Analogical change in progress: accent class shift in Yanbian Korean

The International Conference on Phonetics and Phonology 2013,
the National Institute for Japanese Language and Linguistics,
January 25-27, 2013

Chiyuki Ito
Tokyo University of Foreign Studies, ILCAA
chiyukit@aa.tufs.ac.jp

Introduction

- Yanbian Korean (spoken in north-eastern China; related to Hamgyung dialect in North Korea, Ramsey 1978) has a distinctive accent.
  - Monosyllabic nouns : H(L) vs. L(H)
  - Disyllabic nouns : HL(L), LH(L), LL(H)
  - Trisyllabic: HLL(L), LHL(L), LLH(L), LLL(H)
- The accent in ( ) indicates the accent of a following suffix.
- In native simplex nouns, the default accent class is final H: H, LH, and LLH.

	Antepenult	Penult	Final	Unaccented
Monosyllabic			H (5650)	L (1322)
Disyllabic		HL (2225)	LH (6587)	LL (1134)
Trisyllabic	HLL (253)	LHL (865)	LLH (4016)	LLL (297)

Introduction

- Underlying unaccented class (L, LL, LLL) appears with final accent (H, LH, LLH) in isolation forms, resulting in the neutralization of two accent classes in this environment.
- The correlation between analogical change and type-frequency is discussed in previous literature (Bybee 1995 and references cited, Albright to appear), and a generalization emerging from the literature is that higher type-frequency classes tend to attract words from lower type-frequency classes.

Introduction

- Given the default final accent class and the tonal neutralization in the isolation form, we expect that in the younger generation, more words will shift to the default accent class, in particular from the unaccented class.
- In this study, we test this hypothesis by investigating accentual patterns of monosyllabic native nouns in various native speakers from different generations, and examine how and with what factors they have changed over (apparent) time statistically.
- By doing so, we exemplify a concrete case study of an analogical (lexical) change in progress.

Introduction

- As a reference point for the conservative accent, we take into account Middle Korean (MK, 15th-16th C) accent.
- The data is composed of monosyllabic simplex native nouns, pronouns and numerals.
- The total corpus is 352 words (274 words are attested in Middle Korean).
- The result is compared with the tonal patterns of Sino-Korean words as well, showing that these two lexical classes have different tendencies.

Subjects

- 21 subjects (11 males and 10 females) who were born between 1933 and 1988, from several sub-dialects of Yanbian.

Males

Name	Year of birth	Dialect
LDW	1933	Helong
KSJ	1941	Helong
KCS	1941	Helong
CIB	1943	Yanji
LBS	1954	Yanji
HSW	1955	Yanji
JHW	1958	Longjing
KG	1969	Yanji
LSC	1969	Yanji
LSP	1981	Antu
LGW	1981	Yanji

Females

Name	Year of birth	Dialect
CSJ	1945	Helong
PCO	1949	Helong
SCJ	1955	Yanji
LOH	1956	Wangqing
HY	1973	Longjing
JYH	1975	Longjing
HMH	1978	Antu
KML	1980	Wangqing
PKL	1985	Antu
KMO	1988	Longjing

Subjects

- There are some sub-dialectal differences in accentuation within Yanbian (Kawasuzaki 2010), but the basic accent classes and general tendencies can be treated the same, as far as our subjects are concerned.

1 Yanji
2 Tumen
3 Dunhua
4 Hunchun
5 Longjin
6 Helong
7 Wangqing
8 Antu

http://en.wikipedia.org/wiki/Yanbian_Korean_Autonomous_Prefecture

Middle Korean accent

- In MK, the first high pitch is distinctive and the tonal contour after the first H is predictable depending on the number of following syllables in the same phonological phrase (Kadowaki 1976, Fukui 1985).
- The tonal patterns after the first H are notated as "X".
- R (rise) is assumed to be composed of L + H (Kōno1945).
- Numbers in ( ) indicate the number of attested words. → H is the largest class in MK.

Monosyllabic	Disyllabic	Trisyllabic
H (245)	HX	HXX
L (122)	LH	LHX
R (77)	LL	LLH
	RX	LLL
		RXX

Correspondence with Middle Korean

- Correspondences between Yanbian accent classes of monosyllabic nouns and MK.

Middle Korean	Yanbian
H	H
R	H
L	L

- MK H and R class are merged with H (final) class in Yanbian, resulting in the predominant H class.

Correspondence with Middle Korean

- The regular correspondence rates (= regularity) of MK H and R classes and Yanbian accent are much higher than that of MK L class (94%, 95% vs. 67%).
- Larger class (MK H and R classes) shows a higher regularity rate than smaller class (MK L class), which is assumed to be the result of analogical lexical change to the default H (final) class in Yanbian.

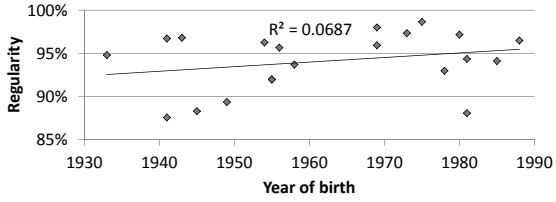
MK/Yanbian	Final	Unaccented	Totals	Regularity
H	3110	192	3302	94%
R	1066	60	1126	95%
L	363	747	1110	67%

Correspondence with Middle Korean

- The regularity in the correspondence between Yanbian accent and MK accent is affected by various factors: age, token-frequency, syllable weight, sex.
- We examine each factor separately below.

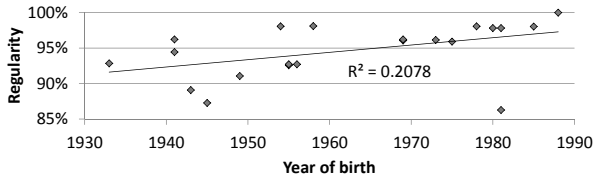
Factor 1: age

- The regularity rate of MK **H class** is quite high for all generations: 88% to 99%.
- A regression analysis using GLM (general linear model) between the regularity rate and year of birth shows no significant result (estimate = 0.0005427, std. error = 0.0004479, t-value = 1.212, p = 0.241).



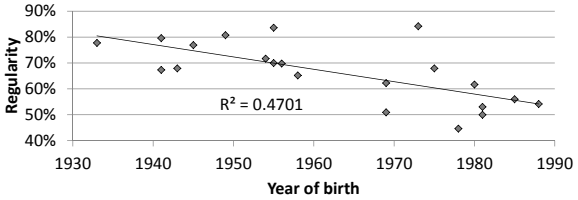
Factor 1: age

- The regularity rate of MK **R class** is also high for all generations: 86% to 100%. Still, a regression analysis using GLM between the regularity rate and year of birth shows slightly significant result (estimate = 0.0010325, std. error = 0.0004625, t-value = 2.232, p = 0.0378).
- Younger speakers tend to pronounce these words with the default final accent class more strongly.



Factor 1: age

- On the other hand, there is a strong correlation between the regularity rate of MK **L class** and year of birth.
- The younger the speakers, the lower the regularity rate of MK L class.



Factor 1: age

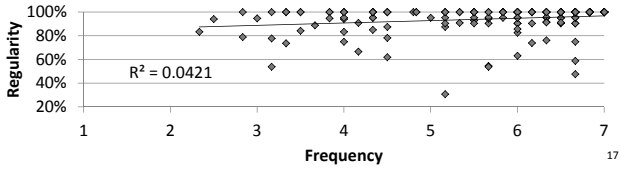
- A regression analysis using GLM between the regularity rate and year of birth shows a significant result (estimate = -0.004785, std. error = 0.001166, t-value = -4.106, p = 0.000602).
- This indicates that many words that are from the MK L class shifted to the final accent class (H) in the younger generation of Yanbian.
- The higher regularity rate in younger generation at least for MK R class can be understood as the same analogical change: some words that irregularly correspond with MK R class in old generation, probably due to some minor dialectal differences between MK and (proto-)Yanbian, are analogically changed to the default final accent class in the younger generation.

Factor 2: token-frequency

- Cross-linguistically, analogically-motivated change tends to affect low-frequency words before high-frequency words (Phillips 1984, Bybee 2001).
- It is expected that low-frequency words tend to change accent class, whereas high-frequency words tend to maintain MK accent.
- In order to confirm the effect of token-frequency, a frequency judgment by six naïve native speakers (three speakers from 1950's and three speakers from 1980's) was conducted for all the words (352) in this study.
- Speakers were asked to judge the frequency of each word from 1 to 7 (1 is least frequent and 7 is most frequent) and the scores by all speakers were averaged.

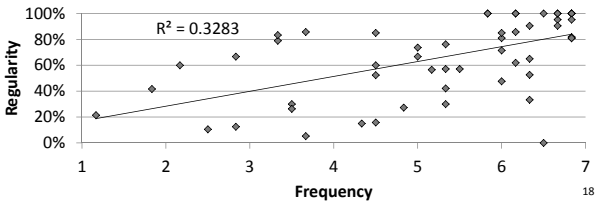
Factor 2: token-frequency

- The correlations between frequency scores and regularity rate in the words that correspond with MK **H class**.
- There is no clear correlation. Regardless of the frequency, most words appear with Yanbian H class.
- Not only high-frequency words but also low-frequency words tend to appear with the regular H class, since H is the default accent class for Yanbian monosyllabic nouns ("ceiling effect").
- The same is true for MK **R class**.



Factor 2: token-frequency

- There is a strong correlation between the regularity rate of MK **L class** and frequency.
- High-frequency words retain MK L class, whereas low-frequency words tend to change accent from more conservative L to the default H class in Yanbian.
- A regression analysis using GLM between the regularity rate and frequency scores shows a significant result (estimate = 0.11506, std. error = 0.02261, t-value = 5.090, p = 4.84e-06).



Factor 3: syllable weight

- In MK, there was a correlation between syllable structure and accent class.
- The observed (O), expected (E) and O/E values: numbers on the left indicate observed numbers, numbers on the right indicate expected numbers, and numbers in parentheses indicate O/E values.
- In both CVV and CV (in contemporary Korean, both CV due to historical monophthongization), L class is underrepresented.
- 6 examples with CVV or CV structures are pronouns.

Structures	H	R	L
CVC	158/168 (0.94)	53/53 (1.01)	93/84 (1.11)
CVV	46/35 (1.32)	12/11 (1.10)	5/17 (0.29)
CVVC	15/25 (0.60)	11/8 (1.41)	19/12 (1.54)
CV	26/18 (1.47)	1/6 (0.18)	5/9 (0.57)

Factor 3: syllable weight

- Given this, we expect that MK L class will change to Yanbian H class, in particular for light syllable words (CV).
- The regularity rate of MK L class with light syllable is only 29%, whereas that of MK L class with heavy syllable is 73%.
- For MK H and R classes, there is no difference between heavy and light syllable words.

Weight/MK accent	H	R	L
Heavy	94%	94%	73%
Light	94%	93%	29%

Factor 4: sex difference

- There are no large sex differences in the regularity rate of MK accent classes.
- Still, the regularity rate of MK L class is slightly higher for females than males.

Sex/MK accent	H	R	L
Male	94%	94%	65%
Female	94%	95%	68%

Correspondence patterns: summary

- As a summary of phonological effects, we conducted a mixed effects logistic regression model with Yanbian accent as the dependent variable, the various factors mentioned above as the independent variables, and item and subject as random factors, by using the lmer function in the lme4 package (Bates, Maechler and Bolker 2011) in R software.
- We concentrate on the words that correspond with MK L class in this analysis.
- As to year of birth, four generations are distinguished (old: 1930-1940's, old-mid: 1950's, mid: 1960-1970's, young: 1980's).
- Token-frequency is divided into high or low based on the median.

Correspondence patterns: summary

- The results show that weight and frequency are highly significant predictors.
- Age is a significant factor as well: among the four different generations, significant differences are observed between young and old, young and old-mid, but no significant difference between young and mid for this change.
- Sex difference is slightly significant.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	3.6503	0.7444	4.903	9.41e-07	***
Weight	-2.5663	0.7073	-3.628	0.000285	***
Frequency	-2.2841	0.5067	-4.508	6.54e-06	***
Age: mid	-0.6940	0.3630	-1.912	0.055905	.
Age: old-mid	-1.5583	0.3764	-4.140	3.47e-05	***
Age: old	-1.9542	0.3698	-5.285	1.26e-07	***
Sex	-0.5544	0.2630	-2.108	0.035056	*

Sino-Korean

- Sino-Korean accents regularly correspond with Middle Chinese tone.
- Tones of c. 5,000 Sino-Korean morphemes are attested (Ito 2007), but not all of them are used as a monosyllabic word.

Middle Chinese	Middle Korean
Level	L
Rising, Departing	R or H (mostly R)
Entering	H

Sino-Korean words

- Data: 398 monosyllabic words (383 words attested as either morphemes or words)
- Collected from 10 Yanbian speakers born between 1941 and 1988. (All of them participated in the investigation of native words as well.)
- Distributional patterns look similar between the two word classes: i.e. final accent class is predominant.
- Still, Pearson's Chi-squared test with Yates' continuity correction shows that there is a significant bias in this distribution ($\chi^2 = 12.4832$, df = 1, p = 0.0004106.)

	Final	Unaccented	Totals
Native	5650 (81%)	1322 (19%)	6972
Sino-Korean	2608 (84%)	497 (16%)	3105

Sino-Korean words

- As in native monosyllabic words, MK H and R classes correspond with Yanbian H (final) class, whereas MK L class corresponds with Yanbian L (unaccented) class in Sino-Korean words as a rule.
- The regular correspondence rate is much higher for MK H and R class (97% and 91%) than MK L class (35%), which is the tendency observed in native monosyllabic words as well.

MK/Yanbian	Final	Unaccented	Totals	Regularity
H	968	35	1003	97%
R	677	70	747	91%
L	658	350	1008	35%

Monosyllabic Sino-Korean words

- Still, compared with native words, the regularity of MK L class is much lower in Sino-Korean words than in native words (67% vs. 35%).

Native

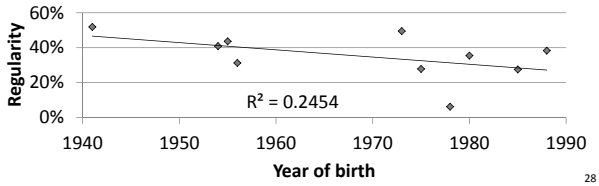
MK/Yanbian	Final	Unaccented	Totals	Regularity
H	3110	192	3302	94%
R	1066	60	1126	95%
L	363	747	1110	67%

SK

MK/Yanbian	Final	Unaccented	Totals	Regularity
H	1076	39	1115	97%
R	750	76	826	91%
L	723	395	1118	35%

Sino-Korean words

- The regularity rate for MK L class in Sino-Korean words seems to decline in younger generations, but a regression analysis using GLM between the regularity rate and year of birth shows no significant result (estimate = -0.004142, std. error = 0.002568, t-value = -1.613, p = 0.14539).
- Even for the oldest speaker, the regularity is c. 50%.



Native vs. Sino-Korean: Hypothesis 1-3

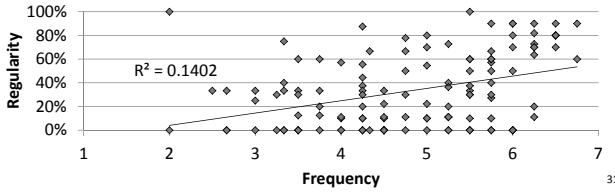
- What factors account for the different regularity rates between native and Sino-Korean words?
- Hypothesis 1: **type-frequency** of MK L class was lower in Sino-Korean than in native words, which accelerated the analogical change to Yanbian default H class more in Sino-Korean than in native words. → In actuality, L class was more frequent in SK words than in native words in MK.
- Hypothesis 2: In MK, the tonal patterns were different between Sino-Korean **morphemes** and Sino-Korean **words**. → Attested accent patterns of Sino-Korean words are more or less the same as those of Sino-Korean morphemes.
- Hypothesis 3: influences from contemporary Mandarin tone. → No statistical effect was observed.

Native vs. Sino-Korean: Hypothesis 4

- Hypothesis 4: **token-frequency** is higher for native words than for Sino-Korean words.
- We conducted the frequency judgment test for Sino-Korean words with 4 native speakers who participated in the frequency judgment test of native words as well.
- The average frequency scores: native words (5.49) vs. Sino-Korean words (4.88).
- This result seems to support the hypothesis 4.

Native vs. Sino-Korean: Hypothesis 4

- Correlation between regularity rate of MK L class and frequency scores in Yanbian Sino-Korean words.
- A regression analysis using GLM between the regularity rate and frequency shows significant result (estimate = 0.10401 std. error = 0.02277, t-value = 4.568, p = 1.14e-05).
- As in native words, high-frequency words tend to retain MK L class more than low-frequency words.



Native vs. Sino-Korean: Hypothesis 5

- Hypothesis 5: syllable shape of **native words** had correlations with accent class, which supported the preservation of the conservative accent.
- As mentioned above, the MK L class was highly underattested in CVV and CV structures.
- Whitman (1994) points out the strong correlation between the MK L class and non-sonorant finals: "Whereas rising monosyllabic nouns appear with the full range of sonorant finals /y/, /n/, /m/, /ng/, and /l/ ([r]), and Class 1.1 (tonic) nouns appear with the full range of vowels and all consonants except /ng/, Class 1.0 nouns include no native Korean stems with a final vowel, no examples with final /m/, and only one example each with final /y/ (poy 'pear') and /n/ (swon 'guest')."

Native vs. Sino-Korean : Hypothesis 5

- This tendency ("**near absence of sonorant finals**" in L class) was confirmed in the data collected by the author as well: there is a strong bias (83%) to obstruent codas (including clusters, all of which end with obstruents) and /ŋ/ in MK L class.

	Obstruent, /ŋ/	Sonorant, zero	Totals
H	94 (42%)	129 (58%)	223
R	22 (31%)	50 (69%)	72
L	90 (83%)	19 (17%)	109

Native vs. Sino-Korean : Hypothesis 5

codas	H/R	L	Totals	L (%)
c	2	1	3	33%
c^h	2	7	9	78%
k	9	6	15	40%
k*	0	1	1	100%
ks	0	2	2	100%
lk	1	2	3	67%
ls	1	0	1	0%
p	8	1	9	11%
p^h	3	3	6	50%
ps	1	0	1	0%
s	22	2	24	8%
t^h	3	7	10	70%
ŋ	4	7	11	64%
zero	63	8	71	11%
m	38	0	38	0%
n	11	1	12	8%
l	51	7	58	12%

- Correlations between codas of contemporary Korean and MK accent class.
- Since some codas were subject to sound/analogical changes from MK to contemporary Korean, and some words are obsolete in contemporary Korean, we examine the correlation between coda types and accentual class based on the coda of contemporary Korean.
- Different degree of biases are observed within obstruents as well: e.g. /-s/ rarely appears with MK L class (8%), whereas /-t^h/ often appears with MK L class (70%).

Native vs. Sino-Korean : Hypothesis 5

- Although the number of words with each coda is quite small, at least some codas show relatively higher correlations with L class, such as /c^h/, /k*/, /ks/, /t^h/, /ŋ/.
- Thus it is assumed that there are local generalizations with regard to the coda types and accent classes: certain codas imply L accent.
- These local generalizations may block the application of analogical changes from L class to the default H class in Yanbian.

Native vs. Sino-Korean : Hypothesis 5

- Sino-Korean morphemes with codas /p/, /k/, /l/, which correspond with Middle Chinese Entering tone, regularly appear with H in MK.
- Sino-Korean morphemes with other codas (/m/, /n/, /ŋ/, zero) appeared with either H/R or L class in MK, which were more or less evenly distributed, with /ŋ/ slightly biased to L.
- Thus Sino-Korean L class does not have a reliable coda generalization.

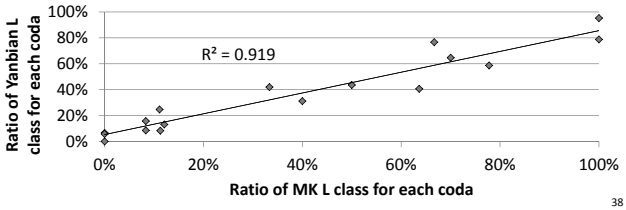
	H/R	L	Totals	L (%)
p	116	1	117	1%
k	517	0	517	0%
l	271	1	272	0%
zero	1376	1057	2433	43%
m	128	141	269	52%
n	457	380	837	45%
ŋ	322	512	834	61%

Native vs. Sino-Korean : Hypothesis 5

- The Entering tone correspondence (/-p, k, l/ → H class) does not support preserving MK **L class**.
- It is hypothesized that the lack of a general association between the syllable shape (coda types) and accent in Sino-Korean L class resulted in the lower regularity rate in Sino-Korean L class words than in native words.

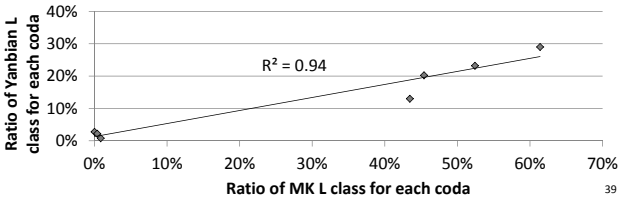
Native vs. Sino-Korean : Hypothesis 5

- In fact, the ratio of MK L class for each coda strongly correlates with the ratio of Yanbian L class for each coda (estimate = 0.80306, std. error = 0.06157, t-value = 13.042, p = 1.37e-09).
- The more L class words each coda contains, the more L class appears.



Native vs. Sino-Korean : Hypothesis 5

- The same is true for Sino-Korean words (estimate = 0.40296, std. error = 0.04552, t-value = 8.852, p = 0.000306).
- However, the ratio of Yanbian L class is in general smaller than that of MK L class, which is because the ratio of L class was originally not large enough for each coda in MK.
- Thus the less reliable correlations between these codas and L class could not prevent the analogical change to the default H.



Implications

- If this hypothesis is true, then it suggests that speakers know some general patterns in the lexicon (minimal generalization (Albright & Hayes 2003), Islands of Reliability (Albright 2002), lexical schema (Bybee 2001)).
- Speakers can learn rules/generalizations based on regular patterns that are supported by higher type-frequency in the lexicon, and apply them in deciding the accent patterns of unfamiliar words.
- Related phenomena are observed in other dialects of Korean as well (Do & Kenstowicz 2013).

Summary and conclusion

- In this study, we investigated accentual patterns of monosyllabic native nouns as well as Sino-Korean words in various native speakers from different generations of Yanbian Korean, and showed an analogical change in progress.
- Old generations (1950's or older) tend to preserve the Middle Korean L accent more regularly than younger generations (1960's or younger) in native words. On the other hand, in Sino-Korean words, even the oldest generation does not retain Middle Korean L accent regularly.
- High-frequency words tend to preserve Middle Korean L accent more faithfully than low-frequency words in both native and Sino-Korean words.

Summary and conclusion

- In general, Sino-Korean words seem to have lower token-frequency than native words.
- Sex difference is slightly significant (females are more conservative, "change from above", Labov 2001).
- The correlation between the coda types (syllable weight) and accent classes in native words may have supported preserving Middle Korean L accent.
- Speakers can learn local rules/generalizations, based on regular patterns that are supported by higher type frequency in the lexicon.

References

- Albright, Adam. (2002). Islands of reliability for regular morphology: Evidence from Italian. *Language* 78, 684-709.
- Albright, Adam. to appear. Modeling analogy as probabilistic grammar. In Juliette Blevins (ed.), *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press.
- Albright, Adam and Bruce Hayes. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90, 119-161.
- Bybee, Joan. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10, 425-455.
- Bybee, Joan. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Do, Youngah and Michael Kenstowicz. (2013). The base in Korean noun paradigms: Evidence from tone, ICPP 2013.
- Fukui, Rei. (1985). Chūki chōsengono akusento taikeini tsuite. [On Middle Korean accentual system]. *Tokyo University Linguistic Papers* '85, 61-71.
- Ito, Chiyuki. (2007). *Chōsen Kanzion Kenkyū* [Sino-Korean Phonology]. Tokyo: Kyuko-shoin.
- Kadowaki, Seiichi. (1976). Chūki chōsengoni okeru seichō kōtaini tsuite [Tonal alternation in Middle Korean]. *Chōsengakuhō* 79, 17-54.

References

- Kawasuzaki, Hideyuki. (2010). [Comparison of the accent systens of Korean spoken in China]. *Tokyo University Linguistic Papers* 29, 103-138.
- Kōno, Rokurō. (1945). *Chōsen hōgengaku shikō: 'Hasami' gokō* [A study in Korean dialectology: A study of the word for 'scissors']. Keijō (Seoul): Tōto Shoseki.
- Labov, William. (2001). *Principles of Linguistic change. Volume II: Social Factors.* Oxford: Blackwell.
- Phillips, Betty. (1984). Word frequency and the actuation of sound change. *Language* 60, 320-42.
- Ramsey, S. Robert. (1978). *Accent and morphology in Korean dialects*. Seoul: Tapchulpansa.
- Whitman, John. (1994). "The accentuation of nominal stems in Proto-Korean". In Young-Key Kim-Renaud (ed.), *Theoretical issues in Korean linguistics*, 425-439, Stanford: CSLI publications.

Acknowledgement: I sincerely thank for all of my Yanbian consultants for their cooperation of providing the data to this study. I am especially grateful to Maeran Kweon and her families for their continual and enthusiastic support in the data construction.