

The effect of word-frequency on consonant gradation in Canadian-Finnish bilingual speech

Pauliina Saarinen

University of Victoria
pauliina@uvic.ca

ABSTRACT

This paper investigates the effects of word-frequency on application of consonant gradation in non-native Finnish speech. Based on the assumption that the same processes that shape native speech are at work also in non-native speech, it is hypothesized that a non-native speaker should produce a more native-like output with frequent lexical items. An experiment was designed to test the hypothesis. Two separate analyses were conducted. In one, the effect of word-frequency, as defined by a Finnish frequency-corpus, was considered and, in the other, the effect of word-familiarity was examined. In the former analysis, word-frequency was found not to have a significant effect on whether or not gradation applies. However in the latter analysis, word-familiarity, defined as individual frequency, was found to have effect on application of gradation. Unfamiliar words were less likely to undergo gradation than familiar words. The findings support the prediction that loss of gradation affects least frequent words first by analogical leveling. It is argued that loss of gradation is a means to reduce grammatical complexity.

Keywords: Finnish; consonant gradation; variation; frequency; word-familiarity.

1 Introduction

In literature on usage-based grammar, it has been established that human language is an outcome of an individual's cognitive organization of linguistic experience (Pierrehumbert 2001; Bybee 2005). The mind's capacity to generalize, discern patterns and build schemas from the accumulated linguistic information is central to the concept of human endowment for language. Memory has an astonishing ability to keep record of one's encounters with word forms and grammatical structures. Frequently encountered patterns become conventionalized in the speaker's grammar, while less frequent patterns do not (Bybee 2005). This individual linguistic experience is evidenced in a speaker's phonetic output as speaker grammar (Pierrehumbert 2000). Naturally, processes which shape speaker grammar in the first language are also at work in non-dominant language. This paper examines the effects of

lexical frequency in non-dominant¹ language, in which the idiosyncratic linguistic experience is even more pronounced. Since word-frequency has been found to affect the shape of words and word strings stored in speakers' mental lexicons (e.g. Hooper 1976; Bybee 2002; Phillips 1984), non-dominant speaker grammar should also exemplify these frequency effects. For example, phonetic change in native speech has been found to progress more quickly in high-frequency items, which typically is evidenced as reductive changes (Bybee 2001; Krug 1998). Like any motor activity, speech becomes more fluid when repeated often (Boyland 1996, in Bybee 2001). With a lexicon less extensive than that of a native speaker and limited exposure to native-like language, a non-dominant speaker should therefore produce more native-like output with frequent lexical items. This hypothesis was tested with the production of consonant gradation in the speech of a non-dominant speaker of Finnish. An experiment was designed to answer if the variance, i.e., the loss of gradation (henceforth GL), which was observed in the subject's speech, is related to word-frequency or not. Word-frequency, as defined by written Finnish word-frequency corpora, had no effect on application of consonant gradation. A relatively high number of standard forms were attested regardless of the frequency of the words. However, when the same data was analyzed in respect to *word-familiarity*, which is defined as a speaker's individual frequency, consonant gradation was found to apply to familiar words significantly more often than to unfamiliar ones. It is argued that *word-familiarity*, rather than word-frequency, accounts for the idiosyncratic nature of mental lexicons and should be taken as a measure of frequency for the individual. As such, results support the prediction that least frequent words are affected first by analogical leveling.

2 Background

In the debate over the diffusion of sound change through the lexicon, the Neogrammarian view has dominated until very recently. According to this view, diachronic sound change is regular, affecting the whole lexicon whenever the phonetic environment is encountered (Phillips 1984; Bybee 2002). However, already in 1885, Schuchardt observed that sound change does not necessarily diffuse through the lexicon at once, but some lexical items might be affected before others:

“The greater or lesser frequency in the use of individual words that plays such a prominent role in analogical formation is also of great importance for their phonetic transformation, not within rather small differences, but within significant ones. Rarely used words drag behind; very frequently used ones hurry ahead. Exceptions to the sound laws are formed in both groups.”(58) [in Bybee 2002, translated version found in Venneman & Wilbur (1972) Schuchardt, the Neogrammarians and the Transformational Theory of Phonological Change. Frankfurt: Athenaum]

¹ Non-dominant language in this paper refers to a language spoken by a person whose dominant language is other than Finnish. The participant was born in Canada, acquired Finnish as the first language at home, and speaks English as a dominant language.

He intuited that frequently used lexical items change faster than less frequently used items. Several studies (Bybee 2000; Jurafsky et al. 2001, and many others) have proven him correct. The next section looks at this evidence more closely and shows that diffusion of sound change through lexicon can indeed be gradual, not only lexically but also phonetically, and that it may progress in two directions, affecting either frequent or infrequent items first. Frequent items, having more use by numerous people, exhibit more innovative pronunciations overall, which renders these lexical items more prone to sound change. The Neogrammarian view is further challenged by the discovery that the gradualness does not only affect lexical items, but also individual segments (Loziewicz 1992). Phillips (1984, 2001) went on to refine the hypothesis that the type of sound changes can be used as a diagnostic for the direction of diffusion of sound change.

2.1 *The two directions of sound change*

2.1.1 Sound change affecting *frequent items* first

Anybody learning a new language can attest to the fact that, with practice, articulation requires less and less effort and that speech gestures become smaller and smoother while the message gets clearer. High frequency therefore has a reductive effect on both articulatory gestures as well as the shape of words and word strings, in non-native or native speech alike. Several studies have recorded that phonetic reduction, stemming from automation of speech, evidences itself as lenition, deletion, or neutralization and, thus, affects frequently used words or phrases first (Hooper 1976; Bybee 2001, 2002, 2005; Phillips 1984). Less frequently used words and phrases are more resistant to change. They may change at a slower pace or retain their full phonetic form. Among such studies is Fidelholtz's (1975) investigation of English vowel reduction in initial syllables of words, which are stressed on the second syllable. He discovered that the more frequent *mistake* has a reduced vowel /ə/ in its first syllable, while the less frequent *mistook* has a full vowel /ʊ/. The more common word *astronomy* has a /ə/ in the initial syllable, while infrequent *gastronomy* has an unreduced /æ/. Hooper (1976) also found that the medial /ə/ in frequent words such as *nursery*, *celery* and *memory* is deleted more often than it is in less frequent *cursor*, *artillery* and *armory*. Deletion is interpreted as stemming from well-rehearsed neuro-motor routines in frequent lexical items (Mackay 1978, in Phillips 1984). Easing of articulation is arguably also behind the deletion of English final /t/ and /d/ following a consonant; they are deleted more often in high-frequency words (Bybee 2002; Jurafsky et al. 2001). Several studies (Browman et al. 1992, in Bybee 2002; Mowrey et al. 1995) have found that speech gestures tend to overlap more due to anticipatory articulation. In English, the consonants following a homorganic nasal, like in *bomb* or *gang*, are deleted systematically, making [mb] or [ng] clusters non-existent in the language. The same erosion may be working on /nd/, as in *grand* (Bybee 2002). Thus, articulatory gestures deemed unnecessary or extreme become smaller in magnitude or are deleted altogether (Mackay 1978, in Phillips 1984). According to Phillips (1984), reductive sound changes of the kind discussed above have physiological motivations due to high-frequency use.

As shown above, sound change can be lexically gradual. It may work its way through the lexicon, eventually affecting all the eligible words. Alternatively, it may never be complete due to idiosyncrasies in frequency of lexemes in individual speakers (Bybee 2002, footnote

1). Sound change in progress is often observable in having two or more competing pronunciations for a word (Phillips 1984, 320), but when the sound change does happen, it is taken to be an abrupt replacement of a phoneme with another one (Labov 1994: 542). However, this may not always be case. Losiewicz (1992) found evidence that sound change may also be phonetically gradual. The English monomorphemic final /t/ and /d/, as in *just*, or *child*, which are prone to deletion (i.e. are in the context for deletion) are shorter than the /t/ and /d/ of the regular past tense. She also found that the duration of the regular past tense /t/ and /d/ is longer in low-frequency verbs than in verbs of high-frequency (ibid.) The findings are supported by the fact that in low-frequency verbs the regular past tense endings are very unlikely to be deleted and, conversely, more prone to deletion in high-frequency verbs (Bybee 2000). Since the stops are in the context for deletion in high-frequency verbs, their duration is also shorter. The findings indicate that the duration of a segment may decrease first before it is completely deleted, thus providing evidence for phonetic gradualness of sound change.

2.1.2 Sound change affecting *infrequent items* first

This section presents examples of frequency effect progressing in the opposite direction, affecting infrequent items first and discusses possible causes for it. Unlike the cases presented in the section above, infrequent items should be more resistant to physiological sound changes to ease articulation, as they do not benefit from rehearsal to the same extent. Thus, sound changes affecting the least frequent items first should have other impetus. Phillips (1984) proposes that sound changes of this sort have cognitive reasons.

Hooper (1976) demonstrated how infrequent irregular verbs, such as *weep* in the past tense, tend to regularize into *weaped*, while frequent verbs, such as *keep*, retain their irregular paradigms. Although irregular, the word *keep* is so ingrained in the speakers' lexicons thanks to its high-frequency use, that *keeped* would sound ungrammatical to a native ear. Infrequent irregular verbs, on the other hand, do not get established in the speakers' grammar due to less use and, thus, are likely to undergo regularization. Exceptional patterns are hard to keep in mind, in which case the stronger tendency in the language wins over the weaker one, a process also known as *analogical leveling*.

Another example of sound change affecting infrequent items comes from Phillips (1984). The Old English diphthong *eo* changed into a mid front rounded vowel *ö* with both a long and a short version in use in the 11th and 12th centuries. In some dialects they were retained until the 14th century, but in Lincolnshire, they changed on to become unrounded and merged with *e(:)*. Interestingly, Ormulum, a manuscript written around 1200AD, captures the change in progress, showing two different spellings for one word, such as *deop* and *dep* for 'deep'. Phillips (Ibid.) discovered that in nouns and verbs, the low-frequency words were more likely to have the innovative spelling with the unrounded vowel. There were no other front rounded vowels in English at the time, so the majority of words would have front vowels unrounded. Thus, *ö* would have had to be learned as a special case. However, front rounded vowels are challenging for children to perceive and learn, and as a result, they are usually acquired later than unrounded vowels (Gilbert et al. 1975, in Bybee 2002). Arguably then, high-frequency use would have aided the acquisition of the special case *ö*, and with less frequent words

children opted for the unrounded equivalent, a pattern that was becoming more natural to them (*ibid.*).

So far, we have established that sound change may be both lexically and phonetically gradual and sometimes it never affects the whole lexicon. These findings suggest that phonetic detail is stored in the lexicon, thus challenging generative views to phonetics and phonology. According to the latter, phonetic detail is applied on an item by means of rules and constraints after its retrieval from the lexicon. The phonetic shape of words is strictly the outcome of what these rules and constraints dictate, treating all eligible words in the same way. However, word-frequency does not fit in the workings of these models although, as shown by the examples presented above, it conditions allophony. Thus, we need an approach, which is suited in conceptualization of word-frequency, another force behind variation in phonetic outcome.

2.2 *Exemplar Dynamics*

This model, as presented in Pierrehumbert (2001), posits that the memory stores an individual's accumulated linguistic experience as exemplars. Linguistic experience means an individual's encounters with full word forms, morphemes, phonemes, word-strings, or prosodic information of which a memory is formed. Whatever category the memory belongs to, it merges with an old one if the difference between the two is too small to notice. Conversely, a token with a noticeable difference is stored in close proximity as a separate category. The distance between the tokens represents similarity or dissimilarity between them. Thus, clusters of related memories emerge. A category label representing the whole cluster is drawn from the essential parameters of the exemplars stored in the mnemonic assemblage. Each category represents parameter space where the properties of a new token are determined and which determines the position of it within the cluster. An exemplar gains strength through activation, i.e., recency and frequency of use. The more recently and frequently a token is activated, the higher its activation level is. Activation level of tokens already stored in the category affect the categorization of a new token entering the same cluster; a new token is more likely to be analyzed as another example of a previously stored token with a high activation level. The number of tokens and their strength comprise the forces that fix the position of a new token in the parameter space. An exemplar may be simultaneously associated with multiple categories dedicated to different aspects of linguistic information. Sparse categories of exemplars represent clusters that are called for less frequently. Similarly, they receive fewer "hits", so their strength tends to be less. These categories are more prone to be forgotten. Exemplar neighborhoods overlap at the outer boundary, allowing room for dialectal or random speaker variation.

The Exemplar model makes it possible for a memory that is deviant from the targeted form to gain strength through frequent or recent activation and get selected. A situation like this could arise in non-dominant speech in which an individual's idiosyncratic grammar is well established. Sociolinguistic contexts for non-dominant speech often reinforce grammar that is deviant from the standard. Moreover, it is the person him or herself whom one hears most often, thereby perpetuating deviant forms (S. Bird, p.c., February 18, 2007). Indeed, in the next paragraphs this model is applied to this very situation, examining the application of consonant gradation in non-dominant Finnish speech. However, before presenting the

experiment in section 3, consonant gradation in Standard Finnish is first discussed in more detail in section 2.2. and variation in its application in non-dominant speech in Section 2.3.. Section 4 presents the results, section 5 the discussion and, finally, section 6 concludes this exposition.

2.3 Consonant Gradation in Finnish²

Consonant gradation is a form of morphophonological process in which the alternation of stem consonants is triggered by inflectional and derivational suffixation. It is a relatively commonly encountered phenomenon in Finnish; of the 1000 most frequent words, a third is subject to consonant gradation. It affects voiceless stops /p, t, k/ when they are preceded by a vowel or liquids. Voiced stops /b, g/ are affected in some slang words. Gradation evidences itself as an alternation of either the length of the stop, i.e. quantitative, as in *rätti-rätin* 'rag-GEN.SG', or the quality of it, as in *koti-kodin* 'home-GEN.SG', and in the case of single stop /k/, its deletion, as in *näky-näyn* 'sight-GEN.SG'. Most frequently affected consonant is /t/, while /p/ is affected least frequently. Table 1 presents consonants subject to quantitative gradation, i.e., shortening or lengthening, and Table 2 those subject to qualitative, leniting or strengthening, gradation.

Table 1

Quantitative Gradation; shortening or lengthening

| | | | |
|---|--------|--------|--------|
| 1 | pp~p | tt~t | kk~k |
| 2 | mpp~mp | ntt~nt | ηkk~ηk |
| 3 | lpp~lp | ltt~lt | lkk~lk |
| 4 | rpp~rp | rtt~rt | rkk~rk |
| 5 | bb~b | | gg~g |

Table 2

Qualitative Gradation; lenition or strengthening

| | | | |
|----|-------|-------|---------|
| 6 | mp~mm | nt~nn | ηk~ηη |
| 7 | lp~lv | lt~ll | lk~lj |
| 8 | rp~rv | rt~rr | rk~rj |
| 9 | | ht~hd | hk~hj |
| 10 | p~v | t~d | k~ - /v |

Gradation has two realizations, STRONG or WEAK. Gradation can be *direct* or *indirect* (i.e., *reverse*) depending on whether the basic form is STRONG or WEAK. In (1), *kukka-kukan* 'flower.NOM.SG-flower.GEN.SG', exemplifies direct gradation since the geminate, STRONG form, is also the basic form. The inflected form is WEAK, gradated form. Conversely, a situation where WEAK form is the basic form is called reverse gradation, as in (2). In this case, WEAK form is the basic form.

² From Hakulinen et al. (2004).

(1) Direct Gradation (STRONG-WEAK)

- a. kuk.ka-ku.kan ‘flower.NOM.SG-flower.GEN.SG’
 b. kier.tää-kier.rän ‘to go around- to go around.1SG’

(2) Reverse Gradation (WEAK-STRONG)

- a. ri.kas:rik.kaan ‘rich- rich.GEN.SG’
 b. ker.rata:ker.taan ‘to review-to review.1SG’

Generally, STRONG variant precedes an open syllable with no coda consonant, while WEAK form is usually found preceding a closed syllable. There are many exceptions to this rule, as in (2); STRONG forms are found preceding closed syllables.

2.4 Variation in Consonant Gradation in non-dominant Finnish speech

Similarities in the process of acquisition of first and second language (Lightbown & Spada 1999) justify the presupposition that the same pressures that shape speech in a dominant language are at work in non-dominant language. It follows that the effects of word-frequency should be observable also in non-dominant speech. This premise led to the present study where the language of a non-dominant speaker of Finnish is examined in respect to the application of consonant gradation. The consultant, a first generation Canadian-born non-dominant speaker of Finnish, uses standard-like gradated forms with some words, while leaving other words ungradated altogether. Table 3 presents some examples of the observed variation.

Table 3

Variation in Consonant Gradation in Canadian-Finnish

| Direct Gradation | (Basic form (SF)) Gradated form (SF) | Non-dominant Finnish | Gradation | Gloss |
|------------------|---|----------------------|-----------|--------------------|
| nt:nn | (ranta) rannalla | rannalla | yes | “on the beach” |
| | (englanti) englanniksi | englantiksi | no | “in English” |
| ŋk:ŋŋ | (helsinki) helsinkiissä | helsinkiissä | no | “in Helsinki” |
| lp:lv | (halpa) halvalla | halpalla | no | “at a cheap price” |
| rt:rr | (kerta) kerran | kerta | no | “one time” |
| | (siirtyä) siirryn | siirryn | yes | “I move (aside)” |

SF=Standard Finnish

Loss of gradation (GL) is unlikely to stem from phonetic motivations to ease articulation. In many cases, applying gradation would have that effect, as it tends to weaken the closure of stops. In cases where it produces a geminate, the consonant is just longer and not necessarily

doubly articulated (Hakulinen et al. 2004, 66). As seen in Table 3, with *nt:nn*³ gradation, the participant produced the standard form in one word (*rannalla*) and a non-standard form in another (*englantiksi*) although the consonants involved are the same in both instances. Since GL may occur in different words containing the same consonant combinations, positing lexical motivation for GL is particularly compelling. Could frequency be the motivation behind the variation in non-dominant Finnish speech? An experiment was set up to answer whether *loss of gradation (GL)* in non-dominant Finnish is affected by word-frequency. If so, two possibilities exist. One is that GL affects frequent items first, and the other is that less frequent words show GL first. It is predicted that GL is correlated with word-frequency, and that due to limited exposure to standard language, a non-native speaker produces more native-like output with frequent lexical items. In other words, the participant is expected to gradate frequent words more successfully than infrequent ones thanks to more repetition and exposure to them.

3 Methodology

3.1 Subject

One Canadian-born Finnish speaker participated in this study. Now 71, she speaks English as her dominant language and Finnish, which she acquired from her immigrant parents as L1. As a child, she was always spoken to in Finnish at home. All her education was in English. She visits relatives and friends in Finland about every five or six years, spending approximately two months in the country on each trip. At home, she does not speak Finnish regularly. She meets her friends and acquaintances from the Finnish community about once a month with whom she speaks Finnish. She reads and understands written Finnish in cards and letters received from her relatives in Finland.

3.2 Tokens

The CSC, or the Finnish Information Technology center⁴, a corpus listing Finnish words in newspaper texts according to descending frequency was used to compile two word lists. The test words were to exhibit direct consonant gradation only and the following frequency requirements. The tokens for the list representing frequently used words of the corpus were to be among the first 1000 most frequent words. 47 tokens were chosen. The frequency of the selected words ranged from 1315 to 136 times in every million words. The tokens

³ According to Hakulinen et al. (2004), *nt~nn* alternation is the second most frequent gradation type with 1432 instances in *Perussanakirja*, a comprehensive dictionary, after *tt~t* alternation which has approximately 2980 instances. *kk~k* alternation comes third with about 1220 instances (73).

⁴ The corpus is accessible at <http://www.csc.fi/kielipankki/aineistot/ssts/frq1/phtml>

representing less common words were to be among 3000 or less frequent words. 41 tokens were chosen for this list⁵. Their frequency ranged between 35 and 11 times in every million words. Distinction between the type of gradation, whether quantitative or qualitative, was not made, and the two lists compiled contained both types of gradations. Similarly, the words on the lists included many types of consonant combinations to ensure a big enough sample within the frequency requirements described above and to gain knowledge of the overall pattern of variation. Table 4 lists the consonant combinations included. The gradated forms of the 88 tokens were to be elicited from the speaker.

Table 4

Consonant Combinations used in the study

| Direct Gradation (STRONG~WEAK) | |
|--------------------------------|------------------------|
| Quantitative (shortening) | Qualitative (lenition) |
| pp~p | mp~mm |
| tt~t | nt~nn |
| kk~k | ŋk~ŋŋ |
| | lt~ll |
| | lk~l |
| | lk~lj |
| | rt~rr |
| | rk~r |
| | ht~hd |
| | p~v |
| | t~d |
| | k~- |

Table 5

Sample of the elicited words from the list representing high-frequency words

| Direct Gradation | Standard Finnish basic:gradated | Frequency/million Words (Rank) | Gloss/phrase elicited |
|------------------|---------------------------------|--------------------------------|------------------------------------|
| kk~k | viikko~viikon | 386 (100.) | "for a week" |
| | tarkka~tarkasti | 142 (856.) | "accurately" |
| | kirkko~kirkossa | 286 (411.) | "in the church" |
| nt~nn | ranta~rannalla | 140 (872.) | "on the beach" |
| | syntyä~synnyin | 588 (157.) | "I was born in Canada." |
| | tuntea~tunnen | 477 (219.) | "I know Matti." |
| | asunto~asunnossa | 246 (483.) | "in an apartment" |
| | palkinto~palkinnon | 150 (819.) | "to win the 1 st prize" |
| | sekunti~sekunnissa | 166 (743.) | "I'll be back in a second." |

⁵ Limiting the study to direct gradation only produced a smaller number of infrequent than frequent test tokens.

Elicitation of tokens was done by prompting the participant for Finnish translations for English words or phrases which were formulated in such a way that the targeted form would have to be used in Standard Finnish. Table 5 provides an example of the word lists used in the experiment.

3.3 *Experiment*

The recording took place at the participant's home in Courtenay, British Columbia in February 2006. The participant was asked to provide a Finnish translation of the words or phrases provided. The participant did not see the word lists. The researcher gave the basic form of a word in case the participant used a synonym rather than the targeted word. The session was recorded using a Sony MZ-B10 portable mini-disc recorder and a Sony EMC-T115 lapel microphone. The data was later transferred onto a CD using "Audacity"-software.

3.4 *Analysis*

Of the 88 words, 6 were discarded due to an error in elicitation or unfamiliarity with a word and thus yielding no translation. All in all 82 words, 44 frequent and 38 infrequent, were analyzed for their correctness, i.e., whether the word form given by the participant contained a gradated form or not. The data was subjected to two separate analyses with respect to two independent variables, word-frequency in the first analysis and word-familiarity in the second one. For both the variables, the occurrence of standard and variant forms was recorded. The rationale for proposing word-familiarity as a variable is that frequent words for native speakers, as represented by word-frequency corpora, do not necessarily correspond to frequent words for individuals, particularly for those who speak a language non-dominantly. Thus, a set of criteria was established to test the subject's familiarity with each token on the two lists. A word was considered familiar if the participant produced an immediate translation without the researcher giving the basic form. Similarly, the word was considered *familiar* if the participant offered a conversational cue, such as "yes" or "uh-huh" which was taken to mean that the word prompted is a familiar one. Conversely, a word was considered *unfamiliar* if the participant, regardless whether or not a conversational cue accompanied the answer, did not come up with a prompt translation, or if the participant hedged upon answering, or produced several forms before settling on one form as an answer. Words for which no answer was given were discarded. The researcher sorted the tokens one by one into familiar and unfamiliar ones using a test-retest-retest method; the tokens were evaluated against the criteria described above three separate times comparing the latest results with previous results (Seliger & Shohamy 2005, 186). This way, 61 familiar and 21 unfamiliar words were found. Then, the occurrence of standard and variant forms was recorded.

4 Results

Table 6 shows the distribution of standard and variant forms based on word-frequency. As can be seen from the table, a relatively high number of standard forms were attested on both lists regardless of their frequency. The participant produced standard forms in 61.36% of the frequent tokens and 78.97% of the infrequent tokens, showing higher number of standard forms with infrequent items than with frequent items.

Table 6

Distribution of standard and variant forms based on **word-frequency**

| Ranking | Standard (%) | Variant (%) | Total |
|--------------------|--------------|-------------|-------|
| <1000 (frequent) | 27 (61.36%) | 17 (38.64%) | 44 |
| >3000 (infrequent) | 30 (78.95%) | 8 (21.05%) | 38 |
| Total | 57 (69.51%) | 25 (30.49%) | 82 |

Conversely, there were more non-standard forms with frequent items (38.64%) than with infrequent items (21.05%). The results prove the trend going in the opposite direction than initially predicted. However, this distribution of standard and variant forms based on word-frequency was found to be insignificant at the .05 level. Thus, it is concluded that word frequency, as defined by the word-frequency corpus, has no significant effect on GL.

Next, the data were subjected to the second analysis, this time using word-familiarity as the independent variable. The variable was evaluated against the criteria described above. Table 7 provides the results for the latter analysis.

Table 7

Distribution of standard and variant forms based on **familiarity**

| | Standard (%) | Variant (%) | Total |
|------------|--------------|-------------|-------|
| Familiar | 49 (80.3%) | 12 (19.6%) | 61 |
| Unfamiliar | 8 (38.09%) | 13 (61.9%) | 21 |
| Total | 57 | 25 | 82 |

This analysis produced statistically significant results ($\chi^2=13.15$, $p \leq 0.001$). Standard forms are considerably more frequent with familiar words than with unfamiliar words, 80.3% as opposed to 38.09%. In turn, the number of variant forms is higher (61.9%) with unfamiliar words than with familiar words (19.6%). The results indicate that there is a relationship between familiarity and GL. Implications of these findings are discussed in the next section.

5 Discussion

In the present study, it was hypothesized that due to limited exposure to standard language, a non-dominant speaker would produce more native-like output with frequent

lexical items. Thus, it was expected that a non-dominant speaker of Finnish would produce more standard-like gradation in commonly encountered words. Conversely, words that are heard or used less often would show more deviant forms. This presentiment was proven correct. However, the results indicate that the definition of word-frequency must be expanded to encompass the idiosyncratic linguistic knowledge that an individual possesses. Although word-frequency, as defined by frequency corpora, might be an approximation of the lexicon in a native-speaker population in general, arguably it does not reflect accurately the status of the lexicon in an individual. The discrepancy between population-wide and an individual's own frequencies is heightened in non-dominant speaker contexts. In its stead, *word-familiarity* is proposed as a measure to reflect idiosyncratic word-frequencies at the level of the individual.

5.1 *Word-familiarity*

The concept of word-familiarity, idiosyncratic word-frequency, is proposed to account for lexical experience an individual has accumulated. This experience is shaped by upbringing, occupation, social class, personal interests and such creating unique lexicons in each individual. Corpus-based word-frequency listings fall short in their ability to capture this variation. Moreover, the discrepancy between individual and language-wide frequencies is heightened in non-dominant language situations.

According to Pierrehumbert (2000, 10), it is the number of different examples of morphophonological alternations that establish a word relationship in the memory. It influences the extent to which it is cognitively real to the speaker. The more often the token is used, the more familiar the speaker/hearer becomes with it. A familiar lexical item triggers appropriate semantic associations reliably and reinforces the relationship between them. This cognitive reality is what word-familiarity represents. With perfectly familiar items, the language user is able to create messages that manifest target-like production in respect to semantics, pragmatics, syntax, and phonology. Anything less than that implies less than perfect acquisition. The activation of a less familiar word is unreliable, triggering fewer or no associations.

Word-familiarity, then, corresponds to word-frequency but reflects an individual's idiosyncratic lexicon, linguistic experience that one gains from exposure to language. Frequently encountered tokens benefit from more frequent feedback and, therefore, show more fluency. Low-frequency tokens, in turn, are less familiar or unfamiliar altogether, showing more variance due to imperfect learning. The participant was significantly more likely to produce standard-like gradation in familiar words, which were produced without delay and/or accompanied by a conversational cue. The words that showed GL were also likely to be accompanied by hedging, delayed answers and/or the lack of conversational cues and, thus, classified as unfamiliar. Familiar items are likely to have been heard and used more often, for which reason their gradated forms were also easier to maintain. Infrequent items, in turn, have not benefited from regular use and have remained less familiar or completely unfamiliar. Neither have their gradated forms received enough reinforcement for the participant to acquire them solidly. These results are interpreted to stem from the participant's limited exposure to standard Finnish; the participant's linguistic experience in

Finnish has been accumulated from infrequent and/or less than perfect feedback in immigrant setting.

Since word-familiarity increases through frequency and recency of use, it can be equated with the concept of *activation level* (Pierrehumbert 2001), discussed in section 2.1. The more recently and frequently an item is called upon, the more familiar the item is and the more reliably it gets activated. The outcome can therefore be expected to be highly idiosyncratic, just as the results show. Consequently, word-familiarity is a gradient continuum ranging from fully acquired familiar words with which the user manifests linguistic dexterity, to moderately familiar words whose activation is unreliable and which are prone to variance, to completely unknown words that remain beyond grammatical manipulation.

5.2 Analogical leveling

The results indicate that GL affects infrequent items first, suggesting cognitive motivations underlying the variation, as discussed in section 2.1.1. GL represents a type of regularization, where unexpected forms are replaced by a more basic pattern and, as such, resembles *analogical leveling*. As discussed earlier, it is interpreted to stem from cognitive motivations, ways of organizing linguistic information in the brain.

As seen in Table 8, the speaker opts for the stem found in the basic form instead of the gradated stem although the latter is standard.

Table 8

Analogical leveling in application of consonant gradation

| Direct Gradation | (Basic form (SF)) Gradated form (SF) | Non-dominant Finnish | Gloss |
|------------------|---|----------------------|--------------------|
| lp:lv | (halpa) halvalla | halpalla | "at a cheap price" |

This tendency was apparent across all lexical categories in the data. *Lauseopin arkisto*, a sentence archive for Standard Finnish, identifies the basic form as the most frequent form across the lexical categories (Hakulinen et al. 2004, 1180-1181). According to the archive, NOMINATIVE and GENITIVE are the two most frequent cases for nouns, occurring at the rate of 26,3% and 27,4% respectively. NOMINATIVE is by far the case most common for adjectives (32,2%), demonstratives (32,1%), and personal pronouns (47,7%). The frequency at which other cases are found in the same sentence archive is far less; INESSIVE is the most common case for nouns and demonstratives with 6,7% and 8,3% respectively, ILLATIVE for adjectives with 5,3%, and ALLATIVE with 7,3% for personal pronouns. For verbs, in turn, the basic form infinitive occurs at the rate of 55,4%. Of the case forms for infinitives, ILLATIVE is the most common with but occurs at a rate considerably lower than the basic form at 16,7%. The above percentages are collected in Table 9.

Table 9

Most frequent case forms in Standard Finnish (%) (Hakulinen et al. 2004)

| | Nouns | Adjectives | Demonstrative Pronouns | Personal Pronouns | Verbs |
|-------------------------|-------|------------|------------------------|-------------------|-------|
| NOM | 26,3 | 32,2 | 32,1 | 47,7 | |
| INE | 6,7 | | 8,3 | | |
| ILL | | 5,3 | | | 16,7 |
| ALL | | | | 7,3 | |
| Basic form (infinitive) | | | | | 55,4 |

The high rates at which the basic forms, i.e., NOMINATIVE and infinitive, occur in standard language may explain the variation in the application of gradation in non-dominant speech. If the basic forms are statistically more common in Standard Finnish, a non-dominant learner would be more likely to have stored more instances of the basic form than the gradated form in the memory. Infrequent, less familiar categories would consist of fewer exemplars, but those would be more likely to contain the basic form consonant combinations due to their high frequency. Therefore, picking one exemplar from these less frequented categories would be likely to contain the basic form stem. Conversely, familiar frequently used exemplars would reside in categories with numerous closely related representatives. Among these exemplars, ones containing standard-like gradation would be well represented. Higher activation level overall in these commonly used categories provide the exemplars containing standard-like gradation with a higher probability to get selected.

As discussed in section 2.2, case suffixes and consonant gradation are intricately related in Finnish, for it is the suffixes that trigger gradation in the stem. See Table 10 for an example.

Table 10

Realization of consonant gradation in INESSIVE in Standard Finnish and non-dominant Finnish

| Direct Gradation | (Basic form) Gradated form; INE; SF | Non-dominant Finnish | Gradation | Gloss |
|------------------|-------------------------------------|----------------------|-----------|--------------------|
| lk~l | (jalka) jalassa | jalassa | yes | “on the foot/feet” |
| ŋk~ŋŋ | (kaupunki) kaupunkissa | kaupunkissa | no | “in the city” |

SF= Standard Finnish

However, based on the results, it is likely that especially in non-dominant speech, case and gradation are subject to their respective frequencies, which dictate whether or not the gradation is present in the output. The participant produced phrases such as *jalassa* ‘on the foot/feet’ and *kaupunkissa* ‘in the city’. The former is a standard form with respect to both case and gradation. The latter is non-standard with respect to gradation. It is proposed that case and gradation are stored as a unit in the same exemplar for familiar and well-established

concepts⁶. However, less common stem-case combinations may be pieced together from one exemplar containing the case suffix and another containing the basic form of the word.

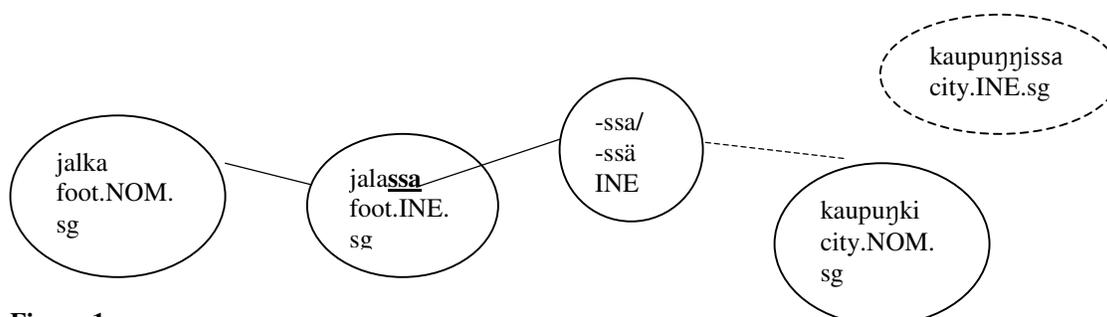


Figure 1

Exemplar associations in the mnemonic network.

This is demonstrated in Figure 1. Familiar concepts, such as *jalassa* ‘on the foot/feet’, are stored and processed as a unit and, thus, show standard gradation. *Jalassa* is associated with its base form *jalka* and the exemplar containing INESSIVE suffix, represented here with a solid line. INESSIVE, due to high-productivity, is assumed to also exist as a separate exemplar and to be added onto stems when needed, as with the phrase ‘in a city’. It consists of exemplars *kaupunki* and INESSIVE, shown with a dashed line. Since the phrase is pieced together from the exemplars that exist separately, gradation is not present in the output. The standard form *kaupungissa*, may also be stored as a unit close by but does not get selected due to its lower activation level, as represented by a dashed oval. The activation levels of the exemplars determine which ones get selected and, consequently, what properties are present in the output.

Analogical leveling is typically known as a phenomenon where a more complex process is replaced by another process, which is perceived as less complicated for the language user. GL resembles analogical leveling in that it regularizes complex grammar by replacing exceptional gradated forms with the most common form of a word, the basic form. Gradated forms are perceived as more complicated, even arbitrary from a point of view of a non-dominant speaker, unless they are well-established in the language. Unique in this regularization is that a non-process takes over a process. The basic form, the most frequent form in the language, overrides a process, consonant gradation. However, it achieves what analogical leveling achieves. It reduces grammatical complexity without risking comprehension. When frequencies of gradated forms in the memory are high enough, regularization ceases. Until then, when the memory fails with infrequent items, the pattern that is stronger in the language is chosen.

⁶ Slips of the tongue due to English interference, e.g., *lipset* for Finnish *huulet* ‘lips’ (huul-et, lip-PLU.NOM) provide evidence that morphology can indeed be lexicalized with the stem to represent a complete concept. The phenomenon is known as the Conserving Effect (Bybee 2005, 2007).

Of the two types of gradation, quantitative is perhaps more predictable than qualitative gradation. Quantitative gradation either shortens or lengthens the existing stem consonants keeping the quality of the consonant constant. On the other hand, qualitative gradation weakens or strengthens the obstruction of the consonant, changing the quality of it, for which reason a non-dominant speaker may perceive qualitative gradation more arbitrary. Acquisition and maintenance of the latter type is expected to be more difficult. Thus, in non-dominant speech, analogical leveling is predicted to cause basic form being used more often than the gradated form in words subject to more subtle qualitative gradation. However, this hypothesis remains to be a topic for future research.

6 Conclusion

From the results obtained from two separate analyses, it is concluded that application of consonant gradation in non-native Finnish as spoken by a second generation Finn as a non-dominant language is closely related to idiosyncratic word-frequency, i.e., word-familiarity. The first analysis considered if word-frequency, as defined by Finnish newspaper corpus, had a significant effect on application of consonant gradation in non-native speech. The results answer in the negative. The participant produced a relatively high number of standard gradation with both frequent and infrequent words. The second analysis, which in turn considered the effect of word-familiarity on application of consonant gradation, found significant results. The number of standard forms was significantly higher with familiar words than with unfamiliar words, while GL was considerably more likely in unfamiliar words than in familiar words. Less frequent items are more unfamiliar and regularized by analogical leveling, while frequent and thus more familiar items benefit from regular feedback showing more standard production. Therefore, it can be concluded that word-frequency is a relative concept that depends on an individual's personal linguistic experience, in native or non-native speech alike. In order to take this idiosyncratic nature of word-frequency into consideration, word-frequency should be defined on an individual basis.

Acknowledgments

I would like to thank Dr. Sonya Bird for her supervision and guidance during this project. This paper has greatly benefited from her insights and suggestions. I also want to thank Sunghwa Lee and Soo-Youn Ham who kindly peer-reviewed my original paper and gave me constructive comments. I also want to express my sincere gratitude to my consultant and friend M.P. for sharing her time and language with me.

Appendix A. Consonant Gradation Data: Frequency <1000

| Direct Gradation | Standard Finnish basic:gradated form | Frequency/million words (number) | Non-dominant Finnish | Gloss |
|------------------|--------------------------------------|----------------------------------|----------------------|------------------------------------|
| | | | | “in “in the end” |
| pp:p | loppu:lopussa | 653 (128.) | lopussa | |
| | helppo:helposti | 253 (461.) | helposti | “easily” |
| | | | | |
| k:j | poika:pojan | 384 (294.) | pojan | “a boy’s (room)” |
| | aika:ajan | 1315 (53.) | - | “for an hour” |
| tt:t | esittää:esitän | 621 (141.) | esitän | “to present” |
| | | | | |
| kk:k | viikko:viikon | 386 (100.) | viikon | “for a week” |
| | tarkka:tarkasti | 142 (856.) | tarkasti | “accurately” |
| | kirkko:kirkossa | 286 (411.) | kirkon sisällä | “in the church” |
| | | | | |
| nt:nn | ranta:rannalla | 140 (872.) | rannalla | “on the beach” |
| | syntyä:synnyin | 588 (157.) | syntyä | “I was born (in Canada)” |
| | tuntea:tunnen | 477 (219.) | tunnen | “I know Matti.” |
| | asunto:asunnossa | 246 (483.) | asunto | “in an apartment” |
| | palkinto:palkinnon | 150 (819.) | palakinto | “to win the 1 st prize” |
| | sekunti:sekunnissa | 166 (743.) | sekuntissa | “(I’ll be back) in a second” |
| | luonto:luonnossa | 165 (748.) | luonnossa | “in the nature” |
| | kunto:kunnossa | 328 (355.) | kunto | “(to be) in shape” |
| | myöntää:myönnän | 385 (291.) | myönnän | “I admit” |
| | tunti:tunnissa | 348 (339.) | tunnin | “in an hour” |
| | hinta:hinnat | 440 (245.) | hinnat | “prices (are up)” |
| | | | | |
| ŋk:ŋŋ | kaupunki:kaupungissa | 1062 (70.) | kaupungissa | “in the city” |
| | Helsinki:Helsingissä | 1133 (63.) | Helsingissä | “in Helsinki” |
| | | | | |
| lt:ll | ilta:illa-n/-lla | 231 (528.) | illan/illalla | “in the evening” |
| | valta:vallassa | 164 (749.) | valtassa | “(a president) in power” |
| | | | | |
| lk:l | jalka:jalassa | 151 (818.) | jalassa | “(socks) on the feet” |
| | | | | |
| lk:l̥j | jälki:jäljet | 136 (904.) | jäljet | “(bear) tracks” |
| | | | | |
| rt:rr | kertoa:kerron | 1851 (30.) | kertoo | “I tell...” |
| | kerta:kerran | 752 (113.) | kerta | “one time” |
| | siirtyä:siirryn | 380 (297.) | siirryn | “I move (aside)” |

| | | | | |
|-------|-----------------|------------|----------------------------------|---------------------------------|
| rk:r | pyrkiä:pyrin | 404 (280.) | pyrkin | “to apply (to school)” |
| | Turku:Turussa | 768 (108.) | Turus | “in Turku” |
| ht:hd | ehtiä:ehdin | 315 (376.) | ehtin | “I was in time for...” |
| | johto: johdossa | 354 (332.) | johtos | “(a twist) in an electric cord” |
| p:v | apu:avusta | 465 (226.) | apusta | “thanks for your help” |
| | luvata:lupaan | 238 (503.) | luvaan | “I promise...” |
| | tapa:tavat | 649 (130.) | tapa | “(good) manners” |
| t:d | äiti:äidille | 244 (488.) | äitille (correct dialectal form) | “for mom” |
| | vaatia:vaadin | 638 (135.) | vaadin | “I demand...” |
| | löytää:löydän | 410 (276.) | löydän | “I find (old things)” |
| | sota:sodassa | 301 (395.) | sodassa | “in the war” |
| | katu:kadulla | 136 (905.) | kadulla | “(met Mari) on the street” |
| | pyytää:pyydän | 198 (620.) | pyydän | “I ask for (more water)” |
| | koti:kodissa | 299 (397.) | kodissa | “in a home” |
| k:- | tuki:tuen | 384 (292.) | tukia | “you support your children” |
| | aikoa:aion | 379 (299.) | - | “I’m going to...” |
| | hakea:haen | 378 (230.) | hajen | “I pick Lisa up” |
| | lukea:luen | 211 (586.) | lujen | “I read (in the evening.)” |
| k:v | luku:luvut | 767 (109.) | - | “numbers” |

Appendix B. Consonant Gradation Data: Frequency >3000

| Direct Gradation | Standard Finnish basic/gradated form | Frequency/million words (number) | Non-dominant Finnish | Gloss |
|------------------|--------------------------------------|----------------------------------|----------------------|----------------------|
| tt:t | rauhottaa:rauhoitin | 34 (3031.) | rauhotin | “I calm (sb down)” |
| | valuutta:valuutat | 34 (3079.) | valuutat | “currencies” |
| kk:k | leikki:leikit | 34 (3010.) | leikit | “(children’s) games” |
| | rannikko:rannikolla | 34 (3083.) | rannikolla | “on the coast” |
| | tupakka:tupakalla | 33 (3101.) | tupakala | “having a cigarette” |
| | lenkki:lenkillä | 31 (3330.) | lenkile | “having a walk” |
| mp:mm | lampi:lammessa | 11 (4973.) | lammessa | “in the pond” |
| nt:nn | tuonti:tuonnit | 35 (3009.) | tuontia | “imports” |

| | | | | |
|--------|------------------------|------------|--------------------------------|--|
| | emäntä:emännälle | 34 (3078.) | emänäle | "to the host" |
| | rinta:rinnassa | 31 (3273.) | rinnassa | "in the chest" |
| | pyyntö:pyynnöt | 31 (3284.) | pyyntöjä | "requests" |
| | Skotlanti:Skotlannissa | 31 (3297.) | Skotlantissa | "in Scotland" |
| | uinti:uinnit | 25 (3951.) | - | "swims (plu)" |
| | ravinto:ravinossa | 25 (3974.) | ravinossa | "in nutrition" |
| | perintö:perinnöt | 24 (4125.) | perintöjä | "inheritances" |
| | | | | |
| ηk:ηη | kenkä:kenjät | 33 (3116.) | kenjät | "shoes" |
| | runko:runnot | 23 (4248.) | runnot | "tree trunks" |
| | sänky:sänjyssä | 20 (4606.) | sänjyssä | "in bed" |
| | | | | |
| lt:ll | kielto:kiellot | 32(3208.) | kiellot | "denials (plu)" |
| | aalto:aallot | 27 (3663) | aallot | waves (in a lake) |
| | puhaltaa:puhallan | 31 (3306.) | puhallan | "I blow (in the soup)" |
| | huolto:huollossa | 30 (3352.) | huollossa | "in (mom)'s care" |
| | | | | |
| lk:l | polku:polulla | 26 (3810.) | polulla | (a bear) on the trail" |
| | | | | |
| lk:l̥j | polkea:poljen | 21 (4499.) | polen (correct dialectal form) | "I pedal (a bike)" |
| | | | | |
| ht:hd | kiirehtiä:kiirehdin | 28 (3516.) | - | "I hurry" |
| | | | | |
| hk:h | pyyhkiä:pyyhin | 14 (6156.) | pyhin | "I wipe (the table.)" |
| | | | | |
| p:v | tupa:tuvassa | 14 (6184.) | tupassa | "in a farm house kitchen" |
| | repiä:revin | 32 (3191.) | revin | "I tore (the letter.)" |
| | kipu:kivussa | 34 (3043.) | kipua | "in pain" |
| | syöpä:syövät | 27 (3701.) | syövät | "cancers" |
| | | | | |
| t:d | paita:paidan | 34 (3037.) | paidan | "(I bought) a shirt" |
| | latu:ladulla | 34 (3074.) | ladulla | "on the ski trail" |
| | satu:sadun | 33 (3017.) | | "(I read) a fairy tale." |
| | sietää:siedän | 32 (3299.) | siedän | "I tolerate" |
| | hauta:haudalle | 31 (3299.) | haudalle | "(to take flowers) to the grave" |
| | aita:aidalla | 29 (3432.) | aidalla | "on the fence" |
| | ruutu:ruudussa | 29 (3461.) | ruudun | "in the square" |
| | muoti:muodissa | 27 (3640) | muodissa | "in fashion" |
| | | | | |
| k:- | reikä:reiän | 19 (4867.) | reikä | "(I repaired) a snag (in the sweater)" |
| | nälkä:nälässä | 26 (3832.) | nälissä | "in hunger/hungry" |
| | pako:paossa | 24 (4081.) | - | "escapes/in exile" |
| | raaka:raa'at | 23 (4244.) | - | "not rype (bananas)" |
| | sika:sian | 19 (4692.) | sian | pig's meat |
| | | | | |

References

- Bybee, J.. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (eds.), *Usage-based models of language*. Stanford: CSLI. 65-85.
- _____. (2001). *Phonology and Language Use*. Cambridge University Press.
- _____. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, Vol. 14, 261-290.
- _____. (2005). From usage to grammar: The mind's response to repetition. *Presidential Address, Oakland*.
- CSC, The Finnish IT center for science (n.d.). Suomen sanomalehtikielen taajuussanasto laskevan taajuuden mukaan (Finnish newspaper language wordlist according to descending frequency). Retrieved February, 13, 2006, from at <http://www.csc.fi/kielipankki/aineistot/ssts/frq1/phtml>
- Fidelholz, J.. (1975). Word frequency and vowel reduction in English. *CLS 11*. 200-213.
- Hakulinen, A., Vilkuna, M., Korhonen, R., Koivisto, V., Heinonen, T.R., and Alho, I. (2004). *Iso Suomen Kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Hooper, J.. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. M. Christie, Jr. (Ed.), *Current progress in historical linguistics*. Amsterdam: North-Holland Publishing Company. 95-105.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins. 229-254.
- Krug, M. (1998). String frequency: a cognitive motivating factor in coalescence, language processing and linguistic change. *Journal of English Linguistics*, Vol. 26, 286-320.
- Labov, W. (1994). *Principles of linguistic change: Internal factors*. Oxford:Blackwell.
- Lightbown, P. & Spada, N.. (1999). *How languages are learned*. Oxford: Oxford University Press.
- Losiewicz, B.C.. (1992). *The effect of frequency on linguistic morphology*. Ph.D. dissertation. Department of Linguistics, University of Texas, Austin.

- Mowrey, R. & Pagliuca, W.. (1995). The reductive character of articulatory evolution. *Rivista di Linguistica* 7:37-124.
- Pierrehumbert, J. B. (2000). *What people know about sounds of language*. Studies in the Linguistic Sciences, 29(2) Fall, 111-120.
- _____. (2001). Exemplar Dynamics: Word frequency lenition and contrast. In J. Bybee and P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137-157). Amsterdam/Philadelphia: John Benjamins.
- Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, Vol. 60, No. 2, 320-342.
- _____. (2001). Lexical diffusion, lexical frequency, and lexical analysis. In J. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins. 123-136.
- Schuchardt, H.. (1885). *Über die Lautgesetze:gegen die Junggrammatiker*. Berlin: Oppenheim. (Reprinted and translated in T. Vennemann & T. Wilbur. [1972]. *Schuchardt, the Neogrammarians and the Transformational Theory of Phonological Change*. Frankfurt: Athenaum).
- Seliger, H. & Shohamy, E.. (2005). *Second Language Research Methods*. Oxford: Oxford University Press.