Frequency, Gradience, and Variation in Consonant Insertion

A Dissertation Presented

by

Young-ran An

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Linguistics

Stony Brook University

August 2010

Stony Brook University

The Graduate School

Young-ran An

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Ellen I. Broselow – Dissertation Advisor
Professor, Department of Linguistics

Marie K. Huffman – Chairperson of Defense
Associate Professor, Department of Linguistics

Robert D. Hoberman
Professor, Department of Linguistics

Andries W. Coetzee
Assistant Professor, Department of Linguistics
University of Michigan

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Frequency, Gradience, and Variation in Consonant Insertion

by

Young-ran An

Doctor of Philosophy

in

Linguistics

Stony Brook University

2010

This dissertation addresses the extent to which linguistic behavior can be described in terms of the projection of patterns from existing lexical items, through an investigation of Korean reduplication. Korean has a productive pattern of reduplication in which a consonant is inserted in a vowel-initial base, illustrated by forms such as *alok-**t**alok* 'mottled,' *otoŋ-**p**otoŋ* 'chubby.' A wide range of consonants may be inserted, with variation both within and across speakers. Based on study of a Korean corpus as well as experiments in which native speakers formed reduplicated versions of nonce words, I argue that the choice of inserted consonants is affected by a complex set of factors, including syllable contact constraints, preference for particular consonant-vowel sequences, and tendency for inserted consonants to be distinct in place of articulation from neighboring consonants.

The analysis in this dissertation shows that there is neither a single preferred consonant nor a random choice among all possible consonants. This phenomenon appears to contradict claims in previous literature concerning the

identity of consonants inserted in reduplication. Contrary to the claim of Alderete *et al*. (1999) that segments in the reduplicant that are not present in the base represent an emergence of the unmarked, the inserted consonant (CI) in Korean reduplication cannot be an unmarked/default consonant because distinct consonants can be inserted in the identical environments, e.g. *alok-**t**alok* 'mottled,' *ulak-**p**ulak* 'wild' where /t/ and /p/ are epenthesized although the bases contain the same set of consonants, /l/ and /k/. Moreover, a particular vowel does not force the occurrence of a particular consonant, e.g. *ulak-**p**ulak* 'wild,' *umuk-**ʧ**umuk* 'unevenly hollowed,' *upul-**k'**upul* 'windingly' in which different CIs are followed by the same vowel /u/.

Examination of the lexical patterns suggests that lexical frequency plays a role in the choice of inserted consonant. First, the frequency of CIs in a word creation experiment correlated significantly with the frequency of word-initial Cs in the Korean corpus. Second, the frequency of consonant combinations CI – C_1 in forms of the shape **CI**V.C_1VC_2 correlated significantly with the frequency of combinations of consonants in CVCV forms in the corpus. Similarly, the frequency of combinations of CI – C_2 in forms of the shape **CI**V.C_1VC_2 correlated with the frequency of combinations of onset C – coda C in the corpus. Third, the frequency of C – V combinations in the experiment correlated significantly with the frequency of lexical C – V combinations in the corpus.

Another factor investigated was the effect of a restriction on syllable contact banning heterosyllabic sequences in which a coda C of a preceding syllable is of lower sonority than a directly following onset C. This restriction has been shown to play a role in Korean phonology, and is potentially relevant to choice of inserted consonant in reduplicants of the form VCVC-**CI**VCVC. This constraint was found to work more strongly for nonce reduplicated words than for the general vocabulary.

The role of the following V on the choice of inserted C was also investigated. Korean speakers' behavior in many psycholinguistic experiments suggested that a CV (body) constituent is prominent for Korean speakers, as opposed to the speakers of English-like languages which evidently have a closer tie between V and C (rhyme).

An additional factor that appeared to affect the choice of CI was identity avoidance. The general vocabulary of Korean was argued to respect an OCP-Place constraint (identity avoidance in place), which does not allow consonants with the same place to co-occur. The dictionary data and the experimental responses also showed significant effects of identity avoidance in place, based on the ratio of observed to expected occurrences of inserted consonants in different contexts. Data from the general lexicon and the reduplication data also revealed a distance effect: co-occurrence restrictions appeared to be stricter for adjacent consonant pairs than for non-adjacent consonant pairs.

Lexical frequency was shown to play a role in the choice of inserted consonants, to some extent; however, individual speakers did not necessarily reflect the lexical patterns. There were two distinct patterns among the speakers with regard to the choice of CI: those who preferred /t/ predominantly over other Cs and those who preferred /ʧ/ predominantly over other Cs. Moreover, within a group of the speakers who chose /t/ most frequently there were some speakers who chose less preferred CIs when the context contained their preferred CI, whereas other speakers stayed with the preferred CI regardless of context.

For my parents,

Gilyoung An and Soonok Lee

TABLE OF CONTENTS

List of Tables

List of Figures

ACKNOWLEDGMENTS

Expressing gratitude enough may not exist especially when it comes to thanking the people who have been involved in the process of writing a dissertation. My advisor Ellen Broselow, who has also been my mentor, has been a blessing to my life, as well as to my career. Ellen has guided me throughout the course of my study at Stony Brook, as well as at every step of my dissertation. I appreciate her thoughtfulness, enthusiasm, and great insights. I would also like to thank the other members of my committee for their feedback and encouragement. Marie Huffman kept me awake with refreshing ideas from a phonetician's perspective. Bob Hoberman called my attention to a morphologist's views. My cordial thanks also go to Andries Coetzee, who gladly agreed to join the committee and continued to encourage me with inspiration while I was writing.

I owe cordial thanks to the faculty members in the Department of Linguistics at Stony Brook. My heartful thank you goes to Richard Larson, who guided me throughout my coursework and writing with wisdom and care. I have enjoyed learning from the professors, John Bailyn, Christina Bethin, Dan Finer, Alice Harris, and Lori Repetti. A special thank you goes to Joy Janzen, who has supported me as a course supervisor and as a friend.

The Linguistics community at Stony Brook has been an invaluable source of my life and research. My deepest thanks go to every one of my colleagues: I dare not to name every one of them, but I will remember their loving hearts and wonderful minds. I have been indebted to Chih-hsiang Shu for his everlasting help in every way. I have enjoyed talking to him about my research and life. My delightful thank you goes to Jiwon Hwang, with whom I could share all concerns and ideas. She has been an amazing officemate. My gratitude also goes to Miran Kim, who never stopped feeding me with motherly care. I appreciate her care and friendship.

I can never express my thanks enough to the professors in Korea: My cordial appreciation is due to Professor YoungEun Yoon, who motivated me to study abroad and continued to support me with encouragement. Professor Young-Seok Kim recommended that I apply for the Linguistics at Stony Brook at first, and I truly appreciate his recommendation.

My heartfelt acknowledgment goes to my friends and my family in Korea, who have been there all the time. A thank you goes to my sisters and brothers, who supported me with all their love. I would like to give my deepest thank you and love to my parents, who have always trusted me no matter what, and I am dedicating this dissertation to them.

Chapter 1
Introduction

This dissertation addresses a fundamental question in linguistics: how much of speakers' linguistic behavior is determined by internalized abstract grammatical principles and how much is influenced by the patterns in their existing lexicon. I specifically explore the role of frequency and the sources of gradience and variation.

The issues of lexical frequency, phonotactic gradience, and phonological variation have traditionally been on the margin of research in phonology. Phonological accounts have focused on qualitative patterns and regularities, and have traditionally assumed that the grammar produces categorical outputs, with quantitative patterns dismissed as irregular or marginal phenomena. However, recent research has uncovered many cases of variation, in both phonology and syntax. For example, in English, we find variants such as "sentim[en]tality ~ sentim[n̩]tality" (Kager 1999) in which the vowel may or may not be reduced, and in syntax we find optionality of a complementizer in structures such as "I know that John likes Mary ~ I know John likes Mary." Furthermore, the likelihood of particular variants may be determined by frequency. For instance, the rate of /t, d/ deletion in English is higher for words with high usage frequency, e.g. *and*, *went*, *just*, contracted *not*, whereas the deletion rate is lower for words with low usage frequency, e.g. *feast*, *mast*, *nest* (Bybee 2000a, 2002; Coetzee 2004, 2006a, b, 2008a, b; Labov 1989; Patrick 1992; Santa Ana 1991, *inter alia*). In addition, speakers tend to exhibit gradient acceptability judgments for novel phonological strings, even among structures that do not occur in their native language. For example, it has been shown that English speakers rate possible but non-occurring nonce forms *blick* [blɪk] as better than nonce forms such as *bwip* [bwɪp], which were in turn rated as more acceptable than *bzarshk* [bzarʃk] (Albright 2006a, b, 2007, among others). Although lexical frequency, gradient phonotactics, and variation do influence speakers' behavior, they have rarely been incorporated into a formal grammar, at least until recently.

In the following sections I discuss evidence that these factors are relevant to linguistic analysis. I also outline the central problem of the dissertation: a reduplication process in Korean in which speakers insert a consonant in vowel-initial bases. A variety of consonants may be chosen for insertion, and the choice does not appear to be fully predictable. In this dissertation I investigate the factors affecting the choice of inserted consonant, using a dictionary study and a set of word creation experiments. I argue that while consonant insertion reveals a large degree of variation both within and across speakers, various factors, including the lexical frequency of different consonants in different positions and the frequency of specific C – C and C – V combinations, affect speakers' choice of consonants for insertion.

1.1. Theoretical Issues

1.1.1. Lexical frequency

The role of lexical frequency in determining speakers' phonological behavior is increasingly apparent in a number of areas, including phonetics (Myers 2007; Pierrehumbert 2002); morpho-phonological processes and optional phonological alternations (Zuraw in press; Zuraw & Ryan 2007); complex patterns of variation (Kang 2002, 2007); speech errors (Stemberger & MacWhinney 1986, 1988); lexical decision (Sereno & Jongman 1997; Alegre & Gordon 1999); and language change (Bybee 1985, 2000a, b, 2001; Bybee & Hopper 2001; Bybee & Slobin 1982; Fidelholtz 1975; Hooper 1976; Phillips 1980, 1983, 1984, 1999, 2001, 2007).

Frequency is particularly important in sound change. As has been noted in the literature of lexical diffusion of sound change, some changes affect the most frequent words first, whereas others affect the least frequent ones first (e.g. Bybee 2002; Hooper 1976). In English, deletion of /t, d/ (*best*, *told* vs. *nest*, *meant*) and vowel reduction (*memory*, *nursery*, *scenery* vs. *mammary*, *cursory*, *chicanery*) are processes that affect high-frequency words (the first group of examples) first. In contrast, the regularization of the past tense affects low-frequency verbs (*weep-wept*, *leap-leapt*, *creep-crept*) more often than high-frequency verbs (*keep-kept*, *sleep-slept*, *leave-left*) (Bybee 2002). According to Hooper (1976), the change in high-frequency words is due to the automation of production (Browman & Goldstein 1992), while the change in low-frequency words is due to imperfect learning, as learners have less exposure to low-frequency words. Bybee (1995a) suggests that more frequently used words become more ingrained or entrenched in memory than less used words. This argument implies that exceptional, low-frequency words are more likely to follow the general rules or constraints (=

general patterns).

1.1.2 Gradience

Regarding the locus for the concept of gradience in grammar, Albright (2006a) outlines the following opposing standpoints, based on how grammar itself is viewed: (i) "Grammar is categorical, but performance is gradient"; (ii) "There is no grammar"; (iii) "Grammar itself is probabilistic and gradient." Concerning the mechanism of why and how gradience effects arise, the first and second views argue that grammar, whether it exists or not, does not have to do with gradient effects. According to these points of view, grammar provides categorical judgments, while gradient effects occur due to the task of processing and judging novel items. Thus gradient effects are merely performance effects: for example, when English speakers distinguish two non-occurring nonce forms *blick* and *bnick* in terms of acceptability, rating only the latter as unacceptable, it is not because there is a grammar that provides rules and constraints determining the acceptability of novel forms. Rather, the acceptability judgments may be attributed to how similar the given sequences are to items in the lexicon, e.g. neighborhood effects (cf. Bybee 2001; Bailey & Hahn 2001).

The third view, however, argues that grammaticality is a continuous function, and tasks like gradient acceptability ratings reflect gradient grammaticality. Therefore, the degree of acceptability for nonce forms like *blick* and *bnick* is based on this probabilistic grammar, which regulates how likely segment sequences are (Albright 2006a, 2007; Albright & Hayes 2003; Coleman & Pierrehumbert 1997; Frisch, Large, & Pisoni 2000; Hammond 2004; Hayes & Wilson 2006). Albright (2006a, 2007) concludes that gradient phonotactic acceptability reflects grammatical effects, not performance effects, based on the results of comparing lexical models and sequential models. The lexical models consider factors like token frequency and neighborhood density in their computation, and the sequential models, which perform better according to Albright, consider factors like type frequency, natural classes, and markedness.

1.1.3 Variation

Variation is also related to the issue of lexical frequency vs. grammar, as is the question of phonotactic gradience, as discussed in the section above. While classical generative phonology has tended to abstract away from variation, there have been models proposed in which variation is not external to the lexicon and grammar, but rather is intrinsic to it (Bybee 2002; Pierrehumbert 1994, 2001, among others). In exemplar-based models, mental representations and the grammatical structure emerge from experience with language; that is, linguistic experiences are categorized with reference to already stored representations, which are also known as exemplar clusters. Such models deem mental representations to be directly formed by speakers' memories of tokens of linguistic items, a stance which does not necessarily presuppose an *a priori* grammar.

Even among grammars assuming abstract mental representations, there have been recent efforts to formalize variation in formal grammars. These approaches within Optimality Theoretic grammars include Partially Ordered Grammars (Anttila 1997), Floating Constraints (Nagy & Reynolds 1997; Reynolds 1994), Constraint Competition (Zubritskaya 1997), Stochastic OT (Boersma 1997; Boersma & Hayes 2001), the Rank-Ordering Model of EVAL (Coetzee 2004, 2006a, b), and Lexically Indexed Variation (Coetzee 2007). These formal approaches argue that variation does not change grammar; rather, grammar accounts for variation. Variation may arise from stochastic constraint rankings (cf. Boersma 1997; Boersma & Hayes 2001) or from the different degrees of constraint violation among non-optimal candidates (cf. Coetzee 2004, 2008a, b, among others).

1.2 A Test Case: Consonant Insertion

The questions of the role of frequency and of the sources of gradience and variation are still controversial. To address these questions, I will look into a specific phenomenon that exhibits gradience and variation, utilizing lexical and grammatical tools.

In my dissertation I focus on a case of consonant insertion, the process of which is attested in many languages. Many languages have been argued to have a single unmarked consonant for epenthesis: (Lombardi 2002; Vaux 2003)[1,2]

---

[1] The references for each language were provided in Vaux (2003), which have been excluded in the table, for exposition: Korean (Kim-Renaud 1975, Hong 1997), Maru (Burling 1966, Blust 1994), Finnish (Anttila 1994), a French aphasic (Kilani-Schoch 1983), Greek (Smythe 1920), Sanskrit (de Chene 1983), Dutch (Booij 1995), German dialects (Ortmann 1998), Buginese (Trigo-Ferre 1988, Lombardi 1997), Inuktitut and East Greenlandic (Mennecier 1995, 1998; Massenet 1986), Basque (Hualde & Gaminde 1998), Japanese (de Chene 1985), Seville Spanish (Martin-Gonzalez, Vaux's p.c.), Bristol English (Wells 1981), Midlands American English (Gick 1999), Motu (Crowley 1992), Polish (Nowak, Vaux's p.c.), Turkish (Underhill 1976), Greenlandic (Rischel 1974), Pishaca (Grierson 1906), various Indic languages (Masica 1991), Arabic (Heath

(1) Table 1.1 Consonants for insertion in different languages

Epenthetic Cs	Languages
ʔ	Tamil, Arabic, Selayarese, German, Ilokano, Czech, Kisar, Malay, Koryak, Indonesian, Gokana, Tunica, English, Cupeño, Persian, Thai
h	Yucatec Maya, Huariapano, Onondaga
t	Axininca, Amharic, Odawa, Algonquian languages, Plains Cree, Korean, French, Maru, Finnish[3]
d	A French aphasic
n	Korean, Greek, Sanskrit, Dutch, German dialects
m	Georgian[4]
ŋ	Buginese
N	Inuktitut and East Greenlandic
r	English, German, Uyghur, Zaraitzu Basque, Japanese, Seville Spanish
l	Bristol English, Midlands American English, Motu, Polish
j	Turkish, Uyghur, Geenlandic, Pishaca, various Indic languages, Arabic, Slavic, Korean[5]
w	Abajero Guajiro, Greenlandic, Arabic
v	Marathi
b	Basque (Markina, Urdiain, Etxarri, & Lizarraga dialects)
ʃ	Basque (Lekeito/Deba & Zumaia dialects)
ʒ	Cretan and Mani Greek, Basque dialects

1987), Slavic (Carlton 1991), Marathi (Bloch 1919; Masica 1991), Cretan and Mani Greek (Newton 1972), Land Dayak (Blust 1994), Dominican Spanish (Morgan 1998).

[2] More languages, which have epenthetic glottals /h, ʔ/, were added after Lombardi (2002).

[3] Amharic, Odawa, Algonquian languages, and Plains Cree were added from Lombardi (2002).

[4] I added Georgian in the table, which prefers {m, b} for insertion, e.g. in the case of reduplication (Alice Harris and Ramaz Krudadze, p.c.).

[5] I added Korean since Korean has a /j/ insertion process, e.g. /pata-**j**-a/ 'sea-vocative,' /hak'jo-e/ ~ /hak'jo-**j**-e/ 'school-in.'

g	Mongolian, Buryat
s/z	French, Land Dayak, Dominican Spanish
x	Land Dayak
k	Maru, (Danish?)

Some languages use more than one segment as an epenthetic consonant, which is problematic for the view that the choice of epenthetic consonant is determined by markedness, whether defined across languages or within languages.

One such case is Korean, which employs different consonants, i.e. {t, n, j}, as epenthetic for the purposes of different processes. I will go over some examples for each epenthetic consonant in the next section.

1.2.1 Consonant insertion in Korean

I present some examples for three processes of epenthesis in Korean, which insert /t/, /n/, and /j/, respectively. First, /t/-epenthesis inserts /t/ between two nouns in a compound (2): [6]

(2) /t/-epenthesis (Kang 2003)

a. /u + os/ [utot] (> [udot˺]) 'top clothes'
b. /k^{h}o + nal/ [k^{h}otnal] (> [k^{h}onnal]) 'tip of a nose'
c. /ki + pal/ [kitpal] (> [kip˺p'al]) 'flag'

The surface realization of /t/ varies depending on context (2a-c).

/n/-epenthesis is a phonological process in which /n/ is optionally inserted before /i/ or /j/ between words in a compound (3) and across words in a phrase (4):

(3) /n/-epenthesis in a compound (Kang 2003; Kang 2005)

a. /pat ilaŋ/ [pat.ni.laŋ] (> [panniɾaŋ]) 'furrow'
b. /hwipal ju/ [hwi.pal.nju] (> [hwipallju]) 'gasoline'
c. /nun jak/ [nun.njak] 'eye drops'

[6] I provide phonemic transcriptions throughout the dissertation, unless phonetic transcriptions become of interest in some occasions.

(4) /n/-epenthesis across words (Kang 2005)

a. /os ip-ko/ [on.nip.k'o] 'wearing clothes'

b. /ʧʰam jep'ɨn jʌʧa/ [ʧʰam.nje.p'ɨn.njʌ.ʧa] 'a very pretty girl'

In /j/-epenthesis /j/ is epenthesized to prevent vowel hiatus (5).

(5) /j/-epenthesis

a. /solmi-a/ [solmija] 'Solmi + vocative'

b. /jʌki-e/ [jʌkije] 'here + in'

We can see that each process of epenthesis above refers to its context: a certain consonant, rather than others, is chosen as an epenthetic segment depending on the context.

However, in Korean there is another process, consonant insertion in reduplication, in which it is not a single consonant that is inserted, but various consonants are inserted. How can we know which consonant to insert? Can we still account for this case by making reference to the context only?

1.2.2 Reduplication in Korean

Korean has a number of ideophones that are usually used to express onomatopoeia. Grammatically, they are adjectival or adverbial. Morphologically, they are formed by two types of reduplication, total and partial. I will give a brief overview for each of these types, and I will move on to the total reduplication, which is the focus of my discussion, later in the following sections.

1.2.2.1 Reduplication patterns

When the reduplicant is smaller than the base, the reduplicant generally constitutes a single syllable, open or closed.[7]

[7] Reduplicants are indicated with an underline.

(6)	a.	k'o-k'otek	'cock-a-doodle-doo'
	b.	tu-tuŋsil	'floatingly'
	c.	ta-tali	'every month'
	d.	p'a-ʧi-ʧik	'with a fizzle'
	e.	p'a-tɨ-tɨk	'with a grinding sound'
	f.	ʧ'ak-ʧ'a-k'uŋ	'agreeableness'
	g.	p'o-tɨ-tɨk	'sound made by something fresh and clean'
	h.	p^{h}ɨ-lɨ-lɨn	'bluish'
	i.	k'o-lɨ-lɨk	'borborygmus'
	j.	nʌpte-te	'flattish'
	k.	jasi-si	'showy'
	l.	pusi-si	'unkemptly'
	m.	pesi-si	'with a smile'
	n.	p^{h}alɨ-lɨ	'shiveringly'

Whether it is prefixation (6a-c), infixation (6d-i), or suffixation (6j-n), all the data in (6) have a reduplicant which is constituted of the universally preferred type of syllable, CV. We also come across examples with a reduplicant made up of CVC:[8]

(7)	a.	t'ek-t'ekul	'rolling'
	b.	kol-kolu	'equally'
	c.	ʌlt'ʌl-t'ʌl	'puzzled'
	d.	ʌʧʌŋ-ʧ'ʌŋ	'equivocal'[9]

In some other instances, the reduplicant is partial, but contains two syllables.

(8)	a.	ali-alilaŋ	'repeated form from a ballad titled "alilang"'
	b.	sɨli-sɨlilaŋ	'a lyric from the ballad "alilang"'

In total reduplication, the reduplicant and the base are generally identical: in one type, the first and second syllables are copied separately:

---

[8] Also see McCarthy (1993) for English examples.

[9] The tensification of an onset in the reduplicant is a separate issue of phonology which is not relevant to the discussion.

(9)	a.	t’it’ip’aŋp’aŋ	‘honking’
	b.	ʧ’ukʧ’ukp’aŋp’aŋ	‘slim and glamorous’
	c.	ʧiʧipepe	‘singing of a swallow’
	d.	kukuʧʌlʧʌl	‘phrase by phrase; clause by clause’

The forms in (9a-b) can be split into *t’it’i* ‘honking’ and *p’aŋp’aŋ* ‘honking’ in (9a) and *ʧ’ukʧ’uk* and *p’aŋp’aŋ* in (9b). They are formed by compounding the two reduplicated forms, which are related in meaning. As for (9c-d), division of the whole into two parts is pointless since neither of the parts is used alone.

A more common pattern of total reduplication involves copying a string of two syllables:

(10)	a.	p^{h}otoŋ-p^{h}otoŋ	‘chubby’
	b.	mik’ɨl-mik’ɨl	‘slippery’
	c.	p^{h}alɨt-p^{h}alɨt	‘verdant’
	d.	pokɨl-pokɨl	‘simmering’
	e.	paŋkɨl-paŋkɨl	‘smilingly’
	f.	aʧaŋ-aʧaŋ	‘toddlingly’
	g.	tekul-tekul	‘rolling’
	h.	holi-holi	‘slim’
	i.	p’ʌn-p’ʌn	‘cheeky’
	j.	ʧol-ʧol	‘trickling; tagging along’
	k.	t’ok-t’ok	‘dripping; knocking; smart’

For this pattern, when the first member of the reduplicated form is vowel-initial, the second member begins with a consonant:

(11)	a.	als’oŋ-**t**als’oŋ	‘confusing’
	b.	oson-**t**oson	‘on good terms’
	c.	oŋki-**ʧ**oŋki	‘densely’
	d.	alok-**t**alok	‘mottled’
	e.	ult^{h}uŋ-**p**ult^{h}uŋ	‘bumpy’
	f.	ulkɨlak-**p**ulkɨlak	‘alternately pale and red’
	g.	ulkɨt-**p**ulkɨt	‘blue and red’
	h.	opul-**k**opul	‘meanderingly’
	i.	olmaŋ-**ʧ**olmaŋ	‘all sorts of little things (in a cluster)’
	j.	ali-**k’**ali	‘confused’

We can also in some cases find a mismatch in vowel qualities (12a-b), consonant properties (12c-d), or both vowel and consonant features (12e-f).[10,11]

(12)	a.	s**i**ŋsuŋ-s**e**ŋsuŋ	'fidgety'
	b.	p**i**ʧaŋ-p**a**ʧaŋ	'even'
	c.	**s**aŋkɨl-**p**aŋkɨl	'all smiles'
	d.	**k**ʌmpul-**t**ʌmpul	'pell-mell'
	e.	**ka**lp^{h}aŋ-**ʧi**lp^{h}aŋ	'at a loss'
	f.	**si**tɨl-**pu**tɨl	'wilted and withered'

In the next section I consider the question of determining which portion is the base and which the reduplicant.

1.2.2.2 Defining the base of reduplication

I look back to some representative examples in which one portion of a reduplicated word is V-initial and the other is C-initial.

(13)	a.	als'oŋ-**t**als'oŋ	'confusing'
	b.	ult^{h}uŋ-**p**ult^{h}uŋ	'bumpy'
	c.	opul-**k**opul	'meanderingly'
	d.	olmaŋ-**ʧ**olmaŋ	'all sorts of little things (in a cluster)'

With respect to these consonants appearing in the total reduplication, the initial question is raised: Are they inserted or deleted? In other words, which portion is the base? I will assume that the vowel-initial portion is the base, for the following reasons. First, the first morpheme in *als'oŋ-tals'oŋ* is from an independent form, *alisoŋ*, and *olmaŋ-olmaŋ* can be used for *olmaŋ-ʧolmaŋ*, *aʧaŋ-aʧaŋ* for *aʧaŋ-paʧaŋ*, *otoŋ-otoŋ* for *otoŋ-potoŋ*, *ukɨl-ukɨl* for *ukɨl-pukɨl*, and *omok-omok* for *omok-ʧomok*, while conveying the same meaning. Second, there is a general tendency that the onset consonant in the base is maintained in the

[10] The first morpheme of (12e), *kalp^{h}aŋ* may come from the morpheme, *ka+l* ('go/do + future tense'), and the second morpheme, *ʧilp^{h}aŋ*, may originate from the morpheme *ʧi+l* ('negation or question + future tense').

[11] Examples like (11-12) are also found in English, e.g. itsy-bitsy, arty-farty, rolly-polly, hokey-pokey, between which I concentrate on the examples like in (11) in the later discussion, but I will also investigate the case of (12) in my future research (cf. Ahn 2005; Parker 2002 for the English reduplication).

reduplicant. It is very unusual to skip the initial consonant of the base in the Korean reduplication process. Therefore, if the second morpheme in (13) were the base, then the reduplicative forms should be *tals'oŋ-tals'oŋ*, *pult^h^uŋ-pult^h^uŋ*, *kopul-kopul*, *ʧolmaŋ-ʧolmaŋ*, rather than *als'oŋ-**t**als'oŋ*, *ult^h^uŋ-**p**ult^h^uŋ*, *opul-**k**opul*, *olmaŋ-**ʧ**olmaŋ*. Third, the consonant-initial portion is phonologically less marked than the onsetless vowel-initial portion. It has been cross-linguistically observed that reduplicants tend to be less marked than their bases (Alderete *et al*. 1999; Kager 1999; McCarthy and Prince 1994, among others). The syllable structure CV is the least marked in the world's languages, and a syllable with an onset is less marked than one without. This argues that the portion with an onset should be the reduplicant in the case of the Korean reduplication. Finally, the motivation for deleting a consonant in word-initial position is not clear. However, if we assume epenthesis, we can argue that the universal tendency to have an onset leads to the insertion of an onset consonant in the onsetless syllable of the base. Therefore, without any compelling evidence to the contrary, I assume that this reduplication involves epenthesis; it is not a case of deletion.

1.2.2.3 Inserted consonants

If consonants are inserted in the onset of the reduplicant, what consonants can be inserted? Table 1.2 gives the consonant inventory of Korean. All of the consonants, except for /ŋ/, can occur in syllable onset position.

(14) Table 1.2 Consonant phoneme inventory of Korean

Place / Manner	**Bilabial**	**Alveolar**	**Palatal**	**Velar**	**Glottal**
Stop	p p^h p’	t t^h t’		k k^h k’	
Affricate			tʃ $tʃ^h$ tʃ’		
Nasal	m	n		ŋ	
Fricative		s s’			h
Approximant	(w)[12]	l	(j)		

In fact, all the onset consonants can also appear as an onset in the reduplicant.

A search of a Korean dictionary revealed 343 entries of total reduplication with an inserted (185 entries) or replaced (158 entries) consonant in the onset of the reduplicant.[13] Korean differentiates obstruents in terms of aspiration and tenseness. Therefore, there are three kinds of [-continuant] obstruents, i.e. lenis, aspirated, and fortis. However, for the time being I treat them as one sound sharing the same place and manner since I will consider two variables, place and manner of articulation, in this dissertation. For instance, /p, p^h, p’/ will be regarded as a single type of consonant.

To investigate the data from the viewpoint of only phonological factors, I excluded 35 out of 185 insertion cases which had meaning association or sound assimilation between the inserted consonant and its neighboring consonants. For instance, *ijʌl-***tʃ^h***ijʌl* ‘Like cures like’ is a set phrase originating from Chinese characters. Thus the second portion, **tʃ^h***ijʌl* ‘cure fire’ cannot be viewed as a pure reduplication of the first portion, *ijʌl* ‘with fire.’ The consonant **tʃ^h** is not inserted but the morpheme **tʃ^h***i* ‘cure’ replaces the whole morpheme *i* ‘with’ in the first portion of the word. In *olɨlak-***n***elilak* ‘rising and falling’ *olɨ* is a stem meaning ‘ascend’ and *neli* is another stem meaning ‘descend.’ Therefore, this cannot be considered to constitute a genuine reduplicated form.[14] As for sound assimilation,

---

[12] Korean glides have been variously considered as consonants and as combinations of two vowels. For the discussion on the status of the palatal glide, see An, Hwang, & Suh 2008.

[13] *Eysseynsu Kwuke Sacen* [*Essence Korean Dictionary*]. 2006. Phacwu, Korea: Mincwungselim Co.

[14] As was pointed out by Ellen Broselow (p.c.), they may be compounds, rather than reduplicative

I regarded examples like *ʌkɨt-**pʌkɨt*** 'uneven,' *ʌsɨt-**pisɨt*** 'similar,' and *ulkɨt-**pulkɨt*** 'colorful' as having assimilation between the last segment of the base and the inserted consonant in the reduplicant. In all the assimilation cases, the preceding consonant was /t/ and the inserted consonant was /p/, in which case /t/ becomes /p/ as in /ʌkɨ**t**-**p**ʌkɨt/ → [ʌkɨ**p**-**p**ʌkɨt].

Examples of each inserted consonant (CI) are provided below. The percentage given for each set of examples indicates the proportion of each group of sounds out of a total of 150 items, which were chosen from the list of 185 for the reason given above.[15,16]

(15) **alveolar stops** **(29.33 %)**

a.	als'oŋ-**t**als'oŋ	'confusing'
b.	oson-**t**oson	'on good terms'
c.	ʌlluŋ-**t**ʌlluŋ	'speckled'
d.	allok-**t**allok	'pied'
e.	otol-**t^h**otol	'hard and lumpy'
f.	ʌʧuŋi-**t'**ʌʧuŋi	'rabble'

(16) **bilabial stops** **(28.67 %)**

a.	ult^{h}uŋ-**p**ult^{h}uŋ	'bumpy'
b.	ʌʧʌŋ-**p**ʌʧʌŋ	'rambling'
c.	ʌli-**p**ʌli	'silly'
d.	uʧil-**p**uʧil	'brusque'
e.	okɨl-**p**okɨl	'bubbling'
f.	otoŋ-**p^h**otoŋ	'chubby'

forms.

15 It was pointed out that a dictionary may hold many archaic words that do not reflect the current grammar (Marie Huffman, p.c.). I looked at the reduplicative forms (V-initial bases) in my dictionary data, and around 10.67 % (16 items out of 150) seems to be less frequently used among speakers, which is judged due to my own personal experience. I do not think it will impact on the current data results.

16 Inserted consonants in reduplicant are marked in bold face.

(17) **palatal affricates** **(25.33 %)**

a.	oŋki-**tʃ**oŋki	'densely'
b.	olmaŋ-**tʃ**olmaŋ	'all sorts of little things (in a cluster)'
c.	ʌls'a-**tʃ**ʌls'a	'delightfully'
d.	ollaŋ-**tʃ^{h}**ollaŋ	'splashing gently'
e.	umul-**tʃ'**umul	'hesitantly'

(18) **velar stops** **(6 %)**

a.	upul-**k**upul	'windingly'
b.	allali-**k'**allali	'bantering sound'

(19) **alveolar fricatives** **(5.33 %)**

a.	alt'ɨl-**s**alt'ɨl	'extremely frugal'
b.	ʌlki-**s**ʌlki	'entangled'

(20) **bilabial nasals** **(2.67 %)**

a.	oŋsoŋ-**m**aŋsoŋ	'hazy'
b.	ʌli-**m**ali	'drowsily'

(21) **palatal approximants** **(2.67 %)**

a.	illʌŋ-**j**allaŋ	'rocking'
b.	iltʃ'uk-**j**altʃ'uk	'from side to side'

The consonants /p', k^{h}, n, s', h, w, l/ happen not to show up in the dictionary examples, but there is no general phonological principle that would prevent them from occurring in onset position. They are theoretically possible, but are empirically rare. I will now consider various hypotheses to account for the choice of inserted consonant.

According to Alderete, Beckman, Benua, Gnanadesikan, McCarthy, and Urbanczyk (1999), if the segments in the reduplicant are not present in the base, then they are either the least marked C or V of the language or a separate morpheme. Thus first I will consider whether the consonant insertion can be predictable based on markedness. Since the choice of inserted consonant varies, we cannot identify a single unmarked consonant, so must define markedness in terms of context:

(22) **Hypothesis 1**
An inserted segment represents the least marked segment possible in a specific context.

The inserted C in the Korean reduplication cannot be an unmarked or default consonant because distinct consonants can be inserted in very similar environments.

(23)
a. alok-**t**alok ‘pied’
b. ulak-**p**ulak ‘wild’
c. umuk-**ʧ**umuk ‘unevenly hollowed’
d. upul-**k’**upul ‘windingly’

/t/ is epenthesized in (24a) but /p/ in (24b) although the bases contain the same set of consonants, /l/ and /k/. Furthermore, the choice of the inserted consonant does not depend on the vowels in the base. /p/, /ʧ/, and /k’/ are epenthesized in (24b-d) respectively, even though they are followed by the same vowel /u/. In this regard, we can see from the following table that there is no clear-cut criterion distinguishing a certain pair of CV from other pairs of CV. For instance, it is hard to argue that it is more likely that /t/ is followed by /ʌ/, /p/ is followed by /u/, and /ʧ/ is followed by /o/. Rather, we may argue that two or more types of vowel are more likely to follow the given consonants, and those vowels happen to be non-front vowels, which may be due to some other factor at work concerning the vowel inventory in Korean. Therefore, a particular vowel does not force the occurrence of a particular consonant.

(24) Table 1.3 CI (/t, p, ʧ/, among others) and its following V combinations in VCVC-**CIV**CVC data from the dictionary (*Eysseynsu Kwuke Sacen*. [Essence Korean Dictionary] 2006. Phacwu, Korea: Mincwungselim; 51 tokens in total)

CI \ following V	/i/	/e/	/ʌ/	/a/	/o/	/u/
/t/	0	1	6	5	4	2
/p/	2	0	3	5	4	7
/ʧ/	2	0	0	1	5	4

An alternative to using markedness to predict the quality of non-copied segments Alderete *et al*. (1999) identify cases like English shm-reduplication (table-shmable) in which they argue that the noncopied material stands alone as an independent morpheme. In the case of Korean, we might hypothesize that several different such morphemes exist corresponding to the different inserted consonants:

(25) **Hypothesis 2**
Separate CIs represent separate morphemes.

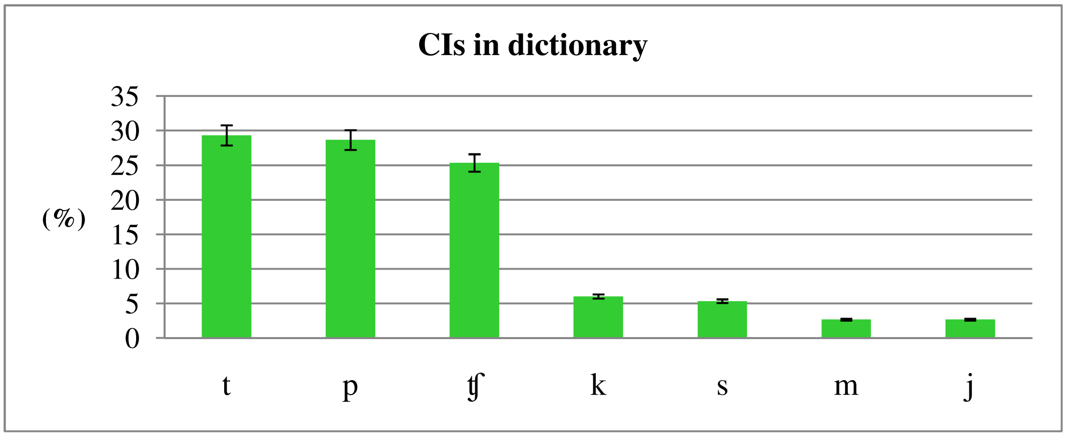
If a segment is a separate morpheme, then it is an affix which must exist in the input. However, there is no evidence that the different inserted Cs carry different elements of meaning or exhibit any differences in behavior. If we simply identify all the possible onset Cs of the language as separate morphemes that may appear in reduplicants, we still have to explain how a speaker chooses from among this set of morphemes in forming the reduplicated version of individual bases.

Another possible alternative is to give up hope of any predictability in the choice of inserted consonants:

(26) **Hypothesis 3**
The choice of inserted consonant is random.

If the choice is randomly made, it is predicted that all the attested CIs should have the same frequency of occurrence. For example, for any given context we expect to detect the same frequency for each possible inserted consonant. However, analysis of all the cases of inserted consonants in biconsonantal bases in the dictionary demonstrates that certain consonants (/t, p, ʧ/) are much more frequently inserted than others (/k, s, m, j/), as shown in Figure 1.1.

(27) Figure 1.1 CI frequency in the dictionary



We do not see random choices, but some patterns: /t, p, ʧ/ are much more frequent than /k, s, m, j/ as CIs. There must be a reason that can account for this pattern. I argue that the choice of CIs is predictable to some extent, although it may not be completely predictable. I examine the factors that are involved in the choice of CIs, in the subsequent chapters.

As attested in the dictionary data, various consonants can be inserted in the reduplicated words; moreover, different consonants can be used as an epenthetic C even in phonologically similar contexts. Furthermore, the CIs are neither unmarked Cs nor separate morphemes in Korean. The reduplication data with a CI (CI-reduplication) involves variation and gradient judgments of acceptability, as will be shown later in the nonce reduplicated forms created by speakers. Based on the analyses of dictionary data and a series of experiments, I will argue that the choice of CI is made lexically, and I will further argue that these apparent lexical effects are in fact grounded in some grammatically determined concepts.

1.3 Overview: Methodology

1.3.1 Dictionary study

To understand the distribution and frequency of CIs in the lexicon, I examined a Korean dictionary (*Eysseynsu Kwuke Sacen* 2006), which revealed 343 entries of total reduplication with an inserted (185 entries) or replaced (158 entries) consonant in the onset of the reduplicant. To investigate the data from the viewpoint of only phonological factors, I excluded 35 out of 185 insertion cases which had meaning association or sound assimilation between the inserted consonant and its neighboring consonants.[17] Thus there were 150 entries that are V-initial reduplicated words. In the later chapters, I analyze the forms with CI=/t, p, ʧ/, which are the most frequent, and they are provided in Appendix 1-A. There were 51 V-initial (VCVC-base) reduplicated words with CI=/t, p, ʧ/.

1.3.2 Behavioral experiments

I conducted several real time behavioral experiments, Experiments 1 – 4, in which Korean speakers were asked to create the most natural reduplicated word with a nonce base (word creation task). In the first two experiments, participants were asked to insert a consonant freely, and in the last two experiments participants were asked to create reduplicative forms by choosing a consonant from a fixed set of particular segments, i.e. /t, p, ʧ/ in Experiment 3, and /t, ʧ/ in Experiment 4. Details on the participants and the stimuli are provided in Appendix 1-B. The participants in these experiments were asked to pronounce aloud the forms they were just making up, and I assumed in my analysis of the responses that the created words were spoken forms, rather than written forms. Any erroneous responses were removed. The participants' choice of CI appeared to correlate with lexical patterns in some aspects; however, it was determined more strictly based on some grammatical principles.

1.4 Dissertation Outline

The discussion of the dissertation is organized as follows. In Chapter 2 I investigate whether the frequency of specific inserted consonants reflects the frequency of those consonants in the lexicon (in all positions, in word-initial position, and in syllable onset), in both the corpus and the first word creation experiment. I also take into consideration the effect of preceding and following context. I argue that the speakers' behavior reflects lexical frequency to some extent; however, the lexical frequency cannot completely explain the choice behavior.

Chapter 3 provides evidence that individual speakers show preferences for certain different consonants as an inserted consonant, which suggests that

[17] This paragraph is an abridged version of what was given in Section 1.2.2.3, in which the excluded examples due to meaning association or sound assimilation were presented.

individual speakers may have different grammars for the consonant choice.

Chapter 4 examines the context for consonant insertion more closely. Since an inserted consonant is followed by a vowel and preceded by a consonant, the consonant to be chosen may be affected by these surrounding segments. I investigate how this can be explicated in the relationships of CI – V and C – CI. I argue that the phonotactic knowledge about a CV constituent plays a role in the consonant choice, and the chosen consonant is subject to the constraints on syllable contact in relation to the preceding consonant.

Chapter 5 considers co-occurrence restriction on consonants which disfavor nearby consonants from sharing place and/or manner of articulation. Based on the Observed/Expected values measured for CI and neighboring Cs, I show that there are restrictions on co-occurring consonants sharing place of articulation in newly created reduplicative words, which also has been argued to exist in the general vocabulary of Korean.

Chapter 6 concludes the dissertation with a summary and remarks on the future directions for research.

1.5 Summary

Questions concerning frequency, gradience, and variation arise in the analysis of various phenomena in numerous languages. As part of the efforts to define the role of frequency in phonology and the sources for gradient intuition and variable data, I investigate consonant insertion, particularly consonant insertion in Korean reduplication. The choice of consonants in this reduplication is not categorically made, and the range of inserted consonants is wide. The investigation of corpus data along with speakers' behavior in word creation experiments helps us to understand the role of lexical frequency and of grammar in linguistic phenomena.

Appendix 1-A Dictionary data

I furnish the reduplicative words with VCVC-bases and CI=/t, p, ʧ/ from the dictionary (*Eysseynsu Kwuke Sacen* 2006). A total of 51 entries are given below, with IPA transcriptions:

Entries

1.	아득바득	atɨk-patɨk
2.	아근바근	akɨn-pakɨn
3.	아등바등	atɨŋ-patɨŋ
4.	아락바락	alak-palak
5.	아장바장	aʧaŋ-paʧaŋ
6.	어뜩비뜩	ʌt’ɨk-pit’ɨk
7.	어칠비칠	ʌʧ^{h}il-piʧ^{h}il
8.	오글보글	ok1l-pok1l
9.	올록볼록	ollok-pollok
10.	오동보동	otoŋ-potoŋ
11.	오동포동	otoŋ-p^{h}otoŋ
12.	우글부글	ukɨl-pukɨl
13.	우락부락	ulak-pulak
14.	우질부질	uʧil-puʧil
15.	울룩불룩	ulluk-pulluk
16.	우둥부둥	utuŋ-putuŋ
17.	우둥푸둥	utuŋ-p^{h}utuŋ
18.	어영부영	ʌjʌŋ-pujʌŋ
19.	어금버금	ʌkɨm-pʌkɨm
20.	어근버근	ʌkɨn-pʌkɨn
21.	어정버정	ʌʧʌŋ-pʌʧʌŋ
22.	알락달락	allak-tallak
23.	알록달록	allok-tallok
24.	아록다록	alok-talok
25.	알롱달롱	alloŋ-talloŋ
26.	아롱다롱	aloŋ-taloŋ
27.	애동대동	etoŋ-tetoŋ
28.	오돌토돌	otol-thotol
29.	오톨도톨	ot^{h}ol-tot^{h}ol
30.	오손도손	oson-toson
31.	오순도순	osun-tosun

32. 우둘투둘 utul-t^hutul
33. 우툴두툴 ut^hul-tut^hul
34. 어룩더룩 ʌluk-tʌluk
35. 얼룩덜룩 ʌlluk-tʌlluk
36. 어룽더룽 ʌluŋ-tʌluŋ
37. 얼룽덜룽 ʌlluŋ-tʌlluŋ
38. 얼럭덜럭 ʌllʌk-tʌllʌk
39. 어런더런 ʌlʌn-tʌlʌn
40. 어빡자빡 ʌp’ak-ʧap’ak
41. 어금지금 ʌkɨm-ʧikɨm
42. 우걱지걱 ukʌk-ʧikʌk
43. 오글쪼글 okɨl-ʧ’okɨl
44. 올랑촐랑 ollaŋ-$ʧ^h$ollaŋ
45. 오밀조밀 omil-ʧomil
46. 오목조목 omok-ʧomok
47. 오롱조롱 oloŋ-ʧoloŋ
48. 우글쭈글 ukɨl-ʧ’ukɨl
49. 우묵주묵 umuk-ʧumuk
50. 우물쭈물 umul-ʧ’umul
51. 울렁출렁 ullʌŋ-$ʧ^h$ullʌŋ

Appendix 1-B Experiments: Participants and stimuli

The participants varied in age, ranging from 20’s to 60’s, who were recruited in Seoul, Korea for Experiment 1 and in Stony Brook, New York, for Experiments 2 – 4. There were 55 participants in Experiment 1 and 15 participants in each of Experiments 2 – 4.

In Experiment 1, there were 3 stimuli containing one C, 19 stimuli containing two Cs, 15 stimuli containing three Cs, and 3 stimuli containing four Cs, in the given word creation task, all of which amount to 40. In more detail, there were 2 of the base form VCV, 1 of VVC, 15 of VCVC, 3 of VCCV, 1 of CVCV, 4 of VCCVC, 9 of CVCVC, 2 of CVCCV, and 3 of CVCCVC. In Experiments 2 – 4, there were 111 VCVC-bases as the stimuli, along with 50 fillers, which were used for the three experiments after randomizations.

Experiment 1

Directions: Each of the following morphemes is part of a reduplicative form. Based on your intuition as a native speaker of Korean, you are requested to fill in each of the blanks with a copied form of the given item. When you create a reduplicant, please make sure that a segment should be different from the correspondent in the given morpheme. Also make sure to read new forms aloud when you are creating them. Feel free to write them in Korean.

Instantiation

a. 옹기종기 [oŋki-ʧoŋki]
b. 알쏭달쏭 [als’oŋ-tals’oŋ]
c. 오손도손 [oson-toson]

Stimuli[18]

1. 언들 ʌntɨl
2. 우술 usul
3. 바직 paʧik
4. 언장 ʌnʧaŋ
5. 두룩 tuluk
6. 오독 otok
7. 살캉 salk^{h}aŋ
8. 오작 oʧak
9. 가삼 kasam
10. 시렁 silʌŋ
11. 울짜 ulʧ’a
12. 곰직 komʧik
13. 아식 asik
14. 아달 atal
15. 빠사 p’asa
16. 우칠 uʧ^{h}il
17. 엉차 ʌŋʧ^{h}a
18. 붕소 puŋso
19. 앗짜 atʧ’a

---

[18] The stimuli were given in Korean for the participants, and the transcriptions provided next to Korean were not in the experiment. Some of the words are from a certain dialect of Korean, which is spoken by none of the participants, and most of them are newly made up for the sake of this experiment.

20. 어울 ʌul
21. 졸사 ʧolsa
22. 옥수 oksu
23. 아주 aʧu
24. 억짱 ʌkʧ’aŋ
25. 어중 ʌʧuŋ
26. 구들 kutɨl
27. 오삼 osam
28. 우끈 uk’ɨn
29. 가만 kaman
30. 아장 aʧaŋ
31. 우설 usʌl
32. 고당 kotaŋ
33. 오공 okoŋ
34. 담풍 tampʰuŋ
35. 아밤 apam
36. 오감 okam
37. 모든 motɨn
38. 후룩 huluk
39. 오롱 oloŋ
40. 온당 ontaŋ

Experiments 2, 3, and 4

The following is the list I used for the three experiments (Experiments 2 – 4), which has 111 stimuli and 50 fillers, with variables: order (C_1C_2, C_2C_1), place (labial, alveolar, velar), manner (stop, fricative, nasal), and vowels (i-i, u-u, a-a) in a base form of VCVC. The stimuli were randomized in each experiment and the instructions were also different in the three experiments: the participants were asked to create the most natural reduplicated form with any C from the consonant inventory of Korean in Experiment 1; with a C from a set of /t, p, ʧ/ in Experiment 2; and with a C from a set of /t, ʧ/ in Experiment 3.

Examples that can help the participants to understand the instructions were given orally to each participant by the experimenter.

The list was provided in Korean only; I provide the IPA transcriptions in the following for ease of understanding.

Stimuli

1. 오굴 okul
2. 어송 ʌsoŋ
3. 오망 omaŋ
4. 웅서 uŋsʌ
5. 이깁 ikip
6. 우군 ukun
7. 아만 aman
8. 아막 amak
9. 아망 amaŋ
10. 엉구 ʌŋku
11. 옹사 oŋsa
12. 아답 atap
13. 우둥 utuŋ
14. 이깅 ikiŋ
15. 이닙 inip
16. 이님 inim
17. 인다 inta
18. 이공 ikoŋ
19. 이닉 inik
20. 우눗 unus
21. 우눕 unup
22. 우눔 unum
23. 아담 atam
24. 우멍 umʌŋ
25. 애싱 esiŋ
26. 우굿 ukus
27. 우뭅 umup
28. 우뭇 umus
29. 우눅 unuk
30. 아강 akaŋ
31. 오닝 oniŋ
32. 아넝 anʌŋ
33. 이믹 imik
34. 우둣 utus
35. 이딘 itin
36. 우둡 utup
37. 아납 anap

38. 오솔 osol
39. 우놀 unol
40. 이닛 inis
41. 우둔 utun
42. 우둑 utuk
43. 우눙 unuŋ
44. 이긴 ikin
45. 이굴 ikul
46. 잉가 iŋka
47. 우분 upun
48. 웅욱 uŋuk
49. 우문 umun
50. 아당 ataŋ
51. 아삽 asap
52. 웅가 uŋka
53. 옹마 oŋma
54. 우숨 usum
55. 이딩 itiŋ
56. 아상 asaŋ
57. 이깃 ikis
58. 우굽 ukup
59. 응지 ɨŋʧi
60. 잉모 iŋmo
61. 아삭 asak
62. 이닝 iniŋ
63. 웅웃 uŋus
64. 잉입 iŋip
65. 잉잇 iŋis
66. 아몽 amoŋ
67. 이상 isaŋ
68. 이빈 ipin
69. 아박 apak
70. 우굼 ukum
71. 이딧 itis
72. 아산 asan
73. 어강 ʌkaŋ
74. 애밍 emiŋ
75. 웅읍 uŋup

76.	우둠	utum
77.	이싱	isiŋ
78.	잉익	iŋik
79.	잉인	iŋin
80.	오성	osʌŋ
81.	어징	ʌʧiŋ
82.	잉임	iŋim
83.	우숩	usup
84.	이딥	itip
85.	아삿	asas
86.	우묵	umuk
87.	아겅	akʌŋ
88.	우잉	uiŋ
89.	아방	apaŋ
90.	아맙	amap
91.	아맛	amas
92.	우뭉	umuŋ
93.	우북	upuk
94.	어상	ʌsaŋ
95.	어응	ʌɨŋ
96.	아닷	atas
97.	이딤	itim
98.	아반	apan
99.	이밉	imip
100.	이민	imin
101.	아긍	akɨŋ
102.	오닝	oniŋ
103.	앙압	aŋap
104.	앙암	aŋam
105.	앙앗	aŋas
106.	아닥	atak
107.	이신	isin
108.	이정	iʧʌŋ
109.	아밍	amiŋ
110.	이심	isim
111.	우붐	upum
112.	아삼	asam
113.	우숙	usuk

114. 우순 usun
115. 울수 ulsu
116. 앵치 eŋʧhi
117. 앙안 aŋan
118. 우숫 usus
119. 아감 akam
120. 이빔 ipim
121. 이빗 ipis
122. 응기 ɨŋki
123. 옹밍 oŋmiŋ
124. 아낙 anak
125. 우붕 upuŋ
126. 웅운 uŋun
127. 아갑 akap
128. 우숭 usuŋ
129. 이농 inoŋ
130. 오벙 opʌŋ
131. 아단 atan
132. 앙악 aŋak
133. 아남 anam
134. 이빙 ipiŋ
135. 이식 isik
136. 아봉 apoŋ
137. 운낭 unnaŋ
138. 아낭 anaŋ
139. 아간 akan
140. 이십 isip
141. 이싯 isis
142. 웅움 uŋum
143. 우솔 usol
144. 이곤 ikon
145. 이빅 ipik
146. 이밍 imiŋ
147. 이밋 imis
148. 아갓 akas
149. 우붓 upus
150. 임모 immo
151. 아을 aɨl

152. 이딕 itik
153. 우궁 ukuŋ
154. 아밤 apam
155. 아낫 anas
156. 이김 ikim
157. 아밧 apas
158. 우걸 ukʌl
159. 응수 ɨŋsu
160. 이잘 iʧal
161. 우발 upal

Chapter 2
Frequency Factor: Lexical Frequency

2.1 Introduction

In reduplication, the choice of inserted consonant is not random. The data from the dictionary have shown that some consonants were more frequently inserted than others. In this chapter, I investigate the extent to which the choice of inserted consonant reflects the frequency of consonants in the lexicon. I will examine whether the frequency of specific inserted consonants reflects the frequency of those consonants in the lexicon (in all positions, in word-initial position, and in syllable onset). I also consider the effect of context: since the inserted consonant always occurred following another consonant and preceding a vowel, I consider whether the C – C and C – V combinations resulting from consonant insertion mirror the frequency of C – C and C – V combinations in the lexicon.

For data on the frequency of individual consonants and consonant-consonant and consonant-vowel sequences, I considered two corpora: the entire corpus (the *Sejong Balanced Corpus* 2007, compiled by the National Institute of the Korean Language) and the reduplication-only corpus (all reduplicated forms in the *Sejong* corpus). All the corpus data are based on the *Sejong* corpus, and if I come to include statistics from other sources like a dictionary, then I will mention it explicitly. The statistics from the corpora were tested against the results from the word creation experiment, to see whether speakers' behavior in the experiment reflected the patterns in the lexicon.

In discussing the frequency-related factors, I consider both type and token frequency. A dictionary provides type frequency, whereas a corpus can provide token frequency as well as type. Bybee (1995a) argues that type frequency plays a major role in morphological patterning, while connectionist models take into account both type and token frequency in capturing morphological well-formedness intuitions. Albright (2004) claims that type frequency is a pivot in building a model for morphological processes, and concludes that addition of token frequency does not improve the performance of a model. I consider both type and token frequency in regard to the lexical statistics of reduplication data; however, I will focus on token frequency when I look into the corpus I am utilizing in this chapter.

The specific questions I am asking are as follows:[19]

(1) Does the frequency of CI reflect the frequency of lexical Cs
  a. overall (in all positions)?
  b. in initial positions?
  c. as first C (= CI) of reduplicants?

(2) Does the frequency of CI – C combinations reflect the frequency of lexical C – C combinations?

(3) Does the frequency of CI – V combinations reflect the frequency of lexical C – V combinations?

Based on these questions with regard to lexical frequency, I compare the experimental results with the data from the entire corpus and the reduplication-only corpus, and I will argue that the pattern shown in the speakers' choice of CIs does not simply replicate the pattern in the existing lexical statistics.

2.2 Testing Hypotheses

(4) **General Hypothesis**
The patterns in the speakers' choice of CIs reflect those in the existing lexical statistics.

I break down this general hypothesis into more specified hypotheses in the subsequent sections, and test them based on the experimental results and the lexical patterns.

---

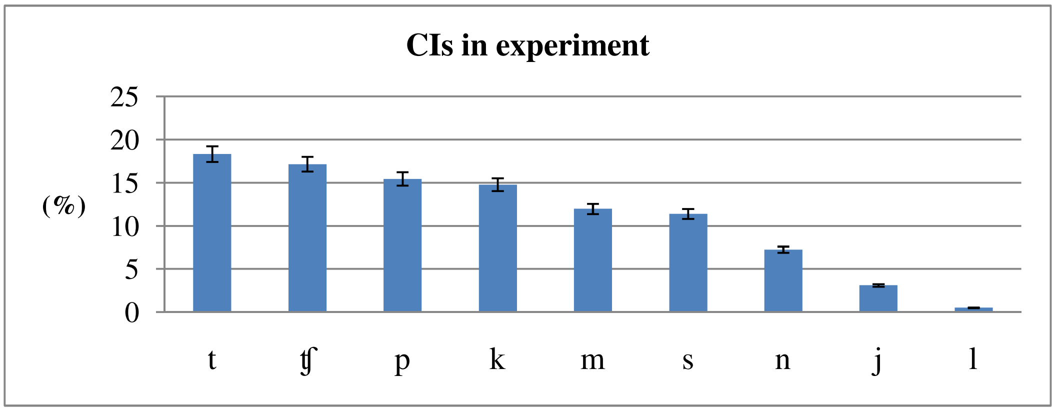
[19] I have not considered the directionality of combining patterns; that is, I examined whether the sequences of CI and the following Cs reflect the sequencing patterns of two Cs in the lexicon, but not whether the sequences of CI and the preceding Cs reflect the sequencing patterns of two Cs in the lexicon. This may be a legitimate question, but I reckon that investigating the C sequences in the rightward direction is more on the right track since the frequency pattern of CIs appears to follow that of word-initial Cs. If CIs were treated as word-initial Cs, then it would be more likely that CI and its following, rather than the preceding, Cs would be immediately relevant in terms of sequencing Cs. (cf. Section 2.2.1.2)

2.2.1 Examination of the entire corpus

2.2.1.1 CIs vs. overall lexical Cs

In the word creation task, not all consonants were equally likely to be chosen for insertion:

(5) Figure 2.1 CI frequency in the experiment (Experiment 1)[20]



What determines the choice of inserted consonant? In this section I consider the following hypothesis:

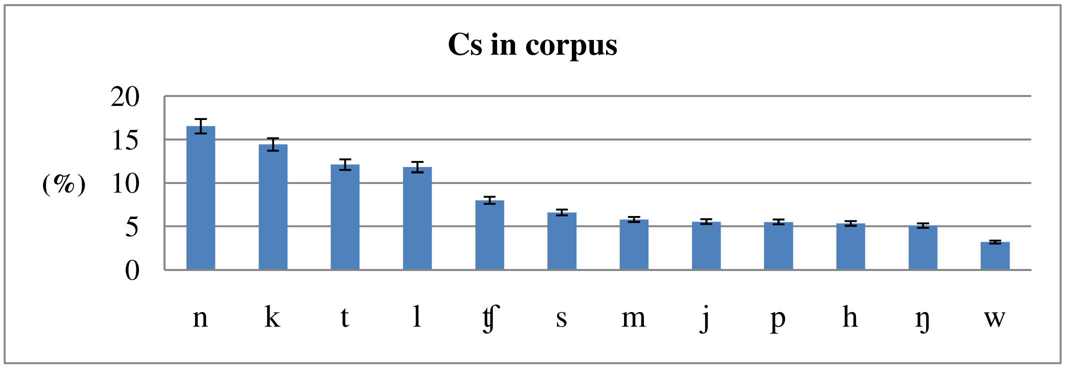
(6) **Hypothesis 1**
The frequency of CIs in the word creation task reflects the frequency of lexical Cs in the entire corpus.

If this is correct, we would expect that when speakers are forced to choose a consonant to insert, the choice of a consonant correlates with its frequency in general vocabulary.

I calculated the overall token frequency of each consonant in the *Sejong Balanced Corpus*. The frequency was computed on the basis of the total number of segments, including consonants and vowels. The percentage of consonants in the corpus amounts to 54.3%, which has been converted to 100% for comparison's sake.

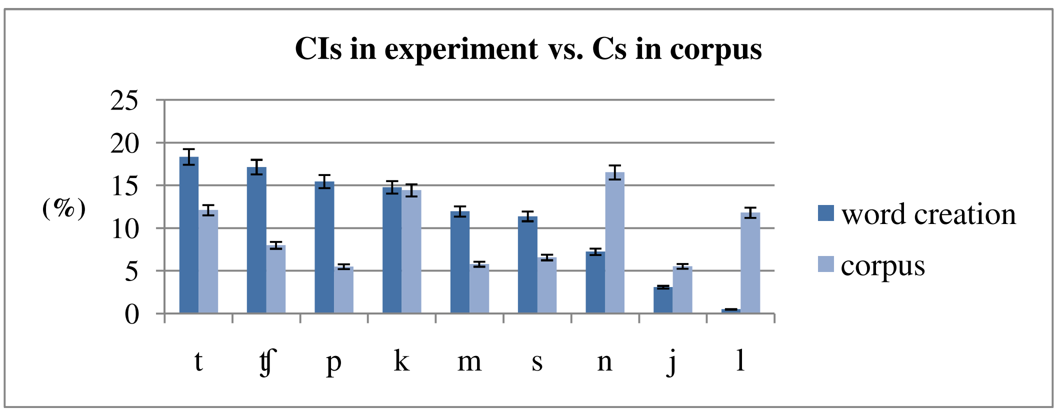
[20] Error bars represent 95% confidence interval of a mean.

(7) Figure 2.2 C frequency in the entire corpus



The comparison between the consonants appearing in the experiment with the counterparts in the corpus shows that the consonant frequencies in the corpus do not match the consonant insertion patterns in the reduplication. /n, k, t, l/ are among the most frequent in the corpus, whereas /t, ʧ, p, k/ were inserted the most frequently in the experiment.

(8) Figure 2.3 CI frequency in the experiment and C frequency in the entire corpus

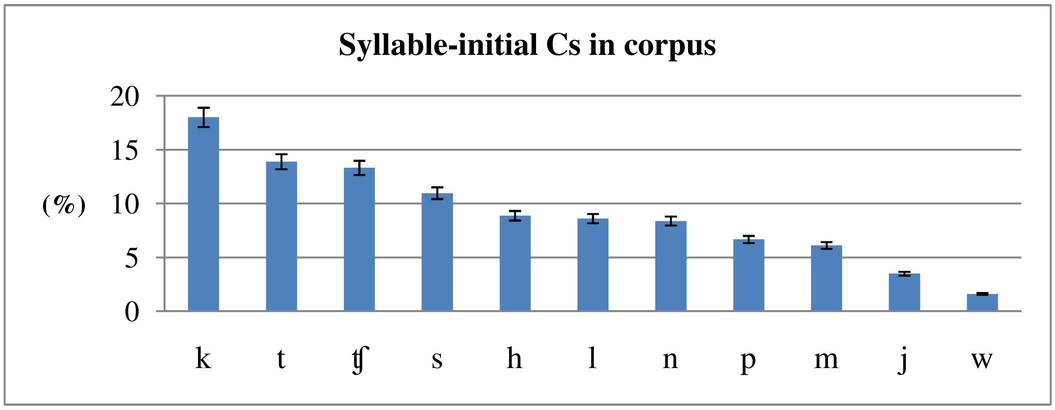


There was no significant correlation between the two patterns of CI frequency in the experiment and C frequency in the entire corpus. That is, the pattern of choosing CIs in the experiment does not replicate the pattern of C occurrences in the entire corpus, and Hypothesis 1 has not been confirmed.

2.2.1.2 CIs vs. lexical Cs in initial position

An examination of the corpus shows an asymmetry between consonant types in syllable-initial position:

(9) Figure 2.4 Frequency of onset Cs in the entire corpus



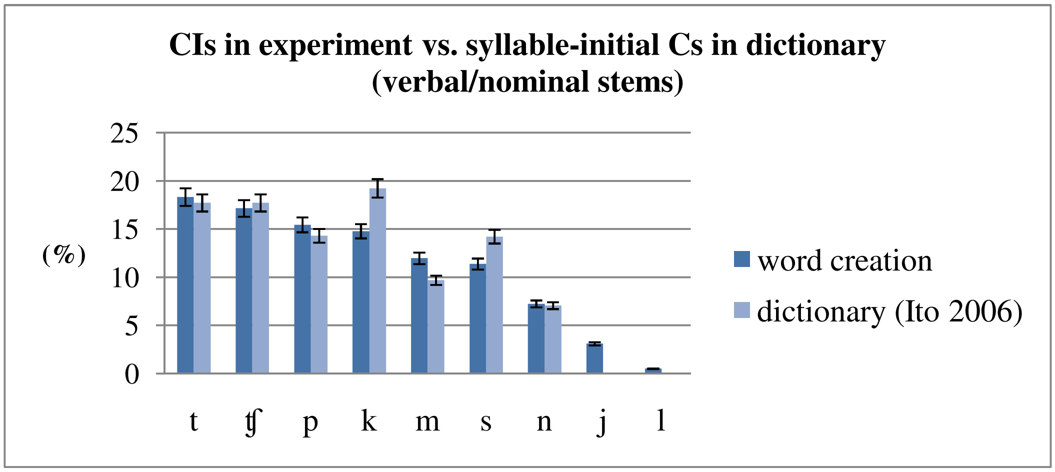
Since the CIs appear in initial position, e.g. of a syllable or a word, it is possible that the CIs are chosen with regard to the frequency of consonants in similar positions:

(10) **Hypothesis 2**
The frequency of CIs reflects the frequency of all lexical Cs in syllable-initial position.

According to Ito's (2006) research on Korean monosyllabic verbal/nominal stems, the attested onset consonants are /k, t, ʧ, p, s, m, n, h/ in the order of frequency, in which /k/ is the most frequent and /h/ is the least frequent, out of a total of 1298 tokens.[21] These results are not the same as those found in my word creation experiment, in which /t/ is the most frequent and /l/ is the least frequent, out of a set of /t, ʧ, p, k, m, s, n, j, l/.

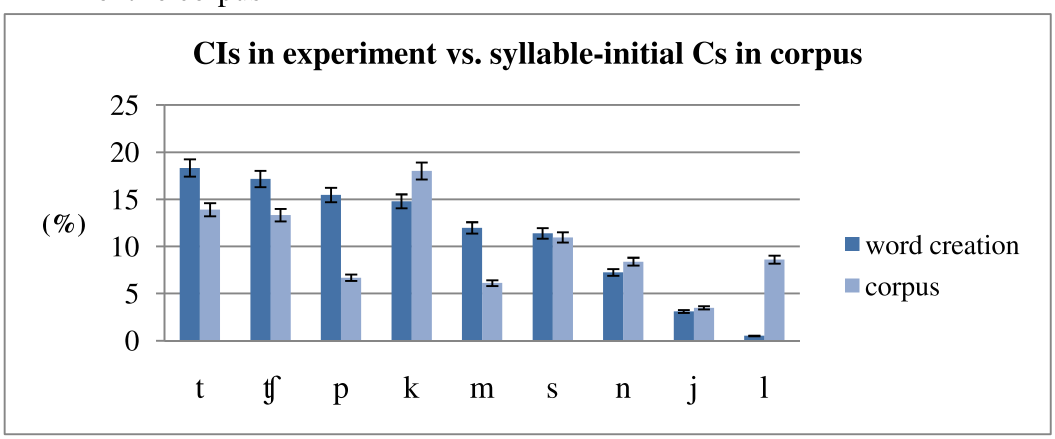
[21] Ito (2006) enumerated the frequencies for each of the Korean phonemes, but I collapsed the three series of lax, aspirated, and tense in order to make them consistent with my experimental responses in which I collapsed those three series. /k/ stands for /k, k^h, k'/, /t/ for /t, t^h, t'/, /ʧ/ for /ʧ, $ʧ^h$, ʧ'/, /p/ for /p, p^h, p'/, and /s/ for /s, s'/.

(11) Figure 2.5 Frequency of CIs in the experiment and onset Cs in the dictionary (Ito 2006)



This frequency was computed based on only onset Cs in the monosyllabic verbal/nominal stems. It may give a clue to all onset Cs in the lexicon; however, it is confined to monosyllabic words, verbal/nominal stems, and dictionary data. Thus I examine the frequency of syllable-initial Cs based on the entire corpus, not just the monosyllabic verbal/nominal stems.[22]

(12) Figure 2.6 Frequency of CIs in the experiment and that of onset Cs in the entire corpus



[22] There is no reason for choosing only verbal/nominal stems in considering whether the frequency of CIs reflects the frequency of onset Cs in the corpus, since normally reduplicative forms do not appear to have to do with verbal or nominal forms.

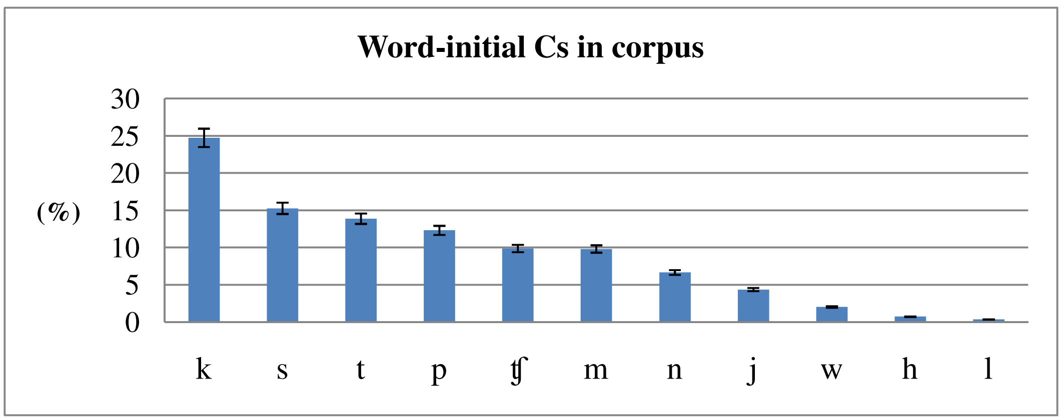
There was no significant correlation between the frequency pattern of CIs and the frequency pattern of syllable-initial Cs in the entire corpus, which means that these two patterns are not alike. Hypothesis 5 is not confirmed.

CIs may be considered to occur in word-initial position, and they can be collated with lexical Cs in word-initial position:

(13) **Hypothesis 3**
The frequency of CIs reflects the frequency of all lexical Cs in word-initial position.

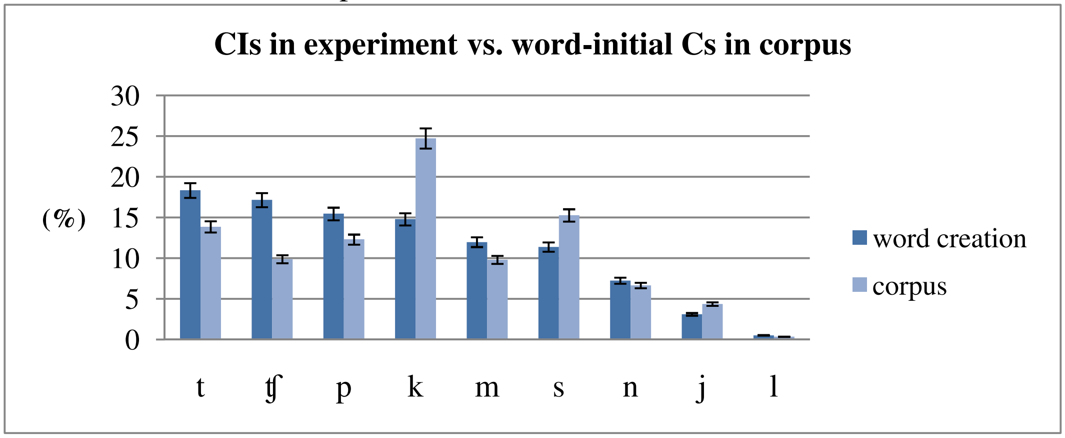
The frequency of word-initial Cs is different from that of syllable onset Cs in the lexicon.

(14) Figure 2.7 Word-initial C frequency in the entire corpus



If we compare the frequency of CIs in the word creation experiment with the frequency of all lexical Cs in word-initial position, we may end up seeing different results from what we saw in the comparison between the frequency of CIs and the frequency of Cs in syllable-initial position.

(15) Figure 2.8 Frequency patterns of CIs in the experiment and of word-initial Cs in the entire corpus



According to the correlation analysis, there was a positive relationship between the frequency of CIs and the frequency of word-initial Cs in the entire corpus, $r_s = .667$, $p < .05$.

(16) Table 2.1 Correlations: frequencies of CIs in the experiment (= expt) and word-initial Cs in the entire corpus

			expt	corpus
Spearman's rho	expt	Correlation Coefficient	1.000	.667(*)
		Sig. (2-tailed)	.	.050
		N	9	9
	corpus	Correlation Coefficient	.667(*)	1.000
		Sig. (2-tailed)	.050	.
		N	9	9

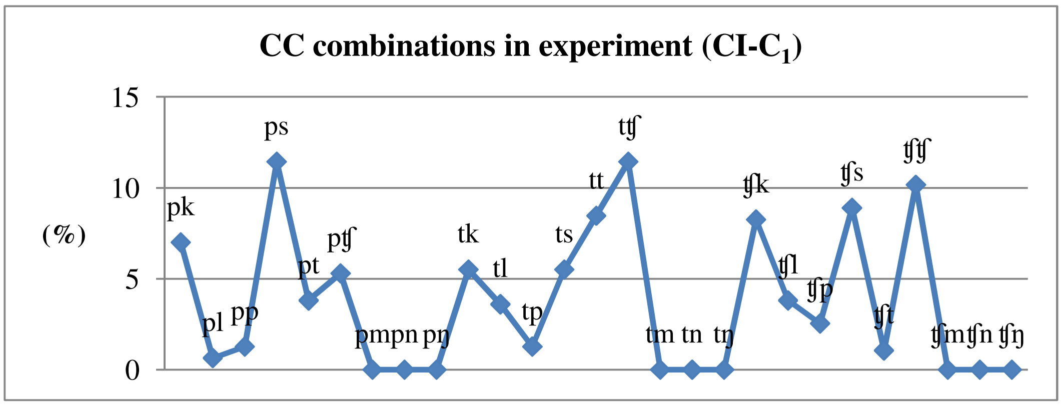
* Correlation is significant at the 0.05 level (2-tailed).

The frequency pattern of CIs in the experiment reflects that of word-initial Cs in the entire corpus. That is, the frequency pattern of CIs in the experiment is similar to that of Cs in word-initial position from the entire corpus, by which Hypothesis 3 has been confirmed. This outcome appears to be due to the fact that the reduplicants are assumed to be words, and the frequency pattern of reduplicant-initial Cs (= CIs) resembles that of word-initial Cs in the entire corpus.

2.2.1.3 CI – C vs. lexical C – C combinations

When a consonant is inserted, it does not occur in an isolated locus; it is rather surrounded by other segments. Thus its insertion is to be influenced by these surrounding sounds. For example, the CIs in the word creation experiment are shown to have co-occurred with certain consonants more often than others.

(17) Figure 2.9 CC combinations in the word creation experiment (VCVC-bases only, CI = /p, t, ʧ/): CI-C_1 combinations in VC_1VC_2-**CI**<u>VC_1VC_2</u>



(18) Figure 2.10 CC combinations in the word creation experiment (VCVC-bases only, CI = /p, t, ʧ/): CI-C_2 combinations in VC_1VC_2-**CI**<u>VC_1VC_2</u>

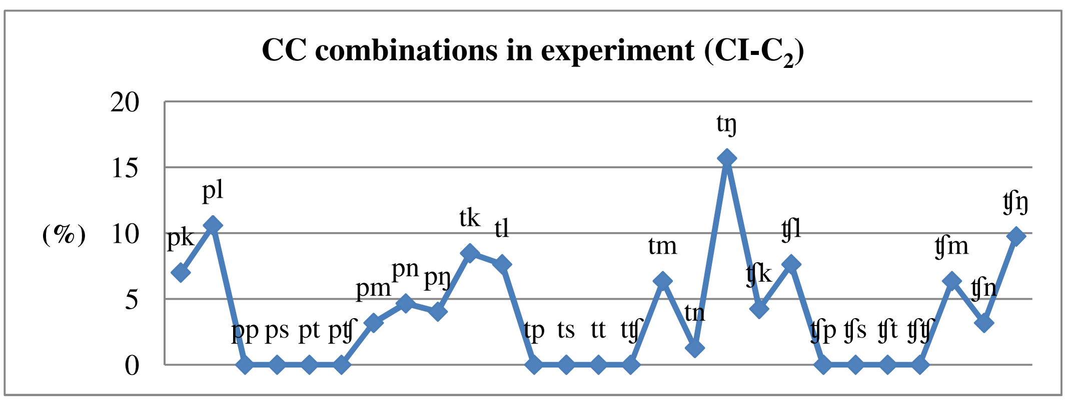
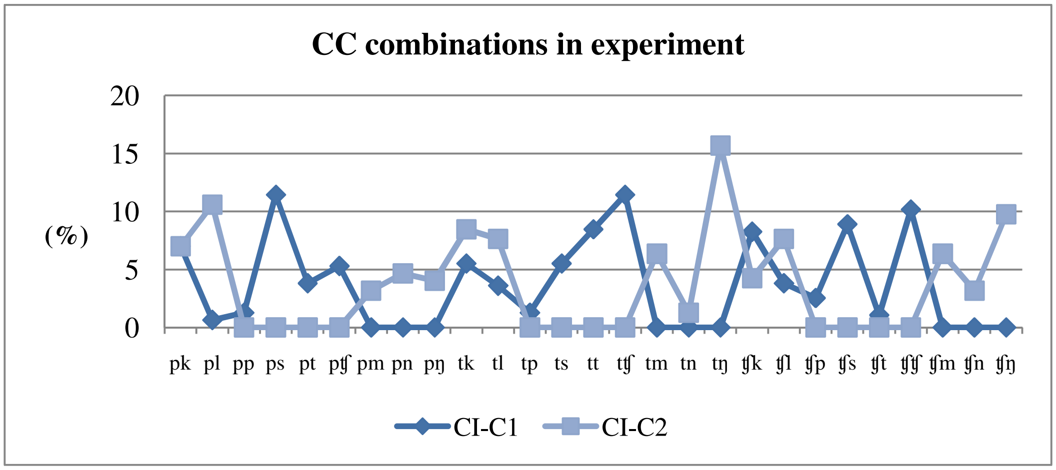


Figure 2.9 and 2.10 show that the patterns of CC combinations between CI and C_1 and those between CI and C_2 are distinct. For example, /ps/, /tʧ/, /ʧʧ/ show the highest frequencies in CI-C_1 combinations, whereas they do not occur at all in CI-

C_2 combinations.[23] The combined figure below shows this non-correspondence between CI-C_1 and CI-C_2 combinations.

(19) Figure 2.11 CC combinations in the word creation experiment (VCVC-bases only, CI = /t, p, ʧ/): CI-C_1 and CI-C_2 combinations in VC_1VC_2-**CI**VC_1VC_2



Note that for CI-C_1 it is either total identity or non-identity, whereas for CI-C_2 it is rather partial or total non-identity that is preferred.[24] For example, there are no combinations of identical CI-C_2, /pp, tt, ʧʧ/, whereas those identical combinations do occur (/pp/ = 1.27%) or they are relatively frequent (/tt/ = 8.47%, /ʧʧ/ = 10.17%) in the CI-C_1 pairs.[25]

(20) **Hypothesis 4**
The frequency of CI – C combinations reflects the frequency of lexical C and C combinations (CC or C.C) in the entire corpus.[26]

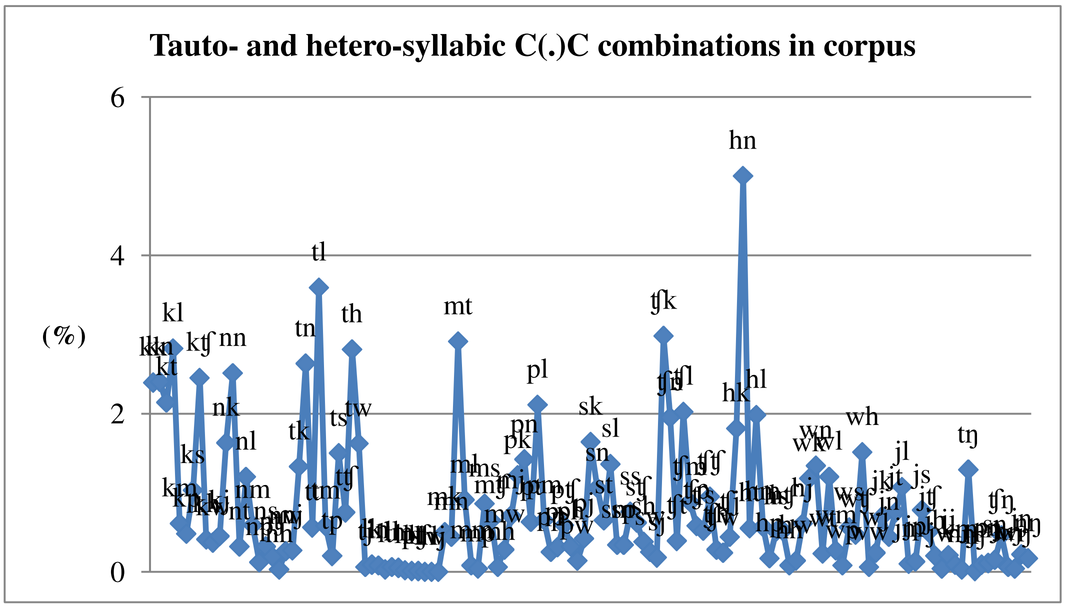
To see if the patterns of CI-C_1 and CI-C_2 combinations in the word creation experiment are actually from those in the corpus, I look into the C and C combinations in the corpus, without an intervening C, e.g. C and C in boldface in

[23] In fact, cooccurrences of /ps, tʧ, ʧʧ/ cannot exist for CI-C_2 since /s/ or /ʧ/ can never come in as C_2 which happens to be a coda; /s, ʧ/ do not occur in coda according to the Korean phonotactics.
[24] This issue will be further discussed in detail in Chapter 4.
[25] Again in the case of /ʧʧ/, we do not expect to see it for CI-C_2 since /ʧ/ is not allowed to occur in coda (C_2).
[26] The dot between Cs marks a syllable boundary.

a form like **C**(V)**C**(V)C… Two consonants can belong to a single syllable or distinct syllables. For the former case the first C is an onset and the second C is a coda or the two Cs make an onset cluster, called tauto-syllabic combination, and for the latter case the first C is an onset and the second C is also an onset, called hetero-syllabic combination.[27]

There are 95 possible CC pairs for tauto-syllabic combinations, and 133 possible CC pairs for hetero-syllabic combinations.[28] Out of these pairs, /hn, tl, ʧk, mt, kl, th/ are among the most frequent, as in the following figures.

(21) Figure 2.12 CC combinations in the entire corpus: Tauto-syllabic (**C**V**C**) and Hetero-syllabic (**C**V.**C**)



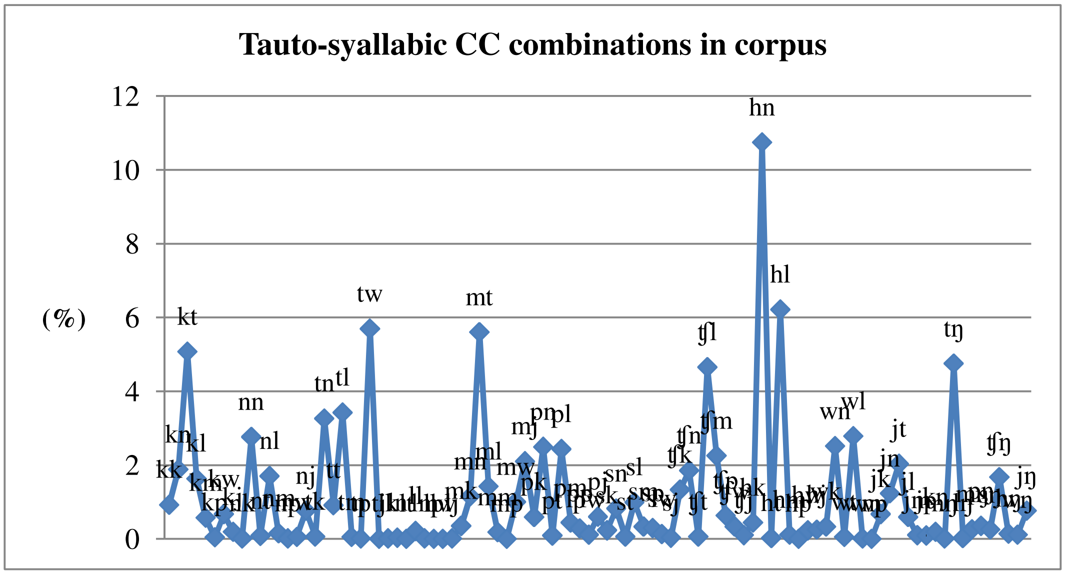
For tauto-syllabic CC combinations, it is the pairs /hn, hl, tw, mt, kt, tŋ, ʧl/ that are the most frequent. For hetero-syllabic C.C combinations, it is /th, tl, ʧk, kʧ, kl,

[27] A tauto-syllabic CC combination normally consists of onset C and coda C in Korean, but it can also be made up of onset Cs, e.g. Cj, Cw, in which approximants /w, j/ immediately follow a C. As for the status of these glides, it has been hard to define it in the literature of Korean linguistics, to the extent that they have been simply called vowels or considered secondary articulation (Suh 2007; An, Hwang, & Suh 2008).

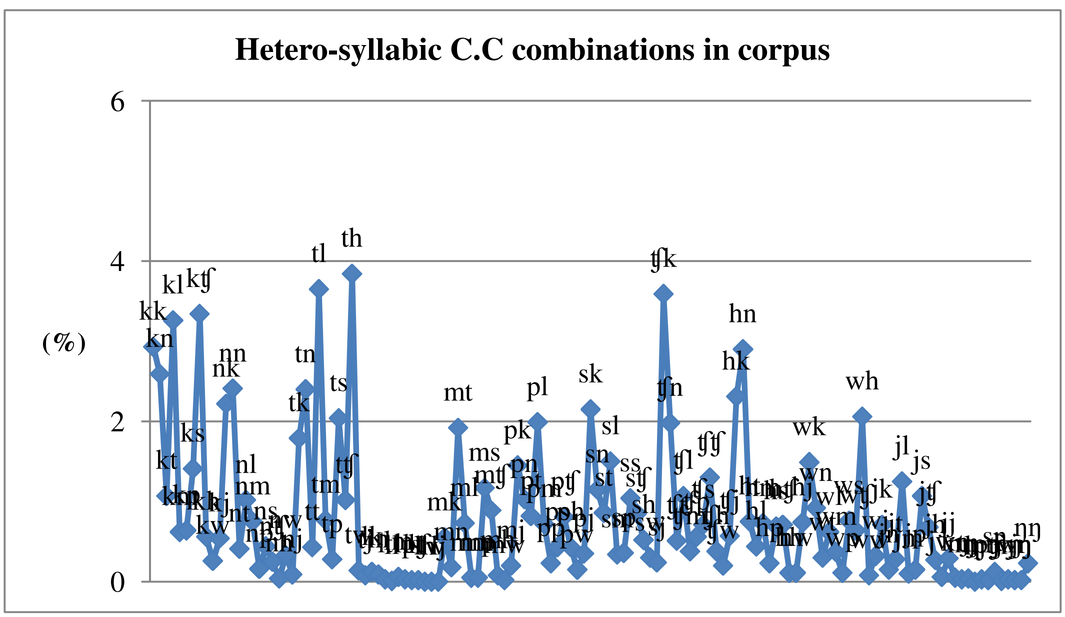
[28] The reason that there is a fewer number of tauto-syllabic pairs than hetero-syllabic pairs is that the coda /t, s, ʧ, h/ in a tauto-syllabic pair is neutralized to /t/. For example, /kt, ks, kʧ, kh/, in which the second consonant indicates coda C, have been merged with a combination of /kt/ in the study.

kk, hn/ that are the most frequent.

(22) Figure 2.13 CC combinations in the entire corpus: Tauto-syllabic (**C**V**C** or **CC**V)

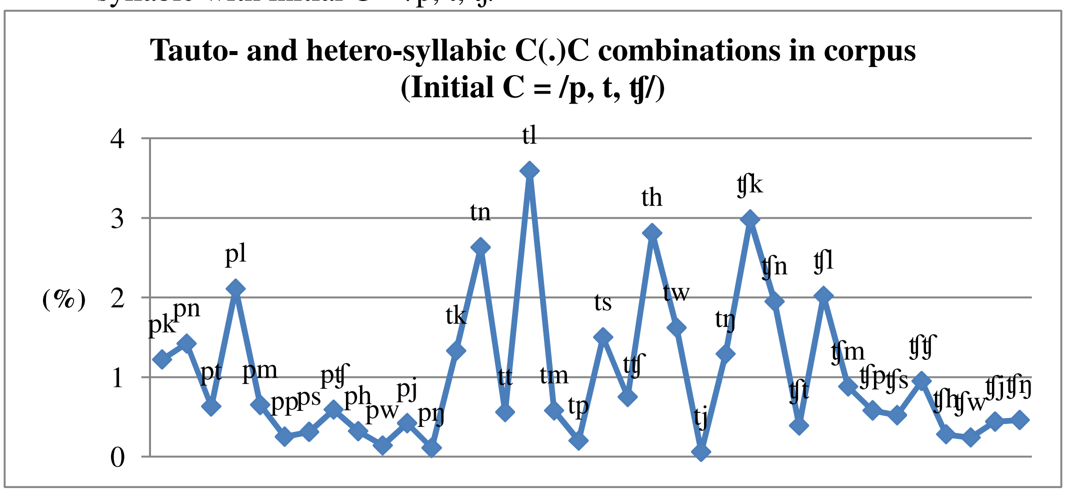


(23) Figure 2.14 CC combinations in the entire corpus: Hetero-syllabic (**C**V.**C**)



These figures however are not commensurable with the figures from the experiment since the experimental results looked into only the data of VCVC-**CI**VCVC forms, with CI = /p, t, ʧ/. Therefore, I consider the hetero-syllabic C.C combinations with initial C = /p, t, ʧ/ in a form of **C**V.**C**… so that I can compare them with CI-C_1 sequences in the reduplicant form of **CI**V.$\mathbf{C_1}$VC_2 from the experiment.

(24) Figure 2.15 C(.)C combinations in the entire corpus: Tauto- and hetero-syllabic with initial C = /p, t, ʧ/



Again, the combined figure of tauto- and hetero-syllabic CC combinations cannot show how these combinations in the corpus relate to CI-C_1 and CI-C_2 combinations in the experiment, and I break them down into tauto- and hetero-syllabic combinations as follows:

(25) Figure 2.16 CC combinations in the entire corpus: Tauto-syllabic with initial C = /p, t, ʧ/

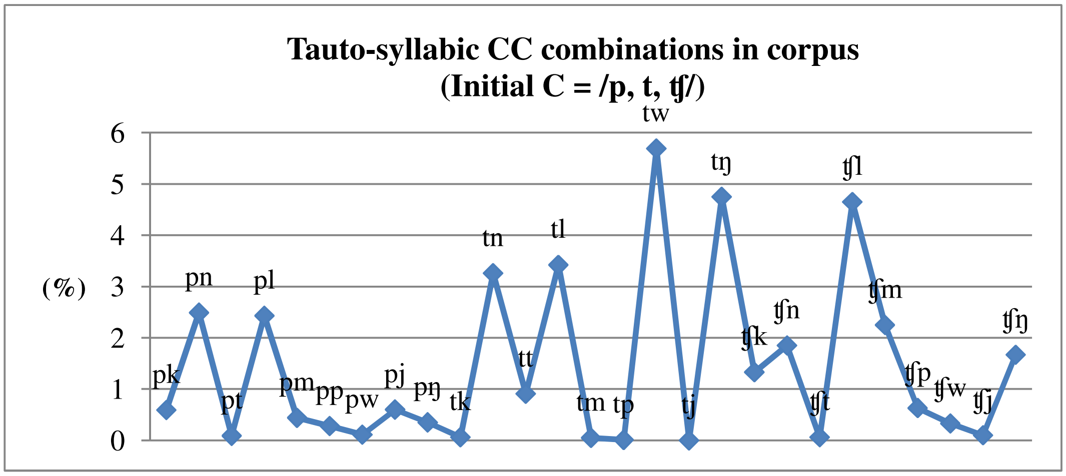
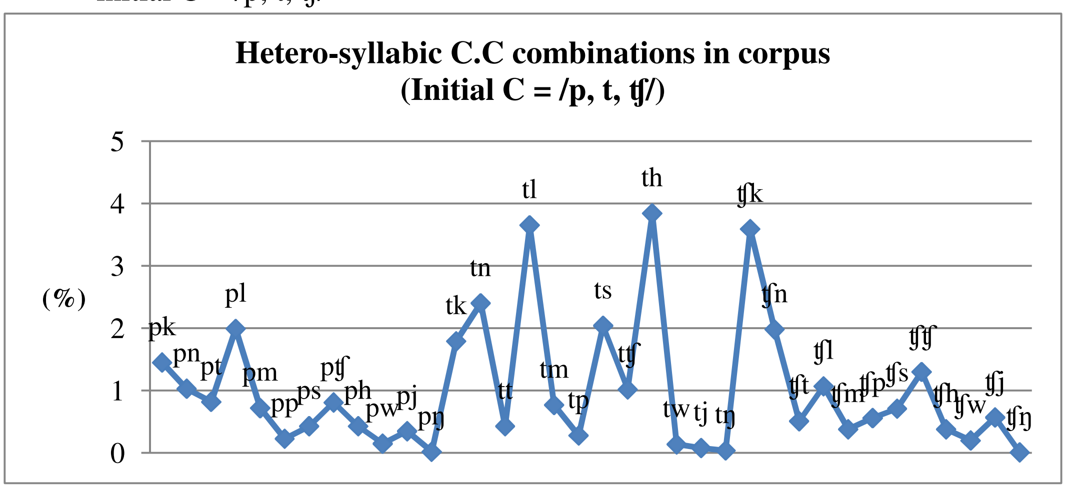


Figure 2.16 shows that among other tauto-syllabic CC combinations between /p, t, ʧ/ and other Cs, the combinations of /pn, tw, ʧl/ are the most frequent in the corpus.

(26) Figure 2.17 C.C combinations in the entire corpus: Hetero-syllabic with initial C = /p, t, ʧ/



The figure for hetero-syllabic C.C combinations show that it is /pl/, /tl, th/, and /ʧk/, respectively, which are the most frequent for each of initial Cs, /p, t, ʧ/. However, it has already been seen that it was /ps/, /tʧ/, and /ʧʧ/, respectively, which were found to be the most frequent in the CI-C_1 pairs with CI = /p, t, ʧ/, from the experiment. The frequencies of /pl/ and /tl/, in particular, in the experiment were much lower (0.64%, 3.6%, respectively) than that of /ps/ and /ttʃ/ (11.44% for each). Therefore, it appears that the experimental results in terms of C.C combinations do not reflect the C.C combination pattern from the entire corpus. A combined figure for all the C(.)C combinations in the experiment and those in the entire corpus, along with the statistics report, is given in the following:

(27) Figure 2.18 C(.)C combinations in the word creation experiment and in the entire corpus, with an initial C = /p, t, ʧ/

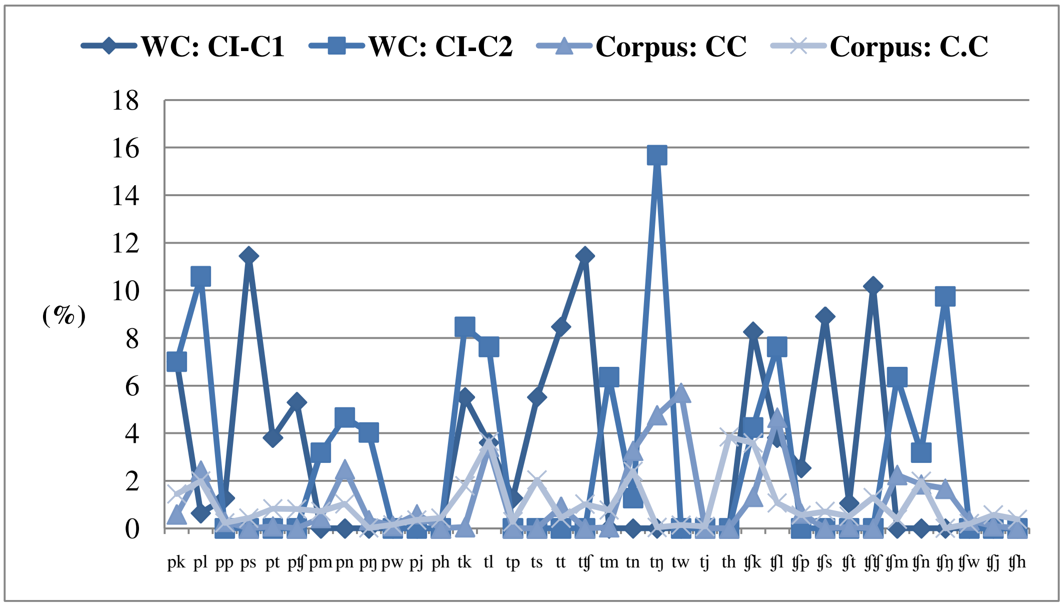


Figure 2.18 looks too complicated to figure out what happens in the relationships of these CC combinations, and the following statistical analysis can furnish information on whether C(.)C combinations in the experiment pattern with C(.)C combinations in the corpus.

(28) Table 2.2 Correlations: CC combinations of CI-C_1 and CI-C_2 in the experiment in a reduplicant form of **CI**VC_1VC_2 and CC combinations of tauto-syllabic and hetero-syllabic consonants in the entire corpus (in which tauto-syllabic CC means that two Cs are in the same syllable with one being an onset and the other being another onset (in rare cases) or a coda, and hetero-syllabic means that two Cs are onsets of adjacent syllables); Initial C = /p, t, ʧ/

			expt: CI-C_1	expt: CI-C_2	corpus: CC	corpus: C.C
Spearman's rho	expt: CI-C_1	Correlation Coefficient	1.000	-.137	-.293	.400(*)
		Sig. (2-tailed)	.	.424	.083	.016
		N	36	36	36	36
	expt: CI-C_2	Correlation Coefficient	-.137	1.000	.627(**)	.219
		Sig. (2-tailed)	.424	.	.000	.199
		N	36	36	36	36
	corpus: CC	Correlation Coefficient	-.293	.627(**)	1.000	.001
		Sig. (2-tailed)	.083	.000	.	.997
		N	36	36	36	36
	corpus: C.C	Correlation Coefficient	.400(*)	.219	.001	1.000
		Sig. (2-tailed)	.016	.199	.997	.
		N	36	36	36	36

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

There was a significant relationship between the patterns of CI-C_1 (in <u>**CI**V.C_1VC_2</u>) combinations in the experiment and heterosyllabic C.C combinations in the entire corpus $r_s = .400$, $p < .05$; between the patterns of CI-C_2 (in <u>**CI**V.C_1VC_2</u>) combinations in the experiment and tautosyllabic CC combinations in the entire corpus, $r_s = .627$, $p < .01$. Therefore, the frequency of CI-C combinations in the experiment reflects the frequency of C(.)C combinations in the entire corpus, and

Hypothesis 8 is confirmed.

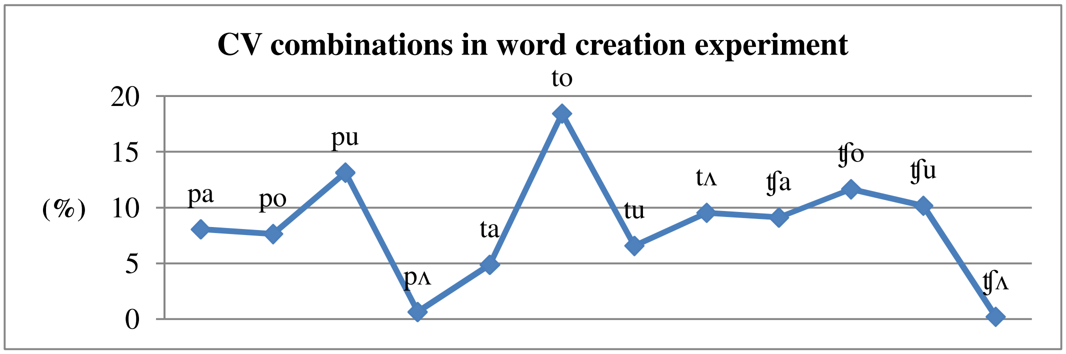
Note that the correlations between the patterns of CI-C_1 and C.C and of CI-C_2 and CC can be attributed to the characteristics of Korean, in which onset Cs combine with possible onset Cs {p, pʰ, p’, t, tʰ, t’, k, kʰ, k’, ʧ, ʧʰ, ʧ’, m, n, ŋ, s, s’, h, (w), l, (j)} on the one hand, and with possible coda Cs {p, t, k, m, n, ŋ, l} on the other. It can be inferred that the CI-C_1 combinations must pattern with C.C combinations since CI and C_1 are onsets of distinct syllables just as the two Cs in the C.C combinations are; also the CI-C_2 combinations pattern with CC combinations since CI is an onset and C_2 is a coda in the reduplicant form just as the first C in the CC combinations is an onset and the second C is a coda.[29] Therefore, it does not appear to be a matter of whether Cs are in a single syllable or in distinct syllables (tauto-syllabic vs. hetero-syllabic), but what seems to be relevant is whether two Cs are in the same syllabic positions or in different syllabic positions (both in onset vs. one in onset and the other in coda).

2.2.1.4 CI – V vs. lexical C – V combinations

If the combining pattern of CI with other consonants in reduplication is similar to the general combining pattern of consonants in lexicon, we can also postulate that the combining pattern of CI and a vowel in reduplication follows the combining pattern of consonant and vowel in the general lexicon. The following figure shows that the co-occurrence of consonant and vowel in the word creation task is not randomly distributed.

[29] Tauto-syllabic CC combinations in the corpus also included a combination like /tw/ which occur in an onset cluster, but other onset clusters like Cj and Cw were few except that the combination of /tw/ was among the most frequent in the corpus. Therefore, this may not have influenced the parallel between the tauto-syllabic (actually, one in onset and the other in coda) combination patterns in the corpus and in the experiment.

(29) Figure 2.19 CV combinations in the word creation experiment (VCVC-bases only, CI = /p, t, ʧ/, V = /a, o, u, ʌ/)



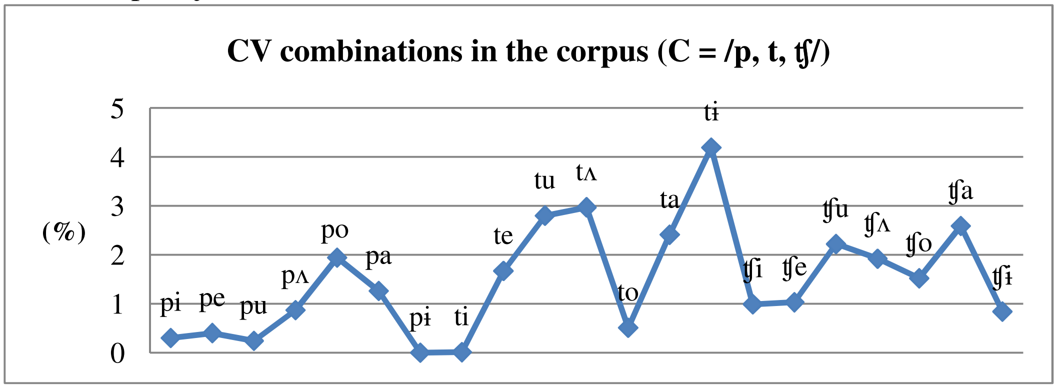
The CIs in the experiment were always followed by a vowel, and there may be some relationship between CI and its following V.

(30) **Hypothesis 5**
The frequency of CI – V combinations reflects the frequency of lexical C and V combinations in the entire corpus.

There can be preferred combinations of lexical Cs and Vs, which are to be reflected in the language users' behavior. The CIs we had in the experiment may reflect a more general preference for certain CV sequences in the lexicon. To test this hypothesis, I investigated the **CI**V sequences in the experiment data, and compare their combining pattern with the CV pattern in the lexicon. It is conceivable that a C and its following V influence each other articulatorily and acoustically, which may lead to preferred combinations of consonants and vowels.

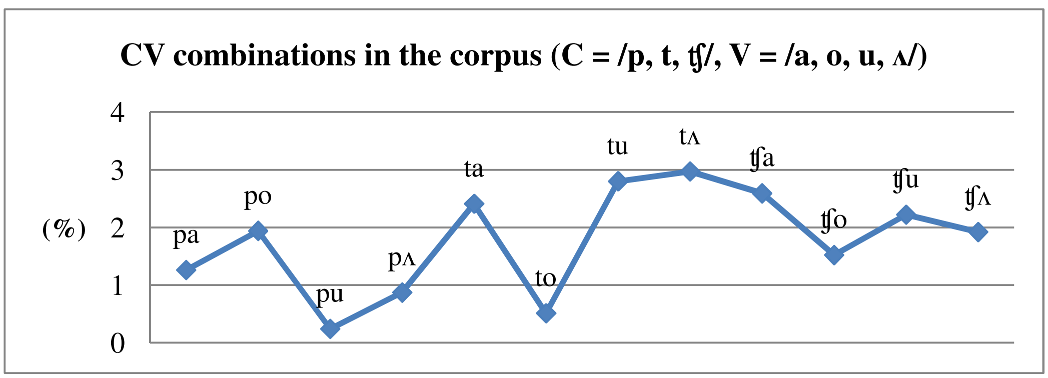
Since I have been focusing on the reduplicated forms with VCVC-bases and CIs of /p, t, ʧ/, I also look into the same data for CV combinations. According to Figure 2.19, /p/ was followed by /u/ most frequently, /t/ by /o/, and /ʧ/ by /o/ in the experiment. However, the frequency pattern of CV combinations in the entire corpus shows a different pattern:

(31) Figure 2.20 CV combinations in the entire corpus (VCVC-bases only, CI = /p, t, ʧ/)



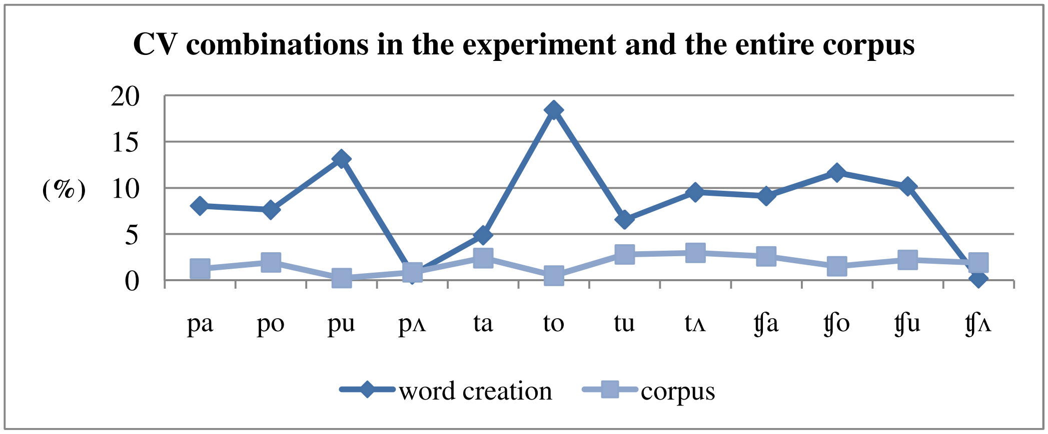
In fact, Figure 2.20 shows the CV combinations in which Vs include the ones which do not appear in the case of word creation experiment. Thus I only consider the Vs, in the CV combinations, which did show up in the experiment, as follows:

(32) Figure 2.21 CV combinations in the entire corpus (VCVC-bases only, CI = /p, t, ʧ/, V = /a, o, u, ʌ/)



What the corresponding CV combinations from the word creation experiment and the corpus show is that the CV combining pattern in the word creation does not reflect that in the entire corpus, as in the following:

(33) Figure 2.22 CV combinations in the word creation experiment and the entire corpus (VCVC-bases only, CI = /p, t, ʧ/, V = /a, o, u, ʌ/)



As can be seen in Figure 2.22, there were no significant correlations between the CV combining patterns from the experiment and from the entire corpus. Hence Hypothesis 5 is not confirmed.

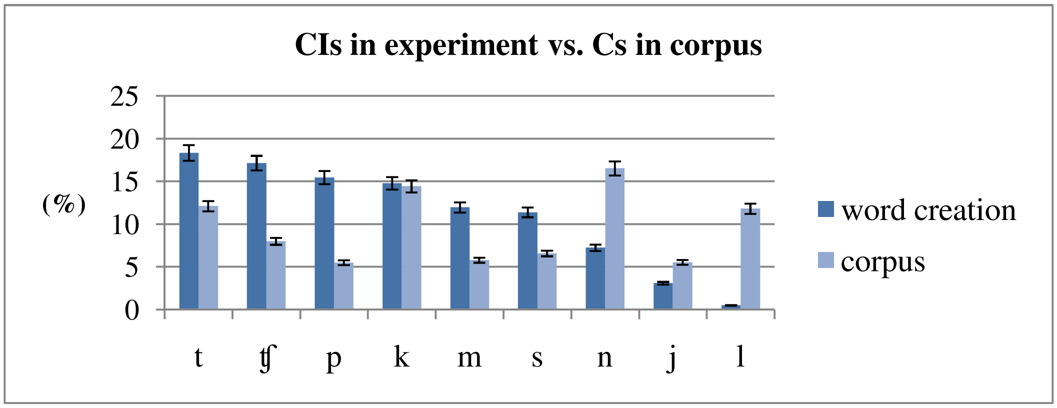
2.2.2 Examination of the reduplication-only corpus

Not every hypothesis presented in the preceding section based on the entire corpus has been confirmed, which gives rise to room for consideration of other possibilities: I will take account of a narrower corpus, reduplication-only corpus, in this section to see if the experimental results actually reflect the statistics from the corpus of reduplicative forms.

2.2.2.1 CIs vs. overall lexical Cs

The unequal chances of different consonants to be chosen as a CI in the word creation experiment imply a possibility that they come from a source in lexicon, which however turned out that they are not originated in the frequency pattern of all lexical Cs in the entire corpus. Therefore, we can look into a specific corpus, reduplication-only corpus, which may be a plausible corpus to examine since the experiment itself consisted of creating reduplicative words.

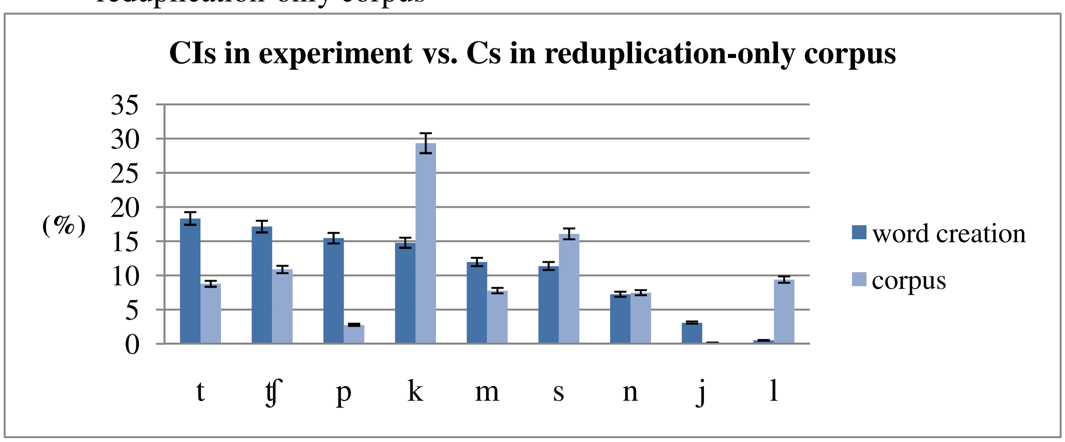
(34) Figure 2.23 CI frequency in the experiment and C frequency in the entire corpus



(35) **Hypothesis 6**
The frequency of CIs in the word creation experiment reflects the frequency of lexical Cs in the reduplication-only corpus.[30]

The frequency pattern of CIs in the experiment is collated with the frequency of lexical Cs in the reduplication-only corpus, as in the following:

(36) Figure 2.24 CI frequency in the experiment and C frequency in the reduplication-only corpus



As Figure 2.24 demonstrates, the frequency pattern of CIs in the experiment does

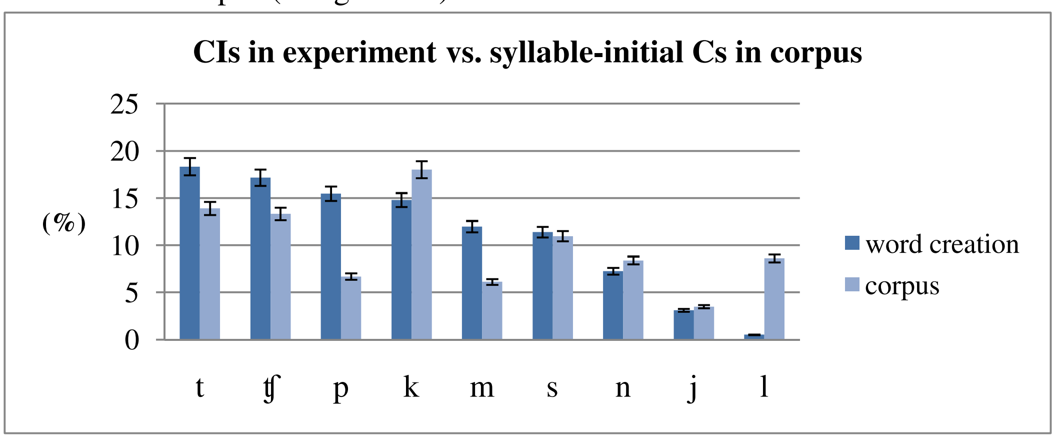
[30] For the lexical Cs, the reduplication-only corpus includes reduplicative forms with V-initial and C-initial bases.

not reflect the frequency pattern of Cs in the reduplication-only corpus. For example, /k/ is just one of the most frequent Cs in the experiment, whereas it is the most frequent C, undefeatable by any other consonants, in the reduplication-only corpus. Therefore, the frequency pattern of choosing CIs in the experiment does not replicate the frequency pattern of C occurrences in the reduplication-only corpus, and Hypothesis 6 has not been confirmed.

2.2.2.2 CIs vs. lexical Cs in initial position

Since CIs themselves occur in onset of a syllable, the frequency pattern of CIs in the word creation task is predicted to follow the frequency pattern of lexical onset Cs in the corpus: The investigation of the entire corpus with respect to the lexical frequency of onset Cs showed that there was no correlation between the frequency pattern of CIs in the experiment and the frequency pattern of lexical onset Cs in the entire corpus.

(37) Figure 2.25 Frequency of CIs in the experiment and that of onset Cs in the entire corpus (= Figure 2.6)



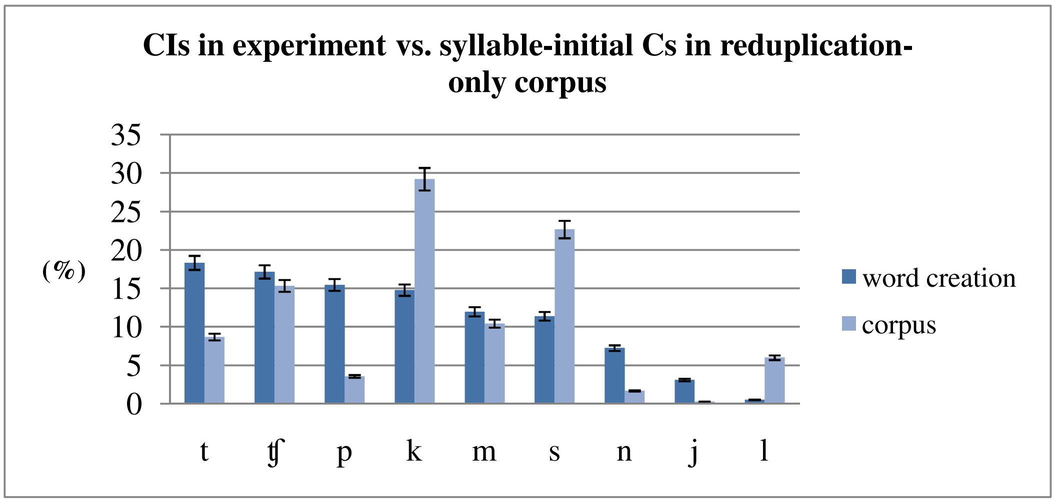
We can see if this is a matter of a size of lexicon: I will consider a smaller-scale corpus, reduplication-only corpus, in this section, and examine whether the frequency of CIs in the experiment reflects the frequency of lexical Cs in syllable onset in the reduplication-only corpus.

(38) **Hypothesis 7**
The frequency of CIs in the word creation experiment reflects the frequency of lexical Cs in syllable-initial position in the reduplication-only

corpus.[31]

The frequency pattern of CIs in the experiment and the frequency pattern of onset Cs in the reduplication-only corpus are given together in the following:

(39) Figure 2.26 CI frequency in the experiment and onset C frequency in the reduplication-only corpus



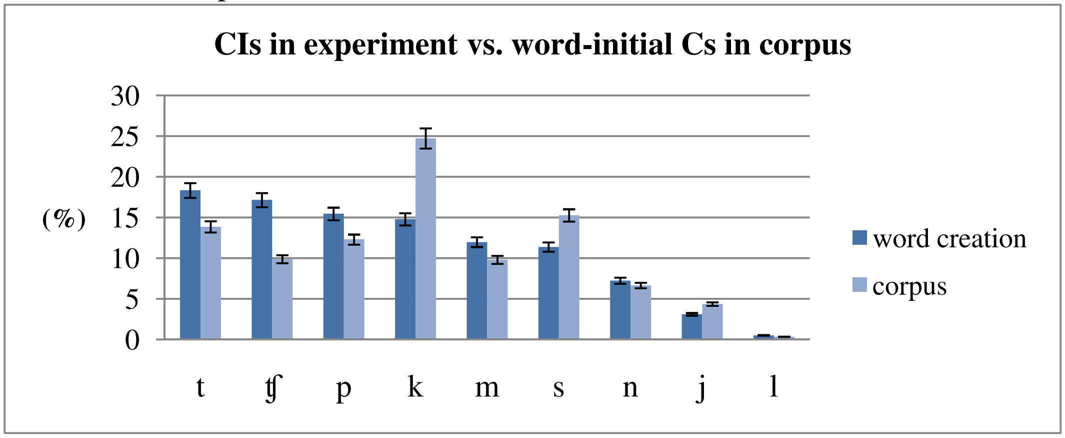
There was no significant relationship between the frequency of CIs in the word creation experiment and the frequency of syllable onset Cs in the reduplication-only corpus. That is, the frequency pattern of CIs is not similar to the frequency pattern of syllable onset Cs in the reduplication-only corpus, and they are patterning differently. Thus Hypothesis 7 is not confirmed.

There have been correlations between the CI frequency in the experiment and the word-initial C frequency in the entire corpus, as was seen in a preceding section:

---

[31] For the lexical Cs in syllable onset position, the reduplication-only corpus includes reduplicative forms with V-initial and C-initial bases.

(40) Figure 2.27 Frequency of CIs in the experiment and word-initial Cs in the entire corpus



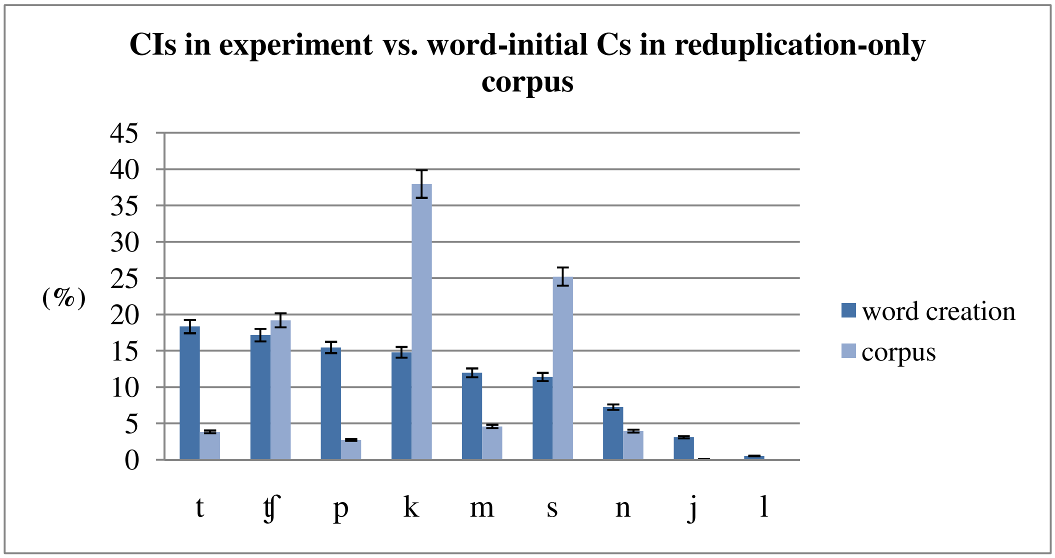
This means that CIs in the experiment are more likely to be counted as word-initial Cs; we further wonder if this assumption holds in any corpus, regardless of the size of corpus. That is, we wonder whether CIs are also considered word-initial Cs in a smaller corpus, reduplication-only corpus, not just in the entire corpus.

(41) **Hypothesis 8**
The frequency of CIs in the word creation task reflects the frequency of lexical Cs in word-initial position in the reduplication-only corpus.[32]

The frequency of CIs in the word creation experiment is shown with the frequency of Cs in word-initial position in the reduplication-only corpus to check for their correlations in patterning:

[32] For the lexical Cs in word-initial position, the reduplication-only corpus includes reduplicative forms with C-initial bases.

(42) Figure 2.28 CI frequency in the experiment and word-initial C frequency in the reduplication-only corpus



The frequencies of /k/ and /s/ are greater than other consonants in word-initial position, both in the entire and the reduplication-only corpus. Furthermore, other word-initial Cs, except for /ʧ/, in the reduplication-only corpus are rarely attested. Speakers' behavior in the choice of CIs is not restricted to the lexicon of reduplicative words only: there was no significant relationship in the frequency pattern of CIs in the word creation experiment and the frequency pattern of word-initial Cs in the reduplication-only corpus. That is, these two patterns were dissimilar and Hypothesis 8 has not been confirmed.

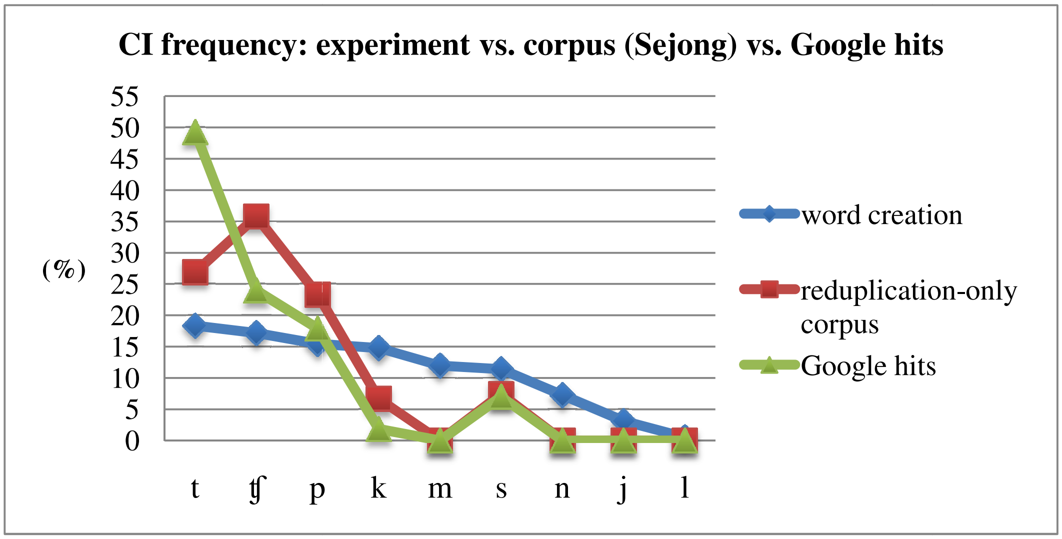
Since CIs occur reduplicant-initially, we can also check the frequency of CIs in the experiment with Cs in reduplicant-initial positions in the reduplication-only corpus.

(43) **Hypothesis 9**

The frequency of CIs in the experiment reflects the frequency of lexical Cs in reduplicant-initial position in the reduplication-only corpus.

To test this hypothesis, I investigated both the reduplicative forms in the *Sejong* corpus and the Google search engine, which is provided as an additional source. I looked into the reduplicated forms with a V-initial base in the *Sejong* corpus and the Google search, because only reduplicative forms with a V-initial base can have a CI in reduplicant.

(44) Figure 2.29 CI frequencies in the experiment and in the reduplication corpora (*Sejong* and Google)



The frequencies are based on 1352, 528, and 6465062 tokens of CI from the word creation experiment, *Sejong* corpus, and Google hits, respectively. The Google search was conducted on April 24, 2008. The frequency patterns shown by the two corpora are very similar, except for the inverse frequency pattern between /t/ and /ʧ/. This inverse pattern in the corpora could be a source for the almost equal frequencies between /t/ and /ʧ/ in the experiment. Since /t/ and /ʧ/ are highly frequent in both the corpora, and one is more frequent in one corpus and the other is more frequent in another, speakers may have about equal preferences for the two Cs in a real-time experiment.

Figure 2.29 shows that the CI frequency pattern in the experiment seems distinct from the C frequency in reduplicant-initial position in the reduplication-only corpora (*Sejong* and Google); however, statistically there were correlations among the frequency patterns of the three sources: between the word creation experiment and the *Sejong* corpus, $r_s = .862$, $p < .01$; between the word creation experiment and the Google hits, $r_s = .921$, $p < .01$; between the *Sejong* corpus and the Google hits, $r_s = .944$, $p < .01$. These three-way correlations indicate that the three frequency patterns from the different sources are similar, and the frequency pattern of CIs in the experiment mirrors the frequency pattern of Cs in reduplicant-initial position in the corpora. This suggests that there are some Cs, as opposed to others which are preferred in reduplicant-initial position. Hypothesis 7 has been confirmed.

(45) Table 2.3 Correlations: CI frequency in the experiment and C frequency in reduplicant-initial position in the *Sejong* corpus and the Google search

			expt	*Sejong*	Google
Spearman's rho	expt	Correlation Coefficient	1.000	.862(**)	.921(**)
		Sig. (2-tailed)	.	.003	.000
		N	9	9	9
	Sejong	Correlation Coefficient	.862(**)	1.000	.944(**)
		Sig. (2-tailed)	.003	.	.000
		N	9	9	9
	Google	Correlation Coefficient	.921(**)	.944(**)	1.000
		Sig. (2-tailed)	.000	.000	.
		N	9	9	9

** Correlation is significant at the 0.01 level (2-tailed).

The data have not considered different laryngeal features on consonants, wherefore /t, t', t^h/ were all categorized as /t/, /p, p', p^h/ as /p/, /k, k', k^h/ as /k/, /ʧ, ʧ', $ʧ^h$/ as /ʧ/, and /s, s'/ as /s/. There are only few cases of the tense and aspirated consonants as a CI, which made them put together with the lenis series. However, we can still see if the results are the same when these laryngeal consonants are taken into consideration.

(46) Figure 2.30 CI frequency in the experiment and C frequency in reduplicant-initial position in the reduplication-only corpora (*Sejong* and Google), in which the laryngeal consonants were separated out in the data

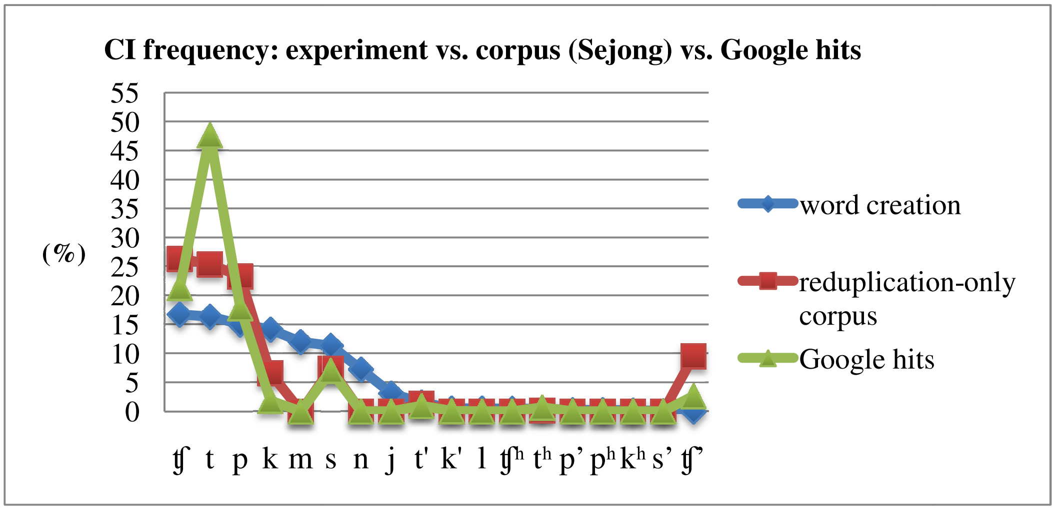


Figure 2.30 also shows that the CI frequency pattern in the experiment is distinct only in some Cs like /m/ and /n/, from the frequency pattern of Cs in reduplicant-initial position in the reduplication-only corpora (*Sejong* and Google).

(47) Table 2.4 Correlations: CI frequency in the experiment and frequency of Cs in reduplicant-initial position in the reduplication-only corpora (*Sejong*, and Google), with laryngeal consonants separated out

			expt	Sejong	Google
Spearman's rho	expt	Correlation Coefficient	1.000	.540(*)	.596(**)
		Sig. (2-tailed)	.	.021	.009
		N	18	18	18
	Sejong	Correlation Coefficient	.540(*)	1.000	.917(**)
		Sig. (2-tailed)	.021	.	.000
		N	18	18	18
	Google	Correlation Coefficient	.596(**)	.917(**)	1.000
		Sig. (2-tailed)	.009	.000	.
		N	18	18	18

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

As has been the case with Figure 2.29, the statistical analysis reports that there are significant correlations among the three patterns (between the word creation experiment and the *Sejong* corpus, $r_s = .540$, $p < .05$; between the word creation experiment and the Google hits, $r_s = .596$, $p < .01$; between the *Sejong* corpus and the Google hits, $r_s = .917$, $p < .01$). Hence with or without considering the laryngeally specified consonants into separate categories, there were significant relationships among the patterns from the different sources. It means that the frequency pattern of CIs in the experiment does reflect the frequency pattern of Cs in reduplicant-initial position in the reduplication-only corpora (*Sejong* and Google).

2.2.2.3 CI – C vs. lexical C – C combinations

The inserted consonant does not occur alone, but it occurs in context. The lexical statistics for the occurrences of a single C can provide a basis for the speakers' choice of CIs in reduplication; in addition, the lexical statistics for the co-occurrences of C and C can serve as a good source for the CI choice behavior. The C – C combining pattern in the lexicon was found to be correlated with the CI – C combining pattern in the experiment (Section 2.2.1.3): in particular, correlations between CI – C_1 (both in onset) and C – C (both in onset); correlations between CI – C_2 (CI in onset and C_2 in coda) and C – C (first C in onset and second C in coda), which appeared to be due to the characteristics of the Korean phonotactics in which there are Cs that can occur in onset and Cs that can occur in coda. All Cs can come in onset, except for /ŋ/, and only seven kinds of Cs /t, p, k, m, n, ŋ, l/ can occur in coda.[33]

We wonder if this is the case with the smaller corpus, reduplication-only corpus: if the correlations that were found in the patterns of C – C co-occurrences in the experiment and in the entire corpus should also be found in the pattern of C – C co-occurrences in the experiment and in the reduplication-only corpus, then it may mean that there is nothing particular about the reduplication-only corpus: the reduplication-only corpus is only a subset of the entire corpus, which has all the properties of the entire corpus. However, if no correlations should be found in the C – C co-occurrence patterns in the experiment and in the reduplication-only corpus, then the reduplication-only corpus itself may have some unique lexical properties, distinct from the entire corpus, but the speakers in the word creation task do not follow those lexical properties of the reduplication-only corpus.

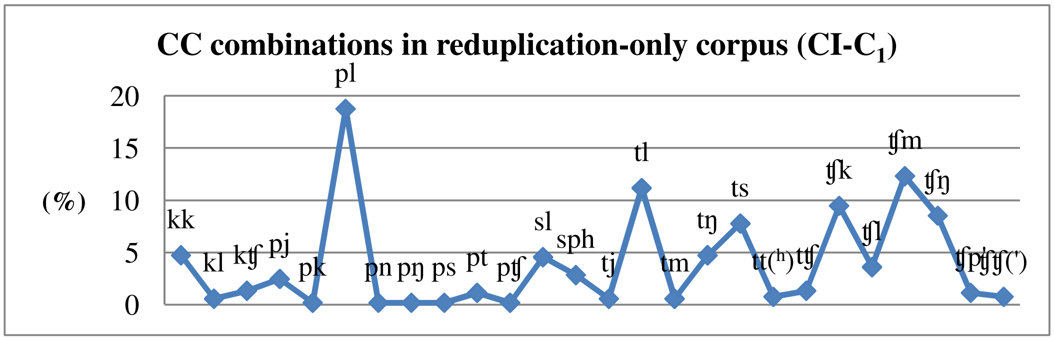
(48) **Hypothesis 10**

The frequency of CI – C combinations reflects the frequency of lexical C and C combinations in the reduplication-only corpus.

Regarding the reduplication-only corpus, I narrow it down to a reduplication-only corpus which contains reduplicated forms with a CI, in which case I can compare the combining patterns of C and C from the experiment and from the reduplication-only corpus on the same ground, i.e. CI – C_1 in the experiment vs. CI – C_1 in the reduplication-only corpus; CI – C_2 in the experiment vs. CI – C_2 in the reduplication-only corpus.

[33] However, this still does not explain why certain combinations of C – C are frequent than other combinations, which will be taken up in Chapter 5.

(49) Figure 2.31 CC combinations in the reduplication-only corpus: CI-C_1



(50) Figure 2.32 CC combinations in the reduplication-only corpus: CI-C_2

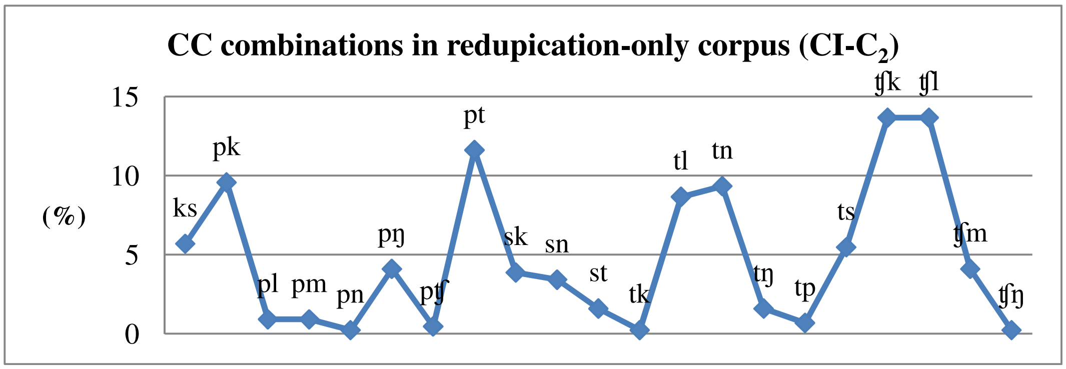
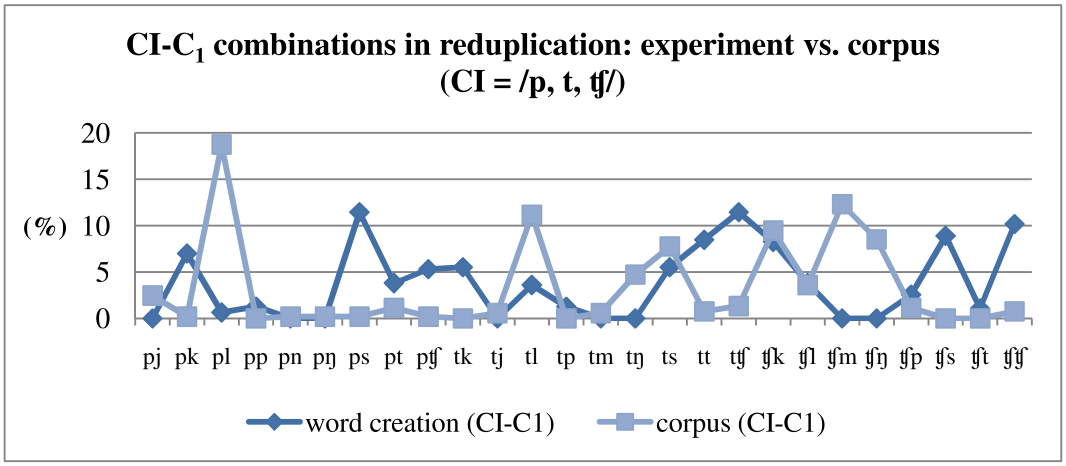


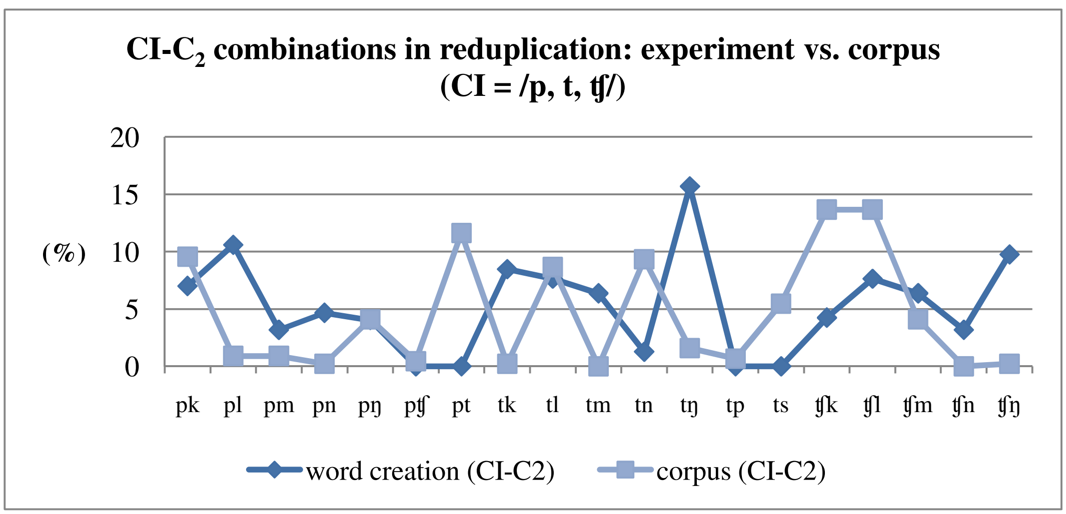
Figure 2.31 and 2.32 show the C and C combinations of CI-C_1 and CI-C_2 in the reduplication-only corpus, and I compare the CI – C combination patterns from the experiment and from the reduplication-only corpus, when the initial C is /p, t, ʧ/ as in the following figures.

(51) Figure 2.33 CC combinations in the experiment and the reduplication-only corpus: CI-C_1 combinations with CI = /p, t, ʧ/



There was no significant relationship between the two frequency patterns of CI-C_1 combinations in the work creation task and the reduplication-only corpus.

(52) Figure 2.34 CC combinations from the experiment and the reduplication-only corpus: CI-C_2 combinations with CI = /p, t, ʧ/

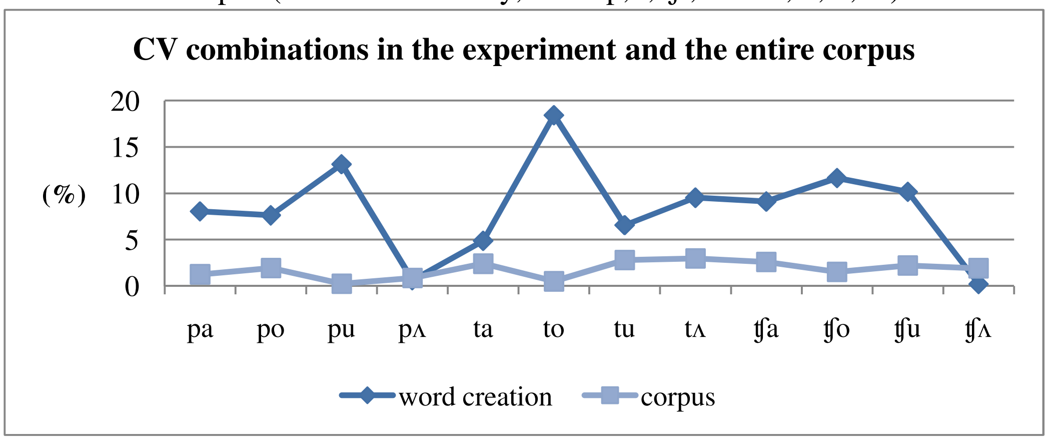


There was no significant relationship between the two frequency patterns of CI-C_2 combinations. The discrepancy in the patterns of C – C combinations in the experiment and the reduplication-only corpus espouses that the frequency of combining C and C in the experiment is not simply copying the lexical frequency of combining C and C in the reduplication-only corpus. Hypothesis 10 is not confirmed.

2.2.2.4 CI – V vs. lexical C – V combinations

There were no correlations between C – V combining patterns in the experiment and in the entire corpus (section 2.2.1.4):

(53) Figure 2.35 CV combinations in the word creation experiment and the entire corpus (VCVC-bases only, CI = /p, t, ʧ/, V = /a, o, u, ʌ/)



If a size of corpus served as a variable, a corpus other than the entire corpus could be examined. The reduplication-only corpus is considered for a C – V combining pattern, which will be compared with a C – V combining pattern in the word creation experiment.

(54) **Hypothesis 11**
The frequency of CI – V combinations reflects the frequency of lexical C and V combinations in the reduplication-only corpus.[34]

[34] For the lexical C and V combinations, the reduplication-only corpus includes reduplicative forms with V-initial bases only.

(55) Figure 2.36 CV combinations in the reduplication-only corpus

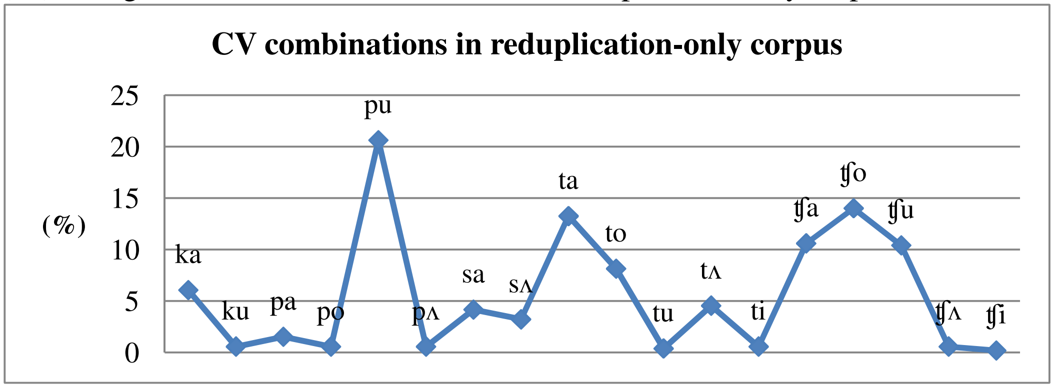
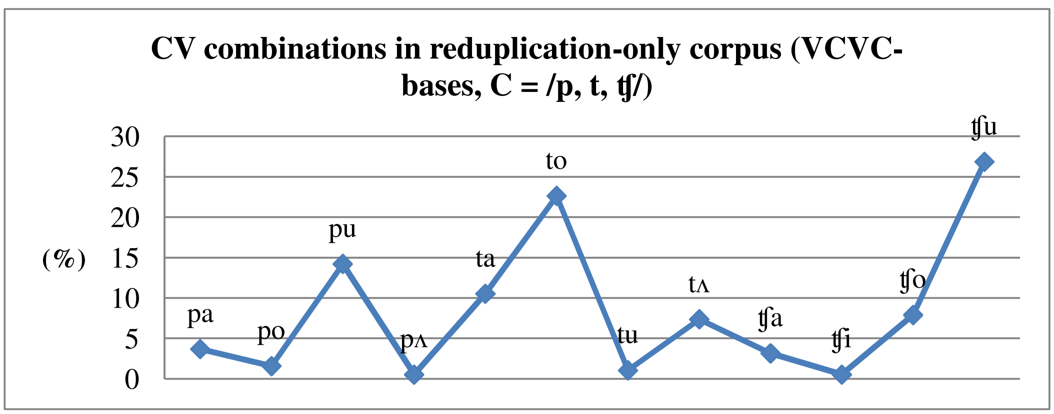


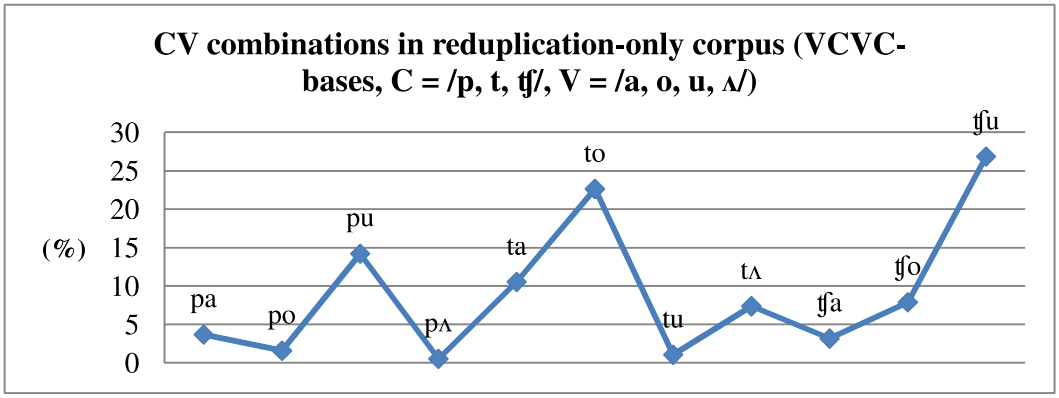
Figure 2.36 shows all CV combinations from the reduplication-only corpus, and we need to consider those combinations that also show up in the experiment. I consider CV combinations when C = /p, t, ʧ/ as I have been focusing on these Cs.

(56) Figure 2.37 CV combinations in the reduplication-only corpus: VCVC-bases, C = /p, t, ʧ/



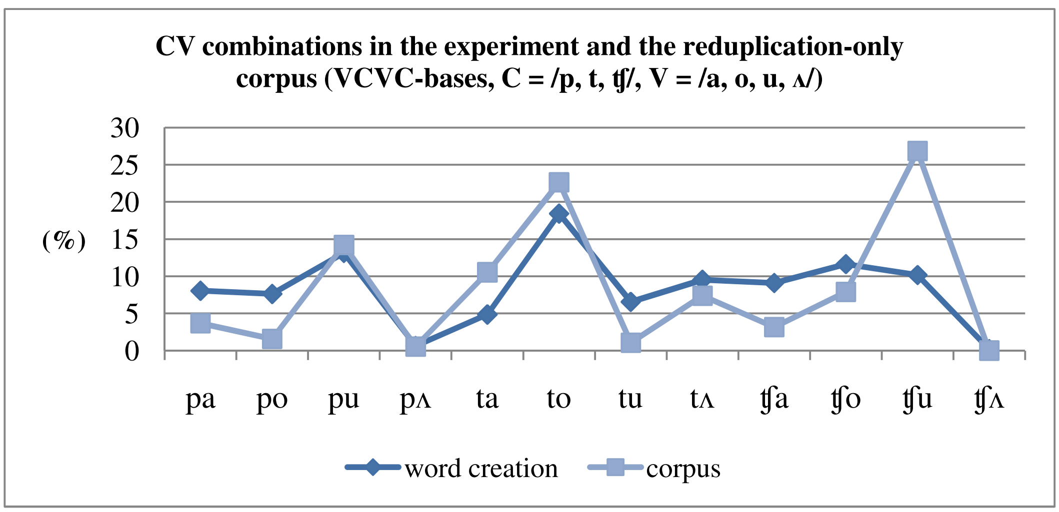
I also need to take account of following Vs which occur both in the experiment and in the reduplication-only corpus for the sake of comparison.

(57) Figure 2.38 CV combinations in the reduplication-only corpus: VCVC-bases, C = /p, t, ʧ/, V = /a, o, u, ʌ/



The CV sequences in the experiment consist of C = /p, t, ʧ/ and V = /a, o, u, ʌ/, and their combining pattern is put together with that from the reduplication-only corpus.

(58) Figure 2.39 CV combinations in the experiment and the reduplication-only corpus: VCVC-bases, C = /p, t, ʧ/, V = /a, o, u, ʌ/



The two patterns of CV combinations from the experiment and the reduplication-only corpus look similar, except that /ʧu/ is much more frequent than /ʧo/ in the corpus, whereas /ʧu/ is a bit less frequent than /ʧo/ in the experiment. As was found out based on the statistical analysis, there was a significant relationship between the two frequency patterns of C and V combination from the experiment and from the reduplication-only corpus, $r_s = .797$, $p < .01$. Thus Hypothesis 11 is confirmed.

(59) Table 2.5 Correlations: CV combinations in the experiment and in the reduplication-only corpus; Initial C = /p, t, ʧ/, V = /a, o, u, ʌ/

			expt	corpus
Spearman's rho	expt	Correlation Coefficient	1.000	.797(**)
		Sig. (2-tailed)	.	.002
		N	12	12
	corpus	Correlation Coefficient	.797(**)	1.000
		Sig. (2-tailed)	.002	.
		N	12	12

** Correlation is significant at the 0.01 level (2-tailed).

Note that the CV combinations in the experiment do not pattern with those in the entire corpus, but they are significantly correlated with those from the reduplication-only corpus. That is, the CV combining pattern in the word creation experiment generally does not follow that of the entire lexicon, but there appears to be some CV co-occurrences that are preferred particularly in reduplication both in the lexicon and in the speakers' creation of reduplicative words.

2.3 Summary

Language users may memorize all patterns they are exposed to in a linguistic environment and merely recapitulate them when they are faced with new forms. To answer the question of whether the Korean reduplication with an inserted consonant, is stored knowledge, which reflects the lexical frequencies, I presented several hypotheses with regard to frequency, token frequency in particular, of consonants in various positions and of consonant-consonant and consonant-vowel co-occurrences, both in the entire corpus and the reduplication-only corpus.

The overall results indicate that in general the speakers' behavior in the experiment does not simply reflect the lexical statistics. The findings showed that there are some correlations, e.g. significant correlations between the frequency pattern of CIs in the experiment and the frequency pattern of lexical Cs in word-initial position in the entire corpus; significant correlations between the frequency patterns of CI – C_1 (in **CIV**.C_1VC_2) in the experiment and C – C (both in onset) and between the frequency patterns of CI – C_2 (in **CIV**.C_1VC_2) in the experiment and C – C (first C in onset and second C in coda) in the entire corpus; significant correlations among the three frequency patterns of CIs in reduplication in the word creation experiment, the *Sejong* corpus and the Google search; and between the frequency pattern of C – V combinations in the experiment and the frequency pattern of C – V combinations in the reduplication-only corpus. The findings in the correlations among the co-occurring patterns (C – C combinations in the experiment and in the entire corpus; C – V combinations in the experiment and in the reduplication-only corpus) are curious, which are to be discussed further in Chapter 4 and 5.

The findings from the lexical statistics suggest that lexical frequency plays a significant role to an extent that some frequency patterns appear to be replicated in the behavioral experiment; however, these statistics do not seem to be a sole factor that works for the CI choice behavior since not all frequency patterns have been shown to be replicated in the nonce word experiment. Therefore, it is necessary to look further into the speakers' CI choice behavior, taking into account other plausible factors.

2.4 Discussion: Lexical Statistics vs. Grammar

There have been debates over how much of phonological and phonetic information is to be included in lexical representations. A model based on construction grammar called Usage-based model concerns how linguistic knowledge is represented and stored in language users' minds. This model especially lays emphasis on the frequency effects of constructions and lexical items (Bybee 1995a, 1995b; Bybee & Hopper 2001). It argues that lexical information is stored in mind, and some constructions and lexical items are more strongly entrenched due to its high frequency in type or token. It opposes the assumption made by generative models that irregular forms are stored in the lexicon while regular forms are generated and processed by means of rules.

One of the main problems that the Usage-based model finds with generative theories is that generative models cannot account for a diachronic tendency that infrequent irregular forms become regular while frequent irregular forms do not. The Usage-based model also presents psycholinguistic evidence that there is difference in production of more frequent vs. less frequent regular words, in terms of response time. Another model, similar more or less to the

Usage-based model, is the Exemplar model (Pierrehumbert 2001, 2002, 2003). According to this approach, phonetic details as well as lexical information are encoded and stored in memory. New stimuli will be categorized on the basis of the existing exemplars which are similar in phonetic information.

Language users' behavioral experiments may shed light on whether they make reference to all and only existing lexical patterns or they use more abstract grammatical generalizations. The comparison of frequencies in the word creation experiment and the corpora in this chapter suggests that speakers do not just memorize what they have in the lexicon and repeat it in their behavior with nonce forms. They may have lexical statistics in mind, but do not necessarily copy/paste them literally when they are faced with new data in real time. This supports an idea that language users do not just live by lexical statistics, and furthermore provides indirect evidence that there must be some other engine, e.g. grammar than lexical statistics. For example, suppose that there is a rule of grammar that tells language users to have the same or a different segment to a context segment. The language users' behavior should be affected by this kind of a rule, encountering non-familiar words, although the existing lexicon of their language does not absolutely prevent them from making a form that does not observe this rule. In this light, it would be useful to consider comparing the statistics based on lexical facts and the statistics based on grammatical constraints, which provide expected values that can be tested against observed values in a real-time experiment. This can help to understand which factor, lexical or grammatical, is more influential in speakers' behavior, which will be discussed in more detail in Chapter 5.

On the other hand, if the internalized rules, as generative theories argue for, are reflected on universal properties in unrelated specific languages, we may find these rules in the same kind of phenomenon in other languages. For example, the consonant insertion patterns I find in the case of Korean reduplication should be attested in other languages. Therefore, both studies in within-a-language and across-languages are necessary in order to understand the issue of lexical statistics vs. abstract rules and constraints. One other yardstick that can help to resolve the controversy over different models would be an issue of learnability, which seeks idealized learning procedures to acquire a grammar based on exposure to language. Whether language learners have internalized rules or resort to existing forms could be more elucidated by further examinations involving learnability.

Chapter 3
Speakers' Preferences in Consonant Choice

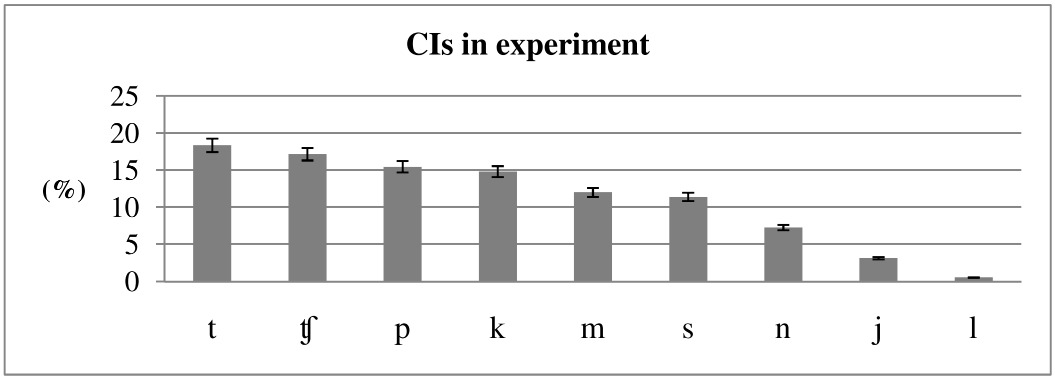
The lexical statistics in the preceding chapter showed that the frequency of CIs in the word creation experiment is significantly correlated with the frequency of Cs in word-initial position in the entire corpus, as well as with the frequency of Cs in reduplicant-initial position in the reduplication-only corpus. The numbers, however, represent the overall frequency based on combined responses of all the participants in Experiment 1. Therefore, the overall frequency does not necessarily mirror the speakers' individual preferences; indeed, there was great variety across speakers. In this chapter I examine individual speakers' choices of CI, proposing that individual preferences for different Cs present an additional source for the asymmetry in the frequency of CIs in Experiment 1.

In addition to the variable frequency patterns of CI among different speakers, I look into whether context plays a role in the choice of CI. I argue that some speakers, if not all, are sensitive to context to the extent that they avoid inserting their preferred C when context contains that same C already.

3.1 Background

In the first word creation task (Experiment 1, N=55), the overall frequency shows that some consonants were preferred to others as CI:

(1) Figure 3.1 CI frequency in Experiment 1[35]



The investigation of lexical frequency in the preceding chapter showed that this uneven distribution among different Cs in consonant insertion reflects the lexical distribution among Cs in word-initial position in the entire corpus.

The reason that the consonant /t/ is among the most popular Cs in the insertion of consonants may be accounted for with an argument that it is a purely preferred segment among others. According to Lombardi (2002), cross-linguistically a glottal stop is the optimal epenthetic consonant, *ceteris paribus*; however, when there are some restrictions, usually morphological, on epenthesis in a language, a coronal like /t/ is the most frequent epenthetic consonant. In particular, /t/ is preferred in onset, due to sonority requirements, when a glottal stop is not available in that language.

However, in Korean, while /t/ was the most frequent CI, a wide range of other consonants were inserted in addition to /t/. In Experiments 1, 2, and 3 (in which speakers had a choice of at least 3 consonants), two consonants emerged as most frequently inserted /t/ and /ʧ/. With respect to this unequal frequency among CIs, I pose the following question:

(2) **General Question**
Do all speakers prefer a single default C for CI? Or do individual speakers have different preferred Cs for CI?

Below I provide tables for all the experiments showing the different frequencies of inserted consonants:

[35] Error bars represent 95% confidence interval of a mean.

(3) Table 3.1 Frequency of CIs from Word Creation 1 (Tokens = 1352)

CI	t	ʧ	p	k	m	s	n	j	l
Tokens	248	232	209	200	162	154	98	42	7
Frequency (%)	18.34	17.16	15.46	14.79	11.98	11.39	7.25	3.11	0.52

(4) Table 3.2 Frequency of CIs from Word Creation 2 (Tokens = 1646)

CI	t	ʧ	k	p	n	s	m	l
Tokens	514	497	252	148	101	90	41	3
Frequency (%)	31.25	30.19	15.31	8.99	6.14	5.47	2.49	0.18

(5) Table 3.3 Frequency of CIs from Word Creation 3 (Tokens = 1665)

CI	t	ʧ	p
Tokens	641	622	402
Frequency (%)	38.5	37.36	24.14

(6) Table 3.4 Frequency of CIs from Word Creation 4 (Tokens = 1662)

CI	t	ʧ
Tokens	806	856
Frequency (%)	48.5	51.5

The results in Experiment 1 are based on reduplication with all V-initial bases (including VCCVC-, VCVC-, VCCV-, VVC-, and VCV-bases), whereas those for Experiments 2, 3, and 4 are based on reduplicative forms with VCVC-bases. Experiment 1 was run with 55 subjects, and Experiments 2, 3, and 4 with 15 subjects. The number of tokens in total excludes responses which did not comply with the directions or were not answered. In Experiments 1 and 2 the participants were requested to make up reduplicative forms by putting in a CI freely, whereas in Experiments 3 and 4 the participants were to pick a preferred consonant from the given set of consonants, /t, p, ʧ/ in Experiment 3 and /t, ʧ/ in Experiment 4. For Experiment 3 I chose /t, p, ʧ/ since they were the three most frequently inserted Cs in Experiment 1, to determine whether one C was more frequently chosen than other two Cs. For Experiment 4 I provided /t, ʧ/ in the hope of discovering a clear driving force behind the choice between /t/ and /ʧ/, the two most popular Cs across the board.

(7) Figure 3.2 Frequency of CI in word creation experiment (= WC) 1, 2, 3, and 4

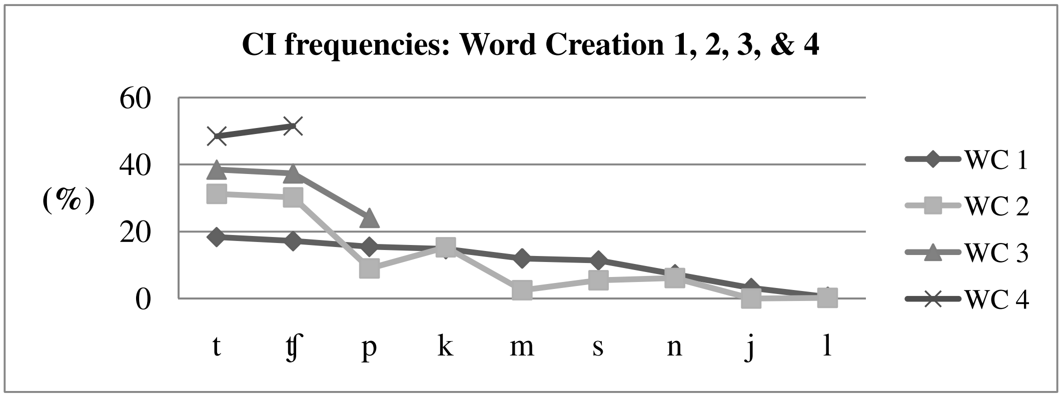


Figure 3.2 shows that the two CIs /t, ʧ/ were almost equally frequent.

3.2 Variation in Consonant Insertion

Analysis of the responses of individual participants revealed different characteristic response patterns. From the different participants in the four word creation experiments, I first chose those who inserted one C predominantly over the others. I next sorted them into different groups according to their preferred CI. The two largest groups contained speakers who preferred to insert /t/ (the *t-dominant pattern*) and those who preferred to insert /ʧ/ (the *ʧ-dominant pattern*). Though not all speakers fell into one of these two groups, the greater popularity of these two patterns makes them best suited for whether speakers with a clear preference for a specific consonant will be affected by context in choosing a consonant for insertion.

To be categorized into either t-dominant pattern or ʧ-dominant pattern, speakers had to prefer either /t/ or /ʧ/, not both; otherwise, they cannot be grouped into either pattern. I used the chi-square test to confirm that the identified speakers in each pattern significantly prefer /t/ or /ʧ/ over any other C. The statistical report of the chi-square test is provided along with the figures which show the percentage of C choices for the individuals who belong to either t-dominant or ʧ-dominant group, in Appendix 3-A at the end of this chapter.

In the sorting, I also found other patterns in which participants preferred to insert other C than /t/ or /ʧ/; among others, there were far more participants in each experiment who cannot be classified into a specific C-dominant group (hence, the *no-C-dominant pattern*). The participants who are grouped into the no-C-dominant pattern did not insert a single C predominantly; rather, they appear to have more than one C that they preferred to insert. In the following table, the number of participants who belong to a group is presented along with the percentage out of a total number of participants in each experiment.

(8) Table 3.5 Other C-dominant groups identified in Experiments 1 – 4

Patterns	Experiment **1** (N=55)	Experiment **2** (N=15)	Experiment **3** (N=15)	Experiment **4** (N=15)
p-dominant group	6 participants (10.91%)	1 participants (6.67%)		
k-dominant group	2 participants (3.64%)	3 participants (20%)		
m-dominant group	2 participants (3.64%)			
no-C-dominant group	31 participants (56.36%)	2 participants (13.33%)	3 participants (20%)	8 participants (53.33%)

The blank cells indicate that there were no participants who can be categorized into such a group. In general, among the participants who preferred a specific C for CI, there were more participants who were grouped into t-dominant or ʧ-dominant pattern than those who were grouped into any other C-dominant pattern.

In the following section, I present the number of speakers who conformed with either of the two identified patterns: there are several speakers per experiment, identified for each pattern. For Experiment 1 the participants were recruited in Seoul, Korea, and for Experiments 2 – 4 the participants were recruited in Stony Brook, New York, and some speakers participated in more than one experiment. Importantly, those who participated in more than one experiment were consistent in their choices of CI; that is, participants who significantly preferred /t/ in one experiment still preferred /t/ in other experiments.

3.2.1 t-dominant and ʧ-dominant patterns

The two groups of participants I identified in each of the four experiments were those who predominantly chose /t/ over other Cs and those who predominantly chose / ʧ/ over other Cs. In the following I present the number of participants (with percentage for a total number of participants) who belong to either t-dominant group or ʧ-dominant group, which can clearly show that there are distinct patterns in the speakers' preference for a specific C as CI.

(9) Table 3.6 t-dominant and ʧ-dominant group identified in Experiment 1

Experiment **1** (N=55)	
t-dominant group	11 participants (20%)
ʧ-dominant group	3 participants (5.45%)

In Experiment 1, 11 participants out of 55 chose /t/ as CI, while 3 participants chose /ʧ/ as CI, predominantly over other consonants. Since it was an open-ended choice, the participants could insert any C in creating a nonce reduplicative word, except for /ŋ/ which is not allowed in onset in Korean.

(10) Table 3.7 t-dominant and ʧ-dominant group identified in Experiment 2

Experiment **2** (N=15)	
t-dominant group	4 participants (26.67%)
ʧ-dominant group	5 participants (33.33%)

In Experiment 2, 4 participants out of 15 were identified as belonging to the t-dominant group, whereas 5 participants were identified as those who have preference for /ʧ/. The participants in each of these groups chose the specific C predominantly over other consonants.

(11) Table 3.8 t-dominant and ʧ-dominant group identified in Experiment 3

Experiment **3** (N=15)	
t-dominant group	5 participants (33.33%)
ʧ-dominant group	7 participants (46.67%)

In Experiment 3, 5 participants out of 15 chose /t/ most often, whereas 7 participants out of 15 chose /ʧ/ most often. In this experiment there were three candidate consonants /t, p, ʧ/ to be chosen as CI, so each consonant had a 33.33% chance of being chosen. The 5 participants in the t-dominant group chose /t/ predominantly over /p, ʧ/, and the 7 participants in the ʧ-dominant group chose /ʧ/ predominantly over /t, p/.

Among the 7 participants identified into the ʧ-dominant group, three also participated in Experiment 2, and they significantly preferred /ʧ/ as CI in that experiment, as well. This consistency in the participants' choice of a specific C confirmed that speakers have their own preferred C for CI.

(12) Table 3.9 t-dominant and ʧ-dominant group identified in Experiment 4

Experiment **4** (N=15)	
t-dominant group	4 participants (26.67%)
ʧ-dominant group	3 participants (20%)

In Experiment 4, the participants were asked to choose either /t/ or /ʧ/ as CI, so each consonant had a 50% chance of being chosen. 4 participants out of 15 chose /t/ in most cases. Interestingly, these 4 speakers also participated in either Experiment 2 or 3, in which they also showed significant preference for /t/ in the choice of Cs. There were 3 participants who were identified as choosing /ʧ/ predominantly over other Cs.

The examination of the participants' choices of CI in all four experiments showed that there is clear distinction among the participants, who preferred different Cs for CI. That is, there is no default C for all speakers to choose as CI. The overall frequency of CIs is not the same as each individual speaker's frequency pattern.

3.3 Context

3.3.1 Experiment 1

Although many speakers showed a preference for a particular consonant, no speakers inserted only one particular consonant in all the experimental nonce forms. I now consider whether the context affects a speaker's likelihood of choosing some other consonant than the preferred consonant.

One possible factor is a tendency to avoid inserting a specific consonant if the same consonant already occurs in the word. Experiment 1 contained two nonce forms which contained /t/ in the base and six nonce forms which contained /ʧ/ in the base.

(13) Table 3.10 t-dominant group: CI choice in context of /t/

t-dominant pattern	CI=/t/	CI=non-/t/
S1	2/2=100	0/2=0
S11	2/2=100	0/2=0
S12	1/2=50	1/2=50
S26	0/2=0	2/2=100
S27	2/2=100	0/2=0
S29	1/2=50	1/2=50
S45	2/2=100	0/2=0
S51	1/2=50	1/2=50

(14) Table 3.11 ʧ-dominant group: CI choice in context of /ʧ/

ʧ-dominant pattern	CI=/ʧ/	CI=non-/ʧ/
S4[36]	3/6=50	2/6=33.33
S23	3/6=50	3/6=50

Table 3.10 and 3.11 cannot present any statistically meaningful findings given the small number of tokens of each type.

3.3.2 Experiments 2 – 4

The stimuli in Experiments 2, 3, and 4 did not have /ʧ/ in the context, which made it possible to focus on contextual effects only in the t-dominant group. That is, I can see if the speakers who prefer /t/ in general would be more likely to insert other Cs when the context contains /t/.

(15) **Hypothesis 1**
Speakers from the t-dominant pattern will be less likely to insert /t/ when /t/ is present in the base.

These three experiments contained 36 items (out of 111) in which /t/ was present in the context. In Experiment 2, Hypothesis 1 predicts that the participants who

36 This participant did not respond to one of the forms that had /ʧ/ in context.

were categorized into the t-dominant group would put other Cs than /t/ in the context of /t/.

For the tables below I computed Observed N, actual number for each speaker's choices of /t/, along with Expected N, estimated number for each speaker's choices of /t/ according to his or her general tendency of choosing /t/ as CI. For example, S5 in Experiment 2 made 41 choices of /t/ in all responses (out of 111), and this general ratio was applied to the choices of /t/ in the context of /t/: out of 36 forms that has /t/ in context, S5 chose /t/ 8 times and is expected to make a choice of /t/ 13.3 times (=41/111 × 36), which is used as an expected N. This computation was used for all the participants who belong to the t-dominant pattern in all experiments, Experiments 2, 3, and 4.

The O/E ratio was calculated with the Observed and Expected values: if the O/E is greater than 1, it means that the number of /t/ occurrences in the context of /t/ is overrepresented than expected; and if the O/E is less than 1, it means that the number of /t/ occurrences in the context of /t/ is underrepresented than expected. Hypothesis 1 predicts underrepresentation of /t/ in the context of /t/ because speakers are not likely to insert the same C as a context C. The cells in tables below are shaded when /t/ is underrepresented significantly.

(16) Table 3.12 Experiment 2: /t/ choice in the context of /t/ (36 words that have /t/ in context)

Participants in the t-dominant group	Observed	Expected	O/E
S5	8	13.3	0.6
S13	12	21.1	0.57
S14	28	27.2	1.03
S15	15	19.5	0.77

In Experiment 2, S5 had a tendency to choose non-/t/ consonants as CI when /t/ exists in the context, which is marginally significant, $\chi^2(1, N = 36) = 3.35$, $p = .067$. S13 was significantly more likely to choose other Cs than /t/ when context contains /t/, $\chi^2(1, N = 36) = 9.44$, $p < .01$. However, participants S14 and S15 still preferred to insert /t/ in spite of the contextual /t/, with no statistically significant difference between the observed and expected values, $\chi^2(1, N = 36) = .09$, $p > .05$; $\chi^2(1, N = 36) = 2.22$, $p > .05$, respectively.

In Experiment 3, the participants were asked to choose a C out of /t, p, ʧ/ for CI: the participants in this experiment, who belong to the t-dominant group and epenthesize /t/ most of the time, are predicted to insert /p/ or /ʧ/ in the context of /t/, according to Hypothesis 1.

(17) Table 3.13 Experiment 3: /t/ choice in the context of /t/ (36 words that have /t/ in context)

Participants in the t-dominant group	Observed	Expected	O/E
S5	17	15.2	1.12
S6	19	18.5	1.03
S9	17	17.2	0.99
S10	18	17.8	1.01
S12	21	25	0.84

All participants S5, S6, S9, S10 and S12 in Experiment 3 still preferred to insert /t/ in spite of /t/ in context, not showing significant difference between the observed and expected values, S5, $\chi^2(1, N = 36) = .35, p > .05$; S6, $\chi^2(1, N = 36) = .03, p > .05$; S9, $\chi^2(1, N = 36) = .00, p > .05$; S10, $\chi^2(1, N = 36) = .00, p > .05$; S12, $\chi^2(1, N = 36) = 2.06, p > .05$.

In Experiment 4, the participants were asked to choose either /t/ or / ʧ/ for CI: the participants in this experiment who belong to the t-dominant group are predicted to insert a non-/t/ segment, i.e. /ʧ/, in the context of /t/, according to Hypothesis 1.

(18) Table 3.14 Experiment 4: /t/ choice in the context of /t/ (36 words that have /t/ in context)

Participants in the t-dominant group	Observed	Expected	O/E
S1	25	22.4	1.12
S3	25	24.6	1.02
S4	27	22.4	1.21
S7	18	22.4	0.8

All participants S1, S3, S4, and S7 in Experiment 4 still preferred to insert /t/ in spite of /t/ in context, not showing statistically significant difference between the observed and expected values, S1, $\chi^2(1, N = 36) = .81, p > .05$; S3, $\chi^2(1, N = 36) = .02, p > .05$; S4, $\chi^2(1, N = 36) = 2.52, p > .05$; S7, $\chi^2(1, N = 36) = 2.26, p > .05$.

Note that these speakers in Experiment 4 in fact participated in Experiment 2 or 3: S3 was in Experiment 2 and S1, S4, and S7 were in

Experiment 3. They all still preferred /t/ in context of /t/, throughout the experiments with different conditions. Thus speakers are consistent in their CI choice behavior.

The examination of the t-dominant groups in Experiments 2, 3, and 4 show that not every speaker who belongs to t-dominant group chose a non-/t/ consonant when they encountered /t/ in the context; however, there *were* some speakers who chose a non-/t/ segment when coming across /t/ in context.

A question arises at this point: Were these participants more likely to choose a non-/t/ when there was /t/ in the context than when there was no /t/ in the context? To see if this tendency was the case, I can count the number of /t/'s in the /t/-context and in the no-/t/ context, predicting that there were more /t/'s in the no-/t/ context than in the /t/-context.

(19) Table 3.15 Experiment 2: /t/ choice in the /t/ context (36 words) and in the no-/t/ context (75 words)

Participants	Observed	Expected	O/E	Observed	Expected	O/E
S5	8	13.3	0.6	33	27.7	1.19
S13	12	21.1	0.57	53	43.9	1.21
S14	28	27.2	1.03	56	56.8	0.99
S15	15	19.5	0.77	45	40.5	1.11

The O/E ratios in the tables above and below are for the /t/ choice in the /t/ context (the O/E ratio to the left) and for the /t/ choice in the no-/t/ context (the O/E ratio to the right). These tables are combined ones for the two contexts, the /t/ context and the no-/t/ context, for comparison's sake. Table 3.15 shows that in general the O/E ratio for the /t/ choice in the no-/t/ context indicates overrepresentation; that is, the participants showed a tendency to insert /t/ more often than expected in the no-/t/ context. This general pattern can be seen in the other two experiments as in Table 3.16 and 3.17 below, which at least means that the /t/ choice is not underrepresented in the no-/t/ context, whereas the /t/ choice is underrepresented in the /t/ context for some participants.

(20) Table 3.16 Experiment 3: /t/ choice in the /t/ context (36 words) and in the no-/t/ context (75 words)

Participants	Observed	Expected	O/E	Observed	Expected	O/E
S5	17	15.2	1.12	30	31.8	0.94
S6	19	18.5	1.03	38	38.5	0.99
S9	17	17.2	0.99	36	35.8	1.01
S10	18	17.8	1.01	37	37.2	0.99
S12	21	25	0.84	56	52	1.08

(21) Table 3.17 Experiment 4: /t/ choice in the /t/ context (36 words) and in the no-/t/ context (75 words)

Participants	Observed	Expected	O/E	Observed	Expected	O/E
S1	25	22.4	1.12	44	46.6	0.94
S3	25	24.6	1.02	51	51.4	0.99
S4	27	22.4	1.21	42	46.6	0.9
S7	18	22.4	0.8	51	46.6	1.09

Hypothesis 1 has been confirmed to some degree, and there is micro-level variation among the speakers who preferred /t/ in the CI choice: some speakers stick to the same preferred segment regardless of context and other speakers avoid inserting their preferred consonant /t/ in the context of /t/, trying not to have identical segments in a word.

3.4 Summary

We have seen that different speakers had different preferred segments as CI. Therefore, although the overall frequencies of CIs reflects the overall frequencies of word-initial consonants in the lexicon, the similar distributions emerge from the participants as a group, and are not necessarily reflected in the productions of each individual speaker. In this chapter we investigated the behavior of two groups: speakers who prefer /t/ and speakers who prefer /ʧ/. The existence of these different groups suggests that different speakers have different subsets of grammar. It also provides explanation for the asymmetric distribution of CIs in the responses from the experiment, which cannot be simply attributed to the overall lexical frequency.

Within a group, there is variation among the speakers, as well. For some speakers, the preferred segment is less likely to be chosen when context contains that segment, which makes room for consideration of a contextual factor. For example, among the speakers who tend to choose /t/ as CI there were some speakers who did not opt for /t/ when /t/ was present in the context. This micro-level variation, in addition to the overall variation among different groups of speakers, makes the issue at hand more complicated. For a group of speakers who choose /t/ as a default or preferred CI, /t/ is usually the preferred CI, but it becomes less acceptable when /t/ exists in context.

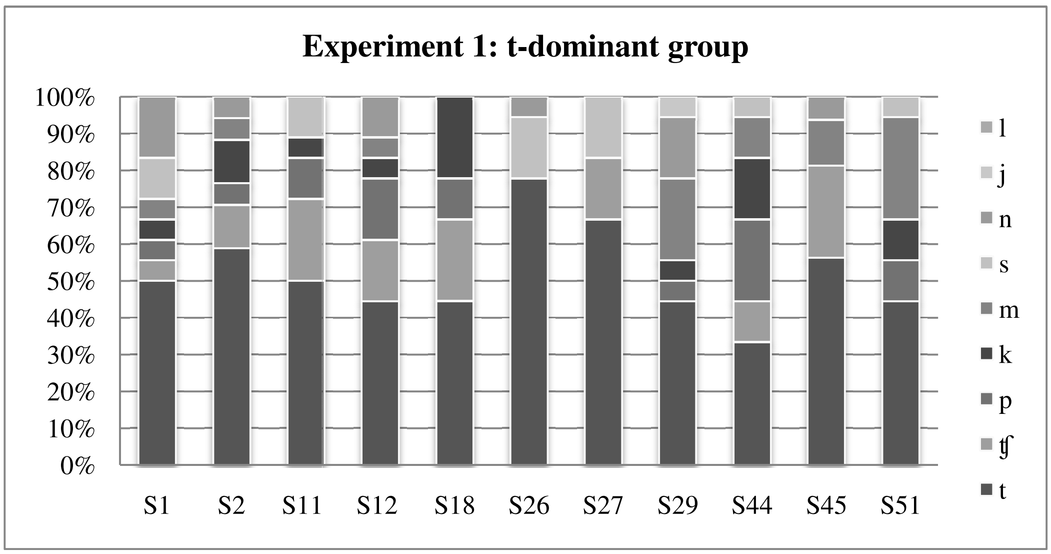
Appendix 3-A Significance values for individual speakers

Individual statistical reports for the t-dominant groups and the ʧ-dominant groups in Experiments 1, 2, 3, and 4:[37]

<u>t-dominant pattern</u>

Experiment 1:

(1) Figure 3.3 t-dominant group: participants who preferred /t/ in Experiment 1

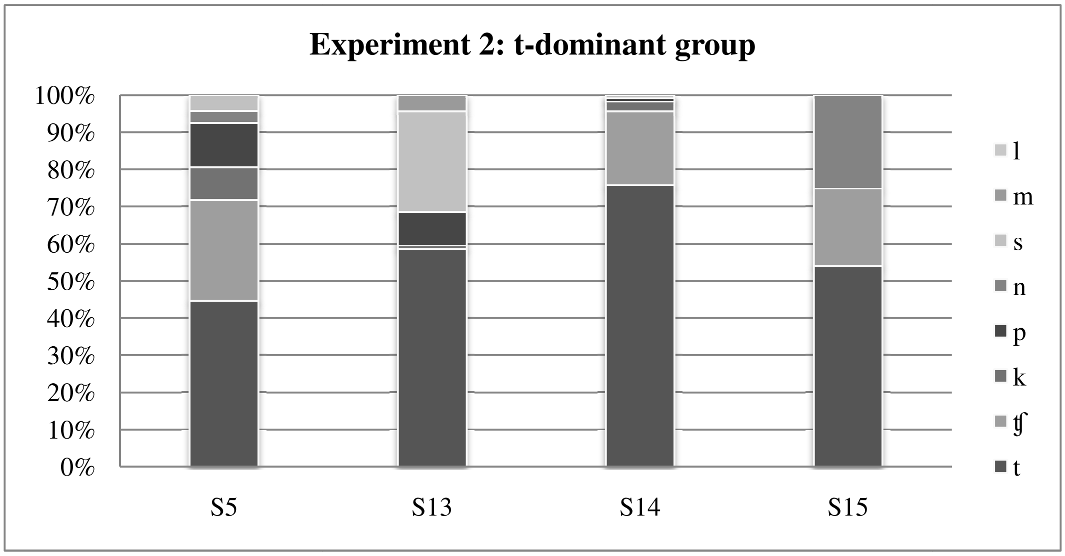


All 11 participants in the t-dominant group chose /t/ significantly more frequently than the expected value (for the computation of the expected value, see Section *3.2.1.1*): for S1 $\chi^2(1, N = 18) = 32.00$, $p < .001$, for S2 $\chi^2(1, N = 18) = 41.51$, $p < .001$, for S11 $\chi^2(1, N = 18) = 32.00$, $p < .001$, for S12 $\chi^2(1, N = 18) = 23.73$, $p < .001$, for S18 $\chi^2(1, N = 18) = 23.73$, $p < .001$, for S26 $\chi^2(1, N = 18) = 91.88$, $p < .001$, for S27 $\chi^2(1, N = 18) = 64.22$, $p < .001$, for S29 $\chi^2(1, N = 18) = 23.73$, $p < .001$, for S44 $\chi^2(1, N = 18) = 10.89$, $p < .01$, for S45 $\chi^2(1, N = 16) = 38.03$, $p < .001$, for S51 $\chi^2(1, N = 18) = 23.73$, $p < .001$.

[37] I only need one choice per base for the Chi square test. Hence I considered only the first choice (= the most preferred choice) of Cs for each item when there are more than one choice of Cs per a reduplicated word.

Experiment 2:

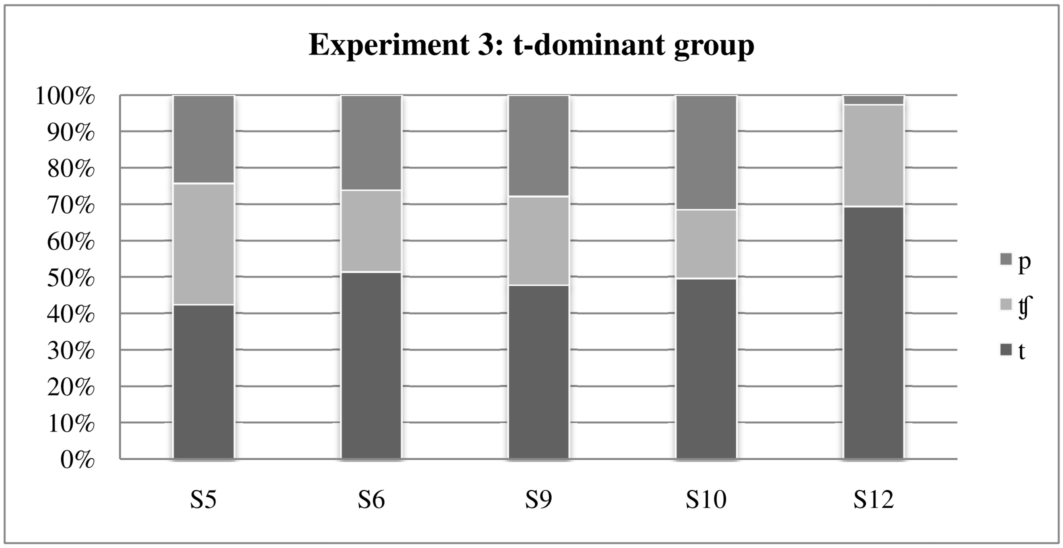
(2) Figure 3.4 t-dominant group: participants who preferred /t/ in Experiment 2



All 4 participants in the t-dominant group chose /t/ significantly more frequently than the expected value (for the computation of the expected value, see Section *3.2.1.1*): for S5 $\chi^2(1, N = 111) = 89.49$, $p < .001$, for S13 $\chi^2(1, N = 111) = 290.81$, $p < .001$, for S14 $\chi^2(1, N = 111) = 531.97$, $p < .001$, for S15 $\chi^2(1, N = 111) = 239.36$, $p < .001$.

Experiment 3:

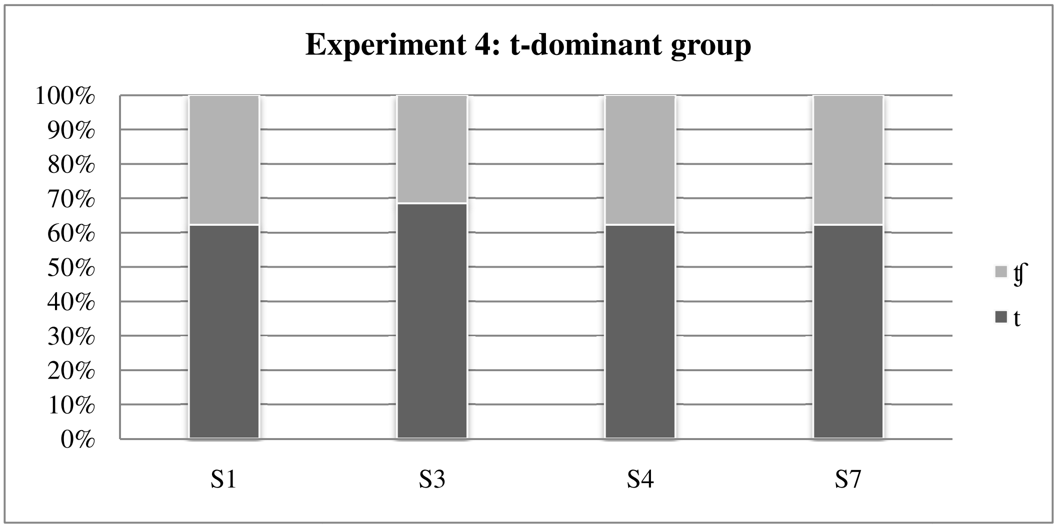
(3) Figure 3.5 t-dominant group: participants who preferred /t/ in Experiment 3



All 5 participants in the t-dominant group chose /t/ significantly more frequently than the expected value (for the computation of the expected value, see Section *3.2.1.1*): for S5 $\chi^2(1, N = 111) = 4.06$, $p < .05$, for S6 $\chi^2(1, N = 111) = 16.22$, $p < .001$, for S9 $\chi^2(1, N = 111) = 10.38$, $p = .001$, for S10 $\chi^2(1, N = 111) = 13.14$, $p < .001$, for S12 $\chi^2(1, N = 111) = 64.88$, $p < .001$.

Experiment 4:

(4) Figure 3.6 t-dominant group: participants who preferred /t/ in Experiment 4

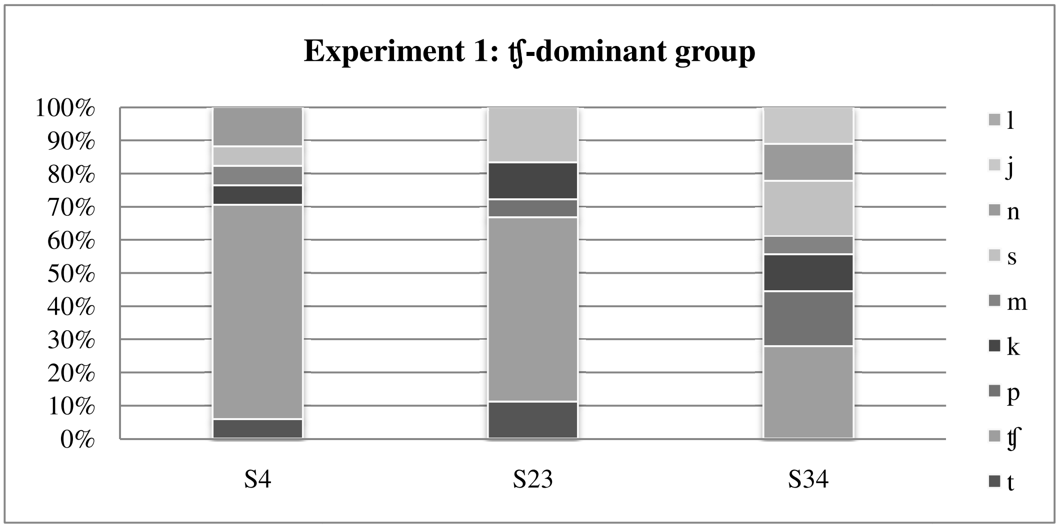


All 4 participants in the t-dominant group chose /t/ significantly more frequently than the expected value (for the computation of the expected value, see Section *3.2.1.1*): for S1 $\chi^2(1, N = 111) = 6.57$, $p < .05$, for S3 $\chi^2(1, N = 111) = 15.14$, $p < .001$, for S4 $\chi^2(1, N = 111) = 6.57$, $p < .05$, for S7 $\chi^2(1, N = 111) = 6.57$, $p < .05$.

<u>ʧ-dominant pattern</u>

Experiment 1:

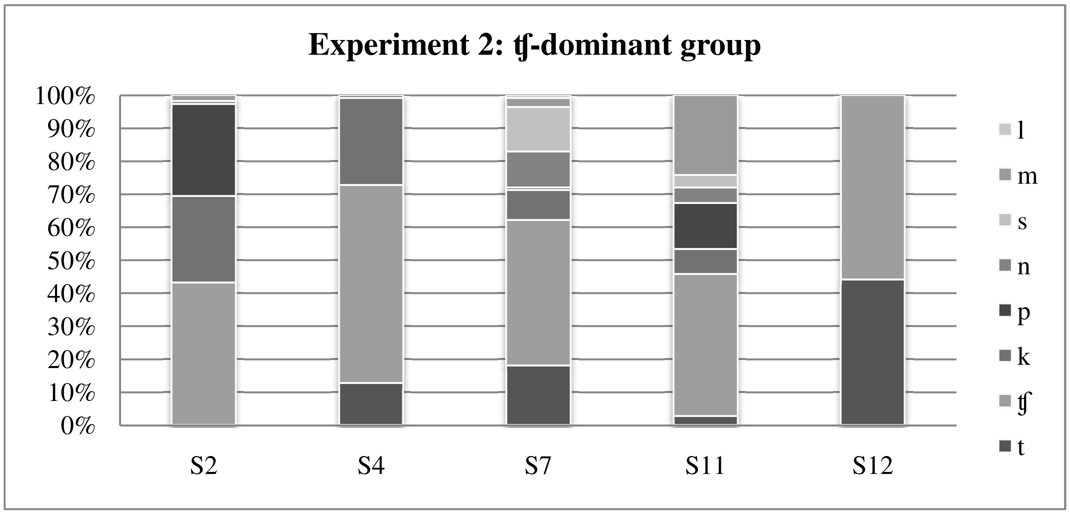
(5) Figure 3.7 ʧ-dominant group: participants who preferred /ʧ/ in Experiment 1



All 3 participants in the ʧ-dominant group chose /ʧ/ significantly more frequently than the expected value (for the computation of the expected value, see Section *3.2.1.2*): for S4 $\chi^2(1, N = 17) = 56.53$, $p < .001$, for S23 $\chi^2(1, N = 18) = 41.51$, $p < .001$, for S34 $\chi^2(1, N = 18) = 6.32$, $p < .05$.

Experiment 2:

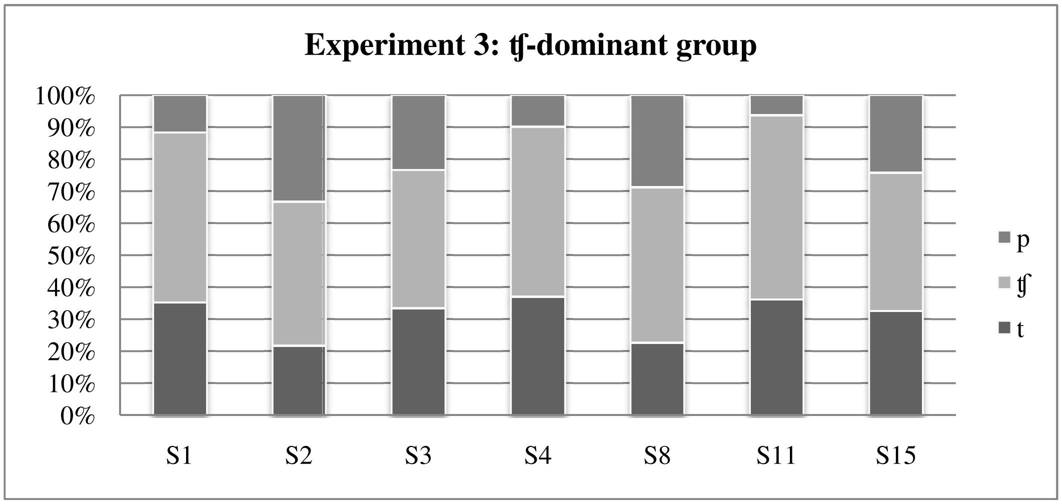
(6) Figure 3.8 ʧ-dominant group: participants who preferred /ʧ/ in Experiment 2



All 5 participants in the ʧ-dominant group chose /ʧ/ significantly more frequently than the expected value (for the computation of the expected value, see Section *3.2.1.2*): for S2 $\chi^2(1, N = 111) = 136.30$, $p < .001$, for S4 $\chi^2(1, N = 111) = 301.70$, $p < .001$, for S7 $\chi^2(1, N = 111) = 143.79$, $p < .001$, for S11 $\chi^2(1, N = 111) = 121.92$, $p < .001$, for S12 $\chi^2(1, N = 111) = 259.34$, $p < .001$.

Experiment 3:

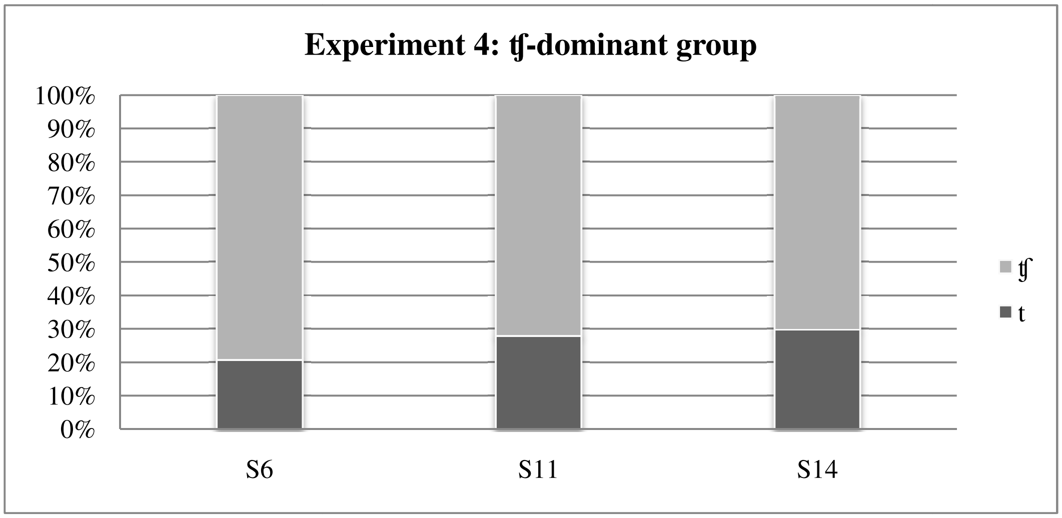
(7) Figure 3.9 ʧ-dominant group: participants who preferred /ʧ/ in Experiment 3



All 7 participants in the ʧ-dominant group chose /ʧ/ significantly more frequently than the expected value (for the computation of the expected value, see Section *3.2.1.2*): for S1 $\chi^2(1, N = 111) = 19.63, p < .001$, for S2 $\chi^2(1, N = 111) = 6.86, p < .01$, for S3 $\chi^2(1, N = 111) = 4.91, p < .05$, for S4 $\chi^2(1, N = 111) = 19.63, p < .001$, for S8 $\chi^2(1, N = 111) = 11.72, p = .001$, for S11 $\chi^2(1, N = 111) = 29.56, p < .001$, for S15 $\chi^2(1, N = 111) = 4.91, p < .05$.

Experiment 4:

(8) Figure 3.10 ʧ-dominant group: participants who preferred /ʧ/ in Experiment 4



All 3 participants in the ʧ-dominant group chose /ʧ/ significantly more frequently than the expected value (for the computation of the expected value, see Section *3.2.1.2*): for S6 $\chi^2(1, N = 111) = 38.06$, $p < .001$, for S11 $\chi^2(1, N = 111) = 21.63$, $p < .001$, for S14 $\chi^2(1, N = 111) = 18.24$, $p < .001$.

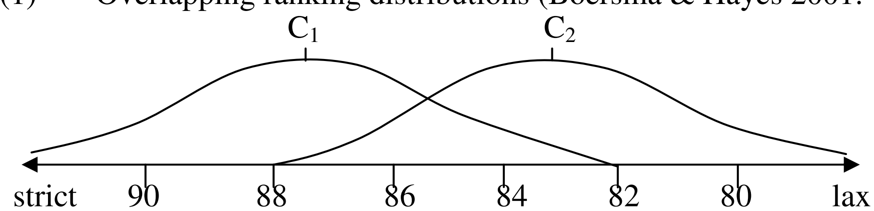
Appendix 3-B A learner model

Gradual Learning Algorithm

A grammar that captures the findings of variation and gradience in speakers' choice of CIs in the word creation experiments must produce variable outputs. The Gradual Learning Algorithm, GLA, has been claimed to be an appropriate algorithm for learning grammars, particularly based on Optimality Theory, OT (Boersma 1997; Boersma & Hayes 2001). It is argued to be better at dealing with free variation, noisy learning data, and gradient well-formedness. The GLA is characterized by its assumption that the constraints are continuous, not discrete, and the grammar is stochastic. Therefore, the GLA allows the grammar to produce variable outputs, which may account for different outputs by different speakers of a language. In this section I will briefly describe how the GLA works and how it can capture variation in grammar.

The conceptual bases for the GLA are a "continuous ranking scale" and "stochastic candidate evaluation." With the continuous scale, the GLA can handle both categorical and gradient rankings. The position of a constraint, at an evaluation time, is perturbed by random noise, before it is finally selected. If the ranges of selection points for constraints do not overlap, then the ranking scale ends up with the traditional categorical ranking. However, if some constraints turn out to have overlapping ranges, then the ranking scale will show free ranking.

(1) Overlapping ranking distributions (Boersma & Hayes 2001: 49)



According to Boersma & Hayes (2001), the hypothetical ranking values for the constraints, C_1 and C_2 are 87.7 and 83.1, respectively. Thus we may see a ranking of $C_2 >> C_1$ at some occasions (5.2%), although $C_1 >> C_2$ should hold most of the time (94.8%).

The GLA locates an appropriate ranking value for a constraint, in the process from the initial state up to the final state. Every constraint starts at the same value. A learning datum consisting of adult surface forms is presented to the algorithm. Then for each constraint a selection point is picked according to the constraint's current ranking value. This generation process follows the standard mechanisms of OT. If a generated form is identical to the learning datum, no further actions will take place. However, if there happens to be a mismatch between the generated form and the learning datum, then the algorithm will take measures in order for the grammar to generate the learning datum. It will basically change the ranking values of the constraints, in which violations matching in the two rival candidates will be cancelled out, as in the following:

(2) Mark cancellation (Boersma & Hayes 2001: 52)

/underlying form/	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
✓ Candidate 1 (learning datum)	*!	*~~*~~	~~*~~		*			~~*~~
☞ Candidate 2 (learner's output)		~~*~~	~~*~~	*		*		~~*~~

Then the adjustment of the ranking values is made repeatedly with further exposure to learning data through the cycle of presentation of learning datum – generation – comparison – adjustment, until the learner's output matches the learning datum.

(3) Adjusting the ranking values (Boersma & Hayes 2001: 53)

/underlying form/	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
✓ Candidate 1 (learning datum)	*→	*→			*→			
☞ Candidate 2 (learner's output)				←*		←*		

The GLA has been relatively successful in dealing with example data with free variation: for instance, the resulting grammar via machine ranking for the variation data, e.g. in Ilokano could generate the predicted variations by running the input underlying forms with the output probabilities. Not only could the grammar generate all and only the correct forms, but it also could produce the matching frequencies in the learning data. That is, when the language has a single output form 100% of the time, the machine grammar also generated only that output, e.g. /paʔlak/ → [pa.lak]. When the language has variation between two forms, 50% of the time each, the machine grammar also generated the two forms with closely matching percents of the time for each form, e.g. /taʔo-en/ → [taw.ʔen] ~ [taʔ.wen]. When the language has variation among three forms, the machine grammar successfully predicted that alternation with about a third of the frequency for each form, e.g. /bwaja/ → [bu:.bwa.ja] ~ [bwaj.bwa.ja] ~ [bub.wa.ja].

An application

In this section I apply the GLA to the reduplication data with CI, to see how learning of the CI choice takes place, which includes variation.[38] I pick some representative examples to illustrate the learning, among the dictionary data and the data from the distinct groups in a series of the experiments. For example, I will show learning of a single, general grammar for the dictionary data, which are categorical and the data from Experiment 2, which displays variation in the choice of CIs (e.g. t and ʧ). In addition, I will show how an individual speaker's grammar can be learned: an individual grammar which prefers a specific C (t or ʧ)

[38] I used the OTSoft, available at http://www.linguistics.ucla.edu/people/hayes/otsoft/ (OTSoft 2.3), for running the GLA.

with no consideration of context; and an individual grammar which prefers a specific C (t or ʧ) that takes context into account.

First, I provide description of the data and explanation of the relevant constraints on the basis of OT, which in turn are to be used in an input file for running the GLA. An input file includes regular OT tableaux with input forms, output candidates and their frequencies, constraints, and constraint violations for the output candidates; a sample input file is provided in Appendix 3-C. Second, I run the input file in the GLA, so that we can see how well the model could learn a grammar for the reduplication case. Then the resulting grammar is assessed in comparison to the targeted grammar.

The learning data

Example data

With respect to the dictionary data, I use the reduplicative forms with VCVC-bases that carry /t, p, ʧ/ as CI:

(4) a. atɨk-**p**atɨk ‘doggedly’
b. okɨl-**p**okɨl ‘bubbling’
c. alok-**t**alok ‘dappled’
d. oson-**t**oson ‘harmoniously’
e. omil-**ʧ**omil ‘elaborate’
f. umul-**ʧ**’umul ‘hesitantly’

The forms with VCVC-bases and CI=/t, p, ʧ/ amounted to 51, the list of which has been provided in Appendix 1-A. The surface forms in the dictionary do not have variants, and they will produce a categorical grammar with only a single optimal output for an underlying form.

The responses in Experiment 2 consist of nonce reduplicative forms with VCVC-bases, which were open to any choice of CI, some of whose examples are:[39]

[39] **C** in the examples indicates a CI, which was actually a blank that is to be filled in by the participants in the experiments.

(5) a. ikip-<u>**C**ikip</u>
b. isim-<u>**C**isim</u>
c. unup-<u>**C**unup</u>
d. ukuŋ-<u>**C**ukuŋ</u>
e. atan-<u>**C**atan</u>
f. apam-<u>**C**apam</u>

I will consider the forms with CI=/t, p, ʧ/, among others, for the sake of comparison with a grammar of the dictionary data. The list of entire stimuli has been furnished in Appendix 1-B. Regarding the response data from Experiment 2, I will obtain three kinds of grammar with the GLA: a single general grammar for all cases with CI=/t, p, ʧ/; an individual grammar for a speaker who prefers /t/ without consideration of /t/ in the context; and an individual grammar for a speaker who prefers /t/ with consideration of /t/ in the context.

OT-based analysis

We have identified two major factors which appear to affect the choice of CI, in this chapter: speakers' preference and consideration of context. Concerning the speakers' preference, I found two distinct groups: t-dominant group and ʧ-dominant group. There is no specific constraint that can impose the tendency of preference; therefore, I will simply use a segmental markedness constraint to show preference for a certain segment.

With regard to the concept of constraints, the classical OT assumes a universal set of ranked and violable output constraints (Prince & Smolensky 1993); however, there is an alternative approach which argues that constraints are learned from language-specific data on the basis of Universal Grammar that consists of a feature set and a constraint format (Hayes & Wilson 2008). I do not necessarily adopt one approach over the other in my analysis; however, I am aware of these different views, and I rather sympathize with Yip's (1995) remark that the strongest claims of universality for constraints may not be too agreeable and reasonable particularly when it comes to the issue of how to handle language-specific and morpheme-specific constraints in OT.[40]

As regards a constraint for the contextual factor (avoiding inserting the same C as a C in the context, in particular), I use a family of *REPEAT constraints, à la Yip (1995), since there was avoidance of repeating the same segment /t/ when /t/ already exists in context. This type of constraint has a long tradition in

[40] This point of discussion was brought to my attention by Andries Coetzee (p.c.).

phonology, which has been called the Obligatory Contour Principle, OCP (Clements & Keyser 1983; Goldsmith 1976; Leben 1973; McCarthy 1979, 1981, 1986; Steriade 1982; Yip 1988, 1995, 1998, among many others). Yip utilized the constraint, *REPEAT in a morphological sense, particularly when discussing reduplication data, and I also use the same constraint in considering the sensitivity to context in this chapter, i.e. avoiding the same C as a context C for CI. I use sensitivity to context interchangeably with identity avoidance which I will use more often in Chapter 5, since the context sensitivity first came up in the previous discussion of this chapter.

(6) Constraints

C:	The reduplicant must begin with a consonant C. (e.g. Constraint, t means that /t/ is inserted in the beginning of the reduplicant.)
*REPEAT(segment):	Output must not contain identical segments. (e.g. *REPEAT(t) requires that /t/ cannot occur repeatedly.)
ONSET:	Syllables must have onsets.
DEP-BR:	Every element of the reduplicant has a correspondent in the base. ("No epenthesis")
MAX-BR:	Every segment of the base has a correspondent in the reduplicant. ("No deletion")
Place-Markedness Hierarchy:	
	*PL/LAB, *PL/DORS >> *PL/COR (Alderete, *et al*. 1999; Prince & Smolensky 1993)
*ONSETV:	The leftmost onset segment in a syllable does not have the specified sonority level. (This constraint family assumes a hierarchy, e.g. *SONV >> *OBSV, which prefers an obstruent onset to a sonorant onset: Lombardi 2002; Smith 2003)

I did not present all constraints that may be at work for this grammar; rather I provided all and only the constraints that are apparently more relevant for the time being in accounting for the choice of CI. For example, I have not listed some constraints like Syllable Contact Law, SYLLCON ("Rising sonority across a syllable boundary is not allowed"), which seems to play a role in this grammar and will be discussed further in the next chapter.

Constraint, C is given along the lines of Yip's constraints for a specific consonant; for example, she used "p" to stand for "the Intensive prefix should end in [p]" for the Turkish reduplication case with a prefix. I use three segmental constraints, t, p, ʧ, to require that the "reduplicant in the CI-reduplication must start with /t/ (or /p/ or /ʧ/)." *REPEAT(segment) could have been presented in other term such as *REPEAT(feature) so that it can show what important role features (e.g. place of articulation and manner of articulation), rather than segments, are playing in avoiding repetition; however, I chose *REPEAT(segment) for the current data since I have not had discussion on avoidance of identical features yet, which will be examined in detail in Chapter 5. Besides, a grammar may become more powerful in its explanatory capacity if we elaborate on the constraint, e.g. *REPEAT(t): we could make it into two separate constraints, *REPEAT(t=C_1) which militates against repetition of a segment that is identical to the first consonant in the base and *REPEAT(t=C_2) which militates against repetition of a segment that is identical to the second consonant in the base. Indeed this idea is sensible enough considering relevant data in the languages that distinguish the influential status of consonants in different positions, but I do not adopt this idea in here yet since I can do without it for the current grammar.

Constraints, ONSET, DEP-BR, and MAX-BR are presented in their canonical sense. As for the Place-Markedness Hierarchy, I will use *LAB, *DOR, *COR, instead of *PL/LAB, *PL/DORS, *PL/COR, respectively, in tableaux henceforth, for convenience's sake.[41] With regard to the *ONSET/X constraints, I use several constraints like the following:

(7) Some sonority cline constraints (Lombardi 2002: 240)

a. *FRICV (prohibits fricative onset)
   *STOPV (prohibits stop onset)
   Universal ranking *FRICV >> *STOPV

b. *SONV (prohibits sonorant onset)
   *OBSV (prohibits obstruent onset)
   Universal ranking *SONV >> *OBSV

I will use *FRICV, *STOPV, *NASV – which bans the occurrence of a nasal in onset, among other sonorants: these constraints can be seen in the example tableaux in this section and in Appendix 3-C. In principle, I can utilize constraints like *GLIDEV (prohibits glide onset) and *LIQUIDV (prohibits liquid onset). I also employed *AffricV (prohibits affricate onset), the sonority value of which is not

[41] Lombardi (2002) also used these simplified forms of constraint for the Place markedness.

completely clear and indicated with (>>) due to the unclear domination relationship with *FRICV and with *STOPV; I used this constraint to demonstrate that there must be some difference between stop /t/ and affricate /ʧ/ although they are both highly frequent as CI in the data.

I employ the above constraints on a necessity basis; i.e. not all of them are shown in tableaux when they are undominated or too low in the ranking hierarchy. For instance, MAX-BR is important but I do not include it in a tableau since it is assumed to be respected by all output candidates I am putting forward. DEP-BR is also part of the grammar, but it is always critically violated by winning candidates, because epenthesis of a consonant must take place to obtain an optimal reduplicative form in our data.

The set of constraints and a grammar may not be perfect as they are presented in this section; my goal of proposing constraints in this section is not to furnish a full-fledged grammar at this point, but rather to see if the current data can be accounted for more or less with a grammar.

(8) Constraints suggested with partial hierarchy
MAX-BR, ONSET >>
*DOR, *LAB >> *COR
*NASV >> *FRICV >> *AFFRICV (>>) *STOPV
t, ʧ, p
*REPEAT(p), *REPEAT(ʧ), *REPEAT(t)
>> DEP-BR

The domination relationship among some of the constraints is not crystal clear, and we can see some examples based on these constraints and their rankings as follows: I instantiate an example from the dictionary data, and a nonword example from a word creation experiment.

(9) Tableau 1. [oson-**t**<u>oson</u>] ‘on good terms’ (=4d)

/oson-RED/	ONSET	*LAB	*COR	*AFFRICV	*STOPV	t	ʧ	p	*REPEAT (p)	*REPEAT (ʧ)	*REPEAT (t)
o.son.-<u>o.son</u>	**!										
☞ o.son.-<u>**t**o.son</u>	*		*		*		*	*			
o.son.-<u>**p**o.son</u>	*	*!			*	*	*				
☹ o.son.-<u>**ʧ**o.son</u>	*		*	*		*		*			

The actual winner in Tableau 1 should be oson-**t**<u>oson</u>, as it is the real output in lexicon; however, there is one more output that is predicted according the given grammar, i.e. oson-**ʧ**<u>oson</u>. This other candidate can be also an optimal output based on the suggested constraints and the ranking hierarchy. This is problematic since the proposed grammar cannot produce a single optimal output; on the other hand, this suggests that there could be more than one output generated by a grammar. It further implies that the strict domination relationship among constraints may not be able to account for every datum, which may need a stochastic ranking of constraints. It is possible that some constraints like t and ʧ above are overlapping in their ranges, by which t can be chosen as an optimal winner sometimes and ʧ can be chosen as an optimal winner some other times.

(10) Tableau 2. [atan-**C**<u>atan</u>] (=5e, nonce word)

/atan-RED/	ONSET	*LAB	*COR	*AFFRICV	*STOPV	t	ʧ	p	*REPEAT (p)	*REPEAT (ʧ)	*REPEAT (t)
a.tan.-<u>a.tan</u>	**!										
a.tan.-<u>ta.tan</u>	*		*		*		*	*			*!
a.tan.-<u>pa.tan</u>	*	*!			*	*	*				
☞ a.tan.-<u>ʧa.tan</u>	*		*	*		*		*			

Tableau 2 shows that /t/ in context does not welcome another /t/ to be inserted by virtue of *REPEAT(t), and therefore, an underlying form that contains /t/ ends up surfacing with the reduplicant containing other CI, /ʧ/ in the example above. This will work out very well if all speakers are sensitive to context all the time. If every speaker detests repetition of a consonant, they would avoid using the same consonant as one of the consonants in the base when they inserted a consonant in the reduplicant. However, things are not that straightforward, and we have seen that speakers tend to have their own preferred segments in consonant insertion, sensitive or insensitive to context. If a speaker, who normally prefers /t/ in consonant insertion, is not sensitive to context and does not pay attention to what there is already in context, then s/he would epenthesize /t/ in spite of existing /t/ in context. In this case, we can make use of a higher-ranked constraint like REPEAT(t), which is from a family of REPEAT constraints, to counteract *REPEAT(t) and stick to /t/ regardless of context. In this regard, we cannot be satisfied with a wholesale grammar; rather, we need to consider individual speakers' preferences.

Consequently, what we need is a stochastic grammar that can capture individual preferences, as well. In the following section, I provide the grammars that I acquired as a result of applying the learning data (from the dictionary and from Experiment 2) to the GLA, an OT-based stochastic model.

The machine-ranked grammars

In this section I will see how well the GLA can perform in generating the data based on its stochastic learning algorithm. First, I made an input file which contains multiple tableaux with underlying forms, output candidates and their frequencies in number, relevant constraints – as were presented in the preceding section, and constraint violations marked. I used the learning data that are from the dictionary data and from Experiment 2 which contains at least two distinct groups of speakers, t-dominant group and ʧ-dominant group. A partial input file for all four learning data is provided in Appendix 3-C.

With regard to the data from Experiment 2, there were three input files altogether: the entire data of the responses, the data from an individual speaker who preferred /t/ irrespective of another /t/ in context (context-insensitive speaker), and the data from an individual speaker who preferred /t/ but avoided inserting /t/ when /t/ exists in context (context-sensitive speaker).

The learning datum from the dictionary consists of 51 reduplicative forms with VCVC-bases and CI =/t, p, ʧ/. There were 204 underlying/surface pairs presented with 4 output candidates per underlying form in the input file.[42] The learning datum from Experiment 2 were also limited to the forms with VCVC-bases and CI =/t, p, ʧ/. The input file for this datum includes 444 underlying/surface pairs with 4 output candidates per underlying form (111 stimuli provided in the experiment). The learning datum for an individual speaker (S15) who is not sensitive to context in Experiment 2 consisted of 444 underlying/surface pairs with 4 output candidates per underlying form: the four output candidates were generated based on the speaker's CI choice range, /t, ʧ, n/. The learning datum for an individual speaker (S13) who is sensitive context in Experiment 2 consisted of 666 underlying/surface pairs with 6 output candidates per underlying form: the six output candidates were generated based on the speaker's CI choice range, /t, p, k, s, m/.

The resulting grammars learned by the GLA for the learning data are given in Appendix 3-D, and I examine the machine rankings for the CI-reduplication data in this section, to see how well they could capture the nature of the data.

The ranking values assigned by the GLA for the dictionary learning datum are as in the following:

[42] An output candidate with no epenthetic C, e.g. [unuk-<u>unuk</u>] (/unuk/), was presented for all learning data in the input files.

(11) Table 3.18 Machine ranking for the dictionary data

Constraint	Ranking Value
ONSET	110.000
*REPEAT(p)	110.000
*REPEAT(ʧ)	110.000
*AFFRICV	96.061
*COR	95.070
*LAB	94.930
*STOPV	93.939
ʧ	93.939
t	90.991
p	-63,219.881
*REPEAT(t)	-108,314.944

Note that the two *REPEAT constraints, *REPEAT(p) and *REPEAT(ʧ), are undominated, whereas another *REPEAT constraint, *REPEAT(t), has almost infinitely low value in ranking (all three *REPEAT constraints shaded in dark gray): first, it means that avoidance of identical consonant are really at work for the dictionary data; second, among the three consonants, /t, p, ʧ/, with high frequencies, repetition of /p/ or /ʧ/ are very unlikely to be tolerated. However, repetition of /t/ can be tolerated fairly well, which makes /t/ common and abundant as CI.[43] In the meantime, the two segmental constraints, ʧ and t (shaded in light gray) are even more highly ranked than the constraint, p, which predicts that it is more likely to have similar frequencies for /ʧ/ and /t/ as CI, whereas /p/ will be much less popular. This implies that although the three consonants /t, p, ʧ/ were among the most frequent Cs in the CI frequencies for the dictionary data, and it is eventually /t, ʧ/, but not /p/ which will be preferred in the data of CI-reduplication. This can tell us why we ended up with /t, ʧ/ that are the most frequently inserted as CI in the experiments with speakers.

The matchup between input frequency and generated frequency showed that all input forms were generated with more or less similar frequencies of CIs, /t, p, ʧ/, albeit the input frequency given only to a single output form.

---

[43] According to the ranking hierarchy, although I did not look into the ʧ-dominant group for the context-sensitivity, it seems that /ʧ/ is much more sensitive to context – in terms of avoidance of repetition, than /t/ is.

(12) Table 3.19 Matchup to input frequencies: e.g. [alok-**t**alok] 'mottled'

/alok/	Input Fr.	Gen. Fr.	Input #	Gen. #
alok-talok	1.000	0.324	198882	324266
alok-alok	0.000	0.000		
alok-palok	0.000	0.420		419747
alok-ʧalok	0.000	0.256		255987

The mismatch between the learning and generated datum, and particularly the generated frequencies divided up for the three candidate CIs hint at a possibility of variation: i.e. /t, p, ʧ/ serve as variants.

The grammar obtained from running the data of Experiment 2 produced the following ranking values, in general:

(13) Table 3.20 Machine ranking for the experimental data (Experiment 2)

Constraint	**Ranking Value**
ONSET	108.000
*REPEAT(ʧ)	100.000
*LAB	97.711
*AFFRICV	96.564
ʧ	95.436
*STOPV	95.436
*REPEAT(t)	95.342
*COR	94.289
p	94.289
t	94.275
*REPEAT(p)	93.298

The constraint, *REPEAT(ʧ) was destined to be inactive in this grammar (given the value of 100, which is an initial default value for a constraint; shaded cell in dark gray) due to the fact that none of the forms in Experiment 2 had /ʧ/ in context. The similar ranking values among the constraints in the middle of the table above (shaded in light gray) indicates that there are chances of variation among the three consonants, /t, p, ʧ/.

(14) Table 3.21 Matchup to input frequencies: e.g. [amat-**C**amat]

/amat/	Input Fr.	Gen. Fr.	Input #	Gen. #
amat-tamat	0.667	0.365	69877	365455
amat-amat	0.000	0.000		
amat-ʧamat	0.333	0.474	34378	474438
amat-pamat	0.000	0.160		160107

An example in the learning datum shows that even though the input frequency did not indicate any occurrence of /p/ as CI, the generated frequency came to have some frequency for /p/. In addition, the input frequency for /t/-inserted form, twice as high as that for /ʧ/-inserted form, decreased about the half, resulting in similar frequencies between the /t/-inserted form and the /ʧ/-inserted form in the generated frequency. This is related to the frequency distribution of CIs in Experiment 2, in which /t/ and /ʧ/ were the most frequently inserted Cs, among others.

The following is the grammar learned by the model when the learning datum is from a speaker who generally inserts /t/ with no consideration of context. Thus this speaker uses his or her preferred C in consonant insertion even if the context has the same C already.

(15) Table 3.22 Machine ranking for the data by a speaker who is *not* sensitive to context (Experiment 2)

Constraint	**Ranking Value**
ONSET	110.000
*LAB	100.000
*REPEAT(ʧ)	100.000
*AFFRICV	97.356
*NASV	96.881
*STOPV	95.763
*REPEAT(t)	95.540
t	94.237
ʧ	92.644
*COR	90.000
p	90.000

The ranking values for the constraints, *LAB, *REPEAT(ʧ), *COR, P indicate that these constraints are inactive: *LAB and P are irrelevant since the speaker did not insert /p/ at all in his or her outputs; *REPEAT(ʧ) is irrelevant since /ʧ/ was not given in context, at all; and *COR is irrelevant since all consonants that this speaker epenthesized in his or her outputs were coronals, /t, ʧ, n/.

The values for *REPEAT(t) and t are close to each other, which suggests that /t/ will be inserted without much consideration of whether /t/ exists in context. Another running of the data can possibly invert the ranking between them, and the generated frequencies generally showed a strong tendency to prefer /t/, regardless of context as in the following:

(16) Table 3.23 Matchup to input frequencies: e.g. [asam-**C**asam]

/asam/	Input Fr.	Gen. Fr.	Input #	Gen. #
asam-nasam	1.000	0.229	90908	228921
asam-sasam	0.000	0.000		
asam-tasam	0.000	0.584		584473
asam-ʧasam	0.000	0.187		186606

Although the speaker actually chose /n/ as CI for the base form of /asam/, which is shown in the input frequency, the GLA recognized the speaker's general tendency to go for /t/, which was learned through the learning datum, and this tendency is shown in the generated frequency.

The grammar resulting from running the GLA for an individual speaker who is sensitive to context is quite different from that for the speaker who is not sensitive to context:

(17) Table 3.24 Machine ranking for the data by a speaker who *is* sensitive to context (Experiment 2)

Constraint	Ranking Value
ONSET	110.000
*DOR	99.969
*NASV	98.153
*LAB	97.138
*REPEAT(t)	96.817
*FRICV	96.781
*STOPV	95.066
t	93.889
*COR	92.893
p	91.014
ʧ	90.000
*REPEAT(p)	-36,312.730

For this speaker, the ranking values between the two constraints, *REPEAT(t) and t are much more apart than for the context-insensitive speaker:

(18) Table 3.25 Machine rankings for a context-insensitive speaker vs. context-sensitive speaker (Experiment 2)

Context-insensitive speaker		Context-sensitive speaker	
Constraint	**Ranking Value**	**Constraint**	**Ranking Value**
ONSET	110.000	ONSET	110.000
*LAB	100.000	*DOR	99.969
*REPEAT(ʧ)	100.000	*NASV	98.153
*AFFRICV	97.356	*LAB	97.138
*NASV	96.881	*REPEAT(t)	96.817
*STOPV	95.763	*FRICV	96.781
*REPEAT(t)	95.540	*STOPV	95.066
t	94.237	t	93.889
ʧ	92.644	*COR	92.893
*COR	90.000	p	91.014
p	90.000	ʧ	90.000
		*REPEAT(p)	-36,312.730

The bigger difference in the ranking values between *REPEAT(t) and t, for this context-sensitive speaker, suggests that their hierarchical relationship is robust enough not to be inverted in any trials of the grammar. Therefore, /t/ is normally preferred as CI for this speaker, but an existing /t/ in context will prohibit an insertion of /t/.

(19) Table 3.26 Matchup to input frequencies: e.g. [akan-**C**akan]

/akan/	Input Fr.	Gen. Fr.	Input #	Gen. #
akan-takan	1.000	0.700	91229	699841
akan-akan	0.000	0.000		
akan-sakan	0.000	0.220		220145
akan-pakan	0.000	0.043		42520
akan-makan	0.000	0.034		34422
akan-kakan	0.000	0.003		3072

The generated frequency does not match the input frequency perfectly well; however, it can show a general tendency to prefer /t/ as in Table 3.25. This tendency gets weakened when /t/ exists in context as in the following table:

(20) Table 3.27 Matchup to input frequencies: e.g. [itip-**C**itip]

/itip/	Input Fr.	Gen. Fr.	Input #	Gen. #
itip-sitip	1.000	0.405	89610	405086
itip-itip	0.000	0.000		
itip-titip	0.000	0.315		314977
itip-pitip	0.000	0.215		214548
itip-mitip	0.000	0.041		40883
itip-kitip	0.000	0.025		24506

The speaker chose /s/ for this specific input, and it does not show his or her overall tendency for inserting Cs in the CI-reduplication. However, the machine learning algorithm could capture the general tendency via the leaning datum of this speaker's. That is, this speaker, who is sensitive to context, does not like repetition of consonants; thus s/he usually inserts /t/ but opts for other C when s/he encounters another /t/ in context (cf. generated frequencies in Table 3.25 vs. Table 3.26).

In all these grammars learned and generated by the GLA, through the learning data of the dictionary and the experiment (Experiment 2: overall grammar, individual grammars (context-insensitive and context-sensitive)), an overall tendency of variation could be captured; however, the generated frequencies did not match the input frequencies to full extent. This may be partly because a complete set of constraints was not provided and/or partly because the number of times to go through underlying forms was not enough for the machine to come up with a perfect grammar.[44] It can be also due to some other factor innate to the algorithm itself, which is to be discussed in the next section.

A caveat: Variation and gradience

We could see that the GLA, which I happened to adopt for my purpose of learning a grammar, can handle data which contain variation: as was seen in the preceding section, the GLA could generate differentiated frequencies for variants. That is, the GLA works well for those cases that have variation in the output for a single input.

We also saw that the GLA appears not to work well for the cases in which any given input is categorically mapped onto the same output. Hence we still came to have variable output forms based on generated frequencies, even for the dictionary data and the experimental results where an input came with an output with no variation. Many linguists are aware of this problem of "equating variation with gradience" (Coetzee, p.c.). After all, these two concepts, variation and gradience, are not the same, and an OT-based stochastic model like the GLA cannot deal with gradient well-formedness in contexts without variation, although it is good at handling data which have variable outputs. There have been alternative approaches to better handling the data with gradient acceptability, e.g. Coetzee & Pater 2008 (based on weighted constraints of Harmonic Grammar), Hayes & Wilson 2008 (maximum entropy model with weighted constraints), *inter alios*.

I do not attempt to find fault with the learning algorithm I utilized in this discussion; nor do I intend to repair the problem found with the algorithm. Rather, it suffices to realize that it was due to the nature of the algorithm *per se* that a correct grammar cannot be reached for categorical mappings between an input and an output. It also suffices to learn that gradient wellformedness, as well as variation, can be captured by a better developed learning model. Therefore, the

[44] It is not likely that it is due to less than enough number of times to run underlying forms through the grammar: it was recommended to run through forms million times, and I used ten million for the number. As for the other possible cause, incomplete set of constraints, I examine more grammatical factors that may affect the choice of CI in the later chapters, Chapter 4 and 5.

data of CI-reduplication, laden with variation and gradience, can be learned; that is, the data can be handled by grammar.

Appendix 3-C Sample input files

The input files I used for the learning of the dictionary data and the experimental responses (Experiment 2: for an overall grammar, for an individual grammar that is not sensitive to context, and for an individual grammar that is sensitive to context) are provided below: only some portion of each data file has been given in the interest of space.

Input file for the dictionary data:

			Onset	*Lab	*Cor	*AffricV	*StopV	t	tʃ	p	*Repeat(p)	*Repeat(tʃ)	*Repeat(t)
			Onset	*Lab	*Cor	*AffricV	*StopV	t	tʃ	p	*Repeat(p)	*Repeat(tʃ)	*Repeat(t)
ollok	ollok-ollok		2										
	ollok-**p**ollok	1	1	1			1	1	1				
	ollok-**t**ollok		1		1		1		1	1			
	ollok-**tʃ**ollok		1		1	1		1		1			
ulluk	ulluk-ulluk		2										
	ulluk-**p**ulluk	1	1	1			1	1	1				
	ulluk-**t**ulluk		1		1		1		1	1			
	ulluk-**tʃ**ulluk		1		1	1		1		1			
ulak	ulak-ulak		2										

	ulak-**p**ulak	1	1	1			1	1	1				
	ulak-**t**ulak		1		1		1		1	1			
	ulak-**ʧ**ulak		1		1	1		1		1			
alak	alak-alak		2										
	alak-**p**alak	1	1	1			1	1	1				
	alak-**t**alak		1		1		1		1	1			
	alak-**ʧ**alak		1		1	1		1		1			

Input file for the Experiment 2 data:

			Onset	*Lab	*Cor	*AffricV	*StopV	t	ʧ	p	*Repeat(p)	*Repeat(ʧ)	*Repeat(t)
			Onset	*Lab	*Cor	*AffricV	*StopV	t	ʧ	p	*Repeat(p)	*Repeat(ʧ)	*Repeat(t)
akam	akam-akam		2										
	akam-takam	7	1		1		1		1	1			
	akam-ʧakam	3	1		1	1		1		1			
	akam-pakam	1	1	1			1	1	1				
akan	akan-akan		2										
	akan-takan	6	1		1		1		1	1			
	akan-ʧakan	5	1		1	1		1		1			
	akan-pakan	1	1	1			1	1	1				

akaŋ	akaŋ-akaŋ		2										
	akaŋ-takaŋ	3	1		1		1		1	1			
	akaŋ-ʧakaŋ	8	1		1	1		1		1			
	akaŋ-pakaŋ	1	1	1			1	1	1				
akap	akap-akap		2										
	akap-takap	7	1		1		1		1	1			
	akap-ʧakap	3	1		1	1		1		1			
	akap-pakap	1	1	1			1	1	1				

Input file for data of an individual who was not sensitive to context (Experiment 2; S15):

			Onset	*Lab	*Cor	*NasV	*AfficV	*StopV	t	ʧ	p	*Repeat (ʧ)	*Repeat (t)
			Onset	*Lab	*Cor	*NasV	*AfficV	*StopV	t	ʧ	p	*Repeat (ʧ)	*Repeat (t)
akam	akam-akam		2										
	akam-takam		1		1			1		1	1		
	akam-ʧakam	1	1		1		1		1		1		
	akam-nakam		1		1	1			1	1	1		
akan	akan-akan		2										
	akan-takan		1		1			1		1	1		
	akan-ʧakan	1	1		1		1		1		1		

	akan-nakan		1		1	1			1	1	1		
akaŋ	akaŋ-akaŋ		2										
	akaŋ-takaŋ		1		1			1		1	1		
	akaŋ-ʧakaŋ		1		1		1		1		1		
	akaŋ-nakaŋ	1	1		1	1			1	1	1		
akap	akap-akap		2										
	akap-takap	1	1		1			1		1	1		
	akap-ʧakap		1		1		1		1		1		
	akap-nakap		1		1	1			1	1	1		

Input file for data of an individual who was sensitive to context (Experiment 2; S13):

			Onset	*Dor	*Lab	*Cor	*NasV	*FricV	*StopV	t	ʧ	p	*Repeat(p)	*Repeat(t)
			Onset	*Dor	*Lab	*Cor	*NasV	*FricV	*StopV	t	ʧ	p	*Repeat(p)	*Repeat(t)
akam	akam-akam		2											
	akam-takam	1	1			1			1		1	1		
	akam-sakam		1			1		1		1	1	1		
	akam-pakam		1		1				1	1	1			
	akam-makam		1		1		1			1	1	1		
	akam-kakam		1	1					1	1	1	1		

akan	akan-akan		2											
	akan-takan	1	1			1			1		1	1		
	akan-sakan		1			1		1		1	1	1		
	akan-pakan		1		1				1	1	1			
	akan-makan		1		1		1			1	1	1		
	akan-kakan		1	1					1	1	1	1		
akaŋ	akaŋ-akaŋ		2											
	akaŋ-takaŋ	1	1			1			1		1	1		
	akaŋ-sakaŋ		1			1		1		1	1	1		
	akaŋ-pakaŋ		1		1				1	1	1			
	akaŋ-makaŋ		1		1		1			1	1	1		
	akaŋ-kakaŋ		1	1					1	1	1	1		
akap	akap-akap		2											
	akap-takap	1	1			1			1		1	1		
	akap-sakap		1			1		1		1	1	1		
	akap-pakap		1		1				1	1	1		1	
	akap-makap		1		1		1			1	1	1		
	akap-kakap		1	1					1	1	1	1		

Appendix 3-D Resulting grammars

Resulting grammars from running the learning data of the dictionary, Experiment 2, and two kinds of data for individual speakers, context-insensitive and context-sensitive, who belong to the t-dominant group in Experiment 2:

Learning of the dictionary data:[45]

```
Result of Applying Gradual Learning Algorithm to otsoft_dictionary_ch3 (1).xls
OTSoft 2.3, release date 5/15/08

1. Ranking Values Found

110.000      Onset
110.000      *Repeat(p)
110.000      *Repeat(ʧ)
96.061       *AffricV
95.070       *Cor
94.930       *Lab
93.939       *StopV
93.939       ʧ
90.991       t
-63,219.881   p
-108,314.944   *Repeat(t)

2. Matchup to Input Frequencies[46]

   /ollok/         Input Fr. Gen Fr.  Input #     Gen. #
   ollok-pollok      1.000   0.420     198419     419747
   ollok-ollok       0.000   0.000
   ollok-tollok      0.000   0.324                324266
   ollok-ʧollok      0.000   0.256                255987

   /ulluk/         Input Fr. Gen Fr.  Input #     Gen. #
   ulluk-pulluk      1.000   0.420     194533     419747
```

[45] In the dictionary data, vowels were not regulated: I did not have interests in vowels, focusing on consonants in the base in relation to CI.
[46] Geminate /ll/ was treated as a singleton /l/.

ulluk-ulluk	0.000	0.000		
ulluk-tulluk	0.000	0.324		324266
ulluk-ʧulluk	0.000	0.256		255987

/ulak/	Input Fr.	Gen Fr.	Input #	Gen. #
ulak-pulak	1.000	0.420	197086	419747
ulak-ulak	0.000	0.000		
ulak-tulak	0.000	0.324		324266
ulak-ʧulak	0.000	0.256		255987

/alak/	Input Fr.	Gen Fr.	Input #	Gen. #
alak-palak	1.000	0.420	198782	419747
alak-alak	0.000	0.000		
alak-talak	0.000	0.324		324266
alak-ʧalak	0.000	0.256		255987

/ʌt’ɨk/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌt’ɨk-pit’ɨk	1.000	0.420	196228	419747
ʌt’ɨk-it’ɨk	0.000	0.000		
ʌt’ɨk-tit’ɨk	0.000	0.324		324266
ʌt’ɨk-ʧit’ɨk	0.000	0.256		255987

/atɨk/	Input Fr.	Gen Fr.	Input #	Gen. #
atɨk-patɨk	1.000	0.420	195913	419747
atɨk-atɨk	0.000	0.000		
atɨk-tatɨk	0.000	0.324		324266
atɨk-ʧatɨk	0.000	0.256		255987

/okɨl/	Input Fr.	Gen Fr.	Input #	Gen. #
okɨl-pokɨl	1.000	0.420	195963	419747
okɨl-okɨl	0.000	0.000		
okɨl-tokɨl	0.000	0.324		324266
okɨl-ʧokɨl	0.000	0.256		255987

/ukɨl/	Input Fr.	Gen Fr.	Input #	Gen. #
ukɨl-pukɨl	1.000	0.420	194477	419747
ukɨl-ukɨl	0.000	0.000		
ukɨl-tukɨl	0.000	0.324		324266

ukʔl-ʧukɨl	0.000	0.256		255987

/uʧil/	Input Fr.	Gen Fr.	Input #	Gen. #
uʧil-puʧil	1.000	0.508	198013	507757
uʧil-uʧil	0.000	0.000		
uʧil-tuʧil	0.000	0.492		492243
uʧil-ʧuʧil	0.000	0.000		

/ʌʧ^{h}il/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌʧ^{h}il-piʧ^{h}il	1.000	0.508	195546	507757
ʌʧ^{h}il-iʧ^{h}il	0.000	0.000		
ʌʧ^{h}il-tiʧ^{h}il	0.000	0.492		492243
ʌʧ^{h}il-ʧiʧ^{h}il	0.000	0.000		

/ʌkɨm/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌkɨm-pʌkɨm	1.000	0.420	195021	419747
ʌkɨm-ʌkɨm	0.000	0.000		
ʌkɨm-tʌkɨm	0.000	0.324		324266
ʌkɨm-ʧʌkɨm	0.000	0.256		255987

/ʌjʌŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌjʌŋ-puʌʔŋ	1.000	0.420	197356	419747
ʌjʌŋ-ujʌŋ	0.000	0.000		
ʌjʌŋ-tujʌŋ	0.000	0.324		324266
ʌjʌŋ-ʧujʌŋ	0.000	0.256		255987

/ʌkɨn/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌkɨn-pʌkɨn	1.000	0.420	196294	419747
ʌkɨn-ʌkɨn	0.000	0.000		
ʌkɨn-tʌkɨn	0.000	0.324		324266
ʌkɨn-ʧʌkɨn	0.000	0.256		255987

/akɨn/	Input Fr.	Gen Fr.	Input #	Gen. #
akɨn-pakɨn	1.000	0.420	196651	419747
akɨn-akɨn	0.000	0.000		
akɨn-takɨn	0.000	0.324		324266
akɨn-ʧakɨn	0.000	0.256		255987

/otoŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
otoŋ-potoŋ	1.000	0.420	195130	419747
otoŋ-otoŋ	0.000	0.000		
otoŋ-totoŋ	0.000	0.324		324266
otoŋ-ʧotoŋ	0.000	0.256		255987

/utuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
utuŋ-putuŋ	1.000	0.420	198166	419747
utuŋ-utuŋ	0.000	0.000		
utuŋ-tutuŋ	0.000	0.324		324266
utuŋ-ʧutuŋ	0.000	0.256		255987

/otoŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
otoŋ-p^{h}otoŋ	1.000	0.420	196665	419747
otoŋ-otoŋ	0.000	0.000		
otoŋ-t^{h}otoŋ	0.000	0.324		324266
otoŋ-ʧ^{h}otoŋ	0.000	0.256		255987

/utuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
utuŋ-p^{h}utuŋ	1.000	0.420	195934	419747
utuŋ-utuŋ	0.000	0.000		
utuŋ-t^{h}utuŋ	0.000	0.324		324266
utuŋ-ʧ^{h}utuŋ	0.000	0.256		255987

/atɨŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
atɨŋ-patɨŋ	1.000	0.420	194609	419747
atɨŋ-atɨŋ	0.000	0.000		
atɨŋ-tatɨŋ	0.000	0.324		324266
atɨŋ-ʧatɨŋ	0.000	0.256		255987

/ʌʧʌŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌʧʌŋ-pʌʧʌŋ	1.000	0.508	196161	507757
ʌʧʌŋ-ʌʧʌŋ	0.000	0.000		
ʌʧʌŋ-tʌʧʌŋ	0.000	0.492		492243
ʌʧʌŋ-ʧʌʧʌŋ	0.000	0.000		

/aʧaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
aʧaŋ-paʧaŋ	1.000	0.508	194487	507757

atʃaŋ-atʃaŋ	0.000	0.000		
atʃaŋ-tatʃaŋ	0.000	0.492		492243
atʃaŋ-tʃatʃaŋ	0.000	0.000		

/allok/	Input Fr.	Gen Fr.	Input #	Gen. #
allok-tallok	1.000	0.324	195218	324266
allok-allok	0.000	0.000		
allok-pallok	0.000	0.420		419747
allok-tʃallok	0.000	0.256		255987

/ʌluk/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌluk-tʌluk	1.000	0.324	193942	324266
ʌluk-ʌluk	0.000	0.000		
ʌluk-pʌluk	0.000	0.420		419747
ʌluk-tʃʌluk	0.000	0.256		255987

/ʌllʌk/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌllʌk-tʌllʌk	1.000	0.324	197695	324266
ʌllʌk-ʌllʌk	0.000	0.000		
ʌllʌk-pʌllʌk	0.000	0.420		419747
ʌllʌk-tʃʌllʌk	0.000	0.256		255987

/ʌlluk/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌlluk-tʌlluk	1.000	0.324	195753	324266
ʌlluk-ʌlluk	0.000	0.000		
ʌlluk-pʌlluk	0.000	0.420		419747
ʌlluk-tʃʌlluk	0.000	0.256		255987

/alok/	Input Fr.	Gen Fr.	Input #	Gen. #
alok-talok	1.000	0.324	198882	324266
alok-alok	0.000	0.000		
alok-palok	0.000	0.420		419747
alok-tʃalok	0.000	0.256		255987

/allak/	Input Fr.	Gen Fr.	Input #	Gen. #
allak-tallak	1.000	0.324	195574	324266
allak-allak	0.000	0.000		
allak-pallak	0.000	0.420		419747

allak-ʧallak	0.000	0.256		255987

/otol/	Input Fr.	Gen Fr.	Input #	Gen. #
otol-t^{h}otol	1.000	0.324	196411	324266
otol-otol	0.000	0.000		
otol-p^{h}otol	0.000	0.420		419747
otol-ʧ^{h}otol	0.000	0.256		255987

/utul/	Input Fr.	Gen Fr.	Input #	Gen. #
utul-t^{h}utul	1.000	0.324	195406	324266
utul-utul	0.000	0.000		
utul-p^{h}utul	0.000	0.420		419747
utul-ʧ^{h}utul	0.000	0.256		255987

/ot^{h}ol/	Input Fr.	Gen Fr.	Input #	Gen. #
ot^{h}ol-tot^{h}ol	1.000	0.324	197311	324266
ot^{h}ol-ot^{h}ol	0.000	0.000		
ot^{h}ol-pot^{h}ol	0.000	0.420		419747
ot^{h}ol-ʧot^{h}ol	0.000	0.256		255987

/ut^{h}ul/	Input Fr.	Gen Fr.	Input #	Gen. #
ut^{h}ul-tut^{h}ul	1.000	0.324	195049	324266
ut^{h}ul-ut^{h}ul	0.000	0.000		
ut^{h}ul-put^{h}ul	0.000	0.420		419747
ut^{h}ul-ʧut^{h}ul	0.000	0.256		255987

/ʌlʌn/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌlʌn-tʌlʌn	1.000	0.324	197421	324266
ʌlʌn-ʌlʌn	0.000	0.000		
ʌlʌn-pʌlʌn	0.000	0.420		419747
ʌlʌn-ʧʌlʌn	0.000	0.256		255987

/alloŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
alloŋ-talloŋ	1.000	0.324	195351	324266
alloŋ-alloŋ	0.000	0.000		
alloŋ-palloŋ	0.000	0.420		419747
alloŋ-ʧalloŋ	0.000	0.256		255987

/ʌluŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌluŋ-tʌluŋ	1.000	0.324	193762	324266
ʌluŋ-ʌluŋ	0.000	0.000		
ʌluŋ-pʌluŋ	0.000	0.420		419747
ʌluŋ-ʧʌluŋ	0.000	0.256		255987

/ʌlluŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌlluŋ-tʌlluŋ	1.000	0.324	197867	324266
ʌlluŋ-ʌlluŋ	0.000	0.000		
ʌlluŋ-pʌlluŋ	0.000	0.420		419747
ʌlluŋ-ʧʌlluŋ	0.000	0.256		255987

/aloŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
aloŋ-taloŋ	1.000	0.324	193102	324266
aloŋ-aloŋ	0.000	0.000		
aloŋ-paloŋ	0.000	0.420		419747
aloŋ-ʧaloŋ	0.000	0.256		255987

/oson/	Input Fr.	Gen Fr.	Input #	Gen. #
oson-toson	1.000	0.324	196736	324266
oson-oson	0.000	0.000		
oson-poson	0.000	0.420		419747
oson-ʧoson	0.000	0.256		255987

/osun/	Input Fr.	Gen Fr.	Input #	Gen. #
osun-tosun	1.000	0.324	196830	324266
osun-osun	0.000	0.000		
osun-posun	0.000	0.420		419747
osun-ʧosun	0.000	0.256		255987

/etoŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
etoŋ-tetoŋ	1.000	0.324	195221	324266
etoŋ-etoŋ	0.000	0.000		
etoŋ-petoŋ	0.000	0.420		419747
etoŋ-ʧetoŋ	0.000	0.256		255987

/ukʌk/	Input Fr.	Gen Fr.	Input #	Gen. #
ukʌk-ʧikʌk	1.000	0.256	196525	255987

ukʌk-uk?k	0.000	0.000		
ukʌk-pikʌk	0.000	0.420		419747
ukʌk-tikʌk	0.000	0.324		324266

/omok/	Input Fr.	Gen Fr.	Input #	Gen. #
omok-ʧomok	1.000	0.256	194854	255987
omok-omok	0.000	0.000		
omok-pomok	0.000	0.420		419747
omok-tomok	0.000	0.324		324266

/umuk/	Input Fr.	Gen Fr.	Input #	Gen. #
umuk-ʧumuk	1.000	0.256	196332	255987
umuk-umuk	0.000	0.000		
umuk-pumuk	0.000	0.420		419747
umuk-tumuk	0.000	0.324		324266

/ʌp’ak/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌp’ak-ʧap’ak	1.000	0.343	196480	342822
ʌp’ak-ʌp’ak	0.000	0.000		
ʌp’ak-pap’ak	0.000	0.000		
ʌp’ak-tap’ak	0.000	0.657		657178

/okɨl/	Input Fr.	Gen Fr.	Input #	Gen. #
okɨl-ʧ’okɨl	1.000	0.256	195326	255987
okɨl-okɨl	0.000	0.000		
okɨl-p’okɨl	0.000	0.420		419747
okɨl-t’okɨl	0.000	0.324		324266

/ukɨl/	Input Fr.	Gen Fr.	Input #	Gen. #
ukɨl-ʧ’ukɨl	1.000	0.256	197279	255987
ukɨl-ukɨl	0.000	0.000		
ukɨl-p’ukɨl	0.000	0.420		419747
ukɨl-t’ukɨl	0.000	0.324		324266

/omil/	Input Fr.	Gen Fr.	Input #	Gen. #
omil-ʧomil	1.000	0.256	195367	255987
omil-omil	0.000	0.000		
omil-pomil	0.000	0.420		419747

omil-tomil	0.000	0.324		324266

/umul/	Input Fr.	Gen Fr.	Input #	Gen. #
umul-ʧ’umul	1.000	0.256	193673	255987
umul-umul	0.000	0.000		
umul-p’umul	0.000	0.420		419747
umul-t’umul	0.000	0.324		324266

/ʌkɨm/	Input Fr.	Gen Fr.	Input #	Gen. #
ʌkɨm-ʧikɨm	1.000	0.256	194797	255987
ʌkɨm-ʌkɨm	0.000	0.000		
ʌkɨm-pikɨm	0.000	0.420		419747
ʌkɨm-tikɨm	0.000	0.324		324266

/oloŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
oloŋ-ʧoloŋ	1.000	0.256	198762	255987
oloŋ-oloŋ	0.000	0.000		
oloŋ-poloŋ	0.000	0.420		419747
oloŋ-toloŋ	0.000	0.324		324266

/ollaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ollaŋ-ʧʰollaŋ	1.000	0.256	195734	255987
ollaŋ-ollaŋ	0.000	0.000		
ollaŋ-pʰollaŋ	0.000	0.420		419747
ollaŋ-tʰollaŋ	0.000	0.324		324266

/ullʌŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ullʌŋ-ʧʰullʌŋ	1.000	0.256	195906	255987
ullʌŋ-ullʌŋ	0.000	0.000		
ullʌŋ-pʰullʌŋ	0.000	0.420		419747
ullʌŋ-tʰullʌŋ	0.000	0.324		324266

3. Tableaux[47]

The following are approximate tableaux for this ranking.

[47] I have not put all the tableaux generated, for reasons of space, showing representative ones. I am also presenting some example tableaux for other resulting grammars, henceforth.

Outputs are derived simply by sorting the constraints by their ranking value, with no stochastic variation.

To diagnose variation, consult two things:

--The candidate frequencies (which are the generated frequencies, not the input frequencies).
--The probability that each constraint outranks the next one down, given directly after the constraint labels.

/otoŋ/:

	Onset (.501)	*Repeat(p) (.501)	*Repeat(?)(1)	*AffricV (.638)	*Cor (.52)	*Lab (.638)	*StopV (.501)	? (.852)	t(1)	p(1)	*Repeat(t)
otoŋ-potoŋ (0.420)	1					1	1	1	1		
otoŋ-totoŋ (0.324)	1				1!		1	1		1	1
otoŋ-?otoŋ (0.256)	1			1!	1				1	1	
otoŋ-otoŋ	2!										

/oson/:

	Onset (.501)	*Repeat(p) (.501)	*Repeat(?)(1)	*AffricV (.638)	*Cor (.52)	*Lab (.638)	*StopV (.501)	? (.852)	t(1)	p(1)	*Repeat(t)
oson-poson (0.420)	1					1	1	1	1		
oson-toson (0.324)	1				1!		1	1		1	
oson-?oson (0.256)	1			1!	1				1	1	
oson-oson	2!										

4. Active Constraints

A constraint is active if it causes the winning candidate to defeat a rival in at least one competition.

Active	Onset
Active	*Repeat(p)
Active	*Repeat(ʧ)
Active	*AffricV
Active	*Cor
Active	*Lab
Active	*StopV
Active	ʧ
Active	t
Inactive	p
Inactive	*Repeat(t)

5. Testing the Grammar: Details

```
   The grammar was tested for 1000000 cycles.
   Average error per candidate:  16.265 percent
   Learning time:  2.240 minutes

6. Parameter Values Used by the GLA

   Initial Rankings

      All constraints started out at the default value of 100.

   Schedule for GLA Parameters

      Stage   Trials   PlastMark  PlastFaith NoiseMark  NoiseFaith
1    12500000    25000002.000      2.0002.000      2.0002.000      2.0002.000      2.000
2    22500000    25000000.159      0.1590.159      0.1592.000      2.0002.000      2.000
3    32500000    25000000.013      0.0130.013      0.0132.000      2.0002.000      2.000
4    42500000    25000000.001      0.0010.001      0.0012.000      2.0002.000      2.000

      There were a total of 10000000 learning trials.

----------
```

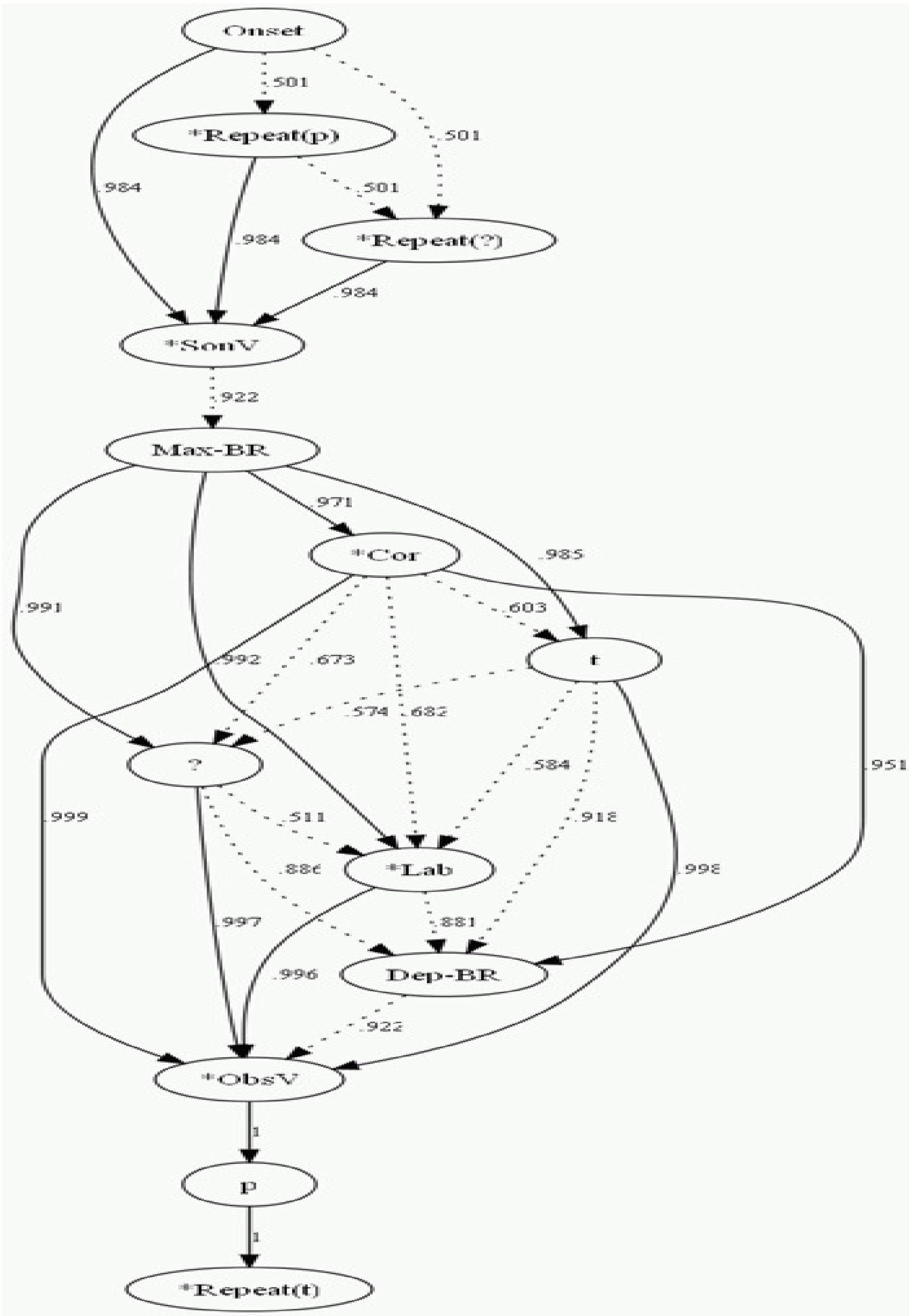
Hasse Diagram that shows the ranking hierarchy: the dictionary data[48],[49]

The stochastic ranking is shown with a labeled arc which indicates a probability that one constraint will outrank another on any given speaking occasion. If the probability is less than .95, a dotted line is used, which means that the opposite ranking is reasonably common; otherwise, a solid line is used.

---

[48] Hsse diagram is provided only for the dictionary data.

[49] A symbol, ? actually indicates ʧ in the diagram: since the diagram is presented in a picture format, I could not correct the symbol.


Onset
.501
.501
*Repeat(p)
.984
.501
.984
*Repeat(?)
.984
*SonV
.922
Max-BR
.971
*Cor
.985
.991
.603
.992
.673
t
.574
.682
?
.584
.951
.999
.511
.918
*Lab
.886
.998
.881
.997
.996
Dep-BR
.922
*ObsV
1
p
1
*Repeat(t)

Learning of the Experiment 2 data:
VCVC-bases, CI=/t, p, ʧ/

Result of Applying Gradual Learning Algorithm to otsoft_expt2_ch3 (1).xls
OTSoft 2.3, release date 5/15/08

1. Ranking Values Found

108.000	Onset
100.000	*Repeat(ʧ)
97.711	*Lab
96.564	*AffricV
95.436	ʧ
95.436	*StopV
95.342	*Repeat(t)
94.289	*Cor
94.289	p
94.275	t
93.298	*Repeat(p)

2. Matchup to Input Frequencies

/akam/	Input Fr.	Gen Fr.	Input #	Gen. #
akam-takam	0.636	0.467	61518	466502
akam-akam	0.000	0.000		
akam-ʧakam	0.273	0.415	25922	414775
akam-pakam	0.091	0.119	9157	118723

/akan/	Input Fr.	Gen Fr.	Input #	Gen. #
akan-takan	0.500	0.467	52847	466502
akan-akan	0.000	0.000		
akan-ʧakan	0.417	0.415	43708	414775
akan-pakan	0.083	0.119	8964	118723

/akaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
akaŋ-ʧakaŋ	0.667	0.415	68618	414775
akaŋ-akaŋ	0.000	0.000		
akaŋ-takaŋ	0.250	0.467	25597	466502

akaŋ-pakaŋ	0.083	0.119	8366	118723

/akap/	Input Fr.	Gen Fr.	Input #	Gen. #
akap-takap	0.636	0.467	60592	466502
akap-akap	0.000	0.000		
akap-ʧakap	0.273	0.415	25483	414775
akap-pakap	0.091	0.119	8543	118723

/akat/	Input Fr.	Gen Fr.	Input #	Gen. #
akat-takat	0.462	0.365	52788	365455
akat-akat	0.000	0.000		
akat-ʧakat	0.462	0.474	52911	474438
akat-pakat	0.077	0.160	8568	160107

/amak/	Input Fr.	Gen Fr.	Input #	Gen. #
amak-ʧamak	0.556	0.415	42466	414775
amak-amak	0.000	0.000		
amak-tamak	0.333	0.467	26082	466502
amak-pamak	0.111	0.119	8989	118723

/aman/	Input Fr.	Gen Fr.	Input #	Gen. #
aman-taman	0.636	0.467	61027	466502
aman-aman	0.000	0.000		
aman-ʧaman	0.364	0.415	35365	414775
aman-paman	0.000	0.119		118723

/amaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
amaŋ-tamaŋ	0.727	0.467	70189	466502
amaŋ-amaŋ	0.000	0.000		
amaŋ-ʧamaŋ	0.273	0.415	26915	414775
amaŋ-pamaŋ	0.000	0.119		118723

/amap/	Input Fr.	Gen Fr.	Input #	Gen. #
amap-tamap	0.583	0.471	60915	471018
amap-amap	0.000	0.000		
amap-ʧamap	0.417	0.416	42899	415509
amap-pamap	0.000	0.113		113473

/amat/	Input Fr.	Gen Fr.	Input #	Gen. #
amat-tamat	0.667	0.365	69877	365455
amat-amat	0.000	0.000		
amat-ʧamat	0.333	0.474	34378	474438
amat-pamat	0.000	0.160		160107

/anak/	Input Fr.	Gen Fr.	Input #	Gen. #
anak-tanak	0.455	0.467	43441	466502
anak-anak	0.000	0.000		
anak-ʧanak	0.455	0.415	44085	414775
anak-panak	0.091	0.119	9209	118723

/aŋak/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋak-ʧaŋak	0.636	0.415	59474	414775
aŋak-aŋak	0.000	0.000		
aŋak-taŋak	0.273	0.467	25175	466502
aŋak-paŋak	0.091	0.119	8882	118723

/aŋam/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋam-taŋam	0.778	0.467	60490	466502
aŋam-aŋam	0.000	0.000		
aŋam-ʧaŋam	0.000	0.415		414775
aŋam-paŋam	0.222	0.119	17553	118723

/anam/	Input Fr.	Gen Fr.	Input #	Gen. #
anam-tanam	0.400	0.467	35959	466502
anam-anam	0.000	0.000		
anam-ʧanam	0.300	0.415	25407	414775
anam-panam	0.300	0.119	26130	118723

/aŋan/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋan-taŋan	0.818	0.467	75691	466502
aŋan-aŋan	0.000	0.000		
aŋan-ʧaŋan	0.091	0.415	8534	414775
aŋan-paŋan	0.091	0.119	8485	118723

/anaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
anaŋ-tanaŋ	0.455	0.467	43203	466502

anaŋ-anaŋ	0.000	0.000		
anaŋ-ʧanaŋ	0.182	0.415	17183	414775
anaŋ-panaŋ	0.364	0.119	34827	118723

/anap/	Input Fr.	Gen Fr.	Input #	Gen. #
anap-tanap	0.556	0.471	43471	471018
anap-anap	0.000	0.000		
anap-ʧanap	0.222	0.416	17584	415509
anap-panap	0.222	0.113	17445	113473

/aŋap/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋap-taŋap	0.667	0.471	51621	471018
aŋap-aŋap	0.000	0.000		
aŋap-ʧaŋap	0.111	0.416	8962	415509
aŋap-paŋap	0.222	0.113	17846	113473

/aŋat/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋat-paŋat	0.400	0.160	35311	160107
aŋat-aŋat	0.000	0.000		
aŋat-taŋat	0.300	0.365	26125	365455
aŋat-ʧaŋat	0.300	0.474	26262	474438

/anat/	Input Fr.	Gen Fr.	Input #	Gen. #
anat-tanat	0.571	0.365	34838	365455
anat-anat	0.000	0.000		
anat-ʧanat	0.286	0.474	16902	474438
anat-panat	0.143	0.160	8613	160107

/apak/	Input Fr.	Gen Fr.	Input #	Gen. #
apak-tapak	0.615	0.471	69272	471018
apak-apak	0.000	0.000		
apak-ʧapak	0.385	0.416	43273	415509
apak-papak	0.000	0.113		113473

/apam/	Input Fr.	Gen Fr.	Input #	Gen. #
apam-tapam	0.818	0.467	77012	466502
apam-apam	0.000	0.000		
apam-ʧapam	0.091	0.415	8839	414775

apam-papam	0.091	0.119	8518	118723

/apan/	Input Fr.	Gen Fr.	Input #	Gen. #
apan-ʧapan	0.583	0.416	60118	415509
apan-apan	0.000	0.000		
apan-tapan	0.417	0.471	43529	471018
apan-papan	0.000	0.113		113473

/apaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
apaŋ-tapaŋ	0.800	0.471	70336	471018
apaŋ-apaŋ	0.000	0.000		
apaŋ-ʧapaŋ	0.100	0.416	9011	415509
apaŋ-papaŋ	0.100	0.113	8230	113473

/apat/	Input Fr.	Gen Fr.	Input #	Gen. #
apat-tapat	0.889	0.370	69771	369666
apat-apat	0.000	0.000		
apat-ʧapat	0.111	0.476	8300	476165
apat-papat	0.000	0.154		154169

/asak/	Input Fr.	Gen Fr.	Input #	Gen. #
asak-tasak	0.500	0.467	52324	466502
asak-asak	0.000	0.000		
asak-ʧasak	0.083	0.415	8510	414775
asak-pasak	0.417	0.119	43775	118723

/asam/	Input Fr.	Gen Fr.	Input #	Gen. #
asam-tasam	0.600	0.467	51339	466502
asam-sasam	0.000	0.000		
asam-ʧasam	0.300	0.415	26167	414775
asam-pasam	0.100	0.119	8555	118723

/asan/	Input Fr.	Gen Fr.	Input #	Gen. #
asan-tasan	0.600	0.467	52092	466502
asan-asan	0.000	0.000		
asan-ʧasan	0.300	0.415	25686	414775
asan-pasan	0.100	0.119	8436	118723

/asaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
asaŋ-ʧasaŋ	0.600	0.415	52191	414775
asaŋ-asaŋ	0.000	0.000		
asaŋ-tasaŋ	0.300	0.467	26253	466502
asaŋ-pasaŋ	0.100	0.119	8852	118723

/asap/	Input Fr.	Gen Fr.	Input #	Gen. #
asap-tasap	0.364	0.467	34944	466502
asap-asap	0.000	0.000		
asap-ʧasap	0.273	0.415	26408	414775
asap-pasap	0.364	0.119	34079	118723

/asat/	Input Fr.	Gen Fr.	Input #	Gen. #
asat-tasat	0.545	0.365	52002	365455
asat-asat	0.000	0.000		
asat-ʧasat	0.182	0.474	17826	474438
asat-pasat	0.273	0.160	25937	160107

/atak/	Input Fr.	Gen Fr.	Input #	Gen. #
atak-tatak	0.333	0.365	25683	365455
atak-atak	0.000	0.000		
atak-ʧatak	0.333	0.474	26397	474438
atak-patak	0.333	0.160	25363	160107

/atam/	Input Fr.	Gen Fr.	Input #	Gen. #
atam-tatam	0.556	0.365	43794	365455
atam-atam	0.000	0.000		
atam-ʧatam	0.333	0.474	26801	474438
atam-patam	0.111	0.160	8674	160107

/ataŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ataŋ-tataŋ	0.400	0.365	35815	365455
ataŋ-ataŋ	0.000	0.000		
ataŋ-ʧataŋ	0.300	0.474	26320	474438
ataŋ-pataŋ	0.300	0.160	25632	160107

/atan/	Input Fr.	Gen Fr.	Input #	Gen. #
atan-tatan	0.625	0.365	43244	365455

atan-atan	0.000	0.000		
atan-ʧatan	0.250	0.474	17191	474438
atan-patan	0.125	0.160	8585	160107

/atap/	Input Fr.	Gen Fr.	Input #	Gen. #
atap-tatap	0.750	0.370	26068	369666
atap-atap	0.000	0.000		
atap-ʧatap	0.000	0.476		476165
atap-patap	0.250	0.154	8738	154169

/atat/	Input Fr.	Gen Fr.	Input #	Gen. #
atat-ʧatat	0.556	0.474	43697	474438
atat-atat	0.000	0.000		
atat-tatat	0.333	0.365	25711	365455
atat-patat	0.111	0.160	8698	160107

/ikim/	Input Fr.	Gen Fr.	Input #	Gen. #
ikim-ʧikim	0.583	0.415	60608	414775
ikim-ikim	0.000	0.000		
ikim-tikim	0.333	0.467	34217	466502
ikim-pikim	0.083	0.119	8775	118723

/ikin/	Input Fr.	Gen Fr.	Input #	Gen. #
ikin-ʧikin	0.700	0.415	60648	414775
ikin-ikin	0.000	0.000		
ikin-tikin	0.300	0.467	26594	466502
ikin-pikin	0.000	0.119		118723

/ikiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ikiŋ-ʧikiŋ	0.667	0.415	51773	414775
ikiŋ-ikiŋ	0.000	0.000		
ikiŋ-tikiŋ	0.333	0.467	26212	466502
ikiŋ-pikiŋ	0.000	0.119		118723

/ikip/	Input Fr.	Gen Fr.	Input #	Gen. #
ikip-ʧikip	0.667	0.416	69384	415509
ikip-ikip	0.000	0.000		
ikip-tikip	0.250	0.471	25528	471018

ikip-pikip	0.083	0.113	8263	113473

/ikit/	Input Fr.	Gen Fr.	Input #	Gen. #
ikit-ʧikit	0.636	0.474	59820	474438
ikit-ikit	0.000	0.000		
ikit-tikit	0.273	0.365	26424	365455
ikit-pikit	0.091	0.160	8798	160107

/imik/	Input Fr.	Gen Fr.	Input #	Gen. #
imik-ʧimik	0.800	0.415	69154	414775
imik-imik	0.000	0.000		
imik-timik	0.200	0.467	17888	466502
imik-pimik	0.000	0.119		118723

/imiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
imiŋ-timiŋ	0.556	0.467	43243	466502
imiŋ-imiŋ	0.000	0.000		
imiŋ-ʧimiŋ	0.333	0.415	25344	414775
imiŋ-pimiŋ	0.111	0.119	8488	118723

/imin/	Input Fr.	Gen Fr.	Input #	Gen. #
imin-ʧimin	0.667	0.415	51982	414775
imin-imin	0.000	0.000		
imin-timin	0.222	0.467	17169	466502
imin-pimin	0.111	0.119	8708	118723

/imip/	Input Fr.	Gen Fr.	Input #	Gen. #
imip-ʧimip	0.667	0.416	51049	415509
imip-imip	0.000	0.000		
imip-timip	0.333	0.471	25799	471018
imip-pimip	0.000	0.113		113473

/imit/	Input Fr.	Gen Fr.	Input #	Gen. #
imit-ʧimit	0.667	0.474	51405	474438
imit-imit	0.000	0.000		
imit-timit	0.333	0.365	26151	365455
imit-pimit	0.000	0.160		160107

/iŋik/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋik-ʧiŋik	0.556	0.415	42289	414775
iŋik-iŋik	0.000	0.000		
iŋik-tiŋik	0.333	0.467	25661	466502
iŋik-piŋik	0.111	0.119	8399	118723

/inik/	Input Fr.	Gen Fr.	Input #	Gen. #
inik-ʧinik	0.500	0.415	53049	414775
inik-inik	0.000	0.000		
inik-tinik	0.333	0.467	34330	466502
inik-pinik	0.167	0.119	17627	118723

/iŋim/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋim-ʧiŋim	0.700	0.415	59676	414775
iŋim-iŋim	0.000	0.000		
iŋim-tiŋim	0.300	0.467	25284	466502
iŋim-piŋim	0.000	0.119		118723

/inim/	Input Fr.	Gen Fr.	Input #	Gen. #
inim-ʧinim	0.636	0.415	59577	414775
inim-inim	0.000	0.000		
inim-tinim	0.364	0.467	35632	466502
inim-pinim	0.000	0.119		118723

/iŋin/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋin-ʧiŋin	0.667	0.415	70082	414775
iŋin-iŋin	0.000	0.000		
iŋin-tiŋin	0.333	0.467	34241	466502
iŋin-piŋin	0.000	0.119		118723

/iniŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
iniŋ-ʧiniŋ	0.545	0.415	52530	414775
iniŋ-iniŋ	0.000	0.000		
iniŋ-tiniŋ	0.273	0.467	26009	466502
iniŋ-piniŋ	0.182	0.119	17172	118723

/inip/	Input Fr.	Gen Fr.	Input #	Gen. #
inip-ʧinip	0.667	0.416	69555	415509

inip-inip	0.000	0.000		
inip-tinip	0.333	0.471	34597	471018
inip-pinip	0.000	0.113		113473

/iŋip/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋip-tiŋip	0.545	0.471	51775	471018
iŋip-iŋip	0.000	0.000		
iŋip-ʧiŋip	0.364	0.416	35079	415509
iŋip-piŋip	0.091	0.113	9128	113473

/iŋit/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋit-tiŋit	0.444	0.365	34663	365455
iŋit-iŋit	0.000	0.000		
iŋit-ʧiŋit	0.444	0.474	34942	474438
iŋit-piŋit	0.111	0.160	8563	160107

/init/	Input Fr.	Gen Fr.	Input #	Gen. #
init-ʧinit	0.636	0.474	62211	474438
init-init	0.000	0.000		
init-tinit	0.182	0.365	16912	365455
init-pinit	0.182	0.160	17737	160107

/ipik/	Input Fr.	Gen Fr.	Input #	Gen. #
ipik-tipik	0.500	0.471	36094	471018
ipik-ipik	0.000	0.000		
ipik-ʧipik	0.500	0.416	34300	415509
ipik-pipik	0.000	0.113		113473

/ipim/	Input Fr.	Gen Fr.	Input #	Gen. #
ipim-ʧipim	0.727	0.416	69181	415509
ipim-ipim	0.000	0.000		
ipim-tipim	0.182	0.471	16936	471018
ipim-pipim	0.091	0.113	8522	113473

/ipiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ipiŋ-tipiŋ	0.556	0.471	42928	471018
ipiŋ-ipiŋ	0.000	0.000		
ipiŋ-ʧipiŋ	0.444	0.416	34149	415509

ipiŋ-pipiŋ	0.000	0.113		113473

/ipin/	Input Fr.	Gen Fr.	Input #	Gen. #
ipin-ʧipin	0.545	0.416	51566	415509
ipin-ipin	0.000	0.000		
ipin-tipin	0.455	0.471	44587	471018
ipin-pipin	0.000	0.113		113473

/ipit/	Input Fr.	Gen Fr.	Input #	Gen. #
ipit-tipit	0.625	0.370	43192	369666
ipit-ipit	0.000	0.000		
ipit-ʧipit	0.250	0.476	17315	476165
ipit-pipit	0.125	0.154	8497	154169

/isik/	Input Fr.	Gen Fr.	Input #	Gen. #
isik-ʧisik	0.636	0.415	60844	414775
isik-isik	0.000	0.000		
isik-tisik	0.182	0.467	17601	466502
isik-pisik	0.182	0.119	16816	118723

/isim/	Input Fr.	Gen Fr.	Input #	Gen. #
isim-ʧisim	0.583	0.415	60778	414775
isim-isim	0.000	0.000		
isim-tisim	0.250	0.467	25449	466502
isim-pisim	0.167	0.119	17250	118723

/isiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
isiŋ-ʧisiŋ	0.714	0.415	43343	414775
isiŋ-isiŋ	0.000	0.000		
isiŋ-tisiŋ	0.143	0.467	9011	466502
isiŋ-pisiŋ	0.143	0.119	9027	118723

/isin/	Input Fr.	Gen Fr.	Input #	Gen. #
isin-ʧisin	0.700	0.415	60625	414775
isin-isin	0.000	0.000		
isin-tisin	0.100	0.467	8567	466502
isin-pisin	0.200	0.119	17715	118723

/isip/	Input Fr.	Gen Fr.	Input #	Gen. #
isip-ʧisip	0.417	0.416	43698	415509
isip-isip	0.000	0.000		
isip-tisip	0.167	0.471	16948	471018
isip-pisip	0.417	0.113	43261	113473

/isit/	Input Fr.	Gen Fr.	Input #	Gen. #
isit-tisit	0.400	0.365	34683	365455
isit-isit	0.000	0.000		
isit-ʧisit	0.300	0.474	25875	474438
isit-pisit	0.300	0.160	26094	160107

/itik/	Input Fr.	Gen Fr.	Input #	Gen. #
itik-ʧitik	0.600	0.474	51750	474438
itik-itik	0.000	0.000		
itik-titik	0.200	0.365	18092	365455
itik-pitik	0.200	0.160	17401	160107

/itim/	Input Fr.	Gen Fr.	Input #	Gen. #
itim-ʧitim	0.500	0.474	51354	474438
itim-itim	0.000	0.000		
itim-titim	0.250	0.365	26702	365455
itim-pitim	0.250	0.160	26242	160107

/itin/	Input Fr.	Gen Fr.	Input #	Gen. #
itin-ʧitin	0.636	0.474	60251	474438
itin-itin	0.000	0.000		
itin-titin	0.091	0.365	8683	365455
itin-pitin	0.273	0.160	26296	160107

/itiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
itiŋ-ʧitiŋ	0.583	0.474	61058	474438
itiŋ-itiŋ	0.000	0.000		
itiŋ-titiŋ	0.250	0.365	25671	365455
itiŋ-pitiŋ	0.167	0.160	17032	160107

/itip/	Input Fr.	Gen Fr.	Input #	Gen. #
itip-ʧitip	0.556	0.476	42702	476165

itip-itip	0.000	0.000		
itip-titip	0.222	0.370	17353	369666
itip-pitip	0.222	0.154	17134	154169

/itit/	Input Fr.	Gen Fr.	Input #	Gen. #
itit-ʧitit	0.667	0.474	50739	474438
itit-itit	0.000	0.000		
itit-titit	0.111	0.365	8157	365455
itit-pitit	0.222	0.160	17066	160107

/ukum/	Input Fr.	Gen Fr.	Input #	Gen. #
ukum-tukum	0.778	0.467	61309	466502
ukum-ukum	0.000	0.000		
ukum-ʧukum	0.222	0.415	16855	414775
ukum-pukum	0.000	0.119		118723

/ukuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ukuŋ-tukuŋ	0.538	0.467	61256	466502
ukuŋ-ukuŋ	0.000	0.000		
ukuŋ-ʧukuŋ	0.385	0.415	43171	414775
ukuŋ-pukuŋ	0.077	0.119	9047	118723

/ukun/	Input Fr.	Gen Fr.	Input #	Gen. #
ukun-tukun	0.462	0.467	52289	466502
ukun-ukun	0.000	0.000		
ukun-ʧukun	0.385	0.415	43462	414775
ukun-pukun	0.154	0.119	16892	118723

/ukup/	Input Fr.	Gen Fr.	Input #	Gen. #
ukup-tukup	0.583	0.471	60124	471018
ukup-ukup	0.000	0.000		
ukup-ʧukup	0.250	0.416	25565	415509
ukup-pukup	0.167	0.113	17275	113473

/ukut/	Input Fr.	Gen Fr.	Input #	Gen. #
ukut-ʧukut	0.455	0.474	43370	474438
ukut-ukut	0.000	0.000		
ukut-tukut	0.273	0.365	24938	365455

ukut-pukut	0.273	0.160	25929	160107

/umuk/	Input Fr.	Gen Fr.	Input #	Gen. #
umuk-tumuk	0.500	0.467	43233	466502
umuk-umuk	0.000	0.000		
umuk-ʧumuk	0.500	0.415	43200	414775
umuk-pumuk	0.000	0.119		118723

/umun/	Input Fr.	Gen Fr.	Input #	Gen. #
umun-tumun	0.727	0.467	69863	466502
umun-umun	0.000	0.000		
umun-ʧumun	0.273	0.415	26494	414775
umun-pumun	0.000	0.119		118723

/umuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
umuŋ-ʧumuŋ	0.556	0.415	42693	414775
umuŋ-umuŋ	0.000	0.000		
umuŋ-tumuŋ	0.444	0.467	34204	466502
umuŋ-pumuŋ	0.000	0.119		118723

/umup/	Input Fr.	Gen Fr.	Input #	Gen. #
umup-tumup	0.462	0.471	52306	471018
umup-umup	0.000	0.000		
umup-ʧumup	0.308	0.416	36024	415509
umup-pumup	0.231	0.113	25608	113473

/umut/	Input Fr.	Gen Fr.	Input #	Gen. #
umut-ʧumut	0.615	0.474	69600	474438
umut-umut	0.000	0.000		
umut-tumut	0.154	0.365	17483	365455
umut-pumut	0.231	0.160	26282	160107

/unuk/	Input Fr.	Gen Fr.	Input #	Gen. #
unuk-tunuk	0.500	0.467	50756	466502
unuk-unuk	0.000	0.000		
unuk-ʧunuk	0.417	0.415	43806	414775
unuk-punuk	0.083	0.119	8744	118723

/uŋuk/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋuk-ʧuŋuk	0.556	0.415	44186	414775
uŋuk-uŋuk	0.000	0.000		
uŋuk-tuŋuk	0.444	0.467	34408	466502
uŋuk-puŋuk	0.000	0.119		118723

/uŋum/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋum-tuŋum	0.500	0.467	52261	466502
uŋum-uŋum	0.000	0.000		
uŋum-ʧuŋum	0.500	0.415	51417	414775
uŋum-puŋum	0.000	0.119		118723

/unum/	Input Fr.	Gen Fr.	Input #	Gen. #
unum-tunum	0.545	0.467	51773	466502
unum-unum	0.000	0.000		
unum-ʧunum	0.364	0.415	34721	414775
unum-punum	0.091	0.119	8554	118723

/uŋun/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋun-tuŋun	0.778	0.467	60132	466502
uŋun-uŋun	0.000	0.000		
uŋun-ʧuŋun	0.111	0.415	8626	414775
uŋun-puŋun	0.111	0.119	8056	118723

/unuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
unuŋ-ʧunuŋ	0.600	0.415	51582	414775
unuŋ-unuŋ	0.000	0.000		
unuŋ-tunuŋ	0.300	0.467	26067	466502
unuŋ-punuŋ	0.100	0.119	8663	118723

/unup/	Input Fr.	Gen Fr.	Input #	Gen. #
unup-tunup	0.500	0.471	52784	471018
unup-unup	0.000	0.000		
unup-ʧunup	0.417	0.416	43404	415509
unup-punup	0.083	0.113	8404	113473

/uŋup/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋup-tuŋup	0.500	0.471	43018	471018

uŋup-uŋup	0.000	0.000		
uŋup-ʧuŋup	0.200	0.416	17279	415509
uŋup-puŋup	0.300	0.113	26689	113473

/uŋut/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋut-ʧuŋut	0.600	0.474	51934	474438
uŋut-uŋut	0.000	0.000		
uŋut-tuŋut	0.300	0.365	26160	365455
uŋut-puŋut	0.100	0.160	8371	160107

/unut/	Input Fr.	Gen Fr.	Input #	Gen. #
unut-ʧunut	0.818	0.474	77773	474438
unut-unut	0.000	0.000		
unut-tunut	0.091	0.365	8611	365455
unut-punut	0.091	0.160	8955	160107

/upuk/	Input Fr.	Gen Fr.	Input #	Gen. #
upuk-tupuk	0.625	0.471	43266	471018
upuk-upuk	0.000	0.000		
upuk-ʧupuk	0.375	0.416	25455	415509
upuk-pupuk	0.000	0.113		113473

/upum/	Input Fr.	Gen Fr.	Input #	Gen. #
upum-tupum	0.600	0.471	52056	471018
upum-upum	0.000	0.000		
upum-ʧupum	0.200	0.416	17749	415509
upum-pupum	0.200	0.113	17905	113473

/upun/	Input Fr.	Gen Fr.	Input #	Gen. #
upun-tupun	0.636	0.471	60461	471018
upun-upun	0.000	0.000		
upun-ʧupun	0.273	0.416	26396	415509
upun-pupun	0.091	0.113	8624	113473

/upuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
upuŋ-tupuŋ	0.636	0.467	60578	466502
upuŋ-upuŋ	0.000	0.000		
upuŋ-ʧupuŋ	0.273	0.415	26509	414775

upuŋ-pupuŋ	0.091	0.119	8522	118723

/uput/	Input Fr.	Gen Fr.	Input #	Gen. #
uput-tuput	0.400	0.370	33430	369666
uput-uput	0.000	0.000		
uput-ʧuput	0.400	0.476	34999	476165
uput-puput	0.200	0.154	17074	154169

/usuk/	Input Fr.	Gen Fr.	Input #	Gen. #
usuk-tusuk	0.556	0.467	43931	466502
usuk-usuk	0.000	0.000		
usuk-ʧusuk	0.333	0.415	25699	414775
usuk-pusuk	0.111	0.119	8620	118723

/usum/	Input Fr.	Gen Fr.	Input #	Gen. #
usum-tusum	0.583	0.467	60222	466502
usum-usum	0.000	0.000		
usum-ʧusum	0.417	0.415	42854	414775
usum-pusum	0.000	0.119		118723

/usun/	Input Fr.	Gen Fr.	Input #	Gen. #
usun-tusun	0.545	0.467	51008	466502
usun-usun	0.000	0.000		
usun-ʧusun	0.364	0.415	34174	414775
usun-pusun	0.091	0.119	9025	118723

/usuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
usuŋ-tusuŋ	0.667	0.467	51514	466502
usuŋ-usuŋ	0.000	0.000		
usuŋ-ʧusuŋ	0.222	0.415	17014	414775
usuŋ-pusuŋ	0.111	0.119	8713	118723

/usup/	Input Fr.	Gen Fr.	Input #	Gen. #
usup-tusup	0.636	0.471	60436	471018
usup-usup	0.000	0.000		
usup-ʧusup	0.182	0.416	17192	415509
usup-pusup	0.182	0.113	17456	113473

/usut/	Input Fr.	Gen Fr.	Input #	Gen. #
usut-tusut	0.600	0.365	51223	365455
usut-usut	0.000	0.000		
usut-ʧusut	0.100	0.474	8761	474438
usut-pusut	0.300	0.160	26389	160107

/utuk/	Input Fr.	Gen Fr.	Input #	Gen. #
utuk-tutuk	0.500	0.365	52981	365455
utuk-utuk	0.000	0.000		
utuk-ʧutuk	0.333	0.474	34676	474438
utuk-putuk	0.167	0.160	17524	160107

/utum/	Input Fr.	Gen Fr.	Input #	Gen. #
utum-ʧutum	0.545	0.474	53049	474438
utum-utum	0.000	0.000		
utum-tutum	0.364	0.365	33672	365455
utum-putum	0.091	0.160	8697	160107

/utun/	Input Fr.	Gen Fr.	Input #	Gen. #
utun-ʧutun	0.636	0.474	60597	474438
utun-utun	0.000	0.000		
utun-tutun	0.273	0.365	26573	365455
utun-putun	0.091	0.160	8618	160107

/utuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
utuŋ-tutuŋ	0.385	0.365	43597	365455
utuŋ-utuŋ	0.000	0.000		
utuŋ-ʧutuŋ	0.385	0.474	42891	474438
utuŋ-putuŋ	0.231	0.160	26200	160107

/utup/	Input Fr.	Gen Fr.	Input #	Gen. #
utup-tutup	0.500	0.370	25319	369666
utup-utup	0.000	0.000		
utup-ʧutup	0.333	0.476	17066	476165
utup-putup	0.167	0.154	8528	154169

/utut/	Input Fr.	Gen Fr.	Input #	Gen. #
utut-ʧutut	0.444	0.474	34337	474438

utut-utut	0.000	0.000		
utut-tutut	0.222	0.365	17724	365455
utut-putut	0.333	0.160	27023	160107

3. Tableaux

The following are approximate tableaux for this ranking. Outputs are derived simply by sorting the constraints by their ranking value, with no stochastic variation.

To diagnose variation, consult two things:

--The candidate frequencies (which are the generated frequencies, not the input frequencies).
--The probability that each constraint outranks the next one down, given directly after the constraint labels.

/asap/:

	Onset (.9985)	*Repeat(?) (.791)	*Lab (.658)	*AffricV (.655)	*StopV (.514)	? (.501)	*Repeat(t) (.646)	*Cor (.501)	p (.503)	t (.636)	*Repeat(p)
asap-tasap (0.467)	1				1	1		1	1		
asap-?asap (0.415)	1			1!				1	1	1	
asap-pasap (0.119)	1		1!		1	1				1	
asap-asap	2!										

/ikit/:

	Onset (.9985)	*Repeat(?) (.791)	*Lab (.658)	*AffricV (.655)	*StopV (.514)	? (.501)	*Repeat(t) (.646)	*Cor (.501)	p (.503)	t (.636)	*Repeat(p)
ikit-tikit (0.365)	1				1	1	1	1	1		
ikit-?ikit (0.474)	1			1!				1	1	1	
ikit-pikit (0.160)	1		1!		1	1				1	
ikit-ikit	2!										

4. Active Constraints

A constraint is active if it causes the winning candidate to defeat a rival in at least one competition.

Active	Onset
Inactive	*Repeat(ʧ)
Active	*Lab
Active	*AffricV

```
   Active     ʧ
   Active     *StopV
   Active     *Repeat(t)
   Active     *Cor
   Active     p
   Active     t
   Active     *Repeat(p)

5. Testing the Grammar:  Details

   The grammar was tested for 1000000 cycles.
   Average error per candidate:  0.670 percent
   Learning time:  3.885 minutes

6. Parameter Values Used by the GLA

   Initial Rankings

      All constraints started out at the default value of 100.

   Schedule for GLA Parameters

      Stage   Trials   PlastMark  PlastFaith NoiseMark  NoiseFaith
1          12500000     25000002.000       2.0002.000       2.0002.000       2.0002.000
2.000
2          22500000     25000000.159       0.1590.159       0.1592.000       2.0002.000
2.000
3          32500000     25000000.013       0.0130.013       0.0132.000       2.0002.000
2.000
4          42500000     25000000.001       0.0010.001       0.0012.000       2.0002.000
2.000

      There were a total of 10000000 learning trials.
```

Learning of individual data:
An individual speaker, who has preference for /t/ (Experiment 2; S15) and is *not* sensitive to context

Result of Applying Gradual Learning Algorithm to otsoft_expt2_ch3 (2).xls
OTSoft 2.3, release date 5/15/08

1. Ranking Values Found

110.000	Onset
100.000	*Lab
100.000	*Repeat(ʧ)
97.356	*AffricV
96.881	*NasV
95.763	*StopV
95.540	*Repeat(t)
94.237	t
92.644	ʧ
90.000	*Cor
90.000	p

2. Matchup to Input Frequencies

/akam/	Input Fr.	Gen Fr.	Input #	Gen. #
akam-ʧakam	1.000	0.187	90828	186606
akam-akam	0.000	0.000		
akam-takam	0.000	0.584		584473
akam-nakam	0.000	0.229		228921

/akan/	Input Fr.	Gen Fr.	Input #	Gen. #
akan-ʧakan	1.000	0.187	91847	186606
akan-akan	0.000	0.000		
akan-takan	0.000	0.584		584473
akan-nakan	0.000	0.229		228921

/akaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
akaŋ-nakaŋ	1.000	0.229	89920	228921
akaŋ-akaŋ	0.000	0.000		
akaŋ-takaŋ	0.000	0.584		584473
akaŋ-ʧakaŋ	0.000	0.187		186606

/akap/	Input Fr.	Gen Fr.	Input #	Gen. #

akap-takap	1.000	0.584	89123	584473
akap-akap	0.000	0.000		
akap-ʧakap	0.000	0.187		186606
akap-nakap	0.000	0.229		228921

/akat/	Input Fr.	Gen Fr.	Input #	Gen. #
akat-takat	1.000	0.423	90208	423155
akat-akat	0.000	0.000		
akat-ʧakat	0.000	0.254		253888
akat-nakat	0.000	0.323		322957

/amak/	Input Fr.	Gen Fr.	Input #	Gen. #
amak-ʧamak	1.000	0.187	89981	186606
amak-amak	0.000	0.000		
amak-tamak	0.000	0.584		584473
amak-namak	0.000	0.229		228921

/aman/	Input Fr.	Gen Fr.	Input #	Gen. #
aman-ʧaman	1.000	0.187	91327	186606
aman-aman	0.000	0.000		
aman-taman	0.000	0.584		584473
aman-naman	0.000	0.229		228921

/amaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
amaŋ-tamaŋ	1.000	0.584	90782	584473
amaŋ-amaŋ	0.000	0.000		
amaŋ-ʧamaŋ	0.000	0.187		186606
amaŋ-namaŋ	0.000	0.229		228921

/amap/	Input Fr.	Gen Fr.	Input #	Gen. #
amap-tamap	1.000	0.584	91283	584473
amap-amap	0.000	0.000		
amap-ʧamap	0.000	0.187		186606
amap-namap	0.000	0.229		228921

/amat/	Input Fr.	Gen Fr.	Input #	Gen. #
amat-tamat	1.000	0.423	90200	423155
amat-amat	0.000	0.000		

amat-ʧamat	0.000	0.254		253888
amat-namat	0.000	0.323		322957

/anak/	Input Fr.	Gen Fr.	Input #	Gen. #
anak-tanak	1.000	0.584	90277	584473
anak-anak	0.000	0.000		
anak-ʧanak	0.000	0.187		186606
anak-nanak	0.000	0.229		228921

/aŋak/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋak-taŋak	1.000	0.584	91307	584473
aŋak-aŋak	0.000	0.000		
aŋak-ʧaŋak	0.000	0.187		186606
aŋak-naŋak	0.000	0.229		228921

/aŋam/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋam-taŋam	1.000	0.584	88358	584473
aŋam-aŋam	0.000	0.000		
aŋam-ʧaŋam	0.000	0.187		186606
aŋam-naŋam	0.000	0.229		228921

/anam/	Input Fr.	Gen Fr.	Input #	Gen. #
anam-tanam	1.000	0.584	90429	584473
anam-anam	0.000	0.000		
anam-ʧanam	0.000	0.187		186606
anam-nanam	0.000	0.229		228921

/aŋan/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋan-taŋan	1.000	0.584	90328	584473
aŋan-aŋan	0.000	0.000		
aŋan-ʧaŋan	0.000	0.187		186606
aŋan-naŋan	0.000	0.229		228921

/anaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
anaŋ-ʧanaŋ	1.000	0.187	88081	186606
anaŋ-anaŋ	0.000	0.000		
anaŋ-tanaŋ	0.000	0.584		584473
anaŋ-nanaŋ	0.000	0.229		228921

/anap/	Input Fr.	Gen Fr.	Input #	Gen. #
anap-tanap	1.000	0.584	89961	584473
anap-anap	0.000	0.000		
anap-ʧanap	0.000	0.187		186606
anap-nanap	0.000	0.229		228921

/aŋap/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋap-taŋap	1.000	0.584	90786	584473
aŋap-aŋap	0.000	0.000		
aŋap-ʧaŋap	0.000	0.187		186606
aŋap-naŋap	0.000	0.229		228921

/aŋat/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋat-taŋat	1.000	0.423	91072	423155
aŋat-aŋat	0.000	0.000		
aŋat-ʧaŋat	0.000	0.254		253888
aŋat-naŋat	0.000	0.323		322957

/anat/	Input Fr.	Gen Fr.	Input #	Gen. #
anat-tanat	1.000	0.423	90687	423155
anat-anat	0.000	0.000		
anat-ʧanat	0.000	0.254		253888
anat-nanat	0.000	0.323		322957

/apak/	Input Fr.	Gen Fr.	Input #	Gen. #
apak-tapak	1.000	0.584	89218	584473
apak-apak	0.000	0.000		
apak-ʧapak	0.000	0.187		186606
apak-napak	0.000	0.229		228921

/apam/	Input Fr.	Gen Fr.	Input #	Gen. #
apam-tapam	1.000	0.584	89898	584473
apam-apam	0.000	0.000		
apam-ʧapam	0.000	0.187		186606
apam-napam	0.000	0.229		228921

/apan/	Input Fr.	Gen Fr.	Input #	Gen. #

apan-ʧapan	1.000	0.187	89950	186606
apan-apan	0.000	0.000		
apan-tapan	0.000	0.584		584473
apan-napan	0.000	0.229		228921

/apaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
apaŋ-napaŋ	1.000	0.229	89324	228921
apaŋ-apaŋ	0.000	0.000		
apaŋ-tapaŋ	0.000	0.584		584473
apaŋ-ʧapaŋ	0.000	0.187		186606

/apat/	Input Fr.	Gen Fr.	Input #	Gen. #
apat-napat	1.000	0.323	91098	322957
apat-apat	0.000	0.000		
apat-tapat	0.000	0.423		423155
apat-ʧapat	0.000	0.254		253888

/asak/	Input Fr.	Gen Fr.	Input #	Gen. #
asak-tasak	1.000	0.584	89773	584473
asak-asak	0.000	0.000		
asak-ʧasak	0.000	0.187		186606
asak-nasak	0.000	0.229		228921

/asam/	Input Fr.	Gen Fr.	Input #	Gen. #
asam-nasam	1.000	0.229	90908	228921
asam-sasam	0.000	0.000		
asam-tasam	0.000	0.584		584473
asam-ʧasam	0.000	0.187		186606

/asan/	Input Fr.	Gen Fr.	Input #	Gen. #
asan-tasan	1.000	0.584	89823	584473
asan-asan	0.000	0.000		
asan-ʧasan	0.000	0.187		186606
asan-nasan	0.000	0.229		228921

/asaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
asaŋ-tasaŋ	1.000	0.584	89219	584473
asaŋ-asaŋ	0.000	0.000		

asaŋ-ʧasaŋ	0.000	0.187		186606
asaŋ-nasaŋ	0.000	0.229		228921

/asap/	Input Fr.	Gen Fr.	Input #	Gen. #
asap-tasap	1.000	0.584	91299	584473
asap-asap	0.000	0.000		
asap-ʧasap	0.000	0.187		186606
asap-nasap	0.000	0.229		228921

/asat/	Input Fr.	Gen Fr.	Input #	Gen. #
asat-nasat	1.000	0.323	90158	322957
asat-asat	0.000	0.000		
asat-tasat	0.000	0.423		423155
asat-ʧasat	0.000	0.254		253888

/atak/	Input Fr.	Gen Fr.	Input #	Gen. #
atak-natak	1.000	0.323	90031	322957
atak-atak	0.000	0.000		
atak-tatak	0.000	0.423		423155
atak-ʧatak	0.000	0.254		253888

/atam/	Input Fr.	Gen Fr.	Input #	Gen. #
atam-natam	1.000	0.229	90109	228921
atam-atam	0.000	0.000		
atam-tatam	0.000	0.584		584473
atam-ʧatam	0.000	0.187		186606

/ataŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ataŋ-nataŋ	1.000	0.323	92301	322957
ataŋ-ataŋ	0.000	0.000		
ataŋ-tataŋ	0.000	0.423		423155
ataŋ-ʧataŋ	0.000	0.254		253888

/atan/	Input Fr.	Gen Fr.	Input #	Gen. #
atan-natan	1.000	0.323	89785	322957
atan-atan	0.000	0.000		
atan-tatan	0.000	0.423		423155
atan-ʧatan	0.000	0.254		253888

/atap/	Input Fr.	Gen Fr.	Input #	Gen. #
atap-natap	1.000	0.323	90270	322957
atap-atap	0.000	0.000		
atap-tatap	0.000	0.423		423155
atap-ʧatap	0.000	0.254		253888

/atat/	Input Fr.	Gen Fr.	Input #	Gen. #
atat-natat	1.000	0.323	90189	322957
atat-atat	0.000	0.000		
atat-tatat	0.000	0.423		423155
atat-ʧatat	0.000	0.254		253888

/ikim/	Input Fr.	Gen Fr.	Input #	Gen. #
ikim-nikim	1.000	0.229	89346	228921
ikim-ikim	0.000	0.000		
ikim-tikim	0.000	0.584		584473
ikim-ʧikim	0.000	0.187		186606

/ikin/	Input Fr.	Gen Fr.	Input #	Gen. #
ikin-nikin	1.000	0.229	90476	228921
ikin-ikin	0.000	0.000		
ikin-tikin	0.000	0.584		584473
ikin-ʧikin	0.000	0.187		186606

/ikiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ikiŋ-ʧikiŋ	1.000	0.187	90750	186606
ikiŋ-ikiŋ	0.000	0.000		
ikiŋ-tikiŋ	0.000	0.584		584473
ikiŋ-nikiŋ	0.000	0.229		228921

/ikip/	Input Fr.	Gen Fr.	Input #	Gen. #
ikip-nikip	1.000	0.229	88431	228921
ikip-ikip	0.000	0.000		
ikip-tikip	0.000	0.584		584473
ikip-ʔikip	0.000	0.187		186606

/ikit/	Input Fr.	Gen Fr.	Input #	Gen. #

ikit-nikit	1.000	0.323	90575	322957
ikit-ikit	0.000	0.000		
ikit-tikit	0.000	0.423		423155
ikit-ʧikit	0.000	0.254		253888

/imik/	Input Fr.	Gen Fr.	Input #	Gen. #
imik-ʧimik	1.000	0.187	90083	186606
imik-imik	0.000	0.000		
imik-timik	0.000	0.584		584473
imik-nimik	0.000	0.229		228921

/imiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
imiŋ-ʧimiŋ	1.000	0.187	89317	186606
imiŋ-imiŋ	0.000	0.000		
imiŋ-timiŋ	0.000	0.584		584473
imiŋ-nimiŋ	0.000	0.229		228921

/imin/	Input Fr.	Gen Fr.	Input #	Gen. #
imin-ʧimin	1.000	0.187	88865	186606
imin-imin	0.000	0.000		
imin-timin	0.000	0.584		584473
imin-nimin	0.000	0.229		228921

/imip/	Input Fr.	Gen Fr.	Input #	Gen. #
imip-ʧimip	1.000	0.187	89149	186606
imip-imip	0.000	0.000		
imip-timip	0.000	0.584		584473
imip-nimip	0.000	0.229		228921

/imit/	Input Fr.	Gen Fr.	Input #	Gen. #
imit-timit	1.000	0.423	88866	423155
imit-imit	0.000	0.000		
imit-ʧimit	0.000	0.254		253888
imit-nimit	0.000	0.323		322957

/iŋik/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋik-ʧiŋik	1.000	0.187	90880	186606
iŋik-iŋik	0.000	0.000		

iŋik-tiŋik	0.000	0.584		584473
iŋik-niŋik	0.000	0.229		228921

/inik/	Input Fr.	Gen Fr.	Input #	Gen. #
inik-ʧinik	1.000	0.187	89204	186606
inik-inik	0.000	0.000		
inik-tinik	0.000	0.584		584473
inik-ninik	0.000	0.229		228921

/iŋim/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋim-niŋim	1.000	0.229	89256	228921
iŋim-iŋim	0.000	0.000		
iŋim-tiŋim	0.000	0.584		584473
iŋim-ʧiŋim	0.000	0.187		186606

/inim/	Input Fr.	Gen Fr.	Input #	Gen. #
inim-tinim	1.000	0.584	91245	584473
inim-inim	0.000	0.000		
inim-ʧinim	0.000	0.187		186606
inim-ninim	0.000	0.229		228921

/iŋin/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋin-ʧiŋin	1.000	0.187	90265	186606
iŋin-iŋin	0.000	0.000		
iŋin-tiŋin	0.000	0.584		584473
iŋin-niŋin	0.000	0.229		228921

/iniŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
iniŋ-tiniŋ	1.000	0.584	89810	584473
iniŋ-iniŋ	0.000	0.000		
iniŋ-ʧiniŋ	0.000	0.187		186606
iniŋ-niniŋ	0.000	0.229		228921

/inip/	Input Fr.	Gen Fr.	Input #	Gen. #
inip-ʧinip	1.000	0.187	90511	186606
inip-inip	0.000	0.000		
inip-tinip	0.000	0.584		584473
inip-ninip	0.000	0.229		228921

/iŋip/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋip-tiŋip	1.000	0.584	90545	584473
iŋip-iŋip	0.000	0.000		
iŋip-ʧiŋip	0.000	0.187		186606
iŋip-niŋip	0.000	0.229		228921

/iɲit/	Input Fr.	Gen Fr.	Input #	Gen. #
iɲit-tiɲit	1.000	0.423	91108	423155
iɲit-iɲit	0.000	0.000		
iɲit-ʧiɲit	0.000	0.254		253888
iɲit-niɲit	0.000	0.323		322957

/init/	Input Fr.	Gen Fr.	Input #	Gen. #
init-tinit	1.000	0.423	92331	423155
init-init	0.000	0.000		
init-ʧinit	0.000	0.254		253888
init-ninit	0.000	0.323		322957

/ipik/	Input Fr.	Gen Fr.	Input #	Gen. #
ipik-ʧipik	1.000	0.187	88956	186606
ipik-ipik	0.000	0.000		
ipik-tipik	0.000	0.584		584473
ipik-nipik	0.000	0.229		228921

/ipim/	Input Fr.	Gen Fr.	Input #	Gen. #
ipim-nipim	1.000	0.229	89688	228921
ipim-ipim	0.000	0.000		
ipim-tipim	0.000	0.584		584473
ipim-ʧipim	0.000	0.187		186606

/ipiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ipiŋ-ʧipiŋ	1.000	0.187	89654	186606
ipiŋ-ipiŋ	0.000	0.000		
ipiŋ-tipiŋ	0.000	0.584		584473
ipiŋ-nipiŋ	0.000	0.229		228921

/ipin/	Input Fr.	Gen Fr.	Input #	Gen. #

ipin-nipin	1.000	0.229	90608	228921
ipin-ipin	0.000	0.000		
ipin-tipin	0.000	0.584		584473
ipin-ʧipin	0.000	0.187		186606

/ipit/	Input Fr.	Gen Fr.	Input #	Gen. #
ipit-tipit	1.000	0.423	89872	423155
ipit-ipit	0.000	0.000		
ipit-ʧipit	0.000	0.254		253888
ipit-nipit	0.000	0.323		322957

/isik/	Input Fr.	Gen Fr.	Input #	Gen. #
isik-tisik	1.000	0.584	89930	584473
isik-isik	0.000	0.000		
isik-ʧisik	0.000	0.187		186606
isik-nisik	0.000	0.229		228921

/isim/	Input Fr.	Gen Fr.	Input #	Gen. #
isim-ʧisim	1.000	0.187	90383	186606
isim-isim	0.000	0.000		
isim-tisim	0.000	0.584		584473
isim-nisim	0.000	0.229		228921

/isiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
isiŋ-nisiŋ	1.000	0.229	90406	228921
isiŋ-isiŋ	0.000	0.000		
isiŋ-tisiŋ	0.000	0.584		584473
isiŋ-ʧisiŋ	0.000	0.187		186606

/isin/	Input Fr.	Gen Fr.	Input #	Gen. #
isin-ʧisin	1.000	0.187	89893	186606
isin-isin	0.000	0.000		
isin-tisin	0.000	0.584		584473
isin-nisin	0.000	0.229		228921

/isip/	Input Fr.	Gen Fr.	Input #	Gen. #
isip-tisip	1.000	0.584	90127	584473
isip-isip	0.000	0.000		

isip-ʧisip	0.000	0.187		186606
isip-nisip	0.000	0.229		228921

/isit/	Input Fr.	Gen Fr.	Input #	Gen. #
isit-tisit	1.000	0.423	90761	423155
isit-isit	0.000	0.000		
isit-ʧisit	0.000	0.254		253888
isit-nisit	0.000	0.323		322957

/itik/	Input Fr.	Gen Fr.	Input #	Gen. #
itik-ʧitik	1.000	0.254	90555	253888
itik-itik	0.000	0.000		
itik-titik	0.000	0.423		423155
itik-nitik	0.000	0.323		322957

/itim/	Input Fr.	Gen Fr.	Input #	Gen. #
itim-nitim	1.000	0.323	89780	322957
itim-itim	0.000	0.000		
itim-titim	0.000	0.423		423155
itim-ʧitim	0.000	0.254		253888

/itin/	Input Fr.	Gen Fr.	Input #	Gen. #
itin-nitin	1.000	0.323	89971	322957
itin-itin	0.000	0.000		
itin-titin	0.000	0.423		423155
itin-ʧitin	0.000	0.254		253888

/itiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
itiŋ-nitiŋ	1.000	0.323	89532	322957
itiŋ-itiŋ	0.000	0.000		
itiŋ-titiŋ	0.000	0.423		423155
itiŋ-ʧitiŋ	0.000	0.254		253888

/itip/	Input Fr.	Gen Fr.	Input #	Gen. #
itip-nitip	1.000	0.323	87920	322957
itip-itip	0.000	0.000		
itip-titip	0.000	0.423		423155
itip-ʧitip	0.000	0.254		253888

/itit/	Input Fr.	Gen Fr.	Input #	Gen. #
itit-nitit	1.000	0.323	89933	322957
itit-itit	0.000	0.000		
itit-titit	0.000	0.423		423155
itit-ʧitit	0.000	0.254		253888

/ukum/	Input Fr.	Gen Fr.	Input #	Gen. #
ukum-tukum	1.000	0.584	90798	584473
ukum-ukum	0.000	0.000		
ukum-ʧukum	0.000	0.187		186606
ukum-nukum	0.000	0.229		228921

/ukuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ukuŋ-tukuŋ	1.000	0.584	90662	584473
ukuŋ-ukuŋ	0.000	0.000		
ukuŋ-ʧukuŋ	0.000	0.187		186606
ukuŋ-nukuŋ	0.000	0.229		228921

/ukun/	Input Fr.	Gen Fr.	Input #	Gen. #
ukun-tukun	1.000	0.584	88857	584473
ukun-ukun	0.000	0.000		
ukun-ʧukun	0.000	0.187		186606
ukun-nukun	0.000	0.229		228921

/ukup/	Input Fr.	Gen Fr.	Input #	Gen. #
ukup-tukup	1.000	0.584	89973	584473
ukup-ukup	0.000	0.000		
ukup-ʧukup	0.000	0.187		186606
ukup-nukup	0.000	0.229		228921

/ukut/	Input Fr.	Gen Fr.	Input #	Gen. #
ukut-tukut	1.000	0.423	88933	423155
ukut-ukut	0.000	0.000		
ukut-ʧukut	0.000	0.254		253888
ukut-nukut	0.000	0.323		322957

/umuk/	Input Fr.	Gen Fr.	Input #	Gen. #
umuk-tumuk	1.000	0.584	89668	584473

umuk-umuk	0.000	0.000		
umuk-ʧumuk	0.000	0.187		186606
umuk-numuk	0.000	0.229		228921

/umun/	Input Fr.	Gen Fr.	Input #	Gen. #
umun-tumun	1.000	0.584	90891	584473
umun-umun	0.000	0.000		
umun-ʧumun	0.000	0.187		186606
umun-numun	0.000	0.229		228921

/umuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
umuŋ-ʧumuŋ	1.000	0.187	90108	186606
umuŋ-umuŋ	0.000	0.000		
umuŋ-tumuŋ	0.000	0.584		584473
umuŋ-numuŋ	0.000	0.229		228921

/umup/	Input Fr.	Gen Fr.	Input #	Gen. #
umup-tumup	1.000	0.584	91004	584473
umup-umup	0.000	0.000		
umup-ʧumup	0.000	0.187		186606
umup-numup	0.000	0.229		228921

/umut/	Input Fr.	Gen Fr.	Input #	Gen. #
umut-tumut	1.000	0.423	90420	423155
umut-umut	0.000	0.000		
umut-ʧumut	0.000	0.254		253888
umut-numut	0.000	0.323		322957

/unuk/	Input Fr.	Gen Fr.	Input #	Gen. #
unuk-tunuk	1.000	0.584	89286	584473
unuk-unuk	0.000	0.000		
unuk-ʧunuk	0.000	0.187		186606
unuk-nunuk	0.000	0.229		228921

/uŋuk/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋuk-tuŋuk	1.000	0.584	90855	584473
uŋuk-uŋuk	0.000	0.000		
uŋuk-ʧuŋuk	0.000	0.187		186606

uŋuk-nuŋuk	0.000	0.229		228921

/uŋum/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋum-tuŋum	1.000	0.584	89574	584473
uŋum-uŋum	0.000	0.000		
uŋum-ʧuŋum	0.000	0.187		186606
uŋum-nuŋum	0.000	0.229		228921

/unum/	Input Fr.	Gen Fr.	Input #	Gen. #
unum-tunum	1.000	0.584	90128	584473
unum-unum	0.000	0.000		
unum-ʧunum	0.000	0.187		186606
unum-punum	0.000	0.229		228921

/uŋun/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋun-tuŋun	1.000	0.584	89327	584473
uŋun-uŋun	0.000	0.000		
uŋun-ʧuŋun	0.000	0.187		186606
uŋun-nuŋun	0.000	0.229		228921

/unuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
unuŋ-tunuŋ	1.000	0.584	89401	584473
unuŋ-unuŋ	0.000	0.000		
unuŋ-ʧunuŋ	0.000	0.187		186606
unuŋ-nunuŋ	0.000	0.229		228921

/unup/	Input Fr.	Gen Fr.	Input #	Gen. #
unup-tunup	1.000	0.584	90428	584473
unup-unup	0.000	0.000		
unup-ʧunup	0.000	0.187		186606
unup-nunup	0.000	0.229		228921

/uŋup/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋup-tuŋup	1.000	0.584	89885	584473
uŋup-uŋup	0.000	0.000		
uŋup-ʧuŋup	0.000	0.187		186606
uŋup-nuŋup	0.000	0.229		228921

/uŋut/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋut-tuŋut	1.000	0.423	90629	423155
uŋut-uŋut	0.000	0.000		
uŋut-ʧuŋut	0.000	0.254		253888
uŋut-nuŋut	0.000	0.323		322957

/unut/	Input Fr.	Gen Fr.	Input #	Gen. #
unut-tunut	1.000	0.423	89660	423155
unut-unut	0.000	0.000		
unut-ʧunut	0.000	0.254		253888
unut-nunut	0.000	0.323		322957

/upuk/	Input Fr.	Gen Fr.	Input #	Gen. #
upuk-tupuk	1.000	0.584	89606	584473
upuk-upuk	0.000	0.000		
upuk-ʧupuk	0.000	0.187		186606
upuk-nupuk	0.000	0.229		228921

/upum/	Input Fr.	Gen Fr.	Input #	Gen. #
upum-tupum	1.000	0.584	91245	584473
upum-upum	0.000	0.000		
upum-ʧupum	0.000	0.187		186606
upum-nupum	0.000	0.229		228921

/upun/	Input Fr.	Gen Fr.	Input #	Gen. #
upun-tupun	1.000	0.584	90349	584473
upun-upun	0.000	0.000		
upun-ʧupun	0.000	0.187		186606
upun-nupun	0.000	0.229		228921

/upuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
upuŋ-tupuŋ	1.000	0.584	90462	584473
upuŋ-upuŋ	0.000	0.000		
upuŋ-ʧupuŋ	0.000	0.187		186606
upuŋ-nupuŋ	0.000	0.229		228921

/uput/	Input Fr.	Gen Fr.	Input #	Gen. #
uput-ʧuput	1.000	0.254	89147	253888

uput-uput	0.000	0.000		
uput-tuput	0.000	0.423		423155
uput-nuput	0.000	0.323		322957

/usuk/	Input Fr.	Gen Fr.	Input #	Gen. #
usuk-tusuk	1.000	0.584	90114	584473
usuk-usuk	0.000	0.000		
usuk-ʧusuk	0.000	0.187		186606
usuk-nusuk	0.000	0.229		228921

/usum/	Input Fr.	Gen Fr.	Input #	Gen. #
usum-tusum	1.000	0.584	88903	584473
usum-usum	0.000	0.000		
usum-ʧusum	0.000	0.187		186606
usum-nusum	0.000	0.229		228921

/usun/	Input Fr.	Gen Fr.	Input #	Gen. #
usun-tusun	1.000	0.584	88992	584473
usun-usun	0.000	0.000		
usun-ʧusun	0.000	0.187		186606
usun-nusun	0.000	0.229		228921

/usuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
usuŋ-tusuŋ	1.000	0.584	89633	584473
usuŋ-usuŋ	0.000	0.000		
usuŋ-ʧusuŋ	0.000	0.187		186606
usuŋ-nusuŋ	0.000	0.229		228921

/usup/	Input Fr.	Gen Fr.	Input #	Gen. #
usup-tusup	1.000	0.584	89825	584473
usup-usup	0.000	0.000		
usup-ʧusup	0.000	0.187		186606
usup-nusup	0.000	0.229		228921

/usut/	Input Fr.	Gen Fr.	Input #	Gen. #
usut-tusut	1.000	0.423	89808	423155
usut-usut	0.000	0.000		
usut-ʧusut	0.000	0.254		253888

usut-nusut	0.000	0.323		322957

/utuk/	Input Fr.	Gen Fr.	Input #	Gen. #
utuk-nutuk	1.000	0.323	91069	322957
utuk-utuk	0.000	0.000		
utuk-tutuk	0.000	0.423		423155
utuk-ʧutuk	0.000	0.254		253888

/utum/	Input Fr.	Gen Fr.	Input #	Gen. #
utum-nutum	1.000	0.323	90821	322957
utum-utum	0.000	0.000		
utum-tutum	0.000	0.423		423155
utum-ʧutum	0.000	0.254		253888

/utun/	Input Fr.	Gen Fr.	Input #	Gen. #
utun-nutun	1.000	0.323	90125	322957
utun-utun	0.000	0.000		
utun-tutun	0.000	0.423		423155
utun-ʧutun	0.000	0.254		253888

/utuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
utuŋ-tutuŋ	1.000	0.423	90409	423155
utuŋ-utuŋ	0.000	0.000		
utuŋ-ʧutuŋ	0.000	0.254		253888
utuŋ-nutuŋ	0.000	0.323		322957

/utup/	Input Fr.	Gen Fr.	Input #	Gen. #
utup-nutup	1.000	0.323	88859	322957
utup-utup	0.000	0.000		
utup-tutup	0.000	0.423		423155
utup-ʧutup	0.000	0.254		253888

/utut/	Input Fr.	Gen Fr.	Input #	Gen. #
utut-nutut	1.000	0.323	91026	322957
utut-utut	0.000	0.000		
utut-tutut	0.000	0.423		423155
utut-ʧutut	0.000	0.254		253888

3. Tableaux

The following are approximate tableaux for this ranking. Outputs are derived simply by sorting the constraints by their ranking value, with no stochastic variation.

To diagnose variation, consult two things:

--The candidate frequencies (which are the generated frequencies, not the input frequencies).

--The probability that each constraint outranks the next one down, given directly after the constraint labels.

/atan/:

	Onset (.99985)	*Lab (.501)	*Repeat(?) (.826)	*AffricV (.567)	*NasV (.654)	*StopV (.532)	*Repeat(t) (.678)	t (.714)	? (.826)	*Cor (.501)	p
atan-tatan (0.423)	1					1	1		1	1	1
atan-natan (0.323)	1				1!			1	1	1	1
atan-?atan (0.254)	1			1!				1		1	1
atan-atan	2!										

/ikip/:

	Onset (.99985)	*Lab (.501)	*Repeat(?) (.826)	*AffricV (.567)	*NasV (.654)	*StopV (.532)	*Repeat(t) (.678)	t (.714)	? (.826)	*Cor (.501)	p
ikip-tikip (0.584)	1					1			1	1	1
ikip-nikip (0.229)	1				1!			1	1	1	1
ikip-?ikip (0.187)	1			1!				1		1	1
ikip-ikip	2!										

4. Active Constraints

A constraint is active if it causes the winning candidate to defeat a rival in at least one competition.

Active	Onset
Inactive	*Lab
Inactive	*Repeat(ʧ)
Active	*AffricV
Active	*NasV
Active	*StopV
Active	*Repeat(t)

```
   Active     t
   Active     ʧ
   Inactive   *Cor
   Inactive   p

5. Testing the Grammar:  Details

   The grammar was tested for 1000000 cycles.
   Average error per candidate:  14.713 percent
   Learning time:  2.719 minutes

6. Parameter Values Used by the GLA

   Initial Rankings

      All constraints started out at the default value of 100.

   Schedule for GLA Parameters
```

	Stage	Trials	PlastMark	PlastFaith	NoiseMark	NoiseFaith
1	12500000	25000002.000	2.0002.000	2.0002.000	2.0002.000	2.000
2	22500000	25000000.159	0.1590.159	0.1592.000	2.0002.000	2.000
3	32500000	25000000.013	0.0130.013	0.0132.000	2.0002.000	2.000
4	42500000	25000000.001	0.0010.001	0.0012.000	2.0002.000	2.000

```
      There were a total of 10000000 learning trials.
```

Learning of individual data:
An individual speaker, who has preference for /t/ (Experiment 2; S13) and *is* sensitive to context

```
Result of Applying Gradual Learning Algorithm to otsoft_expt2_ch3 (3).xls

6-17-2010, 10:25 p.m.

OTSoft 2.3, release date 5/15/08

1. Ranking Values Found
```

110.000	Onset
99.969	*Dor
98.153	*NasV
97.138	*Lab
96.817	*Repeat(t)
96.781	*FricV
95.066	*StopV
93.889	t
92.893	*Cor
91.014	p
90.000	ʧ
-36,312.730	*Repeat(p)

2. Matchup to Input Frequencies

/akam/	Input Fr.	Gen Fr.	Input #	Gen. #
akam-takam	1.000	0.700	88859	699841
akam-akam	0.000	0.000		
akam-sakam	0.000	0.220		220145
akam-pakam	0.000	0.043		42520
akam-makam	0.000	0.034		34422
akam-kakam	0.000	0.003		3072

/akan/	Input Fr.	Gen Fr.	Input #	Gen. #
akan-takan	1.000	0.700	91229	699841
akan-akan	0.000	0.000		
akan-sakan	0.000	0.220		220145
akan-pakan	0.000	0.043		42520
akan-makan	0.000	0.034		34422
akan-kakan	0.000	0.003		3072

/akaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
akaŋ-takaŋ	1.000	0.700	90432	699841
akaŋ-akaŋ	0.000	0.000		
akaŋ-sakaŋ	0.000	0.220		220145
akaŋ-pakaŋ	0.000	0.043		42520
akaŋ-makaŋ	0.000	0.034		34422
akaŋ-kakaŋ	0.000	0.003		3072

/akap/	Input Fr.	Gen Fr.	Input #	Gen. #
akap-takap	1.000	0.700	89285	699841
akap-akap	0.000	0.000		
akap-sakap	0.000	0.220		220145
akap-pakap	0.000	0.043		42520
akap-makap	0.000	0.034		34422
akap-kakap	0.000	0.003		3072

/akat/	Input Fr.	Gen Fr.	Input #	Gen. #
akat-takat	1.000	0.315	88602	314977
akat-akat	0.000	0.000		
akat-sakat	0.000	0.405		405086
akat-pakat	0.000	0.215		214548
akat-makat	0.000	0.041		40883
akat-kakat	0.000	0.025		24506

/amak/	Input Fr.	Gen Fr.	Input #	Gen. #
amak-samak	1.000	0.220	91519	220145
amak-amak	0.000	0.000		
amak-tamak	0.000	0.700		699841
amak-pamak	0.000	0.043		42520
amak-mamak	0.000	0.034		34422
amak-kamak	0.000	0.003		3072

/aman/	Input Fr.	Gen Fr.	Input #	Gen. #
aman-taman	1.000	0.700	90082	699841
aman-aman	0.000	0.000		
aman-saman	0.000	0.220		220145
aman-paman	0.000	0.043		42520
aman-maman	0.000	0.034		34422
aman-kaman	0.000	0.003		3072

/amaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
amaŋ-tamaŋ	1.000	0.700	89157	699841
amaŋ-amaŋ	0.000	0.000		
amaŋ-samaŋ	0.000	0.220		220145
amaŋ-pamaŋ	0.000	0.043		42520
amaŋ-mamaŋ	0.000	0.034		34422

amaŋ-kamaŋ	0.000	0.003		3072

/amap/	Input Fr.	Gen Fr.	Input #	Gen. #
amap-tamap	1.000	0.700	91367	699841
amap-amap	0.000	0.000		
amap-samap	0.000	0.220		220145
amap-pamap	0.000	0.043		42520
amap-mamap	0.000	0.034		34422
amap-kamap	0.000	0.003		3072

/amat/	Input Fr.	Gen Fr.	Input #	Gen. #
amat-samat	1.000	0.405	91650	405086
amat-amat	0.000	0.000		
amat-tamat	0.000	0.315		314977
amat-pamat	0.000	0.215		214548
amat-mamat	0.000	0.041		40883
amat-kamat	0.000	0.025		24506

/anak/	Input Fr.	Gen Fr.	Input #	Gen. #
anak-sanak	1.000	0.220	89599	220145
anak-anak	0.000	0.000		
anak-tanak	0.000	0.700		699841
anak-panak	0.000	0.043		42520
anak-manak	0.000	0.034		34422
anak-kanak	0.000	0.003		3072

/aŋak/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋak-saŋak	1.000	0.220	90140	220145
aŋak-aŋak	0.000	0.000		
aŋak-taŋak	0.000	0.700		699841
aŋak-paŋak	0.000	0.043		42520
aŋak-maŋak	0.000	0.034		34422
aŋak-kaŋak	0.000	0.003		3072

/aŋam/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋam-taŋam	1.000	0.700	90607	699841
aŋam-aŋam	0.000	0.000		
aŋam-saŋam	0.000	0.220		220145

aŋam-paŋam	0.000	0.043		42520
aŋam-maŋam	0.000	0.034		34422
aŋam-kaŋam	0.000	0.003		3072

/anam/	Input Fr.	Gen Fr.	Input #	Gen. #
anam-tanam	1.000	0.700	91017	699841
anam-anam	0.000	0.000		
anam-sanam	0.000	0.220		220145
anam-panam	0.000	0.043		42520
anam-manam	0.000	0.034		34422
anam-kanam	0.000	0.003		3072

/aŋan/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋan-saŋan	1.000	0.220	89682	220145
aŋan-aŋan	0.000	0.000		
aŋan-taŋan	0.000	0.700		699841
aŋan-paŋan	0.000	0.043		42520
aŋan-maŋan	0.000	0.034		34422
aŋan-kaŋan	0.000	0.003		3072

/anaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
anaŋ-sanaŋ	1.000	0.220	90061	220145
anaŋ-anaŋ	0.000	0.000		
anaŋ-tanaŋ	0.000	0.700		699841
anaŋ-panaŋ	0.000	0.043		42520
anaŋ-manaŋ	0.000	0.034		34422
anaŋ-kanaŋ	0.000	0.003		3072

/anap/	Input Fr.	Gen Fr.	Input #	Gen. #
anap-panap	1.000	0.043	89230	42520
anap-anap	0.000	0.000		
anap-tanap	0.000	0.700		699841
anap-sanap	0.000	0.220		220145
anap-manap	0.000	0.034		34422
anap-kanap	0.000	0.003		3072

/aŋap/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋap-taŋap	1.000	0.700	90218	699841

aŋap-aŋap	0.000	0.000		
aŋap-saŋap	0.000	0.220		220145
aŋap-paŋap	0.000	0.043		42520
aŋap-maŋap	0.000	0.034		34422
aŋap-kaŋap	0.000	0.003		3072

/aŋat/	Input Fr.	Gen Fr.	Input #	Gen. #
aŋat-saŋat	1.000	0.405	89092	405086
aŋat-aŋat	0.000	0.000		
aŋat-taŋat	0.000	0.315		314977
aŋat-paŋat	0.000	0.215		214548
aŋat-maŋat	0.000	0.041		40883
aŋat-kaŋat	0.000	0.025		24506

/anat/	Input Fr.	Gen Fr.	Input #	Gen. #
anat-sanat	1.000	0.405	89334	405086
anat-anat	0.000	0.000		
anat-tanat	0.000	0.315		314977
anat-panat	0.000	0.215		214548
anat-manat	0.000	0.041		40883
anat-kanat	0.000	0.025		24506

/apak/	Input Fr.	Gen Fr.	Input #	Gen. #
apak-tapak	1.000	0.700	89767	699841
apak-apak	0.000	0.000		
apak-sapak	0.000	0.220		220145
apak-papak	0.000	0.043		42520
apak-mapak	0.000	0.034		34422
apak-kapak	0.000	0.003		3072

/apam/	Input Fr.	Gen Fr.	Input #	Gen. #
apam-tapam	1.000	0.700	89746	699841
apam-apam	0.000	0.000		
apam-sapam	0.000	0.220		220145
apam-papam	0.000	0.043		42520
apam-mapam	0.000	0.034		34422
apam-kapam	0.000	0.003		3072

/apan/	Input Fr.	Gen Fr.	Input #	Gen. #
apan-tapan	1.000	0.700	89640	699841
apan-apan	0.000	0.000		
apan-sapan	0.000	0.220		220145
apan-papan	0.000	0.043		42520
apan-mapan	0.000	0.034		34422
apan-kapan	0.000	0.003		3072

/apaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
apaŋ-tapaŋ	1.000	0.700	89431	699841
apaŋ-apaŋ	0.000	0.000		
apaŋ-sapaŋ	0.000	0.220		220145
apaŋ-papaŋ	0.000	0.043		42520
apaŋ-mapaŋ	0.000	0.034		34422
apaŋ-kapaŋ	0.000	0.003		3072

/apat/	Input Fr.	Gen Fr.	Input #	Gen. #
apat-sapat	1.000	0.405	90138	405086
apat-apat	0.000	0.000		
apat-tapat	0.000	0.315		314977
apat-papat	0.000	0.215		214548
apat-mapat	0.000	0.041		40883
apat-kapat	0.000	0.025		24506

/asak/	Input Fr.	Gen Fr.	Input #	Gen. #
asak-pasak	1.000	0.043	90190	42520
asak-asak	0.000	0.000		
asak-tasak	0.000	0.700		699841
asak-sasak	0.000	0.220		220145
asak-masak	0.000	0.034		34422
asak-kasak	0.000	0.003		3072

/asam/	Input Fr.	Gen Fr.	Input #	Gen. #
asam-tasam	1.000	0.700	91523	699841
asam-asam	0.000	0.000		
asam-sasam	0.000	0.220		220145
asam-pasam	0.000	0.043		42520
asam-masam	0.000	0.034		34422

asam-kasam	0.000	0.003		3072

/asan/	Input Fr.	Gen Fr.	Input #	Gen. #
asan-tasan	1.000	0.700	88827	699841
asan-asan	0.000	0.000		
asan-sasan	0.000	0.220		220145
asan-pasan	0.000	0.043		42520
asan-masan	0.000	0.034		34422
asan-kasan	0.000	0.003		3072

/asaŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
asaŋ-masaŋ	1.000	0.034	90446	34422
asaŋ-asaŋ	0.000	0.000		
asaŋ-tasaŋ	0.000	0.700		699841
asaŋ-sasaŋ	0.000	0.220		220145
asaŋ-pasaŋ	0.000	0.043		42520
asaŋ-kasaŋ	0.000	0.003		3072

/asap/	Input Fr.	Gen Fr.	Input #	Gen. #
asap-pasap	1.000	0.043	90015	42520
asap-asap	0.000	0.000		
asap-tasap	0.000	0.700		699841
asap-sasap	0.000	0.220		220145
asap-masap	0.000	0.034		34422
asap-kasap	0.000	0.003		3072

/asat/	Input Fr.	Gen Fr.	Input #	Gen. #
asat-tasat	1.000	0.315	89582	314977
asat-asat	0.000	0.000		
asat-sasat	0.000	0.405		405086
asat-pasat	0.000	0.215		214548
asat-masat	0.000	0.041		40883
asat-kasat	0.000	0.025		24506

/atak/	Input Fr.	Gen Fr.	Input #	Gen. #
atak-tatak	1.000	0.315	90950	314977
atak-atak	0.000	0.000		
atak-satak	0.000	0.405		405086

atak-patak	0.000	0.215		214548
atak-matak	0.000	0.041		40883
atak-katak	0.000	0.025		24506

/atam/	Input Fr.	Gen Fr.	Input #	Gen. #
atam-tatam	1.000	0.315	89789	314977
atam-atam	0.000	0.000		
atam-satam	0.000	0.405		405086
atam-patam	0.000	0.215		214548
atam-matam	0.000	0.041		40883
atam-katam	0.000	0.025		24506

/ataŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ataŋ-pataŋ	1.000	0.215	91598	214548
ataŋ-ataŋ	0.000	0.000		
ataŋ-tataŋ	0.000	0.315		314977
ataŋ-sataŋ	0.000	0.405		405086
ataŋ-mataŋ	0.000	0.041		40883
ataŋ-kataŋ	0.000	0.025		24506

/atan/	Input Fr.	Gen Fr.	Input #	Gen. #
atan-tatan	1.000	0.315	90335	314977
atan-atan	0.000	0.000		
atan-satan	0.000	0.405		405086
atan-patan	0.000	0.215		214548
atan-matan	0.000	0.041		40883
atan-katan	0.000	0.025		24506

/atap/	Input Fr.	Gen Fr.	Input #	Gen. #
atap-satap	1.000	0.405	90728	405086
atap-atap	0.000	0.000		
atap-tatap	0.000	0.315		314977
atap-patap	0.000	0.215		214548
atap-matap	0.000	0.041		40883
atap-katap	0.000	0.025		24506

/atat/	Input Fr.	Gen Fr.	Input #	Gen. #
atat-satat	1.000	0.405	90045	405086

atat-atat	0.000	0.000		
atat-tatat	0.000	0.315		314977
atat-patat	0.000	0.215		214548
atat-matat	0.000	0.041		40883
atat-katat	0.000	0.025		24506

/ikim/	Input Fr.	Gen Fr.	Input #	Gen. #
ikim-tikim	1.000	0.700	91089	699841
ikim-ikim	0.000	0.000		
ikim-sikim	0.000	0.220		220145
ikim-pikim	0.000	0.043		42520
ikim-mikim	0.000	0.034		34422
ikim-kikim	0.000	0.003		3072

/ikin/	Input Fr.	Gen Fr.	Input #	Gen. #
ikin-sikin	1.000	0.220	92001	220145
ikin-ikin	0.000	0.000		
ikin-tikin	0.000	0.700		699841
ikin-pikin	0.000	0.043		42520
ikin-mikin	0.000	0.034		34422
ikin-kikin	0.000	0.003		3072

/ikiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ikiŋ-tikiŋ	1.000	0.700	90138	699841
ikiŋ-ikiŋ	0.000	0.000		
ikiŋ-sikiŋ	0.000	0.220		220145
ikiŋ-pikiŋ	0.000	0.043		42520
ikiŋ-mikiŋ	0.000	0.034		34422
ikiŋ-kikiŋ	0.000	0.003		3072

/ikip/	Input Fr.	Gen Fr.	Input #	Gen. #
ikip-tikip	1.000	0.700	91584	699841
ikip-ikip	0.000	0.000		
ikip-sikip	0.000	0.220		220145
ikip-pikip	0.000	0.043		42520
ikip-mikip	0.000	0.034		34422
ikip-kikip	0.000	0.003		3072

/ikit/	Input Fr.	Gen Fr.	Input #	Gen. #
ikit-sikit	1.000	0.405	90174	405086
ikit-ikit	0.000	0.000		
ikit-tikit	0.000	0.315		314977
ikit-pikit	0.000	0.215		214548
ikit-mikit	0.000	0.041		40883
ikit-kikit	0.000	0.025		24506

/imik/	Input Fr.	Gen Fr.	Input #	Gen. #
imik-simik	1.000	0.220	88291	220145
imik-imik	0.000	0.000		
imik-timik	0.000	0.700		699841
imik-pimik	0.000	0.043		42520
imik-mimik	0.000	0.034		34422
imik-kimik	0.000	0.003		3072

/imiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
imiŋ-timiŋ	1.000	0.700	88085	699841
imiŋ-imiŋ	0.000	0.000		
imiŋ-simiŋ	0.000	0.220		220145
imiŋ-pimiŋ	0.000	0.043		42520
imiŋ-mimiŋ	0.000	0.034		34422
imiŋ-kimiŋ	0.000	0.003		3072

/imin/	Input Fr.	Gen Fr.	Input #	Gen. #
imin-timin	1.000	0.700	92141	699841
imin-imin	0.000	0.000		
imin-simin	0.000	0.220		220145
imin-pimin	0.000	0.043		42520
imin-mimin	0.000	0.034		34422
imin-kimin	0.000	0.003		3072

/imip/	Input Fr.	Gen Fr.	Input #	Gen. #
imip-simip	1.000	0.220	90788	220145
imip-imip	0.000	0.000		
imip-timip	0.000	0.700		699841
imip-pimip	0.000	0.043		42520
imip-mimip	0.000	0.034		34422

imip-kimip	0.000	0.003		3072

/imit/	Input Fr.	Gen Fr.	Input #	Gen. #
imit-simit	1.000	0.220	88656	220145
imit-imit	0.000	0.000		
imit-timit	0.000	0.700		699841
imit-pimit	0.000	0.043		42520
imit-mimit	0.000	0.034		34422
imit-kimit	0.000	0.003		3072

/iŋik/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋik-tiŋik	1.000	0.700	89882	699841
iŋik-iŋik	0.000	0.000		
iŋik-siŋik	0.000	0.220		220145
iŋik-piŋik	0.000	0.043		42520
iŋik-miŋik	0.000	0.034		34422
iŋik-kiŋik	0.000	0.003		3072

/inik/	Input Fr.	Gen Fr.	Input #	Gen. #
inik-tinik	1.000	0.700	89461	699841
inik-inik	0.000	0.000		
inik-sinik	0.000	0.220		220145
inik-pinik	0.000	0.043		42520
inik-minik	0.000	0.034		34422
inik-kinik	0.000	0.003		3072

/iŋim/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋim-tiŋim	1.000	0.700	91406	699841
iŋim-iŋim	0.000	0.000		
iŋim-siŋim	0.000	0.220		220145
iŋim-piŋim	0.000	0.043		42520
iŋim-miŋim	0.000	0.034		34422
iŋim-kiŋim	0.000	0.003		3072

/inim/	Input Fr.	Gen Fr.	Input #	Gen. #
inim-tinim	1.000	0.700	89472	699841
inim-inim	0.000	0.000		
inim-sinim	0.000	0.220		220145

inim-pinim	0.000	0.043		42520
inim-minim	0.000	0.034		34422
inim-kinim	0.000	0.003		3072

/iŋin/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋin-tiŋin	1.000	0.700	90038	699841
iŋin-iŋin	0.000	0.000		
iŋin-sʔiŋin	0.000	0.220		220145
iŋin-piŋin	0.000	0.043		42520
iŋin-miŋin	0.000	0.034		34422
iŋin-kiŋin	0.000	0.003		3072

/iniŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
iniŋ-siniŋ	1.000	0.220	90825	220145
iniŋ-iniŋ	0.000	0.000		
iniŋ-tiniŋ	0.000	0.700		699841
iniŋ-piniŋ	0.000	0.043		42520
iniŋ-miniŋ	0.000	0.034		34422
iniŋ-kiniŋ	0.000	0.003		3072

/inip/	Input Fr.	Gen Fr.	Input #	Gen. #
inip-tinip	1.000	0.700	90196	699841
inip-inip	0.000	0.000		
inip-sinip	0.000	0.220		220145
inip-pinip	0.000	0.043		42520
inip-minip	0.000	0.034		34422
inip-kinip	0.000	0.003		3072

/iŋip/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋip-tiŋip	1.000	0.700	90550	699841
iŋip-iŋip	0.000	0.000		
iŋip-siŋip	0.000	0.220		220145
iŋip-piŋip	0.000	0.043		42520
iŋip-miŋip	0.000	0.034		34422
iŋip-kiŋip	0.000	0.003		3072

/iŋit/	Input Fr.	Gen Fr.	Input #	Gen. #
iŋit-siŋit	1.000	0.405	90108	405086

iŋit-iŋit	0.000	0.000		
iŋit-tiŋit	0.000	0.315		314977
iŋit-piŋit	0.000	0.215		214548
iŋit-miŋit	0.000	0.041		40883
iŋit-kiŋit	0.000	0.025		24506

/init/	Input Fr.	Gen Fr.	Input #	Gen. #
init-pinit	1.000	0.215	89651	214548
init-init	0.000	0.000		
init-tinit	0.000	0.315		314977
init-sinit	0.000	0.405		405086
init-minit	0.000	0.041		40883
init-kinit	0.000	0.025		24506

/ipik/	Input Fr.	Gen Fr.	Input #	Gen. #
ipik-tipik	1.000	0.700	89527	699841
ipik-ipik	0.000	0.000		
ipik-sipik	0.000	0.220		220145
ipik-pipik	0.000	0.043		42520
ipik-mipik	0.000	0.034		34422
ipik-kipik	0.000	0.003		3072

/ipim/	Input Fr.	Gen Fr.	Input #	Gen. #
ipim-tipim	1.000	0.700	89327	699841
ipim-ipim	0.000	0.000		
ipim-sipim	0.000	0.220		220145
ipim-pipim	0.000	0.043		42520
ipim-mipim	0.000	0.034		34422
ipim-kipim	0.000	0.003		3072

/ipiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ipiŋ-tipiŋ	1.000	0.700	89651	699841
ipiŋ-ipiŋ	0.000	0.000		
ipiŋ-sipiŋ	0.000	0.220		220145
ipiŋ-pipiŋ	0.000	0.043		42520
ipiŋ-mipiŋ	0.000	0.034		34422
ipiŋ-kipiŋ	0.000	0.003		3072

/ipin/	Input Fr.	Gen Fr.	Input #	Gen. #
ipin-tipin	1.000	0.700	90618	699841
ipin-ipin	0.000	0.000		
ipin-sipin	0.000	0.220		220145
ipin-pipin	0.000	0.043		42520
ipin-mipin	0.000	0.034		34422
ipin-kipin	0.000	0.003		3072

/ipit/	Input Fr.	Gen Fr.	Input #	Gen. #
ipit-sipit	1.000	0.405	90623	405086
ipit-ipit	0.000	0.000		
ipit-tipit	0.000	0.315		314977
ipit-pipit	0.000	0.215		214548
ipit-mipit	0.000	0.041		40883
ipit-kipit	0.000	0.025		24506

/isik/	Input Fr.	Gen Fr.	Input #	Gen. #
isik-pisik	1.000	0.043	90338	42520
isik-isik	0.000	0.000		
isik-tisik	0.000	0.700		699841
isik-sisik	0.000	0.220		220145
isik-misik	0.000	0.034		34422
isik-kisik	0.000	0.003		3072

/isim/	Input Fr.	Gen Fr.	Input #	Gen. #
isim-tisim	1.000	0.700	89817	699841
isim-isim	0.000	0.000		
isim-sisim	0.000	0.220		220145
isim-pisim	0.000	0.043		42520
isim-misim	0.000	0.034		34422
isim-kisim	0.000	0.003		3072

/isiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
isiŋ-misiŋ	1.000	0.034	90221	34422
isiŋ-isiŋ	0.000	0.000		
isiŋ-tisiŋ	0.000	0.700		699841
isiŋ-sisiŋ	0.000	0.220		220145
isiŋ-pisiŋ	0.000	0.043		42520

isiŋ-kisiŋ	0.000	0.003		3072

/isin/	Input Fr.	Gen Fr.	Input #	Gen. #
isin-tisin	1.000	0.700	88753	699841
isin-isin	0.000	0.000		
isin-sisin	0.000	0.220		220145
isin-pisin	0.000	0.043		42520
isin-misin	0.000	0.034		34422
isin-kisin	0.000	0.003		3072

/isip/	Input Fr.	Gen Fr.	Input #	Gen. #
isip-pisip	1.000	0.043	89845	42520
isip-isip	0.000	0.000		
isip-tisip	0.000	0.700		699841
isip-sisip	0.000	0.220		220145
isip-misip	0.000	0.034		34422
isip-kisip	0.000	0.003		3072

/isit/	Input Fr.	Gen Fr.	Input #	Gen. #
isit-tisit	1.000	0.700	87848	699841
isit-isit	0.000	0.000		
isit-sisit	0.000	0.220		220145
isit-pisit	0.000	0.043		42520
isit-misit	0.000	0.034		34422
isit-kisit	0.000	0.003		3072

/itik/	Input Fr.	Gen Fr.	Input #	Gen. #
itik-sitik	1.000	0.405	91944	405086
itik-itik	0.000	0.000		
itik-titik	0.000	0.315		314977
itik-pitik	0.000	0.215		214548
itik-mitik	0.000	0.041		40883
itik-kitik	0.000	0.025		24506

/itim/	Input Fr.	Gen Fr.	Input #	Gen. #
itim-mitim	1.000	0.041	88574	40883
itim-itim	0.000	0.000		
itim-titim	0.000	0.315		314977

itim-sitim	0.000	0.405		405086
itim-pitim	0.000	0.215		214548
itim-kitim	0.000	0.025		24506

/itin/	Input Fr.	Gen Fr.	Input #	Gen. #
itin-pitin	1.000	0.215	90867	214548
itin-itin	0.000	0.000		
itin-titin	0.000	0.315		314977
itin-sitin	0.000	0.405		405086
itin-mitin	0.000	0.041		40883
itin-kitin	0.000	0.025		24506

/itiŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
itiŋ-titiŋ	1.000	0.315	90558	314977
itiŋ-itiŋ	0.000	0.000		
itiŋ-sitiŋ	0.000	0.405		405086
itiŋ-pitiŋ	0.000	0.215		214548
itiŋ-mitiŋ	0.000	0.041		40883
itiŋ-kitiŋ	0.000	0.025		24506

/itip/	Input Fr.	Gen Fr.	Input #	Gen. #
itip-sitip	1.000	0.405	89610	405086
itip-itip	0.000	0.000		
itip-titip	0.000	0.315		314977
itip-pitip	0.000	0.215		214548
itip-mitip	0.000	0.041		40883
itip-kitip	0.000	0.025		24506

/itit/	Input Fr.	Gen Fr.	Input #	Gen. #
itit-mitit	1.000	0.041	90185	40883
itit-itit	0.000	0.000		
itit-titit	0.000	0.315		314977
itit-sitit	0.000	0.405		405086
itit-pitit	0.000	0.215		214548
itit-kitit	0.000	0.025		24506

/ukum/	Input Fr.	Gen Fr.	Input #	Gen. #
ukum-sukum	1.000	0.220	88834	220145

ukum-ukum	0.000	0.000		
ukum-tukum	0.000	0.700		699841
ukum-pukum	0.000	0.043		42520
ukum-mukum	0.000	0.034		34422
ukum-kukum	0.000	0.003		3072

/ukuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
ukuŋ-tukuŋ	1.000	0.700	90344	699841
ukuŋ-ukuŋ	0.000	0.000		
ukuŋ-sukuŋ	0.000	0.220		220145
ukuŋ-pukuŋ	0.000	0.043		42520
ukuŋ-mukuŋ	0.000	0.034		34422
ukuŋ-kukuŋ	0.000	0.003		3072

/ukun/	Input Fr.	Gen Fr.	Input #	Gen. #
ukun-tukun	1.000	0.700	89797	699841
ukun-ukun	0.000	0.000		
ukun-sukun	0.000	0.220		220145
ukun-pukun	0.000	0.043		42520
ukun-mukun	0.000	0.034		34422
ukun-kukun	0.000	0.003		3072

/ukup/	Input Fr.	Gen Fr.	Input #	Gen. #
ukup-tukup	1.000	0.700	89932	699841
ukup-ukup	0.000	0.000		
ukup-sukup	0.000	0.220		220145
ukup-pukup	0.000	0.043		42520
ukup-mukup	0.000	0.034		34422
ukup-kukup	0.000	0.003		3072

/ukut/	Input Fr.	Gen Fr.	Input #	Gen. #
ukut-sukut	1.000	0.405	91074	405086
ukut-ukut	0.000	0.000		
ukut-tukut	0.000	0.315		314977
ukut-pukut	0.000	0.215		214548
ukut-mukut	0.000	0.041		40883
ukut-kukut	0.000	0.025		24506

/umuk/	Input Fr.	Gen Fr.	Input #	Gen. #
umuk-tumuk	1.000	0.700	88912	699841
umuk-umuk	0.000	0.000		
umuk-sumuk	0.000	0.220		220145
umuk-pumuk	0.000	0.043		42520
umuk-mumuk	0.000	0.034		34422
umuk-kumuk	0.000	0.003		3072

/umun/	Input Fr.	Gen Fr.	Input #	Gen. #
umun-tumun	1.000	0.700	89213	699841
umun-umun	0.000	0.000		
umun-sumun	0.000	0.220		220145
umun-pumun	0.000	0.043		42520
umun-mumun	0.000	0.034		34422
umun-kumun	0.000	0.003		3072

/umuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
umuŋ-tumuŋ	1.000	0.700	90784	699841
umuŋ-umuŋ	0.000	0.000		
umuŋ-sumuŋ	0.000	0.220		220145
umuŋ-pumuŋ	0.000	0.043		42520
umuŋ-mumuŋ	0.000	0.034		34422
umuŋ-kumuŋ	0.000	0.003		3072

/umup/	Input Fr.	Gen Fr.	Input #	Gen. #
umup-pumup	1.000	0.043	90768	42520
umup-umup	0.000	0.000		
umup-tumup	0.000	0.700		699841
umup-sumup	0.000	0.220		220145
umup-mumup	0.000	0.034		34422
umup-kumup	0.000	0.003		3072

/umut/	Input Fr.	Gen Fr.	Input #	Gen. #
umut-pumut	1.000	0.215	91005	214548
umut-umut	0.000	0.000		
umut-tumut	0.000	0.315		314977
umut-sumut	0.000	0.405		405086
umut-mumut	0.000	0.041		40883

umut-kumut	0.000	0.025		24506

/unuk/	Input Fr.	Gen Fr.	Input #	Gen. #
unuk-sunuk	1.000	0.220	90948	220145
unuk-unuk	0.000	0.000		
unuk-tunuk	0.000	0.700		699841
unuk-punuk	0.000	0.043		42520
unuk-munuk	0.000	0.034		34422
unuk-kunuk	0.000	0.003		3072

/uŋuk/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋuk-kuŋuk	1.000	0.003	90028	3072
uŋuk-uŋuk	0.000	0.000		
uŋuk-tuŋuk	0.000	0.700		699841
uŋuk-suŋuk	0.000	0.220		220145
uŋuk-puŋuk	0.000	0.043		42520
uŋuk-muŋuk	0.000	0.034		34422

/uŋum/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋum-tuŋum	1.000	0.700	90819	699841
uŋum-uŋum	0.000	0.000		
uŋum-suŋum	0.000	0.220		220145
uŋum-puŋum	0.000	0.043		42520
uŋum-muŋum	0.000	0.034		34422
uŋum-kuŋum	0.000	0.003		3072

/unum/	Input Fr.	Gen Fr.	Input #	Gen. #
unum-tunum	1.000	0.700	90321	699841
unum-unum	0.000	0.000		
unum-sunum	0.000	0.220		220145
unum-punum	0.000	0.043		42520
unum-munum	0.000	0.034		34422
unum-kunum	0.000	0.003		3072

/uŋun/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋun-tuŋun	1.000	0.700	88736	699841
uŋun-uŋun	0.000	0.000		
uŋun-suŋun	0.000	0.220		220145

uŋun-puŋun	0.000	0.043		42520
uŋun-muŋun	0.000	0.034		34422
uŋun-kuŋun	0.000	0.003		3072

/unuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
unuŋ-sunuŋ	1.000	0.220	89415	220145
unuŋ-unuŋ	0.000	0.000		
unuŋ-tunuŋ	0.000	0.700		699841
unuŋ-punuŋ	0.000	0.043		42520
unuŋ-munuŋ	0.000	0.034		34422
unuŋ-kunuŋ	0.000	0.003		3072

/unup/	Input Fr.	Gen Fr.	Input #	Gen. #
unup-tunup	1.000	0.700	90526	699841
unup-unup	0.000	0.000		
unup-sunup	0.000	0.220		220145
unup-punup	0.000	0.043		42520
unup-munup	0.000	0.034		34422
unup-kunup	0.000	0.003		3072

/uŋup/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋup-tuŋup	1.000	0.700	89436	699841
uŋup-uŋup	0.000	0.000		
uŋup-suŋup	0.000	0.220		220145
uŋup-puŋup	0.000	0.043		42520
uŋup-muŋup	0.000	0.034		34422
uŋup-kuŋup	0.000	0.003		3072

/uŋut/	Input Fr.	Gen Fr.	Input #	Gen. #
uŋut-tuŋut	1.000	0.315	89877	314977
uŋut-uŋut	0.000	0.000		
uŋut-suŋut	0.000	0.405		405086
uŋut-puŋut	0.000	0.215		214548
uŋut-muŋut	0.000	0.041		40883
uŋut-kuŋut	0.000	0.025		24506

/unut/	Input Fr.	Gen Fr.	Input #	Gen. #
unut-sunut	1.000	0.405	89418	405086

unut-unut	0.000	0.000		
unut-tunut	0.000	0.315		314977
unut-punut	0.000	0.215		214548
unut-munut	0.000	0.041		40883
unut-kunut	0.000	0.025		24506

/upuk/	Input Fr.	Gen Fr.	Input #	Gen. #
upuk-tupuk	1.000	0.700	91765	699841
upuk-upuk	0.000	0.000		
upuk-supuk	0.000	0.220		220145
upuk-pupuk	0.000	0.043		42520
upuk-mupuk	0.000	0.034		34422
upuk-kupuk	0.000	0.003		3072

/upum/	Input Fr.	Gen Fr.	Input #	Gen. #
upum-tupum	1.000	0.700	89981	699841
upum-upum	0.000	0.000		
upum-supum	0.000	0.220		220145
upum-pupum	0.000	0.043		42520
upum-mupum	0.000	0.034		34422
upum-kupum	0.000	0.003		3072

/upun/	Input Fr.	Gen Fr.	Input #	Gen. #
upun-tupun	1.000	0.700	89892	699841
upun-upun	0.000	0.000		
upun-supun	0.000	0.220		220145
upun-pupun	0.000	0.043		42520
upun-mupun	0.000	0.034		34422
upun-kupun	0.000	0.003		3072

/upuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
upuŋ-tupuŋ	1.000	0.700	90281	699841
upuŋ-upuŋ	0.000	0.000		
upuŋ-supuŋ	0.000	0.220		220145
upuŋ-pupuŋ	0.000	0.043		42520
upuŋ-mupuŋ	0.000	0.034		34422
upuŋ-kupuŋ	0.000	0.003		3072

/uput/	Input Fr.	Gen Fr.	Input #	Gen. #
uput-suput	1.000	0.220	89414	220145
uput-uput	0.000	0.000		
uput-tuput	0.000	0.700		699841
uput-puput	0.000	0.043		42520
uput-muput	0.000	0.034		34422
uput-kuput	0.000	0.003		3072

/usuk/	Input Fr.	Gen Fr.	Input #	Gen. #
usuk-tusuk	1.000	0.700	88927	699841
usuk-usuk	0.000	0.000		
usuk-susuk	0.000	0.220		220145
usuk-pusuk	0.000	0.043		42520
usuk-musuk	0.000	0.034		34422
usuk-kusuk	0.000	0.003		3072

/usum/	Input Fr.	Gen Fr.	Input #	Gen. #
usum-tusum	1.000	0.700	89911	699841
usum-usum	0.000	0.000		
usum-susum	0.000	0.220		220145
usum-pusum	0.000	0.043		42520
usum-musum	0.000	0.034		34422
usum-kusum	0.000	0.003		3072

/usun/	Input Fr.	Gen Fr.	Input #	Gen. #
usun-tusun	1.000	0.700	91896	699841
usun-usun	0.000	0.000		
usun-susun	0.000	0.220		220145
usun-pusun	0.000	0.043		42520
usun-musun	0.000	0.034		34422
usun-kusun	0.000	0.003		3072

/usuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
usuŋ-tusuŋ	1.000	0.700	90651	699841
usuŋ-usuŋ	0.000	0.000		
usuŋ-susuŋ	0.000	0.220		220145
usuŋ-pusuŋ	0.000	0.043		42520
usuŋ-musuŋ	0.000	0.034		34422

usuŋ-kusuŋ	0.000	0.003		3072

/usup/	Input Fr.	Gen Fr.	Input #	Gen. #
usup-tusup	1.000	0.700	89511	699841
usup-usup	0.000	0.000		
usup-susup	0.000	0.220		220145
usup-pusup	0.000	0.043		42520
usup-musup	0.000	0.034		34422
usup-kusup	0.000	0.003		3072

/usut/	Input Fr.	Gen Fr.	Input #	Gen. #
usut-tusut	1.000	0.315	90078	314977
usut-usut	0.000	0.000		
usut-susut	0.000	0.405		405086
usut-pusut	0.000	0.215		214548
usut-musut	0.000	0.041		40883
usut-kusut	0.000	0.025		24506

/utuk/	Input Fr.	Gen Fr.	Input #	Gen. #
utuk-tutuk	1.000	0.315	90234	314977
utuk-utuk	0.000	0.000		
utuk-sutuk	0.000	0.405		405086
utuk-putuk	0.000	0.215		214548
utuk-mutuk	0.000	0.041		40883
utuk-kutuk	0.000	0.025		24506

/utum/	Input Fr.	Gen Fr.	Input #	Gen. #
utum-tutum	1.000	0.315	89579	314977
utum-utum	0.000	0.000		
utum-sutum	0.000	0.405		405086
utum-putum	0.000	0.215		214548
utum-mutum	0.000	0.041		40883
utum-kutum	0.000	0.025		24506

/utun/	Input Fr.	Gen Fr.	Input #	Gen. #
utun-sutun	1.000	0.405	91324	405086
utun-utun	0.000	0.000		
utun-tutun	0.000	0.315		314977

utun-putun	0.000	0.215		214548
utun-mutun	0.000	0.041		40883
utun-kutun	0.000	0.025		24506

/utuŋ/	Input Fr.	Gen Fr.	Input #	Gen. #
utuŋ-tutuŋ	1.000	0.315	90882	314977
utuŋ-utuŋ	0.000	0.000		
utuŋ-sutuŋ	0.000	0.405		405086
utuŋ-putuŋ	0.000	0.215		214548
utuŋ-mutuŋ	0.000	0.041		40883
utuŋ-kutuŋ	0.000	0.025		24506

/utup/	Input Fr.	Gen Fr.	Input #	Gen. #
utup-sutup	1.000	0.405	89049	405086
utup-utup	0.000	0.000		
utup-tutup	0.000	0.315		314977
utup-putup	0.000	0.215		214548
utup-mutup	0.000	0.041		40883
utup-kutup	0.000	0.025		24506

/utut/	Input Fr.	Gen Fr.	Input #	Gen. #
utut-sutut	1.000	0.405	89684	405086
utut-utut	0.000	0.000		
utut-tutut	0.000	0.315		314977
utut-putut	0.000	0.215		214548
utut-mutut	0.000	0.041		40883
utut-kutut	0.000	0.025		24506

3. Tableaux

The following are approximate tableaux for this ranking. Outputs are derived simply by sorting the constraints by their ranking value, with no stochastic variation.

To diagnose variation, consult two things:

--The candidate frequencies (which are the generated frequencies, not the input frequencies).

--The probability that each constraint outranks the next
one down, given directly after the constraint labels.

/amak/:

	Onset (.99985)	*Dor (.74)	*NasV (.641)	*Lab (.546)	*Repeat(t) (.506)	*FricV (.728)	*StopV (.662)	t (.638)	*Cor (.747)	p (.641)	?(1)	*Repeat(p)
amak-tamak (0.700)	1						1		1	1	1	
amak-samak (0.220)	1					1!		1	1	1	1	
amak-pamak (0.043)	1			1!			1	1			1	
amak-mamak (0.034)	1		1!	1				1		1	1	
amak-kamak (0.003)	1	1!					1	1		1	1	
amak-amak	2!											

/atan/:

	Onset (.99985)	*Dor (.74)	*NasV (.641)	*Lab (.546)	*Repeat(t) (.506)	*FricV (.728)	*StopV (.662)	t (.638)	*Cor (.747)	p (.641)	?(1)	*Repeat(p)
atan-satan (0.405)	1					1		1	1	1	1	
atan-tatan (0.315)	1				1!		1		1	1	1	
atan-patan (0.215)	1			1!			1	1			1	
atan-matan (0.041)	1		1!	1				1		1	1	
atan-katan (0.025)	1	1!					1	1		1	1	
atan-atan	2!											

4. Active Constraints

A constraint is active if it causes the winning candidate to defeat a rival in at least one competition.

Active	Onset
Active	*Dor
Active	*NasV
Active	*Lab
Active	*Repeat(t)
Active	*FricV
Active	*StopV
Active	t
Active	*Cor
Active	p
Inactive	ʧ
Inactive	*Repeat(p)

5. Testing the Grammar: Details

The grammar was tested for 1000000 cycles.
Average error per candidate: 4.349 percent
Learning time: 3.253 minutes

6. Parameter Values Used by the GLA

Initial Rankings

All constraints started out at the default value of 100.

Schedule for GLA Parameters

	Stage	Trials	PlastMark	PlastFaith	NoiseMark	NoiseFaith
1	12500000	25000002.000	2.0002.000	2.0002.000	2.0002.000	2.000
2	22500000	25000000.159	0.1590.159	0.1592.000	2.0002.000	2.000
3	32500000	25000000.013	0.0130.013	0.0132.000	2.0002.000	2.000
4	42500000	25000000.001	0.0010.001	0.0012.000	2.0002.000	2.000

There were a total of 10000000 learning trials.

Chapter 4
Local Relationships

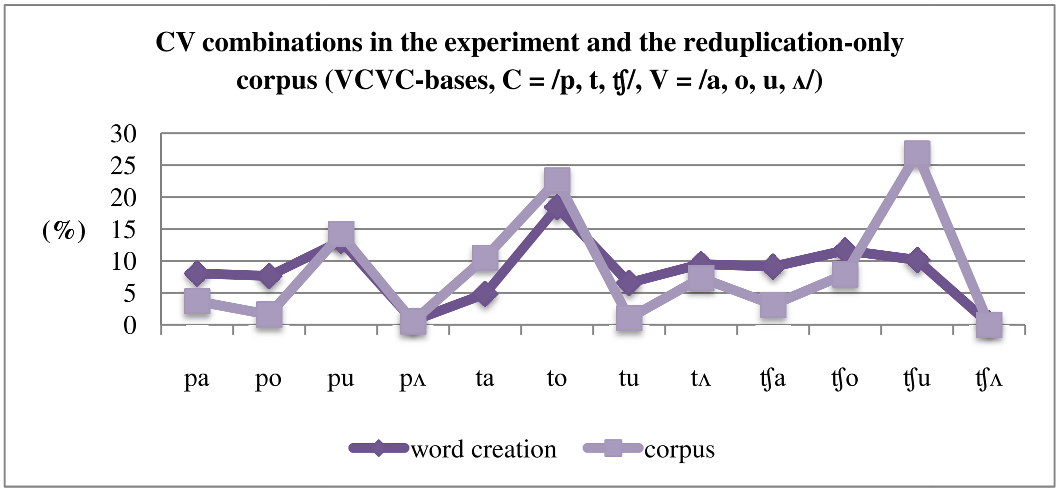
We saw in Chapter 2 that the frequency of inserted consonants in the word creation experiment did show significant correlations with the frequency of consonants occurring in word-initial position in the entire corpus. However, this correlation was not perfect: the participants in the experiment had distinct preferences for different Cs in consonant insertion. Speakers also appeared sensitive to context; that is, when the base contained a consonant that was identical to their generally favored consonant, they had a tendency to choose a different consonant for insertion. This tendency of context-sensitivity invites further investigation of contextual factors. In this chapter I will consider reduplications of the form VCVC-**CI**VCVC, in which CI is flanked by a vowel on the right-hand side and a consonant on the left-hand side. I will examine how the choice of CI is affected by the following V and by the preceding C. I will suggest that the choice of CI is affected by the co-occurrence restriction between C and V in a syllable, based on the argument that languages with an English-type syllable structure do not have C-V cooccurrence restrictions, but have V-C restrictions; whereas languages with Korean-type syllable structure have C-V restrictions, but do not have V-C restrictions. Besides, I argue that the CI choice is also affected by a grammatical requirement governing C.C sequences, the syllable contact constraints.

4.1 C – V Relationship

4.1.1 CV combining patterns

We have seen in Chapter 2 that the frequency of C – V combinations in the word creation experiment correlated with the frequency of C – V combinations in the lexicon, particularly the reduplication-only corpus, as in Figure 4.1.

(1) Figure 4.1 CV combinations in the experiment (Experiment 1) and the reduplication-only corpus: VCVC-bases, C = /p, t, ʧ/, V = /a, o, u, ʌ/



Is the frequency pattern of C – V combinations purely based on the existing lexicon? Why are /p/, /t/, and /ʧ/ more likely to be followed by /u/, /o/, and /o, u/, respectively, than by other vowels? In the following section I discuss evidence that Korean speakers are particularly sensitive to the relationship between the syllable onset and nucleus, and discuss the implications of this for theories of syllable-internal constituent structure.

Korean speakers are sensitive to the CV composition in the general vocabulary, and they are predicted to be alert about the CV co-occurrences in the reduplication when they need to make up new reduplicative words. The frequency of CV co-occurrences will be influential in inserting a new C that is to be followed by an existing V, resulting in a new CV. Just as the CV portion acts as a unit in other experiment tasks like word blending or list recall, so will the cohesive unit of CV serve as an important clue to which C to insert in a word creation task.[50] I argue that speakers make reference to the existing CV co-

[50] The tendency that CV, rather than VC, act together in a closer relationship is also espoused by the articulatory phonology: (Goldstein *et al*. 2008: 4)

> It can also explain macroscopic universal patterns associated with syllable structure, for example, why syllables with onsets (CV) are universal while those with codas are not. This follows from the fact that the in-phase mode is more accessible and more stable. Similarly, it can account for the fact that onsets and Vs combine relatively freely, while combinations of V and coda Cs can be more restricted and for the fact that onsets emerge earlier in phonological development.

occurrences in the lexicon when they are asked to epenthesize a C which will be part of a CV syllable in a newly created word.

In sum, it is the phonotactic probabilities in speakers' mind that affect the choice of CIs. Speakers remember frequent reduplicated forms whose CV combinations are also in store for them to use in creating new reduplicated forms, which means that more frequent forms are more influential in creating novel forms.

Figure 4.1 above shows that the speakers in the word creation task not only followed the overall frequency pattern of CV combinations from the reduplication-only corpus, but they also appeared to be more sensitive to the combinations in more frequent words. The speakers chose /pu/, /to/, /ʧo/ most frequently, among other CV combinations of pV, tV, ʧV, respectively, which is actually confirmed by the fact that the most frequent single reduplicated forms with pV, tV, ʧV are ult^{h}ung-pult^{h}ung 'bumpy,' oson-toson 'harmoniously,' oŋki-ʧoŋki 'thickly' (closely after umul-ʧ'umul 'hesitantly' and aki-ʧaki 'charming') in the reduplication-only corpus. This finding is consistent with the emergent model according to which frequent words, as opposed to infrequent words, are representative of the sound distribution in a language, e.g. English (Dell *et al.* 1993; Lee 2006). For now, it is not clear which is more critical in showing the correlations between the frequency pattern of CV co-occurrences in the experiment and the frequency pattern of CV co-occurrences in the reduplication-only corpus. It is that the most frequent CV combinations of a single word, as well as the overall frequency of CV combinations, have impact on phonotactics.

The research in this section shows that lexical statistics may provide a basis for the phonotactics of a language, which cannot be absolutely determined but rather, probabilistically determined. On the whole, the speakers' behavior mirrors what exists in the lexicon for CV co-occurrences; that is, it reflects the speakers' phonotactical knowledge of the sub-syllabic CV constituent, based on CV combinations. Although this does not account for why a C tends to be followed by a V more frequently than other Vs, but it could show why the attested frequency pattern of CV combinations in the experiment had to be the way they were.

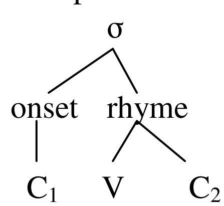
4.1.2 Sub-syllabic restriction

The syllable, which is composed of three parts (onset, nucleus, and coda), has a sub-hierarchical portion that serves as a unit in linguistic behaviors. Behavioral experiments have shown that Korean speakers tend to group C_1 and V

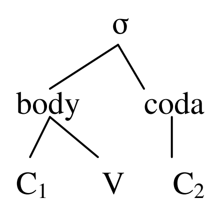
(2b), rather than V and C_2, as a unit in a syllable structure of C_1VC_2, contrary to prediction that V and C_2 will form a sub-syllabic unit on the basis of the traditional syllable structure (2a), onset + rhyme which sub-hierarchically consists of nucleus V and coda C (Chen *et al*. 2004; Derwing *et al*. 1993; Lee 2006). I argue that it cannot be simply the abstract representation of syllable *per se* (2b) body (= onset + nucleus) + coda – in which the term "body" stems from Vennemann (1988) – that motivates and accounts for this grouping of onset and nucleus. Rather, I adopt a claim that there is a sub-syllabic constituent that requires interdependency between C and V, not V and C, to which speakers of the language make reference in their behavior.

(2) Structural representations of syllable (σ)

a.



b.



Note that this syllable structure is implicitly implemented in the spelling norms of Korean. Korean employs a syllable-based spelling system, in which the locus of C_1, in a structure of C_1VC_2, should always be filled up even when only V or VC_2 exist but C_1 does not, whereas the locus of C_2 can be left empty when the sound C_2 does not exist, e.g. 아 /a/, 악 /ak/, 가 /ka/, 각 /kak/.

With regard to the CI choice in the reduplication, I suggest that Korean speakers are implicitly aware that onset and nucleus are tightly connected at the sub-syllabic level. That is, the sub-syllabic units, onset-nucleus are connected, although for now I do not claim that this connection is tighter than the other connection, i.e. nucleus-coda: One might wonder if there is a case of coda C insertion, in which we may be able to seek relationship between inserted coda Cs and the preceding extant Vs. If such a case exists, we can compare the two types of relationship, onset-nucleus vs. nucleus-coda, in terms of CV and VC restrictions. However, I am not aware of such an instance of coda insertion at this point, and I will leave it for future investigation.

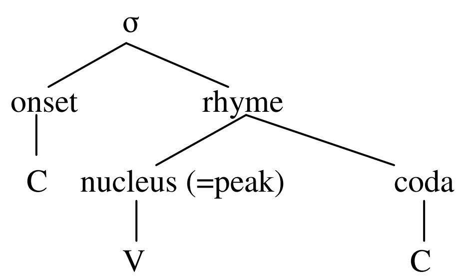
The CV combination patterns both in the word creation experiment and the reduplication-only corpus testify that existing Vs affect the choice of CIs, which indicates an intimate relation between onset C and nucleus V.

4.1.2.1 Rhyme: sub-syllabic grouping of V – C

I suggest that different languages have differing preferences for a sub-syllabic level for a syllable: in my discussion I will distinguish the syllable structures of English-type languages and of Korean-type languages.[51] In this section I look into the syllable structure for English-type languages, based on the English data.

The internal structure of the syllable can be defined by different approaches: syntax-based constituent approach, sonority-based approach, and finite-state approach based on permitted sequences in a given language (Goldsmith 2009: 28). According to the syntax-based view, a syllable is hierarchically constructed with onset + rhyme (= nucleus + coda):

(3) Structural representations of syllable (σ)



This structure has been conceived on the basis of English, which has been found to have stronger co-occurrence restriction between nucleus and coda than between nucleus or coda and the onset. For example, in a single syllable a sequence of simple vowel and sonorant may be followed by any consonant, whereas a sequence of complex vowel (diphthong) and sonorant can be followed only by a coronal consonant (Fudge 1969; Selkirk 1982). However, there are no such collocational constraints attested for the combination of onset and nucleus. Therefore, nucleus and coda, not onset and nucleus, should form a constituent within a syllable.

Psycholinguistic behavioral experiments with regard to syllable structure suggest that a syllable has an internal constituency: it has at least a level of onset and rhyme (Treiman 1983; Treiman *et al*. 2000). In a series of the novel word

[51] Andries Coetzee (p.c.) also suggested this possibility that there is “a parameter available in UG for languages to choose whether they use an English-like (*onset* + *rhyme*) or Korean-like (*body* + *coda*) syllable structure.” [italicized words added by the author]

game experiments with English speakers, the participants showed strong sensitivity to the intermediate units of syllable, onset and rhyme, although they did not show good sensitivity to sub-hierarchical division of nucleus and coda (Treiman 1983). For example, when the English speakers were asked to divide a nonce word consisting of CVC into two groups, they preferred to divide it into C plus VC rather than into CV plus C. This was also shown in an experiment in which subjects were asked to rate different blends of elements from separate syllables:

(4) Blending: flirz + gruns →
 a. fluns (CC/VCC blend)
 b. fruns (C/CVCC blend)
 c. flirs (CCVC/C blend)
 d. flins (CCV/CC blend)

Participants rated the CC/VCC blend (4a) the most natural, followed by the C/CVCC (4b) and the CCVC/C blend (4c). The CCV/CC blend (4d) was rated as least natural. In an experiment in which participants learned a particular blending pattern, the participants showed the most difficulty in learning the CCV/CC blend, and they learned the CC/VCC blend the most easily.

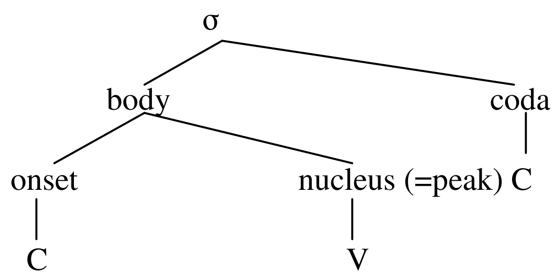
Furthermore, in the experiments with word-likeness rating task, nonword comparison task, and blending task, Treiman *et al*. (2000) found that syllables with uncommon rhymes are harder to remember or pronounce than those with common rhymes, arguing that speakers of English have an implicit knowledge of rhyme frequency.

4.1.2.2 Body: sub-syllabic grouping of C – V

As was hinted in Treiman *et al.* (2000) in which the word-likeness rating task by the speakers of English was influenced by the frequency of the CV as well as by the frequency of the VC, not all languages may have the same internal structure for a syllable: some factor, e.g. lexical frequency, may provide a source for preferring CV, rather than VC, in some languages like Korean, which is the main focus of discussion in this section.

For example, a language game in Korean involves inserting a phonological string after a V, e.g. in a syllable of CVC, not before a V, which dissects a syllable into two sub-components, body and coda: (Derwing 2007)

(5) Structural representations of syllable (σ)



In addition, Korean does not have a tradition of rhyming poetry, unlike other languages, including English. Instead, Korean poetry uses poetic meter based on a syllable count (Derwing 2007; Derwing *et al*. 1993; Yoon & Derwing 2001). The findings from many psycholinguistic experiments, including word-blending tasks, sound similarity judgments, concept formation, and list-recall tasks, revealed that Korean speakers performed significantly better at body-based tasks than at rhyme-based tasks. For example, in an experiment with a list-recall task, when the Korean speakers were asked to recall made-up names after they heard them in a list several times, they could recall better those names with identical bodies than names with identical rhymes (Derwing 2007; Yoon & Derwing 2001).

The case of Korean suggests that the onset-rhyme division of a syllable may not hold in all languages; rather, there may be at least two kinds of sub-syllabic structures, onset-rhyme division as in English vs. body-coda division as in Korean. For example, Minnan Chinese, widely spoken in Taiwan, has nasal harmony that only takes place between onset C and nucleus V, but not between nucleus V and coda C, e.g. ba, bat, ban, mã vs. *bã, *ma (Derwing 2007). That is, nasality never occurs on either onset C or nucleus V independently of each other; however, nasality can occur on either nucleus V or coda C independently of each other. This fact in Minnan indicates a possibility that prefers body-coda division, in which the elements in the body are more tightly bonded. In an experiment with a list recall task, Minnan speaking participants actually showed better performance in recalling the body-sharing items than the rhyme-sharing items.

According to Lee (2006), cross-linguistic differences in sub-syllabic constituency reflect "speakers' sensitivity to the distributional statistics of their language at the sub-syllabic level" (p. 3). For instance, Korean speakers prefer CV to VC grouping because they are implicitly aware of the closer dependency between segments in CV than in VC sequences in the Korean vocabulary. By measuring the strength of association for CV and VC in the CVC words from Korean and English, Lee found that CV is generally more strongly associated in

Korean, whereas VC is more strongly associated in English.

Basically, Lee examined how many syllables with a CVC type are actually attested as real CVC-syllable words in a corpus: he computed possible combinations of onset, nucleus, and coda in Korean, according to which there are 2646 combinations (18 onsets × 21 vowels × 7 codas).[52] The same method of calculation was done for the English CVC-type words. For Korean, there were stronger correlations between onset and vowels than between vowels and codas. That is, there were more CVs, than VCs, that occur significantly than expected by chance. For English, the opposite was the case; that is, there were more VCs, than CVs, that are attested significantly than expected by chance. This phonotactic information is argued to be reflected in such behavioral data as given above. Lee does not refute the existence of a syllable structure; rather, he motivates the existence of differing syllable structures, by looking into the lexical information on the CV and VC combinations in different languages.

Unless there is any robust evidence to the contrary, I suggest that the tight bond between onset C and nucleus V be a source for existing Vs affecting the choice of CI in the reduplication.

4.2 C – C Relationship

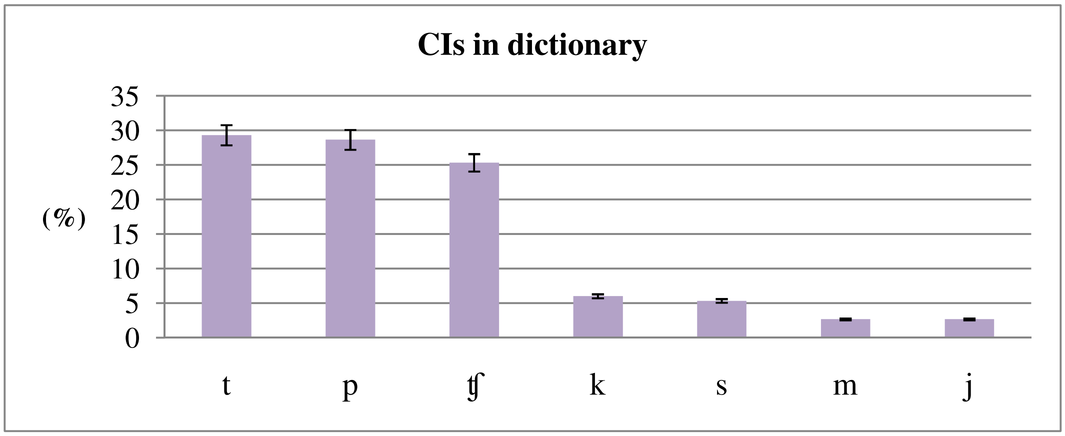
4.2.1 A preliminary question

In this section I examine the relationship between CI and its preceding C in a reduplicative form of VCVC-**CI**VCVC. According to the onset sonority constraints (cf. Chapter 3, Appendix 3-B), the least sonorous consonant is preferred, given a choice among different onsets. For example, obstruents, in a broad distinction of obstruents vs. sonorants, should be the more popular as CI that is to be inserted in the onset of a reduplicant. However, sonorants are not completely excluded as CI:

---

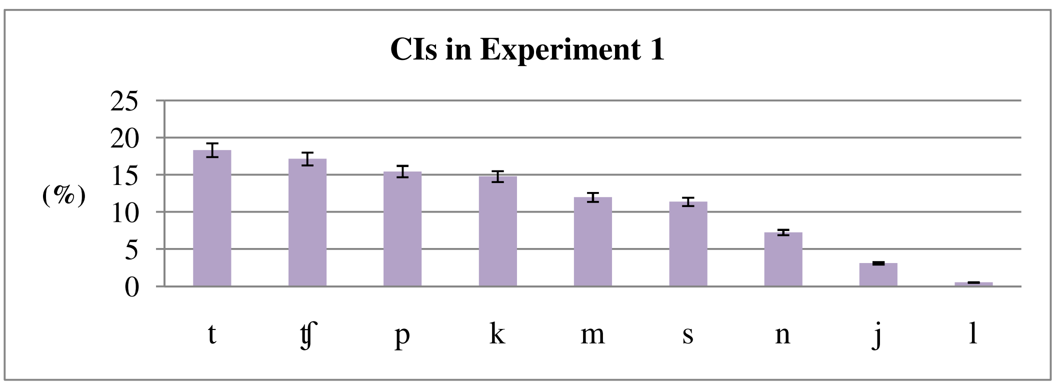
[52] According to Lee (2006), 18 onsets are /k, t, p, ʧ, s, h, l, m, n, k', t', p', ʧ', s', k^h, t^h, p^h, $ʧ^h$/; 21 vowels are /a, e, i, o, ɨ, æ, ɛ, oy, uy, wa, we, wu, wi, wæ, wɛ, ya, ye, yo, yu, yæ, yɛ/; and 7 codas are /k, t, p, l, m, n, ŋ/.

(6) Figure 4.2 CI frequency in the dictionary: Error bars represent 95% confidence interval of a mean.



Despite their relatively small occurrences, sonorants like /m, j/ are attested in the dictionary data of reduplicative words. What is more, in the word creation experiment the participants chose sonorants /m, n, j, l/ with higher frequencies than in the dictionary data.

(7) Figure 4.3 CI frequency in Experiment 1: Error bars represent 95% confidence interval of a mean.



If speakers prefer onsets of minimal sonority, why are sonorants ever chosen for insertion in the onset position?

I propose that the inserted C refers to another C that is across a syllable under the rubric of Syllable Contact Law (8).

(8) Syllable Contact Law (SYLLCON):
Rising sonority across a syllable boundary is not allowed.

The Syllable Contact Law requires that sonority should not incline upward from a coda in a syllable towards an onset of the following syllable (Baertsch & Davis 2003; Davis & Shin 1999; Gouskova 2004; Kang 2004; Vennemann 1988, among others). I will adopt the following sonority scale for the Korean case:

(9) Sonority scale (Clements 1990; Smolensky 1995):
obstruents < nasals < liquids < glides < vowels

There is a general agreement on a sonority scale (9) obstruents < nasals < liquids < glides < vowels, in which the one on the left-hand side of < is less sonorous. However, there has been controversy on the details of the sonority scale. For example, Jesperson (1904) proposed a scale, voiceless stops < voiceless fricatives < voiced stops < voiced fricatives < nasals < laterals < rhotics < glides, whereas Selkirk (1984) suggested such a scale as p, t, k < b, d, g < f, θ < v, z, ð < s < m, n < l < r.

I argue that the constraint SYLLCON allows sonorant CI in onset as long as it is not preceded by a less sonorant consonant in the coda of the preceding syllable.

4.2.2 Sonority-based account

In this section I will investigate how the relationship of CI and its preceding C can be accounted for based on the constraint for syllable contact.

4.2.2.1 Background

The constraint SYLLCON has been argued to be a crucial driving force behind many phonological processes in Korean, including nasalization (10) and lateralization (14) (Davis & Shin 1999; Kang 2002, 2005; Lee 2006; Sohn 2008, among others).

There are several cases of nasalization: an obstruent becomes nasalized before a nasal (10a-c); a lateral becomes nasalized after a non-coronal nasal (10d, e); and a non-coronal obstruent and a lateral become nasals when they occur in sequence (10f, g):

(10) a. /apʰ+nal/ → [am.nal] 'the future'
b. /natʰ+mal/ → [nan.mal] 'word'
c. /kok+mul/ → [koŋ.mul] 'grain'
d. /kjʌŋ-li/ → [kjʌŋ.ni] 'bookkeeping'
e. /tam-lon/ → [tam.non] 'discussion'
f. /tap-lje/ → [tam.nje] 'acknowledgment'
g. /kjʌk-li/ → [kjʌŋ.ni] 'isolation'

In (10a-c) an obstruent in the coda of the preceding syllable meets with a nasal in the onset of the following syllable, resulting in a sonority rise at the syllable contact, which is banned by SYLLCON. The obstruent becomes assimilated to the following nasal, resolving the disallowed sonority contour. In (10d, e) the first syllable ends in a (non-coronal) nasal and the second syllable begins in a lateral, which incurs violation of the syllable contact constraints because of the rising sonority. The nasalization of the lateral in the onset settles the unwelcome sonority rise at the contact.

In (10f, g) a (non-coronal) obstruent coda and a lateral onset come in contact at the syllable boundary, which ends up violating SYLLCON. For this sequence at the syllable contact, nasalizing both the obstruent coda and the lateral onset is the optimal solution: it is better not to have some other changes like only changing the lateral onset into a nasal (which still incurs violation of SYLLCON), lateralizing the obstruent coda (which changes the place of articulation for the non-coronal obstruent drastically), nasalizing the obstruent coda only (which still violates SYLLCON), or obstruentizing the lateral onset (which is not welcome since in general changing the sonorancy of an onset is not deemed as desirable).

The changes occurring in the consonant sequences in (10) provide evidence for the sonority contour at a syllable contact; actually, the co-occurrences of the same consonants at a syllable boundary do not undergo nasalization when they appear in the opposite order: (cf. Davis & Shin 1999)

(11) a. /mantu/ → [man.tu] 'dumpling'
b. /kʌm-kek/ → [kʌm.kek] 'swordsman'
c. /saŋ-toŋ/ → [saŋ.toŋ] 'homology'

In (11) a nasal is followed by an obstruent at a syllable boundary, which does not show rising sonority and therefore does not go through nasalization of the obstruents, which did take place in the obstruent-nasal sequence in (10a-c).

(12)	a. /milmul/	→	[mil.mul]	‘flood tide’
	b. /kalmuli/	→	[kal.mu.li]	‘finishing touches’

A lateral coda followed by a non-coronal nasal in (12), which is the opposite sequence of the same consonants appearing in (10d, e), does not undergo nasalization of the laterals since the sequence at a syllable boundary does not have a sonority rise.

(13)	a. /kalki/	→	[kal.ki]	‘mane’
	b. /kulpi/	→	[kul.pi]	‘dried corvina’

In (13) there is a sequence of a lateral and a non-coronal obstruent at a syllable contact, which is the opposite order of the same consonants shown in (10f, g). There was rising sonority in (10f, g), which was fixed by nasalization, whereas the sonority does not rise in (13), which therefore does not undergo nasalization or any other repair for sonority.

In the process of lateralization /n/ becomes a lateral when it is followed by a lateral (14a-c). Lateralization also occurs in a coronal obstruent-liquid sequence (14d).

(14)	a. /kon+lan/	→	[kol.lan]	‘difficulty’
	b. /sun+li/	→	[sul.li]	‘reasonableness’
	c. /pjʌn+lon/	→	[pjʌl.lon]	‘debate’
	d. /tikɨt liɨl/	→	[ti.kɨl.li.ɨl]	‘letter t(ㄷ) and l(ㄹ)’[53]

The opposite sequence of a nasal coda and a lateral onset in (14a-c) consists of a lateral coda and a nasal onset, which does not incur violation of SYLLCON. Thus we do not expect a change in that sequence; however, we have lateralization again for that sequence, e.g. /pul+napi/ → [pul.la.pi] ‘tiger moth,’ /k^{h}al+nal/ → [k^{h}al.lal] ‘the blade of a knife,’ /t’ʌkkal-namu/ → [t’ʌk.kal.la.mu] ‘oak,’ which is actually due to some other constraint involved in the process. I do not go into detail of this process since it is not directly related to the issue at hand, regarding sonority contour across a syllable boundary.

The sequence of a coronal obstruent and a lateral (14d) can occur in the opposite order:

[53] There are rare cases of a coronal and a liquid coming together across a morpheme boundary. The example (16d) is traditionally presented as a case of lateralization of a /tl/ sequence (Davis & Shin 1999: 308).

(15) a. /mul+toŋi/ → [mul.t’oŋ.i] ‘water jar’
b. /liɨl tikɨt/ → [li.ɨl.ti.kɨt] ‘letter l(ㄹ) and t(ㄷ)’

The sequence of a lateral and a coronal obstruent at a syllable boundary does not have rise in sonority, and it does not have to employ any repair like lateralization.

Such repair strategies as nasalization and lateralization have been argued to apply for the sake of a better sonority profile across syllables, in which the original sonority was rising in violation of the sonority configuration requirement that sonority should drop across a syllable. The strategies fixed the undesirable situation of sonority rise across a syllable boundary, by making the resulting sounds as sonorous as each other.

The repair strategies used in Korean are not peculiar; many languages use some similar strategies to repair a bad sonority contour, e.g. Kazakh, Kirghiz, Faroese, Icelandic, Sidamo, to name a few (Gouskova 2004). For example, Kazakh desonorizes a consonant in onset when it comes in contact with a consonant in coda which happens to be less or equally sonorous:

(16) Kazakh desonorization (Davis 1998; Gouskova 2004)
a. /kol-lar/ kol.dar ‘hands’ cf. al.ma.lar ‘apples’
b. /murin-ma/ mu.rin.ba ‘nose-INT’ cf. kol.ma ‘hand-INT’
c. /koŋɯz-ma/ ko.ŋɯz.ba ‘bug-INT’ cf. ki.jar.ma ‘cucumber-INT’

Kazakh has a requirement that sonority fall, and other languages have different requirements for the sonority slope at a syllable boundary. For instance, Faroese makes use of syllabification to repair a bad sonority configuration; that is, when sonority rises too much, the two consonants are syllabified into a complex onset, whereas when sonority does not rise, the consonant sequence becomes heterosyllabic.

(17) Faroese syllabification (Gouskova 2004)
a. *Sonority rising*: tautosyllabification
a:.^{h}kvamarın ‘beryl’
ai:.^{h}trantı ‘poisonous’
e:.^{h}plı ‘potato’
b. *Sonority non-rising*: heterosyllabification
rɔ^{h}k.tı ‘smoked (SG)’
vɛs.tʊr ‘west’
ɟœr.d̥ɪ ‘did (SG)’

With the Syllable Contact Law prevailing in Korean, as well as in many other languages, I investigate the relationship between CI and its preceding C in another syllable on the basis of the constraint, SYLLCON, and I argue that the C – CI relationship presents a good case of observing this constraint.

4.2.2.2 Syllable Contact Law in consonant insertion

The constraint SYLLCON was respected in 98.28% of the dictionary data, which contained 58 words of the form V.CVC-CV.CVC. In the word creation experiments, the experimental results show that SYLLCON was respected in 97.67% of the responses in Experiment 1 (N=55, 817 tokens); and in 95.63% of the responses in Experiment 2 (N=15, 1672 tokens).[54]

Whether SYLLCON is respected in the data was computed on the basis of the distinction on the sonority scale (9), presented above: obstruents < nasals < liquids < glides < vowels. In Experiments 1 and 2 the participants were asked to freely choose a consonant for CI, whereas in Experiments 3 and 4 the participants chose a CI from a given set of Cs, /t, p, ʧ/ in Experiment 3 and /t, ʧ/ in Experiment 4. Therefore, it is pointless to examine the observance of SYLLCON for the responses from Experiment 3 and 4 since the reduplicative words created by the participants only contained obstruent CIs, which will automatically result in a good sonority profile between CI and its preceding C all the time.

In Experiments 1 and 2, as well as the dictionary data, there were still some cases in which SYLLCON is not observed: 1 case in the dictionary data, 19 cases in Experiment 1, and 73 cases in Experiment 2. These cases include CI = /j, l, m, n/, and all these attested cases in the dictionary and Experiments 1 – 2 are given in Appendix 4-A. I present some of the examples in the following:

[54] When there was more than one response per item in Experiment 2, I also included them in my count: therefore, the number in total is more than expected (1665 tokens = 15 × 111 stimuli).

(18)	a. otok-	→	/otok-**j**otok/	[o.toŋ.njo.tok]
	b. otʃak-	→	/otʃak-**j**otʃak/	[o.tʃaŋ.njo.tʃak]
	c. asik-	→	/asik-**m**asik/	[a.siŋ.ma.sik]
	d. utup-	→	/utup-**m**utup/	[u.tum.mu.tup]
	e. atap-	→	/atap-**n**atap/	[a.tam.na.tap]
	f. umuk-	→	/umuk-**n**umuk/	[u.muŋ.nu.muk]
	g. okam-	→	/okam-**l**okam/	[o.kam.no.kam]
	h. amaŋ-	→	/amaŋ-**l**amaŋ/	[a.maŋ.na.maŋ]
	i. imit-	→	/imit-**l**imit/	[i.mil.li.mit]

Korean has various repair strategies (shaded in the examples) which fix up the instances that do not abide by the Syllable Contact Law, e.g. n-epenthesis (18a, b); obstruent nasalization (18c-f); nasalization of a lateral (18g, h); and lateralization of a coronal obstruent (18i). I do not definitely argue that the cases which appear to violate SYLLCON should be in fact rescued by virtue of the repair strategies, which consequently turn all the SYLLCON-violating cases into SYLLCON-obeying ones. Rather, I remark that these repairs might take place in the speakers' mind since they were asked to pronounce the words when creating new reduplicative forms. I do not argue for anything concerning the SYLLCON-violating instances in the written dictionary data; however, there was only one case of SYLLCON violation in the reduplicative words from the dictionary.

4.2.2.3 SYLLCON on general vs. specific vocabulary

As has been seen in the preceding sections, the Syllable Contact Law is robustly observed both in the general lexicon and the specific set of vocabulary (newly created reduplicative words).

However, there are some remarkable differences between the reduplicated vocabulary and the general vocabulary with respect to the syllable contact requirement: first, reduplicants differ from the rest of the Korean lexicon in their resistance to glide insertion. Glides are cross-linguistically common as an inserted C, e.g. altruizmus [al.tru.**j**iz.mus] 'altruism' in colloquial Slovak (Rubach 2000), kea [ke.**j**a] 'swim' in Lou (Blevins 2008), kokain [ko.ka.**j**in] 'cocaine' in Czech (Zaleska 2008). Glides, /j/ in particular, have a robust status as an inserted C in Korean as well (Kang 1999):

(19) a. /aka+a/ → [a.ka.**j**a] 'baby + vocative'
b. /mintʃu + a/ → [min.tʃu.**j**a] 'Minjoo (proper name) + vocative'
c. /hakkjo + e/ → [hak.kjo.e] ~ [hak.kjo.**j**e] 'school + locative'
d. /namu+e/ → [na.mu.e] ~ [na.mu.**j**e] 'tree + locative'

However, glides were almost never inserted in the reduplicants: only 1.72% in the dictionary data, 1.96% in the experimental responses from Experiment 1, and 0% of the responses in Experiments 2. Therefore, glides are most frequently chosen in Korean as an inserted C in the context of V._V, but they are disfavored in the context of C._V. This is an effect of SYLLCON: glides are the most sonorous Cs, and therefore disfavored in the onset position following coda C.

Furthermore, the reduplicated forms appear to obey an even stronger version of SYLLCON than the lexical entries of the language in general. For example, among the most frequent occurrences for [k.C] combinations in the *Sejong* 2007 corpus, [k.h] occurs in 18.09% of the forms and [k.s] in 16.87%. In contrast, these combinations were extremely rarely produced in the word formation task ([k.h] in 0% and [k.s] in 12.73% in Experiment 1; [k.h] in 1.11% and [k.s] in 7.41% in Experiment 2). If we assume a broad scale of sonority like obstruents < nasals < liquids < glides < vowels (Clements 1990; Smolensky 1995), then this finding may not be interesting; however, if we adopt a finer scale like voiceless stops < voiceless fricatives < voiced stops < voiced fricatives < nasals < laterals < rhotics < glides (Jesperson 1904), this finding in the reduplication data is not trivial.

In addition, [k.t], which obeys SYLLCON, is attested in 4.11% of the corpus data, but in 24.24% of the responses in Experiment 1 and 27.41% of the responses in Experiment 2.

4.3 Summary

This chapter discussed the local relationships between CI and adjacent segments, i.e. between CI and a following V and between CI and a preceding C, in the reduplicative forms VCVC-**CI**VCVC. Significant correlations were found between the frequency of CV combinations in the word creation experiment and the frequency of CV combinations in the reduplication-only corpus (Chapter 2). Korean speakers' behavior in many psycholinguistic experiments suggested that a CV (body) constituent is prominent for Korean speakers, as opposed to the speakers of English-like languages which evidently have a closer tie between V and C (rhyme).

It was also found that there is a certain relation between the CI and its preceding C in the base of reduplicated forms, VCVC-**CI**VCVC. The relationship of C-**C** can be defined by the constraint that defines acceptable syllable contacts, which has been asserted in the literature as one of the essential constraints in Korean. CIs in the reduplication were in general chosen to avoid rising sonority across a syllable boundary. Furthermore, it is noteworthy that the reduplication data follow a stronger version of the Syllable Contact Law than the general lexical entries.

Appendix 4-A SyllCon-violating cases

(1) Table 4.1 Dictionary (58 words): SyllCon-violating combinations

C.-CI at syllable boundary	Number of instances (1)	Examples
ŋ-j	1	illʌŋ-**j**allaŋ

(2) Table 4.2 Experiment 1 (817 tokens): SyllCon-violating combinations

C.-CI at syllable boundary	Number of instances (19)	Examples
l-j	1	usul-**j**usul
	4	atal-**j**atal
m-j	1	osam-**j**osam
	5	apam-**j**apam
ŋ-j	1	atʃaŋ-**j**atʃaŋ
k-j	2	otok-**j**otok
	1	otʃak-**j**otʃak
	1	asik-**j**asik
m-l	1	okam-**l**okam
k-n	2	asik-**n**asik

(3) Table 4.3 Experiment 2 (1672 tokens): SyllCon-violating combinations

C.-CI at syllable boundary	Number of instances (73)	Examples
ŋ-l	1	amaŋ-**l**amaŋ
n-l	1	akan-**l**akan
t-l	1	imit-**l**imit
t-m	1	itit-**m**itit
	1	ipit-**m**ipit
	1	init-**m**init
t-n	2	unut-**n**unut
	1	iŋit-**n**iŋit
	2	asat-**n**asat
	1	aŋat-**n**aŋat
	1	usut-**n**usut
	2	ipit-**n**ipit
	3	isit-**n**isit
	1	imit-**n**imit
	2	anat-**n**anat

	1	utut-**n**utut
	1	ikit-**n**ikit
	1	itit-**n**itit
	1	atat-**n**atat
	2	apat-**n**apat
p-m	2	atap-**m**atap
	1	imip-**m**imip
	1	ikip-**m**ikip
	1	utup-**m**utup
	1	anap-**m**anap
	1	asap-**m**asap
	1	itip-**m**itip
p-n	3	atap-**n**atap
	1	inip-**n**inip
	2	ikip-**n**ikip
	1	unup-**n**unup
	2	utup-**n**utup
	2	asap-**n**asap
	1	uŋup-**n**uŋup
	1	amap-**n**amap
	1	imip-**n**imip
	1	aŋap-**n**aŋap
	1	isip-**n**isip
	1	itip-**n**itip
	1	akap-**n**akap
k-m	1	imik-**m**imik
	1	amak-**m**amak
	1	upuk-**m**upuk
	1	uŋuk-**m**uŋuk
	1	iŋik-**m**iŋik
	1	usuk-**m**usuk
	1	anak-**m**anak
k-n	2	iŋik-**n**iŋik
	1	unuk-**n**unuk
	1	uŋuk-**n**uŋuk
	1	asak-**n**asak
	1	umuk-**n**umuk
	1	upuk-**n**upuk

	1 1 2 1 1	anak-**n**anak ipik-**n**ipik itik-**n**itik utuk-**n**utuk atak-**n**atak

Chapter 5
Identity Avoidance in Consonant Insertion

We saw in preceding chapters that context plays a role in choice of an inserted consonant. Certain CV sequences were more frequent both in the word creation experiment and in the online reduplication-only corpus, and the preference for particular CV sequences was attributed to speakers' knowledge of the cohesive bond between onset C and nucleus V in Korean. The choice of inserted C was also affected by an adjacent C: CI and the preceding C respected the sonority requirement on syllable contacts. Furthermore, some speakers were sensitive to nonadjacent Cs, failing to insert their preferred CI when they encountered the same C in the base.

In this chapter I will further investigate the relationships among Cs in reduplicative forms. I will argue that there is a general tendency for CI to be non-identical to other Cs in the reduplicated form. This tendency to avoid identity among segments in a morphological constituent is attested in various phenomena of other languages. I will present evidence that this co-occurrence restriction is sensitive to distance between the participating consonants.

This chapter covers the following: background concerning the consonant co-occurrence restrictions in many languages; description of the co-occurrence restrictions, i.e. identity avoidance in the Korean vocabulary; and explanation for the identity avoidance in reduplication, based on the measurement of how often consonants that share place and/or manner co-occur.

5.1 Background

Various languages show evidence of a tendency to avoid multiple occurrences of identical or similar segments within a word or morpheme (Berkeley 2000; Clements & Keyser 1983; Coetzee & Pater 2006, 2008; Frisch *et al*. 2004; Goldsmith 1976; Kawahara *et al*. 2006; Leben 1973; McCarthy 1979, 1981, 1986; Mester 1986; Padgett 1995; Steriade 1982; Walter 2007; Yip 1988, 1989, among many others).[55] For example, in Arabic verbal roots, combinations of homorganic consonants in proximity are disfavored. Arabic uses a system of 3-consonant roots, and any pair of root consonants tends not to share the same place of articulation (Greenberg 1950; Coetzee & Pater 2008; Frisch *et al*. 2004; McCarthy 1986, 1988, 1994; Pierrehumbert 1993). For example, verb forms like (2a, b) containing homorganic consonants are extremely rare.[56]

(1) a. /**k**a**t**a**m**a/ 'to conceal; to hide'
    b. /**b**a**h**a**t**a/ 'to be baffled'

(2) a. */**b**a**m**a**h**a/
    b. */**h**a**d**a**t**a/

Furthermore, there is evidence that Arabic speakers are implicitly aware of this avoidance of repetition or similarity. Frisch and Zawaydeh (2001) conducted an experiment using nonce verb forms some of which violated OCP-Place, an identity-avoidance constraint which bans consonants sharing a place of articulation. They found that the forms with OCP-Place violations (3a) were judged by the native speakers of Jordanian Arabic to be significantly less wordlike than the forms without violations (3b).

(3) a. /**t**a**s**a**b**a/
    b. /**t**a**h**a**f**a/

Acceptability ratings varied according to the similarity and the proximity of consonant pairs within the stimuli, with the worst rating for forms like (4a):

(4) a. /**b**a**b**aθa/ (identical)
    b. /θa**b**a**m**a/ (similar adjacent)
    c. /**b**aʃa**f**a/ (similar nonadjacent)
    d. /**b**aʔa**d**a/ (nonhomorganic)

[55] Constraints designed to enforce identity avoidance are often designated as OCP (Obligatory Contour Principle) constraints, based on constraints banning identical adjacent tones.

[56] In fact, the restrictions are more complicated than meets the eye. In the verbal root of $C_1VC_2VC_3$, the cases of C_1 and C_3 or C_2 and C_3 being identical in place, e.g. **l**abe**s**a 'wear,' **t**aha**n**a 'grind,' ka**s**a**d**a 'mean,' ha**s**a**l**a 'obtain,' are more tolerated than the cases in which C_1 and C_2 share the same place (cf. Frisch *et al*. 2004; McCarthy 1979, among others). I set aside this detailed distinction in my discussion.

Similarly, Berent and her colleagues found that Hebrew speakers' linguistic performance was affected by OCP violations (Berent 2002; Berent *et al*. 2001, 2004, *et seq*.).

Parallel patterns can be observed in diverse non-Semitic languages (Berkeley 2000; Frisch *et al*. 2004). In a well-formedness judgment experiment by the native speakers of English, Hay, Pierrehumbert & Beckman (2004) found that nonce forms like /stɹimsi/, which contains two stridents, were rated lower than similar forms like /stɹimpi/, even though words containing two strident obstruents are certainly allowed in English as in 'space' /speɪs/ or 'starch' /stɑɹʧ/. English is also known to have restrictions against co-occurrence of homorganic identical consonants within a word; that is, in a word with a form of sCVC in which the Cs are stops, two /k/'s or two /p/'s are not allowed, e.g. **skeak*, **speap*. However, two /t/'s are allowed, e.g. *state*, *stoat*, *stet* (Coetzee 2005, 2008b; Davis 1991; Fudge 1969). Coetzee (2005) found that listeners in a perception task tended to perceive an ambiguous percept as an OCP-obeying item. This kind of perceptual bias was also found for the forms with two coronal stops, although words with two /t/'s actually exist in the lexicon.

The vocabulary of Dutch was also found to observe OCP-Place; for example, words with two identical coronals or two identical labials are under-represented. There is a further distinction in this pattern; that is, words with non-identical coronals like *steen* 'stone' are just slightly under-represented, whereas words with non-identical labials like *spam* 'spam' are extremely under-represented (Shatzman & Kager 2007). In a lexical decision task, Shatzman & Kager found that listeners rejected nonwords that violated abstract constraints (e.g. OCP-Place) more quickly than they rejected nonwords that obeyed the constraints. They also found that the abstract phonotactic constraints influence speech perception, independently of lexical factors like similarity of the nonwords to existing words in the lexicon.

Another area where OCP effects may be found is in language games. The Cantonese language game "La-Mi" (Yip 1995) typically changes the first consonant of the input into /l/ and the second vowel into /i/ as in (5a). However, when the input already has /l/ or /i/, a different consonant and vowel is used, as in (5b).

(5)	Cantonese La-Mi:		CVC > lVC CiC		
	a.	kɔːŋ	lɔːŋ kiŋ		
		sat	lat sit		
	b.	t'in	lin t'un	*lin t'in	i>u
		lat	k'at lit	*lat lit	l>k'
		lin	k'in lun	*lin lin	l>k' and i>u

Identity avoidance also plays a role in reduplication in various languages. In Javanese Habitual-Repetitive Reduplication (Yip 1995), the final syllable of the first half in the reduplicative form must have an /a/ nucleus, as in (6a), but if the input already contains /a/ as the final syllable of the base, the vowel in the final syllable of the reduplicant will have a different vowel, as in (6b).

(6)	Javanese Habitual-Repetitive Reduplication				
	a.	eliŋ	elaŋ-eliŋ	'remember'	
		tuku	tuka-tuku	'buy'	
		eleʔ	elaʔ-eleʔ	'bad'	
	b.	udan	udan-uden	'rain'	*udan-udan
		kumat	kumat-kumet	'have a relapse'	*kumat-kumat
		edan	edan-eden	'crazy'	*edan-edan

Turkish emphatic adjectives (Demircan 1987; Dobrovolsky 1987; Wedel 1999, 2000; Yip 1995; Yu 1999) are created by prefixing a CVC syllable where the initial CV is identical to the word-initial CV of the base, and the inserted final C comes from the set /p, s, m, r/. This inserted consonant of the reduplicant tends to be non-identical to the first two consonants of the base.

(7)	Turkish Emphatic Reduplication				
	a.	kara	'dark'	ka**p**kara	'pitch black'
	b.	belli	'clear'	be**s**belli	'obvious'
	c.	bejaz	'white'	be**m**bejaz	'bright white'
	d.	temiz	'clean'	te**r**temiz	'spotless'

Wedel (1999, 2000) suggests that out of the candidate Cs /p, s, m, r/ for the inserted consonant, labial stop /p/ is preferred unless either of the first two consonants in the base is labial. He argues on the basis of the research on the online Turkish language lexicon, TELL[57] that the epenthetic C in the Turkish reduplication does not repeat the same segment as one of the base Cs. He also found that when Turkish speakers were asked to create a reduplicated version of adjectives that normally are not subject to emphatic reduplication in Turkish, they avoided inserting a consonant identical to a base consonant. For example, bodur 'squat' → bo**s**bodur (for all participants , mest 'enchanted' → me**p**mest (for all participants), pinti 'very stingy' → pi**m**pinti ~ pi**s**pinti. What is interesting is that the suggested identity avoidance constraint was more strictly observed in the elicited data than in the attested data.

All these data from diverse languages exhibit identity avoidance effects which in turn seem to be an emergence of the unmarked phenomenon.[58] The tendency to avoid identical consonants within a morphological constituent is similar to the patterns that emerged in the case of the inserted Cs in the Korean reduplication. For example, some speakers, who normally preferred to insert /t/ in the reduplicant, avoided epenthesizing /t/ when /t/ was present in the base. In the following section I examine the role of identity avoidance among CI and base Cs in Korean reduplication.

5.2 Identity Avoidance in Korean Reduplication

In this section I examine the effect of a principle of identity avoidance on the choice of inserted consonant in Korean reduplication. I examine data from an existing lexicon (dictionary: *Essence Korean Dictionary*) and an experiment (Experiment 3: word creation task). The methods used in surveying the dictionary and conducting the experiment are those presented in Chapter 1.

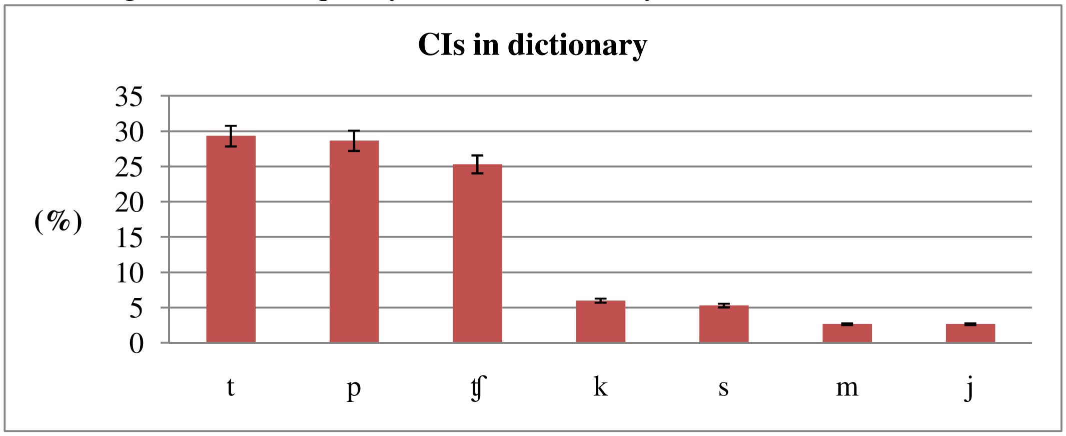
Below I repeat the chart showing the frequency of inserted consonants in reduplicants from the Korean dictionary (reference).

---

[57] http://socrates.berkeley.edu:7037/cgi-bin/TELLsearch.cgi

[58] The emergence of the unmarked follows from the crucial notions of constraint ranking and violation under domination in the Optimality Theory, according to which even dominated constraints can be visibly active, becoming dominant, in such a form as a reduplicant whereas they are dominated by faithfulness constraints in a base. A constraint like Identity Avoidance, if any, must be dominated in the base but becomes undominated in the reduplicant in the presented examples.

(8) Figure 5.1 CI frequency from the dictionary



As the chart indicates, certain consonants were preferred as CIs. However, as discussed in Chapter 3, some speakers tended to avoid inserting their preferred C when encountering the same C in the existing context. We can postulate a hypothesis with respect to identity avoidance:

(9) **Hypothesis 1** (Identity Avoidance Principle)
CIs will differ from their neighboring Cs.

To test this hypothesis, we need a way to measure the identity or difference between CI and its neighboring Cs (e.g. at a featural level), and I first look at restrictions on consonant co-occurrences, if any, in the general vocabulary of Korean.

5.2.1 Background: General vocabulary

As in many other languages, Korean shows a statistical underrepresentation of words containing pairs of consonants that have the same place of articulation. Following the example of previous research on an OCP-Place constraint (Coetzee & Pater 2006; Frisch *et al*. 2004; Greenberg 1950; Kawahara *et al*. 2006; McCarthy 1988, 1994; Pierrehumbert 1993), Ito (2006) categorized the consonants in Korean into co-occurrence classes:

(10) a. Labial = {p, p^h, p', m, P}
b. Coronal Obstruents = {ʧ, $ʧ^h$, ʧ', t, t^h, t', s, s', T, S}
c. Coronal Sonorants = {l, n}
d. Dorsals = {k, k^h, k', ŋ}

Ito examined monosyllabic stems (1420 stems = 664 verbs + 756 nouns), distinguishing onset and coda consonants. In the classes in (10) the capitals P, T, S stand for lenis codas.

Based on the Observed/Expected (O/E) ratio, Ito confirmed that Korean shows effects of an OCP-Place constraint in the general lexicon. The Expected value is the value to be expected when any two consonants can freely combine. A value of O/E greater than 1 indicates that there is no co-occurrence restriction. A value of O/E less than 1 indicates that consonant combinations are underrepresented. The O/E values are given in parentheses in the following table, and the shaded cells indicate the highly underrepresented pairs,

(11) Table 5.1 Co-occurrence restriction (Ito 2006: 11)[59]

Onset \ Coda	Labial	Cor obs	Cor son	Dorsal	Total
Labial	34 (0.66)	62 (1.31)	69 (1.02)	46 (1.04)	211
Cor obs	98. (1.11)	59 (0.73)	106 (0.92)	96 (1.28)	359
Cor son	13 (0.95)	15 (1.19)	22 (1.22)	6 (0.51)	56
Dorsal	46 (1.22)	39 (1.13)	54 (1.09)	15 (0.47)	154
Total	191	175	251	163	780

The cases which do not have an onset C were excluded, and those which have /h/ in onset were also excluded from the calculation since /h/ is very restricted in distribution. In general, co-occurrences of homorganic consonants tend to be underrepresented (χ^2 = 40.39, $p < 0.001$). The combinations of Coronal sonorant onsets and Dorsal codas are underrepresented (O/E = 0.51); however, according to Ito, this is only an accidental gap, due to the fact that /n/ is the only possible coronal sonorant onset. In contrast, the combination of Dorsal onsets and Coronal sonorant codas is overrepresented (O/E = 1.09).

Note that the co-occurrence restrictions in Korean seem to be different from those of other languages in some points. First, Ito remarked that pairs of identical consonants are unrestricted in co-occurrence. This pattern is seen in

[59] Cor obs stands for coronal obstruents, and Cor son stands for coronal sonorants.

other languages that are argued to have consonant co-occurrence restrictions. For example, Kawahara *et al*. (2006) observed in the examination of 4011 roots in Yamato Japanese that pairs of [p], [ɸ], [w], [t], [ʦ], [s], [n], [ʃ], [ʧ], [ç], [j], and [h] all freely occur. Suggesting that total identity may provide an escape hatch from the OCP. Kawahara *et al*. did not distinguish the syllabic position of consonants in the computation of the co-occurrences.

We found that the Korean vocabulary (specifically, nominal/verbal stems, based on Ito's study) is generally restricted in consonant co-occurrences. In the subsequent section, I will examine the CI-reduplication (both from the dictionary and from the word creation experiment, Experiment 3 in particular) in terms of OCP-Manner, as well as OCP-Place. As was noted by Greenberg (1950) and McCarthy (1988, 1994) and discussed in Frisch *et al*. (2004), manner of articulation cannot be excluded from the factors that affect the co-occurrence of consonants; for example, they noted that among the coronal obstruents there are far more roots with one fricative and one stop than roots with two fricatives or two stops. Thus I include manner description, not just place, in my investigation.

5.2.2 Preliminary examination of reduplication data

The identity between the CIs /t, p, ʧ/, the most frequent CIs, and their context consonants was evaluated in terms of the place and manner of articulation. In the study of the dictionary, I considered words with the VC_1VC_2-bases because they were attested most frequently in the CI-reduplication, which amount to 40.8 % (51 items out of 125 V-initial bases with CI classes /t, p, ʧ/ in the dictionary).[60]

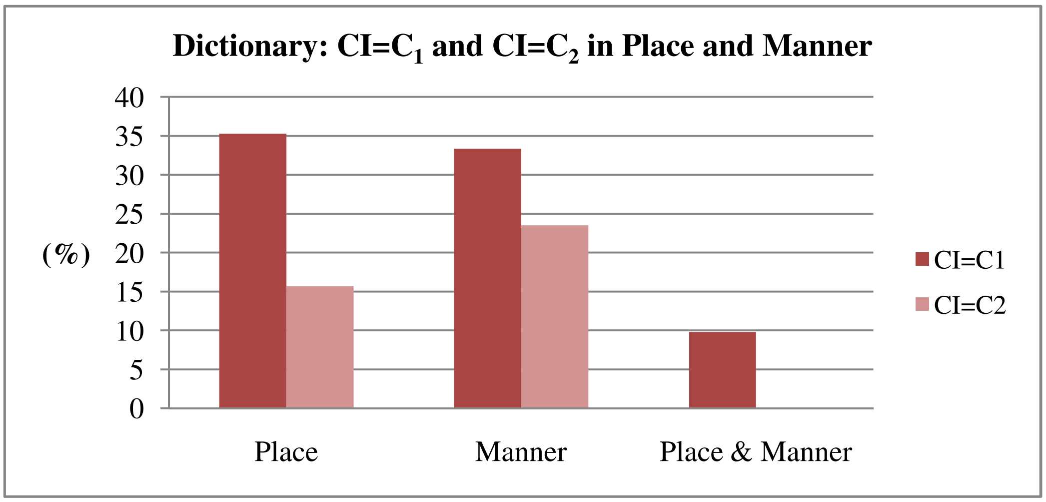
In what follows, C_1 and C_2 stand for the first and second consonant in the base, respectively, which are copied in the reduplicant. $CI=C_1$ and $CI=C_2$ indicate that CI shares a feature (place and/or manner) with C_1 and C_2, respectively. The frequency is presented in percentage (%).

[60] Among the rest of the database (a total of 125 words), 3C-bases amount to 40% (50 items), 2C-bases with non-VCVC amount to 8.8% (11 items), 1C-bases amount to 8% (10 items), and 4C-bases occupy 2.4% (3 items). It may be worth looking at the data of 3C-bases in which I can also see if there is any distance effect between the CI and the base consonants.

(12) Table 5.2 VC_1VC_2-**CI**VC_1VC_2, CI=/t, p, ʧ/ from the dictionary

Dictionary	**CI=C_1 (%)**	**CI=C_2 (%)**
Place	18/51=35.29	8/51=15.69
Manner	17/51=33.33	12/51=23.53
Place & Manner	5/51=9.80	0/51=0

(13) Figure 5.2 Identity: CI=C_1 and CI=C_2 in the dictionary data

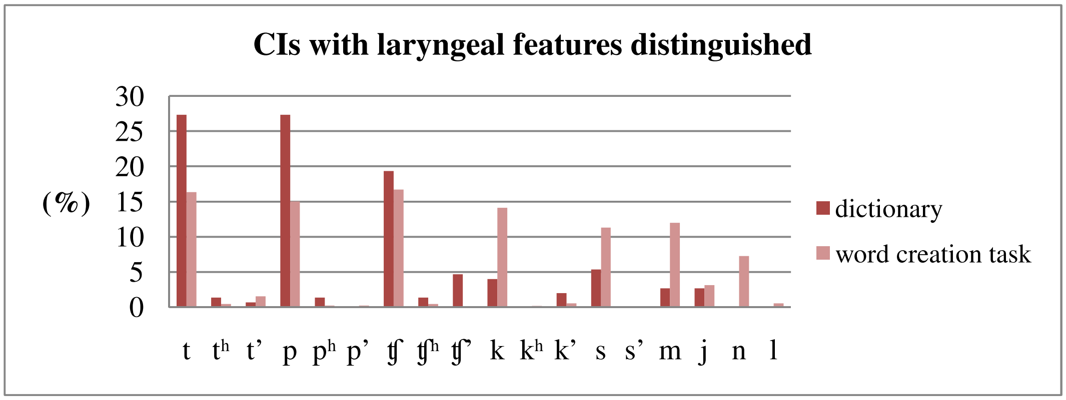


The only instance of two identical sounds in a word occurs in *etoŋ*-***tetoŋ*** 'childlike'. There are five pairs of CI and C_1 which have the same place and manner (9.80%).[61] Four items had /**t**t^h/ or /$\mathbf{t}^h$t/ (*ot^hol*-***tot^hol*** 'rugged,' *ut^hul*-***tut^hul*** 'rough,' *otol*-***t^hotol*** 'uneven,' *utul*-***t^hutul*** 'bumpy'), which are dissimilar in laryngeal properties. Although laryngeal features may be taken into account in computing feature similarity (cf. Frisch *et al*. 2004), I do not consider these

[61] Perceptual salience of the beginning of a word, according to Marie Huffman (p.c.), may explain the tendency in Table 5.2 (and Figure 5.2), in which CI=C_2 (%) is lower than CI=C_1 (%). It might be one of the factors at work, which may explain why the reduplicant-initial CI differs more from C_2 than from C_1 in VC_1VC_2-**C**VC_1VC_2. Also as was pointed out by Andries Coetzee (p.c.), it may be distance that matters; that is, CI is adjacent to C_2 but it is not adjacent to C_1, in which case CI tends to be more different from C_2 (adjacent C) than from C_1 (nonadjacent C). This issue was discussed in Chapter 4 from another perspective, with regard to the C – C relationship. It may be also due to total identity that serves as an escape hatch (with some exceptions) from the identity avoidance effects (cf. Kawahara *et al*. 2006): there are more cases of CI being identical to C_1 (both CI and C_1 in onset) than to C_2 (CI in onset and C_2 in coda) in the form of $V.C_1VC_2$-**C**$V.C_1VC_2$, which is reminiscent of the Yamato Japanese case in which onset Cs were considered in general.

features separately in my analysis. Figure 5.3 shows that the number of inserted consonants with aspiration or tensification is small enough that pooling (for example) /t, t^h, t'/ under /t/ would not give significantly different results from breaking them into distinct classes according to different laryngeal specifications. Note that in the results from the word creation experiment (Experiment 1) as well, aspirated and fortis consonants are rarely attested.

(14) Figure 5.3 CI frequencies from the dictionary and the word creation experiment (Experiment 1), including the laryngeal segments



To see whether identity avoidance really works, I will measure it in the following section. I analyze the dictionary data and the responses from an experiment (Experiment 3), employing the O/E ratio.

5.2.3 Results

I limit my focus to 2C-bases consisting of VC_1VC_2, in order to investigate the exhaustive contextual effect for the choice of CIs, in my analysis. The identity avoidance is measured on the basis of place and manner of articulation for consonants: Place = {Labial, Alveolar, Palatal, Velar}, Manner = {Stop, Affricate, Nasal, Fricative, Approximant}.

The consonants in Korean can be classified as follows, based on place and manner:

(15) Identity classes by Place[62]

a. Labial = {p, p^h, p’, m, w}

b. Alveolar = {t, t^h, t’, n, s, s’, l}

c. Palatal = {ʧ, $ʧ^h$, ʧ’, j}

d. Velar = {k, k^h, k’, ŋ}

(16) Identity classes by Manner

a. Stop = {p, p^h, p’, t, t^h, t’, k, k^h, k’}

b. Nasal = {m, n, ŋ}

c. Affricate = {ʧ, $ʧ^h$, ʧ’}

d. Fricative = {s, s’}

e. Approximant = {l, w, j}

I use these identity classes when testing how significant the identity avoidance is, in terms of place and manner. I could have used more detailed classes with more feature specifications. However, I considered place and manner only for my analysis of the responses from Experiment 3 because the variables I used in the stimuli in Experiment 3 did not have any further distinction; that is, they varied in place (Labial, Coronal, Velar) and in place (Stop, Fricative, Nasal). The candidate CIs were /t, p, ʧ/, from which participants were asked to choose one.

Following other analyses on co-occurrence restrictions, I employ the O/E ratio for the statistical test. Based on the observed numbers (O), I calculated the expected values (E) as the probability that CI occurs, e.g. in a combination of **CI**VC_1VC_2, multiplied by the probability that C_1 (or C_2) occurs, multiplied by the total number of tokens (51 tokens of VCVC-CIVCVC reduplications in the dictionary). From these O and E values the O/E was calculated.

(17) Table 5.3 Observed numbers in the dictionary data

C_1 / CI	Labial	Alveolar	Palatal	Velar	Total
p	0	11	5	5	21
t	0	18	0	0	18
ʧ	5	3	0	4	12
Total	5	32	5	9	51

For instance, the probability of [p] occurring as CI in Table 5.3 is 21/51 = .41.

62 The term, identity classes, is accredited to Yip (1989).

The probability of an alveolar consonant occurring as C_1 is 32/51 = .63. If these two events are independent and their probabilities do not influence each other, then the probability of both events occurring is .41 × .63 = .258. There are 51 pairs in total, and we expect 13.2 pairs (= .258 × 51) that have [p] as CI and an alveolar C as C_1. This is the expected number of the {p, alveolar} pair, and the other expected numbers were computed in the same way, Table 5.4.

(18) Table 5.4 Expected numbers for the dictionary data

C_1 / CI	Labial	Alveolar	Palatal	Velar	Total
p	2.1	13.2	2.1	3.8	21
t	1.8	11.2	1.8	3.2	18
tʃ	1.2	7.7	1.2	2.2	12
Total	5	32	5	9	51

O/E values are calculated by dividing each observed number by the corresponding expected number. If O/E values are smaller than 1, then the consonant pairs are underrepresented; and if O/E values are greater than 1, then the pairs are overrepresented.

5.2.3.1 Dictionary data

I first consider whether pairs of CI – C_1 in reduplicated forms from the dictionary show significant restrictions on their co-occurrence. The following table present O values, O/E ratio, and χ^2 values (indicating whether the O/E ratio differs significantly from 1) for the dictionary reduplicants. Significant underrepresentation, which suggests identity avoidance, is indicated by shading in the cells.

(19) Table 5.5 CI – C_1 pairs: **Place Identity**
(If $\chi^2 > 3.84$, $p < .05$. N = 51)

CI \ C_1	Labial	Alveolar	Palatal	Velar
p	O = 0 O/E = 0 $\chi^2 = 3.81$	O = 11 O/E = 0.83 $\chi^2 = 1.61$	O = 5 O/E = 2.38 $\chi^2 = 7.77$	O = 5 O/E = 1.32 $\chi^2 = 0.9$
t	O = 0 O/E = 0 $\chi^2 = 4.69$	O = 18 O/E = 1.61 $\chi^2 = 16.46$	O = 0 O/E = 0 $\chi^2 = 3.04$	O = 0 O/E = 0 $\chi^2 = 5.84$
ʧ	O = 5 O/E = 4.17 $\chi^2 = 17.83$	O = 3 O/E = 0.39 $\chi^2 = 9.51$	O = 0 O/E = 0 $\chi^2 = 1.63$	O = 4 O/E = 1.82 $\chi^2 = 2.58$

Since I do not consider /h/ in my analysis due to its distributional peculiarity, I have four places of articulation for C_1, labial, alveolar, palatal, and velar. What I examine in Table 5.5 is whether there are significant co-occurrence restrictions against CIs /p, t, ʧ/ and context Cs (C_1, which is onset C) in terms of place.

Assuming a tendency toward identity avoidance, I predict that the pairs {p, labial}, {t, alveolar}, and {ʧ, palatal} will be underrepresented significantly since each of the pairs shares place of articulation. The pair {p, labial} appears underrepresented at a marginally significant level. The pair {ʧ, palatal} is not significantly underrepresented, despite O/E = 0, because there was not much difference between observed and expected values in which no pair of {ʧ, palatal} was observed and only 1.2 pair of {ʧ, palatal} was expected. The pair {t, alveolar} is actually significantly overrepresented, which seems due to many occurrences of coronal sonorants, /l/ in particular, in the existing reduplicative forms.

Interestingly, {t, labial} and {t, velar} are significantly underrepresented, which is not predicted by identity avoidance. Also, {ʧ, alveolar} is significantly underrepresented, which is not surprising, considering that /ʧ/ and alveolar consonants can come under the category of coronals.

I now consider cooccurrence of CI and C_2.

(20) Table 5.6 CI – C_2 pairs: **Place Identity**
(If $\chi^2 > 3.84$, $p < .05$. N = 51)

CI \ C_2	Labial	Alveolar	Velar
p	O = 1 O/E = 1.25 $\chi^2 = 0.08$	O = 6 O/E = 0.87 $\chi^2 = 0.37$	O = 14 O/E = 1.06 $\chi^2 = 0.22$
t	O = 0 O/E = 0 $\chi^2 = 1.15$	O = 7 O/E = 1.19 $\chi^2 = 0.38$	O = 11 O/E = 0.98 $\chi^2 = 0.02$
ʧ	O = 1 O/E = 2 $\chi^2 = 0.73$	O = 4 O/E = 1 $\chi^2 = 0$	O = 7 O/E = 0.91 $\chi^2 = 0.14$

Table 5.6 presents combinations of CI = /t, p, ʧ/ and context Cs (C_2, which is coda C) for the classes of labial, alveolar, and velar. The place of palatal was excluded from the table, because a palatal cannot occur as C_2 (coda C) in Korean. None of the pairs in Table 5.4 show significant underrepresentation; specifically, the pairs {p, labial} and {t, alveolar} do not appear to be underrepresented, contrary to prediction.

Therefore, in terms of place, co-occurrence restrictions against pairs of CI – C_1 (both in onset) appear to be stricter than restrictions against pairs of CI – C_2 (CI in onset and C_2 in coda).

In terms of manner, we expect that {p, stop}, {t, stop}, and {ʧ, affricate} would be underrepresented because these pairs share manner of articulation. However, none of them show significant underrepresentation, suggesting that there are no co-occurrence restrictions against combining two Cs with the same manner of articulation. This seems to be the case with the pairs of CI – C_1 as in Table 5.7 and with the pairs of CI – C_2 as in Table 5.8.

(21) Table 5.7 CI – C_1 pairs: **Manner Identity**
(If $\chi^2 > 3.84$, $p < .05$. N = 51)

CI \ C_1	Stop	Nasal	Affricate	Fricative	Approximant
p	O = 12 O/E = 1.33 $\chi^2 = 2.89$	O = 0 O/E = 0 $\chi^2 = 3.04$	O = 4 O/E = 2.35 $\chi^2 = 5.95$	O = 0 O/E = 0 $\chi^2 = 1.4$	O = 5 O/E = 0.65 $\chi^2 = 2.77$
t	O = 5 O/E = 0.65 $\chi^2 = 2.73$	O = 0 O/E = 0 $\chi^2 = 2.26$	O = 0 O/E = 0 $\chi^2 = 2.26$	O = 2 O/E = 2.86 $\chi^2 = 3.82$	O = 11 O/E = 1.67 $\chi^2 = 6.72$
ʧ	O = 5 O/E = 0.94 $\chi^2 = 0.03$	O = 4 O/E = 4 $\chi^2 = 13.38$	O = 0 O/E = 0 $\chi^2 = 1.31$	O = 0 O/E = 0 $\chi^2 = 0.67$	O = 3 O/E = 0.67 $\chi^2 = 0.96$

(22) Table 5.8 CI – C_2 pairs: **Manner Identity**
(If $\chi^2 > 3.84$, $p < .05$. N = 51)

CI \ C_2	Stop	Nasal	Approximant
p	O = 6 O/E = 0.92 $\chi^2 = 0.14$	O = 11 O/E = 1.17 $\chi^2 = 0.78$	O = 4 O/E = 0.8 $\chi^2 = 0.4$
t	O = 6 O/E = 1.09 $\chi^2 = 0.07$	O = 8 O/E = 1 $\chi^2 = 0$	O = 4 O/E = 0.93 $\chi^2 = 0.04$
ʧ	O = 4 O/E = 1.05 $\chi^2 = 0.03$	O = 4 O/E = 0.73 $\chi^2 = 0.89$	O = 4 O/E = 1.38 $\chi^2 = 0.87$

For the pairs of CI – C_2 (coda C), affricates and fricatives were left out since they do not come as coda Cs in Korean.

All in all, the predictions of identity avoidance did not seem clearly borne out in the pairs of CI and context Cs in terms of place and manner. However, among others, the pairs of CI – C_1 appear to be more compliant to co-occurrence restrictions on place than the pairs CI-C_2.

The sample size for the above analysis was 51, and identity avoidance effects may have shown up in a larger sample. In addition, cell sizes were not adequate for the chi-square test, which normally assumed 5 or more in 80% of cells when there are more than 4 cells, but no cells with zero count. This can be a potential problem for the statistical test because significant results, which in fact exist, may not come out. I will look into a bigger sample, which was collected in a word creation experiment, in the following section.

5.2.3.2 Word creation data

In this section I consider whether identity avoidance account plays a role in the shape of newly created words, specifically nonce reduplicative forms. To address this question, I looked at results from Experiment 3, a word creation experiment. I chose Experiment 3 for two reasons. First, it has one of the largest samples, which is important for valid statistical results; Experiment 3 produced 1665 tokens (vs. 472 for Experiment 1, 1159 for Experiment 2, and 1662 for Experiment 4). Second, all the CI responses in Experiment 3 consists of /t, p, ʧ/ only, which can help to show speakers' tendency, controlling for potential third variables when other CIs were also involved. The participants in Experiments 1 and 2 chose a CI freely out of all possible Cs in Korean; the participants in Experiment 4 could only choose between /t/ and /ʧ/ for CI.

The place and manner of articulation that were used for context in Experiment 3 are Place = {Labial, Alveolar, Velar} and Manner = {Stop, Nasal, Fricative}:

(23) Identity classes by Place
- a. Labial = {p, m}
- b. Alveolar = {t, n, s}
- c. Velar = {k, ŋ}

(24) Identity classes by Manner
- a. Stop = {p, t, k}
- b. Nasal = {m, n, ŋ}
- c. Fricative = {s}

The identity classes in (23-24) were all used for the CI – C_1 pairs; however, the identity classes by manner were reduced to stops and nasals for the CI – C_2 pairs since C_2 occurs in coda, where fricative /s/ is neutralized as stop /t/ in Korean.[63]

[63] In spelling out coda consonants, it is more natural to use /s/ instead of /t/, so I used "ㅅ"

The following tables show O values, O/E ratio, and chi-square values. The cells that exhibit significant underrepresentation are shaded.

(25) Table 5.9 CI – C_1 pairs: **Place identity**
(If $\chi^2 > 3.84$, $p < .05$. N = 1665)

CI \ C_1	Labial	Alveolar	Velar
p	O = 64 O/E = 0.59 $\chi^2 = 33.61$	O = 216 O/E = 1.18 $\chi^2 = 13.23$	O = 122 O/E = 1.13 $\chi^2 = 3.17$
t	O = 190 O/E = 1.11 $\chi^2 = 9.31$	O = 292 O/E = 1 $\chi^2 = 0.24$	O = 159 O/E = 0.93 $\chi^2 = 2.83$
ʧ	O = 196 O/E = 1.18 $\chi^2 = 10.41$	O = 257 O/E = 0.91 $\chi^2 = 8.84$	O = 169 O/E = 1.02 $\chi^2 = 0.18$

According to Hypothesis 1 on identity avoidance, we predict that pairs {p, labial} and {t, alveolar} should be underrepresented at a significant level because these pairs share place of articulation, which must be avoided in consideration of identity avoidance. Table 5.9 shows that the pairs of {p, labial} are significantly underrepresented as was predicted. The pairs of {t, alveolar} are not significantly underrepresented; however, they are not significantly overrepresented, either. Hence this does not disconfirm the hypothesis for identity avoidance. Meanwhile, the {ʧ, alveolar} pairs are significantly underrepresented, which was also the case with the dictionary data in the preceding section. This again can be explained in terms of sharing the place: both /ʧ/ and alveolar consonants come under the rubric of coronals. Therefore, Hypothesis 1 is confirmed with regard to the place of articulation for CI – C_1 pairs.

Table 5.10 below shows the pairs of CI – C_2 in terms of place:

(standing for /s/ in Korean) in spelling out the stimuli. However, the participants were asked to read aloud when they create new reduplicative words; therefore, I assume that coda /s/ was neutralized as /t/, in the process. However, I do not entirely exclude the spelling effects, which can be a confounding factor in any such experiments.

(26) Table 5.10 CI – C_2 pairs: **Place Identity**
(If $\chi^2 > 3.84$, $p < .05$. N = 1665)

CI \ C_2	Labial	Alveolar	Velar
p	O = 145 O/E = 1.13 $\chi^2 = 3.96$	O = 121 O/E = 0.86 $\chi^2 = 6.7$	O = 136 O/E = 1.06 $\chi^2 = 1.17$
t	O = 214 O/E = 1.06 $\chi^2 = 1.12$	O = 225 O/E = 1.02 $\chi^2 = 0.72$	O = 202 O/E = 1 $\chi^2 = 1.1$
ʧ	O = 181 O/E = 0.92 $\chi^2 = 5.83$	O = 239 O/E = 1.11 $\chi^2 = 5.49$	O = 202 O/E = 1.02 $\chi^2 = 0.68$

We also predict that the pairs of {p, labial}, {t, alveolar}, and {ʧ, alveolar} – when /ʧ/ and alveolar Cs are considered coronals – should be significantly underrepresented if the Cs in a pair tend not to be identical in place. However, they are not underrepresented; rather, the pairs of {p, labial} and {ʧ, alveolar} are even overrepresented at a significant level. This goes against Hypothesis 1, and I conjecture that this may be ascribed to different requirements for different positions in a syllable in Korean. The occurrence of coda Cs is far more restricted than that of onset Cs: coda Cs are limited to {p, t, k, m, n, ŋ, l}, while onset Cs can be any C – except for /ŋ/ – from the consonant inventory.

The following two tables exhibit the O/E ratios with regard to co-occurrence restrictions for CI and context Cs (C_1 and C_2) in terms of manner.

(27) Table 5.11 CI – C_1: **Manner Identity**
(If $\chi^2 > 3.84$, $p < .05$. N = 1665)

CI \ C_1	Stop	Nasal	Fricative
p	O = 171 O/E = 1 $\chi^2 = 0.33$	O = 148 O/E = 0.9 $\chi^2 = 3.38$	O = 83 O/E = 1.3 $\chi^2 = 8.04$
t	O = 292 O/E = 1.07 $\chi^2 = 2.51$	O = 257 O/E = 0.99 $\chi^2 = 0.42$	O = 92 O/E = 0.61 $\chi^2 = 2.96$
ʧ	O = 257 O/E = 0.97 $\chi^2 = 1.33$	O = 270 O/E = 1.07 $\chi^2 = 3.72$	O = 95 O/E = 0.96 $\chi^2 = 0.91$

No pair of CI – C_1 was significantly underrepresented in term of manner, according to Table 5.11.

(28) Table 5.12 CI – C_2: **Manner Identity**
(If $\chi^2 > 3.84$, $p < .05$. N = 1665)

CI \ C_2	Stop	Nasal
p	O = 219 O/E = 1.07 $\chi^2 = 2.92$	O = 183 O/E = 0.93 $\chi^2 = 2.34$
t	O = 315 O/E = 0.98 $\chi^2 = 2.32$	O = 326 O/E = 1.05 $\chi^2 = 2.32$
ʧ	O = 321 O/E = 1.02 $\chi^2 = 0.28$	O = 301 O/E = 1 $\chi^2 = 0.28$

Table 5.12 also shows that no pair of CI – C_2 was significantly underrepresented in term of manner. From Table 5.11 and 5.12 above, I suspect that the manner of articulation may not be an appropriate predictor for co-occurrence restrictions.

Thus far, we have seen that place, rather than manner, can be a better predictor for consonant co-occurrence restrictions. In addition, we could see identity avoidance operating for the pairs which occur in the same grammatically defined positions, e.g. both Cs in a pair occurring in onset. This further requirement may be due to the fact that Korean has drastically different sets of possible onset and coda consonants.

5.2.4 Discussion

The general patterns involving CI and the neighboring Cs (C_1 and C_2) in reduplicants of the form of VC_1VC_2-**CI**$\underline{VC_1VC_2}$ appeared to show some effect of identity avoidance. With regard to pairs of CI and C_1 which share place of articulation, for the dictionary data, O/E values showed that the pair {p, labial} was underrepresented at a marginally significant level; the pair {ʧ, palatal} was underrepresented, but not at a significant level, because of not much difference between the observed and expected values; and the pair {ʧ, alveolar} (both of which are coronals) were significantly underrepresented. For the experimental responses, O/E values showed that the pair {p, labial} was significantly underrepresented; the pair {t, alveolar} was neither significantly underrepresented nor significantly overrepresented; and the pair {ʧ, alveolar} (both coronals) was significantly underrepresented.

There was no significant underrepresentation for pairs of CI and C_1 which share manner of articulation. For pairs of CI and C_2 which share both place and manner of articulation, there was no significant underrepresentation. These results indicate that place is more relevant to consonant co-occurrence restrictions than manner.

Support for an OCP-Place restriction in Korean vocabulary was provided by Ito (2006), who noted that pairs of consonants in verbal/nominal stems of the form CVC tend not to share place of articulation. Therefore, it is confirmed that identity avoidance in place exists in Korean, both in the general lexicon and in a specific lexicon (reduplicated words). Furthermore, the investigation in this chapter could affirm that the identity avoidance in place is also obeyed in newly created words.

Ito's examination of O/E ratio was for the co-occurrences of onset C and coda C, which showed significant underrepresentation with onset C – coda C pairs sharing place. However, the examination of O/E in this chapter did not show significant underrepresentation with onset C – coda C pairs sharing place; instead, it showed significant underrepresentation with onset C – onset C pairs which share place.

It is noteworthy that Ito looked into the consonant co-occurrences within a syllable, with one C in onset and the other in coda in a form of CVC, whereas I looked at the consonant co-occurrences across a syllable, i.e. between CI and onset C and between CI and coda C in an adjacent syllable in the reduplicant form of **CI**V.C_1VC_2 (CI – C_1; CI – C_2). Therefore, if we acknowledge that identity avoidance in place is respected also between onset C and coda C in general, then the apparent disobedience of identity avoidance in a pair of onset C – coda C in the studied reduplicated forms may not be attributed to the matter of syllabic positions. That is, I doubt if I can argue that Cs in the same syllabic positions, such as onset C – onset C, are more subject to identity avoidance than Cs in disparate syllabic positions like onset C – coda C. Rather, I presume that there are distance effects. C_1 is adjacent to CI, whereas C_2 is non-adjacent to CI. Identity avoidance in place is generally applied to the former type of pairs more rigorously than to the latter type of pairs (cf. Frisch & Zawaydeh 2001 on Jordanian Arabic).

The correlations that were found in Chapter 2 between the CI – C_1 combinations (both in onset) and lexical C – C combinations (both in onset), and between CI – C_2 combinations (CI in onset and C_2 in coda) and lexical C – C combinations (first C in onset and second C in coda) appear consistent with the results in this chapter, in that seemingly different degrees of restriction on co-occurring consonants are applied to onset C – onset C and to onset C – coda C. However, the findings of co-occurrence restrictions in the general lexicon and the reduplicated words based on O/E ratio in this chapter further implies that the correlations may not stem from distinct syllabic positions, but may be due to a matter of distance between consonants.

5.3 Summary

The general vocabulary (particularly, nominal/verbal stems) of Korean was argued to respect an OCP-Place constraint which does not allow consonants with the same place to occur together. Experimental responses also supported this type of constraint (identity avoidance in terms of place) in that consonants which share place of articulation tend not to co-occur in newly created reduplicative words.

The results are consistent with various phenomena in other languages, in which there are stronger restrictions against co-occurrences of consonants that share place, rather than manner. Furthermore, the current results suggest that distance between consonants matters in the co-occurrence restrictions for consonants, which is also consistent with the restrictions found in other languages.

Chapter 6
Conclusions and Future Directions

6.1 Summary and Conclusions

Thus far in this dissertation I have studied consonant insertion, particularly with reference to the Korean reduplication. The specific reduplication studied frequently expresses an onomatopoeic function, e.g. *t'ok-t'ok* 'dripping; knocking; smart.' While the reduplicant and the base are generally identical, as in *p^{h}oŋtaŋ-p^{h}oŋtaŋ* 'with splashes,' *tekul-tekul* 'rolling,' when the first member of the pair begins with a vowel, the second member has a consonant inserted, e.g. *als'oŋ-**t**als'oŋ* 'confusing,' *opul-**k**opul* 'meanderingly,' *olmaŋ-**ʧ**olmaŋ* 'all sorts of little things (in a cluster).' Almost any consonant out of the Korean phoneme inventory can be inserted in reduplication.

I have investigated both corpus data and speakers' behavior in online word formation with a specific question in mind: to what extent lexical frequency and grammar, respectively or collaboratively, will account for linguistic phenomena. I have shown that the choice of inserted consonant shows a great deal of variation.

This phenomenon appears to contradict claims in previous literature concerning the identity of consonants inserted in reduplication. Contrary to the claim of Alderete *et al*. (1999) that segments in the reduplicant that are not present in the base represent an emergence of the unmarked, the inserted consonant (CI) in Korean reduplication cannot be an unmarked/default consonant because distinct consonants can be inserted in identical environments, e.g. *alok-**t**alok* 'mottled,' *ulak-**p**ulak* 'wild' where /t/ and /p/ are epenthesized although the bases contain the same set of consonants, /l/ and /k/. Moreover, a particular vowel does not force the occurrence of a particular consonant, e.g. *ulak-**p**ulak* 'wild,' *umuk-**ʧ**umuk* 'unevenly hollowed,' *upul-**k'**upul* 'windingly' in which different CIs are followed by the same vowel /u/.

While Korean reduplication may insert a wide range of consonants, the choice of inserted consonant is not entirely randomly made. Analysis of all the cases of inserted consonants in biconsonantal bases in a Korean dictionary demonstrated that certain consonants (/t, p, ʧ/) are much more frequently inserted than others (/k, s, m, j/). Thus there is neither a single preferred consonant nor a random choice among all possible consonants. I investigated both the lexical patterns in existing reduplications and speakers' behavior in forming new reduplicated words.

The dictionary and experimental results showed different sets of CIs, {/t/, /p/, /ʧ/, /k/, /s/, /m/, /j/} for the dictionary and {/t/, /p/, /ʧ/, /k/, /s/, /m/, /j/, /n/, /l/} for an experiment (Experiment 1). Furthermore, in the dictionary the most likely CIs were from the classes /t, p, ʧ/, whereas in the experiment the frequencies of CIs were more widely distributed among various consonants, /t, p, ʧ, k, s, m/. Hence the frequencies of CIs in the word creation task did not simply reflect the frequencies of CIs in the lexicon.

However, examination of the lexical patterns suggests that lexical frequency plays a role in the choice of inserted consonant. First, the frequency of CIs in a word creation experiment correlated significantly with the frequency of word-initial Cs in the Korean corpus. Second, the frequency of consonant combinations CI – C_1 in forms of the shape **CI**V.C_1VC_2 correlated significantly with the frequency of combinations of consonants in CVCV forms in the corpus. Similarly, the frequency of combinations of CI – C_2 in forms of the shape **CI**V.C_1VC_2 correlated with the frequency of combinations of onset C – coda C in the corpus. Third, the frequency of C – V combinations in the experiment correlated significantly with the frequency of lexical C – V combinations in the corpus.

Another factor investigated was the effect of a restriction on syllable contact banning heterosyllabic sequences in which a coda C of a preceding syllable is of lower sonority than a directly following onset C. This restriction has been shown to play a role in Korean phonology, and is potentially relevant to choice of inserted consonant in reduplicants of the form VCVC-**CI**VCVC. This constraint was found to work more strongly for nonce reduplicated words than for the general vocabulary.

The role of the following V on the choice of inserted C was also investigated. Korean speakers' behavior in many psycholinguistic experiments suggested that a CV (body) constituent is prominent for Korean speakers, as opposed to the speakers of English-like languages which evidently have a closer tie between V and C (rhyme).

An additional factor that appeared to affect the choice of CI was identity avoidance. The general vocabulary of Korean was argued to respect an OCP-Place constraint (identity avoidance in place), which does not allow consonants with the same place to co-occur. The dictionary data and the experimental responses also showed significant effects of identity avoidance in place, based on the ratio of observed to expected occurrences of inserted consonants in different contexts. Data from the general lexicon and the reduplication data also revealed a distance effect: co-occurrence restrictions appeared to be stricter for adjacent consonant pairs than for non-adjacent consonant pairs.

Lexical frequency was shown to play a role in the choice of inserted consonants, to some extent; however, individual speakers did not necessarily reflect the lexical patterns. There were two distinct patterns among the speakers with regard to the choice of CI: those who preferred /t/ predominantly over other Cs and those who preferred /ʧ/ predominantly over other Cs. Moreover, within a group of the speakers who chose /t/ most of the time there was microvariation: some speakers chose less preferred CIs when the context contained their preferred CI, whereas other speakers stayed with the preferred CI regardless of context.

The findings in this dissertation cast light on the question of how much of speakers' behavior reflects lexical statistics and grammar. The lexical frequency is clearly reflected in the speakers' linguistic behavior. However, many of the lexical effects I found in the nonce reduplicated words are in fact founded in grammatically determined concepts that have been shown to play a role in many languages, such as syllable contact restrictions, identity avoidance, and preference for particular CV combinations. Unlike English-type languages that have stricter co-occurrence restrictions for the VC (rhyme) constituent in the sub-hierarchy of a syllable, Korean has a penchant for the CV (body) constituent, to which Korean speakers pay attention to in the choice of CI; hence the correlations between the word creation and the corpus in terms of the C – V combining pattern. Furthermore, Korean in general has co-occurrence restrictions on consonants which affect the choice of CI, as well; hence the correlations in the C – C combining pattern between the word creation task and the corpus. These patterns suggest that speakers' behavior is not determined solely by lexical statistics, but also by grammatical principles.

6.2 Future Directions

A number of factors were identified that affect Korean speakers' choice of inserted consonants. Further investigation is necessary to determine the relative influence of the different factors. Examination of reduplication data with C change, e.g. ***k**ʌmpul-**t**ʌmpul* 'pell-mell,' in addition to C insertion, may furnish deeper insights into the consonant insertion behavior.

Furthermore, we are in need of an appropriate model that can deal with the variable and gradient data in a more precise way. While a preliminary stochastic grammar is included in an Appendix to Chapter 3, a thorough investigation of a model that is capable of predicting the data with variation and gradience remains to be done. Recent years have seen various proposals for models to capture variation and gradience in a formal framework. The Korean reduplication data provides a case against which such models can be tested.

References

Abney, Steven. 1997. Statistical methods and linguistics. *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. Judith L. Klavans and Philip Resnik (eds.). Cambridge, MA: MIT Press. pp. 1-26.

Ahn, Miyeon. 2005. A Phonetically Based Account of English Reduplication. MA thesis. Seoul National University.

Albright, Adam. 2004. The lexical bases of morphological well-formedness. Ms. University of California, Los Angeles.

Albright, Adam. 2006a Gradient phonotactic effects: lexical? grammatical? both? neither?, The talk handout at the annual meeting of Linguistic Society of America. Albuquerque. January 7.

Albright, Adam. 2006b. A grammatical account of gradient phonotactics. Presentation handout in the colloquia series of Stony Brook University, Spring 2006.

Albright, Adam. 2006c. Explaining universal tendencies and language particulars in analogical change. Ms. MIT.

Albright, Adam. 2007. Gradient phonological acceptability as a grammatical effect. Ms. MIT.

Albright, Adam. 2008a. From clusters to words: Grammatical models of nonce word acceptability. Handout of talk presented at 82nd LSA. Chicago.

Albright, Adam. 2008b. How many grammars am I holding up? Discovering phonological differences between word classes. In Charles B. Chang and Hannah J. Hanyie (eds.), Proceedings of the 26th West Coast Conference on Formal Linguistics. Somerville, MA: Cascadilla Proceedings Project. pp. 1-20.

Albright, Adam. In prep. Gradient phonological acceptability as a grammatical effect. Ms. MIT.

Albright, Adam and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/ experimental study. *Cognition* 90: 119–161.

Albright, Adam and Bruce Hayes. 2006. Modeling productivity with the Gradual Learning Algorithm: The problem of accidentally exceptionless generalizations. In G. Fanselow, C. Féry, R. Vogel, and M. Schlesewsky (eds.). *Gradience in Grammar: Generative Perspectives*. Oxford University Press. pp. 185–204.

Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, John McCarthy, and Suzanne Urbanczyk. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30-3: 327-364.

Alderete, John and Stefan Frisch. 2007. Dissimilation in grammar and the lexicon. In de Lacy, Paul (ed.). *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press. pp. 379-398.

Alderete, John, Paul Tupper, and Stefan Frisch. 2009. Phonotactic learning without a priori constraints: A connectionist analysis of Arabic cooccurrence restrictions. ROA, http://roa.rutgers.edu/files/1055-1109/1055-ALDERETE-0-0.PDF

Alegre, Maria and Peter Gordon. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40: 41-61.

Allen, Scott W. and Lee R. Brooks. 1991. Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General* 120-1: 3-19.

An, Young-ran. 2006. Identity avoidance in Korean reduplication. Ms. Stony Brook University.

An, Young-ran, Jiwon Hwang, and Yunju Suh. 2008. Phonetic status of Cj combinations in Korean and Spanish. Poster presentation at the 156th Meeting of the Acoustical Society of America (ASA). Miami, Florida. November 10-14.

Anttila, Arto. 1995. Deriving variation from grammar: A study of Finnish genitives. Ms. Stanford University.

Anttila, Arto. 1997. Deriving variation from grammar. In Frans Hinskens, Roeland van Hout and W. Leo Wetzels (eds.). *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins. pp. 35-68.

Anttila, Arto and Young-mee Yu Cho. 1998. Variation and change in Optimality Theory. *Lingua* 104: 31-56.

Archangeli, Diana and Douglas Pulleyblank. 2007. Harmony. In de Lacy Paul (ed.). *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press. pp. 351-378.

Baayen, Harald. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer.

Baertsch, Karen. 2002. An Optimality Theoretic Approach to Syllable Structure: The Split Margin Hierarchy. Ph.D. dissertation. Indiana University.

Baertsch, Karen and Stuart Davis. 2003. The split margin approach to syllable structure. In ZAS Papers in Linguistics 32: 1-14.

Bailey, Todd M. and Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44: 568-591.

Bao, Zhiming. 1990. Fanqie languages and reduplication. *Linguistic Inquiry* 21-3: 317-350.

Baroni, Marco and Stefan Evert. 2006. Counting words: An introduction to Lexical Statistics. Course material of Introductory Course, Language and Computation Section at the 18th European Summer School in Logic, Language, and Information (ESSLLI). July 31 – August 11, 2006, Málaga, Spain.

Battistella, Edwin L. 1996. *The Logic of Markedness*. Oxford: Oxford University Press.

Benua, Laura. 2004. Transderivational identity: phonological relations between words. In John McCarthy (ed.). *Optimality Theory in Phonology: A Reader*. pp. 419-437.

Berent, Iris. 2002. Identity avoidance in the Hebrew lexicon: Implications for symbolic accounts of word formation. *Brain and Language* 81: 326-341.

Berent, Iris, Daniel L. Everett, and Joseph Simron. 2001. Do phonological representations specify variables?: Evidence from the Obligatory Contour Principle. *Cognitive Psychology* 42: 1-60.

Berent, Iris, Joseph Shimron, and Vered Vaknin. 2001. Phonological constraints on reading: Evidence from the Obligatory Contour Principle. *Journal of Memory and Language* 44-4: 644-665.

Berent, Iris, Vered Vaknin, and Joseph Shimron. 2004. Does a theory of language need a grammar? Evidence from Hebrew root structure. *Brain and Language* 90: 170-182.

Berg, Thomas and Jussi Niemi. 2000. Syllabification in Finnish and German: Onset filling vs. onset maximization. *Journal of Phonetics* 28-2: 187-216.

Berkley, Deborah Milam. 2000. Gradient Obligatory Contour Principle Effects. Ph.D. dissertation, Northwestern University.

Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge University Press.

Blevins, Juliette. 2008. Consonant epenthesis: Natural and unnatural histories. In Jeff Good (ed.). *Language Universals and Language Change*. Oxford: Oxford University Press. pp. 79–109.

Boersma, Paul. 1997. How we can learn variation, optionality, and probability. Ms. University of Amsterdam.

Boersma, Paul. 1998. The OCP in functional phonology. Ms. University of Amsterdam. Also in Paul Boersma (1998), Chapter 18, *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives*. pp. 415-440.

Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45–86.

Broe, Michael. 1993. Specification Theory: The Treatment of Redundancy in Generative Phonology. PhD. Dissertation. The University of Edinburgh.

Broselow, Ellen. 1984. Default consonants in Amharic morphology. *MIT Working Papers in Linguistics* 7: 15-32.

Broselow, Ellen. 2009. Stress adaptation in loanword phonology: perception and learnability. In Paul Boersma and Silke Hamann (eds.). *Phonology in Perception*. Mouton de Gruyter. pp. 191-234.

Browman, Catherine P. and Goldstein, Louis M. 1992. Articulatory phonology: An overview. *Phonetica* 49:155–180.

Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.

Bybee, Joan L. 1995a. Regular morphology and the lexicon. *Language and Cognitive Processes* 10: 425-455.

Bybee, Joan L. 1995b. The semantic development of past tense modals in English. In J. Bybee and S. Fleischman (eds.). *Modality in Grammar and Discourse*. Amsterdam: John Benjamins. pp. 503-517.

Bybee, Joan L. 2000a. The phonology of the lexicon: Evidence from lexical diffusion. In Michael Barlow and Suzanne Kemmer (eds.). *Usage-Based Models of Language*. Stanford: CSLI Publications. pp. 65-85.

Bybee, Joan L. 2000b. Lexicalization of sound change and alternating environments. In Michael B. Broe and Janet M. Pierrehumbert (eds.). *Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press. pp. 250–268.

Bybee, Joan L. 2001. *Phonology and Language Use*. Cambridge Studies in Linguistics 94.

Bybee, Joan L. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14: 261-290.

Bybee, Joan L. 2007. *Frequency of Use and the Organization of Language*. Oxford University Press.

Bybee, Joan L. and Carol Lynn Moder. 1983. Morphological classes as natural categories. *Language* 59-2: 251-270.

Bybee, Joan L. and Dan I. Slobin. 1982. Rules and schemes in the development and use of the English past tense. *Language* 58-2: 265-289.

Bybee, Joan L. & Paul Hopper. 2001. *Frequency and the Emergence of Linguistic Structure.* Amsterdam: John Benjamins Publishing Company.

Bybee, Joan L. and Dan I. Slobin. 1982. Rules and schemas in the development and use of the English past tense. *Language* 58: 265-289.

Chae, Wan. 2003. *Onomatopoeia in Korean [hankwuke-uy uysenge-wa uytaye]*. Seoul: Seoul National University.

Chen, Train-Min, Gary Dell, and Jenn-Yeu Chen. 2004. *A cross-linguistic study of phonological units: Syllables emerge from the statistics of Mandarin Chinese, but not from the statistics of English*. Poster Presented in the 26th Annual Meeting of the Cognitive Science Society. August 5-7, Chicago, IL.

Cho, Sung-moon. 2001. An Optimality-theoretic analysis on the change of the vowel harmony in Korean. Ms.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row Publishers.

Clements, George N. 1990. The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (eds.). *Papers in Laboratory Phonology I: between the Grammar and Physics of Speech.* Cambridge: Cambridge University Press. pp. 282-333.

Clements, George N & Keyser, Samuel J. 1983. *CV Phonology*. Cambridge, Mass.: MIT Press.

Coetzee, Andries W. 2004. What It Means to Be a Loser: Non-Optimal Candidates in Optimality Theory. Ph.D. dissertation, University of Massachusetts at Amherst.

Coetzee, Andries W. 2005. The Obligatory Contour Principle in the perception of English. In Sónia Frota, Marina Vigário and Maria João Freitas (eds.). *Prosodies*. Berlin: Mouton de Gruyter. pp. 223-245.

Coetzee, Andries. 2006a. Grammar is both categorical and gradient. ROA # 864 at www.roa.rutgers.edu.

Coetzee, Andries W. 2006b. Variation as accessing "non-optimal" candidate. *Phonology* 23: 337-385.

Coetzee, Andries W. 2007. A lexical theory of variation. Handout from the Workshop on Variation, Gradience and Frequency in Phonology. July 6-8, Stanford.

Coetzee, Andries. 2008a. Phonological variation and lexical frequency. Ms. University of Michigan.

Coetzee, Andries W. 2008b. Grammaticality and ungrammaticality in phonology. *Language* 84-2: 218-257

Coetzee, Andries W. 2009. An integrated grammatical/non-grammatical model of phonological variation. In Young-Se Kang, Jong-Yurl Yoon, Hyunkyung Yoo, Sze-Wing Tang, Yong-Soon Kang, Youngjun Jang, Chul Kim, Kyoung-Ae Kim, and Hye-Kyung Kang (eds.). *Current Issues in Linguistic Interfaces. Volume 2*. Seoul: Hankookmunhwasa. pp. 267-294.

Coetzee, Andries W. and Joe Pater. 2005. Gradient phonotactics in Muna and Optimality Theory. Ms.

Coetzee, Andries W. and Joe Pater. 2006. Lexically ranked OCP-Place construction in Muna. ROA # 842 at www.roa.rutgers.edu.

Coetzee, Andries W. and Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26: 289-337.

Cohn, Abigail 1992. The consequences of dissimilation in Sundanese. *Phonology* 9: 199-220.

Coon, Jessica and Gillian Gallagher. 2008. Similarity and Correspondence in Chol Mayan. Ms. MIT.

Cutler, Anne, Nuria Sebastián-Gallés, and Brit van Ooijen. 2000. Constraints of vowels and consonants on lexical selection: Cross-linguistic Comparisons. *Memory and Cognition* 28: 746-755.

Davidson, Lisa. 2007. The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology* 24-2: 261-286.

Davis, Stuart. 1989. On a non-argument for the Rhyme. *Journal of Linguistics* 25: 211-17.

Davis, Stuart. 1991. Coronals and the phonotactics of nonadjacent consonants in English. In Carole Paradis and Jean-François Prunet (eds.). *The Special Status of Coronals: Internal and External Evidence*. San Diego: Academic Press. pp. 49-60.

Davis, Stuart. 1998. Syllable contact in Optimality Theory. Korean *Journal of Linguistics* 23: 181-211.

Davis, Stuart and Seung-Hoon Shin 1999. The syllable contact constraint in Korean: An Optimality-theoretic analysis. *Journal of East Asian Linguistics* 8: 285-312.

de Lacy, Paul. 2002. The Formal Expression of Markedness. Ph.D. dissertation. University of Massachusetts at Amherst. MA: GLSA Publications.

de Lacy, Paul. 2006. *Markedness: Reduction and Preservation in Phonology*. Cambridge Studies in Linguistics 112. Cambridge University Press.

de Lacy, Paul. 2007. *The Cambridge Handbook of Phonology*. Cambridge University Press.

de Lucca, Manuel. 1987. *Diccionario Aymara-Castellano Castellano-Aymara*. La Paz, Cochalamba: Editorial Los Amigos del Libro.

Dell, Gary S., Cornell Juliano, and Anita Govindjee. 1993. Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science* 17:149-195.

Demircan, Ozgur. 1987. Emphatic reduplication in Turkish. In Hendrik E. Boeschoten and Ludo Th. Verhoeven (eds.). *Studies on modern Turkish: Proceedings of the 3rd conference in Turkish Linguistics.* Tilburg University Press.

Derwing, Bruce L., Yeo Bom Yoon, and Sook Whan Cho. 1993. The organization of the Korean syllable: Experimental evidence. In Patricia M. Clancy (ed.). *Japanese/Korean Linguistics, Volume 2*. Stanford, CA: Center for the Study of Language and Information.

Derwing, Bruce L. 2007. What's in CVC-like things?: Ways and means to look at phonological units across languages. In Maria-Josep Solé, Patrice Speeter Beddor, and Manjari Ohala (eds.) *Experimental Approaches to Phonology*. Oxford: Oxford University Press. pp. 325-338.

Deza Galindo, Juan Francisco. 1989. *Jaya Mara Aru: Nuevo Diccionario Aymaracastellano, Castellano-aymara*. Lima, Peru: Graphos 100 Editores.

Dienhart, John. 1999. Stress in reduplicative compounds: Mish-mash or hocus-pocus? *American Speech* 74: 13-38.

Dobrovolsky, Michael. 1987. Why CVC in Turkish reduplication? In Plilw and M. Saari (eds). *The Nordic Languages and Modern Linguistics* 6. Helsinki. pp. 131-146.

Downing, Laura. 2000. Morphological and prosodic constraints on Kinande verbal reduplication. *Phonology* 17: 1-38.

Ellias-Ulloa, Jose. 2007. Stress adjustments in Shipibo loanwords from Spanish. Ms. Stony Brook University.

Essence Korean Dictionary [Eysseyns Kwuke Sacen]. 2006. Phacwu, Korea: Mincwungselim Co.

Evanini, Keelan. 2006. How to formalize variation: Stochastic OT models and /s/ deletion in Spanish. Ms. University of Pennsylvania.

Fidelholtz, James. 1975. Word frequency and vowel reduction in English. In *CLS (Chicago Linguistic Society)* 11: 200-213.

Finley, Sara and William Badecker. 2007. Towards a substantively biased theory of learning. Ms.

Flemming, Edward. 1995. Auditory Representations in Phonology. Ph.D. dissertation. University of California at Los Angeles.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Robert Kirchner and Donca Steriade (eds.), *Phonetically Based Phonology*. Cambridge: Cambridge University Press. pp. 232-276.

Frisch, Stefan A. 1996. Similarity and Frequency in Phonology. Ph.D. dissertation. Northwestern University, Evanston, Illinois.

Frisch, Stefan A. 2004. Language processing and segmental OCP effects. In Hayes, Kirchner, and Steriade (eds.), *Phonetically-based phonology*. Cambridge University Press. pp. 346-371.

Frisch, Stefan A., Michael B. Broe, and Janet B. Pierrehumbert. 1997. Similarity and phonotactics in Arabic. Bloomington, IN and Evanston, IL: Indiana University and Northwestern University, Ms. ROA-223, Rutgers Optimality Archive, http://ruccs.rutgers.edu/roa.html.

Frisch, Stefan A., Nathan R. Large, and David B. Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42: 481-496.

Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language & Linguistic Theory* 22: 179-228.

Frisch, Stefan A. and Bushra Zawaydeh. 2001. The psychological reality of OCP-Place in Arabic. *Language* 77: 91-106.

Fudge, Erik. C. 1969. Syllables. *Journal of Linguistics* 5: 253–287. [Reprinted as Chapter 19 in Goldsmith (1999).]

Gallagher, Gillian 2008. The role of contrast in laryngeal cooccurrence restrictions. Paper presented to the 27th West Coast Conference on Formal Linguistics. UCLA.

Glowacka, Dorota. 2001. Unstressed vowel deletion and new consonant clusters in English. *Poznan Studies in Contemporary Linguistics* 37:71-94.

Gnanadesikan, Amalia. 1997. Phonology with Ternary Scales. Ph.D. dissertation. University of Massachusetts at Amherst.

Gnanadesikan, Amalia. 2004. Markedness and faithfulness constraints in child phonology. In Kager, Pater, and Zonneveld (eds.). *Constraints in Phonological Acquisition*. pp. 73-108.

Göksel, Asli and Celia Kerslake. 2005. *Turkish: A Comprehensive Grammar*. London and New York: Routledge.

Goldsmith, John. 1976. Autosegmental Phonology. Ph.D. dissertation, MIT.

Goldsmith, John. 2009. The syllable. Ms. University of Chicago. A chapter in the forthcoming the second volume of the *Handbook of Phonological Theory*.

Goldstein, Louis, Hosung Nam, Elliot Saltzman, and Ioana Chitoran. 2008. Coupled oscillator model of speech timing and syllable structure. Special session on Coordination of Speech Gestures. International Symposium on Phonetic Frontiers and Phonetics Congress of China. Beijing.

Gouskova, Maria. 2004. Relational hierarchies in Optimality Theory: The case of syllable contact. *Phonology* 21: 201-50.

Greenberg, Joseph H. 1950. The patterning of root morphemes in Semitic. *Word* 6: 162–181.

Guy, Gregory R. 1980. Variation in the group and the individual: The case of final stop deletion. In William Labov (ed.). *Locating Language in Time and Space*. New York: Academic Press. pp. 1-36.

Hall, Nancy. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23: 387-429.

Hall, Nancy. 2007. R-dissimilation in English. Ms. University of Haifa.

Hammond, Michael. 2004. Gradience, phonotactics, and the lexicon in English phonology. *International Journal of English Studies* 4: 1–24.

Hansson, Gunnar Ólafur. 2001. The phonologization of production constraints: Evidence from consonant harmony. Paper presented at the 37th Meeting of the Chicago Linguistic Society.

Hansson, Gunnar Ólafur. 2007. On the evolution of consonant harmony: The case of secondary articulation agreement. *Phonology* 24: 77-120.

Hay, Jennifer, Janet Pierrehumbert, and Mary Beckman. 2004. Speech perception, well-formedness and the statistics of the lexicon. In Local, J., R. Ogden and R. Temple (eds.). *Phonetic Interpretation: Papers in Laboratory Phonology VI.* Cambridge University Press. pp. 58-74.

Hayes, Bruce. 1986. Inalterability in CV phonology. *Language* 62: 321-351.

Hayes, Bruce. 1997. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In John McCarthy (eds.). *Optimality Theory in Phonology*. Malden, Massachusetts: Blackwell Publishing Ltd. pp. 290-309.

Hayes, Bruce. 1998. Gradient well-formedness in Optimality Theory. Ms. University of California at Los Angeles.

Hayes, Bruce, Robert Kirchner and Donca Steriade. 2004. *Phonetically Based Phonology*. Cambridge: Cambridge University Press.

Hayes, Bruce and Zsuzsa Cziraky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23: 59-104.

Hayes, Bruce and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379-440.

Hooper, Joan Bybee. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In W. Christie (ed.). *Current Progress in Historical Linguistics*. North Holland.

Hume, Elizabeth. 1996. Coronal consonant, front vowels parallels in Maltese. *Natural Language and Linguistic Theory* 14-1: 163-203.

Hume, Elizabeth. 2003. Language specific markedness: The case of place of articulation. A talk given at Seoul conference. *Studies in Phonetics, Phonology and Morphology*.

Hume, Elizabeth and Georgios Tserdanelis. 2002. Labial unmarkedness in Sri Lankan Portuguese creole. *Phonology* 19: 441-458.

Hume, Elizabeth and Ilana Bromberg. 2005. Predicting epenthesis: An information-theoretic account. Ms.

Inkelas, Sharon and Cheryl Zoll. 2005. *Reduplication: Doubling in Morphology*. Cambridge University Press.

Ito, Chiyuki. 2006. Gradient morpheme structure and co-occurrence restrictions in Korean monosyllabic Stems. Ms. Tokyo University of Foreign Studies, ILCAA.

Itô, Junko and Armin Mester. 1995. The core-periphery structure of the lexicon and constraints on reranking. *Papers in Optimality Theory: University of Massachusetts Working Papers in Linguistics* 18: 181-210.

Itô, Junko and Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.). *The Handbook of Japanese Linguistics*. Malden, Massachusetts: Blackwell. pp. 62-100.

Jespersen, Otto. 1904. *Lehrbuch der Phonetik*. Leipzig & Berlin: Teubner.

Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles*, *Part VI: Morphology*. London: George Allen and Unwin Ltd.

Jun, Jongho. 1995. Perceptual and Articulatory Factors in Place Assimilation: An Optimality Theoretic Approach. Ph.D. dissertation. University of California at Los Angeles.

Jun, Jungho. 2004. Place assimilation. In Bruce Hayes, Robert Kirchner, and Donca Steriade (eds.). *Phonetically-based Phonology*. Cambridge University Press. pp. 58-86.

Jun, Jongho. 2006. Variable affix position in Korean partial reduplication. Handout presented at LSA 2006. Albuquerque. January 2006.

Jun, Jongho. 2007. Variation in Korean stem-final obstruents. Ms. Seoul National University.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kang, Eungyeong. 2002. Edge-integrity and the syllable structure in Korean. *Language, Information, and Computation: Proceedings of the 16th Pacific Asia Conference*. January 31 – February 2. Jeju, Korea. pp. 135-146.

Kang, Eungyeong. 2004. Edge-Demarcation in Phonology. Ph.D. dissertation. Cornell University.

Kang, Eungyeong. 2005. Consonant copy in Korean. *Studies in Phonetics, Phonology and Morphology* 11-2: 3-20.

Kang, Ongmi. 1998. A correspondence theoretic analysis on Korean partial reduplication. *Studies in Phonetics, Phonology, and Morphology* 4: 1-17. [in Korean]

Kang, Ongmi. 1999. A Correspondence analysis on hiatus resolution in Korean, *Studies in Phonetics, Phonology, and Morphology* 5: 1-24.

Kang, Ongmi. 2003. *Korean Phonology [Hankuke Umwunlon]*. Seoul: Thaehaksa

Kang, Yoonjung. 2002. Frequency of use and analogical change in Korean nouns: A case against abstract representation. Poster presented at NELS. MIT.

Kang, Yoonjung. 2007. Frequency effect and regularization in Korean nouns. Handout from the Workshop on Variation, Gradience and Frequency in Phonology. July 6-8, Stanford.

Kaun, Abigail R. 2004. The typology of rounding harmony. In Bruce Hayes, Robert Kirchner and Donca Steriade (eds.). *Phonetically Based Phonology*. Cambridge: Cambridge University Press. Chapter 4.

Kawahara, Shigeto, Hajime Ono, and Kiyoshi Sudo. 2006. Consonant cooccurrence restrictions in Yamato Japanese. Paper presented to the *Japanese/Korean Linguistics* 14. CSLI Publications. pp. 27-38.

Kean, Mary-Louise. 1992. Markedness: An overview. In William Bright (ed.). *International Encyclopedia of Linguistics*, Vol. 2. New York: Oxford University Press. pp. 390-391.

Kennard, Catherine Hicks. 2004. Copy but don't repeat: The conflict of dissimilation and reduplication in the Tawala durative. *Phonology* 21: 303-323.

Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.

Kessler, Brett. 2001. *The Significance of Word Lists*. Stanford: CSLI Publications.

Kessler, Brett and Rebecca Treiman. 1997. Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language* 37: 295-311.

Kim, Hyung-Soo. 2003. A new look at partial reduplication in Korean. *Studies in Phonetics, Phonology, and Morphology* 9-2: 353-380.

Kim, Hyung-Soo. 2009. The full-to-partial reduction in Korean and Turkish reduplication. *Linguistic Research* 26-2: 121-148.

Kim, Soohee. 1996a. Emergence of the unmarked: Korean partial reduplication. *CLS* 32, Main Session. pp. 179-189.

Kim, Soohee. 1996b. Emergence of the markedness constraint NO-LAR and its split behavior in the reduplicative domain. *Proceedings of WECOL.*

Kim, Soohee. ______. Reduplication and the mimetic stratum in Korean. Ms. University of Washington.

Kim, Sun-Hoi. 2001. Multiple patterns in reduplication. *Eoneohag* 29: 109-131.

Kim, Young-Suk. 2007. Phonological awareness and literacy skills in Korean: An examination of the unique role of body-coda units. *Applied Psycholinguistics* 28: 69-94.

Kingston, John and Paul de Lacy. 2006. Synchronic explanation. Ms. University of Massachusetts at Amherst and Rutgers University.

Kingston, John and Randy Diehl. 1994. Phonetic knowledge. *Language* 70: 419-453.

Kirchner, Robert. 2000. Geminate inalterability and lenition. *Language* 76: 509-545.

Labov, William. 1989. The child as linguistic historian. *Language Variation and Change* 1: 85-97.

Laeufer, Christinane. 1995. Effects of tempo and stress on German syllable structure. *Journal of Linguistics* 31: 227-266.

Leben, William. 1973. Suprasegmental Phonology. PhD dissertation. MIT.

Lee, Minkyung. 2006. Gyeongsang Korean /n/-insertion revisited. *Studies in Phonetics, Phonology and Morphology* 12-3: 623-641.

Lee, Yongeun. 2006. Sub-syllabic Constituency in Korean and English. Ph.D. dissertation. Northwestern University.

Lembaga Basa & Sastra Sunda.1985. *Kamus Umum Basa Sunda, Cetatan Kelima*. Bandung, Indonesia: Penerbit Tarate Bandung.

Lombardi, Linda. 1991. Laryngeal Features and Laryngeal Neutralization. Ph.D. dissertation. University of Massachusetts at Amherst.

Lombardi, Linda. 2002. Coronal epenthesis and markedness. *Phonology* 19: 219-251.

Łubowicz, Anna. 2005. Locality of Conjunction. Paper presented at the 24th West Coast Conference on Formal Linguistics. pp. 254-262.

MacEachern, Margaret. 1997. Laryngeal similarity effects in Quechua and Aymara. *Proceedings of the West Coast Conference on Formal Linguistics* 15: 337-351.

MacEachern, Margaret. 1999. *Laryngeal Cooccurrence Restriction*. New York: Garland.

Martin, Andrew T. 2007. The Evolving Lexicon. PhD dissertation. University of California, Los Angeles.

McCarthy, John J. 1979. Formal Problems in Semitic Phonology and Morphology. Ph.D. dissertation. MIT.

McCarthy, John J. 1981. A prosodic theory of non-concatenative morphology. *Linguistic Inquiry* 12: 373-418.

McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17: 207-263.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45: 84–108.

McCarthy, John J. 1989. Linear order and phonological representation. *Linguistic Inquiry* 20: 71-100.

McCarthy, John J. 1993. A case of surface constraint violation. *Canadian Journal of Linguistics* 38-2: 169-195.

McCarthy, John. J. 1994. The phonetics and phonology of semitic pharyngeals. In P. Keating (ed.). Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III. Cambridge: Cambridge University Press. pp. 191-233.

McCarthy, John J. 2007. Consonant harmony via correspondence: Evidence from Chumash. Bateman, Leah, Michael O'Keefe, Ehren Reilly, and Adam Werle (eds.). *Papers in Optimality Theory III: University of Massachusetts Occasional Papers 32*. ROA 879.

McCarthy, John J and Alan Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. In M. Gonzàlez (ed.). *Proceedings of the North East Linguistic Society* 24. pp. 333-379.

McCarthy, John and Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.). *University of Massachusetts Occasional Papers 18: Papers in Optimality Theory*. pp. 249-384.

Mester, Arnim. 1986. Studies in Tier Structure. Ph.D. dissertation, University of Massachusetts, Amherst.

Minkova, Donka. 2000. Reduplication and syllable well-formedness: The criss-crossing of phonology and semantics. Abstract of a talk given at International Conference on English Historical Linguistics.

Minkova, Donka. 2002. Ablaut reduplication in English: The criss-crossing of prosody and verbal art. *English Language and Linguistics* 6-1: 133-169.

Morelli, Frida. 1999. The Phonotactics and Phonology of Obstruent Clusters in Optimality Theory. Ph.D. dissertation, University of Maryland at College Park.

Morén, Bruce. 2003. The Parallel Structures Model of Feature Geometry. *Working Papers of the Cornell Phonetics Laboratory* 15. Cornell, Ithaca. pp. 194-270.

Myers, J. Scott. 1994. OCP effects in Optimality Theory. Rutgers Optimality Archive 6. http://ruccs.rutgers.edu/roa.html.

Myers, James. 2007. Frequency effects in lenition and the challenge of lexicalized markedness. Handout from the Workshop on Variation, Gradience and Frequency in Phonology. July 6-8, Stanford.

Nagy, Naomi, and Reynolds, Bill. 1997. Optimality Theory and variable word-final deletion in Faetar. *Language Variation and Change* 9: 37-55.

Nevins, Andrew Ira. 2005. Overwriting does not optimize in nonconcatenative morphology. *Linguistic Inquiry* 36-2: 275-287.

Nevins, Andrew and Bert Vaux. 2003. Metalinguistic, shmetalinguistic: The phonology of shm-reduplication. In *CLS* 39.

Odden, David. 1986. On the role of the Obligatory Contour Principle in phonological theory. *Language* 62: 353-383.

Odden, David. 1988. Anti-antigemination and the OCP. *Linguistic Inquiry* 19: 451-475.

Padgett, Jaye. 1995. *Stricture in Feature Geometry*. Stanford: CSLI Publications.

Park, Hae-Il. 2000. English Reduplication: An Optimality Theoretic Account. MA thesis. Seoul National University.

Park, Hansang. 2003. Productivity of Korean partial reduplication. *Studies in Phonetics, Phonology, and Morphology* 9-2: 393-414.

Parker, Stephen G. 2002. Quantifying the Sonority Hierarchy. PhD dissertation. University of Massachusetts at Amherst.

Pater, Joe. 2000. Non-uniformity in English stress: The role of ranked and lexically specific constraints. *Phonology* 17-2: 237-274.

Pater, Joe, Christopher Potts, and Rajesh Bhatt. 2007. Harmonic Grammar with linear programming. Ms, University of Massachusetts at Amherst.

Patrick, Peter L. 1992. Linguistic Variation in Urban Jamaican Creole: A Sociolinguistic Study of Kingston, Jamaica. Ph.D. dissertation. University of Pennsylvania.

Phillips, Betty. 1980. Old English an/on: A new appraisal. *Journal of English Linguistics* 14: 20-23.

Phillips, Betty. 1983. Lexical diffusion and function words. *Linguistics* 21: 487-499.

Phillips, Betty. 1984. Word frequency and the actuation of sound change. *Language* 60: 320-342.

Phillips, Betty S. 1999. The mental lexicon: Evidence from lexical diffusion. *Brain and Language* 68:104-09.

Phillips, Betty. 2001. Lexical diffusion, lexical frequency, and lexical analysis. In Joan Bybee and Paul Hopper (eds.). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins. pp. 123-136.

Phillips, Betty. 2007. Frequency effects in the lexical diffusion of phonological change. Handout from the Workshop on Variation, Gradience and Frequency in Phonology. July 6-8, Stanford.

Pierrehumbert, Janet. 1993. Dissimilarity in the Arabic verbal roots. *Proceedings of the North East Linguistics Society* 23. pp. 367-381.

Pierrehumbert, Janet. 1994. Syllable structure and word structure. In P. Keating (ed.). *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*. Cambridge: Cambridge University Press. pp. 168-190.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee and P. Hopper (eds.). *Frequency Effects and the Emergence of Lexical Structure*. Amsterdam: John Benjamins. pp. 137-157.

Pierrehumbert, Janet. 2002. Word-specific phonetics. *Laboratory Phonology VII*. Berlin: Mouton de Gruyter. pp. 101-139.

Pierrehumbert, Janet. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*. 46: 115-154.

Prasada, Sandeep and Steven Pinker. 1993. Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* 8: 1-56.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. MIT Press.

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in Generative Grammar*. Blackwell Publishing.

Pulleyblank, Douglas. 1988. Vocalic underspecification in Yoruba. *Linguistic Inquiry* 19: 233-270.

Reynolds, Bill. 1994. Variation and Phonological Theory. Ph.D. dissertation, University of Pennsylvania.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80: 475-531.

Rubach, Jerzy. 2000. Glide and glottal stop insertion in Slavic languages: A DOT analysis. *Linguistic Inquiry* 31-2: 271-317.

Santa Ana, Otto. 1991. Phonetic Simplification Processes in the English of the Barrio: A Cross-Generational Sociolinguistic Study of the Chicanos of Los Angeles. Ph.D. dissertation, University of Pennsylvania.

Sejong Balanced Corpus. 2007. Seoul, Korea: The National Institute of the Korean Language.

Selkirk, Elizabeth. 1982. Syllables. In Harry van der Hulst and N. Smith (eds.). *The Structure of Phonological Representations*. Dordrecht: Foris. pp. 337-383.

Selkirk, Elizabeth. 1984. On the major class features and syllable theory. In Mark Aronoff and Richard T. Oehrle (eds.). *Language Sound Structures*. Cambridge, MA: MIT Press. pp. 107-136.

Selkirk, Elizabeth. 2004. The prosodic structure of function words. In John McCarthy (ed.). *Optimality Theory in Phonology*. Blackwell. pp. 464-482.

Sereno, Joan A. and Allard Jongman. 1997. Processing of English inflectional morphology. *Memory and Cognition* 25-4: 425-437.

Shatzman, Keren B. and René Kager. 2007. A role for phonotactic constraints in speech perception. *ICPhS* XVI: 1409-1412.

Smith, Jennifer L. 2003. Onset sonority constraints and subsyllabic structure. Revised version of paper presented at the 9th International Phonology Meeting. University of Vienna, November 3, 2002. Rutgers Optimality Archive #608.

Smolensky, Paul. 1995. On the internal structure of the constraint component Con of UG. Ms. UCLA. ROA-86.

Sohn, Hyang-Sook. 2008. Phonological contrast and coda saliency of sonorant assimilation in Korean. *Journal of East Asian Linguistics* 17: 33-59.

Stemberger, Joseph P. and MacWhinney, Brian J. 1986. Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition* 14: 17–26.

Stemberger, Joseph P. and MacWhinney, Brian J. 1988. Are inflected forms stored in the lexicon? In M. Hammond and M. Noonan (ed.). *Theoretical Morphology*. New York: Academic Press.

Steriade, Donca. 1982. Greek Prosodies and the Nature of Syllabification. Ph.D. dissertation. MIT.

Steriade, Donca. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5: 73-155.

Steriade, Donca. 1995. Positional neutralization. Ms. University of California at Los Angeles.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume and Keith Johnson (eds.). *The role of speech perception in phonology*. San Diego: Academic Press. pp. 219-250.

Struijke, Caro. 2000. Existential Faithfulness: A Study of Reduplicative TETU, Feature Movement, and Dissimilation. Ph.D. dissertation. University of Maryland at College Park.

Suh, Yunju. 2007. Consonant+labiovelar glide combinations in Spanish and Korean. Proceedings of ICPhS 16. Saarbrücken, Germany.

Suzuki, Keiichiro. 1999. Identity avoidance vs. identity preference: The case of Sundanese. Paper presented at the 73rd Annual Meeting of the Linguistic Society of America. Los Angeles.

Treiman, Rebecca. 1983. The structure of spoken syllables: Evidence from novel word games. *Cognition* 15: 49-74.

Treiman, Rebecca, Brett Kessler, Stephanie Knewasser, Ruth Tincoff, and Margo Bowman. 2000. English speakers' sensitivity to phonotactic patterns. In Michael B. Broe and Janet B. Pierrehumbert (eds.). *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press. pp. 269-282.

Truckenbrodt, Hubert. 2007. The syntax-phonology interface. In Paul de Lacy (ed.). *The Cambridge Handbook of Phonology*. pp. 435-456.

Urbanczyk, Suzanne. 2006. Reduplicative form and the root-affix asymmetry. *Natural Language and Linguistic Theory* 24-1: 179-240.

Urbanczyk, Suzanne. 2007. Reduplication. In Paul de Lacy (ed.). *The Cambridge Handbook of Phonology*. Cambridge University Press. pp. 473-493.

Ussishkin, Adam. 2006. Affix-favored contrast inequity and psycholinguistic grounding for non-concatenative morphology. *Morphology* 16: 107-125.

Ussishkin, Adam. 2007. Morpheme Position. In Paul de Lacy (ed.). *The Cambridge Handbook of Phonology*. Cambridge University Press. pp. 457-472.

Vaux, Bert. 2003. Consonant epenthesis and hypercorrection. Ms. Harvard University.

Vennemann, Theo. 1988. *Preference Laws for Syllable Structure*. Berlin: Mouton de Gruyter.

Vitevitch, Michael. S. and Paul A. Luce. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers* 36: 481-487.

Walter, Mary Ann. ______. Hypocoristic formation in Cairene Arabic. Ms. MIT.

Walter, Mary Ann. 2007. Repetition Avoidance in Human Language. Ph.D. dissertation, MIT.

Wedel, Andrew. 1999. Turkish emphatic reduplication. *Linguistics Research Center: Phonology at Santa Cruz.*

Wedel, Andrew. 2000. Preservation of contrast in Turkish reduplicative paradigms. Ms. University of California, Santa Cruz.

Wedel, Andrew. 2007. Feedback and regularity in the lexicon. *Phonology* 24: 147-185.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner and Donca Steriade (eds.). *Phonetically Based Phonology*. Cambridge: Cambridge University Press. pp. 34-57.

Yip, Moira. 1988. The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19: 65-100.

Yip, Moira. 1989. Feature geometry and co-occurrence restrictions. *Phonology* 6: 349-374.

Yip, Moira. 1995. Repetition and its avoidance: The case of Javanese. In K. Suzuki and D. Elzinga (eds.). *Proceedings of South Western Optimality Theory Workshop 1995 Arizona Phonology Conference Vol 5: U. Of Arizona, Department of Linguistics Coyote Papers.* Tucson AZ. pp. 238-262.

Yip, Moira. 1998. Identity avoidance in phonology and morphology, In Steven G. Lapointe, Diane K. Brentari, and Patrik M. Farrell (eds.). *Morphology and its Relation to Phonology and Syntax*. Stanford: CSLI. pp. 216-246.

Yip, Moira. 2000. Segmental unmarkedness versus input preservation in reduplication. In Linda Lombardi (ed.). *Segmental Phonology in Optimality Theory*. Cambridge: Cambridge University Press.

Yoon, Yeo Bom and Bruce Derwing. 1994. Sound similarity judgments for Korean words by Korean and English speakers. University of Toronto: Toronto Working papers in Linguistics.

Yoon, Yeo Bom and Bruce Derwing. 2001. A language without a rhyme: Syllable structure experiments in Korean. *Canadian Journal of Linguistics* 46: 187-237.

Yu, Alan. 1999. Dissimilation and Allomorphy: The case of Turkish Emphatic reduplication. Ms. University of California, Berkeley.

Zaleska, Joanna. 2008. Glide and glottal stop insertion in Czech: A non-serialist account. Presented at Generative Linguistics in Poland 6, April 5 – 6.

Zubritskaya, Katya. 1997. Mechanism of sound change in Optimality Theory. *Language Variation and Change* 9: 121-148.

Zuraw, Kie. 2002a. Vowel reduction in Palauan reduplicants. *Proceedings of AFLA 8: The Eighth Meeting of the Austronesian Formal Linguistics Association. MIT Working Papers in Linguistics* 44. pp. 385-398.

Zuraw, Kie. 2002b. Aggressive reduplication. *Phonology* 19: 395-439.

Zuraw, Kie. 2003a. Probability in language change. In Rens Bod, Jennifer Hay, and Stefanie Jannedy (eds.). *Probabilistic Linguistics*. Cambridge, MA: MIT Press. pp. 139-176.

Zuraw, Kie. 2003b. Optimality Theory in linguistics. In Michael Arbib (ed.). *Handbook of Brain Theory and Neural Networks*, 2nd edition. Cambridge, MA: MIT Press. pp. 819-822.

Zuraw, Kie. 2009. Frequency influences on rule application within and across words. *Proceedings of CLS (Chicago Linguistic Society)* 43-2. pp. 283-309.

Zuraw, Kie and Kevin Ryan. 2007. Frequency influences on phonological rule application within and across words. Handout from the Workshop on Variation, Gradience and Frequency in Phonology. July 6-8, Stanford.