

**The Finnish Language in Post-Utopian Sointula:  
The Effects of Frequency on Consonant Gradation**

by

Pauliina Saarinen  
B.A., University of Victoria, 2001

A Thesis Submitted in Partial Fulfillment  
of the Requirements for the Degree of

MASTER OF ARTS

in the Department of Linguistics

© Pauliina Saarinen, 2009  
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

## **Supervisory Committee**

### **The Finnish Language in Post-Utopian Sointula: The Effects of Frequency on Consonant Gradation**

by

Pauliina Saarinen  
B.A., University of Victoria, 2001

#### **Supervisory Committee**

Suzanne Urbanczyk, Department of Linguistics  
**Supervisor**

Sonya Bird, Department of Linguistics  
**Supervisor**

Tae-Jin Yoon, Department of Linguistics  
**Departmental Member**

Hiroko Noro, Department of Pacific and Asian Studies  
**Outside Member**

## Abstract

### **Supervisory Committee**

Suzanne Urbanczyk, Department of Linguistics

Supervisor

Sonya Bird, Department of Linguistics

Supervisor

Tae-Jin Yoon, Department of Linguistics

Departmental Member

Hiroko Noro, Department of Pacific and Asian Studies

Outside Member

This research investigated the effect of frequency of language use on the production of consonant gradation by non-dominant speakers of Finnish in the immigrant community of Sointula, BC. Three types of frequency – word-frequency, suffix-frequency, and stem-frequency – were tested. It also investigated whether quantitative or qualitative gradation is more successful in producing gradation than the other and, finally, whether immigrant generation can explain the variation between participants. A translation task was administered to the six participants across three generations.

Based on the framework of exemplar-driven cognitive grammar (Bybee 2001; Pierrehumbert 2001), the frequency-effects were assumed to be contingent upon the mode of lexical access; frequent complex words, presumably accessed as wholes thanks to frequent usage, would not exhibit as many gradation errors as infrequent words, which would be accessed via their composite parts due to infrequency.

The anticipated frequency-effects were not found. Both frequent and infrequent words manifested some gradation loss as an analogical change. This suggests that all words are infrequent. While Bybee's model assumes high-volume language use over time in dominant language contexts, lack of volume appears to suppress the differential behavior between frequent and infrequent words in Sointula. However, correct gradation

was predictable based on suffix-use, which in turn was determined partly by semantics of suffixes; those Finnish suffixes that are semantically mappable to equivalent morphemes in English were better preserved than GEN object-markers, which do not have corresponding morpheme in English. With the atrophy of the GEN object-marker also gradation becomes redundant. This may arise from the tendency to mark syntactic constituency with word-order alone in English-influenced Finnish. Thus, semantics of suffixes proves to be a better predictor of gradation than frequency.

Gradation loss increased with each generation born abroad; by G3, it has all but disappeared. Consonant gradation is not preserved through the generations. Qualitative gradation disappears before quantitative gradation. The above findings are sensible in a context of reduced language-functionality.

Against expectation, little evidence for storing sub-word morphemes and decomposed access was found. Instead, the data suggests that most stored lexical items are whole words and that gradation is associated with whole complex forms.

## Table of Contents

|  |      |
|--|------|
| Supervisory Committee .....  | ii   |
| Abstract .....   | iii  |
| Table of Contents .....  | v    |
| List of Tables .....   | vii  |
| List of Figures .....  | viii |
| Acknowledgements .....   | ix   |
| Dedication .....   | xi   |
| Chapter One Introduction .....   | 1    |
| 1.1 Background .....   | 1    |
| 1.2 Thesis outline .....   | 3    |
| Chapter Two Literature Review .....  | 4    |
| 2.1 Consonant Gradation in Standard Finnish .....                                      | 5    |
| 2.1.1 Gradation types: Quantitative and qualitative .....                              | 6    |
| 2.1.2 The direction of gradation: Direct and indirect gradation .....                  | 8    |
| 2.1.3 Unpredictability of qualitative gradation .....                                  | 10   |
| 2.1.4 Implications for acquisition .....   | 11   |
| 2.2 Language loss in minority language contexts .....                                  | 14   |
| 2.2.1 Inter-generational language loss: Language shift .....                           | 14   |
| 2.2.2 Intragenerational language loss: language attrition and imperfect learning ..... | 15   |
| 2.3 Cognitive view of language: Usage-based grammar .....                              | 19   |
| 2.3.1 Frequency effects of frequent and infrequent items .....                         | 21   |
| 2.3.2 Exemplar Dynamics .....  | 25   |
| 2.4 Summary .....  | 32   |
| 2.5 Research statement and summary of research questions .....                         | 34   |
| Chapter Three Methodology .....  | 36   |
| 3.1 Participants .....   | 36   |
| 3.2 Instruments .....  | 39   |
| 3.2.1 Stimuli: Sentence list .....   | 39   |
| 3.2.2 Questionnaire .....  | 45   |
| 3.3 Procedures .....   | 46   |
| 3.4 Data-processing .....  | 48   |
| 3.4.1 Independent variables .....  | 52   |
| 3.4.2 Dependent variables and measurements .....                                       | 53   |
| Chapter Four Results .....   | 57   |
| 4.1 Frequency of use .....   | 58   |
| 4.1.1 Word-frequency .....   | 58   |
| 4.1.2 Suffix-frequency .....   | 65   |
| 4.1.3 Stem-frequency .....   | 69   |
| 4.1.4 Summary of frequency-tests .....   | 70   |
| 4.2 Gradation-type .....   | 71   |
| 4.3 Individual speakers and immigrant generation .....                                 | 74   |
| 4.4 Chapter summary .....  | 77   |

|  |     |
|--|-----|
| Chapter Five Discussion .....  | 80  |
| 5.1 Research predictions and reseach findings .....                        | 81  |
| 5.1.1 <i>Frequency of use and lack of expected frequency-effects</i> ..... | 86  |
| 5.1.2 <i>Atrophy of GEN object marker</i> .....                            | 91  |
| 5.1.3 <i>Gradation-types</i> .....   | 99  |
| 5.1.4 <i>Individual participants and immigrant generations</i> .....       | 100 |
| 5.2 Implications for Cognitive Grammar .....                               | 104 |
| 5.2.1 <i>Accounting for standard forms</i> .....                           | 104 |
| 5.2.2 <i>Accounting for non-standard forms</i> .....                       | 111 |
| Chapter Six Conclusion .....   | 113 |
| 6.1 Summary of Research .....  | 113 |
| 6.2 Limitations .....  | 116 |
| 6.2 Future research .....  | 117 |
| 6.4 Contributions .....  | 117 |
| Bibliography .....   | 119 |
| Appendix A Stimuli: Sentence List .....                                    | 127 |
| Appendix B The Finnish Language in Sointula Background Questionnaire ..... | 129 |
| Appendix C Glossary of terminology .....                                   | 133 |
| Appendix D The breakdown of suffixes in the stimuli .....                  | 135 |
| Appendix E Consent script .....  | 136 |
| Appendix F Participant Consent Form .....                                  | 137 |
| Appendix G Flow-chart for Suffix-frequency tests .....                     | 140 |

## List of Tables

|  |    |
|--|----|
| Table 2-1: Finnish consonant gradation, a morpho-phonological process .....  | 6  |
| Table 2-2: Quantitative gradation, i.e. shortening or lengthening .....  | 7  |
| Table 2-3: Qualitative gradation, i.e. leniting or strengthening .....   | 7  |
| Table 2-4: Direct gradation .....  | 8  |
| Table 2-5: Indirect (reverse) gradation .....  | 9  |
| Table 2-6: Approx. rates of occurrence of gradation types in the Comprehensive<br>Dictionary .....                   | 10 |
| Table 2-7: Diabological K .....  | 12 |
| Table 2-8: The van Els taxonomy .....  | 15 |
| Table 2-9: Analogical levelling .....  | 23 |
| Table 2-10: Variation in consonant gradation in non-dominant Finnish .....   | 24 |
| Table 2-11: Most frequent case forms in Standard Finnish (%) .....   | 25 |
| Table 2-12: Research questions and predictions .....   | 35 |
| Table 3-1: The properties of the selected word-forms .....   | 44 |
| Table 3-2: An excerpt of the sentence list. ....   | 45 |
| Table 3-3: Coding parameters summarized .....  | 50 |
| Table 3-4: Independent and dependent variables .....   | 56 |
| Table 4-1: Word-frequency * Whole-word-accuracy crosstabulation (with NOM.sg<br>word-forms) .....                    | 59 |
| Table 4-2: Word-frequency * Whole-word-accuracy crosstabulation (without NOM.sg<br>word-forms) .....                 | 60 |
| Table 4-3: Word-frequency * Gradation-accuracy crosstabulation .....   | 62 |
| Table 4-4: Word frequency * Suffix-accuracy crosstabulation .....  | 63 |
| Table 4-5: Suffix-frequency * Suffix-accuracy crosstabulation .....  | 66 |
| Table 4-6: Suffix-frequency * Gradation-accuracy crosstabulation .....   | 68 |
| Table 4-7: Stem-frequency * Gradation-accuracy crosstabulation .....   | 70 |
| Table 4-8: Gradation-type * Gradation-accuracy crosstabulation .....   | 72 |
| Table 4-9: Gradation-type * Gradation-accuracy crosstabulation in words with gradation-<br>triggering suffixes ..... | 73 |
| Table 5-1: Predictions and answers for research questions .....  | 83 |
| Table 5-2: GEN object atrophy .....  | 93 |

## List of Figures

|   |    |
|---|----|
| Figure 2-1: Lexical associations of the word 'unbelievable' .....   | 20 |
| Figure 2-2: Exemplar associations in the mnemonic network .....   | 30 |
| Figure 4-1: The count of instances of gradation loss by Participant in the frequent and<br>infrequent categories of word-forms..... | 62 |
| Figure 4-2: The effect of Participant on Gradation-accuracy .....   | 75 |
| Figure 4-3: The effect of Generation on Gradation-accuracy .....  | 76 |

## Acknowledgements

The completion of this thesis would not have been possible without the many individuals who have dedicated their time and effort for this project.

I am very grateful to my supervisory committee, Drs Suzanne Urbanczyk, Sonya Bird, Tae-Jin Yoon and Hiroko Noro, who guided me through this project as it evolved from an idea into a finished thesis. It was for Sonya's great phonology class that the idea for this thesis took root. A big thank you, Sonya, for not having your baby before I defended my proposal. I thank you, Su, for taking over supervising the project. Your expertise in morphology, language endangerment and field methods was most helpful as were your comments on the drafts. You also provided encouragement, understanding, patience and kleenex when it was most needed. Many thanks to Dr Hiroko Noro. I found your advice, comments and interest most encouraging. I also want to thank you Dr Tae-Jin Yoon for your insights in developing the stimuli, advice in coding my data, relentlessness in teaching me the secrets of SPSS, and for valuable comments on the contents of this thesis.

I also want to express my gratitude to the many linguists in Finland whom I approached at different stages of research, Antero Niemikorpi, Eero Vitie, Paula Lyytinen, Matti Leiwo, Päivi Torvelainen, and Maisa Martin. Your knowledge of corpus-Finnish, consonant gradation and child-language was very useful and much appreciated.

I am grateful for my fellow Finns in Victoria and Sointula who participated in this project, either by testing the workings of my research tools or providing me with their language samples. I was honored to visit and have coffee and pulla with you. Special

thanks go to Gloria Williams, Ingrid Belveal and Tom Roper, my liaisons in Sointula, who answered my many questions. Thank you John Taylor for hosting me during my stay in Sointula. Sointula and Sointula's Finns will always remain in my heart.

I would also like show my appreciation for the institutions that have helped me over the course of my MA program. I appreciate the financial help from the Department of Linguistics through employment as a teaching assistant (2005-2008) and from the University of Victoria by means of a MA fellowship (2007-2008). The Faculty of Graduate Studies provided me with travel-grants to help cover the travel costs to Sointula for data-collection and to Seattle, WA for a conference. I appreciate my employer Pacific Gateway International College and specifically Director Megan Domenichelli for making many allowances in accommodating my graduate studies.

I want to also thank the people — the faculty, staff and fellow students — at the Linguistics Department at UVic. You create a wonderful atmosphere in the department. Special thanks to Janet Leonard for great walks and talks, Carolyn Pytlyk for proofreading and the whole Ling-gang in general for relaxing evenings over meals in good supportive company. Thank you to my family in Finland and on Vancouver Island for cheering me on.

I have a special thank you for you, Andrew Arnold, for being a tireless sounding-board, a proofreader, an editor and a born linguist. Your endless support, encouragement and love made writing this thesis possible.

Last but not least, I want to thank Woof and Phillip for insisting on walks. They knew that sniffing (fresh air) is good for the mind.

## Dedication

*Rakkaille vanhemmilleni*

# Chapter One

## Introduction

### 1.1 Background

Sointula, an island community on the west coast of British Columbia was established in 1901 as a utopian colony by a group of Finnish socialists. The dream did not last but the Finns remained. Today a few second- and later-generation Finns, some of whom are descendants of the original pioneers, continue to live in the community together with some newcomers from Finland and Canadians. Sointula is home to about 600 people and is still known as a “Finnish community”. A mere 60 years ago, there was a lively Finnish speaking community there. Since then, close contact with the mainstream culture has been eating away at the vitality of Sointula’s Finnish and English has become the dominant language for the succeeding immigrant generations. While today approximately 80% of the community’s total population has Finnish heritage (Välakangas 2004), there are only about 50 Finnish speakers remaining in the community (G. Williams, p.c., August 21, 2007). Finnish in Sointula is a severely threatened language.

This research emerged from a pilot study conducted as a term project in 2006. The language of a second-generation Finn, who had learned Finnish as a L1 at home but subsequently become English-dominant, was examined with respect to the application of consonant gradation. Consonant gradation is a feature of Finnish grammar that affects either the quantity (quantitative gradation) or the quality (qualitative gradation) of the consonants /p, t/ and /k/ in the stems of morphologically complex words (i.e. words made of many parts, for example, English *un-believ-able* or *knive-s*). Quantitative gradation

shortens a long consonant without changing the quality of the consonant. For example, long consonants, written in orthography as *pp*, *tt*, *kk*, become short, *p*, *t*, *k*. Conversely, direct Qualitative gradation weakens the closure of the consonant in the suffixed form by changing the quality of the consonant into another. For example, a consonant /t/ can turn into /n, l, r/ or /d/. Gradation is triggered by suffixes of the shape –C(onsonant) or –CCV(owel) (Hakulinen et al. 2004, p. 109) (see Section 2.1). The consultant's language manifested standard-like gradation in some words while in others gradation was lost and what should have been a gradated, stem was replaced by an ungraded NOMINATIVE singular (NOM.sg) case stem. Corpus-based word-frequency was investigated as the conditioner of the attested variation in gradation because frequency of whole words, as defined by corpus-based word-frequency rankings, has been found to affect the phonetic shape of individual words stored in the mental lexicon (Bybee 2001, 2002; Phillips 1984). The pilot project concluded that, at least for the one non-dominant Finnish speaker consulted, it was idiosyncratic frequency (i.e. word-familiarity) that conditioned the production of gradation, rather than corpus-based frequency (Saarinen, to appear).

This research elaborates upon the topic of the term paper by increasing both the number of participants and the complexity of the questions asked. It studies the speech of six non-dominant Finnish speakers from Sointula across three generations. Its purpose is to determine whether frequency of words, suffixes and stems affects the production of consonant gradation. It also investigates if one gradation-type, quantitative or qualitative, is more successful than the other. Lastly, it answers whether immigrant generation can explain the variation in the production of gradation between the participants.

Findings from this investigation conclude that expected frequency-effects cannot be found in Sointula Finnish. All words in Sointula behave as infrequent words as manifested by the patterning of gradation loss in the data. Moreover, semantics of suffixes is a better predictor of gradation-accuracy than frequency. Finally, gradation is not sustained through immigrant generations with qualitative gradation being lost before quantitative gradation.

## **1.2 Thesis outline**

This thesis is presented in six chapters. Chapter Two reviews the relevant literature that pertains to the following areas of linguistics: 1) Finnish consonant gradation, 2) language loss in minority language contexts, 3) Usage-based Grammar, Frequency-effects and Exemplar Model. It finishes by presenting the Research statement, Research questions and Research hypotheses, which emerge from the literature reviewed. Chapter Three provides a detailed description of the Methodology used in this research. The results of the data collected are presented in Chapter Four. The findings isolated by this research are discussed in detail in the light of the research questions and hypotheses in Chapter Five. Finally, Chapter Six concludes this thesis by summarising the results, discussing some of the limitations, weighing its contributions and offering some directions to take in future research.

## **Chapter Two Literature Review**

The goal of this research was to investigate whether frequency of use affects the phonetic output (i.e. the pronunciation of words as heard in speech) in the non-dominant language context of Sointula. It was particularly interested in examining if there is a relationship between the frequency of use and preservation of consonant gradation. Three types of frequency were considered — word-frequency, suffix-frequency and stem-frequency. In addition, this research set out to examine if one gradation type, direct quantitative or direct qualitative, is more successful than the other. Finally, this research examined to what extent individual participants are affected by loss of gradation and whether immigrant generation can explain variation between the participants. Relevant literature reviewed in this chapter relates to the following areas of linguistics: Section 2.1 introduces the properties of Finnish consonant gradation; Section 2.2 discusses language loss in minority language communities; Usage-based grammar, frequency-effects and Exemplar Dynamics model are presented in Section 2.3; Section 2.4 presents the summary of the literature reviewed. The chapter finishes with presenting Research statement, Research questions and Anticipated answers in Section 2.6.

## 2.1 Consonant Gradation in Standard Finnish

The following paragraphs outline the mechanics of consonant gradation of Finnish in necessary detail (from Hakulinen et al. 2004) and establish why consonant gradation may pose a challenge for non-dominant speakers.

Consonant gradation is a frequently encountered phenomenon in every-day written and spoken Finnish. Approximately a third of the 1000 most frequent words are subject to it. Consonant gradation is a morpho-phonological process in which the alternation of stem consonants is triggered by inflectional and derivational suffixation<sup>1</sup>. Affected are voiceless stop consonants /p, t/ and /k/ when they are preceded by a vowel or a voiced consonant. Suffixes of the shape –C or –CCV, which create a closed syllable in the previous stem, initiate the consonant-alternation. Vowel-initial suffixes that create an open syllable in the previous stem do not trigger the alternation (see Table 2-1 below). Voiced stops /b, g/ are affected in some slang words. Syllable boundaries are marked with | in the tables below.

---

<sup>1</sup> In general, inflectional morphology modifies a word to fit in a sentence context and derivational morphology creates new words.

| stem.NOM.sg                          | Gradation | -C/-CCV suffix (closed syllable)                         |  | V-initial suffix (open syllable)                         |  |
|--------------------------------------|-----------|--|--|--|--|
|                                      |           | Stem-GEN.sg<br>“of a(n) x”                               | Stem-ELA.sg<br>“from a(n) x”   | Stem-PART.sg<br>“a(n) x (obj)”                           | Stem-ILL.sg<br>“in(to) a(n) x”                             |
| a. <i>tonttu</i><br>‘an elf’         | tt-t      | <i>ton<u>l</u>tu-n</i><br>* <i>tont<u>tt</u>u-n</i>      | <i>ton<u>l</u>tu-s<u>l</u>ta</i><br>* <i>tont<u>tt</u>u-s<u>l</u>ta</i>      | <i>tont<u>l</u>tu-a</i><br>* <i>tontu-a</i>              | <i>tont<u>l</u>tu-un</i><br>* <i>tontu-un</i>              |
| b. <i>palkka</i><br>‘a salary’       | kk-k      | <i>pall<u>k</u>a-n</i><br>* <i>palk<u>kk</u>a-n</i>      | <i>pall<u>k</u>a-s<u>l</u>ta</i><br>* <i>palk<u>kk</u>a-s<u>l</u>ta</i>      | <i>palk<u>k</u>a-a</i><br>* <i>palka-a</i>               | <i>palk<u>k</u>a-an</i><br>* <i>palka-an</i>               |
| c. <i>tur<u>p</u>a</i><br>‘a muzzle’ | p-v       | <i>tur<u>l</u>v<u>a</u>-n</i><br>* <i>tur<u>p</u>a-n</i> | <i>tur<u>l</u>v<u>a</u>-s<u>l</u>ta</i><br>* <i>tur<u>p</u>a-s<u>l</u>ta</i> | <i>tur<u>l</u>p<u>a</u>-a</i><br>* <i>tur<u>v</u>a-a</i> | <i>tur<u>l</u>p<u>a</u>-an</i><br>* <i>tur<u>v</u>a-an</i> |

Note. N(OMINATIVE), GEN(ITIVE), PART(ITIVE), ILL(ATIVE)

**Table 2-1: Finnish consonant gradation, a morpho-phonological process**

Stems subject to gradation have two stem-morpheme alternants. Referring to the above table, *tonttu* ‘an elf’, for instance, has the allomorphs *tonttu-* and *tontu-*. The allomorphs of *palkka* ‘a salary’ are *palkka-* and *palka-*, and the allomorphs of *turpa* ‘a muzzle’ are *turpa-* and *turva-*. The addition of a suffix of the shape –C or –CCV to a stem creates a phonetic environment (i.e. a closed syllable) that calls for the use of the stem-allomorph containing gradated consonants.

### 2.1.1 Gradation types: Quantitative and qualitative

Consonant gradation is an alternation of either the quantity (i.e. length) or the quality of the stop-consonant. For example, the alternation of *tt~t* in *rätti~rätin* ‘rag.NOM.sg~rag.GEN.sg’ is called *quantitative* gradation because the quantity of the consonant in question is affected through shortening of the geminate-consonant [t:] of the NOM.sg stem into the singleton [t] in the GEN.sg. Conversely, the alternation of *rt~rr* in *virta~virran* ‘stream.NOM.sg~stream.GEN.sg’ or the deletion of single stop /k/, as in

*näky~näyn* ‘sight.NOM.sg~sight.GEN.sg’ exemplify *qualitative* gradation since the quality of the consonants in question changes. In the former, /t/ changes into [r] and in the latter, /k/ deletes (changes into Ø). The most frequently affected consonant is /t/, while /p/ is affected the least often. Proper nouns, a few recent two-syllable coinages, and loan words in which a single stop consonant is between vowels remain exempt from gradation (e.g. *laku* ‘liquorice’, *söpö* ‘adorable, cute’, and *auto* ‘a car’). This word-shape in old vocabulary would be subject to qualitative gradation. Single stop consonants are less likely to undergo gradation in general than geminate stop consonants, which are subject to quantitative gradation. Quantitative gradation is overall a more productive process than qualitative gradation and even new words are not usually exempt from it (p. 75). Table 2-2 presents the consonants subject to quantitative gradation and Table 2-3 those subject to qualitative gradation.

|   | /p/    | /t/    | /k/    |
|---|--------|--------|--------|
| 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~rk | rkk~rk |
| 5 | bb~b   |        | gg~g   |

**Table 2-2: Quantitative gradation, i.e. shortening or lengthening**

|   | /p/   | /t/   | /k/     |
|---|-------|-------|---------|
| 1 | mp~mm | nt~nn | ηk~ηη   |
| 2 | lp~lv | lt~ll | lk~l/lj |
| 3 | rp~rv | rt~rr | rk~r/rj |
| 4 |       | ht~hd | hk~h/hj |
| 5 | p~v   | t~d   | k~ø/v   |

**Table 2-3: Qualitative gradation, i.e. leniting or strengthening**

In the above tables, a tilde indicates that gradation can proceed in either direction. Thus, quantitative gradation (Table 2-2) can mean either shortening or lengthening of the consonants. Similarly, qualitative gradation (Table 2-3) affects the consonants either by leniting (i.e. weakening) them or by strengthening them. The following section provides details regarding the direction of gradation.

### 2.1.2 The direction of gradation: Direct and indirect gradation

Gradation can be *direct* or *indirect* (or *reverse*) depending upon the direction of gradation, that is, whether the bare base form without any morphology (i.e. NOM.sg case) contains a stronger or a weaker obstruction than does its suffixed form. For example, *kukka~kukan* ‘flower.NOM.sg~flower.GEN.sg’ (see Table 2-4 below), exemplifies *direct* gradation because the base form contains a geminate [kk] while the suffixed form contains a short stop [k] (the former of the two being a stronger obstruction of airflow). The stronger obstruction is marked in bold in Table 2-4 for easy reference. Exceptions to the general patterns are marked with an asterisk (\*) after the word.

| Direct gradation (basic NOM.sg form ~ suffixed form) |  |                                  |
|--|--|----------------------------------|
| a. <b>kuk</b> ka ~ kulkan                            | ‘flower.NOM.sg ~ flower.GEN.sg’          | ~a flower (obj/poss)             |
| b. kier <b>l</b> tää ~ kierrän                       | ‘to.go.around.INF ~ go.around.pres.1sg’  | ~I go around.                    |
| c. hen <b>l</b> to ~ henlon                          | ‘slender.NOM.sg ~ slender.GEN.sg’        | ~slender (sg, obj/poss)          |
| d. hyl <b>p</b> iä ~ hylpin                          | ‘to.jump.INF ~ jump.pres.1sg’            | ~I jump.                         |
| e. tak <b>k</b> i ~ takkiin*~ talkin                 | ‘coat.NOM.sg ~ coat.ILL.sg ~coat.GEN.sg’ | ~into a coat ~ a coat (obj/poss) |

**Table 2-4: Direct gradation**

Conversely, a situation in which the obstruction is weaker in the basic form than in the suffixed form is called *indirect (reverse) gradation* (see Table 2-5 below). For example in 2-5 (a), the bilabial nasal geminate [mm] in the base form of *lumme* ‘water lily’ allows airflow through the nasal cavity which in the suffixed form *lumpeen* ‘water lily (object/possessive)’ is blocked by the bilabial stop [p]. In this case, the obstruction strengthens in the suffixed form.

| Reverse gradation (basic NOM.sg form ~ suffixed form) |  |   |   |
|---|--|---|---|
| a.  | lumme* ~ lumpeen*                          | ‘waterlily.NOM.sg ~ waterlily.GEN.sg’       | ~ a waterlily (obj/poss)                |
| b.  | rejlhas ~ rejlkka.i.lsiin ~ rejlkka.i.tta* | ‘tire.NOM.sg ~ tire.plu.ILL ~ tire.plu.ABE’ | ~ into the tires ~ without tires        |
| c.  | rilkas ~ riklkaan* ~ riklka.i.lta          | ‘rich.NOM.sg ~ rich.GEN.sg ~ rich.plu.PART’ | ~ rich (sg, obj/poss) ~ rich (plu, obj) |
| d.  | kerlra.lta* ~ kerltaan*                    | ‘to.review.INF ~ review.pres.1sg’           | ~ I review.                             |

**Table 2-5: Indirect (reverse) gradation**

The direction of gradation is easily seen in the following examples representing *kk~k* alternation. *Palkka~palkan* ‘salary.NOM.sg~salary.GEN.sg’ represents weakening gradation while *hake\*~hakkeen\** ‘(wood)chips.NOM.sg~(wood)chips.GEN.sg’ represents strengthening gradation. Regardless of the direction of gradation, a stronger consonant obstruction often precedes an open syllable (with no coda-consonant) while a weaker consonant obstruction is usually found preceding a closed syllable (with a coda-consonant). Exceptions to this general rule are commonplace (see cases marked with an asterisk) in both directions but more numerous in indirect gradation. Not only is indirect gradation less predictable but also less frequent than its direct counterpart. Table 2-6

below is an excerpt adapted from Hakulinen et al. (2004) showing an approximate number of entries in the *Comprehensive Dictionary* that are subject to direct and indirect gradation. An example of each alternation is in the parentheses. This thesis project excluded indirect (reverse) gradation and instead focused on direct quantitative and direct qualitative gradation only.

| Direct gradation (example)                             | # of entries | Indirect gradation (example)                               | # of entries | Total           |
|--|--------------|--|--------------|-----------------|
| pp:p<br>(hyppiä~hypin<br>'jump.INF~jump.pres.1sg')     | 274          | p:pp<br>(opas~oppaan<br>'guide.NOM.sg~guide.GEN.sg')       | 106          | 380             |
| mp:mm<br>(empiä~emmin<br>'hesitate.INF~hesitate.1sg')  | 25           | mm:mp<br>(kammata~kampaan<br>'comb.INF~comb.pres.1sg')     | 15           | 40              |
| nt:nn<br>(ranta~rannan<br>'beach.NOM.sg~beach.GEN.sg') | 1432         | nn:nt<br>(vanne~vanteen<br>'rim.NOM.sg~rim.GEN.sg')        | 114          | Approx.<br>1550 |
| ηk:ηη<br>(auriηko~auriηnon<br>'sun.NOM.sg~sun.GEN.sg') | 120          | ηη:ηk<br>(reηηas~reηηkaan<br>'tire.NOM.sg~tire.GEN.sg')    | 17           | Approx.<br>140  |
| lp:lv<br>(halpa~halvan<br>'cheap.NOM.sg~cheap.GEN.sg') | 13           | lv:lp<br>(kelvata~kelpaa<br>'satisfice.INF~satisfice.3sg') | 5            | 18              |

**Table 2-6: Approx. rates of occurrence of gradation types in the Comprehensive Dictionary**

### 2.1.3 Unpredictability of qualitative gradation

No simple rules can be postulated for consonant gradation. Although both types of gradation, quantitative and qualitative, have many exceptions to their general patterns, qualitative is more unpredictable than is quantitative gradation. Karlsson (1974, in Leiwo 1984) proposes that the unpredictability of qualitative gradation results from the process of lexicalisation. This is to say that qualitative gradation, particularly the indirect kind, is associated with specific words and gradation would be included in the mental

representations of the words which are subject to it. Lexicalisation of gradation would be able to explain how various exceptions emerge in the language and, as linguist Martin points out (p.c., October 16, 2008), why so much variation exists in its production even among monolingual adults. In Standard Finnish, individual variation occurs in many words, such as in *Riika—Riian/Riikan* when referring to the capital of Latvia, or *säkä—sään/säkän* “(horse’s) whithers.GEN.sg”. One limitation of Karlsson’s view is that lexicalisation applies to qualitative gradation only. Quantitative gradation is more predictable and can be better accounted for by postulating gradation rules; a unified explanation of both types of gradation would be preferable.

A growing number of exceptions to qualitative gradation indicate an increasing instability of this type of gradation in Standard Finnish (M. Leiwo, p.c., September 29, 2008; Karlsson 1974, in Leiwo 1984). Whole classes of nouns (left unspecified by Leiwo) are exempt from qualitative gradation (Leiwo 1984, p. 64). The phenomena imply that the language is changing to eliminate qualitative gradation.

#### **2.1.4 Implications for acquisition**

From the point of view of a learner of Finnish, consonant gradation is undoubtedly a difficult feature of grammar to master. Consonants seem to alter almost arbitrarily without any salient cue as to what they change into. The phoneme /k/, for instance, can lengthen, change into [v, j] or [ŋ], remain unchanged, or disappear altogether (Hakulinen et al. 2004) (see Table 2-7 below). For this reason it is known in some grammars as “a diabolical K”.

|    |                               |  |                             |
|----|-------------------------------|--|-----------------------------|
| a. | palkata~ palkkaan             | ‘hire.INF ~ hire.pres.1sg’                   | ~ I hire                    |
| b. | puku~puvut                    | ‘dress.NOM.sg ~ dress.NOM.plu                | ~ dresses, suits            |
| c. | kulkea~kuljen                 | ‘wander.INF ~ wander.pres.1sg                | ~ I wander (move aimlessly) |
| d. | aurinko~auri <del>ij</del> on | sun.NOM.sg ~ sun.GEN.sg                      | ~ the sun (poss/obj)        |
| e. | leka~lekat                    | sledgehammer.NOM.sg<br>~sledgehammer.NOM.plu | ~ sledgehammers             |
| f. | reikä~reiät                   | hole.NOM.sg ~ NOM.plu                        | ~ holes                     |

**Table 2-7: Diabolical K**

Despite the complexities, Finnish children learn by the age of five the principles of consonant gradation in the vocabulary they use (P. Lyytinen, p.c., September 30, 2008). Before gradation is mastered, child speech typically contains word-forms in which the suffix is correct but in which gradation in the stem has not yet been observed (ibid.). The error is induced by applying the suffix mechanically to new word bases. Children may also over-apply gradation and produce a non-standard forms, for example *\*muin* (cf. *mukin* in Standard Finnish) meaning “mug.GEN.sg” (Martin, M., p.c., October 16, 2008). Leiwo (1984) reports as typical that a child at the age of 18.5 months first produces correct quantitative forms with the object case, having learned them by rote but a month later starts experimenting with productivity and forming complex words by combining the suffix with different forms of gradation. The use of the suffix is mastered at this stage but gradation is not. At the age of 20 months, standard gradated forms reappear (Leiwo 1984, p. 67). In general, qualitative gradation causes difficulties for a longer period of time than does quantitative gradation (Leiwo 1984, p. 67). Evidence from Finnish child-language, Leiwo argues, indicates that both qualitative and quantitative gradation are acquired through experimentation with the patterns of productivity the child has observed

(ibid.). It is clear that both available strategies for forming complex words, whole-word access and access via combining parts (discussed in Section 2.3 below), are utilised at different stages of acquisition. Regardless of what strategy is used, exposure to gradated words is needed for children to experiment and eventually fully acquire consonant gradation.

In language contexts where exposure to native-like speech is reduced, learning consonant gradation can be expected to be challenging. Saarinen (to appear) conducted a preliminary investigation of the production of consonant gradation by one non-dominant speaker in an immigrant setting. The consultant for this pilot study was a second-generation speaker of Finnish, who had learned Finnish as a first language (L1) at home as a child but had later shifted to English as a dominant language. While speaking Finnish, the consultant produced native-like gradation with some words while leaving some ungradated. The observation is in accordance with Martin (1998; p.c., October 16, 2008) who also found gradation loss in the speech of immigrant Finns in the United States. She observed that more loss occurs with qualitative than with quantitative gradation. The complexities of Finnish consonant gradation and previous work done among immigrant Finns motivate the current research that investigates the possibility of a link between observed variation in gradation and frequency of use. Since the language of non-dominant speakers of Finnish often exhibits gradation-loss, it is reasonable to suggest that reduced exposure to native-like speech in minority language contexts might explain the variation.

This literature on consonant gradation gives rise to a research question *is direct quantitative gradation more successful than direct qualitative gradation?* This is

Research Question Five of the total of six questions. It was postulated to investigate the effect of gradation-types on gradation accuracy. Quantitative gradation is more common and predictable of the two types of gradation. Because direct (weakening) quantitative gradation keeps the quality of the consonant constant and only shortens the alternating consonants (e.g. *kk~k* alternation in *kirkko~kirkot* 'church.NOM.sg'~'church.NOM.plu', it is assumed that direct quantitative gradation manifests less gradation loss than direct (weakening) qualitative gradation, which changes the quality of the consonant to another (as *rp~rv* alternation in *turpa~turvat* 'muzzle.NOM.sg~muzzle.NOM.plu').

## **2.2 Language loss in minority language contexts**

### **2.2.1 Inter-generational language loss: Language shift**

Unless it is a rare language contact situation of balanced bilingualism, coexisting languages, in immigrant and indigenous contexts alike, develop patterns of strength and compete for domains of use (de Bot & Weltens 1991, p. 42). Typically, the language of the majority (L2) intrudes over time into the domains where the ancestral language (L1) was previously spoken as the dominant language. Over several generations, L2 becomes dominant in the speech community while L1 is heard less and less. This inter-generational language loss phenomenon is known as *language shift* (e.g. Dorian 1981; Fishman 2001; Köpke & Schmid 2004; Sasse 1992). In the process of losing language domains to L2, L1 gradually loses its functionality within the speech community; L2 becomes the language of daily communication while the use of L1 is no longer found practical. Many studies, such as in the case of French in Louisiana (Landry, Allard &

Henry 1996) and on Turkish in France (Yağmur & Akinci 2003), reported that shifting from L1 is apparent among the young and that the preservation of the L1 language communities in the future is threatened due to low linguistic vitality. Socio-economic and socio-psychological pressures contributing to the speed at which language shift advances vary from one minority context to the next but, as a general rule, the shift from dominant L1 to dominant L2 is complete within two to three generations (Alba, Logan, Lutz & Stults 2002; Fishman 2001; Gonzo & Saltarelli 1983; Sasse 1992).

### **2.2.2 Intragenerational language loss: language attrition and imperfect learning**

Decrease in L1 use at the societal level leads to loss in the grammatical structure of L1 at the level of the individual (Seliger & Vago 1991; Waas 1983; Yağmur, de Bot & Korzilius 1999). A contributing factor is *language attrition* or, as defined by Köpke & Schmid (2004, p. 5), “the non-pathological decrease in proficiency in a language that had previously been acquired by an individual, i.e. intragenerational loss”. The term is used to refer to the loss of both L1 and L2 in native or non-native contexts, as shown below in Table 2-8 (from Köpke & Schmid 2004).

| Language Lost | Linguistic environment |                                     |
|---------------|------------------------|-------------------------------------|
|               | L1                     | L2                                  |
| L1            | Dialect loss           | L1 attrition                        |
| L2            | L2 attrition           | Language reversion (in the elderly) |

**Table 2-8: The van Els taxonomy.**

L2 attrition in L1 contexts refers to foreign language-learning, in which language learners lose their L2 due to lack of use. As another type of language loss, *dialect loss*

refers to overall leveling of dialects in L1 contexts, for example, due to a more prestigious language variety taking over. Language reversion, in turn, affects elders in immigrant settings where L2 erodes for example due to dementia. The present project falls in the highlighted area of study, focusing on L1 attrition in an L2 environment. According to Pavlenko (2004), language attrition differs from other language contact phenomena, such as borrowing or code-mixing, in that it refers to more or less permanent language loss, or decrease in language skills. Other language contact phenomena do not necessarily replace the L1 grammar but could also enrich or transform it (p. 56). Also *imperfect learning* is likely to be a contributing factor in the structural loss. L1 is passed on to the later generation in an affected form due to the fact that the donor generation never acquired the ancestral language fully (de Bot & Weltens 1991, p. 42).

Three decades of attrition research has investigated the involvement of several socio-cultural, psycholinguistic and linguistic factors in trying to establish a set of reliable predictors of attrition. Of various socio-cultural factors, *education level*, *age at the onset of bilingualism or of attrition*, and *contact with L1* have emerged as key variables linked to attrition, which also have psycholinguistic and linguistic repercussions. Despite the fact that the variable “education level” has produced some ambivalent findings (likely to due the methodological difficulties in controlling the variable), research by Jaspaert and Kroon (1989), Waas (1996), and Yağmur, de Bot, and Korzilius (1999) among others has found that formal schooling in L1 appears to increase the level of awareness regarding a language and therefore protect the implicit language knowledge from L2 interference. In the above studies, higher education correlated with better performance in text editing, vocabulary, and verbal fluency tasks. Although worthy

of mention here as one of the determinants of attrition, *education level* is not a variable of interest in the present project. Education level and the role of school as a preserver of a language is not relevant for the purposes of this project. In Sointula, making a living has traditionally relied on fishing, forestry, crafts, and manufacturing. For this reason, many Sointulans did not need formal schooling beyond primary education to make ample living (P. Kiiskilä, p.c., March 19, 2008)<sup>2</sup>. Primary education on Malcolm Island has been offered in English since the late 1920s or early 1930s (I. Belveal, p.c., February 4, 2009).

The second factor, *age at contact*, appears to largely determine to what extent an individual's language is affected by attrition. Findings such as those of de Bot and Cline (1994), who found that some German speakers in Australia retain their fluency despite the language contact, suggest that a threshold period exists during which L1 becomes fixed. If language skills are maintained during that time, they might be less vulnerable and saved from attrition even if the language environment changes. Therefore, a variable *age at the onset of bilingualism* (also known as *age at the onset of attrition* in some research) has bearing in the attrition process (Köpke & Schmid 2004, p. 20). The research has established that the earlier the language environment changes, the more severe (e.g. Seliger & Vago 1991) and the quicker is (Hamideh & Hamideh 2006) L1 attrition.

*Contact with L1* — the frequency of opportunities that a person has to use his or her L1 — is the last and most relevant of the socio-cultural variables discussed here. It includes language exposure through interpersonal communication, reading and writing, and contact with media (Landry, Allard, & Henry 1996). The amount of language contact is invariably considered in attrition research as a determiner of language skills in minority

---

<sup>2</sup> In 1970, Sointula had the highest per capita income of any community in Canada (Crawford 2008).

language communities under pressure. Limited L1 contact has been shown by many to lead to atrophy of L1 structure (cf. Köpke & Schmid 2004 for a list of references).

Within psycholinguistics, Andersen (1982) found that low-frequency items are more prone to attrition. This has been supported by experimental data from lexical retrieval (Hulsen 2000; Yağmur et al. 1999). Studies have found that frequently activated memory items need less priming than infrequently activated items to reactivate them, leading to the Activation Threshold Hypothesis in 1985 by Paradis (Paradis 2004). This theory maintains that each activation of a memory increases its relative strength while infrequently used items become weaker and fade from memory. Attrition, according to Paradis, is “long-term lack of stimulation” (p. 28).

Another research question emerges from the above literature on language loss. Research Question Six was formulated to examine the relationship between language loss due to reduced language contact and gradation-accuracy: *does immigrant generation explain the variation in the gradation-accuracy between the participants?* Three generations are included in the present project, G(eneration)2, G2.5, and G3. The participants in G2 have two Finland born parents and represent the first generation born abroad. G2.5 participants have one Finland-born parent and one parent (of Finnish heritage) born abroad. Both parents (of Finnish heritage) of G3 participants were born abroad (see Section 3.1). Since exposure to native-like speech reduces in every generation born abroad, participants’ gradation accuracy is assumed to reflect the amount of language contact the participants have received. Earlier generations are assumed to produce more instances of standard gradation than later generations, having received

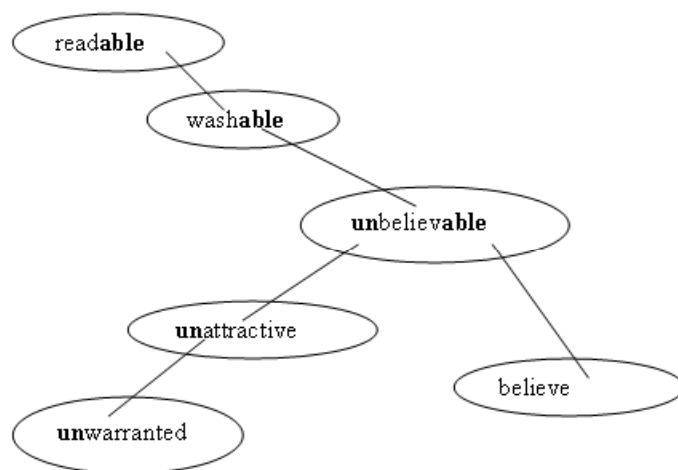
more exposure to native-like speech at home. G2 is hypothesized to be least affected by gradation loss while G3 participants are expected to show most atrophy in gradation.

Psycholinguistic research among *dominant* language speakers has arrived at the same conclusion with L1 attrition research among *non-dominant* speakers that exposure is needed to keep memories alive and active. Although the link between human cognition and frequency of use is hardly deniable, the notion of frequency has only recently started gaining acceptance within linguistics. This approach to grammar is now known as usage-based grammar, which will be introduced next.

### 2.3 Cognitive view of language: Usage-based grammar

Frequency does not fit well in the workings of traditional generative, rule-based, formal models of grammar because they postulate that the structure of human language is independent of language use and that the phonetic shape of words is strictly an outcome of rules and/or constraints which apply **after** the retrieval of the word from the lexicon. Once the conditions for change are met, the rules apply without exceptions. Thus, the pronunciation outcome of a word is considered to be separate from the mental representation for that word. These approaches view linguistic variation as idiosyncratic, as rules cannot account for it. In contrast, cognitive grammarians argue that phonetic detail must be **stored in the lexicon** associated with the corresponding lexical items because the frequency of language use affects the pronunciation, organization, storage and access of memories of words/word-strings (Bybee 2001, p. 20).

Usage-based grammar posits that the structure of language arises from language use. The mind has an astonishing ability to keep record of one's encounters with word forms and grammatical structures (see Bybee 2001 for an excellent overview of Usage-based grammar and studies done in the field; see also Barlow & Kemmer 2000; Langacker 2000, 2002). The mind's capacity to make generalisations, discern patterns and create connections between lexical items from the gathered linguistic experience is "the human endowment for language" (Bybee 2001). Similar memories are linked to each other through a mental web (see Figure 2-1 below from Bybee 2002, p. 273). Some memories overlap and take less memory space. Thus, the mental web is highly embedded. Lexical connections between the frequently encountered patterns become well established in the speaker's grammar, while less frequent patterns do not. New words can be constructed from frequently occurring parts, using other words in the lexicon as models.



**Figure 2-1: Lexical associations of the word 'unbelievable'**

The content of mental representations and the connections between memory tokens are coloured by speakers' own linguistic experience. This experience can be observed in speech (Pierrehumbert 2000). Thus, what gets stored and what shape the stored items take depend on language use. Variation is thus inherent in language. Many studies (Bybee 2000; Jurafsky, Bell, Gregory & Raymond 2001; Phillips 1984) have shown that much linguistic variation is motivated by frequency of use. This previous research also underlies the current project. Researchers studying the effects of frequency on words<sup>3</sup> have been able to show that frequency of language use affects the phonetic shape of words and word-strings either by introducing variation into their mental representations or preventing variation from entering into them. The two opposite processes are known together as frequency-effects.

### **2.3.1 Frequency effects of frequent and infrequent items**

Frequently used lexical items are said to have high *token-frequency* (i.e. the rate at which a word-form or a word-string is encountered in a language) because many people use them often. High token-frequency has been found to affect the phonetic shape of words in three ways (Bybee 2007, pp. 9-14). First, frequent words tend to exhibit more innovative pronunciations and are generally more likely to undergo reductive changes<sup>4</sup> than are infrequent memory tokens (e.g. “I dunno” for “I don’t know”). Second, high-frequency use strengthens the mental representations of lexical items and makes them

---

<sup>3</sup> A ‘word’ is a gradient concept. It is a unit of storage whose size is largely determined by usage. Thus, a string of words counts as a word if the word-string is perceived as a unit.

<sup>4</sup> This type of sound change arguably has physiological motivations. Rehearsal of speech gestures increases the fluidity of speech, as it would affect any neuromotor activity, e.g. piano-playing, showing the benefits of high frequency use (Pagliuca & Mowrey 1997).

more accessible. Speakers do not have difficulties retrieving high-frequency tokens from memory and producing them even if they contain irregularities and exceptional material. For instance, most English irregular verbs are high-frequency words whose irregular paradigms are easily accessible (e.g. be – was – were – been). Thus, high-frequency use supports preservation of irregular forms. Third, frequent items gain independence (i.e. autonomy) from other items in the lexicon (Bybee 2007, pp. 9-14; Bybee & Brewer 1980, p. 50). The more frequent the lexical item is, the less connected it needs to be with other items in the lexical network. A fully autonomous lexical item is a whole memory unit with its own lexical representation and it can be accessed independently from other tokens (Bybee 2007, p. 14). Thus, autonomous lexical items do not contribute to the network — even if they were “either words of the same paradigm or words of the same lexical class” (Bybee 2007, p. 13). They may also break off from other related tokens, as *be going to* has broken off from other uses of *go*. The second and third effects of high frequency are relevant in this study; Finnish complex word-forms might be accessed as wholes in the lexicon if they are high-frequency tokens with strong mental representations. As such, frequent complex word-forms should also produce accurate gradation because gradation would be accessed together with the whole independently of other stored tokens in the lexicon.

Infrequent items, in turn, do not benefit from rehearsal to the same extent as do frequent items. Thus, infrequent items tend not to exist in the lexicon as whole word-tokens or their mental representations tend to possess less strength. According to Bybee (2002), low-frequency tokens “may not be sufficiently available in [linguistic] experience to be acquired and entrenched” (p. 270). Consequently, memory often fails to remember

their exact phonetic shape or the word altogether. If possible, they have to be constructed from commonly occurring parts but in the process, they may undergo sound changes that are “based on the general patterns of the language” (ibid.). Infrequent words tend to be replaced by a more common paradigm, a phenomenon also known as *analogical levelling* (Bybee 2001, p. 108) because exceptional patterns are hard to remember. Thus, sound changes affecting infrequent items have *cognitive motivations* (Phillips 1984) rather than physical motivations of frequent words, which tend to undergo reductive changes due to increased fluency. Hooper (1976) demonstrated how infrequent irregular verbs, such as *weep* in the past tense, tend to regularize by analogy into *weaped* (cf. *wept*) since *-ed* is more regular and the most frequent pattern to form the past tense in English. Infrequent *weep* contrasts with frequent verbs such as *keep* and *sleep*, which resist analogical changes. Their past tense forms *kept* and *slept* are so ingrained in the speakers’ lexicons thanks to their high-frequency use that *keeped* and *sleeped* would sound ungrammatical to a native ear. According to Bybee, an analogical formation always uses the base form as its basis, as shown in Table 2-9 (adapted from Bybee 2001, p. 108). The base form is the unmarked or the most frequent form.

| Original forms |             | Analogical forms   | Non-occurring forms |
|----------------|-------------|--------------------|---------------------|
| Base form      | Stem change | Base form + suffix |                     |
| roof           | rooves      | roofs              | *roove (singular)   |
| heave          | heft        | heaved             | *hef (present)      |
| weep           | wept        | weaped             | *wep (present)      |

**Table 2-9: Analogical levelling**

When the phonetic environment of a word changes (e.g. through suffixation) and triggers an alternation in the stem, the base form is chosen as the most basic stem as it no longer has the alternation (p. 114). The base form stem is combined with the suffix (e.g. weep + ed = weeped). Thus, the use of analogical formations implies that a complex form is pieced together from its constituent parts via decomposition.

The data from the pilot project (examining a non-dominant speaker’s production of gradation (Saarinen, to appear)) manifested gradation loss as an analogical formation. Instead of using gradated stems, the consultant chose to use non-gradated base forms as stems in unfamiliar multi-morpheme words. Table 2-10 below illustrates variation (in bold) in gradation in the consultant’s speech (SF = Standard Finnish).

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

**Table 2-10: Variation in consonant gradation in non-dominant Finnish**

In the column marked as ‘non-dominant Finnish’ in the above table, it can be seen that the variant word-forms use the base-form as the stem to which the suffix is attached. The base form found in the INF(INITIVE) (verbs) and NOM(INATIVE) (other word classes) is the most frequent case in Standard Finnish. Table 2-11 below lists the most

frequent cases in Finnish (adapted from Hakulinen et al. 2004). The highest percentages for each word class are bolded.

|              | Nouns       | Adjectives  | Demonstrative<br>Pronouns | Personal<br>Pronouns | Verbs                |
|--------------|-------------|-------------|---------------------------|----------------------|----------------------|
| NOM(INATIVE) | <b>26.3</b> | <b>32.2</b> | <b>32.1</b>               | <b>47.7</b>          | n/a                  |
| INE(SSIVE)   | 6.7         | 4.5         | 8.3                       | 0.6                  | 8.6/3.0 <sup>5</sup> |
| ILL(ATIVE)   | 6.3         | 5.3         | 6.9                       | 1.0                  | 16.7                 |
| ALL(ATIVE)   | 2.4         | 1.6         | 1.7                       | 7.3                  | n/a                  |
| infinitive   | n/a         | n/a         | n/a                       | n/a                  | <b>55.4</b>          |

**Table 2-11: Most frequent case forms in Standard Finnish (%)**

In choosing NOM case base-forms as stems, the consultant chose the pattern that is the most frequent and therefore presumably the strongest in the mind and perceived as the least complicated option. Thus, analogical leveling of infrequent words is highly pertinent to the current study, which sets out to examine non-dominant Finnish speech again.

The following section presents a theoretical model that is capable of conceptualizing the fact that language use affects the mental representations of words and speakers' grammar.

### **2.3.2 Exemplar Dynamics**

Exemplar Dynamics was initially proposed as a model of perception but developed later to apply to production (Hintzman 1986; Johnson 1997; Pierrehumbert

<sup>5</sup> The first number refers to so called E-infinitive (e.g. syödEssä 'while eating'), while the second number refers to MA-infinitive (e.g. syöMÄssä 'to be eating').

2001). It posits that an individual stores gathered linguistic experience as memory tokens (or exemplars). Linguistic experience refers to a person's encounters with full word forms, morphemes, phonemes, word-strings, or prosodic information (i.e. pitch, loudness, tempo and rhythm) of which a memory is formed. Whatever category an incoming stimulus belongs to, it merges with an old one if the difference between the two is too small to notice. Conversely, a related token with a noticeable difference is stored in close proximity. Completely new items form new memories. The distance between memory tokens represents similarity or dissimilarity between them. Memory tokens close to each other represent all the stored variants that are perceived to belong to a single category. Distance to another collection of tokens signifies the difference between the two categories. Consequently, clusters of related memories emerge. Each cluster represents a different category. A category functions as a parameter space with different values within it. An individual compares the properties, or the exact quality, of a new token entering the cluster against the values within the parameter space and determines a position for the new token within the cluster. A category label for the whole cluster is drawn from the essential properties of the tokens stored in the mnemonic assemblage, as if calculating an average of a set of different values<sup>6</sup>.

A memory token gains strength through activation. The more recently and frequently a token is activated, the higher its activation level is. A new token entering the cluster is likely to be analyzed as another example of an existing memory token with a high activation level. The number of tokens within the cluster, their exact quality, and their activation level (or strength) comprise the forces that fix the position of a new token

---

<sup>6</sup> This mechanism is known as 'entrenchment'. Entrenchment stabilizes the amount of variation present in a speech community at any given time so that the perceived and produced categories of individuals match for the most part. Input that cannot be identified is ignored (Pierrehumbert 2001, p. 150).

in the parameter space. A token may be simultaneously associated with multiple categories dedicated to different aspects of linguistic information. Sparse clusters of tokens represent categories 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.

According to the Exemplar Model, perception of speech is essentially categorization of an incoming speech signal. A listener evaluates the quality of an incoming stimulus against the quality of the category labels. The category that represents the stimulus most closely wins the competition and the incoming token adds to the overall strength of the category either by reinforcing an existing exemplar or by forming a novel memory in the exemplar cluster (which is also known as *neighborhood*) (Pierrehumbert 2001, p. 144).

In speech production, an individual selects and activates a category label. The production target is drawn from a selection of stored exemplars within the targeted category. Social and stylistic factors guide the selection as does the activation level of the tokens in the neighborhood from which the sampling is taken. The exemplars within the selection are averaged to get the production target (p. 150).

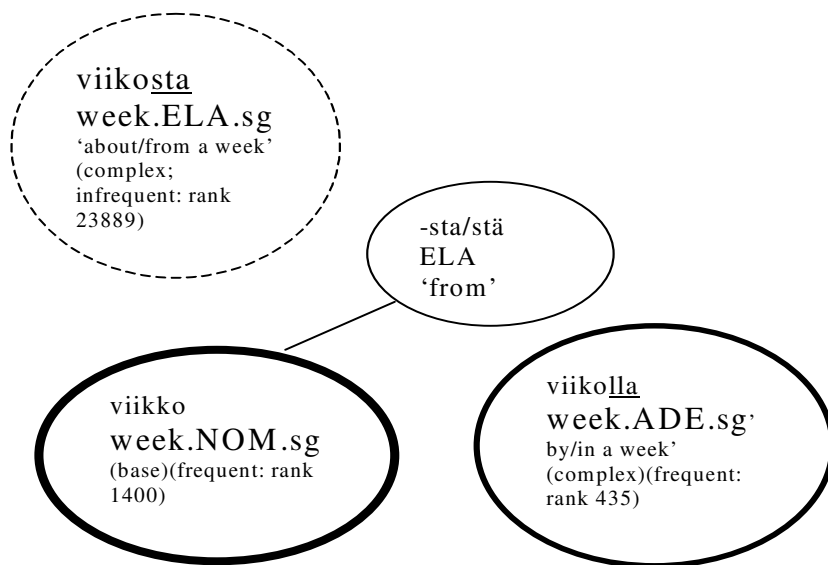
Through the mechanisms explained above, a memory deviating from the norm, for example a word affected by gradation loss, may gain strength through frequent and/or recent activation and be selected. If other speakers in the speech community use the same deviant form, its representations grow stronger. An individual can also add to the strength of their own memory tokens; if deviant forms are well established in a speaker’s own grammar, they get stronger if the representations are not updated with standard forms. A

situation such as this could well arise in non-dominant language contexts. Exemplar Theory, as presented above, is adopted for the present project since it “predicts precisely the kind of increased variability observed in language shift situations” (Bird 2008, p. 389).

Four research questions have arisen from the literature on usage-based grammar. Research Question One asks *does whole word-frequency affect achieving whole word-accuracy?* Whole word-frequency refers to the frequency of a single-unit word-form. The words for ‘beach’, *ranta* ‘beach.NOM.sg’ (infrequent; rank: 7297), and for ‘on the beach’ *rannalla* ‘beach.ADE.sg’ (infrequent; rank: 3139) (see section 3.2.1 for details on frequency-categories), are examples of whole words. According to Bybee (2001, 2007), frequently activated high frequency word-forms have stronger mental representations than infrequent word-forms. Consequently, frequent word-forms gain autonomy and can afford to have weaker connections to other stored items in the lexicon than infrequent word-forms. While autonomous frequently used word-forms can be accessed as wholes, infrequent word-forms are composed of parts because no whole-word mental representation exist or can be found for them in the lexicon. Common patterns and regularities shared by other stored items in the lexicon are used as models in constructing complex words from pieces (see Figure 2-1). Thus, infrequent words are prone to analogical changes.

Applied to non-dominant Finnish, Bybee’s predictions would mean that high-frequency complex word-forms including gradated stems and suffixes should be accessed as wholes and thus produce more accurate whole word responses than infrequent words. Frequent use has strengthened the mental representations of frequent complex words to

the extent that they exist as whole word memory tokens independently from other lexical items. Conversely, low-frequency complex word-forms do not exist as whole word representations in the mental lexicon or whole word representations cannot be found. Therefore, low-frequency complex words have to be constructed from parts, i.e. non-gradated NOM stems (as the most frequent word-shape) and suffixes. This would predict more instances of gradation loss as analogical change. Figure 2-2 below help demonstrate the link between frequency and lexical access.



**FIGURE 2-2: Exemplar associations in the mnemonic network**

*Viikko* is the ungradated, unsuffixed base form meaning ‘week’. The complex word-form *viikolla* contains the gradated stem *viiko-* and the suffix *-lla*, meaning “by/in a week”. Both are high-frequency words and therefore stored as whole units, indicated by solid dark circles around them (base word-forms, e.g. *viikko*, are single morpheme words and automatically accessed as whole). The base form *viikko*, as a NOM form (the most common case), has the strongest mental representation represented by the darkest circle.

Crucially, *viikolla* as a frequent complex word is accessed as a whole independently from other tokens and gradation within it is preserved thanks to whole-word access. Therefore it does not need to be connected in the network. Conversely, the gradated whole word representation for “from a week”, *viikosta*, is presumably not available as it is an infrequent word, indicated by a dotted circle around it. It has to be built from parts. In order to do that, the most common case type, un-gradated and unsuffixed base form ‘week’, *viikko*, is combined with the suffix *-sta* meaning ‘from’. However, the resultant complex word *\*viikkosta* is a non-standard form manifesting analogical leveling because the base word-form is non-gradated. In other words, gradation is lost in the process of piecing the complex word together from parts. In the current research, gradation with its presence or absence is taken to indicate what mode of access, decomposed or whole word access, was used to produce the word.

Based on these frequency-effects, it is expected that frequent whole words produce more accurate whole word-form responses than infrequent whole words. Participants’ responses are compared with the targeted whole word-forms. Two types of targeted word-forms will be tested, NOM.sg base word-forms and their complex counterparts. Testing the accuracy of the whole word-unit determines a) how accurately the participants can produce the targeted whole word-forms in their sentence contexts and b) whether the gradated stem-allomorph (of complex words) is erroneously used as a word in isolation, for example *ranna* instead of the targeted *ranta* ‘beach.NOM.sg’ (cf. *rannalla* ‘on the beach’).

Research Question Two asks *does whole word-frequency affect achieving gradation-accuracy?* Following Bybee’s predictions above (see Research Question One),

frequent whole words are assumed to manifest more accurate gradation than are infrequent whole words. Gradation is assumed to be lexicalised in frequent whole words and accessed with the whole. Conversely, infrequent whole words are assumed to have weaker mental representations and be more prone to analogical leveling. Thus, they are expected to manifest more gradation loss than frequent whole words. This question focuses on the accuracy of gradation with respect to the stem only and disregards suffixes. Examining gradation independently of suffix-choice makes assessment of gradation-accuracy with various suffix-errors possible.

Research Question Three asks *does suffix-frequency affects achieving gradation-accuracy?* The presence of a gradation-triggering suffix is needed for gradation to take place in the stem. This motivates measuring the effect of suffix-frequency on gradation-accuracy. However, it is unlikely that case suffixes, whether frequent or infrequent, help in achieving gradation in the stem since the consonant alternation is not predictable from the suffix and some stems inherently remain outside gradation, as was discussed in Section 2.1.

Research Question Four asks *does stem-frequency affects achieving gradation-accuracy.* NOM.sg stem word-forms are independent word-forms, e.g. *ranta* “beach.NOM.sg”. Their gradated allomorphs are found with a suffix in complex word-forms. For example, the complex word-form *rannalla* ‘beach.ADE.sg’ meaning ‘on the beach’ consists of the gradated stem *ranna-* and the suffix *-lla*. The research question asks if the non-gradated stem word-forms aid in accessing, or prime, gradated allomorphs. Bybee (2001) writes that phonological and semantic similarity induces priming effects. Thus, *teach* primes *teacher* (p. 25). However, it is unlikely that

ungradated NOM.sg case stem word-forms, e.g. *poika* ‘boy’, help prime gradated complex words, e.g. *pojasta* ‘from/about a boy’, to a significant degree because the allomorphs, *poika* and *poja-*, differ phonetically and also by their positions in sentences. Moreover, complex gradated word-forms may not be available in the lexical network, frequent words due to autonomy and infrequent words due to infrequency and low activation levels. Gradation-accuracy is expected to run with whole words and not with their parts, stems or suffixes.

## 2.4 Summary

Finnish consonant gradation cannot be easily explained through rules because the alternating consonants have a multitude of outcomes which vary in their quantity or quality. Many exceptions to the general patterns further obscure the process. Therefore, the acquisition of consonant gradation can be challenging in language contexts where exposure to native-like speech is reduced, such as in Sointula, BC, or in other immigrant communities outside Finland.

Reduced exposure has been linked to language loss (Köpke & Schmid 2004) in speech communities which are experiencing language shift. Language shift refers to a process in which the ancestral language (L1) is being replaced by the language of the majority (L2) over several generations (Brezinger & Dimmendaal 1992; Dorian 1981; Fishman 2001; Sasse 1992). At the level of the individual, language loss manifests itself as structural changes in speaker grammar in which both language attrition (i.e. atrophy of language skills within a generation) and imperfect learning (i.e. incomplete language acquisition) are likely to have a role (de Bot & Weltens 1991). Lack of language

exposure in non-dominant language contexts and the resultant variance in speaker grammar due to reduced frequency of use can be linked with variation studies that examine frequency of use as a force shaping human speech. However, frequency of use does not factor in the studies of language structure conducted in traditional generative fashion. Thus, Usage-Based Grammar and the Exemplar Model were adopted.

Human cognition refers to our ability to perceive, categorise and remember information. Linguistic information is not an exception. Usage-based grammar is *cognitive grammar* in maintaining that linguistic structure emerges from language use through human cognition (Barlow & Kemmer 2000; Bybee 2001; Langacker 2000, 2002). Frequency of exposure is the key notion in the process. Several studies have shown that frequency of use has an effect on the mental representations of words and word strings (e.g. Bybee 2000, 2001, 2006; Hooper 1976; Phillips 1984). Because frequency of use affects word-shape, word-specific phonetic detail must be stored in the lexicon with the particular lexical item. Three frequency-effects are most relevant to the present project: 1) high-frequency use reinforces mental representations of words and, although they may contain exceptional material, they are easily accessed and produced, 2) frequent items tend to be stored and accessed as wholes, and 3) infrequent items manifest generalizations by analogy because exceptional forms are hard to remember. Analogical formations always use the base-form (the most common form) as their basis.

This project takes the position that a speaker's grammar acquired through experience is mapped to memory exemplars and organised into clusters of related memories. Exemplars gain strength through activation. The Exemplar Model is well

suited for the present project as it can predict variation that stems from reduced exposure to L1 in minority language contexts (Bird 2008).

The current study arose from the literature reviewed in the above sections and set out to examine if a link exists between frequency of use and the preservation of consonant gradation in the speech of Sointula Finns.

## **2.5 Research statement and summary of research questions**

The purpose of this study was to determine whether the frequency of word-forms, suffixes and stems affects the production of consonant gradation in the immigrant community of Sointula. It also investigated if one gradation-type, quantitative or qualitative, is more successful than the other. Lastly, it examined whether immigrant generation can explain the variation in the production of gradation between the participants.

For the purposes of this study the following research questions have been formulated. They are summarised in Table 2-12 below.

|  |
|--|
| <ol style="list-style-type: none"><li>1. Does whole word-frequency affect achieving whole word accuracy? <i>Yes. More standard gradation is expected with frequent words than with infrequent words.</i></li><li>2. Does whole word- frequency affect achieving gradation-accuracy? <i>Yes. More standard gradation is expected with frequent words than with infrequent words.</i></li><li>3. Does suffix-frequency affect achieving gradation-accuracy? <i>No. It is expected that no significant difference in gradation is found between frequent and infrequent suffixes.</i></li><li>4. Does stem-frequency affect achieving gradation-accuracy? <i>No. It is expected that no significant difference in gradation is found between frequent and infrequent stems.</i></li><li>5. Is direct quantitative gradation more successful than direct qualitative gradation? <i>Yes. Direct quantitative gradation is more successful in producing gradation than direct qualitative gradation.</i></li><li>6. Does immigrant generation explain the variation in the production of gradation between the participants? <i>Yes. More gradation can be expected with earlier generations than later generations.</i></li></ol> |
|--|

**Table 2-12: Research questions and predictions**

## **Chapter Three Methodology**

This project examined consonant gradation in the speech of Sointula Finns who are losing their previously dominant heritage language to English. The aim of this research was to determine whether frequency of use of whole word-forms, suffixes and stems affects the preservation of gradation in complex words consisting of multiple morphemes. An experiment was conducted in the community for this purpose. This chapter describes in detail the methodology used for the collection and the analysis of the data. Section 3.1 describes the recruitment process and provides the participants' background information. Section 3.2 describes the instruments, the sentence list and the questionnaire. Section 3.3 outlines the procedures of investigation. Section 3.4 describes coding of the data. Lastly, Section 3.5 outlines the variables formulated for the purposes of the present project and describes the statistical analyses performed on the data.

### **3.1 Participants**

There were two sets of participants for this project. The participants for both the pilot study and the main study in Sointula were selected on the basis of the following six criteria. First, all participants were to be second or later generation Finns with Finnish blood-heritage from both parents not older than 100 years. It was reasoned that family origins older than a century might no longer play an active enough role in the lives of later immigrant generations. In the current study, second generation (G2) refers to the children of the actual immigrant generation G1. G2 is the first generation born outside

Finland, with both parents born in Finland. G2.5 has one parent born in Finland while the other parent was born abroad. G3 participants have two parents born abroad. Second, if the parents and grandparents of all participants did not immigrate directly to Sointula, they must have come via an English-speaking country. Third, the participants for the main study were to be born and raised in Sointula, BC and have learned Finnish at home. Fourth, the participants were also expected to be above the age of 16, at which age parents' consent for participation in this study is no longer needed. Maximum age was not set. Fifth, the participant was to be able to participate in the study unaided and be free from any known medical conditions that would affect language performance. Finally and most importantly, the participants were to possess a language proficiency level that could be characterized as less than that of a native speaker but at the same time good enough to allow the participants formulate Finnish sentences and comprehend when spoken to in Finnish.

Participants in the study were found through the researcher's connections in the Finnish community in Victoria and in Sointula. The target was to find three participants for piloting of the instruments in Victoria. However, the piloting of the instruments was conducted with two participants due to an unexpected last-minute cancellation. The volunteers, a female and a male aged 57 and 64 respectively, represented generations G2 and G2.5. For data collection in Sointula, the target was to find 6-8 participants. No compensation was given for participation. Prospective participants were contacted by phone. The phone call was conducted in Finnish in order to get a first impression of the potential participants' language ability while inquiring about their willingness to participate in the study.

In Sointula, a total of 17 individuals were contacted, 14 of whom the researcher organized to meet during a 6-day field trip to the community in late-March, 2008. Four of these individuals had native-like command of Finnish. Three of these exceptional speakers represented G3, while one was a G2 speaker. These elderly Sointulans' ages were in their 70s and 80s, which made it clear that the Finnish community in Sointula in the first half of the 1900s was strong enough to ensure language transfer from one generation to the next as far as G3 and to maintain the acquired language skills beyond the critical period for language. These elders were not suitable participants for the present project due to their impeccable language skills. The number of participants reduced further as one individual had to be excluded from the study due to the fact that Finnish was not spoken at home after all. This individual had learned Finnish as an adult during an extended stay in Finland<sup>7</sup>. Yet another participant's language was in such an advanced state of attrition that this participant was unable to recall vocabulary needed to construct the test sentences. Inclusion of this individual's data was not possible. A mishap in recording in one case and a medical condition in another prevented the participation of two more individuals. The remaining six participants' language data was used in the study. These participants represented the three generations targeted, G2, G2.5 and G3.

Due to the unexpected decrease in the number of participants, balancing the number of women and men in Sointula was not possible. The six participants, all born and raised in Sointula, included 4 women and 2 men, of whom two represented G3, two G2.5, and two G2. G2 included an individual from G2.5 who received G2 exposure to Finnish as a child from a monolingual Finland-born grandmother living at home. Two of

---

<sup>7</sup> Interestingly, his language differed from the rest in that contrastive consonant length was missing from his speech; all of his consonants were short to the naked ear. Other participants, who had learned Finnish at home, were able to distinguish between long and short consonants.

the participants were mother and daughter. The mean age of the participants was 67.6 years.

A control group of dominant Finnish speakers was not included in the study. The researcher is a native of Finland and fluent Finnish-speaker, who emigrated from Finland in her mid-20s. Her knowledge of Standard Finnish was taken as a base line to which each participant's production of consonant gradation was compared. Variation in gradation is easy for a native speaker to detect auditorily — as easy as it would be for a native English speaker to detect the difference between *beat* and *bit*.

## **3.2 Instruments**

The instruments include a sentence list and a questionnaire. They are included in Appendices A and B.

### **3.2.1 Stimuli: Sentence list**

The development of the sentence list started with the creation of a list of suitable words from a Finnish word-frequency corpus. With the help of linguist, Eero Vitie, at *Kotimaisten kielten tutkimuskeskus (The Research Institute for the Languages in Finland)*, a Finnish word-frequency corpus (Nykysuomen taajuussanansto [Modern Finnish Lexicon]) was found online at <http://kaino.kotus.fi/sanat/taajuuslista/parole.php>. The corpus consists of 1,339,787 word-form tokens compiled from newspapers, periodicals, and magazines (i.e. a parole corpus) between 1996-1998. The corpus lists Finnish words by decreasing frequency. Two smaller versions of the large corpus were

also available. The smallest version, from which words appearing only once or twice had been removed, was sufficient for the purposes of this study. It contained 362,514 word-forms. For each word, the corpus provided the rank, frequency of occurrence in every million words, and the proportion that the word represented of all the words in the corpus. The first objective was to find 60 complex word-forms that met pre-determined requirements regarding word-class, direction of gradation, the frequency of the word-form as a whole, frequency of the suffix and frequency of the stem, and gradation-type.

Suitable complex word-forms were nouns subject to direct (weakening) gradation. Word-forms that ranked between 1 and 1,558 were considered frequent and those with a rank above 3,099 were infrequent. The most common frequent word on the word-list *helsingissä* ‘helsinki.INE.sg’ is encountered 11,662 times per every million words. It is the 91. most common word in modern Finnish. The least common frequent word on the list, *pojan* ‘boy.GEN.sg’, is mentioned 1,111 times per every million words. A space of 1,541 words was left between frequent and infrequent words to make the frequency limits more defined. No word-forms were chosen from this intermediate range. The most common infrequent word is *loppu* ‘end.NOM.sg’, ranking as the 3,099. most frequent word. It is mentioned 596 times per every million word. The most infrequent word tested *keitossa* ‘soup.INE.sg’, which is mentioned only 4 times per every million words, ranks as the 230,650. most frequent word in modern Finnish.

Suitable complex word-forms contained either frequent or infrequent gradation-triggering suffixes. Suffixes that do not trigger gradation were not included. Hakulinen et al. (2004) was consulted in choosing the suffixes (see Table 2-11 in Chapter Two).

NOMINATIVE plural<sup>8</sup> and GENITIVE cases were selected as frequent gradation-triggering suffixes. NOM represents 37.4% and GEN 21.7% of all suffixes in Finnish parole. Infrequent suffixes used were INESSIVE, ADESSIVE, ELATIVE, ABLATIVE, ALLATIVE, and TRANSITIVE, whose use in Finnish parole ranges from 5.8 to 1.0%. The number of frequent and infrequent suffixes was balanced. However, it was not possible to balance the use of each case within those broad frequency categories as there would have not been enough word-tokens within the first 1500 most frequent words to meet all the requirements if case was included in them. For example, in order to find enough words to balance specific suffixes within suffix-frequency categories, satisfy the direction of gradation, and balance stem-frequency, the range of frequent words would have had to be extended so much that not all words would have qualified as frequent words.

The stems of suitable complex word-forms could fall within the frequency-limits of frequent, intermediate or infrequent word-forms. The complex words were chosen so that the number of frequent and infrequent stems could be balanced. Within the frequent category of words, there was a small number of word-forms available that would meet all the requirements (see Table 3-1 below). By necessity, eight complex word-forms in the data contained intermediate-range stems. For example, the frequent word-form *kirkon* ‘church.GEN.sg’ (rank: 872) contains an intermediate range stem *kirkko* ‘church.NOM.sg’ (rank: 2,018) and a frequent suffix. Stems whose rank fell in the intermediate zone were not included in the word-list and coding although the complex word that contained the intermediate stem could qualify on the word list.

---

<sup>8</sup> NOM.sg is the base form without case-marking and consonant gradation. Pluralization triggers gradation.

Suitable complex word-forms represented qualitative and quantitative gradation in balanced numbers. The gradation-types included gradation of all three consonants, /p, t/ and /k/. Balancing the word-forms representing each of the consonants was unnecessary.

Some words did not qualify on the list as their behavior varied with respect to gradation. Some western dialects of Finnish leave some words ungradated while in other dialects they are gradated (A. Niemikorpi, p.c., December 27, 2007). Since controlling the dialectal areas of participants and their ancestors was not feasible, only stems that behaved in a unified manner with respect to gradation were chosen. Accommodating the various criteria for word-forms explains the seemingly odd-numbered cut-off points for frequency limits. For reasons already mentioned above, the limits had to be somewhat flexible and move ahead until a sufficient number of suitable word-forms had been found.

A few selected word-forms proved to be problematic in that they had multiple meanings although they were identical in form, that is, they were homophones. One type of homophony encountered was in which one word-form had two distinct meanings (i.e. two different lexemes shared one word-shape). A case in point is the word *lippu* (rank: 8776), which has the meanings, '*flag.NOM.sg*' or '*ticket.NOM.sg*'. Computer generated word-frequency corpora do not differentiate between the multiple meanings of a single word-form; they count the frequency of occurrence of one word-shape only. However, the different lexemes had to be separated, as a participant could be familiar with one meaning but not with the other. Thus, an assumption was made that the two lexemes occur at the same frequency as indicated by the corpus-ranking although this may not be the case in actual language use. The two meanings were included on the word list as two separate words with the same frequency ranking.

The second type of homophony encountered was a single word-form, or lexeme, sharing multiple grammatical functions. For example, GENITIVE case in Finnish, represented with the suffix *-n*, can have four different grammatical functions. It can function as an object, an object of adposition, a possessive, or an expression of quantity. This is an example of syncretism, where the difference between grammatical functions is neutralised and consequently represented by one shared form. The different grammatical functions of the GEN suffix *-n* were not balanced. For instance, the word-shape *pojan* ‘boy.GEN.sg’, could be on the word list as an object and a possessive with the same frequency ranking. The two instances were considered different word-forms. There were four word-shapes (*viikon*, *kirkon*, *pojan*, *lopusta*) that had two grammatical functions on the word list.

The word-list was updated to reflect the changes from the inclusion of the five homophonous words. This caused a slight imbalance in suffixes and gradation-types. Thus, the inclusion of homophones will affect suffix and gradation-type results to some extent. Table 3-1 below presents the properties of the word-forms on the finalised word list. There were 33 frequent and 32 infrequent words on the list, totaling 65 words. The word-forms included 34 quantitative word-forms (20 in the frequent category, 14 in the infrequent category) and 31 qualitative word-forms (13 in the frequent category, 18 in the infrequent category). Collectively, the 65 word-forms contained 34 different stems, 17 frequent and 17 infrequent, and 65 suffixes, of which 34 were frequent and 31 infrequent. Frequent and infrequent stems and suffixes and word-forms representing the two gradation-types spread across both frequency-categories of word-forms. That is why the number of frequent word-forms and frequent suffixes (in the table below) do not match.

|              | Frequent | Infrequent | Total           |
|--------------|----------|------------|-----------------|
| Word-forms   | 33       | 32         | 65              |
| Quantitative | 20       | 14         | 34              |
| Qualitative  | 13       | 18         | 31              |
| Stems        | 17       | 17         | 34 <sup>9</sup> |
| Suffixes     | 34       | 31         | 65              |

**Table 3-1: The properties of the selected word-forms.**

The selected word-forms were then inserted into frame-sentences. Whole sentences were intended to provide a context to the word-forms of interest and thought to produce targeted forms more successfully than eliciting word-forms alone or in phrases. The sentences were simple in form. They contained only main clauses with 1 – 2 arguments, the subject only or the subject and the object and optionally one adverbial. Most of the verbs used ranked among the first 500 most common verbs in Finnish. Special attention was given to choosing verbs that were common in every day life to avoid the verbs' causing difficulties to the participants. Some examples of verbs used are *be, like, cost, make, have, know, buy, want, give, see* and *taste*. Then, the Finnish sentences were translated into English. The English sentences were formulated in such a way that the targeted word-form would have to be used when the sentences were translated back into Finnish by the participants. Special attention was given to maintaining the meaning and naturalness of the sentences. There were 80 sentences in total, each sentence testing 1 – 3 word-forms. Finally, the sentences were randomized.

<sup>9</sup> One of the five homophones contains an intermediate range stem, which was not included.

The sentence list was for the use of the researcher only. It was used to elicit spoken language data from the participants. Table 3-2 presents an example of the sentences created.

| #  | Sentence  | Finnish sentence                                    | Words tested               |
|----|---|---|----------------------------|
| 12 | I got <b>the tickets</b> (for the concert).     | (Minä)Sain/ostin <b>liput</b> konserttiin.          | liput (tickets) (obj)      |
| 14 | The store is closed <i>for a week</i> .         | <b>Kauppa</b> on kiinni <b>viikon</b> .             | kauppa,viikon (quantifier) |
| 15 | <b>Stores</b> are closed on Sundays in Finland. | <b>Kaupat</b> on/ovat kiinni sunnuntaisin Suomessa. | kaupat                     |
| 16 | A boy has <i>a hole</i> in his shoe.            | Pojalla on <i>reikä</i> kengässä.                   | pojalla, reikä, kengässä   |

**Table 3-2: An excerpt of the sentence list**

### 3.2.2 Questionnaire

The background questionnaire was used to gather socio-biographical and language related information from the participants. Hulsen's (2000) socio-linguistic questionnaire — used to study four generations of Dutch immigrants in New Zealand — was used as a model. The questionnaire was intended to provide an impression of the participants and the questions addressed information such as sex, age, generation, occupation, number of visits to Finland, language use and language background. It was not subjected to statistical analysis. Language-related questions pertained to participants' perception of their own language proficiency, patterns of language use, inter-generational language transmission, and contact with Finnish in daily life. The questionnaire was

written in English to even out the possible differences between the participants in Finnish reading and writing skills. There were 32 mostly close-ended questions on the questionnaire. The participants were instructed to tick boxes in response to each question, for example,

19. Do you have children?

Yes  No  If you have no children, go to question 26.

20. If yes, what language do you use mostly when talking to your children?

Finnish  English  Mixed  Other  What language? .....

1. Do you consider it important to maintain the Finnish language for yourself?

very important  important   
 rather important  not very important   
 hardly important  unimportant

### 3.3 Procedures

In preparation for data collection in the community, the application for ethical approval was submitted in January 2008. Once the ethical approval was received in early March, piloting of the instruments and testing procedures took place in Victoria in mid-March, 2008. The pilot sessions were useful in finalizing the instruments, becoming comfortable with using the recording equipment and learning to elicit the sentences. The volunteers gave feedback on the sentences and questionnaire. Revisions to the instruments were made after each test session.

The data collection sessions in Sointula were conducted in the same fashion as the pilot sessions in Victoria. Each participant was interviewed at his/her own home. Data collection sessions started by going over the consent form with each participant. Upon

recruitment, the participants were told about the experiment, but now the researcher gave a more detailed description of what the session entailed and answered any questions that were asked. The researcher used Finnish in conversations with the participants before the experiment to help them acclimatize to the language. However, the elicitation of sentences was done in English to avoid unintended prompting by the researcher.

The translation task was always conducted before administering the questionnaire to avoid conditioning the participants in any way before the experiment. The participants were asked to translate English sentences into Finnish as accurately as possible. Translation of the sentences began with two or three practice sentences before starting on the actual test sentences. The sessions were recorded with an M-Audio Microtrack digital recorder and Sony Electric Condenser Microphone ECM-MS907. The recorder rested on the table in front of the participant. Because each sentence was unrelated and the themes in the sentences varied, the researcher briefly set up the context for each sentence by saying in English, for example, “Now imagine that you are looking out the window into the ocean. How would you say in Finnish *I can see the ocean from my place?* Doing so helped the participant stay focused and understand the intended context for each sentence. In case the participant could not remember a word, the researcher helped by eliciting further or by saying it, unless it was a test word. The participants regarded the data-collection sessions as an opportunity to learn. The atmosphere was casual and the discussion often trailed off the topic. Test items for which the participant did not give answers the first time around were elicited again at the end of the experiment. Each sentence elicitation session took about 1.5 – 2 hours to complete. Six elicitation sessions (one per participant) amounted to about 10 hours of recordings for data-analysis.

### 3.4 Data-processing

The list of targeted word-forms was entered in an Excel data-base. The targeted words were coded for their word-frequency, suffix-frequency, stem-frequency<sup>10</sup>, and gradation-type (independent variables; see Section 3.5.1 for details). The targeted word-forms represented the usage in Standard Finnish to which the participant responses were compared. After the participants' sentences were transcribed, the words of interest were entered into the database as heard on the recording. Any variation from the targeted form rendered a participant's answer non-targeted. Note that a *standard* form was not necessarily a *targeted* form. A targeted response meant the very word-form that was sought in elicitation. Therefore, targeted forms were also standard forms. A non-targeted response referred to any deviation from the targeted form, even if the response was standard. Non-targeted responses took many shapes depending on whether they were whole word-forms, suffixes, stems or gradation.

The dependent variable Whole word-form accuracy was coded whether or not the whole word-form was targeted; the label "1" was used for non-targeted if the response did not match the targeted form with respect to the stem, suffix and/or gradation. A non-targeted answer was taken to mean that a participant did not know the appropriate word for the sentence context but offered another one, which may or may not have contained gradation. Non-targeted responses were further coded in terms of stem-, gradation-, and/or suffix accuracy depending on where the error occurred. Errors coded in various ways were the dependent variables. The following guidelines were used in the coding of errors (see Table 3-3 below for the summary of coding parameters).

---

<sup>10</sup> Stems that fell in the intermediate range were marked as "no entry".

The dependent variable Stem-accuracy refers to NOM.sg case base word-forms, which are non-suffixed and non-gradated, and included in tests measuring *Whole word accuracy* (see Section 4.1.1 in Chapter Four). Stem-accuracy was coded whether or not the participant produced the stem that was targeted. In many cases, non-targeted stems were coined “Finglish” words. They are words heavily influenced by English and used by Finnish immigrants living in English-speaking countries. An example in Sointula Finnish is *molikka*, meaning “soup”. This coined word originates from “Mulligan stew”. Another example is *tiketti* for “ticket”. Coinages were coded in whole word tests as non-targeted responses but as ‘no entry’ with respect to other variables. For instance, a coined complex word *molikassa* could not be evaluated for gradation or the stem based on frequency because no frequency-ranking is associated with it. Even though the suffix in the coinage could be targeted or non-targeted, it was entered as ‘no entry’ in order to be able to compare gradation-accuracy and suffix-accuracy at the time of data-analysis.

The dependent variable suffix-accuracy was coded whether or not the suffix was targeted. Non-targeted suffixes were labelled “1”. They were broken down further into different types of suffix-errors: 1) a non-targeted gradation-triggering suffix was used, 2) a non-targeted, non-gradation-triggering suffix was used (PART), and 3) no suffix at all was used but a NOM.sg stem (unsuffixed) was used instead. Each type of suffix-error was given a label (“1”, “2”, and “3” respectively) so that the specific types of errors could be isolated within non-targeted suffixes.

Finally, the dependent variable gradation-accuracy was used to code whether or not the stems used were correctly gradated. The only kind of gradation error observed was gradation loss, that is, lack of gradation when it was called for. Other types of

gradation-errors (i.e. gradation took deviant permutations or it was applied when it was not needed) were not found. Gradation was always either perfectly applied or not at all. The following errors lead to errors in gradation, a) failing to gradate the stem although a gradation-triggering suffix was present, b) the use of the non-gradation-triggering PART suffix (produced a standard but non-targeted word-form) and c) the use of the NOM.sg shape base-form (a standard but non-targeted word-form). The coding parameters are summarised in Table 3-3 below.

|           | ✓<br><b>Targeted “0”</b> | ×<br><b>Non-targeted “1”</b>            |   |   |
|-----------|--------------------------|---|---|---|
| Word-form | Matches target form      | Does not match target form              |   |   |
| Stem      | Targeted stem used       | Other than targeted stem used           |   |   |
| Gradation | Applied when called for  | Not applied when called for             |   |   |
| Suffix    | Targeted suffix used     | Non-targeted suffix used                |   |   |
|           |                          | Non-targeted grad-triggering suffix “1” | Non-targeted, non-grad-triggering suffix (PART) “2” | No suffix used (Non-targeted NOM.sg base word-form) “3” |

**Table 3-3: Coding parameters summarized**

There was a small number of unexpected responses to which the above coding guidelines did not apply. In two cases, a participant produced the infinitive (verb) *käyttää* “to use” as a response to the targeted *käytössä* “in use”. The stem in these words was accepted as targeted since the meaning of the stem is accurate and subject to gradation. The response was coded as non-targeted in terms of whole word-form, gradation and the suffix.

Another unforeseen case was one participant's response *paikasta* (rank: 3158) that suited equally well as the targeted form *paikalta* (rank: 4891) in the given sentence context. Both can be translated as "from a place". Since both the targeted and the alternate response fall within the infrequent word- and suffix-categories, the unexpected response was accepted as targeted.

Quite often a participant gave multiple answers for one word-form, one of which had to be chosen as the response. If a participant, being uncertain as to the shape of the appropriate answer, gave both standard and variant forms one after another, the response was judged as targeted or non-targeted depending on what form the participant settled on. If none of the forms stood out above others as the chosen response, that response was taken as non-standard regardless whether the participant mentioned the targeted form or not.

The collected data also included variation that was not of interest in this project. For example, variation in vowel harmony was ignored. Thus, a participant's response *euroopiissa*, as opposed to the standard *euroopassa* "in Europe", is considered as standard despite the deviant vowel, since the response contains standard gradation, the targeted stem and the targeted suffix. Similarly, anglicized stems, such as *mjuusikki* for *musiikki* "music" or *juuroppa* for *eurooppa* "Europe" were considered as standard for the meaning of the stem is a match with the targeted one.

The above coding decisions followed from the variables that were postulated for the study. The following two sections describe in detail the variables used.

### **3.4.1 Independent variables**

The first independent variable was a nominal variable *word-frequency*, or *the frequency category of a whole word representing the rate at which the word is encountered in Standard Finnish*. Its two levels, frequent and infrequent, were based on corpus frequency rankings for word-forms obtained from the Finnish parole corpus used for this study. Words that ranked between 1 – 1,558 were considered frequent and those that ranked above 3,099 were considered infrequent. An intermediate range of 1,541 words was left in between. No words were selected from the intermediate range. The two frequency ranges were assigned a number: “1” for frequent words and “2” for infrequent words.

The second independent variable was *suffix-frequency*. It was defined as *the frequency category of a suffix in a complex word representing the rate at which the suffix is encountered in Standard Finnish*. It too was a nominal variable and had two levels, frequent and infrequent, labeled “1” and “2” respectively. The labels corresponded to the rates at which suffixes are encountered in the parole corpus of standard Finnish (cited in Hakulinen et al. 2004). GENITIVE and NOMINATIVE plural, the two most frequent gradation-triggering suffixes in Finnish parole, represented frequent suffixes in the current study. They represent 37.4 and 21.7% of all suffixes in the parole corpus respectively. The proportion of the selected infrequent suffixes of all the suffixes in the same parole corpus ranged between 5.8% and 1.0%. The infrequent suffixes used were INESSIVE (5.8%), ELATIVE (4.0%), ADESSIVE (4.0%), ABLATIVE (1.0%), ALLATIVE (2.2%) and TRANSITIVE (1.6%) represented infrequent suffixes.

The third independent variable was *stem-frequency*. It was defined as *the frequency category of a NOM.sg stem of a complex word representing the rate at which the NOM.sg shape word is encountered in Standard Finnish*. It had two levels, frequent and infrequent, which were labelled as “1” and “2” respectively. The frequency limits for stems were the same as for whole word-forms. If the targeted whole word was a NOM.sg word, stem-frequency and whole word-frequency were the same.

The fourth independent variable was *gradation type*, or *the type of alternation of consonants inside stems subject to gradation that is triggered by C- or CCV-shape suffixation*. It had two levels, direct quantitative and direct qualitative. The levels were given labels “1” and “2” respectively. Direct quantitative gradation referred to consonant alternation which shortens the stem consonants while keeping the quality of the consonant same. Direct qualitative gradation weakens the obstruction of the consonants in the stem by changing the quality of the stem consonant (Hakulinen et al. 2004).

The fifth independent variable was a categorical variable *participant*. It had six levels which referred to *the individual participants in the study*. They were labeled P1, P2, P3, P4, P5, and P6.

The sixth and last independent variable was *generation*. It referred to *the immigrant generation of a participant in the study*. This categorical variable had three levels, G2, G2.5, and G3, which referred to generations 2, 2.5 and 3 respectively.

### **3.4.2 Dependent variables and measurements**

The effects of word-frequency, suffix-frequency, and/or stem-frequency were tested on some or all of the three dependent variables. The first dependent variable was

*Whole-word-accuracy, or a participant's whole word answer in the sentence context.*

Whole-word-accuracy was regarded as a single word unit, either a complex word-form or a NOM.sg stem. Errors in suffixes or stems rendered whole complex word-forms non-targeted because stems and suffixes are context-sensitive. A participant who did not know the targeted word-form could offer another word, which may or may not contain correct gradation. Therefore, if the suffix or the stem failed to match the targeted word-form, gradation could not be accepted in those words because the word received was not the one targeted. With regards to NOM.sg stems, whole-word accuracy determined whether the participant knew the NOM.sg stem-allomorphs contained in the complex word-forms. Furthermore, whole-word-accuracy determined whether gradated stem-allomorphs were incorrectly used as independent word-forms as answers to NOM.sg stems. Whole-word-accuracy was only used as the dependent variable with the independent variable Word-frequency. Whole-word-accuracy was not used with other independent variables because it contained variation in stems, gradation and/or the suffix. The dependent variable Gradation-accuracy (discussed below) was more informative, being isolated to the stem and so most tests were conducted on gradation-accuracy.

The second dependent variable *Gradation-accuracy* referred to *a participant's gradation-answer independent of the sentence context*. Failure to gradate when necessary rendered gradation non-targeted. Evaluating gradation independently of the sentence context allowed examination of gradation-accuracy as correlated with suffix-accuracy. The variable Gradation-accuracy was the primary means of assessing gradation-success in stems independent of context-sensitive suffixes; the use of a wrong

suffix created a word that was inappropriate for the sentence context. The focus of the thesis was the outcome of this variable.

The third and final dependent variable was *Suffix-accuracy* which referred to *the suffix-answer given by a participant*. Any deviation in suffixation from the targeted form rendered suffix-response non-targeted. Suffix-errors included the use of a non-targeted gradation-triggering suffix, a non-targeted non-gradation-triggering suffix (PART), or the absence of a suffix (use of ungradated NOM.sg shape stem).

The experimental design included six independent and three dependent variables. To test the proposed hypotheses, a series of Pearson's chi-square tests ( $\chi^2$ ) was performed to compare the number of cases in each category to the expected number of cases and to determine whether there is a significant difference between the frequencies of occurrence. The significance level is set at 0.05, so that any p-value less than 0.05 is considered statistically significant.

The variables used in the study are summarized in Table 3-4 below.

### **Independent variables**

Word-frequency - *The frequency category of a whole word representing the rate at which the word is encountered in Standard Finnish.*

Suffix-frequency - *The frequency category of a suffix in a complex word representing the rate at which the suffix is encountered in Standard Finnish.*

Stem-frequency - *The frequency category of a NOM.sg stem of a complex word representing the rate at which the NOM.sg shape word is encountered in Standard Finnish.*

Gradation-type - *The type of alternation of consonants inside stems subject to gradation that is triggered by C- or CCV-shape suffixation.*

Participant - *The individual participants in the study.*

Generation - *The immigrant generation of a participant in the study.*

### **Dependent variables**

Whole-word-accuracy - *A participant's whole word answer **in the sentence context.***

Gradation-accuracy - *A participant's gradation-answer **independent of the sentence context.***

Suffix-accuracy - *The suffix-answer given by a participant.*

## **Chapter Four Results**

The first half of the thesis, namely Chapters One, Two and Three, have provided the theoretical background for this project and described the methodology that evolved out of the theory. Chapter Four, Five and Six focus on the outcome of that research, reporting on the results and discussing the findings. This chapter presents the results of the statistical analyses performed on the data in three sections. The first section, 4.1, focuses on the results relating to frequency of use, the main theme in this research, and seeks to answer whether or not frequency conditions the application of consonant gradation in non-dominant Finnish speech. This general question breaks down into three specific frequencies, Word-frequency, Suffix-frequency and Stem-frequency, which are tested on gradation in sections 4.1.1, 4.1.2 and 4.1.3 respectively. In addition to searching for the effects of frequency in non-dominant Finnish speech, this project sets out to answer if one Gradation type, quantitative or qualitative, is more successful than the other. The results of gradation type will be presented in section 4.2. Last section, section 4.3, examines individual speakers' performance in achieving standard gradation and whether immigrant Generation might explain the variance between individuals.

## 4.1 Frequency of use

### 4.1.1 *Word-frequency*

Two of the six Research Questions formulated for this research involves Word-frequency as the independent variable. Research Question One, which asks *does whole word-frequency affect achieving whole word-accuracy?*, is addressed in Section 4.1.1. Section 4.1.2 answers Research Question Two, *does whole word accuracy affect achieving gradation-accuracy?* Word-frequency is also tested on Suffix-accuracy although no specific research question was formulated for it. The test is included so that all three components of complex targeted word-forms, that is, the answer as a whole, gradation, and the suffix, are tested. The following paragraphs provide the results of the statistical analyses.

**Word-frequency and Whole-word-accuracy.** In order to answer Research Question One, data-analysis uses whole word-accuracy as the dependent variable. Participants' responses that conform to the targeted whole-words were coded as targeted (i.e. correct). Any deviation from the targeted whole word-forms rendered participants' responses non-targeted (i.e. incorrect) (see coding parameters in Table 3-3). The answer could deviate with respect to the stem, gradation and/or the suffix. Targeted responses are of two types 1) complex word-forms as single units such as *rannalla* 'on the beach', in which the gradated stem *ranna-* and the suffix *-lla* are fused together and inseparable, and 2) simple NOM.sg word-forms, which contain neither gradation nor suffixes. The data are analysed in two ways. In the first test, both complex and simple word-forms are included. It was hypothesized that participants would be more successful in producing

targeted word-forms if they were frequent ones, having supposedly benefited from hearing and using them more often. Infrequent word-forms, in turn, were hypothesized to show more variation due to more infrequent usage.

Table 4-1 presents the crosstabulation for the test, listing the number of instances of targeted forms and non-targeted forms in the two frequency categories of words and the percentages they represent (%). A total of 535 word-form tokens are considered in the test.

| Word frequency | Targeted     | Non-targeted | Total |
|----------------|--------------|--------------|-------|
| Frequent       | 162<br>62.5% | 97<br>37.5%  | 259   |
| Infrequent     | 195<br>70.7% | 81<br>29.3%  | 276   |

**Table 4-1: Word-frequency \* Whole-word-accuracy crosstabulation (with NOM.sg word-forms)**

A chi-square test is performed to determine if a relationship exists between the two categorical variables, the independent variable Word-frequency and the dependent variable Whole word accuracy. Significance would mean the frequency of a word-form affects its production by the participants.

It can be seen from the above table that of the total of 259 frequent word-forms, 62.5% is targeted. Of the 276 infrequent words, in turn, 70.7% is targeted. Infrequent word-forms contain significantly more accurate whole word responses than frequent word-forms ( $\chi^2 (1) = 3.95, p < 0.05$ ). In conclusion, there is a significant association between the frequency of word-forms and the correctness of the participants' responses but in the opposite direction than anticipated.

The same test is run the second time without the NOM.sg bases to measure the effect of word-frequency on complex word-forms only. The rationale for this is that the NOM.sg base word-forms are less complex than suffixed word-forms since they contain neither gradation nor suffixes. Suffixed word-forms present more chances for errors than simple word-forms.

Table 4-2 lists the number of instances of targeted and non-targeted forms in the two frequency categories of word-forms, and the percentages they represent (%). Note that with the exclusion of stems, the total number of tokens considered in this test has dropped to 363 from 535<sup>11</sup>.

| Word frequency | Targeted     | Non-targeted | Total |
|----------------|--------------|--------------|-------|
| Frequent       | 101<br>55.8% | 80<br>44.2%  | 181   |
| Infrequent     | 110<br>60.4% | 72<br>39.6%  | 182   |

**Table 4-2: Word-frequency \* Whole-word-accuracy crosstabulation (without NOM.sg word-forms)**

Without NOM.sg base word-forms, both frequency categories are relatively balanced in producing targeted complex word-form responses ( $\chi^2(1) = 0.802$ ,  $p = 0.370$ ). Thus, the significant result in the first test can be attributed to the success in the production of infrequent NOM.sg base word-forms. The insignificant result from the latter test (NOM.sg bases excluded) is considered as the more important of the two tests because the production of stems is not as challenging as the production of complex word-

<sup>11</sup> The difference is 172 tokens. There were 14 coined simple word-form answers that are included in the first test but not in the second one ( $172 + 14 = 186$  NOM.sg word-forms).

forms. The test provides ‘no’ answer for Research Question One; whole word-frequency does not affect success at producing targeted whole word-forms.

**Word-frequency and Gradation-accuracy.** The previous test compared the participants’ answers to the targeted whole word-forms. It now becomes important to separate the components of a targeted complex word-form and consider if the frequency of a word-form has an effect on its parts, that is, gradation and the suffix. First, a test is run to determine if word-frequency produces a significant effect on gradation-accuracy alone. It was hypothesized that consonant gradation would be more successful in frequent than in infrequent word-forms as it is assumed that the participants have benefited from more frequent usage with frequent word-forms than with infrequent word-forms.

Table 4-3 presents lists the number of instances of targeted forms and non-targeted forms in the two frequency-categories and the percentages they represent (%). The non-targeted gradation column indicates the number of cases of gradation loss due to the use of the NOM<sup>12</sup>. The total number of tokens considered in this test is 338<sup>13</sup>.

---

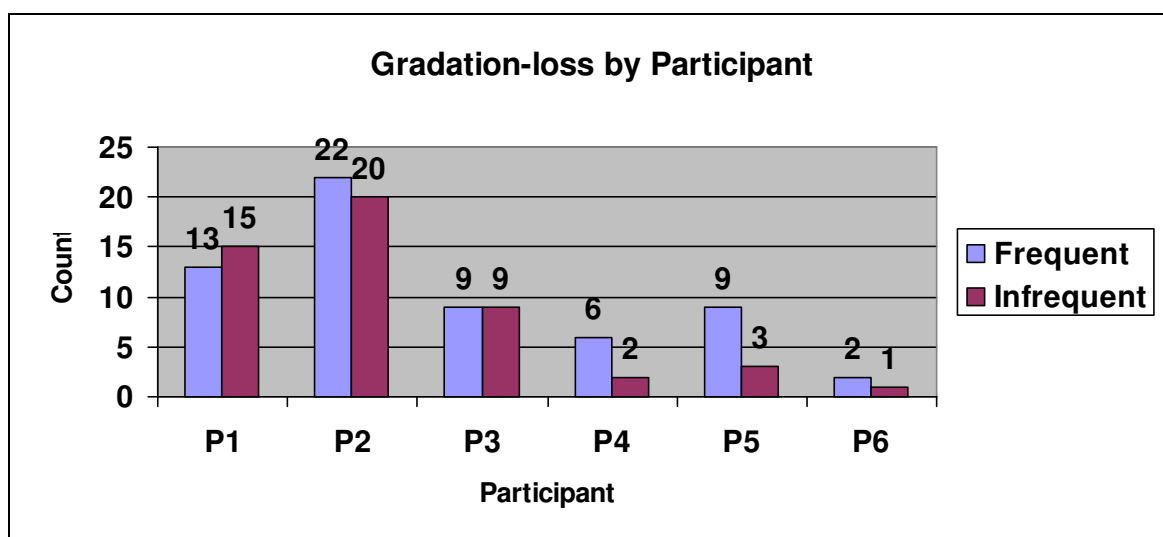
<sup>12</sup> Gradation loss refers to gradation-errors due to the use of the plain NOM.sg word-form or NOM.sg shape stem with a gradation-triggering suffix as an answer (i.e. failure to gradate the stem despite the presence of a gradation-triggering suffix).

<sup>13</sup> Coined (no frequency-ranking) responses are excluded. Eight PART responses (freq: 7 instances; infreq: 1 response) are also excluded from this test (contain non-gradation triggering suffix) measuring the total use of the NOM.sg stems. Plain NOM.sg stem word-forms are automatically excluded because they are not gradated.

| Word frequency | Targeted     | Non-targeted | Total |
|----------------|--------------|--------------|-------|
| Frequent       | 113<br>64.9% | 61<br>35.1%  | 174   |
| Infrequent     | 114<br>69.5% | 50<br>30.5%  | 164   |

**Table 4-3: Word-frequency \* Gradation-accuracy crosstabulation**

The crosstabulation table above shows that frequent and infrequent categories have a relatively balanced number of instances of targeted and non-targeted gradation ( $\chi^2(1) = 0.799, p = 0.371$ ). Of the total of 174 instances of gradation-responses in the frequent category, 35.1% is non-targeted and, conversely, 30.5% of the total of 164 gradation-responses in the infrequent category are non-targeted. Gradation-loss did not affect words only in the infrequent category, contrary to expectations. Figure 4-1 below illustrates gradation loss by participant. Gradation-loss is predicted by generation rather than frequency.



**Figure 4-1: The count of instances of gradation loss by Participant in the frequent and infrequent categories of word-forms**

The participants P1 and P2 represent G3, and participants P3 and P4 represent G2.5, and P5 and P6 are from G2. There is an overall tendency for gradation loss to increase from G2 to G3. Since G3 shows more gradation loss than G2.5 and G2, the patterning of gradation loss in the data is not a coincidence but arguably related to the participants' immigrant generation. The patterning of gradation loss will be discussed in Section 5.1.4 of Chapter Five as one of the points of interest in this research.

**Word-frequency and Suffix-accuracy.** The above sections have measured the effect of word-frequency on accuracy of whole word-form and on gradation in the stem of a complex word. This last supplementary word-frequency test aims to answer whether word-frequency affects the accuracy of the suffix. The hypothesis tested is that frequent words should manifest more targeted suffix responses than infrequent words.

Table 4-4 lists the number of instances of targeted suffixes and non-targeted suffixes in the two frequency categories of words and the percentages they represent (%). A total of 346 suffixes are considered in this test<sup>14</sup>.

| Word frequency | Targeted     | Non-targeted | Total |
|----------------|--------------|--------------|-------|
| Frequent       | 119<br>66.9% | 59<br>33.1%  | 178   |
| Infrequent     | 129<br>76.8% | 39<br>23.2%  | 168   |

**Table 4-4: Word frequency \* Suffix-accuracy crosstabulation**

<sup>14</sup> Coined words are excluded in order to be able to compare these results to the following gradation-results. Simple NOM.sg base word-forms were automatically excluded since they are not suffixed.

Table 4-4 shows that of the total of 178 suffix-responses in the frequent word-category, 66.9% is targeted. Targeted answers in the infrequent category account for 76.8% of the total of 168 suffix-responses in the infrequent category. The results indicate that the participants found targeting the suffixes significantly easier in the infrequent word-category than in the frequent word-category ( $\chi^2 (1) = 4.199, p < 0.05$ ). It can be concluded that Word-frequency and Suffix-accuracy are correlated but in the opposite direction than expected; the suffixes in infrequent words are generally easier to target than the suffixes in frequent words. This issue is addressed in Section 5.1.2 of Chapter Five as another point of interest that this research uncovered.

**Section summary.** The first set of tests measured the effect of Word-frequency on Whole-word-accuracy, Gradation-accuracy, and Suffix-accuracy. The results can be summarized with four points:

1. Word-frequency does NOT produce a significant effect on Whole-word-accuracy after filtering out NOM.sg base word-forms; errors in both frequency-groups are equally common.
2. Word-frequency does NOT produce a significant effect on Gradation-accuracy; gradation loss occurs throughout the data, regardless of the frequency-rankings of word-forms.
3. Gradation loss tracks with the participants' immigrant generation.
4. Word-frequency produces a significant effect on Suffix-accuracy but in the unexpected direction; the participants are less successful in targeting the suffix of frequent words than of infrequent words.

Word-frequency was anticipated to correlate with whole word accuracy but the initial hypothesis is not confirmed by the results. As for the effect of Word-frequency on Gradation-accuracy, the finding of no relationship is equally surprising. However, the distribution of gradation loss across both frequency categories is related to immigrant generation. Contrary to the expectations, the participants' performance in targeting the suffix is significantly better in infrequent words than in frequent words.

#### **4.1.2 Suffix-frequency**

Having completed the tests that measure the effect of Word-frequency on Whole-word-, Gradation-, and Suffix-accuracy, the focus can now move on to Suffix-frequency. The tests in this section, those pertaining to suffix-frequency, become particularly interesting since the presence of a gradation-triggering suffix is needed for gradation to take place in the stem.

Suffix-frequency is tested on two dependent variables, Suffix-accuracy and Gradation-accuracy. Research Question Three asks if suffix-frequency has an effect on achieving gradation-accuracy in the stem. Although not initially included as a research question, the effect of Suffix-frequency on Suffix-accuracy is included after the discovery that a relationship existed between Word-frequency and Suffix-accuracy. Arguably it is the frequency of the suffix in a complex word-form that largely determines whether the whole word-form is frequent or infrequent. This set of tests begins with determining if Suffix-frequency conditions the accurate production of targeted suffixes. Then, a test measuring the effect of Suffix-frequency on Gradation-accuracy is run.

**Suffix-frequency and Suffix-accuracy.** Recall from 4.1.1 that Suffix-accuracy was significantly higher for infrequent words than for frequent words. On the assumption that Word-frequency is directly related to Suffix-frequency, it can be hypothesized that Suffix-frequency would also produce a significant effect on Suffix-accuracy. Moreover, it can be expected that infrequent suffixes would be easier to target than frequent suffixes, in accordance with the previous results regarding the direction of the effect. If the hypothesis on the direction of the effect is also supported, properties of suffixes must be examined more closely to find an explanation for the unexpected results.

Table 4-5 below lists the number of instances of targeted suffixes and non-targeted suffixes in each category of suffix-frequency, and the percentages they represent (%). A total of 346 suffixes are considered in this test<sup>15</sup>.

| Suffix-frequency | Targeted     | Non-targeted | Total |
|------------------|--------------|--------------|-------|
| Frequent         | 126<br>67.0% | 62<br>33.0%  | 188   |
| Infrequent       | 122<br>77.2% | 36<br>22.8%  | 158   |

**Table 4-5: Suffix-frequency \* Suffix-accuracy crosstabulation**

Table 4-5 indicates that of the total of 188 suffix-responses in the frequent category, 33.0% is non-targeted. There are 158 suffix-responses in the infrequent category of which 22.8% is non-targeted. These results are significant ( $\chi^2(1) = 4.394, p < 0.05$ ) and they confirm that the participants made significantly fewer suffix-errors with infrequent suffixes than with frequent suffixes. The effects of Suffix-frequency and

<sup>15</sup> NOM.sg bases are automatically excluded since they are not suffixed. Coinages are removed. Included in this test are targeted suffix-responses and non-targeted suffix-responses (gradation-triggering, PART, and NOM.sg shape responses).

Word-frequency are in the same direction as was predicted. In order to find a satisfying explanation to the counter-intuitive suffix-results, suffixes are put under scrutiny in Section 5.1.2 of Chapter Five and discussed as another point of interest in this research.

**Suffix-frequency and Gradation-accuracy.** The answer to Research Question Three, *does suffix-frequency affects achieving gradation-accuracy*, was anticipated to be ‘no’. Suffix-frequency was not expected to have bearing on gradation since the nature of gradation (qualitative or quantitative) of the stem consonants /p, t,/ and /k/ is not predictable from suffixes. Gradation type is largely a property of the stem. With a gradation-triggering suffix<sup>16</sup>, a base word-form that is subject to gradation produces a unique gradated stem, dictated by the consonant quality in the stem.

However, both Word-frequency and Suffix-frequency were found to have a significant effect on the production of the targeted suffix. Suffix-frequency may have a significant effect on Gradation-accuracy, for it is the presence of a suffix that is needed for gradation to take place in the stem. Given the direction of the effects in earlier findings, one might expect better performance with infrequent suffixes than with frequent ones.

Table 4-6 lists the number of instances of targeted and non-targeted gradation in two frequency-categories of suffixes and the percentages they represent (%). A total of 346 word-forms were considered in this test<sup>17</sup>.

---

<sup>16</sup> Recall that to gradate or not to gradate is, as a general rule, phonologically conditioned. C- or CCV-shape suffixes trigger gradation while vowel-initial suffixes that produce an open syllable do not. Some recent 2-syllable coinages or loans that have a single stop consonant between vowels are left outside gradation, e.g., *huki* “turn.NOM.sg” (Hakulinen et al. 2004, p. 74) (*huki-C*, *huki-V*, *\*hui-C*, *\*hui-V*) (see section 2.1.1 for more information).

<sup>17</sup> NOM.sg word-forms are automatically excluded since they do not contain gradation.

| Suffix-frequency | Targeted     | Non-targeted | Total |
|------------------|--------------|--------------|-------|
| Frequent         | 110<br>58.5% | 78<br>41.5%  | 188   |
| Infrequent       | 117<br>74.1% | 41<br>25.9%  | 158   |

**Table 4-6: Suffix-frequency \* Gradation-accuracy crosstabulation**

Table 4-6 shows that gradation is accurate in 58.5% of the total of 188 gradation-responses in the frequent category. In the infrequent category, there are 158 gradation-responses, of which 74.1% is accurate. Again, the infrequent category produces significantly fewer errors than the frequent category ( $\chi^2 (1) = 9.188, p < 0.005$ ). With regards to the counter-intuitive suffix-frequency tests, the suspicion is that the property of suffixes that correlates with suffix-accuracy and gradation-accuracy may not be suffix-frequency but some other property that is associated to a greater extent with infrequent suffixes with than frequent suffixes. Chapter Five will return to this issue.

**Section summary.** The second set of tests measured the effect of Suffix-frequency on Suffix-accuracy and on Gradation-accuracy. The results can be summarized as follows:

1. Suffix-frequency has a significant effect on Suffix-accuracy.
2. Suffix-frequency has a significant effect on Gradation-accuracy.
3. Contrary to expectation but consistent with findings on the effect of word-frequency, the infrequent category of suffixes produces significantly fewer errors in both suffixes and gradation.

Based on the results, it can be concluded that gradation is associated with suffix-frequency in Sointula Finnish. Whether it is the frequency or another property of the suffixes that is associated with suffix-accuracy and gradation is not yet known. However, the results indicate that infrequent suffixes are more successful than frequent suffixes in both targeting the suffix and gradation. The examination of suffixes continues in Chapter Five (Discussion) in order to find an explanation for the surprising suffix-results.

The above sections have completed the tests for Word-frequency and Suffix-frequency. The effect of Stem-frequency on gradation remains to be tested in examining the effects of frequency of use in Sointula Finnish.

#### **4.1.3 Stem-frequency**

To remind the reader, stem-frequency refers to the frequency of the stem inside a complex word-form. Because stems alone cannot stand as words, stem-frequency refers to the corresponding NOM.sg word-forms which have their own frequency ranking. Thus, stem-frequency is the same as the word-frequency of simple base word-forms. For example, the complex frequent word-form *markan* ‘mark.GEN.sg’ (rank: 172) contains an infrequent stem *markka* ‘mark.NOM.sg’ (rank: 3099)<sup>18</sup>. The effect of Stem-frequency is tested on Gradation-accuracy, answering Research Question Four, *will stem-frequency have an effect on achieving standard gradation?* It was hypothesized that the answer would be ‘no’ since NOM base word-forms and their gradated allomorphs are phonetically dissimilar and are used in different sentence contexts. Thus, the frequency of

---

<sup>18</sup> Stems (i.e. NOM.sg bases) were subject to the same frequency requirements as complex word-forms (Frequent words rank between 1 – 1,558 while infrequent words rank above 3,099). Stems that fall within the intermediate range were excluded.

a NOM.sg word-form is assumed to not reinforce the representation of its gradated allomorph to a significant degree.

Table 4-7 lists the number of instances of targeted gradation and non-targeted gradation in two categories of stem-frequency, and the percentages they represent (%). A total of 304 word-form tokens are considered in this test.

| Stem-frequency | Targeted    | Non-targeted | Total |
|----------------|-------------|--------------|-------|
| Frequent       | 99<br>67.8% | 47<br>32.2%  | 146   |
| Infrequent     | 99<br>62.7% | 59<br>37.3%  | 158   |

**Table 4-7: Stem-frequency \* Gradation-accuracy crosstabulation**

The above results are insignificant ( $\chi^2 (1) = 0.886$ ,  $p = 0.347$ ). This means that stem-frequency does not have a significant effect on gradation-accuracy.

#### **4.1.4 Summary of frequency-tests**

A series of Pearson Chi-square tests were conducted to measure participants' performance in the production of targeted word-forms, gradation and suffixes in frequent and infrequent categories of words, suffixes and stems. The results yield five findings:

1. Word-frequency does NOT have a significant effect on Whole-word-accuracy.
2. Word-frequency does NOT have a significant effect on gradation-accuracy.
3. Gradation loss spreads across both frequent and infrequent word categories.
4. Suffix-frequency has a significant effect on both suffix-accuracy and gradation-accuracy but not in the expected direction; gradation and suffix-

accuracy are more successful in words containing infrequent suffixes than in words containing frequent suffixes.

5. Stem-frequency does NOT have a significant effect on consonant gradation.

Except for the hypothesis regarding stem-frequency, the initial hypotheses formulated for the frequency-tests are not supported by the results. It was anticipated that the frequency of a word-form would affect its phonetic shape in such a way that frequent categories would show more targeted answers than infrequent categories. Contrary to expectations, no relationship can be established between Word-frequency and the dependent variables Whole-word-accuracy and Gradation-accuracy. The finding that gradation loss affects word-forms in both frequency-categories of complex words is considered important and will be discussed in detail Chapter Five as the first topic of interest in this research.

The significant relationship between Word-frequency and Suffix-accuracy foreshadowed surprising suffix-results<sup>19</sup>; suffix-frequency is found to be related to both Suffix-accuracy and Gradation-accuracy against the initial hypothesis. The infrequent category of suffixes produces more targeted suffixes and gradation. The surprising direction of these effects motivates a closer examination of the suffixes in Chapter Five.

## 4.2 Gradation-type

---

<sup>19</sup> The significance between Word-frequency and Suffix-response is attributable to the suffix in a complex word because the frequency of a complex word is largely determined by the frequency of the suffix it contains.

Having completed the tests measuring the effect of frequency of whole word-forms, suffixes and stems on gradation accuracy, this chapter now shifts its focus onto Gradation-type. Research Question Five, *is direct quantitative gradation is more successful than direct qualitative gradation*, was formulated for this purpose. It was hypothesized in section 2.1.1 that direct quantitative gradation (affects the quantity, or length, of the consonant) is more successful than direct qualitative gradation (affects the quality of the consonant). With the latter type of gradation, a non-dominant speaker with a limited exposure to standard language might not know what consonants /p, t,/ and /k/ change into.<sup>20</sup>

Table 4-8 lists the number of instances of targeted gradation and non-targeted gradation in two categories of gradation, quantitative and qualitative, and the percentages they represent (%). A total of 346 tokens are considered in this test.

| Gradation-type | Targeted     | Non-targeted | Total |
|----------------|--------------|--------------|-------|
| Quantitative   | 128<br>70.3% | 54<br>29.7%  | 182   |
| Qualitative    | 99<br>60.4%  | 65<br>39.6%  | 164   |

**Table 4-8: Gradation-type \* Gradation-accuracy crosstabulation**

The table 4-8 indicates that of the total of 182 instances of quantitative gradation, 70.3% is targeted. Conversely, 60.4% is targeted of the total 164 instances of qualitative gradation. These results are approaching significance ( $\chi^2 (1) = 3.796$ ,  $p = 0.051$ ), favoring quantitative gradation.

<sup>20</sup> In section 3.2.1, it was noted that all of the participants in the current project were able distinguish the difference between long and short consonants although they all showed loss of gradation to a varying degree.

The previous test includes all types of participant responses, that is, word-forms with gradation-triggering suffixes, non-gradation triggering PART suffixes and NOM.sg base (non-suffixed) word-forms. The two latter types of suffix-responses cannot trigger gradation in the stem. The test was run the second time without PART suffixes and NOM.sg base word-forms to determine if a difference exists in gradation-performance between gradation-types in words with gradation-triggering suffixes only. Table 4-9 below is similar to Table 4-8 but includes gradation-triggering suffixed forms only. The table lists the number of instances of targeted gradation and non-targeted gradation in two categories of gradation, and the percentages they represent (%). With the exclusion of non-gradation triggering responses, the total of tokens considered drops from 346 to 267<sup>21</sup>.

| Gradation-type | Targeted     | Non-targeted | Total |
|----------------|--------------|--------------|-------|
| Quantitative   | 128<br>92.1% | 11<br>7.9%   | 139   |
| Qualitative    | 99<br>77.3%  | 29<br>22.7%  | 128   |

**Table 4-9: Gradation-type \* Gradation-accuracy crosstabulation in words with gradation-triggering suffixes**

These results are highly significant ( $\chi^2(1) = 11.371, p = 0.001$ ). Of the total of 139 instances of quantitative gradation, 92.1% is targeted while only 7.9% non-targeted. Conversely, of the total 128 instances of qualitative gradation, 77.3% is targeted and

<sup>21</sup> The difference is 79 tokens, which includes 71 NOM.sg base word-forms and eight PART word-forms.

22.7% non-targeted. The results prove correct the hypothesis that quantitative gradation is more successful of the two gradation-types.

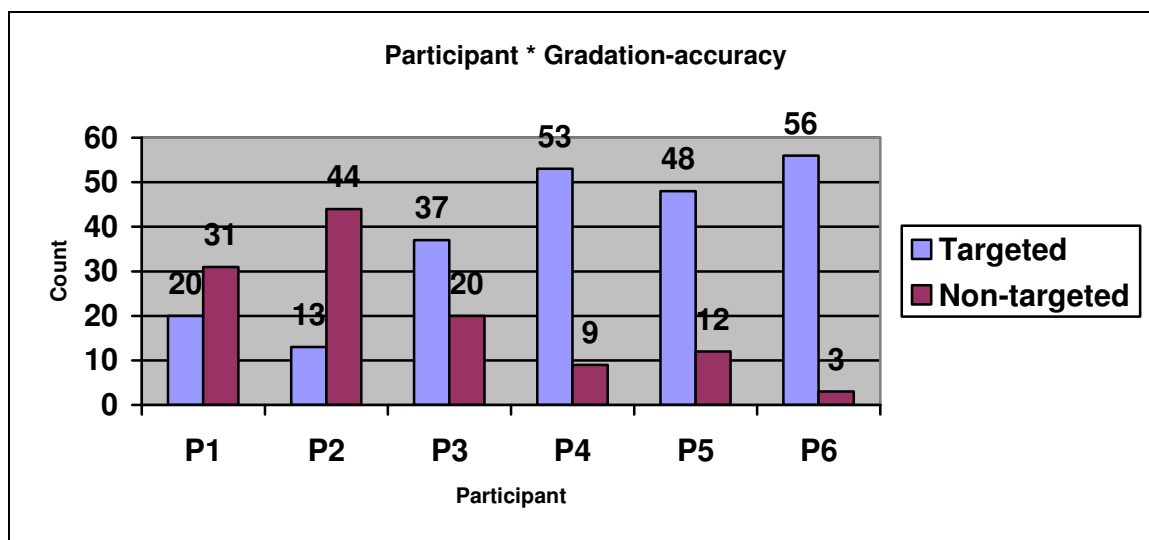
### 4.3 Individual speakers and immigrant generation

The third and last section considers the individuals as the determiners of gradation and measures the extent to which each participant is affected by loss of gradation. All participants are known to illustrate loss of gradation to a varying degree. After a comparison of the participants, the effect of immigrant generation is tested on gradation accuracy in order to answer Research Question Six, *does immigrant generation explain the variation in the production of gradation between the participants?* Thus, the independent variables Participant and Generation were tested on the dependent variable Gradation-accuracy in two separate chi-square tests.

The differences between participants are more easily seen if presented in graph-format, thus Figure 4-2 below. The numbers above the bars indicate the number of targeted and non-targeted gradation responses. A total of 346 tokens were considered in this test<sup>22</sup>.

---

<sup>22</sup> NOM.sg shape word-forms are automatically excluded being non-gradated. PART responses are included as their use leads to gradation-errors.

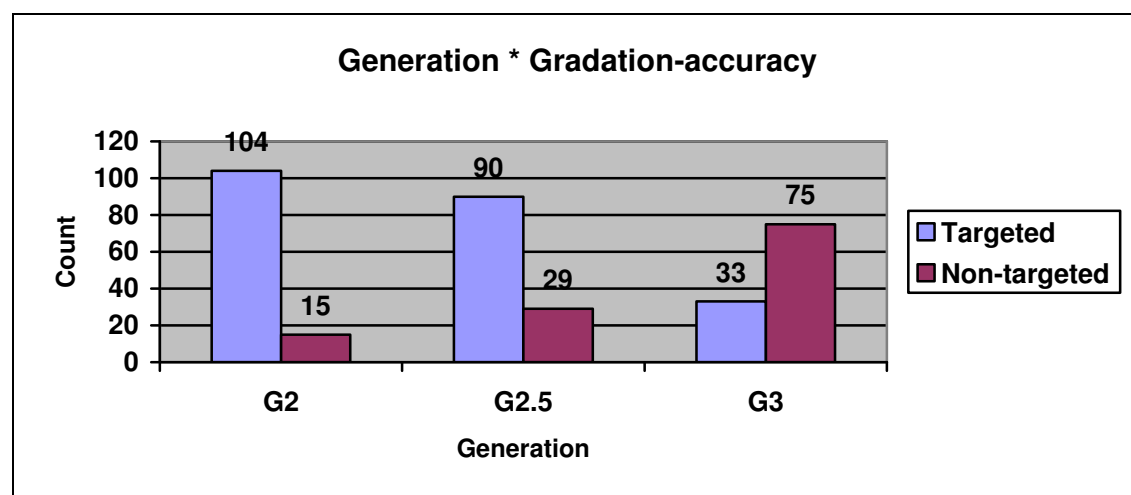


**Figure 4-2: The effect of Participant on Gradation-accuracy**

The differences between participants are highly significant ( $\chi^2(5) = 100.854$ ,  $p < 0.001$ ). A chi-square test is unable to indicate where the significances within the participants lie. However, the above figure shows that the participants vary greatly in gradation accuracy. P6 performed most accurately with the highest number of cases of standard gradation ( $n = 56$ ) and P2 least accurately with the lowest number of standard gradation ( $n = 13$ ). P1 and P2 differ from the rest of the participants in that they have more instances of non-targeted than targeted gradation. P3 through P6 show more targeted gradation than P1 and P2. As already mentioned in Section 4.1.1, P1 and P2 represent G3 while the participants P3 and P4 belong to G2.5 and P5 and P6 to G2. Thus, the participants' performance in gradation appears to reflect their immigrant generation in such a way that G3 participants, namely P1 and P2, perform least accurately in gradating consonants while P3, P4, P5, P6, who represent generations 2 and 2.5, perform better. Finer distinctions between generations are made below.

Although already shown in Section 4.1.1 and in the previous paragraph that immigrant generation appears to explain the differences in gradation-accuracy between participants, the following test measures the effect of the independent variable Generation on gradation-accuracy. It was hypothesized at the beginning that the immigrant generation has bearing on the overall gradation-success since the exposure to native-like speech reduces in each generation born abroad. This would predict that G2, being the first generation born abroad, is the least affected by gradation loss while G3 with two parents born abroad is the most affected by it. G2.5, with one Finland-born parent, would fall in between G2 and G3.

Figure 4-3 below is similar to Figure 4-2 but collapses the six individual speakers into the three generations. The numbers above the bars indicate the number of targeted and non-targeted gradation responses in each immigrant generation. A total of 346 tokens were considered in this test<sup>23</sup>.



**Figure 4-3: The effect of Generation on Gradation-accuracy**

<sup>23</sup> NOM.sg word-forms are automatically excluded being non-gradated. PART responses are included.

The above figure clearly shows that the predictions regarding the patterning of generations are confirmed; G2 manifests the largest number of targeted gradation responses and the smallest number of gradation-errors. Conversely, G3 shows least targeted gradation responses and the most gradation-errors. The differences between the three generations were statistically highly significant ( $\chi^2 (2) = 89.139, p < 0.001$ ). Again, the chi-square test does not allow comparing significances between generations but this does not seem necessary. It can be concluded from the results that immigrant generation affects gradation-accuracy. This result provides a likely explanation for the variance in gradation-accuracy between individual participants; the earlier the immigrant generation, the better their performance in consonant gradation.

#### **4.4 Chapter summary**

This project was designed to measure the effect of three types of frequency, namely word-frequency, suffix-frequency and stem-frequency, on consonant gradation in Sointula Finnish. Separate chi-square tests were used to measure the effect of each independent variable on some or all of the three dependent variables, Whole-word-accuracy, Gradation-accuracy, and Suffix-accuracy. In addition to examining frequency of use, this research sought to answer if one type of gradation, quantitative or qualitative, was more successful than the other. Lastly, this research measured to what extent individual speakers showed loss of gradation and whether the variance between the participants was attributable to their immigrant generation. Although the results are somewhat surprising, they are consistent with one another.

The tests results pertaining to frequency of use of whole word-forms and suffixes does not support the hypotheses formulated for this project. Word-frequency is not associated with whole-word-accuracy or Gradation-accuracy. However, the patterning of gradation loss is shown to be meaningful and worthy of discussion in Chapter Five.

The hypothesis that no relationship would be found between Suffix-frequency and Gradation-accuracy is not supported either. The results indicate that the frequency of the suffix has bearing not only on gradation-accuracy but also on suffix-accuracy. Since the presence of a gradation-triggering suffix is needed for gradation to take place in the stem, significance between Suffix-frequency and Gradation-accuracy, in hind-sight, is not a great surprise, although it was contrary to what was anticipated at the beginning of this research journey. Earlier tests had already given an indication as to what direction the effect was likely to run and indeed, the results show that infrequent suffixes are more successful than frequent ones. The unexpected direction of the effect needs an explanation and Section 5.1.2 of Chapter Five will return to this topic.

Gradation-type results confirm that quantitative gradation is more successful than qualitative gradation. The difference became obvious only after filtering out PART and NOM.sg shape responses which do not trigger gradation in the stem.

Finally, the results show that the differences between participants are associated with immigrant generation in such way that earlier immigrant generations perform better in gradation than later immigrant generations.

The above results are exciting in that they lead to several interesting questions:

1. Are there frequent words in non-dominant language contexts?
2. Why are the anticipated frequency effects not found in non-dominant language context?
3. How well does the corpus-model apply in non-dominant language contexts?
4. What property of suffixes makes infrequent suffixes easier to target?
5. How are complex words stored in the memory and retrieved?
6. Why is consonant gradation not preserved through generations?

These questions and other issues will be discussed at length in the next chapter, Chapter Five.

## Chapter Five Discussion

The main goal of this thesis was to examine if frequency of use affects the phonetic shape of complex words-forms, those that consist of more than one morpheme. The study tested three types of frequency. These types cover the frequency of word-forms and their component parts: 1) whole word frequency, referring to targeted complex forms or NOM.sg stems (e.g. *rannalla* ‘on the beach’ (rank: 3134); *ranta* ‘beach.NOM.sg (rank: 7297), 2) suffix-frequency (i.e. frequency of the suffix in a complex word, e.g. *-lla* (infrequent) and 3) stem-frequency (i.e. frequency of the stem in a complex word-form in its base form, e.g. *ranta* ‘beach’). This research was searching for frequency effects that could explain the variation often manifested in the production of consonant gradation in non-dominant Finnish. The expected frequency effects were not found. Instead, the frequency results suggest that all words in both frequency-categories are affected by gradation-loss, not only infrequent words. Moreover, the semantics of suffixes turns out to be a better predictor of gradation than frequency, explaining the surprising suffix-results. These two main discoveries and other findings are addressed and discussed here in Chapter Five. The chapter is organized into two major sections. Section 5.1, discusses the findings. Frequency of use is addressed in Section 5.1.1. Semantics of suffixes, in particular the atrophy of the GEN object marker and the resultant gradation loss are discussed in Section 5.1.2. Section 5.1.3 focuses on the findings relating to gradation-types; Section 5.1.4 examines individual participants and their immigrant generation. A summary of the frequency findings is provided in Section 5.1.5. The second major

section, Section 5.2, discusses the implications of the above findings for Cognitive grammar. See Appendix C for the glossary of the terminology used throughout the chapter.

## **5.1 Research predictions and reseach findings**

Six research questions (see Section 2.7 of Chapter Two for details) were formulated for this project (see Table 5-1 below). The following paragraphs review the predictions for the research questions.

This thesis applies the framework of Usage-based grammar to consonant gradation, as outlined in Chapter Two (in Sections 2.3 and 2.4). Bybee (2001, 2007) claims that high-frequency memory tokens are independent of other memory tokens and accessed as units. Accessing high-frequency complex word-forms as wholes would not be any more cognitively taxing than accessing non-gradated complex words or monomorphemic words, which contain only one morpheme. Low-frequency tokens are presumably pieced together from their component parts using other stored word-forms as models. In the context of this research, presence or absence of consonant gradation is assumed to indicate which mode of lexical access was used to produce a particular lexical item: whole-word or decomposed access. It was assumed that if gradation in a complex word-form is standard-like, it is likely to have been accessed as a whole. Conversely, loss of gradation in a complex word-form is taken to indicate that the complex form is accessed via its component parts. Thus, the link between gradation and frequency is an indirect one; gradation is assumed to be dependent on the mode of lexical access, which is presumably conditioned by frequency of use.

Based on these assumptions, it was expected that Whole word-frequency affects Whole word accuracy so that frequent whole words produce more instances of accurate whole word responses than infrequent whole words. This addresses Research Question One. Similarly, more instances of accurate gradation are expected with frequent than with infrequent whole words, answering Research Question Two. Whole word-frequency was also used to predict Suffix-accuracy. Although no research question was formulated for it at the outset, frequent whole word-forms should produce more targeted suffixes than infrequent whole word-forms. With regards to frequency of suffixes (Research Question Three) and stems (Research Question Four), neither was expected to have an effect on gradation. Based on the literature on consonant gradation and cognitive grammar, it was proposed that the consonant alternations are not predictable from the suffix. Ungraded NOM.sg case stem word-forms (independent words) do not aid in recovering gradation in complex words because NOM case base word-forms and their graded allomorphs are phonetically dissimilar and used in different sentence-contexts. Moreover, complex words may not be available in the network due to autonomy (frequent word-forms) or low activation levels (infrequent word-forms).

With regards to gradation types (Research Question Five), it was anticipated that qualitative gradation would be less successful than quantitative. As was discussed in Chapter Two, qualitative is more unpredictable, less common and acquired later than its quantitative counterpart (Leiwo 1984, p. 67).

Lastly, immigrant generation was expected to affect gradation accuracy (Research Question Six) because the amount of exposure to native-like speech reduces in every generation born abroad. The next following paragraphs outline the findings. The

predictions and answers to the research questions are summarised in Table 5-1 below for convenience.

#### Research Questions, Predictions and Answers

1. Does whole word-frequency affect achieving whole word accuracy? *Yes. More accurate whole word responses are expected with frequent words than with infrequent words.*  
**Answer:** —No (unexpected result).
2. Does whole word-frequency affect achieving gradation-accuracy? *Yes. More accurate gradation is expected with frequent words than with infrequent words.*  
**Answer:** —No (unexpected result).
3. Does suffix-frequency affect achieving gradation-accuracy? *No. It is expected that no significant difference in gradation-accuracy is found between frequent and infrequent suffixes.*  
**Answer:** —Yes (unexpected result).
4. Does stem-frequency affect achieving gradation-accuracy? *No. It is expected that no significant difference in gradation-accuracy is found between frequent and infrequent stems.*  
**Answer:** —No (expected result).
5. Is direct quantitative gradation more successful than direct qualitative gradation? *Yes. Direct quantitative gradation produces more accurate gradation than direct qualitative gradation.*  
**Answer:** —Yes (expected result).
6. Does immigrant generation explain the variation in gradation-accuracy between the participants? *Yes. More accurate gradation can be expected with earlier generations than later generations.*  
**Answer:** —Yes (expected result).

**Table 5-1: Predictions and answers for research questions**

The first test, which measured the effect of Word-frequency on Whole-word-accuracy, answered Research Question One, *does whole word-frequency affect achieving*

*whole word-accuracy?* Participants' answers were judged in terms of the accuracy of a word-form for the sentence context. Any deviation in stems, gradation and/or suffixation from the targeted word-forms rendered the participants' answers non-targeted. This yielded a statistically insignificant result (see Table 4-2) and a "no" answer to Research Question One.

Research Question Two asked *does whole word-frequency affect achieving gradation-accuracy?* The answer is "no". Both frequency-categories of complex word-forms produced incidences of gradation in balanced numbers (see Table 4-3). Although the result did not confirm the hypothesis, the patterning of gradation loss in the data was shown in Section 4.1.1 to be meaningful as gradation-loss tracks with immigrant generation. This suggests that language use in the non-dominant language context of Sointula differs from that in Finland where Finnish is spoken as a dominant language. The finding is discussed in detail in Section 5.1.1 as one of the topics of interest that emerged from this research.

It is important to draw the reader's attention to the total number of words that manifest gradation-errors. Of 346 tested complex word-forms, 227 word-forms (66%) manifested standard gradation while 119 showed gradation-loss (34%). Of the 119 instances of gradation-loss, eight instances are due to the use of the PART case, 71 due to the use of the plain NOM.sg word-form without a suffix, and 40 (12%) due to the use of NOM.sg stem with a gradation-triggering suffix. While only the last represents classic analogical leveling, the two latter uses will both be taken as manifestations of analogical leveling until a finer distinction between two is made in Section 5.1.2.

The supplementary test Word-frequency \* Suffix-accuracy produced a significant result (see Table 4-4). The effect ran contrary to what was anticipated; infrequent words produced more targeted suffix-responses than frequent words. This curious result foreshadows the fact that there is more to suffixes than first thought as discussed in 5.1.2 below.

Research Question Three considered *does suffix-frequency affect achieving gradation-accuracy?* Suffix-frequency was found to have a significant effect on Gradation-accuracy (see Table 4-6). Contrary to the initial hypothesis, there were significantly fewer errors with word-forms containing infrequent suffixes rather than frequent suffixes. Thus, the answer to Research Question Three is ‘yes’. This finding is related to the previous one. The properties of suffixes become a point of interest in search of an explanation to this surprising result. It turns out that semantics influences the mental representations of suffixes, which in turn affects gradation-performance. This will be discussed as another finding in Section 5.1.2.

The last frequency-test, which measured the effect of Stem-frequency on Gradation-accuracy, answered Research Question Four, *does stem-frequency affect achieving gradation-accuracy?* The test did not show statistical significance (see Table 4-7). Recall that stem-frequency refers to the frequency of the stem in a complex word in a base-form. For instance, the complex word-form *rannalla* ‘beach.ADE.sg’ meaning *on the beach* contains a gradated stem *ranna-*, which means ‘beach’. Gradated stems cannot stand as words in isolation but its base-form *ranta* ‘beach.NOM.sg’ can and has a frequency ranking associated with it. The frequency of the stem made no difference to gradation success, as was expected. Thus, the answer to Research Question Four is ‘no’.

To answer Research Question Five, *is direct quantitative gradation more successful than direct qualitative gradation?*, the independent variable Gradation-type was tested on Gradation-accuracy. The overall results showed that quantitative gradation is more successful than qualitative gradation, as was predicted (see Table 4-9). This provides a “yes” answer to Research Question Five. These results are discussed in Section 5.1.3.

The answer to Research Question Six, *is immigrant generation able to explain the variation in gradation-accuracy between participants?*, is “yes”. The effect of Generation on Gradation-accuracy was found highly significant (see Figure 4-3). The generations fall in the anticipated order; G2 performs above all other generations while G2.5 performs above G3. These results are discussed as the last findings in Section 5.1.4. These interesting results are discussed in the following sections in detail.

### **5.1.1 Frequency of use and lack of expected frequency-effects**

Results showed that word-frequency, as defined by corpus-rankings, did not have an effect on achieving the targeted, correctly gradated word-forms, and neither did suffix-frequency. This is contrary to expectation and requires explanation. The only apparent frequency-effect was that of suffix-frequency but, as discussed in 5.1.2 below, this is different from the expected frequency-effects in that it arises from semantics, not from frequency of use.

Currently, no frequency-corpora exist for non-dominant languages, which leaves majority language-corpora as the only baseline by which to compare minority-speakers’ lexical knowledge. It is possible that a dominant-language word-frequency corpus is an

inappropriate tool for a non-dominant language context. Language use in Sointula may differ so much from that in Finland that corpus-rankings based on Standard Finnish cannot be used in Sointula to construct even broad and well-separated frequency categories of frequent and infrequent words<sup>24</sup>. That is, there may be a group bias in the non-dominant speaker population; isolated from Finnish spoken in Finland, Sointula Finnish may have evolved on its own to a distinct language variety that cannot be represented by a Finnish frequency-corpus. Alternatively, individual variation may remain too great to be represented by corpus-frequencies. If that is the case, the data may not match corpus-frequencies because the sample size is too small due to too few participants. As explained in Chapter Three (Methodology), attempts have been made to minimise this risk but the possibility still exists.

Another possibility that might explain the lack of frequency effects in Sointula data is the way frequency categories were quantified in the present study. The boundaries of frequency categories are always arbitrary by nature. What defines words as frequent and infrequent is the differential behavior words exhibit with respect to a phenomenon under investigation. Bybee (2002), for instance, describes the variation in the deletion of final /t/ and /d/ after a consonant in Chicano English speakers in Los Angeles to be observable at a cut-off point of 35 times per million words (p. 264). At that point frequent words behave differently from the rest of the words. This current study, being the first one applying corpus-frequencies to a non-dominant language context, had no previous work on which to base the frequency boundaries. Admittedly, the words used in both the frequency categories represented relatively frequent words if compared to Bybee's study.

---

<sup>24</sup> It is expected that the Finnish frequency categories based on rankings in Finland match those in Sointula. The word-rankings within the categories are not expected to match perfectly.

However, this study focused on non-dominant speakers. Thus, cut-off points as low as in Bybee's 2002 study could not be used in Sointula in order to receive data. The intermediate range of 1,541 words was left in between the frequency categories to alleviate this dilemma of relative frequency of the tested words in the present study. Within the intermediate range (from which no words were tested), frequency of use dropped from 1,111 times per million (the least common of the frequent words) to 596 times per million (the most common of the infrequent words). Another mitigating factor is the relative infrequency of words in Finnish in general. This is due to the enormous morphological complexity of words in Finnish; according to Järviö et al. (2006), 95% of the morphologically complex word stock in Finnish occurs less than once per million (p. 398). For instance, one word may occur in hundreds or thousands of variants in Finnish (Bertram, Schreuder & Baayen 2000) and thus the frequency of use of that word is distributed among these variants. In morphologically less complex languages, frequency of use is distributed among fewer variants, for which reason each word is used more frequently. This means that words that are frequent in morphologically less complex languages may be far less frequent in morphologically complex languages, such as Finnish.

Assuming that the experimental design is sound and that the measuring tool, corpus-rankings, is applicable in non-dominant language contexts, then language use in the non-dominant language setting of Sointula is a likely explanation for the discrepancy between the expected and the attested frequency-effects. The results suggest that word-frequency does not predict the phonetic shape of words in Sointula; gradation loss as an

analogical change<sup>25</sup> was found in both frequency-categories with no statistically significant difference between the two frequency-categories of word-forms. Thus, all words in Sointula behave as infrequent words do in dominant contexts. Recall that analogical changes are typical of infrequent words (Bybee 2001, p. 14). Without previous studies in non-dominant language settings to give a baseline of the expected error-rate attributable to analogical leveling, it cannot be known how strongly these results argue for infrequency. However, the fact that frequent words are affected by gradation-loss is alone suggestive of infrequency of all words in Sointula. It appears as a better explanation for why words did not differ in their behavior than the alternative, which is that all words are frequent.

In dominant language contexts, frequent and infrequent words behave differently and corpus-frequency is an accurate enough filter to show the behavioural difference. However, frequency effects are presumably contingent upon actual language use; a sufficient volume of language use over a time span is needed to manifest the frequency effects. An hourglass analogy helps demonstrate the point. Two hourglasses, one representing the dominant language context in Finland and another the non-dominant language context in Sointula, contain, say, a million words each. The speed at which the words flow in the Finnish hourglass is much greater thanks to high-volume use than the speed of the word-flow in the Sointula hourglass due to reduced exposure to the Finnish language. Thus, the flow of words in the two settings does not have the same effect on

---

<sup>25</sup> The unmarked or the most common form replaces exceptional forms (Bybee 2001). Applied to the present project, it means that gradated forms are replaced with non-gradated forms. Of the total number of 346 complex word-forms, 12% manifested analogical leveling (the use of the base-form together with a gradation triggering suffix). The analogical forms spread across both the frequency categories.

the speaker. All words in Sointula remain below some threshold-level of language use and therefore cannot be distinguished from each other by frequency.

Phonetic detail that is difficult to remember is leveled by means of analogy so that the phonetic shape of word-forms conforms to the more frequent and simpler patterns in the language (i.e. the use of the NOM.sg non-gradated stem). Overall exposure to words (i.e. token-frequency, or the frequency at which a word-form is encountered in the language) may not be sufficient to train the Finnish speakers in Sointula to the level of accuracy that speakers are able to attain in Finland. This is not an unreasonable assumption in a community where Finnish is no longer used as a means of daily communication and opportunities to hear and speak the language are few. That NOM case is the default case as the most common case type in Finnish leads to an interesting conjecture that type-frequency effects, or generalisations made of patterns in the language, may appear sooner than token-frequency effects (due to high frequency count for a word-token) in low-functionality language contexts.

Corpora are widely used amongst researchers adhering to usage-based approaches to language in particular. They have proven to be powerful tools in isolating the patterns of sound change, allowing researchers to make generalizations about the whys and hows of the progress of sound change in the lexicon. Corpus-based frequency effects (discussed in Section 2.3.1) have been proven to hold in various languages across the globe and have applied so consistently to various language phenomena that they have appeared as the widely held truth among usage-based grammarians (Bybee 2001; Phillips 1984 among others). Corpus-based frequency effects are so successful in describing the spread of sound change in the lexicon because they have been mostly, if not exclusively,

drawn from data from languages spoken in dominant language contexts and applied back to them. It was not possible to find other studies where the applicability of frequency effects had been tested in minority language contexts. It appears as if a high-volume language use over time is built into the corpus-frequency model. The current project has tested whether the corpus-model and the frequency effects hold in a low-volume language context. That these effects are not evident in the current study highlights the fact that high-volume language use over a time span can have been taken for granted in corpus-studies conducted in dominant language contexts. However, while the results do not show a differential behavior between frequent and infrequent words in Sointula, they suggest that analogical leveling pertains to Sointula data suggesting that all words behave as infrequent words do in dominant contexts. Previous work done on frequency effects in dominant language contexts allow analogical changes of Sointula Finnish to be linked with infrequency. Further research on analogical leveling in similar language settings is necessary to verify this result.

### **5.1.2 Atrophy of GEN object marker**

Attention now turns from word-frequency to suffix-frequency. Recall that suffix-frequency did produce a frequency-effect, correlating with Gradation-accuracy. However, the effect ran counter to intuition; infrequent suffixes showed significantly more standard gradation than frequent suffixes (see Table 4-6). When the effect of Suffix-frequency on the accuracy of the suffix was measured in a supplementary test (see Table 4-5), the direction of the effect was the same. Infrequent suffixes again show significantly more

targeted suffix-responses<sup>26</sup> than frequent suffixes. What property, or the lack thereof, in frequent suffixes makes them more difficult to target than infrequent suffixes? The answer to the question appears to lie — at least partly — in the semantics of suffixes within a sentence context.

Infrequent suffixes on Finnish nouns have specific functions that largely correspond to those of prepositions in English sentences. Therefore, they convey meanings such as ‘in’, ‘on’, ‘into’, ‘onto’, ‘from’, and ‘off’. Of the frequent suffixes, NOM plural marks the plural, as implied by its name. In addition to functioning as a quantifier and a possessive, GEN marks objects in the language. While the NOM plural case-marker has one easily comprehensible task in the language, GEN has four functions, exemplifying a linguistic phenomenon known as *syncretism* (see also section 3.2.1 in Chapter Three). It appears that not all functions of GEN are equally easy for non-dominant speakers. While quantifier and possessive functions of the GEN suffix *-n* have a semantic, or meaning-related, component associated with its form, the GEN suffix *-n* in the object functions does not.

Quantifying and possessive functions are fundamentally different from the object functions of GEN, as the following examples demonstrate. For instance, in *Nämä ovat äidin rahoja* “These are mom.GEN.sg money.PART.plu” meaning *This is mom’s money*, the GEN suffix *-n* marks the possessive. Hakulinen et al. (2004) identifies this function of the GEN as the most salient in meaning (p. 1184). The possessive GEN can be compared with the quantifying GEN, in which the *-n* suffix marks the expression of quantity, as in *Latu on kilometrin pituinen* “skitrail.NOM.sg is kilometre.GEN.sg long”

---

<sup>26</sup> Recall that the use of a wrong suffix or the failure to use one (i.e. the use of NOM.sg form) were coded as suffix-errors. (The coined words were removed from the analysis so as to make the data comparable with respect to suffix-responses and gradation.)

meaning ‘The ski trail is a kilometre long’ or in *Mari on vuoden ikäinen* ‘Mari is year.GEN.sg old’ meaning ‘Mari is a year old’. In addition to being semantically salient, both possessive and quantifying functions can be tagged on to an equivalent morpheme in English, the ‘s. In the sentence *He lives an hour’s walk away*, the GEN functions as a quantifier while in *The dog’s coat is soft*, it is a possessive. Thus, the possessive and quantifying functions of GEN suffix *-n* have an equivalent morpheme in English whose meaning supports the form.

GEN as a marker of grammatical objects does not have a salient semantic meaning but rather a structural function within a sentence. Its form can only be understood in relation to other sentence constituents. While NOM usually marks the subject in Finnish, GEN marks the object. The difficulty in producing the Finnish GEN object marker is likely to result from the fact that the object-function is not marked on full nouns in English. In English, the only way to keep track of who did what to whom is by keeping the relative order of the constituents in a sentence fixed. Object marking in Finnish, in turn, allows freedom in word order. All word-orders (S)ubject(V)erb(O)bject, SOV, VOS, VSO, are legitimate although some are more marked than others.

Larmouth (1974) showed that in four generations of Finnish speakers in the U.S., the word order becomes identical to English word order. In immigrant-populations object markers are elided and the word order becomes fixed, as illustrated in Table 5-2 below (an example from Larmouth 1974: 14).

|     |  |
|-----|--|
| G1: | Lintu lensi ikkunan läpi.<br>bird.NOM.sg fly.past.3sg window.GEN.sg through<br>‘The bird flew through the window.’ |
| G2: | Lintu lensi ikkuna(n) läpi.  |
| G3: | Lintu lensi ikkuna läpi. ~ läpi ikkuna.  |
| G4: | Lintu lensi läpi ikkuna.   |

**Table 5-2: GEN object atrophy**

In G1, the GEN *-n* (in bold) marking the object of the postposition *läpi* is invariably present while in G2 it is becoming variable, marked by the parentheses around the suffix. In G3 and G4, the GEN suffix is completely absent. Similarly, the word order is becoming unstable in G3, *läpi* being either treated as a postposition or a preposition. By G4, the word order has become identical to that in English. Thus, it may be that the GEN as the marker of the object function in Finnish is elided because the same goal can be achieved by a word order alone in English-influenced non-dominant speech.

A hypothesis arises from the above observations. Suffixes that are supported by semantics and have equivalent morphemes in English may be mappable to each other and retained better in English-influenced non-dominant Finnish than frequent suffixes. In other words, language *transfer* between English and Finnish would aid in preservation of Finnish suffixes in non-dominant Finnish. Of the frequent suffixes, the NOM.plu marker and the GEN suffix, marking possessive and quantifying functions, are arguably more semantically salient (mappable to English 's) than GEN in object functions. The forms of those infrequent suffixes that have directional and locative functions (tested in the current study) are also supported by English morphemes with close enough meanings. They should be retained relatively well. If this was the case, semantic suffix-salience happens to be associated to a greater extent with the infrequent group of suffixes than with the frequent group due to the stimuli<sup>27</sup> used in this study. In addition, what appears to be a frequency-effect would be a result of the semantics of suffixes. The hypothesis of semantic salience can be tested.

---

<sup>27</sup> See Appendix D for the breakdown of suffixes used in the stimuli.

If indeed the participants resort more often to the plain NOM.sg stem word-forms with words containing less semantically salient suffixes, they should do this with words which target the GEN suffix in object functions.<sup>28</sup> If this is the case, the difference between both the suffix-accuracy and gradation-accuracy in the two frequency-categories of suffixes should disappear if the object functions are filtered out. To determine whether this assumption is accurate, two post-hoc tests were run where the effect of Suffix-frequency was measured on 1) suffix-accuracy and 2) gradation-accuracy but this time all 95 object functions (30 objects of adposition and 65<sup>29</sup> objects of verbs) were filtered out<sup>30</sup>. The test for suffix-accuracy produced a null-result ( $\chi^2(1)=0.123$ ,  $p=0.725$ ). It can be concluded that the participants make more suffix-errors with less semantically salient GEN object marker than with other suffixes with more salient meanings. The result is attributable to the atrophy of the GEN case object marking in the immigrant population of Finnish speakers since the significance disappears with the removal of the GEN object functions.

This effect of suffix-frequency on suffix-accuracy also affects gradation-accuracy; the atrophy of the GEN object marker may also explain why frequent suffixes showed more gradation loss than infrequent suffixes. If participants resort to NOM.sg stems in object functions, then gradation will be absent in these responses as NOM.sg case bases are ungraded, regardless of frequency. On this take, the removal of the objects would also erase the counter-intuitive effect of suffix-frequency on gradation. Thus, another test

---

<sup>28</sup> To follow the reasoning for the following post-hoc tests, see Appendix G for a flow-chart.

<sup>29</sup> The total of 72 verbal objects were tested, for seven of which no answer was received. Thus, 65 verbal objects were left in the data to be filtered out.

<sup>30</sup> Word-forms for which no answer were given, stem word-forms, and coinage words (which initially were coded for the correctness of the suffix but not for gradation due to lack of frequency-rank) were also filtered out. The tests included targeted (0s), non-targeted grad-triggering suffix-responses (1s), PART (2s) and NOM.sg (3s) responses.

was run in which gradation-accuracy was measured in the two frequency-categories of suffixes (i.e. Suffix-frequency\*Gradation-accuracy) but all the word-forms tested in object-functions were filtered out. After the removal of the objects, the significance in the latter test was indeed reduced from  $\chi^2(1)=9.188$ ,  $p=0.002$  (see Table 4-6) to  $\chi^2(1)=4.482$ ,  $p=0.034$  but not erased altogether. Gradation in the infrequent category still remains more successful than in the frequent category. An unknown property (or properties) in other suffixes render(s) gradation in infrequent word-forms and suffixes more successful than in frequent word-forms and suffixes. It is reasonable to suggest that other word-forms (or case-types) in addition to GEN in object functions may have developed usage that differs from Standard Finnish due to language shift. This might explain the observed results. However, it is not possible to examine case-types beyond objects within the limits of this study.<sup>31</sup>

Since suffix-salience does not fully explain the unexpected behavior of gradation with regards to suffix-frequency test (i.e. that infrequent suffixes are more successful in producing gradation than frequent suffixes), another line of investigation was pursued examining the *appropriateness of the suffix* in the sentence context. Consider standard word-forms *rannalla* ‘**on** the beach’ and *rannasta* ‘**from** the beach’, for example. It is the suffix that makes these words different from each other and context-sensitive. The suffix determines whether the word is appropriate for the sentence or not. The question of interest is whether a difference in gradation-accuracy exists between words with a targeted suffix (i.e. the word fits the intended sentence context) and with a non-targeted suffix (i.e. the word offered is a standard word-form but ungrammatical in the sentence

---

<sup>31</sup> Balancing of cases within the frequency categories was not possible due to lack of suitable (controlled) words within the frequency-ranks 1 – 1500.

context). For this purpose two more post-hoc tests were run. The first test tallied the instances of correct gradation in words containing a targeted suffix. The test showed no difference in gradation-accuracy between the two frequency-categories of targeted suffixes ( $\chi^2 (1)=0.902$ ,  $p=0.342$ ). This implies that the difference in gradation-accuracy across frequency categories results from gradation-errors in words with non-targeted suffixes. The second test, which counted gradation only in words with non-targeted suffixes, confirms this with significant results ( $\chi^2 (1)=6.924$ ,  $p=0.009$ ). In sum, there is no difference in gradation-accuracy between the two frequency-categories of **targeted** suffixes. Instead, gradation-errors arise with the use of **non-targeted** suffixes<sup>32</sup>. The use of non-targeted suffixes leads to gradation-errors in one of three ways: a) using a non-targeted gradation-triggering suffix and leaving the stem ungraded, b) using the PART suffix which, being V-initial and creating an open syllable, does not trigger gradation; c) using a NOM.sg case stem, which being non-suffixed does not trigger gradation. Therefore, non-targeted suffixes are worthy of some more perusal.

The remaining question is whether a difference in gradation-accuracy exists between the two frequency-categories of **non-targeted** suffixes. Of the non-targeted suffixes, PART and NOM.sg do not trigger gradation. Word-tokens with PART and NOM.sg are left out and only **non-targeted gradation-triggering suffixes** are tested once more. It turns out that there is no difference in gradation-accuracy between the two frequency-categories of non-targeted gradation-triggering suffixes ( $\chi^2 (1)=0.166$ ,  $p=0.683$ ). The participants appear to know gradation equally well in words that contain a non-targeted and targeted gradation-triggering suffix .

---

<sup>32</sup> Non-targeted suffixes include 1) non-targeted gradation triggering suffixes, 2) PART suffix (does not trigger gradation), and 3) NOM.sg (lack of suffix and gradation).

By the process of elimination, the cause of the difference in gradation-accuracy between frequency-categories of suffixes can be narrowed down to the inappropriate use of PART and NOM.sg in Sointula's Finnish. The inappropriate use of PART and NOM.sg is greater in the frequent than in the infrequent category of suffixes and as a result, gradation errors are more common in the frequent group. Lack of suffix-salience with GEN objects can explain some of this use but beyond that no other pattern becomes obvious. A possible interpretation is that other cases (in addition to GEN in object functions) are replaced with NOM.sg and PART in Sointula's Finnish. This is plausible as Sointula's Finnish speakers, particularly in G2.5 and G3, are no longer fluent in their ancestral language. It is noteworthy that NOM.sg and PART cases are the two most frequent cases overall in Standard Finnish (see Table 2-11 in Chapter Two) and both of them are non-gradation triggering. NOM.sg appears to affect more words (or whole case-types) than just those which should have had GEN in object functions; while there were only eight instances of PART responses in total in the data (seven in the frequent category and one in the infrequent category of words), NOM.sg shape responses accounted for 71 of the total of 98 non-targeted responses. Other non-targeted gradation triggering suffixes were used in 19 instances.

The use of the plain NOM.sg stem word-form without a suffix should be distinguished from those cases in which the NOM.sg stem is combined with a gradation-triggering suffix. While both cases manifest gradation loss, the underlying reasons appear different. Gradation loss with gradation-triggering suffixes represents true analogical leveling, which arises from low frequency according to Bybee (2001). Gradation loss as the use of the plain NOM.sg stem seems to be indicative of

morphological attrition. In this case, loss arises from absence of gradation-triggering suffix, that is, GEN object marker, due to the atrophy of these suffixes in a non-dominant language setting. Lack of semantic salience can explain some of the morphological atrophy in the data.

The above Section 5.1.2 has discussed in detail the second main research finding, that is, the atrophy of GEN object marker and its affect on gradation accuracy in Sointula Finnish. Results pertaining to gradation-types and individual participants' and their immigrant generation remain to be discussed in the following sections.

### **5.1.3 Gradation-types**

The hypothesis regarding gradation-types was confirmed. After filtering out non-gradation-triggering responses, namely PART and NOM.sg answers, gradation-type results show that qualitative gradation is significantly less successful than quantitative gradation, as was initially predicted. Of the total of 40 gradation-errors in suffixed forms, 29 (72.5%) occur in word-forms subject to qualitative gradation (see Table 4-9). These findings parallel the results from Martin's (1998) work among Finnish immigrants in the United States; she also found qualitative gradation less successful than quantitative gradation.

As was discussed in Chapter Two, qualitative gradation is more unpredictable, less common and acquired later than its quantitative relative in dominant Finnish (Hakulinen et al 2004; Leiwo 1984). Karlsson (1974, in Leiwo 1984) argues that these characteristics of qualitative gradation indicate that it is disappearing from Standard Finnish. In this light, it is not surprising that qualitative gradation would erode faster than

quantitative gradation in non-dominant Finnish as well. An interesting conjecture is that the instability of qualitative gradation, and perhaps other linguistic change due to reduced language use, may be observable within a shorter time-span in non-dominant than in dominant language settings.

#### **5.1.4 Individual participants and immigrant generations**

The background questionnaire administered after the elicitation provided qualitative data on the participants. The study tested six individuals, known as P1 - P6. All participants had learned Finnish as their first language at home and used Finnish in daily life until they reached school age. Four had (ex)partners brought up in Finnish, but everybody uses/used English as the language of communication with their partners. Today, all use predominantly English in their daily life. The participants have little contact with Finnish TV or radio. Only P4 (from G2.5) reported reading Finnish magazines often. All of the participants comprehended when spoken to in Finnish and were able to formulate sentences in Finnish although all manifested deviant grammar to varying degrees. Generations reflected their language skills in such a way that G2 was the most fluent and comfortable generation speaking Finnish and showed least deviant forms. The participants in G3 were the least comfortable and least fluent speakers of Finnish. Their language also deviated from the standard to the greatest degree of all participants. All were able to recall most of the vocabulary needed with the exception of few words. Lexical loss was most apparent in P1, who represents G3.

The results confirmed the initial hypothesis regarding immigrant generations; G3 manifested more gradation-loss than G2 and G2.5. G2 was the least affected by gradation-

loss. It becomes obvious from the above results that consonant gradation is not sustained through generations in the immigrant community of Sointula. In the light of earlier findings, these results are hardly surprising.

In Finnish as a dominant language, consonant gradation functions as a cue that helps the listener anticipate the grammatical function of a word (Leiwo 1984). The function is often, determinable from the shape of the particular stem, whether it is strong or weak (p. 72). The atrophy of the GEN marker (discussed in Section 5.1.2) may indicate that in an English-influenced variety of Finnish, the grammatical constituency of nouns is recognised by the relative order of nouns alone, just as in English. There may no longer be the need for suffixes on full nouns as markers of grammatical functions of objects. In the same process, gradation in the stem would simultaneously become superfluous. The results show incremental gradation loss by immigrant generation. By G3, it has all but disappeared from use. These findings are in line with those of Larmouth (1974) who showed that by the third generation born abroad, Finnish syntax has become English-like. The prediction, that the need for suffixation and gradation becomes lost once the word order becomes more fixed with each passing generation, is testable by controlling for different word-orders and left to be determined by future research.

Also Yagmur, de Bot, and Korzilius's (1999) attested syntactic loss and difficulties in lexical retrieval among G1 speakers of Turkish in Australia. They found that low frequency and highly marked lexical items (i.e. specific vocabulary) are vulnerable to loss (p. 65). If gradation is encoded in whole words, then it can be expected that with loss of lexical items, gradation is lost as well. More words would have to be composed of parts, coined or replaced by base forms. In order to preserve exceptional

material in a language, high volume of use for words, types of patterns and word-strings is needed. Language use in Sointula was found to be insufficient to preserve the full phonetic form of complex words in Sointula.

Leiwo (1984) writes that maintaining unproductive patterns in a language is a strain. He argues that over time languages work toward more simplified grammars and toward freeing themselves of unnecessary exceptional grammatical features (p. 19). On this take, the disappearance of gradation from the non-dominant variety of Finnish can be expected. With fewer and fewer opportunities to use Finnish, there is little support and need for gradation overall in Sointula – and even less for qualitative gradation.

**Summary of results.** The above sections have discussed in detail the findings of this research. The first finding is that none of the anticipated frequency-effects were found. Nonetheless, the results are interesting. They suggest that all words behave as infrequent words do in dominant contexts. This was manifested in that gradation-loss as an analogical change occurred in both frequency-categories at a comparable frequency. The results also highlighted that high-volume language use over time is assumed in the corpus-model. While corpus-frequency rankings are capable of showing differential behavior between words in dominant language contexts, the lack of high-volume language use over time causes all words in Sointula behave like infrequent words do in dominant language contexts.

The second finding is that most of the gradation-errors arise from the atrophy of the GEN in object functions. The GEN object-marker (more numerous in the frequent category than other frequent case-suffixes) was shown to be less semantically salient than

infrequent suffixes because it is not supported by an English equivalent morpheme. For this reason it was not well preserved in English-influenced non-dominant Finnish. This was manifested as lower levels of suffix-accuracy in the frequent suffix-category. In turn, this affected gradation. Further examination of suffixes revealed that gradation-errors resulted from the erroneous use of the NOM.sg and PART cases. Since both are non-gradation-triggering forms, gradation was affected. Although lack of saliency of GEN in object functions can partly explain gradation-loss in the data, it did not account for all of the gradation-errors in the frequent suffix-category. It seems likely that some individual words (or whole case-types) appear to have been replaced with the NOM. The following generalisation can be made: correct gradation is predictable based on suffix-use, which in turn is determined in part by salience of the suffix. Thus, the semantics of suffixes turns out to be a better predictor of gradation than frequency in a non-dominant context.

The results confirmed the initial hypothesis that qualitative gradation is less successful than quantitative gradation in Sointula Finnish. In light of the observation (by Karlsson 1974, in Leiwo 1984) from dominant Finnish, that qualitative gradation is not stable in Standard Finnish, disappearance of qualitative gradation before quantitative gradation from Sointula Finnish can be expected.

The last section confirmed that immigrant generation can explain the variance between the participants in the production of consonant gradation. Loss of gradation increased with each generation born abroad; by G3, it has all but disappeared. It becomes clear from the results that consonant gradation is not sustained through the generations. These findings can be compared with those of Larmouth (1974) who showed that by the third immigrant generation born abroad, sentence structure becomes fixed and identical to

English. It is possible that the atrophy of the GEN object marker is linked to the tendency in English-influenced Finnish speech to mark the contrasts of syntactic constituency with word order alone. In this process, the need for suffixation and gradation would become neutralised.

Thus far this thesis has discussed the main findings that were isolated from the Sointula data. In the final section, Section 5.2, the discussion evaluates evidence to infer how complex word-forms might have been accessed by the participants, either as wholes or as parts.

## **5.2 Implications for Cognitive Grammar**

### ***5.2.1 Accounting for standard forms***

Based on Bybee (2001), it was proposed that standard gradation would result from whole word access facilitated by frequent language use (see Sections 2.3.1 and 2.7 in Chapter Two). In Sointula, however, no word-forms can be described convincingly as frequent. Regardless of frequency, standard gradation was attested in balanced numbers across both frequency categories of complex word-forms. In principle, there are three ways through which standard gradation in the stem could have been achieved: 1) the gradated stem is retrieved and combined with the targeted suffix, 2) gradation rules modify the consonants of a NOM.sg case base and produce a gradated stem, which is then combined with the targeted form, and 3) complex words in which gradation and the suffix are embedded are accessed as wholes. How likely are the above options to have

applied and produced the incidences of standard gradation in the data? Each of the possibilities is evaluated in the following paragraphs.

With regards to the first option, if it is assumed that plausible cognitive entities (i.e. units of perception and production) in isolation have representation in the lexicon (Bybee 2001, p. 30), then gradated stems could have storage in the memory. However, storing gradated stems is highly unlikely. The NOM case is by far the most frequent case-type (see section 2.3.1 of Chapter Two) and thus the basis of analogical formations. In other words, NOM has the highest type-frequency. The meanings of the bases and gradated stems overlap. This is similar to English in which the stem allomorphs [**naif**] and [**naiv**] in *knife* and *knives* do not differ in meaning. As a result, gradated stem-shapes could easily go unnoticed. In addition, one would expect that, if stored, gradated stems would occasionally surface in isolation. In other words, the participants could erroneously choose a gradated allomorph in contexts where the ungraded stem is appropriate, e.g. \**ranna* instead of *ranta* “beach.NOM.sg”. In addition, gradated stems could surface in combination with non-gradation triggering suffixes (i.e. vowel-initial suffixes), such as PART *-A* meaning ‘of a/an N(oun)’ or partial object or ILL *-OON*<sup>33</sup> meaning ‘into’. However, errors such as these were never encountered in the data. No evidence can be found that gradated stems are stored as memory tokens and used in combination with suffixes. Therefore, it is unlikely that standard gradation was achieved by activating a memory token representing a gradated stem.

As per the second possibility, one could propose that standard gradation was achieved through applying structuralist gradation rules that modify the base consonants.

---

<sup>33</sup> The capitalisation means that the vowel quality in the suffix changes according to vowel harmony. The allomorphs of PART are *-a* or *-ä* and ILL *-oon* and *-öön*.

This is also highly unlikely given the following evidence. Recall that gradation is highly stem-specific, depending on the consonants involved. In addition, exceptions to the gradation-patterns are common (see Section 2.1 in Chapter Two). That gradation outcome largely depends on the stem provides evidence against the proposal of gradation rules. An array of rules would have to account for the various gradation outcomes with different consonants in different structural contexts. Moreover, proposing rules leaves the problem of how to explain the exceptions. Rules should apply every time an appropriate context is encountered. However, certain Finnish words do not get gradated even though they have the structural and the phonetic context for change.

The fact that no incidence of incorrect application of gradation was encountered in the data provides a further argument against gradation rules. If gradation-rules existed, in a non-dominant language context speakers could be expected to make errors in applying these rules in inappropriate places due to lack of exposure to them. Gradation was either perfectly applied or not at all. For example, there were no instances where Qualitative gradation was applied when Quantitative was called for, or vice versa. Nor were there errors where gradation was attempted but the quality of the consonant involved was amiss. Thus, it is unlikely that standard gradation was achieved through application of gradation-rules.

The third option proposes that standard gradation in the data was achieved through accessing complex word-forms as wholes. Based on Bybee (2001), whole word access was proposed to be associated with frequent words only. However, despite the fact that words in Sointula largely appeared to behave as infrequent words do in dominant language contexts, the results from this study suggest that they were accessed as wholes

regardless of frequency, contrary to Bybee. A comparison of gradation accuracy with **targeted** and **non-targeted gradation-triggering suffixes** (see also Section 5.1.2) reveals that gradation-accuracy does not vary<sup>34</sup> with suffix-accuracy. Gradation accuracy remains relatively stable with gradation-triggering suffixes, regardless whether they are the right or wrong suffixes (i.e. words) for the sentence context. This is intriguing as it suggests that the words are, for the most part, whole units in which no separation between gradation and the word exists. Gradation and the suffix appear to get activated together as one whole unit.

Finding evidence for whole word access in Sointula is surprising given that, as already mentioned, all words in Sointula can be interpreted as infrequent<sup>35</sup>. According to Bybee (2001), infrequent words are affected by analogical leveling because they are difficult to keep in mind (e.g. *dream* + *ed*; cf. *dreamt*). They are constructed from parts because whole word representations do not exist for them or cannot be found due to low frequency of use. In other words, infrequency predicts decomposition according to Bybee. However, the data show that of 346 word-forms which should have all been gradated, 66% (227 instances) show standard gradation. How were these accessed? An argument was made above against storing gradated stems and gradation-rules. Thus, whole word access emerges as the best explanation of how targeted forms were achieved.

Proposing whole word access eliminates the need to differentiate between irregular and regular formations or between simple or complex words (Hay & Baayen 2005, p. 343). In this view, traditionally problematic morpho-phonological alternations,

---

<sup>34</sup> Cf. Gradation in words with targeted suffixes (84.7%) and with non-targeted suffixes (89.5%). Note that both suffix-uses result in standard forms.

<sup>35</sup> Gradation loss as a true analogical change (40 instances/346 complex words) also occurred in balanced numbers (Freq: 21; Infreq: 19) across both word-frequency categories of words.

such as Finnish consonant gradation, would be accounted for. When comparing morphemes and whole words as preferred units of storage, Hay & Baayen (2005, p. 343) argue that full words have stronger representations than morphemes. Langacker (2000) writes that speakers are more aware of whole words than of their parts. According to him, a large portion of complex words are initially learned as wholes and analyzed later, if at all (p. 46)<sup>36</sup>. Intuitively, language users' preference for whole words is likely to stem from the fact that in natural language word-forms exist in sentence contexts where they occupy their slots. The focus on the shape of the whole word ensures that it fits in the sentence. Inflectional morphology, such as Finnish case-suffixes tested here, modifies a word for its sentential context. The word-forms in this study were also elicited in sentences, in the most natural context they can be.

Someone skeptical of whole word access could argue that if semantically salient suffixes in complex words can be recognised as units of specific meaning, as the results of this study suggest, complex words must then be made of parts. It need not be so. Langacker (2000) argues that, in general, complex words do not get their full meaning from their parts. This is particularly true for derivational morphology (i.e. morphology to create new words). For instance, words such as *propeller*, *ruler*, and *stretcher* do not get their meaning by composition but are initially learned as unanalysed wholes. This is despite the fact that the suffix *-er* is easily isolatable in all of them. Successful use of those words does not require them to be linked with their bases *propel*, *rule*, or *stretch*. Even if they were connected, it is not important to what degree the composite expressions activate their components. Furthermore, considerable variation in this respect can be

---

<sup>36</sup> During the elicitation for the current project, one participant (from G2.5) produced the targeted form *illalla* "in the evening" without hesitation but then asked "is it correct?" as if having noticed for the first time that the cited form does not accurately match the parts (e.g. *ilta* 'evening.NOM.sg' + *lla*).

expected between individuals (Langacker 2000, p. 47)<sup>37</sup>. Just the same it is possible, but not necessary, for a native speaker of English to isolate the *un-* prefix in *unruly* and not think of the residue *ruly* to be related to *rule* — which it is. *Ruly* is a bound stem, as are gradated stems.

Similarly in inflectional morphology, complex gradated word-forms can be successfully used independently of their ungraded counterparts. Even if semantically salient suffixes could be extracted from complex word-forms, the remaining gradated stems are not necessarily easily recognisable without their suffixes; so different are the phonetic shapes of gradated stems from their non-graded allomorphs. For instance, a native speaker of Finnish cannot be certain that the meaning *ranta* “beach.NOM.sg” or *poika* “boy.NOM.sg” can be linked to the gradated stems *ranna* or *poja* if they were separated from their suffixes (e.g. *rannalla* “on the beach” or *pojalla* “in the possession of the boy”). Assigning meaning to gradated stems in isolation is difficult since such situations never arise; gradated stems are rarely, if ever, analysed alone as they do not exist in isolation. Therefore, the ability to isolate a suffix, or an affix in general, from a complex word does not necessarily mean losing the sight of the whole and that affixed forms are not accessed as wholes.

Based on the argument in the previous paragraphs, whole word access emerges as the most satisfying explanation of the three possible ways to achieve standard forms in the data. This is despite the fact that no words cannot be convincingly called frequent in Sointula. A whole approach to morphology called *Word and Paradigm Morphology*

---

<sup>37</sup> For example I, as an English-language learner, never came to think until very recently that the word “dependent” is inside the word “independent”. To me, they were unrelated entries in the lexicon both of which I could use successfully without having noticed that *in-* in “independent” is a prefix. Prior to that revelation, I had certainly taken an advantage of the productive uses of the prefix *in-* in other words.

(*WPM*) exists which views that lexicons are made of whole words. It is presented below as a further argument for how standard forms were achieved in the data.

*Word and Paradigm Morphology (WPM)* is built around the notion that full words are basic units of lexicons (Matthews 1972; Blevins 2006). Blevins's version of the model is compatible with other models of cognitive grammar in that Blevins also proposes **a network of exemplar representations for words**, just as Bybee's network model<sup>38</sup>, along the lines of what has been adopted here (see Figure 2-1 in Chapter Two), but takes the concept a step further by claiming that less than a word-size morphemes do not exist as "building blocks" from which words can be constructed (Blevins 2006, p. 536). According to WPM, less than word-size units cannot provide enough detail to predict the shape of words and recover their word-forms. Blevins (2006) builds his argument partly on Saami and Estonian, both languages related to Finnish which also have consonant gradation in their grammars. He argues that a whole-word based approach "avoids the problem of reconstituting forms from their less informative parts by retaining word forms, and using them as the basis for generalizations and analogy" (p. 568). However, Hay & Baayen (2005) think that the no morphemes-view is too strong. According to them, a less than word-size part can accrue lexical strength. The amount of lexical strength depends not so much on the frequency of the sub-word part as on how much its form is supported by other sub-word forms in similar positions in other whole words. This support is called *paradigmatic analogy* (p. 343). They propose that the more paradigmatic support a part receives, the more it is "present" in a complex word.

---

<sup>38</sup> Bybee views that affixes are in the lexicon attached to words and analysed through lexical connections and available for productivity, or word-formation (1995, p. 635).

There is little evidence in this study for storage of less-than-word-size morphemes. Therefore, WPM seems to better account for the findings than the corpus-model, which proposes mental representations for both whole words and sub-word morphemes based on frequency. However, the participants did make some errors in gradation. The following section speculates on the ways by which these gradation-errors might have occurred.

### **5.2.2 Accounting for non-standard forms**

As was mentioned in Section 5.1.2, there were two types of non-targeted forms in the data. The first type was the plain NOM.sg word-forms, which appeared to arise from morphological attrition. The second type of non-targeted form refers to the true manifestations of analogical leveling<sup>39</sup>, that is, loss of gradation despite the presence of gradation-triggering suffixes. Although all words in Sointula appear to behave like infrequent words in dominant language contexts, only 12% of the total number of 346 complex forms in the data manifested analogical leveling. Two possibilities exist by which these gradation-errors could have taken place.

One is that the gradation-errors in the suffixed forms are due to decomposition, as predicted by analogical leveling (see Section 2.3.1). Non-standard word-forms manifesting gradation loss could have been achieved by combining bases with suffixes. Recall that piecing complex forms from their composite parts predicts more gradation loss than accessing complex forms as wholes because the stems are not gradated. It is

---

<sup>39</sup> Eliminating exceptional material by replacing it with the most common pattern in the language, e.g. *sweep* – *swept*; cf *swept*.

possible that this mode of access was used when no whole-word mental representation existed or could be found for a complex word.

The other is that whole-word exemplars exist for the levelled (non-standard) word-forms in the lexicons of Sointula's non-dominant Finnish speakers and were activated. Gradation-errors in some words could have become so entrenched that these words exist as whole exemplars in the lexicons of Finnish speakers in Sointula. As predicted by the Exemplar Model (discussed in Section 2.4 of Chapter Two), non-standard forms could gain high activation levels if exemplars are not updated with standard forms. It may be that two whole-word exemplars, a standard form and a non-standard form, were competing at some point in a speaker's lexicon. Since the exposure to standard language is reduced in Sointula, a non-standard form may have won. The more listeners and speakers share the same non-standard forms, the more these exemplars gain in strength and the more likely they are to be selected.

Since non-standard word-forms accessed via their parts and whole-word exemplars representing non-standard words are phonetically the same, it not possible to distinguish which mode of lexical access was used to retrieve them from the lexicon. However, it can be concluded is that decomposition must have occurred at some point in time for gradation-errors to enter in the data.

## **Chapter Six Conclusion**

The final chapter presents four concluding sections that complete this research project. The chapter begins with a summary of the research project and its findings. The second section evaluates the limitations of this study. The third section briefly presents some directions for future studies. The last section weighs the possible contributions this project makes to the knowledge on usage-based grammar and cognitive linguistics.

### **6.1 Summary of Research**

The main goal of this thesis was to investigate whether frequency of use affects the production of consonant gradation by non-dominant Finnish speakers in the immigrant community of Sointula, BC, Canada. In the search for frequency-effects which would explain variation in gradation in non-dominant Finnish, the effect of three types of frequency, word-frequency, suffix-frequency, and stem-frequency, were tested on the phonetic output of the participants. The variables of phonetic output were Gradation-accuracy, Whole-word-accuracy, and Suffix-accuracy. In addition to testing frequency of use, this study also investigated whether one gradation type, either Quantitative or Qualitative, is more successful in producing gradation. Finally, this study examined to what extent individual participants manifest loss of gradation and whether immigrant generation can explain the variation between participants.

To investigate these questions, an experiment and then a background questionnaire were administered to the participants during a trip to the community. The

speakers translated English sentences which contained words of interest into Finnish. The words were either NOM.sg case bases or their complex forms and met the pre-set criteria for controls. The words were inserted into Finnish sentences which were then translated into English, randomised, and elicited from the participants. The data used for this study consisted of the speech samples of six individuals who were all born and raised in the community and represented second or later immigrant generations. Two participants represented each of the generations, G2, G2.5, and G3. The number of targeted and non-targeted responses were tallied and subjected to Chi-square tests to determine statistical significances.

It was expected that frequent complex words would manifest more standard gradation than infrequent complex words. The frequency-effects were assumed to be contingent upon the mode of lexical access; frequent complex words would be accessed as wholes thanks to frequent usage and as a result would not exhibit many gradation errors. On the other hand, infrequent words would be accessed via their composite parts and consequently were expected to exhibit more gradation errors. However, the frequency-effects were not found. In Sointula, frequent words could not be differentiated from infrequent words by their behavior. Words in both frequency-categories behaved as infrequent words do in dominant language contexts, manifesting gradation loss as an analogical change. The corpus-model assumes high-volume language use over time in dominant language contexts. Lack of volume in Sointula is hypothesized to suppress the differential behavior between frequent and infrequent words.

Frequency was not as good a predictor of targeted suffix-responses as semantics. Suffixes that have semantically salient meanings<sup>40</sup> and corresponding morphemes in English appear to be retained better than those that do not. The participants had difficulties in retaining GEN in object functions. It is possible that the atrophy of the GEN object marker is linked to the tendency in English-influenced Finnish speech to mark the contrasts of syntactic constituency with word order alone, as reported by Larmouth (1974). Gradation loss with the use of the NOM.sg case word-forms is taken to indicate morphological attrition (i.e. absence of trigger for gradation) in non-dominant Finnish. This type of gradation loss differs from gradation loss as an analogical change due to low frequency.

The results confirmed the hypothesis that qualitative gradation is less successful than quantitative gradation in Sointula Finnish. Disappearance of qualitative gradation before quantitative gradation from Sointula Finnish is not surprising given that qualitative gradation is unstable in dominant Finnish (Karlsson 1974, in Leiwo 1984) also.

The results also showed that consonant gradation is not sustained through the generations; loss of gradation increased with each generation born abroad and by G3, it has all but disappeared. If gradation were coded lexically, lexical loss among immigrants (also found by Yagmur, de Bot, & Korzilius 1999) would predict more gradation loss among non-dominant speakers of Finnish.

The above findings were evaluated in terms of the mode of lexical access likely to have produced them. Although no words in Sointula appears to qualify as frequent by the standards of dominant language contexts, most forms in the data manifested standard

---

<sup>40</sup> Semantically salient meanings include the possessive, quantifying, locative and directional functions of the tested suffixes.

gradation. Whole word access emerges as the most likely way by which they were achieved. Consonant gradation is best explained as a property of whole word-forms since it cannot be reliably recovered if words were composed of parts. Moreover, the data show that gradation performance does not vary with suffix-accuracy. This implies that no separation between gradation and the suffix exists and that gradation is associated with individual lexical items. Few gradation errors arise from true cases of analogical change (i.e. the use of the NOM.sg shape stem with gradation-triggering suffix)<sup>41</sup>. This is against predictions made based on Bybee (2001, 2007). All evidence combined, the data provides a stronger argument for storing whole words than sub-word morphemes in the lexicon, regardless of frequency.

## 6.2 Limitations

While attention to detail was attempted at every stage of this research and careful consideration preceded every decision, taking care of the following two issues that emerged later would have had a favorable effect on the research outcome. The first issue involves the properties of suffixes. Case-suffixes were controlled with respect to their frequency. Ideally, their grammatical functions would also have been controlled to avoid weighing one suffix-type more than the other. The second issue concerns the number of participants. Due to unexpected circumstances, the speech samples of six individuals of the total of 14 interviewed could be used in the research. With six participants only, the results are limited in their generalisability.

---

<sup>41</sup> There were 19 instances of gradation-loss in suffixed forms (i.e. analogical levelling), 71 in NOM.sg shape word-forms (i.e. morphological attrition). (Eight instances of gradation loss in the data arise in PART but a PART word is a standard form.)

## 6.2 Future research

This research is the first study to investigate the effects of frequency with more than one non-dominant speaker. Replication of the study with more participants and refined methodology would further clarify if frequency, and/or other factors such as semantics, are associated with the production of morpho-phonological alternations in complex words in non-dominant speakers. Another interesting extension of this research would be to examine whether semantically salient suffixes are available for productivity to a greater extent than non-salient suffixes. If semantically salient suffixes are parsable, it should also make them more productive, following the predictions of Hay and Baayen (2005). In addition, lexical encoding of complex word-forms in non-dominant speakers remains an interesting empirical question. Future research could examine if priming affects the retrieval of complex words and whether allomorphy, or morpho-phonemic irregularities, block the recognition and retrieval of morphemes. Similarly, complex words may be targeted more reliably if they occur together with another word that often accompanies it, e.g. verb + object sequences, reminding us that *string frequency* and *lexical bundles* (Tremblay, Derwing & Libben 2007) are also relevant in lexical retrieval of complex words. Yet another interesting area of research would be to examine the language knowledge of aphasics with respect to morpho-phonological alternations in complex words. Seeing how structures break down in aphasia would help us better understand how language is organised by the brain.

## 6.4 Contributions

The language knowledge of non-dominant speakers has sparked little interest outside language contact and attrition research. This pioneering research has demonstrated that non-dominant speakers' language knowledge can be valuable in that it produced interesting and relevant findings for usage-based grammarians. The findings prove the application of the corpus-model with its frequency-effects into a non-dominant language environment a useful exercise but show that frequency corpora require high volume of language use, which is not attainable in non-dominant language settings. Furthermore, this research made attempts to explain how multi-morpheme words might be stored in the brain and, by doing so, entered in the complex area of lexical encoding from an angle that is, to the author's knowledge, original.

## Bibliography

- Alba, R., Logan, J., Lutz, A. and Stults, B. (2002). Only English by the third generation? Loss and preservation of the mother tongue among the grandchildren of contemporary immigrants. *Demography*, 39 (3), 467-484.
- Andersen, R. W. (1982). Determining the linguistic attributes of language attrition. In R. D. Lambert & B. F. Freed (Eds.), *The Loss of Language Skills* (pp. 83-117). Rowley, MA: Newbury House.
- Anderson, A. (1958). *History of Sointula*. Sointula: Sointula Centennial Committee.
- Aoyama, K. (2002). *Quantity contrasts in Japanese and Finnish: Differences in adult production and acquisition*. In Y. Shirai, H. Kobayashi, S. Miyata, K. Nakamura, T. Ogura & H. Sirai, *Papers from the Second Annual Conference of the Japanese Society for Language Sciences* (pp. 185-206). Tokyo: Kuroshio.
- Appel, R. and Muysken, P. (1987). *Language contact and bilingualism*. London: Edward Arnold.
- Barlow, M. & Kemmer, S. (2000). *Usage-based models of language*. Standford, CA: CSLI.
- Bauer, L. (2001). *Morphological productivity*. Cambridge: Cambridge University Press.
- Berez, A. L. & Gries S. T. (May 3, 2008). *In defence of corpus-based methods: A behavioral profile analysis of polysemous 'get' in English*. Paper presented at North West Linguistic Conference 2008, University of Washington, Seattle, WA.
- Bertram, R., Schreuder, R. & Baayen, H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Learning, Memory and Cognition* 26 (2), 489-511.

- Bird, S. (2008). An Exemplar Dynamic approach to language shift. *Canadian Journal of Linguistics* 53 (2/3), 387-398.
- Blevins, J. P. (2006). Word-based morphology. *Linguistics*, 42, 531-573.
- Bot, K. de & Cline, M. (1994). A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. *Journal of Multilingual and Multicultural Development*, 15 (1), 17-28.
- Bot, K. de & Weltens, B. (1991). Recapitulation, regression, and language loss. In H. W. Seliger & R. Vago (Eds.), *First language attrition* (pp. 31-51). Cambridge: Cambridge University Press.
- Brezinger, M. & Dimmendaal, G. J. (1992). Social Contexts of language death. In M. Brenzinger (Ed.), *Language death: Factual and Theoretical explorations with Special reference to East Africa* (pp. 3-5). New York: Mouton de Gruyter.
- Brown, J.D. (1988). Chapter 5: Critiquing statistical studies. *Understanding Research in Second language Learning: A Teacher's Guide to Statistics and Research Design*. (pp. 43-62) New York: Cambridge University Press.
- Brown, J.D. (1991). Statistics as a foreign language – Part 1: What to look for in reading statistical language studies. *TESOL Quarterly*, 25 (4), 569-586.
- Brown, J.D. (1992). Statistics as a foreign language – Part 2: More things to consider in reading statistical language studies. *TESOL Quarterly*, 26 (4), 573-609.
- Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65-85). Stanford: CSLI.
- Bybee, J. (2001). *Phonology and Language Use*. Cambridge University Press.
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261-290.

- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82 (4), 711-733.
- Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- Bybee, J. & Brewer, M. A. (1980). Explanation in Morphophonemics: Changes in Provençal and Spanish Preterite Forms. In J. Bybee (Ed.), *Frequency of use and the organization of language* (pp. 41-73). Oxford: Oxford University Press.
- Bybee, J. & Newman, J. E. (1995). Are stem changes as natural as affixes?. *Linguistics* 33, 633-654.
- Carroll, D. W. (2008). *Psychology of Language*. Belmont: Thomson Wadsworth.
- Crawford, K. (2008, August 6). North Island dreams of better days. *The Tye*. Retrieved from <http://thetyee.ca/News/2008/08/06/NorthIsland/> on Feb. 5, 2009.
- Dorian, N. (1981). *Language death: The life cycle of a Scottish Gaelic Dialect*. Philadelphia: University of Pennsylvania Press.
- Fishman, J. A. (2001). *Reversing Language Shift: Theoretical and empirical foundations of assistance to threatened languages*. Clevedon: Multilingual Matters.
- Gonzo, S. and Saltarelli, M. (1983). Pidginization and linguistic change in emigrant languages. In R. Andersen (Ed.), *Pidginization and creolization as language acquisition* (pp. 181-197). Rowley, MA: Newbury House Publishers.
- Gregory, M., Raymond W.D., Bell, A., Fosler-Lussier, E. & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society* 35, 151-166.
- Hakulinen, A., Vilkuna, M., Korhonen, R., Koivisto, V., Heinonen, T.R., & Alho, I. (2004). *Iso Suomen Kielioppi [The Large Grammar of Finnish]*. Helsinki: Suomalaisen Kirjallisuuden Seura.

- Hamideh, J. & Hamideh, M. (2006). L1 Persian attrition. *The Linguistics Journal* 1 (1). Retrieved from [http://www.linguistics-journal.com/January2006\\_V1.pdf](http://www.linguistics-journal.com/January2006_V1.pdf) . ISSN Print: 1718-2301.
- Hatch, E. & Farhady, H. (1982). Chapter 2: What is a variable? *Research Design and Statistics for Applied Linguistics* (pp. 12-17). Rowley, MA: Newbury House Publishers.
- Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics*, 39 (6), 1041- 1070.
- Hay, J. (2003). *Causes and Consequences of Word-structure*. New York: Routledge.
- Hay, J. & Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, 9 (7), 342-348.
- Helimski, E. (1997). Factors of Russianization in Siberia. In Shoji, H. & Janhunen, J. (Eds.), *Northern Minority Languages* (pp. 77-91). Osaka: The National Museum of Ethnology.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93, 411-428.
- 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* (pp. 95-105). Amsterdam: North-Holland Publishing Company.
- Hulsen, M. (2000). *Language loss and language processing: Three generations of Dutch immigrants in New Zealand*. Unpublished Doctoral Dissertation, Nijmegen: Katholieke Universiteit Nijmegen. Retrieved May 26, 2008, from [http://webdoc.ubn.kun.nl/mono/h/hulsen\\_m/langloanl.pdf](http://webdoc.ubn.kun.nl/mono/h/hulsen_m/langloanl.pdf)
- Janhunen, J. (Ed.) (1997). *Northern Minority Languages*. Osaka: The National Museum of Ethnology, 147-159.

- Janse, M. (2003). Language death and language maintenance: Problems and prospects. In M. Janse and S. Tol (Eds.), *Language Death and Language Maintenance: Theoretical, practical and descriptive approaches* (pp. vii-xvii). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Jaspaert, K. & Kroon, S. (1989). Social determinants of language loss. *I.T.L.: Review of Applied Linguistics*, 83/84, 75-98.
- Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson & J. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-165). San Diego: Academic Press.
- 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* (pp. 229-254). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Järvikivi, J., Bertram, R. and Niemi, J. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. *Language and Cognitive Processes*, 21 (4), 394-431.
- Karlsson, F. (1974). Centrala problem i finskans böjningsmorfologi, morfofonematis och fonologi [Central problems in Finnish declension morphology, morphophonetics and phonology]. *Suomi*, 117 (2).
- Karvonen, P. (2001). *The People of Sointula*. [Videorecording] Edmonton: Karvonen Films.
- Krug, M. (1998). String frequency: a cognitive motivating factor in coalescence, language processing and linguistic change. *Journal of English Linguistics*, 26, 286-320.
- Köpke, B. & Schmid, M. S. (2004). Language attrition: The next phase. In M. S. Schmid, B. Köpke, M. Keijzer and L. Weilemar (Eds.), *First language attrition: Interdisciplinary perspectives on methodological issues* (pp. 1-43). Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Landry, R., Allard, R. & Henry, J. (1996). French in South Louisiana: Toward Language Loss. *Journal of Multilingual and Multicultural Development*, 17 (6), 442-468.
- Larmouth, D. W. (1974). Differential Interference in American Finnish Cases. *Language*, 50 (2), 356-366.
- Lainio, J. (1989). Spoken Finnish in Urban Sweden. *Uppsala Multiethnic Papers 15*. Uppsala: Centre for Multiethnic Research, Uppsala universitet.
- Langacker, R. (2000). A dynamic usage-based model. In M. Barlow and S. Kemmer (Eds.), *Usage-based models of language* (pp. 1-63). Stanford, CA:CSLI.
- Langacker, R. (2002). *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin/New York: Mouton de Gruyter.
- Leiwo, M. (1984). Kieliopillinen ja psykolingvistinen produktiivisuus [Grammatical and psycholinguistic productivity]. In K. Sajavaara, J. Tommola & M. Leiwo, *Psycholinguistic Papers IV* (37), 63-75.
- Maandi, K. (1989). Estonian among immigrants in Sweden. In N. Dorian (Ed.), *Investigating obsolescence: Studies in language contraction and death* (pp. 227-241). Cambridge: Cambridge University Press.
- Martin, M. (1998). *Amerikansuomen morfologiaa ja fonologiaa* [The morphology and phonology of American-Finnish]. Suomen kielen lisensiaatin työ: Jyväskylän yliopisto [University of Jyväskylä], Finland.
- Matthews, P. H. (1972). *Inflectional morphology: a theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.
- McKinnie, M. & Priestly, T. (2004). Telling Tales out of School: Assessing Linguistic Competence in Minority Language Fieldwork. *Journal of Multilingual and Multicultural Development*, 25 (1), 24-40.
- Mowrey, R. & Pagliuca, W. (1995). The reductive character of articulatory evolution. *Rivista di Linguistica* 7, 37-124.

- Newman, P. & Ratliff, M. (2001). *Linguistic Fieldwork*. Cambridge: Cambridge University Press.
- Nykysuomen taajuussanasto [Modern Finnish Lexicon]. Aineistopalvelu Kaino. Kotimaisten kielten tutkimuskeskus (KKTK) [Research Institute for the Languages in Finland]. Retrieved from <http://kaino.kotus.fi/sanat/taajuuslista/parole.php>
- Paradis, M. (2004). *A neurolinguistic Theory of Bilingualism*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Pavlenko, A. (2004). L2 influence and L1 attrition in adult bilingualism. In M. S. Schmid, B. Köpcke, M. Keijzer, and L. Weilemar (Eds.), *First Language Attrition: Interdisciplinary perspectives on methodological issues* (pp. 47-59). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Pierrehumbert, J. B. (2000). What people know about sounds of language. *Studies in the Linguistic Sciences*, 29 (2) Fall, 111-120.
- Pierrehumbert, J. B. (2001). Exemplar Dynamics: Word frequency lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137-157). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, 60 (2), 320-342.
- Poplack, S., Wheeler, S. & Westwood, A. (1989). Distinguishing language contact phenomena: evidence from Finnish-English bilingualism. In K. Hyltenstam & L. Obler (Eds.), *Bilingualism across the lifespan: aspects of acquisition, maturity, and loss* (pp. 132-154). New York: Cambridge University Press.
- Saarinen, P. (to appear). *Effects of word-frequency on consonant gradation in Canadian-Finnish bilingual speech*. To be published in the 23<sup>rd</sup> North West Linguistic Conference (February 18-19, 2008) proceedings. University of Victoria, Victoria, BC, Canada.

- Salo, A. H. (1978). *The Kalevan Kansa Colonization Company: Finnish-Canadian millenarian movement in British Columbia*. Unpublished Masters thesis. University of British Columbia, Vancouver, BC, Canada.
- Sasse, H.-J. (1992). Theory of language death. In M. Brenzinger (Ed.), *Language death: Factual and Theoretical Explorations with Special Reference to East Africa* (pp. 7-30). Berlin: Mouton de Gruyter.
- Selinger, H. W. & Vago, M. (Eds.). (1991). *First language attrition*. Cambridge: Cambridge University Press.
- Schmid, M., Köpcke, B., Keijzer, M. & Weilemar, L. (Eds.). (2004). *First language attrition: Interdisciplinary perspectives on methodological issues*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Tremblay, A., Derwing, B. L. & Libben, G. (2007). *Are lexical bundles stored and processed as single units?*. Paper presented at the 23<sup>rd</sup> North West Linguistic Conference, February 18-19, 2008.
- Waas, M. (1996). *Language attrition downunder*. Frankfurt: Peter Lang.
- Wild, P. (1995). *Sointula: Island utopia*. Madeira Park: Harbour Publishing.
- Välikangas, A. (Producer). (2004). *Ulkosuomalaisen tarina: Sointula* [Television series]. Helsinki: Filmihalli Oy.
- Yağmur, K., de Bot, K. & Korzilius, H. (1999). Language attrition, language shift and ethnolinguistic vitality of Turkish in Australia. *Journal of Multilingual and Multicultural Development*, 20 (1), 51-69.
- Yağmur, K. & Akinci, M. (2003). Language use, choice, maintenance, and ethnolinguistic vitality of Turkish speakers in France: Intergenerational differences. *International Journal of Sociology of Language*, 164, 107-128.

## Appendix A

### Stimuli: Sentence List

| Random assignment | Sentence number | Sentence  | Words tested           |
|-------------------|-----------------|---|------------------------|
| 1                 | 1               | A boy likes food.   | poika/ruoasta          |
| 2                 | 40              | It's a warm evening.  | ilta                   |
| 3                 | 8               | It costs one mark. (e.g. scoop of ice-cream)  | markan                 |
| 4                 | 83              | From the rest, ... (mom makes soup). (What happens to the rest of a roasted, say, chicken?) | lopusta                |
| 5                 | 68              | A small group of friends arrives in the evening.  | joukko/illalla         |
| 6                 | 81              | I liked the end (of the movie/book).  | lopusta                |
| 7                 | 59              | In the first chapter, ... (What happens in the first chapter of your fav book?)             | luvussa                |
| 8                 | 7               | For the duration of the movie, have your phone off.   | ajan                   |
| 9                 | 17              | In this food, there is fish.  | ruoassa                |
| 10                | 15              | In Europe, prices are high. Prices are on the rise in Europe.                               | hinnat/euroopassa      |
| 11                | 20              | The president knows this place.   | presidentti/paikan     |
| 12                | 70              | I got the tickets (for the concert).  | liput (tickets)        |
| 13                | 65              | I know his nature. (character, human nature)  | luonnon                |
| 14                | 9               | The store is closed for a week.   | kauppa/viikon          |
| 15                | 57              | Stores are closed (on Sunday in Finland).   | kaupat                 |
| 16                | 3               | A boy has a hole in his shoe.   | pojalla/reika/kengassa |
| 17                | 52              | A million is a large figure/number.   | luku                   |
| 18                | 43              | The bridge is new.  | silta                  |
| 19                | 85              | The heel of the shoe is high.   | kengän                 |
| 20                | 16              | This machine is not in use.   | kaytossa               |
| 21                | 82              | In the end (of the movie) they got married. (When?)   | lopussa                |
| 22                | 18              | Grand-pa turned the music down.   | musiikin               |
| 23                | 45              | One hour is 60 minutes/enough.  | tunti                  |
| 24                | 50              | Food tastes/is good.  | ruoka                  |
| 25                | 63              | The bird has a ring on its feet.  | jalassa                |
| 26                | 61              | To finish off/conclude, he laughed aloud.   | lopuksi                |
| 27                | 29              | One ticket costs 10 marks.  | lippu (ticket)/markkaa |
| 28                | 54              | The boy drives through Europe.  | euroopan               |
| 29                | 12              | The president's mansion/castle is in Helsinki. /The president's car is black.               | presidentin            |
| 30                | 47              | The flower is red.  | kukka                  |
| 31                | 76              | The boy saw the bread the store.  | leivan/kaupassa        |
| 32                | 24              | The boy's home is in Helsinki.  | pojan/helsingissa      |
| 33                | 49              | This place is Sointula.   | paikka                 |
| 34                | 14              | At the end of this week I am busy. There's no more pulla.                                   | viikon/lopussa         |
| 35                | 33              | Finland's capital ( ) is a nice city.   | helsinki               |
| 36                | 21              | At this place/location there was a bank.  | pankki/paikalla        |
| 37                | 80              | Mom fixed the hole (in my socks) in an hour.  | reian/tunnissa         |
| 38                | 55              | The school is behind the bank.  | pankin                 |
| 39                | 51              | My shoe is small.   | kenka                  |
| 40                | 75              | Pick me up from my place.   | paikalta               |
| 41                | 72              | In that event/situation, the use of the elevator is dangerous.                              | tilanteessa/kaytto     |
| 42                | 25              | Smoking is a bad habit. / That's bad manners.   | tapa                   |
| 43                | 32              | I want a Finnish flag.  | lipun (flag)           |
| 44                | 41              | The soup is/tastes good. (fishsoup)   | keitto                 |

|    |    |  |                                     |
|----|----|--|-------------------------------------|
| 45 | 77 | Thank you for this evening.  | illasta                             |
| 46 | 64 | I got/bought it for one mark.  | markalla                            |
| 47 | 2  | He asked the girl what time it is.   | tytolta                             |
| 48 | 22 | The girl knows the beginning of the story. I'm also familiar with the beginning of it. | alun                                |
| 49 | 39 | There is a bird on the roof of the church  | kirkon/lintu                        |
| 50 | 42 | The bread is/tastes good.  | leipa                               |
| 51 | 46 | His salary is good./He has a good salary. (a pay day)                                  | palkka                              |
| 52 | 78 | In the bread, there are seeds.   | leivassa                            |
| 53 | 5  | She was in Europe once.  | euroopassa/kerran                   |
| 54 | 66 | Mark is no longer in use in Finland.   | markka/kaytossa                     |
| 55 | 26 | I don't like that habit. (talking about smoking)                                       | tavasta                             |
| 56 | 23 | This time is good. (an appointment, say 6 p.m.)  | aika                                |
| 57 | 69 | A bird likes bread.  | leivasta                            |
| 58 | 35 | Evenings are long already.   | illat                               |
| 59 | 58 | The boy gave the flowers to the girl.  | poika/kukan/tytölle                 |
| 60 | 53 | A week is seven days.  | viikko                              |
| 61 | 4  | I know the boy.  | pojan                               |
| 62 | 48 | My foot is sore.   | jalka                               |
| 63 | 38 | I know that church./ I saw the church.   | kirkon                              |
| 64 | 44 | The river runs under the bridge.   | sillan                              |
| 65 | 31 | The Finnish flag is blue and white.  | lippu (flag)                        |
| 66 | 34 | I know that bird.  | linnun                              |
| 67 | 56 | We talked about this week.   | viikosta                            |
| 68 | 13 | The streets of Helsinki are busy.  | helsingin                           |
| 69 | 27 | He likes his salary.   | palkasta                            |
| 70 | 36 | He goes over the bridge.   | sillan                              |
| 71 | 62 | She mingled/walked among the guests.   | joukossa                            |
| 72 | 71 | He is a politics's expert.   | politiikan                          |
| 73 | 73 | That situation was dangerous./ That was a dangerous situation.                         | tilanne                             |
| 74 | 74 | He closed the store.   | kaupan                              |
| 75 | 37 | The movie has a good ending.   | loppu                               |
| 76 | 11 | A teacher of English teaches English. He taught English to the students for an hour.   | englannin/tunnin                    |
| 77 | 6  | This week I am in Helsinki.  | viikolla/helsingissa                |
| 78 | 79 | There is a fly in the soup. (Where is the fly?)  | keitossa                            |
| 79 | 30 | The price of a ticket is 10 marks.   | lipun (of the ticket)/hinta/markkaa |
| 80 | 67 | I saw the numbers.   | luvut                               |

## Appendix B

### The Finnish Language in Sointula Background Questionnaire

In this questionnaire I want to get an impression of the different uses and functions of the Finnish language in daily life of second and later generation Finns in Sointula, BC. There are no right or wrong answers. The questionnaire consists of 32 questions.

1. What is your age? .....
2. Female  Male
3. Where were you born (town, province/state, country)? .....
4. If born outside Sointula, when did you come to Sointula? .....
5. What part(s) of Finland is your ancestry from? .....  
.....
6. How many generations ago did your ancestors move away from Finland? Who were they? .....  
.....  
.....
7. Have you visited Finland?  
Yes  No  If no, go to question 10.
8. If yes, how many times? .....
9. If yes, approximately how long was/were your visit(s)? .....
10. What is your occupation? .....
11. As a child, what language did you use mostly at home when talking with your parents?  
Finnish  English  Mixed   
Other  What language? .....
12. How would you rate your proficiency in Finnish before you moved away from home?  
very good  good   
quite good  rather poor   
poor  very poor
13. Has your use of Finnish decreased, remained the same, or has it increased since you moved away from home? When did it happen (e.g. went to English speaking school)? Why did it happen?  
  
Finnish use has decreased   
Since when? .....  
Why? .....  
  
Finnish use has remained the same

Finnish use has increased

Since when? .....

Why? .....

14. Do you think that your Finnish proficiency has changed since you moved away from home?  
 very much  much   
 pretty much  a little   
 very little  not at all
15. What is your marital status?  
 married/common-law  widowed  divorced   
 single  If single, go to question 17.
16. What language was your (ex)partner brought up with? .....
17. What is/was your (ex)partner's occupation? .....
18. What language do/did you use mostly when talking to your (ex)partner?  
 Finnish  English  Mixed   
 Other  What language? .....
19. Do you have children?  
 Yes  No  If you have no children, go to question 26.
20. If yes, what language do you use mostly when talking to your children  
 Finnish  English  Mixed   
 Other  What language? .....
21. Which language do your children use mostly when talking to you?  
 Finnish  English  Mixed   
 Other  What language? .....
22. Which language do your children use mostly when talking to each other?  
 Finnish  English  Mixed   
 Other  What language? .....
23. Do you encourage your children to speak and write in Finnish?  
 very often  often   
 quite often  sometimes   
 rarely  never
24. Do you correct your children's Finnish at home?  
 regularly  seldom  never
25. Which language do you use mostly when talking with your grandchildren?  
 Finnish  English  Mixed   
 Other  What language? .....
26. Do you think Finnish plays a vital role in the cohesion between the members of your family?  
 very much so  to a considerable degree

somewhat  not at all

27. In which situations **outside your family** do you use Finnish?

|   | always | often | sometimes | seldom | never | not applicable |
|---|--------|-------|-----------|--------|-------|----------------|
| friends/neighbors                       |        |       |           |        |       |                |
| workplace                               |        |       |           |        |       |                |
| school                                  |        |       |           |        |       |                |
| church                                  |        |       |           |        |       |                |
| clubs                                   |        |       |           |        |       |                |
| shops                                   |        |       |           |        |       |                |
| in letters to relatives in Finland      |        |       |           |        |       |                |
| in letters to relatives outside Finland |        |       |           |        |       |                |
| to pets                                 |        |       |           |        |       |                |
| other, namely<br>-----                  |        |       |           |        |       |                |

5. Do you consider it important to maintain the Finnish language for yourself?

very important  important   
 rather important  not very important   
 hardly important  unimportant

6. Do you watch Finnish videos or TV?

very often  often   
 sometimes  not often   
 hardly ever  never

7. Do you listen to Finnish radio programs?

very often  often   
 sometimes  not often   
 hardly ever  never

8. Do you read Finnish newspapers, magazines, or books?

very often  often   
 sometimes  not often   
 hardly ever  never

9. How would you rate your knowledge of the Finnish language?

very good  good

rather good  rather poor

poor  very poor

Are there any relevant remarks you would like to add? Please use the space below to do so.

**Thank you kindly. Your contribution is greatly appreciated.**

## Appendix C

### Glossary of terminology

|  |   |
|--|---|
| Base (or plain NOM.sg stem)            | A free standing simple word in NOM.sg case as listed in the corpus, e.g. <i>ranta</i> ‘beach’. It does not contain gradation or suffixation. In suffixed word-forms they are gradated and thus are called gradated stems.   |
| Gradated stem                          | In a suffixed word-form, such as <i>ranna</i> — meaning ‘beach’. It cannot stand alone as a word but needs a suffix to make a word.   |
| Loss of gradation                      | Loss of gradation can occur with or without suffixation: the non-standard use of the NOM.sg stem either as an independent word-form or as a non-gradated stem when the gradated stem is called for.   |
| Phonetic output                        | The production of speech by the participants. In this study it refers to the dependent variables <i>Response</i> , <i>Gradation</i> and <i>Suffix-accuracy</i> .  |
| Whole-word-accuracy                    | A dependent variable. A whole word-form answer given by a participant that corresponds to a targeted word-form, either a suffixed or plain NOM.sg stem word-forms. Evaluated in terms of accuracy for the sentence context.   |
| Stem-frequency                         | The frequency of the gradated stem as a NOM.sg case base. E.g., <i>markalla</i> ‘for a mark’ contains the base <i>markka</i> (rank: 3099.) It has two levels, frequent and infrequent.  |
| Suffixed word-form (Complex word-form) | A word-form consisting of many parts, e.g. <i>rannalla</i> ‘on the beach’. Contains a gradated stem <i>ranna</i> — ‘beach’ and the suffix <i>-lla</i> meaning ‘on’. In English, e.g. <i>knives</i> , containing an allomorph of [ <i>naiv</i> ] and the plural suffix [ <i>z</i> ]. In this study, they are always gradated and suffixed. |

|                                       |   |
|---------------------------------------|---|
|                                       |   |
| Suffix-frequency                      | The frequency of the suffix in a complex word-form as listed in Hakulinen et al. (2004). It has two levels, frequent and infrequent.  |
|                                       |   |
| Suffix-accuracy                       | A suffix-response produced by a participant. A dependent variable measuring correctness of suffix.  |
|                                       |   |
| Targeted word-form (whole)            | A whole word-form sought after in this research and elicited from the participants. They are listed and ranked in the corpus. Either complex word-forms or plain NOM.sg stem word-forms, e.g. <i>rannalla</i> 'on the beach' or <i>ranta</i> 'beach'.                             |
|                                       |   |
| Token-frequency                       | The frequency at which a word-token is encountered in a language. E.g. <i>rannalla</i> is encountered 590 times in every million words. It ranks as the 3134. most common word in Finnish.  |
|                                       |   |
| Type-frequency                        | The frequency at which a pattern is encountered in a language (or the number of lexical items that participate in the process). Examples include the English –ed past tense as the most frequent past tense type in English or that most words in Finnish represent the NOM case. |
|                                       |   |
| Whole word-form                       | Complex or plain NOM.sg stem word-forms; targeted word-forms.   |
|                                       |   |
| Word-frequency (whole word-frequency) | An independent variable. Frequency ranking of targeted word-forms, both suffixed and plain NOM.sg stem word-forms. Two levels, frequent and infrequent.   |

**Appendix D**  
**The breakdown of suffixes in the stimuli**

|                            |                      | <i>Number of<br/>Tokens (N)</i> | <i>Total (x6<br/>participants)</i> |
|----------------------------|----------------------|---------------------------------|------------------------------------|
| <b>Frequent suffixes</b>   | GEN                  | 34                              | 204                                |
|                            | Quantifier           | 29                              | 174                                |
|                            | Object               | 4                               | 24                                 |
|                            | Object of adposition | 12                              | 72                                 |
|                            | Possessive           | 5                               | 30                                 |
| <b>Infrequent suffixes</b> | NOM.pl               | 8                               | 48                                 |
|                            | INE                  | 5                               | 30                                 |
|                            | ADE                  | 31                              | 186                                |
|                            | TRANS                | 13                              | 78                                 |
|                            | ALL                  | 5                               | 30                                 |
|                            | ELAT                 | 2                               | 12                                 |
|                            | ABL                  | 1                               | 6                                  |

## Appendix E Consent script

For recruiting potential participants by phone for the actual study in Sointula

My name is Pauliina Saarinen. I am doing a research study at the University of Victoria on the Finnish language in Sointula. The topic interests me very much because I am a native of Finland. My acquaintances in Sointula, BC provided me with your name as someone who might be interested in participating in my study. I am looking for a few Sointula residents with Finnish heritage for my study. I contact you because you are a second or later generation Finn and speak Finnish. I ask you to participate in my study.

I am asking you to take part in a research study because I am trying to learn more about variation in Finnish word forms as spoken in Sointula. Your participation is voluntary, and you can stop being part of it at any time without any consequences to you. The research session takes place in Sointula in March 18 – 23, 2008 and could be scheduled to take place at your convenience within this time period where ever you feel the most comfortable, at your home, for instance. Your participation would involve translating English sentences to Finnish, which are audio-recorded, and completing a survey of your language background, and will also involve going over the informed consent form, as part of the university's ethical review process. If you agree to participate, the whole research session would take about 2 hours. To fight fatigue, we can take breaks in the middle. There are no risks involved. You will not be paid for participating in this research study.

The data, the audio-recordings of the translations and the questionnaire information, will be used to write my Masters thesis. I will give copies of the audio-recordings to the Sointula Museum if you so wish. Similarly after having finished the thesis, I will give a copy of it to the museum. I may also present the findings in presentations at scholarly meetings and/or in scholarly publications and use the same data for another research project in the future.

If you have any questions or concerns about the research, please feel free to contact me at (250) 216-1617. You can also reach my supervisor Dr. Suzanne Urbanczyk at University of Victoria at (250) 721-7431 or [urbansu@uvic.ca](mailto:urbansu@uvic.ca). If you have questions regarding your rights as a research subject, contact the Human Research Ethics at the University of Victoria at (250) 472-4545 or [ethics@uvic.ca](mailto:ethics@uvic.ca).

Date of Preparation: January 24, 2008 – Verbal Consent Script

## **Appendix F Participant Consent Form**

You are being invited to participate in a study entitled **The Finnish language in Sointula** that is being conducted by me, Pauliina Saarinen.

I am a graduate student in the department of Linguistics at the University of Victoria and you may contact me if you have further questions by phoning (250) 216-1617 or by emailing [pauliina@uvic.ca](mailto:pauliina@uvic.ca).

As a graduate student, I am required to conduct research as part of the requirements for a Masters degree in Linguistics. It is being conducted under the supervision of Dr. Suzanne Urbanczyk. You may contact my supervisor at (250) 721-7431.

The purpose of this research project is to investigate variation in Finnish word forms as spoken in Sointula. Research of this type is important because the results will advance the state of knowledge of linguistic theory in general by providing evidence on the validity of competing models of the theory of grammar. Similarly, Sointula offers a rare context in which the preservation of grammatical features of Finnish can be studied. Research findings interest linguists working on other languages related to Finnish. Languages, such as North Sámi, Votic and Karelian, make extensive use the same grammar and are all on the verge of extinction. The present study also adds to the corpus of knowledge on the preservation and loss of distinctive aspects of grammar of threatened languages.

You are being asked to participate in this study because you are a second or later generation Finnish speaker living in Sointula, BC.

If you agree to participate in this research, your participation will include two different parts. In the first part, you would be translating English sentences to Finnish, which will be audio-recorded. For example, you would be asked to translate sentences such as “The boy saw the girl” or “The boy saw apples in the store” in Finnish. There will be 85 sentences. The second part involves filling out a small questionnaire, asking questions about your language background and language use. The whole research session will take about 2 hours of your time.

Participation in this study may cause some inconvenience to you, including committing 2 hours of your time for the research session. There is a risk that you might also feel uncomfortable sharing your language with me and/or that you might become fatigued during the experiment. To avoid causing you emotional discomfort, I would ease you into the translation task gradually by practicing with few example sentences. To avoid fatigue,

having breaks during the experiment prevent you from becoming fatigued. Feel free to ask for a break at any time during the research session.

Your voluntary participation would help find out how word-forms have preserved in Sointula's Finnish and how humans store words in their memory. The findings will also advance the state of linguistic knowledge in general. Moreover, you would be able to use your Finnish speaking skills, for which opportunities in Sointula are diminishing. You are not compensated for your participation.

Your participation in this research must be completely voluntary. If you do decide to participate, you may withdraw at any time without any consequences or any explanation.

If you do withdraw from the study, your data will not be used for the analysis.

In terms of protecting your anonymity, you are not asked to disclose your identity for the purposes of this study. Your confidentiality and the confidentiality of the data will be protected by coding your data, so that your real identity remains unknown. Data from this study will be stored in password-secured computer files indefinitely.

I will use the findings of this research to write my Masters thesis. It is also anticipated that the results of this study will be shared with others in presentations at scholarly meetings and/or in scholarly publications. I will give copies of the recordings to the Sointula Museum if you so wish. Similarly, after having finished the thesis, I will give a copy of it to the museum.

If you have any questions or concerns about the research, please feel free to contact me ((250) 216-1617). You can also reach my supervisor Dr. Suzanne Urbanczyk at University of Victoria ((250) 721-7431 or [urbansu@uvic.ca](mailto:urbansu@uvic.ca)). If you have questions regarding your rights as a research subject or if want to verify the ethical approval of this study, contact the Human Research Ethics at the University of Victoria at (250) 472-4545 or [ethics@uvic.ca](mailto:ethics@uvic.ca). In addition, you may raise any concerns you might have, by contacting the Associate Vice-President, Research at the University of Victoria (250-472-4545).

Do you want your audio-recordings to be given to the Sointula Museum? \_\_\_\_\_

Yes. \_\_\_\_\_ No.

Your signature below indicates that you understand the above conditions of participation in this study and that you have had the opportunity to have your questions answered by the researchers.

---

*Name of Participant*

---

*Signature*

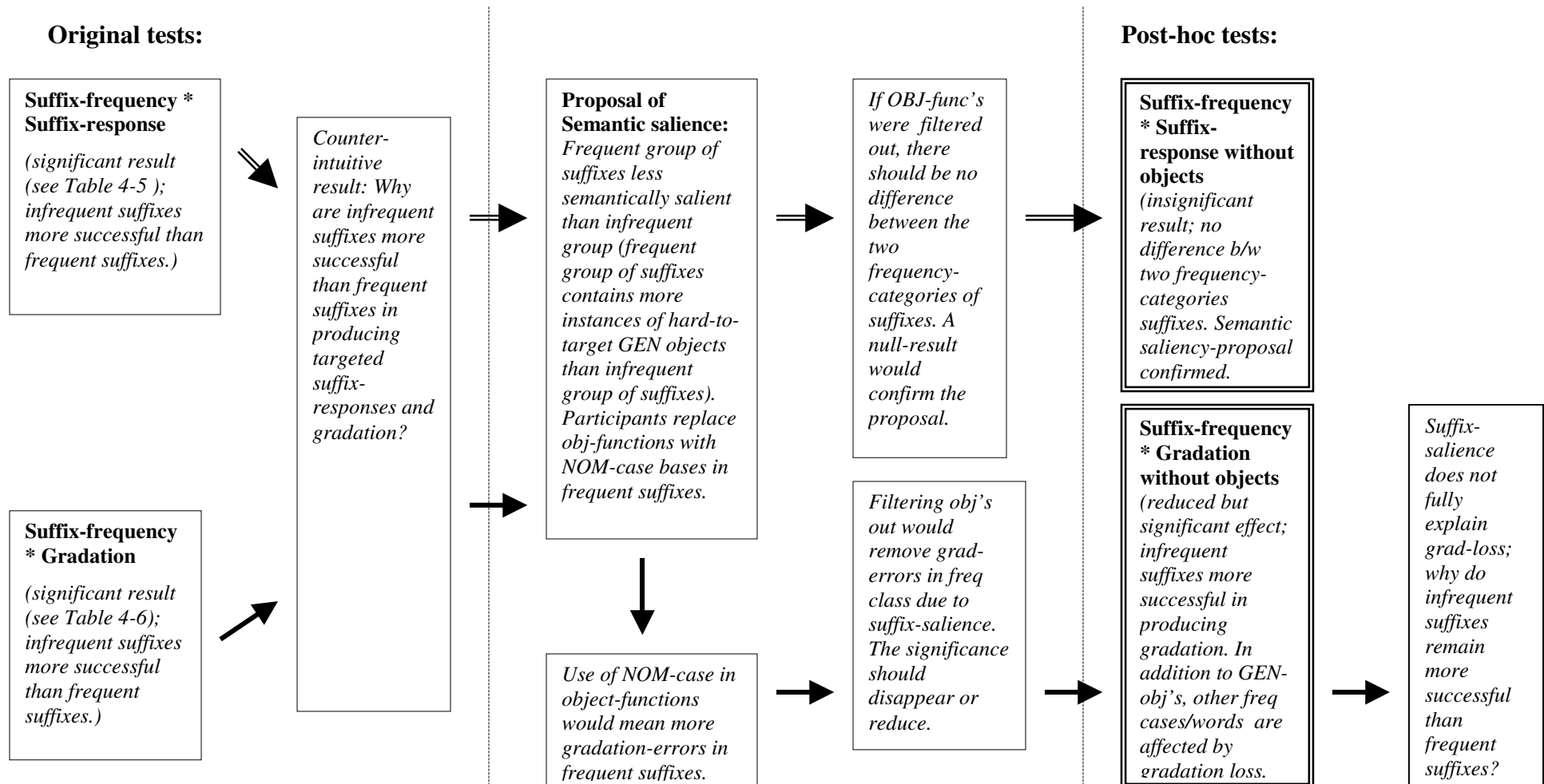
---

*Date*

***A copy of this consent will be left with you, and a copy will be taken by the researcher.***

## Appendix G Flow-chart for Suffix-frequency tests

Flow-chart for the Suffix-frequency tests discussed in Section 5.1.2  
(For Suffix-accuracy, follow  $\Rightarrow$  and Gradation-accuracy  $\rightarrow$ .)



Examination of gradation-performance and its connection with the atrophy of GEN object marker.

