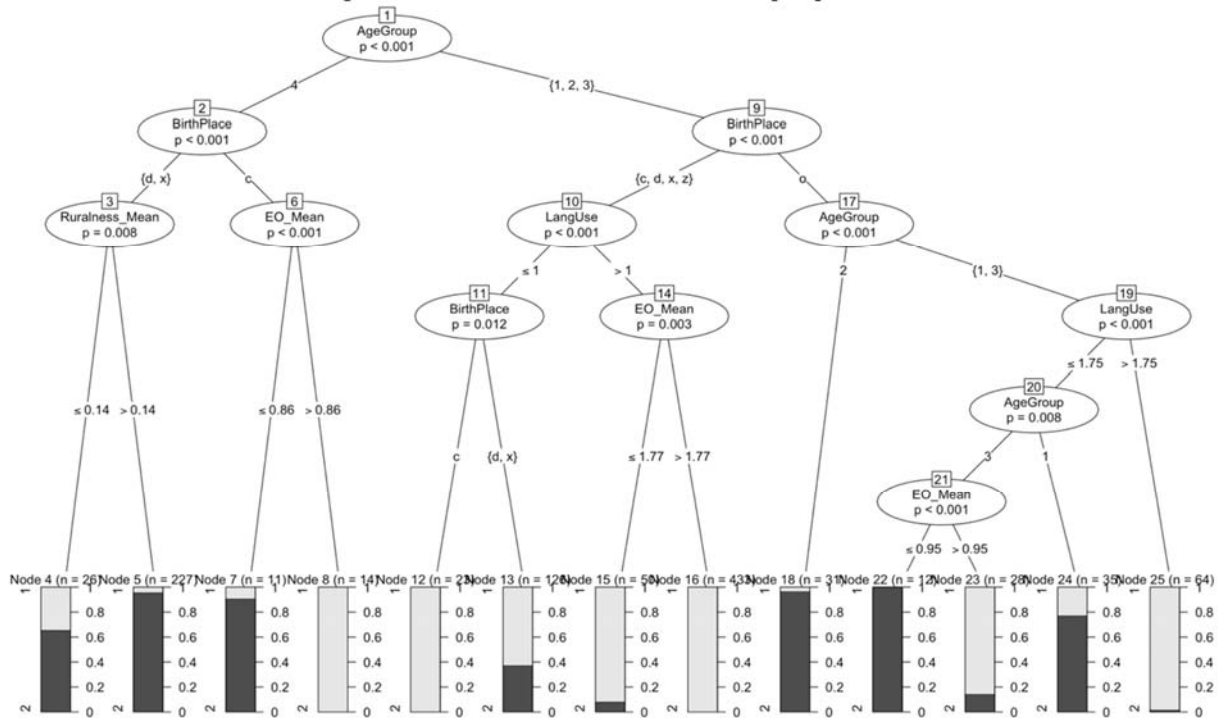
### 5.1.3. Language use

Language use variable concerns how often individuals use Ho Ne in everyday life. Among the ethnic Ho Ne, those who were not born in the innermost 5K village and used Ho Ne less frequently tend to delete [-w-]. The interaction effect between language use and age indicates that the oldest generation who use Ho Ne less frequently may also choose to delete [-w-].

### 5.1.4. Ethnic orientation

Ethnic orientation (EO) functions as a proxy for ethnic (culture) integration [6]. Individuals with a high EO index score were likely to either possess knowledge of or have actively participated in some traditional Ho Ne activities, i.e., hunting, foraging, etc.). EO factor only emerges when interacting with other factors. For example, those fourth-generation speakers who were born and raised in the innermost 5K village and obtained a higher EO index score tend to produce more consonant cluster. In the meantime, while the second-generation outsider speakers prefer the simplified variant, only speakers those who use Ho Ne less frequently and obtained a lower EO index score prefer the simplified initials among the third generation,

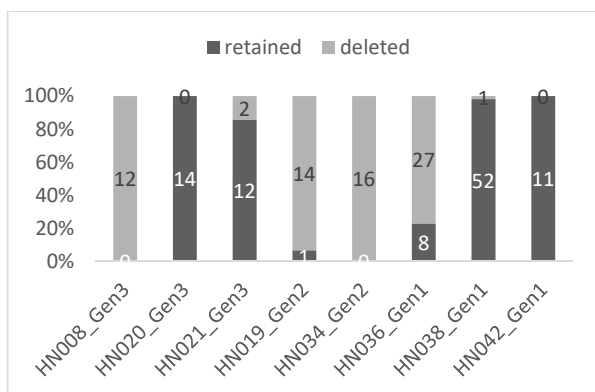
Figure 2: Conditional inference tree for [-w-] deletion



## 6. DISCUSSIONS: GENDER, LOCALE, AND ETHNICITY

The results in section 5 indicate that initial simplification is strongly preferred by two groups of speakers: the youngest generation and non-native speakers (i.e., in-marrying women). Then our question became: who initiated this change? Figure 3 shows that these eight speakers' choices are almost categorical. But not all in-marrying categorically prefer the simplified variant. How could we explain the internal variability among the in-marrying women? A closer look at the individual's personal history could better assist us in unveiling this mystery.

**Figure 3: Tokens and percentage of [-w-] retention and deletion by the eight in-marrying women**



Speaker HN008, HN019, HN034, and HN036 represent Hakka-speaking in-marrying women of all three generations. Compared to the other four in-marrying women, these four Hakka native speakers show preference of the simplified variant. What set HN020 and HN021 apart from was their language background and place of origin. Both HN020 and HN021 came from Guangxi where Hakka is not a common language. When they married into the community, Ho Ne was learned with little Hakka influence. As for HN038 and HN042, although they were married into the village, they both started speaking Ho Ne before their adolescence. HN038 moved in the current Ho Ne village when she was only nine years old. HN042 was born to a Ho Ne mother. But after her Han father diseased, her mother married back into the 5K village to a Ho Ne man. To some extent, they linguistic behaviours are similar to their Ho Ne cohorts. Therefore, it is interesting to compare patterns observed in HN038 and HN042's data with their Hakka-speaking cohort HN036—a "true" in-marrying woman and a fluent speaker of Ho Ne.

Speaker HN036 married into the village she was 24-year-old. Even after almost 50 years living in the Ho Ne village, she tends to simplify the cluster. It is also worth noting that speaker HN036 obtained an nearly perfect

score (97.5%) on the vocabulary test (performed to evaluate non-native speakers' Ho Ne proficiency).

In today's Ho Ne community, children still received most of their input of Ho Ne from their care-takers during their early childhood. In an exogamous family, if the mother acquires Ho Ne and is willing to use it as a family language, she is more likely to play a predominate role in her children's acquisition of Ho Ne. In other words, children might learn Ho Ne language primarily from their non-native Ho Ne-speaking mothers. Consequently, children may carry their mothers' accent. Therefore, if we exclude the two-non-native speaker of Hakka (HN020, HN021) and the near-native speakers of Ho Ne (HN038, HN042), those four Hakka-speaking wives cross all three generations (most Gen4 speakers are not married yet) behave consistently and tend to simplify the initial cluster. This lends evidence for us to paint the following picture: (1) the cluster simplification could be initiated by native Hakka-speaking in-marrying women; (2) in-marrying women who are fluent in Ho Ne who could also be responsible for the transmission this variant to the next generation. The fact that neither of the two local Hakka varieties allows initial cluster [4] further support that Hakka-speaking in-marrying women are leaders of the change.

## 7. DISCUSSION AND CONCLUSIONS

The current study is yet another manifestation quantitative variationist paradigm in SLSCs [1][13][14] in a contact-induced change [5] [7] [8] [17]. It demonstrates how a variationist study in SLSCs could benefit from ethnography-based fieldwork in all its research stages: from establishing a variable to sorting out the relevant social factors, and interpreting the results. The findings show that the k(h)w-k(h) variable is an innovation led by the Hakka-speaking in-marrying women who acquired Ho Ne as a second language regardless of their age groups. In contrast, in-marrying women who were not Hakka native speakers do not participate in this change. Lastly, the influence of Hakka in-marrying women magnified among the second and the third generation as they became the majority of Ho Ne men's spouses; those fluent speakers who are also active users of Ho Ne are probably could be responsible for propagating the simplified variant through passing it down to the next generation. In future studies, more in-marrying women will be recruited to verify the current findings. Language socialization study will also be carried out to examine language transmission in Ho Ne families and communities.

## 8. ACKNOWLEDGEMENT

This work is supported by the Humanity and Social Science Foundation of Ministry of Education of China (18YJC740011).

## 9. REFERENCES

- [1] Chirkova, K., J. Stanford & D. Wang. 2018 A long way from New York City: Socially stratified phonological convergence in Ganluo Ersu. *Language Variation and Change*. 20:1,
- [2] Chambers. J. K. and Trudgill P. 1998. *Dialectology*. Cambridge, UK:Cambridge Univ. Press.
- [3] Dorian, N. (eds). 2014. *Small-language fates and prospects: lessons of persistence and change from endangered languages: collected essays*. Brill.
- [4] Gan, C.Y. 2011. *Boluo Sheyu yanjiu*. Tianjin: Nankai daxue chubanshe. 209.
- [5] Gao, K. 2017. *Dynamics of language contact in China: Ethnolinguistic diversity and variation in Yunnan*. Ph.D. dissertation. University of Hawai'i
- [6] Hoffman, M. F., and J.A. Walker. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change* 22:1, 37-67.
- [7] Jamsu Reynolds, J. 2012. *Language Variation and Change in an Amdo village: Gender, education, and resistance*. Ph.D. dissertation. Georgetown University.
- [8] Li, Y.M. and W. Li. 2015. *The language situation in China*. Vol.3. Boston: Walter de Gruyter; Beijing: The Commercial Press.
- [9] Mansfield, J. and J. Stanford. Documenting sociolinguistic variation in lesser-studied indigenous communities: Challenges and practical solutions. 207. In Hildebrandt, Kristine A., Carmen Jany, and Wilson Silva. (eds), *Documenting Variation in Endangered Languages. Language Documentation & Conservation* Special Publication Honolulu: University of Hawai'i Press. 13. 116-136
- [10] Mao, Z. W. and C. J. Meng. 1986. "A Sketch Grammar of She" Beijing: Minzu Chubanshe. 1986.
- [11] Meyerhoff, M. 2017. Writing a linguistic symphony: Analyzing variation while doing language documentation." *Canadian Journal of Linguistics/Revue canadienne de linguistique*: 1-25.
- [12] Ratliff, M. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics
- [13] Stanford, J.N. 2017. A call for more diverse sources of data: Variationist approaches in non-English contexts. *Journal of Sociolinguistics* 20(4): 525-541.
- [14] Stanford, J.N. & D. R. Preston. 2009. (eds), *Variation in indigenous minority languages*. Amsterdam & Philadelphia, PA: John Benjamins Publishing Company.
- [15] Strobl, C., J. Malley, and G. Tutz. 2009. "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests." *Psychological methods* 14.4, 323.
- [16] Tagliamonte, S. A., and R. H. Baayen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language variation and change* 24.2, 135-178.
- [17] Zhou, M.L. (eds). 2012. *The contact between Putonghua and minority languages in China*. Special Issue. *International Journal of Sociology of Language*. 215.