

UC Berkeley

Proceedings of the Annual Meeting of the Berkeley Linguistics Society

Title

Homophony and Contrast Neutralization in Southern Min Tone Sandhi Circle

Permalink

<https://escholarship.org/uc/item/1rp098rv>

Journal

Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 41(41)

ISSN

2377-1666

Author

Tsui, Tsz-Him

Publication Date

2015

DOI

10.20354/B4414110008

Peer reviewed

PROCEEDINGS OF THE FORTY-FIRST ANNUAL MEETING OF THE
BERKELEY LINGUISTICS SOCIETY

February 7-8, 2015

General Session

Special Session
Fieldwork Methodology

Editors

Anna E. Jurgensen
Hannah Sande
Spencer Lamoureux
Kenny Baclawski
Alison Zerbe

Berkeley Linguistics Society
Berkeley, CA, USA

Berkeley Linguistics Society
University of California, Berkeley
Department of Linguistics
1203 Dwinelle Hall
Berkeley, CA 94720-2650
USA

All papers copyright © 2015 by the Berkeley Linguistics Society, Inc.

All rights reserved.

ISSN: 0363-2946

LCCN: 76-640143

Contents

Acknowledgments	v
Foreword	vii
<i>The No Blur Principle Effects as an Emergent Property of Language Systems</i>	
Farrell Ackerman, Robert Malouf.	1
<i>Intensification and sociolinguistic variation: a corpus study</i>	
Andrea Beltrama	15
<i>Tagalog Sluicing Revisited</i>	
Lena Borise	31
<i>Phonological Opacity in Pendau: a Local Constraint Conjunction Analysis</i>	
Yan Chen	49
<i>Proximal Demonstratives in Predicate NPs</i>	
Ryan B. Doran, Gregory Ward	61
<i>Syntax of generic null objects revisited</i>	
Vera Dvořák.	71
<i>Non-canonical Noun Incorporation in Bzhedug Adyghe</i>	
Ksenia Ershova	99
<i>Perceptual distribution of merging phonemes</i>	
Valerie Freeman	121
<i>Second Position and “Floating” Clitics in Wakhi</i>	
Zuzanna Fuchs.	133
<i>Some causative alternations in K’iche’, and a unified syntactic derivation</i>	
John Gluckman	155
<i>The ‘Whole’ Story of Partitive Quantification</i>	
Kristen A. Greer	175
<i>A Field Method to Describe Spontaneous Motion Events in Japanese</i>	
Miyuki Ishibashi	197

<i>On the Derivation of Relative Clauses in Teotitlán del Valle Zapotec</i> Nick Kalivoda, Erik Zyman	219
<i>Gradability and Mimetic Verbs in Japanese: A Frame-Semantic Account</i> Naoki Kiyama, Kimi Akita.	245
<i>Exhaustivity, Predication and the Semantics of Movement</i> Peter Klecha, Martina Martinović.	267
<i>Reevaluating the Diphthong Mergers in Japono-Ryukyuan</i> Tyler Lau	287
<i>Pluractionality and the stative vs. eventive contrast in Ranmo</i> Jenny Lee	307
<i>Labial Harmonic Shift in Kazakh: Mapping the Pathways and Motivations for Decay</i> Adam G. McCollum	329
<i>Reference to situation content in Uyghur auxiliary ‘bolmaq’</i> Andrew McKenzie, Gülnar Eziz, Travis Major	353
<i>Case-Marking in Estonian Pseudopartitives</i> Mark Norris	371
<i>Discourse Coherence and Relativization in Korean</i> Sang-Hee Park.	397
<i>Negotiating Lexical Uncertainty and Speaker Expertise with Disjunction</i> Christopher Potts, Roger Levy	417
<i>Patterns of Misperception of Arabic Consonants</i> Chelsea Sanker.	447
<i>The Imperative Split and the Origin of Switch-Reference Markers in Nungon</i> Hannah Sarvasy	473
<i>Asymmetries in Long-Distance QR</i> Misako Tanaka.	493
<i>The cross-linguistic distribution of sign language parameters</i> Rachael Tatman	503

<i>Homophony and contrast neutralization in Southern Min tone sandhi circle</i> Tsz-Him Tsui 515
<i>Cultural Transmission of Self-Concept from Parent to Child in Chinese American Families</i> Aya Williams, Stephen Chen, Qing Zhou 533
<i>Fruits for Animals: Hunting Avoidance Speech Style in Murui</i> Katarzyna Izabela Wojtylak 545
<i>A Quest for Linguistic Authenticity: Cantonese and Putonghua in Postcolonial Hong Kong</i> Andrew D. Wong. 563

Acknowledgments

As the Executive Committee of the 41st Annual Meeting of the Berkeley Linguistics Society, we would like to express our gratitude to the conference participants, volunteers, session chairs, faculty, and staff members for their participation. It was your contributions that made the conference a success. We are especially grateful for the patience, hard work, and support of Paula Floro and Belén Flores, without whom BLS 41 would not have been possible. We would also like to thank the following departments and organizations of the University of California, Berkeley for their generous financial support:

Department of Linguistics
Student Opportunity Fund
Graduate Assembly
Department of Philosophy
Department of Psychology
Department of Ethnic Studies
French Department
Department of Spanish and Portuguese
Berkeley Language Center
Office of Academic Affairs Vice President

Foreword

This monograph contains a number of the talks given at the 41st Annual Meeting of the Berkeley Linguistics Society, held in Berkeley, California, February 7-8, 2015. The conference included a General Session and the Special Session *Fieldwork Methodology*. The 41st Annual Meeting was planned and run by the second-year graduate students of the Department of Linguistics at the University of California, Berkeley: Kenny Baclawski, Anna Jurgensen, Spencer Lamoureux, Hannah Sande, and Alison Zerbe.

The original submissions of the papers in this volume were reviewed for style by Anna Jurgensen and Hannah Sande. Resubmitted papers were edited as necessary by Anna Jurgensen and Kenny Baclawski, and then compiled into the final monograph by Anna Jurgensen. The final monograph was reviewed by Spencer Lamoureux. The endeavor was supported by Alison Zerbe's management of the Berkeley Linguistic Society's funds for publications.

The BLS 41 Executive Committee
July 2015

Homophony and contrast neutralization in Southern Min tone sandhi circle

Tsz-Him Tsui

*The Ohio State University**

1 Introduction

Southern Min Chinese has a number of varieties, such as Xiamen (or Amoy) spoken in Fujian Province in China, as well as Taiwanese in Taiwan. The tone sandhi circle in Southern Min involves five lexical tones in a tonal chain shift. The surface forms of these tones alternate between a citation/juncture form that appears at the right edge of a tone group, and a sandhi form that appears elsewhere (Myers & Tsay 2008). The length of the tone group, i.e. the domain of the sandhi, can vary from a single noun phrase to a complete sentence (Chen 1987; Lin 1994). The following Taiwanese examples are taken from Myers & Tsay (2008)¹.

(1) Taiwanese tone sandhi circle

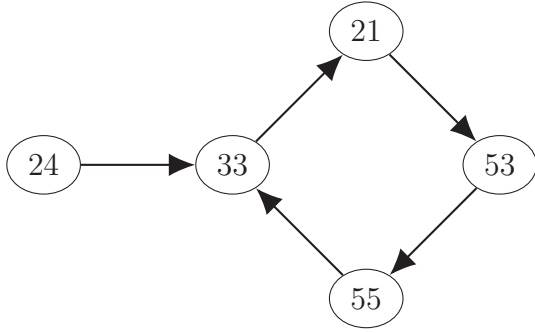
- a. $\underline{\text{si}}^{24} \rightarrow \underline{\text{si}}^{33}\text{-kan}^{55}$
“time” “time; time span”
- b. $\underline{\text{si}}^{33} \rightarrow \underline{\text{si}}^{21}\text{-tsiŋ}^{53}$
“temple” “temple monk”
- c. $\underline{\text{si}}^{21} \rightarrow \underline{\text{si}}^{53}\text{-tiam}^{53}$
“four” “four o’clock”
- d. $\underline{\text{si}}^{53} \rightarrow \underline{\text{si}}^{55}\text{-laŋ}^{24}$
“dead” “dead people”
- e. $\underline{\text{si}}^{55} \rightarrow \underline{\text{si}}^{33}\text{-bun}^{24}$
“poetry” “poetry and prose”

The schematic diagram below illustrates the tone sandhi in Southern Min in chain forms. The contrast between two tones – the rising 24 and the high-level 55 – are neutralized, as they both become the mid-level 33 in non-phrase final positions.

(2) Schematic flowchart of the Southern Min tone sandhi circle

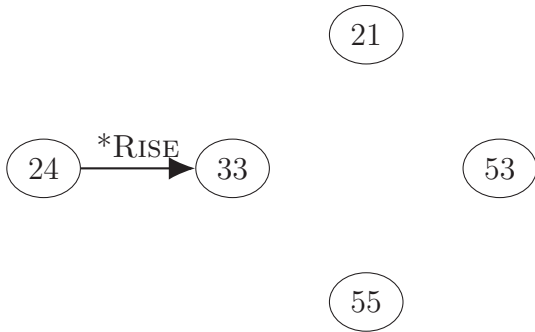
*I would like to thank Jane Tsay and James Myers for their assistance in this project, as well as comments and suggestions from the Phonics discussion group at the Ohio State University. All remaining mistakes are my own.

¹5 represents the highest tone level, and 1 represents the lowest tone level. The current paper focuses on neutralization of the lexical tones, and will not discuss the 21↔53 sandhi for syllables with plosive endings.



Recent constraint-based approaches (Hsieh 2005; Barrie 2006; Thomas 2008) assumed a need to alter the rising tone 24 in Southern Min. For example, Hsieh (2005) and Thomas (2008) suggested that it is phonetically more difficult to realize a full rising tone in a non-final syllable, as it is phonetically shorter than a final syllable. Such proposal concurred with previous phonetic and typological evidence posited by Zhang (2002). Therefore, the rising tone 24 becomes a mid-level tone 33 to satisfy the constraint *RISE.

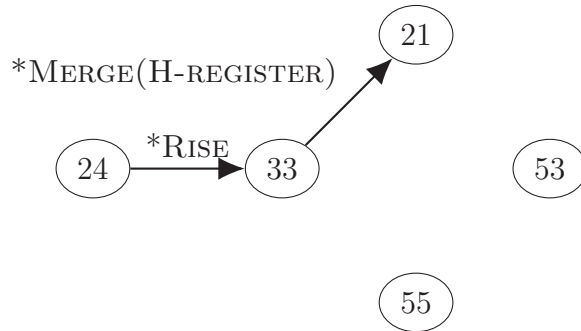
- (3) *RISE
Assign a violation to each occurrence of rising tone.
- (4) Markedness constraint changes the rising tone



Hsieh (2005), Barrie (2006) and Thomas (2008) suggested that the current neutralization of tones 24 and 33 violates the principle of contrast preservation (Lubowicz 2003), which penalizes the mapping of two different inputs to an identical output. Once again, such principle may be modeled using OT constraints. For example, we may treat the tones 55, 53 and 33 as H-register tones, versus the L-register tones 24 and 21 (Hsieh 2005; Barrie 2006). A constraint, *MERGE(H-REGISTER), would then penalize the mapping of the H-register tone 33 and another tone (i.e. the rising tone 24 in our Southern Min example) to the H-register tone 33 (Hsieh 2005). To avoid such mapping, the H-register tone 33 is therefore changed into the L-register tone 21. Such second tone sandhi would also satisfy the *MERGE(H-REGISTER) constraint by not merging into another H-register tone.

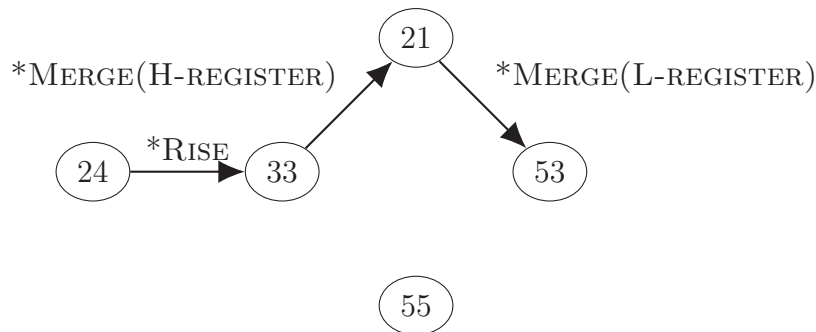
- (5) *MERGE(H-REGISTER)
Assign a violation to each mapping of a H-register tone and another tone to that particular H-register tone. (Hsieh 2005)

- (6) *MERGE(H-REGISTER) penalizes the first neutralization of tones 24 and 33, motivates the second neutralization of tones 33 and 21



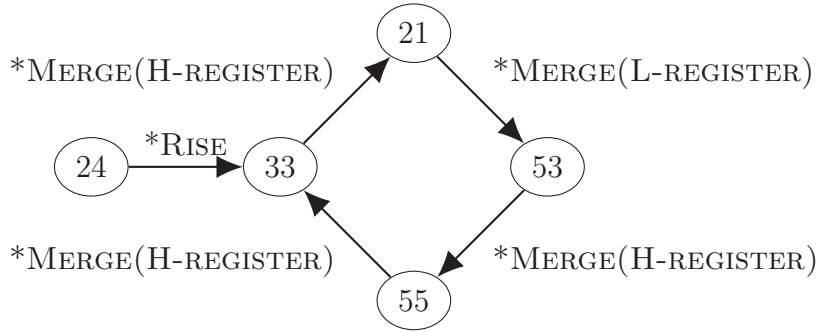
This second neutralization of tones 33 and 21, once again, violates the principle of contrast preservation. A second constraint *MERGE(L-REGISTER) would then penalize the mapping of the L-register tone 21 and another tone (i.e. tone 33) to the L-register tone 21. Such neutralization of two tones is avoided by another tone sandhi, which turns the L-register tone 21 into the H-register 53.

- (7) *MERGE(L-REGISTER)
Assign a violation to each mapping of a L-register tone and another tone to that particular L-register tone. (Hsieh 2005)
- (8) *MERGE(L-REGISTER) motivates further tone sandhi



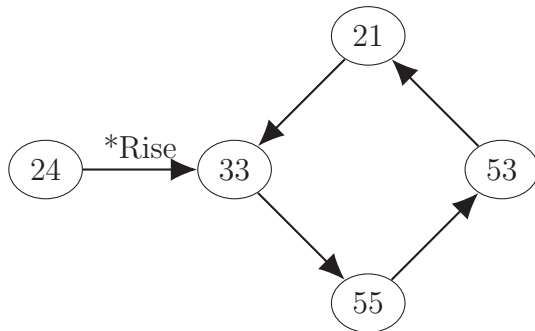
While the merger of the L-register tone 21 to 53 satisfies *MERGE(L-REGISTER), it violates *MERGE(H-REGISTER). These repeated violations of contrast preservation constraints prompt further tone sandhis, and generate the full tone sandhi circle.

- (9) Southern Min tone sandhi circle, under the constraint ranking
*RISE >> *MERGE(L-REGISTER) >> *MERGE(H-REGISTER)



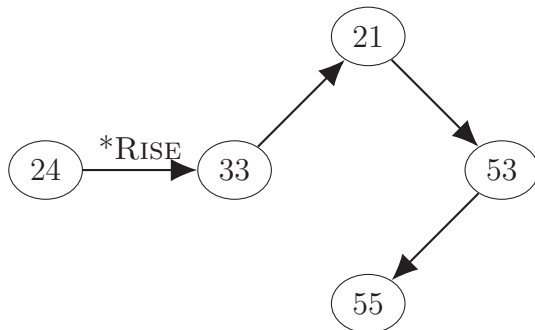
The recent accounts vary in the exact specification of the contrast preservation constraints. Hsieh (2005) utilized the **MERGE(L-REGISTER)* and **MERGE(H-REGISTER)* constraints, as specified above. Barrie (2006) suggested different constraints that aimed to preserve either register and pitch of the tones. On the other hand, Thomas (2008) proposed a single **MERGE* constraint that penalizes any kind of mergers. While the exact specification of the constraints differ, the motivation is clear: they all penalize merger of tones. Such constraint-based approaches imply that the attested sandhi circle, which merges tones 24 and 55 into tone 33, is more optimal than other possible-yet-unattested circles, such as the reversed circle in (10), where the direction of tone sandhis were opposite from the attested pattern.

(10) Possible-yet-unattested pattern: reversed tone sandhi circle



The attested tone sandhi circle should also be more optimal than other unattested non-circular chain shifts, such as the one shown in (11).

(11) Possible-yet-unattested pattern: non-circular tonal chain shift



Finally, if we assume the original motivation of the tone sandhi was to eliminate the rising tone 24, we may simply do so by having one single merger, as in the unattested pattern in (12). The constraint-based approaches, therefore, imply that it is more optimal to have the attested tone sandhi circle than not having any tonal chain shifts at all.

(12) Possible-yet-unattested pattern: no tonal chain shift



In sum, the recent constraint-based approaches suggested that the attested tone sandhi circle in Southern Min satisfies the principle of contrast preservation, and are more optimal than the unattested tone sandhi patterns. However, the reasons for preserving certain contrasts and not others are also not elaborated. For example, it is not explained why the rising tone 35 should merge into the mid-level tone 33, instead of one of the other three tones. In other words, it remains unclear *in what ways* is the attested circle more optimal than the unattested patterns, besides that the attested circle satisfies the proposed constraint rankings. These recent proposals lack independent evidence that supports the claim that contrast preservation is the motivation behind the attested Southern Min tone sandhi pattern.

The current study attempts to examine independent evidence for and/or against invoking the principle of contrast preservation in the explanation of the Southern Min tone sandhi circle. Section 2 reviews recent studies which suggested that phonological neutralizations avoid creation of homophones. Section 3 describes a series of simulations that compares the amount of homophony created by the attested and unattested tone sandhi patterns in Southern Min. Section 4 concludes the paper, and suggest that the Southern Min tone sandhi circle cannot be solely motivated by the principle of contrast preservation.

2 Contrast preservation and anti-homophony pattern

It would be very difficult for language users to communicate if their language has a lot of homophones – listeners would have to rely heavily on context to distinguish semantically-different homophones. The current section reviews recent studies which argued that phonological neutralization avoids high-functional contrasts. By preserving phonological contrasts that are often used to distinguish minimal pairs, the attested neutralization patterns create less homophony than the unattested ones, and therefore avoid creating confusion in the listeners.

Two recent studies by Silverman (2010) and Kaplan (2011) focused on the consonant neutralization patterns in Korean, and suggested that the attested neutralization patterns

create less homophony than other possible-yet-unattested patterns. For example, the coda neutralization eliminates the distinction between aspirated and plain stops. This is illustrated in (13) below. The underlying /t^h/ and /t/ codas both become [t̚] in the surface forms. Therefore, /pat^h/ ‘heritage’ and /pat/ ‘field’ share a homophonous surface form [pat̚]. The plain /k/ coda also becomes unreleased, and therefore /pak/ ‘gourd’ becomes [pak̚].

(13) Example of attested Korean coda neutralization (Silverman 2010; Kaplan 2011)

- a. /pat^h/ → [pat̚]
“heritage”
- b. /pat/ → [pat̚]
“field”
- c. /pak/ → [pak̚]
“gourd”

We may also conceive other possible coda neutralization patterns for Korean. For example, instead of neutralizing the stops by place, we may set up a ‘toy Korean’ that neutralizes the /t/ and /k/ codas. In this toy example, /pat/ ‘field’ and /pak/ would share a homophonous surface form, while /pat^h/ ‘heritage’ would have a different surface form.

(14) Toy Korean coda neutralization

- a. /pat^h/ → [pat^h]
“heritage”
- b. /pat/ → [pat̚]
“field”
- c. /pak/ → [pat̚]
“gourd”

The amount of homophony before and after the coda neutralization can be quantified in terms of 1.) number of word types, 2.) number of homophonous word tokens, and 3.) information content, in terms of Shannon’s entropy in bits (Kaplan 2011). The calculation of homophony is exemplified in the toy example (15) below.

(15) Quantifying homophony in the toy example

	Before neutralization:			After neutralization:	
Tokens	/pat ^h /	/pat/	/pak/	/pat ^h / → [pat ^h]	/pat, pak/ → [pat̚]
	1	1	4	1	1+4 = 5
	Word types = 3 (/pat ^h /, /pat/, /pak/)			Word types = 2 ([pat ^h], [pat̚])	
	Homophonous word tokens = 4 (from /pak/)			Homophonous word tokens = 5 (from [pat̚])	
	Information content = 1.252			Information content = 0.650	

Phonological neutralization decreases the number of word types, while increasing the number of homophonous word tokens. The information content of a corpus also decreases after the neutralization. In the toy example (15) above, coda neutralization decreased the number of word types from 3 to 2, while it increased the number of homophonous word tokens from 4 to 5. The information content of this tiny corpus also dropped from 1.252 to 0.650.

Using Korean corpora, Silverman (2010) and Kaplan (2011) simulated the attested consonant neutralization pattern (such as (13)), as well as other possible-yet-unattested patterns (such as (14)). They compared the amount of homophony before and after the attested and unattested patterns, and found that the attested patterns produced less homophony than the unattested ones. The attested patterns eliminated fewer word types, generated fewer homophonous word tokens, and removed less information content than the unattested patterns. A higher amount of homophony would “[make] communication more difficult”, as listeners would have to rely on context to distinguish two semantically-different homophones (Kaplan 2011:p.3). Therefore, avoiding homophony in consonant neutralization would avoid bringing extra difficulties to the listeners. In other words, it appeared that the attested consonant neutralization patterns in Korean targeted the less-functional phonological contrasts, and had less impact on the communication system than the unattested patterns. Other corpus studies have also suggested that it is rare for high-functional phonological contrasts to be eliminated in historical sound changes, as such neutralization would create confusions in the listeners (Wedel et al. 2013b,a).

As discussed in the section 1, previous studies suggested that the Southern Min tone sandhi circle is motivated by the principle of contrast preservation (Hsieh 2005; Barrie 2006; Thomas 2008). It was more optimal to preserve certain tonal contrasts over others, and thus chain shifts were formed to avoid neutralization of particular tonal contrasts. If we assume that the reason for certain phonological contrast to be preserved over others is to avoid homophony (Silverman 2010; Kaplan 2011; Wedel et al. 2013b), and that the Southern Min tone sandhi circle is motivated by contrast preservation, we should expect that the attested tone sandhi circle generates less homophony than other possible-yet-unattested patterns. The next section describes three simulations, in which we compared the amount of homophony generated by attested and unattested tone sandhi patterns.

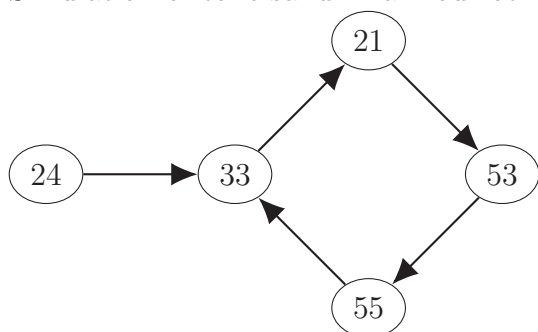
3 Homophony in Southern Min tone sandhi circle

Assuming contrast preservation is driven by homophony avoidance, we expect the attested Southern Min tone sandhi circle to produce less homophony than the unattested patterns – the attested pattern should eliminate fewer word types, generate fewer homophonous word tokens, and remove less information content than the unattested patterns. In this study, the attested and unattested tone sandhi patterns were stimulated, using a portion of the Taiwanese Spoken Corpus (Tsay & Myers 2005). The Taiwanese Spoken Corpus contains transcripts of radio talk shows in Taiwanese, and the current portion contained 78683 word tokens.

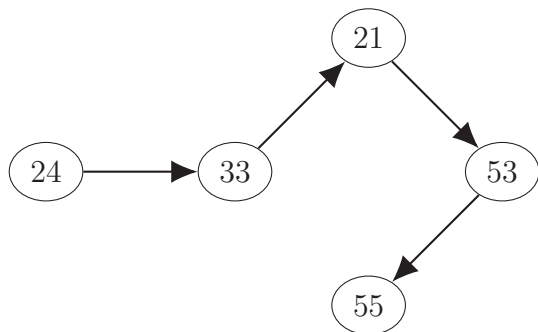
The attested and unattested tone sandhi patterns can be simulated by searching and replacing the corresponding tones in the Taiwanese Spoken Corpus. The amount of ho-

mophony can be quantified by measuring 1.) the number of word types, 2.) the number of homophonous tokens, and 3.) information content in the corpus, as shown in the toy example (15) above. Following previous studies (Hsieh 2005; Barrie 2006; Thomas 2008), the need to eliminate the rising tone 24 was assumed to be the original trigger of the tone sandhi. The attested pattern, along with 87 unattested patterns, were simulated. These include 24 circular patterns ($4 \times 3 \times 2 \times 1 \times 1$) that involve all four other tones (besides the rising tone 24, such as (16a)), 24 non-circular patterns ($4 \times 3 \times 2 \times 1$) that involve all four other tones (e.g. (16b)), 24 patterns ($4 \times 3 \times 2$) that involve three other tones (e.g. (16c)), 12 patterns (4×3) involving two other tones (e.g. (16d)), and 4 patterns that involve one other tone (e.g. (16e)).

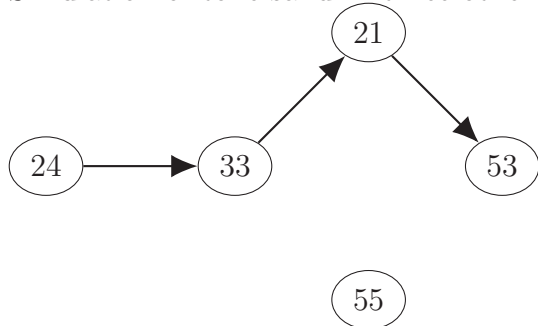
- (16) a. Simulation of tone sandhi: all four other tones (circular)



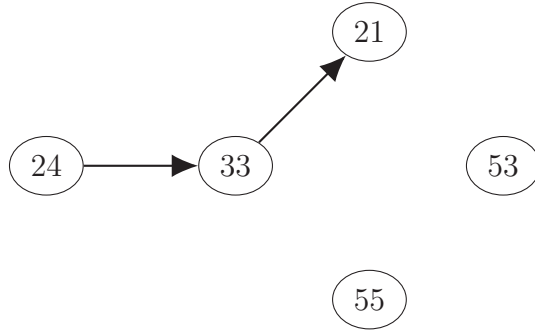
- b. Simulation of tone sandhi: all four other tones (non-circular)



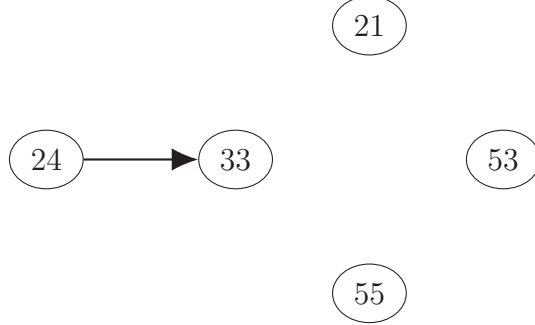
- c. Simulation of tone sandhi: three other tones



- d. Simulation of tone sandhi: two other tones



e. Simulation of tone sandhi: one other tone



One of the variables in the simulations is the domain for tone sandhi. The Southern Min tone sandhi circle is sensitive to syntax and semantics (Chen 1987; Lin 1994). Only the final tone in a tone group remains unchanged, while the tone group could be as short as a single word, and as long as a complete sentence. Therefore, the amount of tone sandhi can vary greatly when we assume different tone group sizes. Consider the following example (17), taken from the Taiwanese Spoken Corpus. When we assume individual words to be tone groups of their own, only one tone is in the non-final position, and thus subject to tone sandhi. However, when we assume the whole sentence to be one single tone group, five tones are in non-final positions, and all five of them would undergo tone sandhi.

(17) a. Original ample from the Taiwanese Spoken Corpus

lai³³bin³³ long⁵³ tsuan²⁴ si³³ po⁵³
 inside every/all all is treasure
 “Everything inside is a treasure.”

b. Tone group = word

(lai²¹bin³³) (long⁵³) (tsuan²⁴) (si³³) (po⁵³)
 inside every/all all is treasure

c. Tone group = sentence

(lai²¹bin²¹ long⁵⁵ tsuan³³ si²¹ po⁵³)
 inside every/all all is treasure

To the best of my knowledge, no current Southern Min corpora contain any phrasal / tone group information. Therefore, simulations of tone sandhi patterns were performed independently, using the smallest-possible (word) and largest-possible units (sentence) as tone groups. The small tone group (word) situation will provide a lower-bound measure of homophony after the tone sandhis, while the large tone group (sentence) situation will provide an upper-bound measure.

A second concern arises when we assume the size of tone groups to be larger than words. While tone sandhis may involve neutralization of tonal contrast and produce homophony, since lexical items do not undergo tone sandhi in the phrase-final position, two allomorphs are created for the same lexical item – one with the sandhi tone and one with the citation tone (Tsay & Myers 1996). Therefore, the amount of homophony may actually decrease in the surface, against our prediction, as new surface forms for the lexical items are created. Such phenomena is illustrated in the example (18) below. The lexical items ‘know’ /tsai⁵⁵/ and ‘scholar’ /tsai²⁴/ had different citation forms. The underlying tonal contrast is neutralized in the non-phrase-final position, where both lexical items become [tsai³³]. At the same time, however, two allomorphs are formed for each of the lexical items: ‘know’ can be [tsai³³] or [tsai⁵⁵], while ‘scholar’ can be [tsai³³] or [tsai²⁴]. These allomorphs decrease the amount of homophony by increasing the number of word types and decreasing the number of homophonous word tokens in the surface. As a result, when we assume the tone groups to be sentences in our simulation of tone sandhi, we need to evaluate the amount of homophony and consider the effect of allomorphy at the same time.

- (18) Homophony and allophony in Southern Min tone sandhi
- a. (...tsai⁵⁵→³³... ...tsai⁵⁵)
 know know
 - b. (...tsai²⁴→³³... ...tsai²⁴)
 scholar scholar

There are at least two solutions for quantifying homophony in light of allomorphy. The first is to ignore the final word in the sentences, and focus on the amount of homophony generated by the tone sandhi. Such method, while leaving out the final word, would allow us to more accurately estimate the amount of homophony generated by the tone sandhi process. The amount of homophony generated by the attested and unattested tone sandhi patterns can be compared. Once again, our prediction would be that the attested tone sandhi pattern produces less homophony than the unattested ones, and thus create less impact on the communication system.

We can also include the sentence-final words in our calculation of homophony. Once again, the impact of homophony *and* allomorphy together on the communication system may be quantified by measuring the changes to the number of word types, number of homophonous word tokens, and information content in the corpus. Due to allomorphy, we expect that the number of word types to increase, and the number of homophonous word tokens to decrease after tone sandhis. The information content should increase as well. In the discussion of homophony in section (2), we assumed that neutralization patterns that create the least amount of homophony are more optimal because they produce the least amount of impact to the communication system. Similarly, in the current situation where both homophony and

allomorphy appear, we may assume that tone sandhi patterns that produce the least amount of impact to the communication system to be more optimal. The more allomorphs there are in the surface, the more surface forms the language users would have to learn. Therefore, when tone sandhi in sentences causes both homophony and allomorphy, the attested tone sandhi pattern should generate fewer new word forms and remove fewer homophonous word tokens than other unattested patterns if the attested pattern is indeed an optimal pattern.

3.1 Simulation 1: Tone sandhi in words

This section describes the results from the simulations of attested and unattested tone sandhi patterns in the Taiwanese Spoken Corpus, using individual words as tone groups. This is a lower bound measure of the amount of homophony generated by the tone sandhi patterns. Assuming that the attested sandhi pattern is more optimal under the principle of contrast preservation, we predict that the attested pattern should generate less homophony than the unattested patterns: the attested pattern should 1.) remove fewer word types, 2.) generate fewer homophonous word tokens, and 3.) decrease less information content than the unattested patterns.

Figures 1, 2, and 3 below show, respectively, propotional change in number of word types, number of homophonous word tokens, and information content after the attested and unattested tone sandhis. The attested tone sandhi circle removed more word types than 10 other unattested sandhi patterns, and increased the number of homophonous word tokens more than 45 other unattested patterns. The attested tone sandhi circle also removed more information content in the corpus than 26 other unattested patterns.

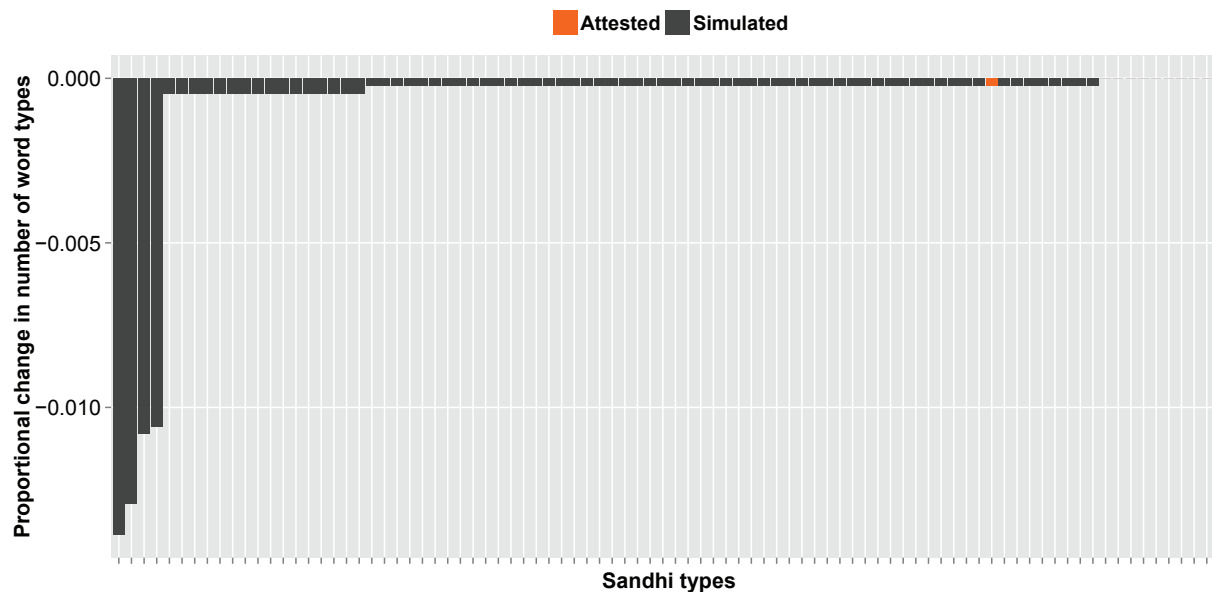


Figure 1: Proportional change in number of word types after attested and unattested tone sandhi, assuming individual words to be tone groups.

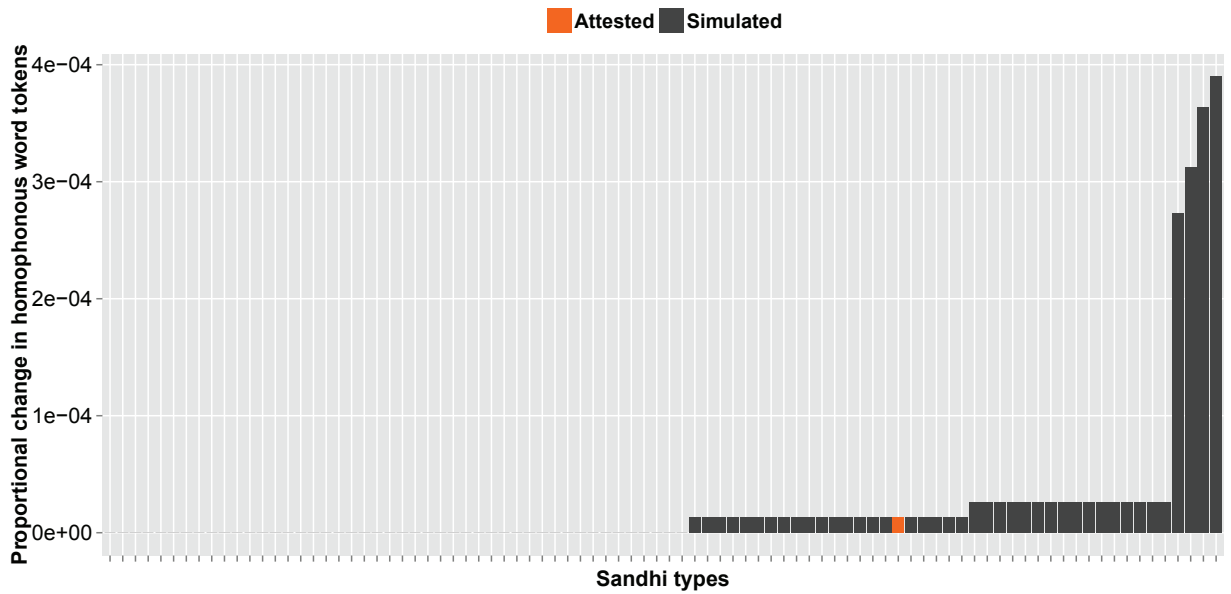


Figure 2: Proportional change in number of homophonous word tokens after attested and unattested tone sandhi, assuming individual words to be tone groups.

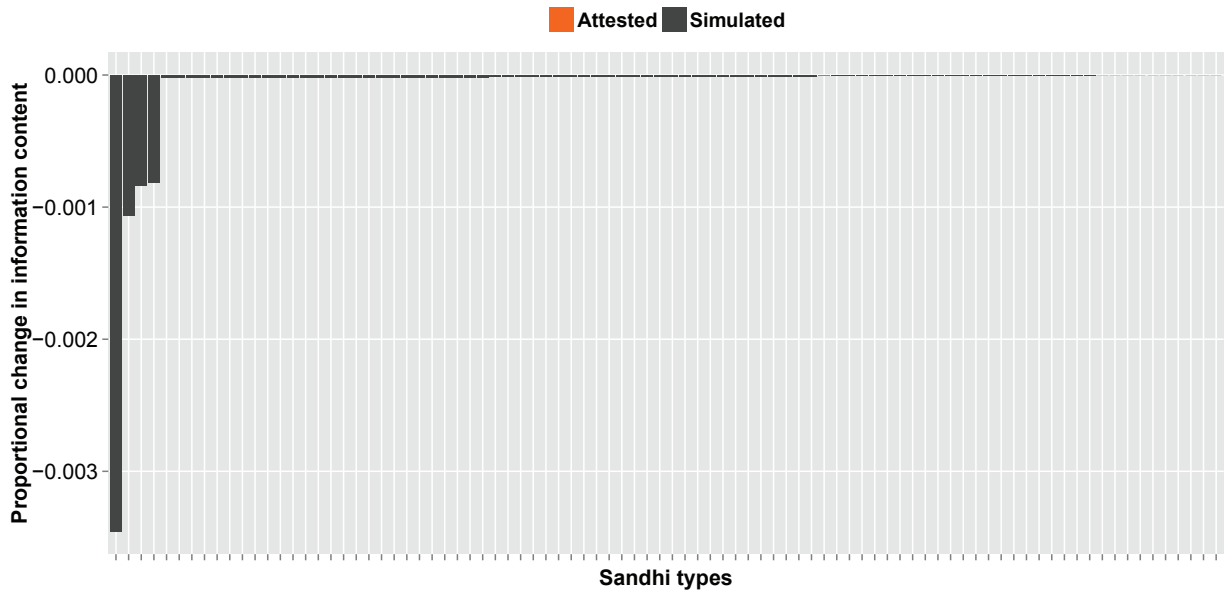


Figure 3: Proportional change in information content after attested and unattested tone sandhi, assuming individual words to be tone groups.

Based on the simulations of tone sandhis above, which used individual words as tone groups, it is unclear whether the attested tone sandhi circle in Southern Min is indeed the optimal pattern. Should we assume that contrast preservation is driven by homophony avoidance, the attested pattern is creating more homophony than at least 10 other possible

sandhi patterns. In other words, based on the simulations above, the attested tone sandhi circle appears to be less optimal than at least 10 other possible patterns.

3.2 Simulation 2: Tone sandhi in sentences, without final words

Attested and unattested tone sandhi patterns were also simulated in the Taiwanese Spoken Corpus, using sentences as tone groups. This session describes the simulation results without including the sentence-final words. By excluding the final words, we can exclude allomorphy in the lexical items, and produce an upper bound measure of the amount of homophony generated by the tone sandhi patterns.

Once again, assuming that the attested sandhi pattern is more optimal under the principle of contrast preservation, we predict that the attested pattern should generate less homophony than the unattested patterns: the attested pattern should 1.) remove fewer word types, 2.) generate fewer homophonous word tokens, and 3.) decrease less information content than the unattested patterns.

Figures 4, 5, and 6 below show, respectively, proportional change in number of word types, number of homophonous word tokens, and information content after the attested and unattested tone sandhis. The attested tone sandhi circle removed more word types than 14 other unattested sandhi patterns, and increased the number homophonous word tokens more than 7 other unattested patterns. The attested tone sandhi circle also removed more information content in the corpus than 14 other unattested patterns.

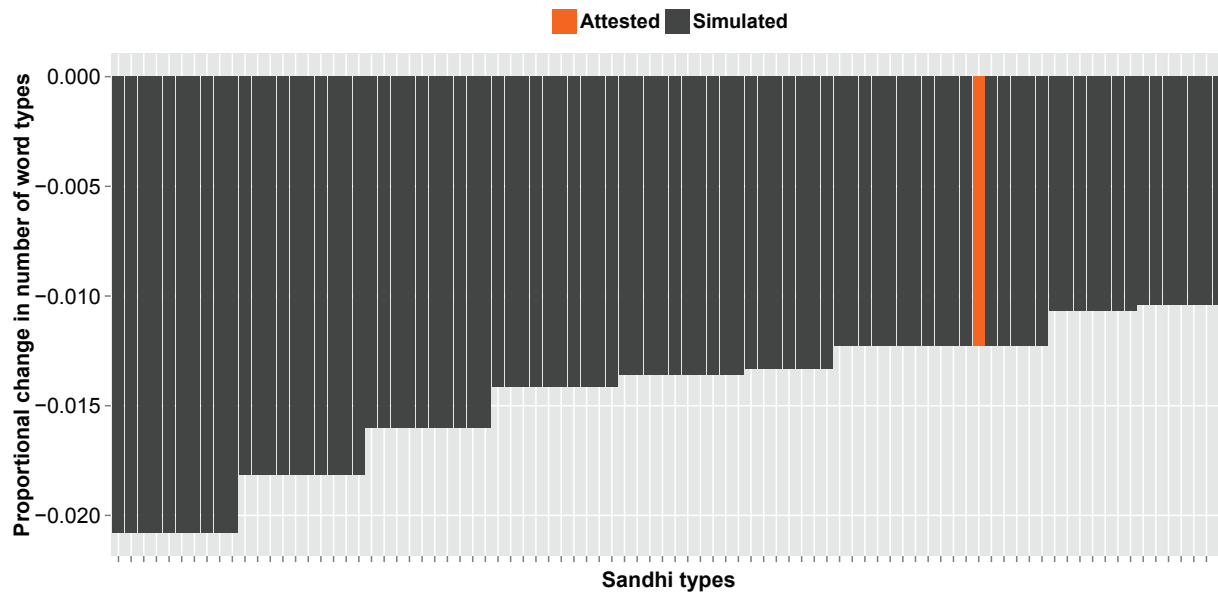


Figure 4: Proportional change in number of word types after attested and unattested tone sandhi, assuming sentences to be tone groups. Final words were excluded.

Based on the simulations of tone sandhis above, that used sentences as tone groups, it is again unclear whether the attested tone sandhi circle in Southern Min is indeed the optimal pattern. Should we assume that contrast preservation is driven by homophony avoidance, the

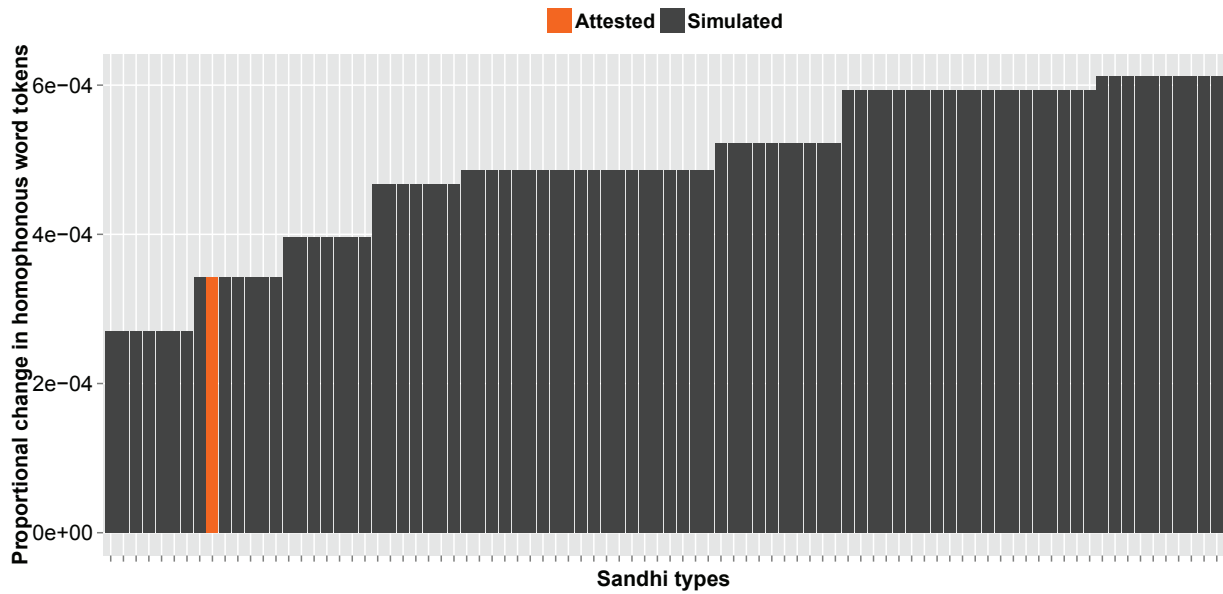


Figure 5: Proportional change in number of homophonous word tokens after attested and unattested tone sandhi, assuming sentences to be tone groups. Final words were excluded.

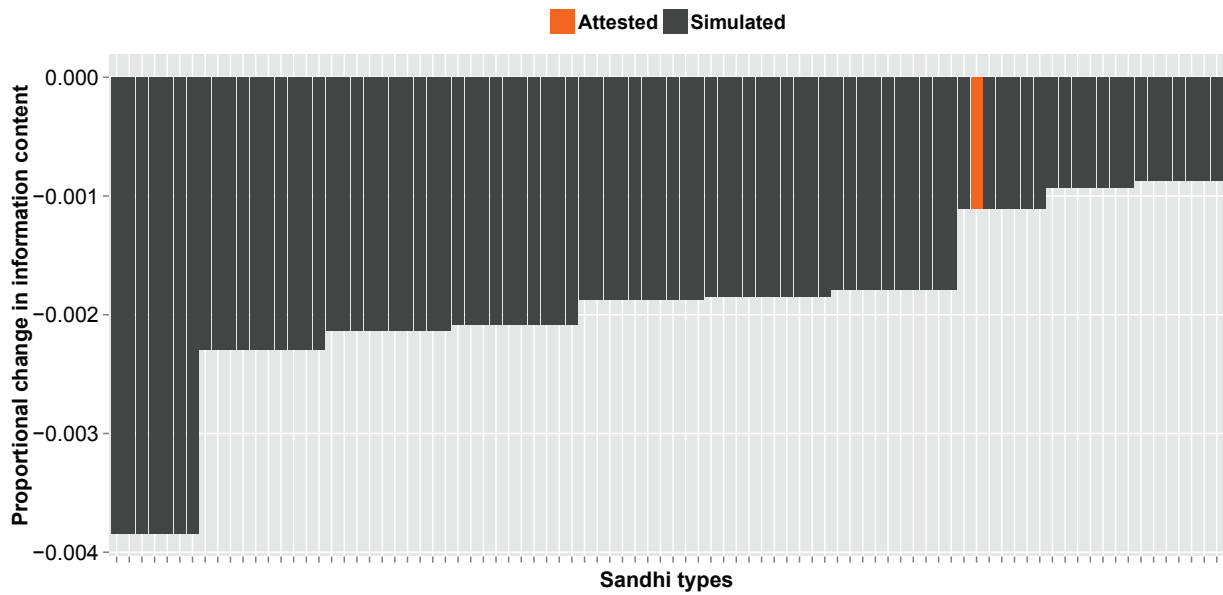


Figure 6: Proportional change in information content after attested and unattested tone sandhi, assuming sentences to be tone groups. Final words were excluded.

attested pattern is creating more homophony than at least 7 other possible sandhi patterns. In other words, based on the simulations above, the attested tone sandhi circle appears to be less optimal than at least 7 other possible patterns.

3.3 Simulation 3: Tone sandhi in sentences, including final words

Finally, the attested and unattested tone sandhi patterns were simulated, using sentences as tone groups. This time, the quantification of homophony includes the sentence-final words. Allomorphy in the lexical items is thus included. The assumption here is that the attested sandhi pattern is more optimal if it produces less impact on the communication system than other unattested patterns. Therefore, we predict that the attested pattern should generate less changes to the homophony in Taiwanese than the unattested patterns: the attested pattern should 1.) produce fewer word types, 2.) remove fewer homophonous word tokens, and 3.) increase less information content in the corpus than the unattested patterns.

Figures 7, 8, and 9 below show, respectively, proportional change in number of word types, number of homophonous word tokens, and information content after the attested and unattested tone sandhis. The attested tone sandhi circle produced more new word types than 78 other unattested sandhi patterns, and removed more homophonous word tokens than 82 other unattested patterns. The attested tone sandhi circle also removed more information content in the corpus than 75 other unattested patterns.

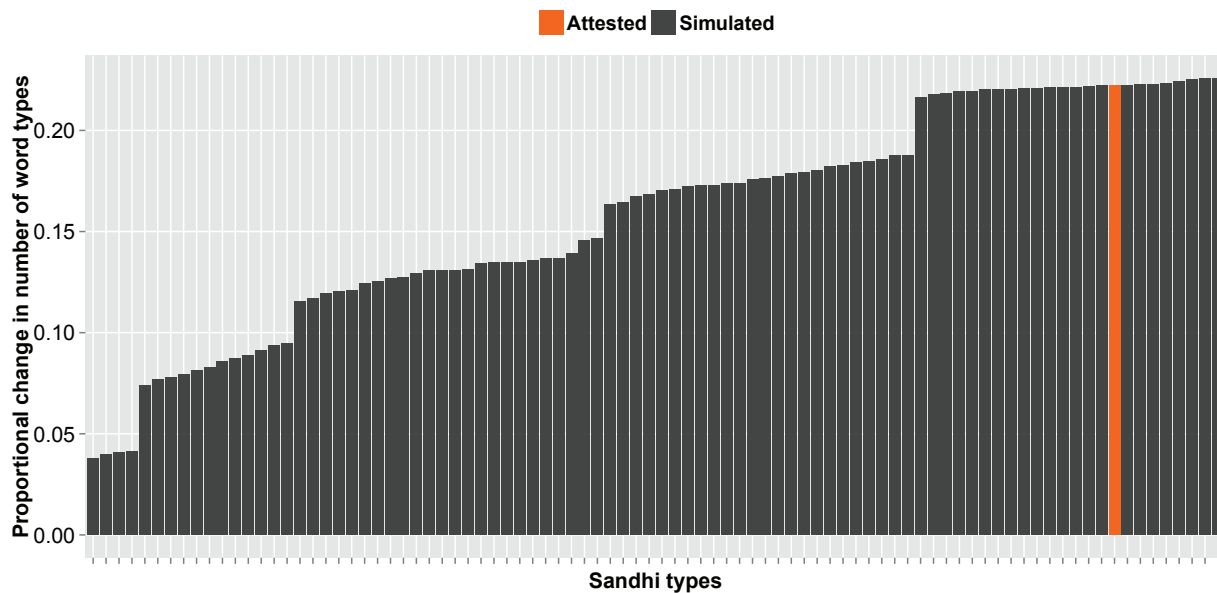


Figure 7: Proportional change in number of word types after attested and unattested tone sandhi, assuming sentences to be tone groups.

Based on the simulations of tone sandhis above, it is yet again unclear whether the attested tone sandhi circle in Southern Min is indeed the optimal pattern. When we include allomorphy in the simulations here, we expected the attested sandhi would bring in the least impact to the communication system. In other words, we expected the attested sandhi to create fewer word types and remove fewer homophonous word tokens than the unattested sandhi patterns. We also expected the attested sandhi to increase less information content than the unattested patterns. However, the attested pattern is generating a larger impact to the communication system than at least 75 of other possible patterns. Therefore, if we

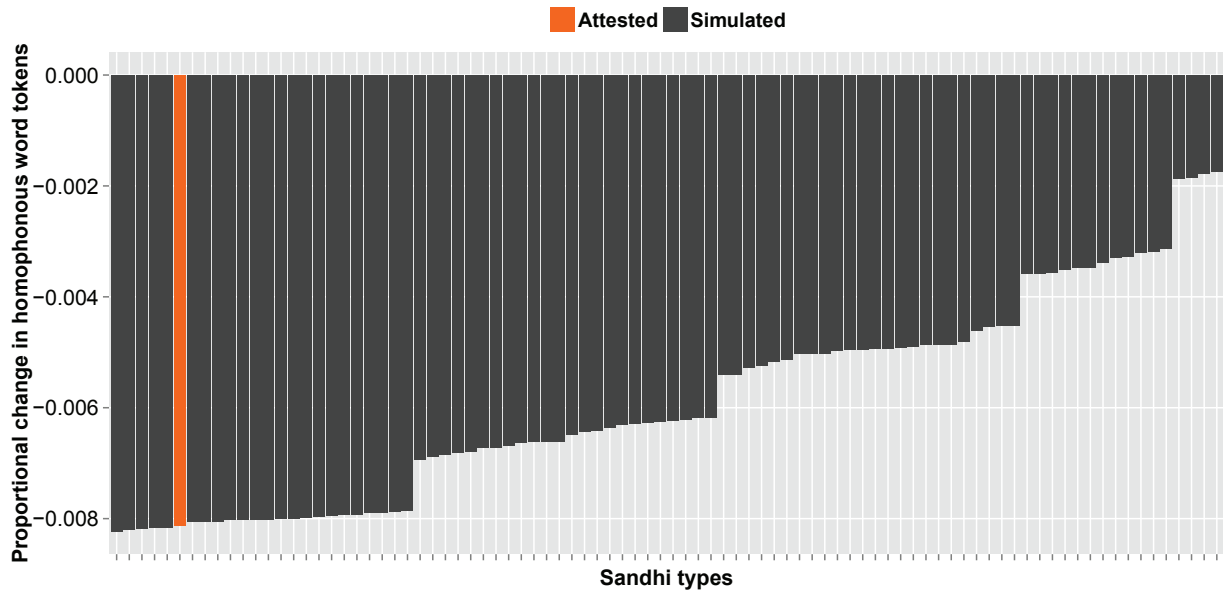


Figure 8: Proportional change in number of homophonous word tokens after attested and unattested tone sandhi, assuming sentences to be tone groups.

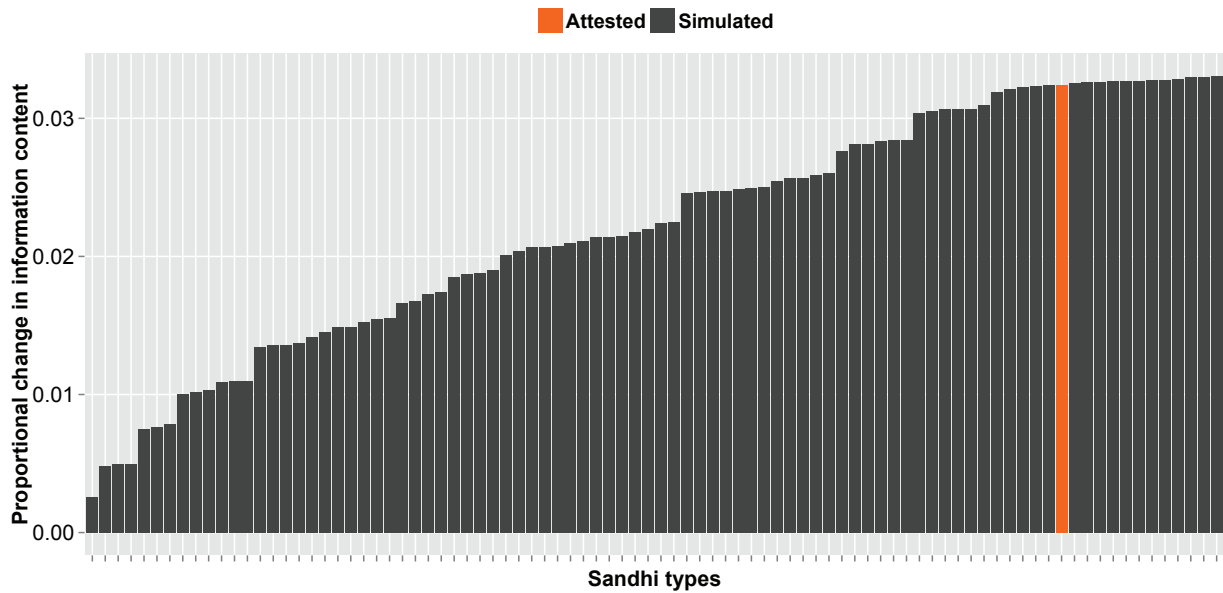


Figure 9: Proportional change in information content after attested and unattested tone sandhi, assuming sentences to be tone groups.

include allomorphy in our quantification here, the attested tone sandhi circle appears to be less optimal than at least 75 other possible patterns.

4 Discussion

Based on our results from the simulations in sections 3.1 and 3.2, it is unclear whether the tone sandhi circle in Southern Min is motivated by the principle of contrast preservation. Assuming that phonological contrasts are preserved to avoid homophony, a neutralization pattern is optimal if it generates less homophony than other possible patterns. However, the attested tone sandhi circle is generating more homophony than at least 7 other plausible patterns. Therefore, the Southern Min tone sandhi circle cannot be motivated by the principle of contrast preservation alone.

Results from 3.3 also cast doubts on whether the tone sandhi circle can be solely motivated by the principle of contrast preservation. For many lexical items, the tone sandhi circle generates two allomorphs: the sandhi version in non-phrase final positions, and the other version in phrase final positions. The conjecture here is that the tone sandhi pattern should create less impact on the communication system than other possible-yet-unattested patterns, if the attested sandhi pattern is indeed an optimal one. However, the attested pattern generates more surface word forms than at least 75 other unattested patterns. If our conjecture about homophony and allomorphy is false, and it is in fact optimal for the tone sandhi pattern to generate *more* surface forms instead of fewer, the attested sandhi pattern is still not the optimal pattern. This is because the attested sandhi pattern is removing fewer homophonous word tokens than 5 other patterns. It is, therefore, difficult to support the claim that the attested tone sandhi pattern in Southern Min is driven by contrast preservation.

In conclusion, while previous studies (e.g. Hsieh 2005; Barrie 2006 and Thomas 2008) suggested that the Southern Min tone sandhi circle is motivated by contrast preservation, the current study failed to find any supporting evidence. We assume that phonological neutralization should target non-functional contrasts and produce less homophony than other possible patterns (Silverman 2010; Kaplan 2011). Using a Taiwanese corpus, the attested tone sandhi pattern was simulated along with 87 possible-yet-unattested patterns, and the attested sandhi pattern produces more homophony than at least 7 other unattested patterns. Therefore, it appears that the Southern Min tone sandhi circle cannot be motivated by the principle of contrast preservation only.

The current study quantified the amount of homophony before and after tone sandhi in Southern Min, and examined whether this particular neutralization phenomenon generates less homophony than other possible-yet-unattested patterns of neutralization. This methodology can be certainly be extended to other neutralization phenomena in other languages, whenever a sizable corpus is available. A cross-linguistic corpus study can provide further insights to the relationship between attested versus unattested neutralization in terms of the amount of homophony that they generate (cf. Wedel et al. 2013a,b).

5 References

- BARRIE, MICHAEL. 2006. Tone Circles and Contrast Preservation. *Linguistic Inquiry* 37.131–141.
- CHEN, MATTHEW Y. 1987. The syntax of Xiamen tone sandhi. *Phonology* 4.109.

- HSIEH, FENG-FAN. 2005. Tonal Chain-shifts as Anti-neutralization-induced Tone Sandhi. *University of Pennsylvania Working Papers in Linguistics* 11.99–112.
- KAPLAN, ABBY. 2011. How much homophony is normal? *Journal of Linguistics* 47.631–671.
- LIN, JO-WANG. 1994. Lexical government and tone group formation in Xiamen Chinese. *Phonology* 11.237.
- LUBOWICZ, ANNA. 2003. *Contrast preservation in phonological mappings*. University of Massachusetts Amherst PhD dissertation.
- MYERS, JAMES, and JANE TSAY. 2008. Neutralization in Taiwanese Southern Min tone sandhi. *Interfaces in Chinese Phonology: Festschrift in Honor of Matthew Y. Chen on His 70th Birthday*, ed. by Yuchau E. Hsiao, Hui-chuan Hsu, Lian-Hee Wee, and Dah-an Ho, no. W8 in Language and Linguistics Monograph Series, 47–78. Taipei, Taiwan: Institute of Linguistics, Academia Sinica.
- SILVERMAN, DANIEL. 2010. Neutralization and anti-homophony in Korean. *Journal of Linguistics* 46.453–482.
- THOMAS, GUILLAUME. 2008. An analysis of the xiamen tone circle. *Proceeding of the 27th West Coast conference on Formal Linguistics*, 422–430.
- TSAY, JANE, and JAMES MYERS. 1996. Taiwanese tone sandhi as allomorph selection. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 22.394–405.
- TSAY, JANE, and JAMES MYERS. 2005. *Taiwanese spoken corpus*. Taiwan: National Chung Cheng University.
- WEDEL, ANDREW; SCOTT JACKSON; and ABBY KAPLAN. 2013a. Functional Load and the Lexicon: Evidence that Syntactic Category and Frequency Relationships in Minimal Lemma Pairs Predict the Loss of Phoneme contrasts in Language Change. *Language and Speech* 56.395–417.
- WEDEL, ANDREW; ABBY KAPLAN; and SCOTT JACKSON. 2013b. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128.179–186.
- ZHANG, JIE. 2002. *The effects of duration and sonority on contour tone distribution*. Outstanding Dissertations in Linguistics. New York: Routledge.