CODEN: UM/PHON/R-05/0010

Reports in Phonetics
Umeå University

September 2005

10

Phonological Quantity in Swedish Dialects

Typological Aspects, Phonetic Variation and Diachronic Change

Felix Schaeffler

PHONUM

Felix Schaeffler: Phonological Quantity in Swedish Dialects - Typological Aspects, Phonetic Variation and Diachronic Change

PHONUM 10
Department of Philosophy and Linguistics
Umeå University
SE-901 87 Umeå, Sweden

ISBN 91-7305-935-8
ISSN 1101-2714

Cover by Alexander Pankow
Printed by Print & Media, Umeå 2005

PHONUM is distributed on the basis of mutual exchange and is also available via the Department's Internet pages: http://www.ling.umu.se/

Abstract

This study investigates the realisation of phonological quantity in the dialects of Modern Swedish, based on a corpus containing recordings from 86 locations in Sweden and the Swedish-speaking parts of Finland. The corpus was recorded as part of the national SweDia project.

The study is explorative in character. Quantity structures in Swedish dialects and their geographical distribution, as described in the dialectological literature, are compared to the results of a data-driven categorisation (cluster analysis). The results reveal an overall good correspondence of the data driven and the traditional categorisation, although with some deviations in the detail.

The study is divided into two parts. The first part lays the foundation for the data-driven categorisation, which is then described in the second part. First, the phonology and phonetics of quantity in Swedish are described in terms of durational distinctions and vocalic quality differences that typically accompany the durational differences. Preaspiration, which appears to be a normative feature in some dialects, is covered as well. An overview of the historical development of the Swedish quantity system is provided, with special emphasis on a phonological interpretation of quantity changes. Thereafter, dialectological evidence is combined with phonological and typological considerations to develop a categorisation of Swedish dialects.

The second part explains the methodology of cluster analysis and applies this method to vowel and consonant durations from one contrastive word pair, in order to obtain an alternative dialect categorisation. Analyses of vowel quality and preaspiration are performed in addition to the durational analyses. Hypotheses derived from the cluster analysis are then tested on one additional word pair recorded in 75 locations and on three additional word pairs recorded in four locations.

The general pattern emerging from the cluster analysis is a categorisation of the dialects into three main types, a Finland-Swedish, a Northern and a Southern type. This categorisation shows a good geographical agreement with the categorisation that is derived from the analysis of the dialectological literature. Therefore, the durational patterns of the three types are interpreted as reflections of three different phonological systems: 4-way systems with vocalic and consonantal quantity, 3-way systems with vocalic quantity and with consonantal quantity only after short vowels, and 2-way systems with complementary quantity. From the historical perspective, the 4-way system constitutes the most conservative and the 2-way system the most recently developed system.

Finally, it is argued that the historical development is one of the factors behind occasional mismatches between the data-driven and the dialectological categorisation. Data from one of the dialects, which has recently abandoned a 4-way system but has obviously retained the durational properties of the older system, is used as an example to illustrate this historical hypothesis.

Acknowledgements

I would like to thank all the people who helped to make this thesis happen. First of all I would like to thank my main supervisor Eva Strangert for inspiring discussions that clarified my thinking, for her support of my ideas, for her thoroughness and for endless help during the final work-marathon. I am also grateful to my second supervisor Anders Eriksson for useful advice, inspiration and his help and critique in the final stages.

Then, there's my fellow doctorand Fredrik Karlsson. Thank you for interesting discussions and endless help with LaTeX, R and Linux. You kept me several times from throwing my computer out of the window. I would like to thank Kirk Sullivan for lots of advice in practical things and creative coffee breaks, joined by Fredrik (see above) and Erik Eriksson whom I'd like to thank for many intriguing conversations and bracing irony.

Furthermore, I'd like to thank all colleagues at the department for their support, and for the warm welcome at the department. A special thanks to Pär Wretling, who helped me to settle in and was a great colleague.

I am very grateful to The Bank of Sweden Tercentenary Foundation and the Department of Philosophy and Linguistics for financing my life during my four years in Umeå.

I am also much obliged to all my current and former colleagues from the SweDia project. I'd like to thank even those I don't know but who criss-crossed Sweden on the hunt for speech samples for the SweDia corpus. Naturally, I am also very thankful to all those helpful people who participated in the recordings.

Furthermore I'd like to thank Jan van Doorn, Jenny Sullivan Hellgren and Luke van Doorn for their support, encouragement and Swedish and Australian treats.

And, finally, I'd like to thank you, Sonja, for everything! Without you, there is nothing.

Contents

List of Figures

List of Tables

Part I

Background

Chapter 1

Introduction

1.1 The Aims and Structure of the Thesis

This work studies phonological quantity — that is the phonological use of sound duration — in the dialects of modern Swedish. The study has two parts, a theoretical one and an empirical one. The theoretical part provides an overview over general aspects of phonological quantity, discusses the historical background of phonological quantity in Swedish and describes the different quantity patterns found in the Swedish dialects today, based on dialectological descriptions found in the literature. The empirical part analyses a corpus of recordings of quantity minimal pairs from 86 different locations in Sweden and the Swedish speaking parts of Finland. The first, explorative, step analyses durational and qualitative relations between the different sounds in question and shows that a data-driven approach results in a categorisation which is similar to that proposed in the traditional accounts.

Important findings of the explorative study are then tested on other parts of the corpus in the second step of the empirical part. The final chapter gives an overview over the findings, discusses typological aspects and potential connections to historical changes. Phonological quantity in Swedish (and other Germanic languages) has been extensively discussed in a diachronic perspective (e.g. Riad, 1992). Therefore, an exclusively synchronous perspective on quantity in Swedish dialects would leave out important aspects. Dialectological descriptions of Swedish dialects usually apply a historical perspective, treating dialects with a rich quantity system as conservative dialects that have kept parts of an ancient system lost elsewhere. Therefore, this work does also consider diachronic aspects. The phonetic patterns or "types" which are found in the data base under study will be compared to the historical analyses and descriptions found in the various dialectological and historical linguistics sources.

The remaining part of Chapter 1 discusses terminological issues and describes notational conventions used in the text. Chapter 2 will focus on phonological descriptions of quantity in Swedish and phonetic aspects of quantity, that is the acoustics and perception of durational contrasts. With respect to vowel quantity, the feature of "tenseness" has extensively been discussed in the literature. These accounts are discussed separately in Chapter 3. Chapter 4 will take up the historical aspects of quantity in Swedish. Chapter 5 describes the different types of quantity systems found in the Swedish dialects. Chapter 6 will present accounts of quantity typologies and typological observations regarding vocalic and consonantal length.

Third, it will use the findings of the first part for the formulation of a preliminary typology of the Swedish dialects.

The second part of the study consists of a series of investigations of a large corpus of quantity data from Swedish dialects, which was recorded within the SweDia project (see e.g. Bruce et al., 1999) between 1998 and 2000. This part has an explorative orientation. It applies a data-driven classification method to the SweDia data. The resulting classification is then compared to the typology which was worked out in the first part. The aims and structure of the second part are described in more detail in Section 7.1.

1.2 Terminology and Conventions

1.2.1 Quantity

In the phonological literature, the term "quantity" is used in at least two senses. The first use refers to phonological contrasts between two sounds which are expressed by differences in duration. If the term "quantity" or "segmental quantity" is used in the present study, it refers to this meaning. The term "length" is used as well, as a synonym for "quantity". Both these terms are kept apart from "duration", which is reserved for a phonetic use, that is when referring to the measured or perceived duration of speech sounds.

The second use of the term "quantity" refers to the weight of syllables. For example, phonological theory describes the stress system of many languages as "quantity sensitive". This means that syllables attract stress depending on their weight. The weight of a syllable is in many cases determined by the type of segments it is composed of, and the segmental quantity of certain segments contributes to this weight, but it is often not the only factor influencing the weight. This use of the term quantity with respect to the weight of syllables will be specified as "syllabic quantity" or "prosodic quantity" in the present study, if the meaning is not apparent from the context.

1.2.2 Tonal accents

Swedish is a language with tonal accents. Stressed syllables in words with more than one syllable can be associated with two different tonal patterns. The precise realisation of the tonal patterns varies across dialects (see Bruce and Gårding, 1978). In Standard Swedish, the "acute accent", or "accent I", has a low tone associated with the stressed syllable, while the "grave accent", or "accent II", has a high tone. In focussed position, both accents show an additional high tone, which results in two tonal peaks for accent II and one late tonal peak for accent I. A detailed account of the tonal patterns in Swedish can be found in Bruce (1977). If reference is made to the tonal accents in the present study, the terms "accent I" and "accent I" will be used.

1.2.3 The use of the term "dialect"

This thesis uses the term "dialect" for the varieties of modern Swedish under study. This stresses the focus on geographical variation of the Swedish language, not on other types of variation, for example social variation, although differences between age and gender groups will be referred to whenever of importance. In dialectology, but also in everyday language, the term "dialect" is often reserved for the most "genuine" variety at a certain geographical location. It has to be stressed that this is not necessarily the denotation of the term in the present study. Within the SweDia project, extreme care was taken to encourage the participants to use the same register as in conversations with locals. While this does not guarantee that participants use varieties which are identical in every detail, it should ensure that their speech is sufficiently representative of the locally used varieties. The term dialect, as it is used here, thus refers to anyone of these commonly used local varieties.

For practical reasons, comparisons are sometimes made to Standard Swedish, often represented by the variety of Central Standard Swedish, as spoken in the Stockholm region. In the present study, this term refers to the form of Swedish described in dictionaries and grammars, but also to material from Stockholm speakers.

1.3 Conventions

Frequently in the present study, reference is made to geographical locations in Sweden and Finland (e.g. names of villages or regions). In all cases, the original Swedish names for these locations have been used, even if English translations exist.

Phonological and phonetic notation follows the usual conventions. The categories phonologically “long vowel”, “short vowel”, “long consonant” and “short consonant” are denoted by V:, V, C: and C.

Chapter 2

Phonological and Phonetic Aspects of Quantity

The first section of this chapter will describe how quantity has been analysed in Standard Swedish. In the second section, phonological models of quantity across languages will be discussed. Such models incorporate data from different languages and are therefore important prerequisites for a discussion of typological aspects of quantity, which are addressed in more detail in Chapter 6. In the following sections, acoustic aspects of quantity distinctions are addressed, as well as the perception of quantity.

2.1 Quantity in Swedish

Standard Swedish is normally described as a language with "complementary quantity". Elert (1964) describes this in the following way:

> "*The main characteristic of quantity in isolated words in Swedish, according to the conventional interpretation, is that either the vowel or the immediately following consonant in syllables with main stress or secondary stress is long.*" Elert (1964), p. 39

That is to say, in stressed position there exist long and short vowels as well as long and short consonants, but they are distributed in a complementary relationship: If a long vowel is followed by a consonant, the consonant is short, while a short vowel is followed by a long consonant or a consonant cluster. Obviously, there is an asymmetry in the distribution of long and short vowels. The minimal word in Swedish consists of either a long vowel or a short vowel plus a long consonant. At least in mono-syllabic words, there are thus no short vowels in open stressed syllables. With two syllables or more, the situation is slightly more complicated as the definition of an open syllable depends on the definition of the syllabic border. Conventionally, the long consonants are treated as ambi-syllabic, and the phonological structure of a word like *penna* ('pencil') is thus assumed to be <u>CVC</u>.CV, with stress on the first syllable and a syllabic border within the geminate. If short vowels are followed by a consonant cluster, at least one consonant is always analysed as belonging to the coda of the first syllable. If these preconditions are accepted, then there are, as a consequence, no short vowels in open stressed syllables in Standard Swedish. This leads furthermore to the assumption that all stressed syllables in Swedish have the same weight: they are always "heavy". Syllable weight is often expressed in terms of the mora. A light syllable has one mora and a heavy one has two. Long vowels are generally seen as bi-moraic, and short vowels as mono-moraic (cf. for example Riad, 1992). In the case of stressed syllables with short vowels, the second mora is thus contributed by the following consonant. This moraic theory will play an important role for the historical discussion in Chapter 4, where its consequences will be outlined more thoroughly.

Standard Swedish is not the only language described with complementary quantity. Other Germanic languages include Norwegian, Faroese and Icelandic, and similar phonological issues have been discussed for these languages (see e.g. Árnason, 1980). Complementary quantity has also been described for a German dialect (Bannert, 1976) and for the Scots dialect on the Shetland Islands (van Leyden, 2002; van Leyden, 2004).

The complementary relationship between vowels and consonants in Swedish may be given several different phonological interpretations. Elert (1964), see pp. 39–43, provides a summary of the different positions. He distinguishes at least 11 different positions, and uses two dimensions to describe them. The first dimension is based on the position of the distinctive element: in the vowel, in the consonant, or in both segments. The second dimension distinguishes different ways of expressing the quantity contrast: the contrast can be based on duration, gemination (duplication) of a segment, diphthongisation, accent placement or vowel quality. According to Elert, the most common interpretation of complementary quantity is a length mark for both segments (/viːt/ – /vitː/). Morphological arguments have led to representations that mark the distinction only as a gemination of the consonant (/vit/ – /vitt/). This also corresponds to the way in which quantity is represented in Swedish orthography. An additional possibility is the diphthongisation of the vowel (/vijt/ – /vit/), but Elert stresses that this solution has no phonetic basis, as the diphthongisation or glide in the second half occurs in Swedish only with high or mid-high vowels. Some authors, and even Swedish handbooks and dictionaries use a notation where the position of the stress marker indicates the length pattern (/viˈt/ – /vitˈ/). Elert finds no phonetic or structural motivation for this representation. However, a possibility that he considers adequate, at least in terms of economy, is a representation expressing the distinction only by vowel quantity (/viːt/ – /vit/):

> *"The functional load borne by the length of consonants is very small, and the number of unpredictable features in Swedish speech attributable to consonant length is so inconsiderable that it should be possible for practical purposes . . . to count with vowel length only."*Elert (1964), p. 42

The fact that the long and short vowels differ in timbre has been used in purely phonemic interpretations. In this type of analysis there is no need for a separate quantity prosodeme. Elert (1964) rejects this possibility, however, since it would lead to an unusually large number of phonemes in comparison to other languages. (see pp. 42–43). Choosing to treat long and short vowels as different phonemes constitutes a lower level of abstraction, according to Elert, and *". . . is justifiable for defining categories for limited practical phonetic tasks"* (p. 43). For a phonological analysis, however, Elert rejects this solution. Instead, he treats quantity as a "prosodeme". Following this structural interpretation, long and short vowels are allophones of a single vowel phoneme that can be combined with the prosodeme "length". This analysis is motivated by the pair-wise arrangement of long and short vowels in the Swedish grammar.

A generative analysis of Swedish quantity has been presented by Eliasson and La Pelle (1973). In their paper, the authors argue for underlying consonant quantity in Swedish. Another generative account of Swedish quantity, which also argues for underlying consonant quantity (Riad, 1992), will be discussed in detail in Chapter 4. Generative accounts will, therefore, not be considered here.

2.2 Phonological Quantity across Languages

While complementary quantity seems to be a rather rare feature across the world's languages, phonological distinctions based on segment duration are quite common (see also Chapter 6). Trubetzkoy (1939) analysed a wide range of languages and derived several types of quantity languages. The first type includes languages where long syllable nuclei should be interpreted as geminates, i.e. as a succession of two short segments. He states five criteria for those languages:

1. A morphemic border can occur within the lengthened segment. In Finnish, for example, suffigation can lead to long vowels.

2. Long nuclei behave like diphthongs. As an example, Slovak is cited, where long nuclei directly after long nuclei are shortened. Long vowels behave like diphthong nuclei in these cases.

3. Rhythmical effects at the word level show that long nuclei equal two short nuclei. As an example the word stress rule of Latin is given. Word stress falls on the heavy penultima, or on the antepenultima, if the penultima is light. According to Trubetzkoy, the rule is that stress falls on the second to last mora before the last syllable.

4. Long nuclei show two different types of accent. Example languages comprise Japanese, Slovenian and Lithuanian.

5. Languages show a "push" phenomenon (Danish "stød") that splits a long vowel into two parts. Example languages are Danish and Latvian.

All languages in this group are called mora-counting, as the syllable is not the smallest prosodic unit. In contrast to mora-counting languages, syllable-counting languages have long nuclei that have to be interpreted as separate units and not as a combination of two short nuclei. Syllable counting languages show the following criteria:

1. Syllabic nuclei are always mono-phonemic. Example languages are Hungarian and some dialects of Czech.

2. Long segments do not behave like polyphonematic diphthongs. An example language is Standard Czech.

Quantity oppositions are according to Trubetzkoy (1939) always "privative". That is to say, the opposition is characterised through the presence versus the absence of a feature. In this sense, one of the members of the opposition is always "marked". In the mora-counting languages, the long segments are the marked ("feature carrying") segments. In the syllable-counting languages, either the long or the short segments can be the marked member of the opposition. Trubetzkoy (1939) cites German, Dutch and English as examples. In these languages, the quantity distinction is neutralised in stressed open final syllables. Here, only long vowels are structurally permitted. The duration difference between the vowels in these languages is dependent of the "syllable cut correlation" (see below).

Trubetzkoy's analysis points towards several typologically relevant aspects of quantity. At the structural level, the relation between long and short segments can be different from language to language, and presumably also from dialect to dialect. It is important to note that Trubetzkoy assumes differences in terms of the markedness of long and short segments from language to language. According to his analysis, the long segment is the marked member of the opposition in some quantity languages, as it consists of two short segments. In other languages, the short segment is the marked member of the opposition, as can for example be derived from a more restricted distribution of the short segment. Trubetzkoy (1939) does not recognize "complementary quantity" as a separate category.

If Standard Swedish should be classified according to Trubetzkoy's categories, the result is not totally unequivocal. In some paradigms in Swedish, for example the indication of the neuter gender in adjektives, a /t/ is added to the word stem. In the case of words ending in /t/, the result is generally interpreted as a /t:/, resulting in a geminate that occurs within the segment. According to Trubetzkoy's system, this would identify Swedish as a mora-counting language where long segments are composed of two short segments. This, however, applies only to the consonants. There are no grammatical paradigms resulting in poly-morphemic long vowels in Standard Swedish. Furthermore, Swedish shares the neutralisation of the quantity distinction in stressed open final syllables with other Germanic languages. Thus, Swedish shows similarities with "syllable cut" languages.

2.3 Quantity and the Syllable-Cut Theory

The syllable cut is a concept that dates back to Sievers (1893). The general idea behind the syllable cut is a difference in the vowel to consonant transition. Under "abrupt cut", the vowel is cut off by the following consonant while it still has maximal energy. Under "smooth cut" the vowel energy has decayed before the consonant sets in (see e.g. Trubetzkoy, 1939, p. 196). A phonological and diachronic interpretation of the syllable cut has been performed by Vennemann (see e.g. Vennemann, 1995; Vennemann, 2000). Vennemann's work focusses on German, English and Icelandic, but because of the obvious similarities between these languages and Swedish with regard to syllable structure and the distribution of long and short vowels, his arguments are of interest here. Vennemann denies languages like German the status of a quantity language. Instead, he argues that the historical development has led Germanic languages from being quantity languages to becoming syllable cut languages.

In short, this change proceeded in the following way: Old High German was an accent based quantity language (Vennemann, 1995, p. 190). The term "quantity language" applies to languages that have a syllable weight definition that incorporates vowel length and consonant length. Syllable weight in quantity languages is defined in the following way: syllables can be short or long and light or heavy. Short syllables are open and have a short vowel. Other syllables are long. Light syllables are either short or not relevant for weight. All other syllables are heavy. The distinction between long or short and light or heavy is irrelevant for stressed syllables. For unstressed syllables, there is a difference between syllable-based and accent-based quantity languages. In syllable-based quantity languages, unstressed syllables can be heavy. In accent-based languages, unstressed syllables can be long or short, but are always light in Vennemann's system. Obviously, syllable based quantity languages are those languages that allow long–short distinctions even in unstressed positions. This would include languages like Finnish, but according to Vennemann also Latin and Ancient Greek, and probably the early stages of all Germanic languages. The Old High German and Middle High German stage was, however, representative of an accent-based quantity language: there were no length distinctions in unstressed syllables.

The change that affected Old High German and lead to the change into a syllable-cut language is the transformation of open syllables into smooth-cut syllables, and the transformation of closed syllables into abrupt-cut syllables. However, this rule describes only a tendency (it is a "preference law" in Vennemann's terminology). Whether a closed syllable becomes abruptly cut or an open syllable retains a smooth cut is also dependent on the strength of the consonant[1]. Weaker — more vowel-like — consonants are rather unsuitable for abrupt cuts. Strong consonants like voiceless plosives tend to produce abrupt cuts, even in open syllables. For example, a Middle-High German word like *gate* with a CV.CV structure became *Gatte* with an abruptly cut vowel and an ambi-syllabic consonant.

Consequently, all syllables were smoothly cut before the change, even those with a short vowel. Generally, "real" quantity languages have usually smoothly cut syllables. Vennemann (2000), p. 252, cites Czech as an example of a modern language with long and short vowels in smoothly cut syllables. After the change, some syllables have developed abrupt cut, according to the preference law and the influence of consonant strength and intrinsic vowel length. Thus, the real structural change affected those short vowels that ended up in an abruptly cut syllable. They went from smooth cut to abrupt cut. The vowels that remain smoothly cut get lengthened under accent. In their underlying form, they remain light. The evidence for this assumption comes from the behaviour of smoothly cut syllables with respect to the so-called penult rule: in German, accent can not move beyond a heavy penult.

It is obvious that the interpretation here deviates from the accounts given for Swedish (see Section 4). The lengthening process assumed for Swedish for a comparable word led to a CVC̣:V structure, that is an ambi-syllabic long consonant. As the phonological quantity structure of Middle High German is obviously identical to the structure of Swedish before the Quantity Shift, there were VC: sequences in Middle High German. With regard to the *gate* example, Vennemann states:

> *"The result of the deviant kind of development thus is the same as for geminates, which naturally produce abrupt cut and then wind up as ambi-syllabic consonants by way of de-gemination. . . "* Vennemann (2000), p. 264

If Vennemann's analysis is interpreted correctly here, it states that CV.CV evolved directly into smoothly or abruptly cut syllables with a short but ambi-syllabic consonant in the latter case. CVC̣:V structures developed first into abruptly cut syllables with a long consonant, and then the consonant was degeminated.

In any case, the outcome of the change is a phonological structure where abruptly cut syllables are underlyingly heavy, while smoothly cut syllables are underlyingly light and become heavy under accent. In Chapter 4, an account will be presented that analyses the Swedish quantity pattern from a historical perspective as well, but in a different framework.

[1] With "consonant strength", Vennemann (2000) denotes obviously a kind of inverse sonority scale.

2.4 Quantity and Duration

Extensive investigations concerning durational aspects of Standard Swedish have been performed by Elert (1964). He investigated utterances of 11 speakers of Stockholm Swedish, using material consisting of read word lists. The first list contained 118 mono- and bi-syllabic words. In the case of the bi-syllabics, words with tonal accent I as well as accent II were included to examine the influence of the tonal distinction on the durational patterns. All words contained either a V:C or a VC: sequence. The durations of the vowels and the following consonants were measured. The second list consisted of 290 sentences. Every sentence contained one target word in a fixed position (the second to last stressed word). The duration of the vowels and the consonants were measured for this list as well. The structure of the words was the same as in the word list in most of the cases.

Table 2.1: A summary of the results of Elert (1964) for the one-word list (pp. 114 and 143).

V	*C*	*V dur (ms)*	*C dur (ms)*	*V+C dur (ms)*	*V/C*	*V:/V (V/V:)*	*C:/C (C/C:)*
/iː/	t	149	233	382	0.64	1.64 (0.61)	1.15 (0.87)
/ɪ/	t	91	268	359	0.34		
/eː/	t	190	203	393	0.94	1.68 (0.59)	1.28 (0.78)
/e/	t	113	260	373	0.43		
/ɛː/	t	189	212	401	0.89	1.75 (0.57)	1.25 (0.8)
/ɛ/	t	108	265	373	0.41		
/yː/	t	153	238	391	0.64	1.7 (0.59)	1.14 (0.88)
/ʏ/	t	90	272	362	0.33		
/ʉː/	t	124	251	375	0.49	1.2 (0.83)	0.98 (1.02)
/ɵ/	t	103	245	348	0.42		
/øː/	t	169	201	370	0.84	1.69 (0.59)	1.2 (0.83)
/œ/	t	100	241	341	0.41		
/uː/	t	128	245	373	0.52	1.23 (0.81)	1.11 (0.9)
/ʊ/	t	104	271	375	0.38		
/oː/	t	170	206	376	0.83	1.65 (0.61)	1.24 (0.81)
/ɔ/	t	103	255	358	0.4		
/ɑː/	t	187	214	401	0.87	1.56 (0.64)	1.24 (0.81)
/a/	t	120	265	385	0.45		

The mean long vowel duration in Elert's material was 163 ms for the word list (n=116) and 158 ms for the sentence list (n=721). Short vowels had a duration of 108 ms (word list, n=116) and 103 ms in the sentence list (n=740). The overall V: to V ratios are thus approximately 1.5:1. As table 2.1 shows, these ratios vary across vowels. Some vowel pairs show V:/V ratios as low as 1.23. Elert gives no overall values for consonant durations. Mean C:/C ratios for a range of consonants, however, do not exceed 1.2 in the word list (see p. 142). The durational difference between the long and the short consonants is thus considerably lower than the difference between long and short vowels. This can also be derived from the ratios for /t:/ versus /t/ in 2.1. Concerning intrinsic durations, Elert's data showed a general tendency for low vowels to be longer than high vowels. Rounded versus unrounded vowels showed no significant differences in terms of duration. For the long vowels, back vowels were slightly longer than front vowels. Influences of preceding consonants on vowel duration were not tested by Elert on his material. Vowels before voiced consonants were generally longer than vowels before unvoiced consonants, although this result might have been obscured by segmentation criteria (cf. Elert, 1964, p. 133–135). Besides voicing there were effects of the manner of articulation: vowels before fricatives tended to be longer than vowels before plosives and nasals. Concerning place of articulation, vowels tended to be longest before retroflexes. Then followed, in descending order: dentals, labials and palatals/velars. These tendencies held more or less for all manners of articulation. The vowels were longer in mono-syllabic words than in di-syllabic words. This effect was about the same for long and short vowels. There was thus no effect on the V:/V-ratio in this case (cf. Elert, 1964, p. 138). In di-syllabic words, vowels in words with accent I were longer than vowels in words with accent II. Elert does not address this question directly, but his measurements indicate that the effect is about the same for long and short vowels (cf. Elert, 1964, table 5.32, p. 138). In both cases, the ratio is about 1.1:1.

Regarding the consonants, Elert's data showed longer durations for voiced than for unvoiced consonants. Effects of the place of articulation included a tendency for velars to be longer than dentals and retroflexes. Labials were mostly the shortest consonants, at least when unvoiced. Regarding manner of articulation, no clear tendencies were observed. Consonants after high vowels were longer than consonants after low vowels. This effect was stronger for phonologically long vowels. Other vowel-features (lip-rounding, front-back etc.) had no clear effect.

Word length and tonal accents affected consonant duration, as well. Consonants in monosyllabic words were longer than the same consonants in di-syllabic words, and consonants were longer in accent II words than in accent I words, especially in the word list.

The results presented by Elert (1964) show some effects that could be seen as a tendency to keep the duration of the VC-sequence constant, be it V:C or VC:. For example, high vowels, which are intrinsically comparatively short, are followed by longer consonants than the intrinsically rather long low vowels. A similar relation exists for vowels before voiced and unvoiced consonants: voiced consonants are shorter than unvoiced consonants, but vowels are longer before voiced consonants. Another compensatory relationship can be seen for vowels and consonants in accent I and accent II words. Considering the combined effect of these tendencies Elert Elert (1964) draws the following conclusion:

> "*There is a tendency for the total duration of a vowel plus an immediately following consonant to remain constant.*" Elert (1964), p. 151

There is one effect which influenced vowels and consonants in the same direction: both were longer in mono-syllabic words than in disyllabic words.

> "*. . . if account is taken of the variations caused by the intrinsic duration of vowels and consonants, of the influence of neighbouring phonemes on the duration, and of the variations among speakers, there is a strong correlation between the relative duration of vowels and consonants and phonemic quantity. The correlation is so regular that it strongly supports the assumption that the listener uses the relationship between the duration of stressed vowels and the immediately following consonant as a cue to the identification of phonemic quantity.*" Elert (1964), p. 174

In Section 2.6.2 this hypothesis will be considered further.

2.5 Factors of Durational Variation

Phonological length is not the only source of durational variation in speech. Also the following factors influence the duration of a speech sound (see e.g. Lehiste, 1970; Malmberg, 1944):

- Sound duration changes with speech rate.
- There are intrinsic effects on sound duration. For example, high vowels are shorter than low vowels.
- There are context effects on sound duration. For example, vowels before voiced consonants are longer than vowels before unvoiced consonants.
- Duration and accentuation are connected. Sound duration decreases with increasing distance from the stressed syllable.
- Duration and word length are connected. Sounds in longer words are shorter than sounds in shorter words.
- Duration and syllable structure are connected. Vowels in open syllables are longer than vowels in closed syllables.

In the context of quantity, mainly speech rate and stress influences have been discussed. Speech rate, above all, has always been viewed as a potentially disturbing factor for quantity perception. If perception would rely solely on absolute duration, variation in speech rate alone could shift quantity perception from one category to the other. Acoustic and perception studies have therefore often looked for "higher order invariants" (Pind, 1995) in durational terms, that is for relations between quantity bearing segments and surrounding segments, so that, even if a certain duration is not maintained in absolute terms, the distinction is still maintained in relative terms. Several authors have suggested the relationship between a vowel and a following consonant as a potentially important cue for quantity perception across different speech rates. The evidence for this hypothesis from perception studies will be discussed in Section 2.6, in this section, the results of acoustic studies will be reported.

For Swedish, Sock et al. (1996) showed that the relative durations of vocalic phases in stressed V:C and VC: remained stable across two different speech rates (normal and fast). The vowel in V:C sequences constituted about 58% of the whole sequence, while the vowel in VC: sequences constituted about 35% of the phase. In other words, the V:/C ratio was about 1.4, the V/C: ratio was about 0.5. The same test with a language with vocalic quantity without complementarity (Gambian Wolof) showed a less stable relationship. Here, the vocalic phase was compressed with respect to the consonantal phase in the fast speech condition. Sock et al. (1996) interpreted the results as proof for a more stable quantity distinction in Swedish. It has to be mentioned, however, that their results were obtained from only one subject per language. Pind (1999) investigated changes in absolute and relative segment duration in Icelandic, another language with complementary quantity. He investigated stimuli from six speakers in a short and a long sentence context with three different speech rates (fast, normal and slow). His results do not reveal any invariant vowel-to-consonant relationship across speech rates, especially not for V/C:. Even if only the results for the better controlled short sentence context are taken into account, the V/C: ratio varies from about 0.5 to 0.3 from the fast to the slow speech rate. V:/C shows less variation, from 1.3 to 1.2, or from 1.25 to 1.14, depending on the vowel context. His results clearly reveal the different elasticity of segments. For the short vowels, the slow-to-fast ratio does not exceed 1.2, while long vowels in the slow condition are about 1.5 times longer than in the fast condition. Short consonants lie in the same order of magnitude as the long vowels, or slightly higher (1.5–1.6). Long consonants show the highest amount of variation, being nearly two times longer in the slow than in the fast condition.

Similar effects were observed by Heldner and Strangert (2001) for Swedish, although not induced by speech rate variation but by variations in focal accent. The comparison of words with non-focal and focal accent revealed that the lengthening under focus was “non-linear”. While V: and C in V:C were lengthened to similar extents (approximately 20–25%), the lengthening of V in VC: was only about 5%, and C: was lengthened about 35%. Thus, focal accentuation seems to have similar effects as speech rate changes. V: and C are lengthened to similar extents, keeping the V/C ratio stable, while the short vowel is more resistant to durational changes. The C: consonant, on the other hand, is the most “elastic” segment.

The results by Pind (1999) and by Heldner and Strangert (2001) are clearly different from those presented by Sock et al. (1996). It has to be kept in mind, however, that Sock et al. (1996) only compared normal to fast speech rate and did not include slow speech rate. Most of the non-linear changes in Pind (1999) occurred between normal and slow speech rate, and focal accentuation can also be seen as a change towards a slowing-down.

Several authors (e.g. Bannert, 1969; Fant et al., 1991) have pointed out that durational quantity contrasts increase under accent. The non-linear lengthenings clearly support the durational distinctions. If V: is lengthened under accent, but V remains stable, the durational difference between the segments is increased. The same holds for C and C:, if C: is lengthened to a greater extent than C.

Heldner and Strangert (2001) leave the question open as to why the short vowel is resistant to lengthening. They point out that the distinction could be maintained even if the short vowel would be lengthened linearly. In a follow-up study (Heldner, 2001), the author comes to the conclusion that the shortness of the short vowel is important for naturalness judgements and suggests the existence of an expandability constraint that “is certainly not physiologically determined” (p. 108). Thus, a phonological motivation for the lack of lengthening is conceivable.

2.5.1 Quantity and Adaptive Dispersion

Engstrand and Krull (1994) investigated the realisation of phonological quantity in spontaneous speech for three different languages: Swedish, Finnish and Estonian. Both Finnish and Estonian have a more complex quantity system than Swedish. Engstrand and Krull (1994) interpreted their data in the framework of “adaptive dispersion”.

This theory (see e.g. Lindblom, 1987) takes bottom-up and top-down processes for the decoding of messages on the part of the listener into account and assumes that the speaker adapts to these processes. In other words, the speaker provides the listener with a contrast that is just sufficient. If the communicative situation provides the listener with additional information about the content of the message, the precision of articulation on the part of the speaker is low. When the speaker presents new, unknown information, articulation is more precise.

The “sufficient-contrast principle” has also been used to predict the position of vowels in a vowel system (Liljencrants and Lindblom, 1972). Engstrand and Krull (1994) transferred the concept to durational contrast and formulated three hypotheses for their investigation. The first hypothesis is based on adaptive dispersion and suggests that Finnish and Estonian speakers maintain the durational distinctions to a greater extent than the Swedish speakers. The authors point out that this is not self-evident. Motor control of quantity production could also be governed by rather peripheral processes. This is the basis for the second and the third hypothesis. Either, disruptive factors like stress and speaking rate are compensated for on the motor execution level, leading to the same amount of maintenance in all three languages. Or, durational differences based on quantity are simply overridden by factors like stress and speech rate, leading towards frequent reductions in all three languages. The two latter processes would not allow the speaker to adapt to the needs of the listener and to shape his utterances accordingly.

In all three languages, the authors found significant duration differences between long and short segments, in spite of the great durational variation usually occurring in spontaneous speech. The comparison of the languages indicated that Estonian and Finnish speakers maintain their durational distinctions more clearly than Swedish speakers. Thus, the authors see the adaptive dispersion hypothesis confirmed. They assume that the effect occurs for two reasons: Both Finnish and Estonian have a more "densely populated" quantity dimension than Swedish, and the accompanying quality differences of Swedish are largely lacking in Finnish and Estonian. Both characteristics force the speakers of Finnish and Estonian to maintain the durational contrasts to a greater extent.

2.6 The Perception of Quantity

Perception studies of Swedish quantity have mainly focussed on two aspects: The importance of vowel duration and quality cues for the perception of vocalic quantity, and the role of the following consonant for the perception of vocalic quantity.

2.6.1 Durational and Spectral Cues

Hadding-Koch and Abramson (1964) investigated the relation between spectral and durational cues in 3 Swedish vowel pairs: /ɛː/ versus /ɛ/, /øː/ versus /œ/ and /ʉː/ versus /ɵ/.

This choice of vowel pairs was based on the varying degree of quality difference between the members of the pair. /ɛː/ versus /ɛ/ was chosen as representative of a small quality difference, /ʉː/ versus /ɵ/ was chosen as representative of a large quality difference, while the difference in /øː/ versus /œ/ was considered to fall somewhere in between.

The authors constructed test stimuli with these three vowel pairs by embedding them in minimal pairs (Swedish *väg* — *vägg*, *ful* — *full* and *stöta* — *stötta*). Recordings of a male Southern Swedish speaker were used. The /ɛː/ in *väg* was reduced from 240 ms (original length) to 135 ms (length of /ɛ/ in *vägg*) in seven steps. The reduction cuts were made in the central part of the vowel. The manipulated stimuli and the two original utterances of *väg* and *vägg* were played in a carrier sentence in randomised order to 13 subjects of Southern Swedish origin, who judged each stimulus several times in a forced-choice test.

The results revealed a clear effect of vowel duration. The four longest stimuli, including the original, resulted in *väg* between 100% and 79%. Between the fourth and the third stimuli, the *väg* judgements dropped from 80% to 25%. The vowel duration of stimulus four was ca. 185 ms, that of stimulus three ca. 175 ms, as estimated from Figure 1 on page 100 in Hadding-Koch and Abramson (1964). The *väg* judgements for the stimuli two, one and the original *vägg* were close to 0%. The clear durational effect was accompanied by small differences in vowel formant frequencies between the original versions of the long and the short vowel. F1 differed 10 Hz, F2 40 Hz while the F3 values measured were identical.

For the /ʉː/ versus /ɵ/ vowel pair, two different sets of stimuli were created, as the /ʉː/ vowel showed lower intensity and a tendency towards a glide in the second part of the vowel. Therefore, the vowel in one of the sets was cut between the beginning and the midpoint, and in the other set between the midpoint and the endpoint. In each case, six stimuli were produced, including the non-manipulated version of the word *ful*. The duration ranged from 180 to 95 seconds in 15 ms steps. Three stimuli were shorter than the original /ɵ/ in *full*, which had a duration of 140 ms. Fifteen subjects judged the first set and 11 subjects the second set of stimuli. The results revealed a much lower influence of duration for this vowel pair. Cuts in the glide part lead to fewer changes in perception than cuts in the steady-state part. For the glide-cut stimuli, even the shortest stimulus still received around 70% long-vowel responses. If the steady-state portion was cut, around 55% of the listeners still reported long-vowel perceptions. Formant measurements from the original utterances supported the dipthongised character of the /ɵ/ and revealed considerable differences for all three formant values. The greater quality difference for the /ʉː/ versus /ɵ/ vowel pair was thus confirmed and agreed with a much lower perceptual influence of durational cues. The durational ratio of these vowels also was lower than of the other two vowel pairs (see Table 2.2). This corresponds with the low V:/V ratios in Elert (1964) for this vowel (see Table 2.1). The authors suggest a trading relationship between durational and quality cues:

> "*We might . . . speculate that, whatever the historical situation in Swedish was, as cue value shifted from length to quality in this pair, the constraint upon speakers to maintain a clear durational difference lessened.*" Hadding-Koch and Abramson (1964), p. 105, footnote 1

Their general conclusion concerning the experiment is:

> "*. . . the present study supports the traditional view that length is a distinctive feature of Swedish vowels, although it obviously does not pervade the whole vowel system.*" Hadding-Koch and Abramson (1964), p. 105.

In other words, there is clear evidence that some "long and short" vowel pairs in Swedish are primarily distinguished by duration. Other pairs rest mainly on quality cues. The results also suggest a trading relationship between these two features, which can be interpreted as a consequence of adaptive dispersion.

Table 2.2: Mean vowel durations and ratios from Hadding-Koch and Abramson (1964)

Vowel pair	*V: duration*	*V duration*	*V:/V (V/V:) ratio*
/ɛː/ versus /ɛ/	240	135	1.78 (56%)
/ʉː/ versus /ɵ/	175	140	1.25 (80%)
/øː/ versus /œ/	185	100	1.85 (54%)

Hadding-Koch and Abramson (1964) mention consonant duration only briefly in their paper. They observe longer consonants after short vowels, "although the difference was often slight” (p. 97). After performing a pilot experiment with two listeners that did not change their linguistic judgement of a word pair when the durations of the post-vocalic consonants were reversed, the authors decided not to include post-vocalic consonant duration as a controlled variable in their experiment.

Behne et al. (1996) performed a similar perception experiment as Hadding-Koch and Abramson (1964) with an extended set of stimuli and a resynthesis technique. They investigated the three vowel pairs /iː/ versus /ɪ/, /oː/ versus /ɔ/ and /ɑː/ versus /a/. The vowels were presented in a nonsense carrier word ending in a voiceless plosive. The category judgements were obtained via a rhyme test. The authors manipulated durational as well as spectral cues. Using recordings of a male Stockholm speaker, durational as well as spectral cues (F1 and F2) were manipulated in ten steps from the long to the short version of the stimuli, resulting in 10x10 stimuli per vowel pair. Postvocalic consonant duration was set to the mean value of the duration after long and short vowels. As all stimuli were produced from manipulations of the long vowel version, the authors expected a bias towards the long vowel.

The results showed a clear dominance for durational cues for /iː/ versus /ɪ/ and /oː/ versus /ɔ/. The mean long vowel judgements for these two pairs were unaffected by the spectral manipulations. For both vowels, the category boundary seemed to appear between the sixth and the seventh stimulus, where long vowel judgements fell from around 80% to around 40% for /iː/ versus /ɪ/ and from around 90% to around 40% for /oː/ versus /ɔ/. However, the overall manipulation effect was less successful for the back vowel. The /oː/ manipulations still led to around 20% long-vowel judgements even with the shortest stimulus. For /iː/, the long-vowel judgements were close to 0 for the shortest stimulus.

For /ɑː/ versus /a/ the results suggested a combination of durational and spectral cues. Spectral manipulation alone changed the mean long vowel judgements from around 70% to around 30%. In this case, the changes in judgement were rather continuous, showing an S-shaped curve. The durational manipulations changed the long vowel judgements from 90% to 10%, again with a clear drop between the sixth and seventh stimulus from around 60% long vowel judgements to around 25%.

The results of the reported investigations generally suggest an important role of durational cues in Standard Swedish and Southern Swedish. For some vowel pairs, spectral cues influence the judgements.

2.6.2 The Influence of the Post-Vocalic Consonant

In all experiments described so far, postvocalic consonant duration was either not controlled or kept constant. Behne et al. (1998) investigated the influence of postvocalic consonant duration on the perception of Swedish vowel length. The test procedure was similar to the one used in the previously described experiment, but this time postvocalic consonant duration was manipulated, while vowel durations and spectral cues were kept constant.

The vowel durations and spectral values were averaged over long and short productions and the vowels in the test stimuli were set to these mean values. Postvocalic closure duration was changed in 10 equidistant steps from the duration of a typical consonant duration after a long vowel to the typical duration after a short vowel. The stimuli included the three vowel pairs mentioned above and /t/ as well as /d/ as post-vocalic consonants. Twenty-one subjects were tested with these stimuli in a randomised forced-choice rhyme test.

The results showed virtually no effect of the post-vocalic consonant. For the stimuli with the voiceless plosive, average long-vowel responses were around 80% – 95% across all stimuli. For the condition with the voiced plosive, average long-vowel responses were more variable across vowels, but again, the responses did not change depending on consonant duration. In most cases, the average number of long-vowel responses in this experiment matched the long-vowel responses for comparable stimuli in the previous experiments by the same authors. The authors conclude from their study that post-vocalic consonant duration is irrelevant for quantity perception.

However, there are reasons to believe that the procedure used by Behne et al. (1998) might have missed certain aspects of the V-C relationship: In terms of relative duration, it is questionable whether the set of stimuli used in the experiment did span the usual range from long to short vowels. For example, the /iː/ vowel had a duration of 108 ms and the postvocalic closure duration of /t/ was changed in 6 ms steps from 102 ms to 154 ms. For the longest consonant, the V/C ratio was thus 0.7. In the material studied by Elert (1964), V/C ratios in comparable contexts were about 0.35 – 0.4. The experimental procedure did thus not include the relative values usually observed in VC: contexts, at least for the voiceless consonant condition.

Thus, for quantity judgements it is likely that the listener takes also context information into account. In research on voiceless versus voiced distinctions, the ratio of the consonant to the vowel (C/V) has been suggested as an important auditory cue for the distinction (see Kluender et al., 1988; Kingston and Diehl, 1994). Dommelen (1999) transferred this concept to the complementary quantity pattern in Norwegian. Like Behne et al. (1998), he investigated the influence of consonant length on the perception of the vowel, but with a different procedure. Dommelen (1999) used a bi-syllabic word pair (Norwegian *mate*, 'to feed', versus *matte*, 'mat'). The duration of the stressed vowel, the intervocalic consonant and the word-final schwa were manipulated.

His results suggest that there is an influence of the following consonant on the categorical perception of the vowel quantity. But the effect is lower than the V/C ratio would suggest. Dommelen (1999) varied the vowel duration from 214 to 80 ms in 8 ms steps. These values were tested before two different consonant durations: 170 ms ("short consonant") and 250 ms ("long consonant"), and two different schwa durations, 72 ms ("short schwa") and 143 ms ("long schwa"). In the long schwa condition, the long consonant shifted the long/short vowel perceptual boundary by -8.2 ms. Before the short consonant the border occurred at 136.2 ms, before the long consonant, it occurred at 144.4 ms. For the V + short C condition, the V/C ratio at the category boundary was 0.8, for the V+ long C condition, it was 0.58. Thus, the category boundary is clearly not a function of the V/C ratio here. For the V + long C condition there were at least seven stimuli in the experiment where the V/C ratio was below 0.8 without a categorical change from long vowel to short vowel. In other words, if one would predict the results of the long C condition from the absolute vowel length in the short C condition, only one stimulus would have been predicted incorrectly (the step size in vowel change was 8 ms, the size of the effect was -8.2). If the V/C ratio would have been used as a predictor, at least seven stimuli would have been categorised incorrectly. Exactly the same argument applies to the vowel length as a percentage of the whole sequence (V/(V+C)): These values reflect the category boundary worse than the absolute durations of the vowel alone, at least under the conditions of the experiment of Dommelen (1999).

In the short schwa condition, the effect of the consonant was stronger, shifting the perceptual boundary of the vowel from 116 ms to 135.3 ms, thus the size of the effect would be -19.3 ms in this case. Consequently, three vowel durations would be categorised incorrectly if the category boundary in the long consonant condition would be predicted from the one in the short consonant condition. Again, nothing is gained by using the V/C ratio. The ratio for the short consonant condition is 116/170 = 0.68, for the long consonant condition it is 135.3/250 = 0.54. For the 250 ms consonant, a 0.68 V/C ratio results from a vowel duration of 170 ms, a difference of about 35 ms from the "true" category boundary. This, again, is a worse estimation of the boundary than the absolute duration of the vowel.

It is therefore likely that the V/C ratio overstates the influence of the consonant on the vowel. Traunmüller and Krull (2003) investigated the influence of speech rate changes on the perception of the quantity opposition in Estonian. Their results suggest that all segments in the vicinity of the quantity-carrying segment have an influence on quantity perception. The 'window' of segments used for estimations of durations is considerably larger than just the following consonant. In a follow-up study, Krull et al. (2003) compared Finnish, Estonian and Norwegian. The authors showed that the window length for quantity estimations is likely to be different for different languages. Estonian and Finnish use a larger window for quantity perception than Norwegian. Still, their results indicate that the V/C ratio overstates the influence of the consonant on the vowel in Norwegian.

2.7 Conclusion

This chapter discussed the phonological and phonetic aspects of quantity in Swedish and other languages. Phonologically, a number of different solutions have been suggested for quantity distinctions like those in Standard Swedish. In some, the vowel has been treated as the distinctive element, in others, the consonant. Alternatively, quantity, has been viewed, for example, as a prosodeme, length, that combines either with the vowel or the consonant, while the long and short segments are seen as allophones.

In addition, analyses of quantity in various languages have shown that "long" versus "short" distinctions may be interpreted in different ways, in terms of duration and/or spectral differences but also in terms of the "syllable-cut distinction". The syllable-cut theory, suggests that — in some languages — the durational distinction is a secondary effect of a difference in the articulatory coupling between a vowel and the following consonant.

Phonetic studies of complementary quantity in Standard Swedish, have — in addition to the detailed description of durational aspects given by Elert (1964) — also resulted in very important findings of how the quantity pattern is affected by stress and speech rate variation. For example, Engstrand and Krull (1994) in a comparative study of Finnish, Swedish and Estonian, showed that the durational distinctions are maintained to a greater degree in languages with a more complex quantity systems — Finnish and Estonian — than in Swedish and interpreted this finding as supporting the hypothesis of adaptive dispersion.

Speech rate variation appears to affect the perception of quantity distinctions as well and experimental data show that the perception of quantity in Swedish (and related languages) is influenced by the duration of surrounding sounds. But, as even in languages with complementary quantity, the V/C ratio overstates the perceptual influence of consonant duration on vowel duration, it was argued against a simple V/C ratio interpretation of quantity perception.

The interaction of durational and spectral cues is another issue taken up. Here, investigations have shown that some "long" versus "short" vowel pairs in Swedish are mainly distinguished by duration, while others show a greater influence, or even a primacy of spectral cues.

Therefore, in the next chapter, the connection between spectral and durational cues will be described and analysed with regard to the tense/lax feature.

Chapter 3

Accounts of Tenseness, Quantity and their Relationship

The acoustic and perceptual studies reported in Chapter 2 have shown that Swedish long versus short vowel pairs differ in the amount of spectral differences accompanying the length distinction. Such combined spectral and durational differences between vowel pairs have been observed for a range of languages and have led to several different phonological interpretations. Frequently, the combination of durational and spectral cues has been viewed in terms of a tense/lax feature. In the following, this approach is discussed, as the interpretation of the "combined differences" is of primary importance for the phonological interpretation of Standard Swedish and its dialects.

3.1 The Tense-Lax Feature in the Literature

The tense/lax feature is one of the notoriously difficult concepts in phonology. Therefore, some of the seminal works on this matter are reviewed here. The terms 'tense' and 'lax' suggest an articulatory definition. Chomsky and Halle (1968), for example, describe tenseness in the following way:

> "*The feature 'tenseness' specifies the manner in which the entire articulatory gesture of a given sound is executed by the supraglottal musculature. Tense sounds are produced with a deliberate, accurate, maximally distinct gesture that involves considerable muscular effort...* " Chomsky and Halle (1968), p. 324

In Jakobson et al. (1951), the following definition is given:

> "*Tense phonemes are articulated with greater distinctness and pressure than the corresponding lax phonemes. The muscular strain affects the tongue, the walls of the vocal tract and the glottis. The higher tension is associated with a greater deformation of the entire vocal tract from its neutral position. This is in agreement with the fact that tense phonemes have a longer duration than their lax counterparts.*" Jakobson et al. (1951), p. 38

In addition, they describe acoustic effects of the articulatory differences. As a consequence of the greater deformation of the vocal tract, the sum of the deviations of a tense vowel's formants from the neutral position is greater than that of a corresponding lax vowel (see Jakobson et al., 1951, p. 36). The longer duration of the tense sounds is described as a *result* of the articulatory settings. Jakobson et al. (1951) do, however, not abandon the idea of a long-short feature. They divide distinctive features into two classes: inherent features and prosodic features (although most of their work describes the inherent features). The difference between these two classes of features lies in the dependence on context. An inherent feature (like tense/lax) needs no reference to the context it appears in. Prosodic features, on the other hand, are only definable with respect to the context. In the description of prosodic features, Jakobson et al. (1951) refer explicitly to the long versus short distinction. This distinction is thus treated as a prosodic one, and the long versus short opposition is claimed to be based entirely on relative, not absolute duration (see Jakobson et al., 1951, p. 14).

In later editions of "Preliminaries to Speech Analysis", the editors added an article about "Tenseness and Laxness" by Roman Jakobson and Morris Halle. This article contains a survey of different theories about the distinction. It further provides a detailed list of the proposed correlates of the tense/lax distinction. In addition to the already mentioned features, the authors list differences in tongue and pharynx shape, increased sub-glottal pressure for tense vowels, and a longer phase of culmination for tense sounds.

With regard to the relationship of tense/lax and long/short, the authors refer again to the distinction between inherent and prosodic features, this time explicitly mentioning the syllable as a possible domain of the prosodic features:

> "*The prosodic length of a vowel is inferred from the contrast of long and ceteris paribus short vowels in a syllabic sequence, whereas length as a component of the tenseness feature is intrinsically connected with the other, qualitative manifestations of the given feature within the same phoneme.*" Jakobson and Halle (1961), p. 59

With regard to the sound structure of Dutch, where the lax vowels are usually described as shorter than the tense vowels, they state[1]:

> "*. . . yet for the identification of these phonemes shortness is hardly decisive, since however much one stretches/a²/ in /rà²t/, rad, 'wheel', it does not change into /rà¹t/, raad, 'council'.*" Jakobson and Halle (1961), p. 59

This statement relates the difference between tense/lax and long/short to perceptual effects like the ones tested by Hadding-Koch and Abramson (1964). If lengthened short vowels (and shortened long vowels) do not lead to perceptual changes (the perception of another word), the primacy of the tense/lax distinction is claimed. It is evident that also the reverse should be true: if lengthening or shortening leads to changes in perception, the long/short feature should be of primary importance. We will return to this discussion below.

3.2 Acoustic Correlates of Tenseness

Mooshammer (1998) reviews several investigations that address the tense-lax distinction, with a special focus on German. She reports the already mentioned centralisation of the formant

[1]The superscripts '1' and '2' are used by the authors for marking 'tense' and 'lax', respectively.

values for lax vowels compared to their tense counterparts. It is interesting to note that, in German, the three lax vowels /ɪ/, /ʏ/ and /ʊ/ lie closer to the tense vowels /e/, /ø/ and /o/ than to their normally claimed counterparts /i/, /y/ and /u/ in the F1-F2 chart.

Furthermore, Mooshammer (1998) reports several works which examine the behaviour of tense vowels in lexically unstressed positions. Jessen (1993) and Jessen et al. (1995) for example examined word pairs with morphological changes that moved the tense or the lax vowel from a stressed into an unstressed position. Their results point towards the existence of tense unstressed vowels in German, but also to a neutralisation of the contrast under many circumstances. The vowels seem to differ in the probability of neutralisation. The /i/ versus /ɪ/ contrast is often reported to be neutralised. The neutralisation and non-neutralisation is obviously also connected to the recording situation. While unstressed "tense" vowels can be found in "lab speech", the contrasts are often neutralised in spontaneous speech (see Mooshammer, 1998). This suggests a continuous process.

For Swedish, Engstrand and Nordstrand (1983) presented data on the durational and spectral reduction of the vowel pairs /iː/ versus /ɪ/ and /ʉ/ versus /ɵ/. The authors conducted an experiment with one subject (the first author Engstrand). The two vowel pairs were recorded in two stress (stressed – unstressed) and two speech rate (slow – fast) conditions. For /iː/ versus /ɪ/, the qualitative difference in the vowel pairs was maintained for all conditions, which was mainly observable in the F3 frequencies. Durational differences were greatest in the slow, stressed cases, and lowest in the fast, unstressed cases, where the duration difference between long and short (tense and lax) vowels was not significant anymore. Furthermore, the data showed that stress alone had a significant effect on the quality of the vowels, while rate variations seemed to influence the formant values to a lesser degree. Engstrand and Nordstrand (1983) conclude from their experiment that durational and spectral qualities of vowels are quite independent. For a single vowel type, they observed great variations in duration combined with minor quality variations, and minor durational variations combined with large quality variations. Thus, the vowel spectrum is not a simple function of the vowel duration. For /ʉ/ versus /ɵ/ the authors observed a rather small duration difference even in the slow stressed condition, in accordance with the reports of Elert (1964) and Hadding-Koch and Abramson (1964) for this vowel pair. The qualities of these two vowels are thus not strongly influenced by stress variation. The influence of rate variations is stronger, both in terms of duration and spectral changes. Thus, the effects are different for the two vowel pairs.

For both vowel pairs, formant dynamics were not the same in the different conditions, which on the one hand points towards undershoot (see below) and on the other hand towards an adaptation of initial formant frequency to medial formant frequency.

The authors interpret their results as evidence for a better maintenance of spectral than of durational correlates with the tense – lax distinction. It has to be taken into account, however, that the data basis for this conclusion was rather small, as only one subject and only two vowel pairs were investigated. One of these vowel pairs had already shown rather exceptional durational and spectral differences in previous investigations.

3.3 Articulatory Correlates of Tenseness

In the Swedish tradition (e.g. Elert, 1964; Gårding, 1974; Bannert, 1976), the length distinction has often been viewed as the primary distinction. Accompanying spectral differences between vowels have been interpreted as a consequence of the durational distinction. Gårding (1974) and Bannert (1976) for example favour an explanation resting on the articulatory "undershoot hypothesis". In the words of Gårding:

> "*När talaren har tid att utföra ett långt i eller e blir vokalen mera sluten. För de öppna vokalerna är däremot käköppningsrörelsen viktigast. För en lång, öppen vokal hinner talaren öppna käken mera och långt a blir därför öppnare än kort.*[2]
> Gårding (1974), pp. 29–30

Both Gårding (1974) and Bannert (1976) refer to Lindblom (1963). Lindblom's study dealt with the behaviour of Swedish *short* vowels under different stress and speech rate conditions. He concluded from his results that the speaker tries to reach a certain duration- and context-independent target (described as target values for F1 and F2) for each vowel but fails to reach it when duration is decreased. The effects of de-stressing and increased speech rate were quite similar in this respect. Although the study was performed in the acoustic domain, Lindblom provides articulatory interpretations of his data:

> "*As a vowel becomes shorter, there is less and less time for the articulators to complete their 'on-' and 'off-glide' movements within the CVC syllable. Provided that the neural events corresponding to the phonemes actually stay invariant, the speech organs fail, as a result of the physiological limitations, to reach the positions that they assume when the vowel is pronounced under ideal steady-state conditions.*" Lindblom (1963), p. 1779

It is important to note that Lindblom did not refer to the long vowel versus short vowel distinction in his study. However, the acoustic differences between the long and short variants of vowels resemble those found in his and other reduction studies. It thus appears to be quite straightforward to apply his results to quantity distinctions and to think of a short vowel as a "reduced long vowel".

However, articulatory investigations have challenged this view. Hoole and Mooshammer (2002) have investigated the kinematic characteristics of German tense and lax vowels in a CVC context under different stress and speech rate conditions. They used electromagnetic midsagittal articulography (EMMA) with four sensors on the tongue. The CVC sequence was segmented by velocity patterns of the sensor closest to the consonant articulator (the onset C and the coda C were always identical). A CV phase was distinguished from a nucleus and a VC phase. The onsets and offsets of the CV and the VC phase were determined by a 20% threshold criterion of the tangential velocity signal. In other words, the beginning of the CV phase was set to the point where the velocity of the sensor exceeded a certain threshold, the end was set to the point where the value fell below this threshold. The same criterion was used for the VC phase. The nucleus was defined as the phase between CV offset and VC onset.

Their results showed that the nucleus phase of the tense vowels was much longer than the nucleus phase of the lax vowel in the stressed condition and with a normal speech rate. For tense vowels, the sensor remained in a "holding phase" for some time. For the lax vowel, the time when the sensor stood still in the nucleus phase was much shorter, showing basically only

[2] *When the speaker has time to produce a long /i/ or /e/, the vowel becomes more closed than for the short counterparts. For the open vowels, however, the jaw-opening movement is the most important. In the production of a long open vowel, the speaker has time for a greater jaw opening, and a long /a/ therefore becomes more open than a short one.* (Translation by Felix Schaeffler)

a turning point in the movement pattern. Furthermore, the nucleus phases of the tense and lax vowels showed an interesting behavior for stress and speech rate changes. The nucleus phase of tense vowels was compressed to a considerable extent if the vowel was de-stressed. Speech rate changes reduced the nucleus phase of tense vowels as well, but the effect was not as strong. The nucleus phase of the lax vowels was however only slightly affected by the speech rate or the stress condition.

Another interesting effect could be seen in the context of movement amplitudes for the tense and lax vowels. Both de-stressing and speech rate increase reduced the movement amplitudes for the vowels, as might be expected. A comparison of the movement amplitudes for tense and lax vowels showed, however, an unexpected effect: For front vowels at least, the movement amplitudes for lax vowels were actually larger than for tense vowels. Hoole and Mooshammer (2002) comment on this in the following way:

> *"... 'lax' clearly suggests that the articulatory system has, in some sense, less to do than in tense vowels — which at first sight fits in with the obvious fact that tense vowels are more peripheral. But this viewpoint really looks at the vowels as isolated sounds. As soon as the vowels are produced in valid German syllable structures ... then generally, more movement is required in the lax case."* Hoole and Mooshammer (2002), p. 21

The application of the "undershoot hypothesis" to the tense/lax or long/short distinction has thus a certain weakness: the "undershoot" of the lax or short vowel would predict less movement during the articulation of this vowel. The results of Hoole and Mooshammer (2002) suggest the opposite, at least for some vowel pairs. Hoole and Mooshammer (2002) interpret their results in terms of "underlying force". Tense vowels use a "distributed force input", lax vowels a "pulsatile force input". The authors suggest that lax vowels differ from tense vowels in the strength of the coupling of the movements from the onset consonant to the coda consonant. For a lax vowel, the coupling is rather tight. Thus there is not much place for variation in the duration of the whole gesture complex. One could think of a process controlled by a single impulse. For the tense vowel, the connection between the gestures from the onset consonant to the coda consonant is rather loose, and thus the "vocalic nucleus" (in the sense of Hoole and Mooshammer, 2002) can be varied in duration. This account could provide a possible explanation for the syllable structure constraints for lax vowels: To induce the typical articulatory pattern for a lax vowel, the vowel has to occur between two consonants.

3.4 Tenseness and the Syllable Cut

The results of Hoole and Mooshammer (2002) can be interpreted as phonetic evidence in favour of the syllable cut concept, at least with respect to the non-existence of short vowels in stressed open syllables.

Other authors have looked for alternative phonetic correlates of the syllable cut on the acoustic level. Greisbach (2001) investigated three potential acoustic correlates of the syllable cut in Northern Standard German. Comparisons of the slope of the intensity contour at the end of German /aː/ and /a/ did not reveal any significant results. Likewise, a perception test with different vowel durations and manipulated intensity slopes did not show any significant influence of the intensity slope on the perception of the vowel as "long" or "short". The two other investigated acoustic correlates were the formant transitions to the following consonant and the fundamental frequency contour at the end of the vowel.

Significant effects were only observed for the formant transitions. When the vowel was truncated from the end, thereby eliminating the formant transitions to the consonant (a voiceless plosive), the long-to-short category boundary was reached significantly later than when periods in the middle of the vowel were cut out, thereby leaving the formant transitions intact. Greisbach (2001) interprets this observation as evidence for the syllable cut, in spite of the fact that the observed results were actually the direct opposite to the expected results: the original hypothesis was that the truncation of the formant transitions would model abrupt cut, thus leading to an earlier categorical change. Furthermore, the observed effect with manipulated vowel-consonant transitions does not provide any evidence that long and short vowels in fact show differences in formant transition to the following consonant.

3.5 Conclusion

This chapter has shown that all accounts of the combined spectral and durational cues in Swedish and similar languages have certain weaknesses. Proponents of the tense-lax feature recognised the existence of an additional feature "length", which was not connected to the tense-lax feature, in spite of the obvious correspondences in durational behaviour. Experiments like those reported in Hadding-Koch and Abramson (1964) showed, however, that the primacy of spectral and durational cues can vary within one and the same vowel system. Consequently, one would have either to accept the use of both features (tense/lax, long/short) for one and the same vowel system with negative consequences on feature economy and the identification of natural classes, or the definition of the tense-lax feature should be revised.

Proponents of a long-short feature have suggested that spectral differences between long and short vowels are a secondary effect of the length distinction and assumed an "undershoot" effect. Articulatory evidence suggests, however, that "undershoot" is not an appropriate model for the distinction, as the predicted effects (less tongue movement for lax vowels) are not borne out.

An alternative theory suggests that tense and lax vowels show different articulatory dynamics which demand a following consonant for lax, but not for tense vowels.

Up to now, only synchronic accounts have been considered. Delattre (1962) has presented an alternative view of the matter. He argues against the common notion of tense and lax and views the modern English tense-lax contrasts as the result of a former long-short contrast:

> "*. . . /i/ is longer than /I/ today not because it is less open . . . but because of the survival of a former (Middle-English) distinctive feature* long/short */i:/i/ which gradually changed to a rather* less central/more central *articulatory distinction /i/I/ with attenuation (but not extinction) of the old* long/short *distinction. Thus it seems to be distinctive shortness of the vowel that caused centering (and indirectly opening) and not the opposite.*" Delattre (1962), p. 1142

For the Swedish quantity system a diachronic perspective is important and will be discussed in the next chapter.

Chapter 4

The History of the Swedish Quantity System

This chapter provides a description of the development of Modern Swedish. After a short survey of the main diachronic processes from Proto-Nordic to Modern Swedish, the details of the diachronic quantity changes are described on the basis of Riad (1992) in the framework of metrical phonology. The main topic in this chapter will be the "quantity shift" and the changes in segmental and syllabic quantity associated with it. Important changes before the quantity shift are described only as far as they are relevant for the quantity changes that followed.

4.1 A Short Historical Summary

Table 4.1 provides an overview over suggested periods in the development of Modern Swedish. It lies in the nature of such classifications that they are approximative and theory-dependent. The table shows that different authors (Haugen, 1976; Pettersson, 1996; Riad, 1992) use different schemes, and that the use of the term "Swedish" is in the early stages rather a convention than an indication for the existence of a distinct Swedish language.

Traditional accounts of Germanic languages state that all Germanic languages are derived from a common ancestor. This proto-language had a quantity system that used vowel and consonant duration in stressed and unstressed syllables. Stress fell normally on the initial syllable. From about 500 AD, a distinct Scandinavian language evolves. This is for example characterised by the deletion of vowels in unstressed position. This period is therefore often called the "syncope period" (Riad, 1992, p. 151).

A further change following the syncope period is the monophthongisation of diphthongs. This change affected mainly Eastern Scandinavian (which developed into Swedish and Danish), but not Western Scandinavian (which developed into Norwegian and Icelandic), and this change is thus for the development of the different modern languages.

After these changes, "Old Swedish" had the following properties with respect to quantity: Vocalic and consonantal quantity were still independent, but vocalic quantity was restricted to stressed positions. Consonants however, could still show a long versus short distinction word-finally, although this distinction was also deteriorating. According to Riad (1992, p. 253) it is unclear at which point exactly consonant quantity was lost in unstressed position, but it is usually assumed that the quantity system immediately before the quantity shift was a system with vocalic and consonantal quantity, although only in stressed position.

Table 4.1: Periods and changes in the history of the Swedish language from Haugen (1976), Petterson (1996) and Riad (1992). (For Pettersson (1996), the terminology is given in Swedish.)

<table>
<tr><th>time frame</th><th>Haugen</th><th>Pettersson</th><th>Riad</th><th>Changes</th></tr>
<tr><td>before 500</td><td>Proto-Scandinavian</td><td>Urgermanska</td><td>Early Proto-Nordic (–600)</td><td></td></tr>
<tr><td>500–800</td><td>Common Scandinavian</td><td>Urnordiska</td><td rowspan="2">Proto-Nordic</td><td>Syncope period</td></tr>
<tr><td>800–1000</td><td>Old (East) Scandinavian</td><td rowspan="2">Runsvenska</td><td rowspan="2">Monophtongisation</td></tr>
<tr><td>1000–1200</td><td>Old East Scandinavian</td><td rowspan="2">Early Old Swedish</td></tr>
<tr><td>1200–1300</td><td>Old Swedish</td><td rowspan="2">Klassisk Fornsvenska</td><td></td></tr>
<tr><td>1300–1400</td><td rowspan="2">Middle Swedish</td><td rowspan="3">Late Old Swedish</td><td rowspan="5">Quantity Shift</td></tr>
<tr><td>1400–1500</td><td>Yngre Fornsvenska</td></tr>
<tr><td>1500–1600</td><td rowspan="3">Modern Swedish</td><td rowspan="2">Äldre Nysvenska</td></tr>
<tr><td>1600–1700</td><td rowspan="2"></td></tr>
<tr><td>from 1700</td><td>Yngre Nysvenska</td></tr>
</table>

4.1.1 The Quantity Shift

The quantity shift is usually dated around the year 1200, in the later Old Swedish period (Riad's "Late Old Swedish"). As already mentioned, the quantity shift resulted in the loss of V:C: and VC sequences. The loss of V:C: forms is usually only briefly mentioned in the literature. The main attention has been concentrated on the lengthening of the VC sequences. Here, two phenomena have received considerable attention: on the one hand, whether VC was changed by consonant or vowel lengthening, and on the other hand, how the change proceeded in words with different numbers of syllables.

There are clear dialectal differences in the lenghtening patterns. Southwestern Swedish dialects lengthened almost always the short vowels, and sometimes also post-vocalic /m/. Central and northern dialects, on the other hand, lengthened either the vowel or the following consonant. Hesselman (1902) has given detailed descriptions of the changes, and his findings are often referred to as "Hesselman's laws". These laws state that lengthening occurred first in mono-syllabic stems and that the phonetic quality of sounds played a role in the lengthening processes: Low vowels were always lengthened. High vowels were also lengthened, unless they were followed by a voiceless stop, /s/ or sometimes /r/. In these cases, the consonants lengthened (cf. also Riad, p. 271).

4.2 Riad's Account of the Quantity Shift

There are several alternative theories of the quantity shift, for example those of Haugen (1976) and Widmark (1998). However, the following discussion will be restricted to the alternative given in Riad (1992). Riad interprets the quantity shift as a consequence of changes in the stress system. Furthermore, he connects the quantity shift, especially its details in time course, with another Scandinavian phenomenon, namely balance. In general, he describes the quantity shift as a transition from a segmental quantity system to a prosodic quantity system.

It is the main objective of this section to describe the background of Riad's assumption that consonant quantity is underlying — and vowel length only derived by rule — after the quantity shift. Therefore, not all details in Riad's theory will receive equal attention.

According to traditional accounts, three types of syllable quantities existed before the quantity shift: light syllables, heavy syllables and superheavy syllables. In mono-syllabic words, VC rhymes are interpreted as light, V:, V:C and VC: rhymes are interpreted as heavy and V:C: rhymes are interpreted as superheavy. In bi-syllabics, the first syllable of CV.CV is light, CV:.CV and CVC.CV are heavy and CV:C.CV is superheavy.

Riad criticises this traditional account because of its disregard of syllable structure. In mono-syllabic words the consonant(s) following the vowel are part of the syllable, while in bi-syllabic words a stressed light syllable contains only a short vowel ('CV.CV). This structure does not exist in mono-syllabic words: Here, short vowels do not exist in open syllables. This minimal word condition is normally explained by the fact that Germanic content words have to be bi-moraic.

Therefore Riad suggests that all mono-syllabic words are heavy. His definition of light, heavy and overlong syllables refers to moraicity: Light syllables contain one mora, heavy syllables contain two morae, either a long vowel or a short vowel plus a consonant or the first part of a geminate. Overlong syllables contain three morae: a long vowel plus the moraic part of a geminate.

The historical analysis in Riad (1992) begins before the syncope period, hence at a stage where a clear separation between the Germanic languages is not assumed. Riad calls this stage "Early Proto-Nordic" (see Riad, 1992, pp. 35–36). The description of quantity and stress at this time should apply to the ancestor language of all Germanic languages.

4.2.1 The Situation Before the Syncope Period

In Early Proto-Nordic, Quantity is phonematic on the segment level. V contrasts with V: and C contrasts with C: in stressed and unstressed position. Hence, VC, V:C, VC: and V:C: are all possible. The main-stress position of words is marked in the lexicon, and is always stem-initial. Main stress on light syllables is therefore possible, but not preferred. Riad interprets the existence of light syllables in bi-syllabic words as an effect of four, partly conflicting, principles: Prokosch's law, the minimal word constraint, the onset principle and onset maximisation. Prokosch's law is the *tendency* of Germanic to make stressed syllables bi-moraic. The minimal word constraint states that words must have at least two morae[1]. The onset principle states that every syllable should have an onset. Onset maximisation states that the onset should be maximal, that means that, as long as phonotactic and other constraints allow this, out of a sequence of consonants between two vowels as many consonants as possible should form the onset of the second syllable. As a consequence of Prokosch's law, light stressed syllables should be excluded. But in the case of CVCV, the onset principle overrides Prokosch's law. The result is thus <u>CV</u>.CV with a light stressed syllable, and not <u>CVC</u>.V with a heavy stressed syllable, followed by a second syllable without an onset.

The heavy-light distinction was different for main stressed and final (unstressed) syllables. In main stressed syllables, any segment could be moraic. In final syllables, only sonorant segments could be moraic. The rather rare mono-syllabic words in early Proto-Nordic behaved like final syllables. As the minimal word constraint demanded two morae, early Proto-Nordic mono-syllabic words had always two sonorant segments (see Riad, 1992, p. 145).

4.2.2 The Situation After the Syncope Period

As mentioned before, during the syncope period vowels in unstressed position were lost or reduced. The detailed discussion of the processes occurring during the syncope period in Proto-Nordic is left out here. It suffices to say that the application of syncope in open unstressed syllables and closed syllables ending in an obstruent provides the evidence for the heavy/light distinctions in non-main stressed syllables. Furthermore, during the syncope period vowel length distinctions in unstressed position had been lost and the number of heavy syllables in unstressed position was generally reduced. Consequently, the evidence for heavy syllables outside the main-stressed position diminished after the syncope period.

After the syncope period a re-interpretation of the stress system occurred. As a first step, the interpretation of mono-syllabic words changed. Before the syncope period, they were treated as "final syllables". Thus, the sonority of the segments in the syllable contributed to its weight. After the syncope period, all mono-syllabic words became heavy by the assignment of main stress.

The new stress system is characterised by the following rules (see Riad, 1992, p. 251): Underlyingly long segments still have a pre-linked mora. The string of segments in a word

[1] Note that a CV.CV word satisfies this constraint.

is syllabified with core syllables that consist of an onset and a mora. Segments outside this domain are not syllabified. Then the bi-moraic main-stress foot is assigned. If main stress falls on a syllable which does not have two morae (underlyingly or via core syllabification), and where the target syllable is not directly followed by another prosodic structure, a mora is inserted (see Table 4.2, stage A).

In other words, a main-stress syllable only gets a mora by mora insertion, if the syllable does not yet contain two morae, and the syllable is followed by an unsyllabified segment. Thus, main-stress syllables with underlying morae do not receive a mora by mora-insertion (they already have two, one underlyingly and one from core syllabification). Neither do CV.CV structures receive a mora by mora insertion, because there is no unsyllabified segment following the main-stress syllable. Mora insertion is however applied to CVC syllables, if both segments are underlyingly short. In this case, core syllabification adds one mora, and main-stress assignment adds another mora, which goes to the unsyllabified C. This mechanism predicts the neutralisation of length distinctions in the coda of mono-syllabic words, as exactly the same structure surfaces if the vowel is underlyingly short and the consonant is underlyingly long (see again Table 4.2, stage A). Riad argues that this was precisely the situation in Swedish at this time and presents orthographic evidence for it (cf. Riad, 1992, p. 288). Furthermore, he argues that word-final geminates were not very frequent even at earlier stages.

The mora insertion rule then laid the foundations for the lengthening processes during the quantity shift. This process starts with the CVC words, where consonantal quantity was neutralised and Riad assumes that the consonant surfaced short.

4.2.3 Balance

Balance is a phenomenon that, according to Riad's analysis, started to occur after the syncope period (Riad, 1992, p. 171). Balance describes the fact that the vowel quality of the second syllable in bi-syllabic words depends on the quantity of the first syllable. If the first syllable is light, the vowel quality is different ("more pronounced", "sharper"), than if the first syllable is heavy. Balance is often also described as a stress phenomenon: the two syllables of a light word seem to have level stress (Riad, p. 172). Riad analyses balance as a Nordic innovation. Balance dialects have a different structural description of light stems at the foot level, which originates from an alternative interpretation of the bi-moraic condition (Riad, 1992, p. 193). Balance is just another step towards a generalisation of Prokosch's law. One of the first steps was syncope, often making light stressed syllables heavy (for example when CV.CV became CVC). The prosodic structure of balance is described by Riad in the following way:

> "*Suppose now that at some point a re-interpretation of the bi-moraic condition takes place, such that the new generation associate main-stress as such with two moras. At this point, one of (at least) two possible developments would take place. Either lengthening rules are introduced to see to it that a heavy syllable is always provided for the main-stress, or main-stress would proceed to claim two moras for its realization, disregarding syllable structure. The former procedure would be the quantity shift, while the latter would be balance.*" Riad (1992), p. 194

In other words, balance results from level stress, which in turn is the association of an unipositional stress foot with two morae in two different syllables.

4.2.4 The Quantity Shift in Balance Dialects

Balance and non-balance dialects differ in the implementation of the quantity shift. Riad looks upon balance in combination with the previously mentioned and well known observation, that VC lengthening in Northern and Middle Sweden proceeds in a different manner than in Southern Sweden. According to Riad, balance is the cause of the earlier lengthening in mono- than in bi-syllabic words and the reason why Southern (non-balance) dialects lengthen only vowels and not consonants.

The lengthening of (heavy) CVC mono-syllabic words occurs in three steps (see Riad, 1992, p. 301). The first step is "spontaneous" /a/-lengthening. It is different from the other lengthenings as it affects /a/ (and /ɛ/) also in other heavy syllables than CVC (e.g. CVCC) and shows no connection to the quality of the following consonant. The second step lengthens non-low vowels. Here, the quality of the following consonant influences the lengthening process (see above). The third step is consonant lengthening. In the following, these sub-stages of the quantity shift will be disregarded. The focus will lie on the difference between mono- and bi-syllabic stems.

Thus, only two stages are distinguished in Table 4.2. The first stage (B1) shows the derivation in balance dialects when mono-syllabic lengthening occurs but bi-syllabic light stems are still possible (due to balance). Stage B2 shows the derivation in dialects that have lost balance. Initially, balance keeps bi-syllabic stems from lengthening, because balance is an alternative to lengthening (see above). As soon as balance is lost, the bi-syllabic stems are lengthened as well.

It is crucial for Riad's system that the lengthening started in exactly those environments were the mora insertion rule was applied. Again, a re-interpretation is taking place. The lengthening requires that final non-moraic consonants become extra-metrical. As a consequence, a mora contributed by mora insertion cannot be linked to the consonant any more. Hence, the mora goes to the vowel which becomes bi-moraic and surfaces long. At the same time, long-short distinctions in the coda of mono-syllabic words are surfacing again (in other contexts they were never gone), as the difference between underlyingly moraic and non-moraic consonants is not longer neutralised by mora insertion. Thus, as in some other phonological accounts (e.g. Eliasson and La Pelle, 1973), consonant quantity is underlying in Riad's model and vocalic quantity is derived.

Stage B2 in Table 4.2 shows the situation after the quantity shift in balance dialects, when balance is gone. The loss of balance abandons the possibility of a two-positional main stress. Instead, the main-stressed syllable has to contain two morae. The transition is realised by a process called "main-stress retraction" (cf. Riad, 1992, p. 316). Main-stress, previously dominating two syllables, retracts to the initial syllable. This requires the addition of a mora. Riad claims that this process makes no predictions about which segment should receive the "new" mora:

> *"The new mora is not unambiguously assigned to one syllable. Rather it comes into being as main-stress shifts all of its weight to the initial syllable. The formulation of mora-insertion in a balance prosody therefore makes no statement with regard to which syllable should receive the mora."* Riad (1992), p. 317

Thus, main-stress retraction can direct the mora either to the vowel or to the following consonant, thereby triggering either vowel or consonant lengthening (see stage B2 in Table 4.2).

The attraction of the mora is segment dependent and follows obviously Hesselman's laws (see 4.1.1). Again, Riad argues that after the loss of balance only consonant quantity can be underlying. In table 4.2 a theoretical stage D is given, which, as we hope, should illustrate the main effects of underlying vowel quantity. Two effects can be observed. In mono-syllabic words, moraic consonants occur even with underlying vowel quantity, as an effect of mora insertion. However, Riad argues that it is not predicted that these consonants surface long. As can be seen, the underlying contrast lies in the vowel, and on the surface, the contrast is also expressed in the vowel in terms of vowel duration. Thus, durational distinctions in the consonant should be neutralised, as the system expresses the distinction in the vowel. In the bi-syllabic case, underlying vowel quantity would never trigger long consonants.

Table 4.2: Development of quantity in balance and non-balance dialects, following Riad (1992). Underlying form (UL), core syllabification (CS), main-stress assignment (MS) and surface form (SF) are shown. Parentheses indicate unsyllabified segments, squared brackets indicate extrametricality, underlining indicates main stress.

A. Before the quantity shift

UL	CVC	$CV_\mu C$	CVC_μ	CVCV	$CV_\mu CV$	$CVC_\mu V$
CS	$CV_\mu.(C)$	$CV_{\mu\mu}.(C)$	$CV_\mu.(C_\mu)$	$CV_\mu.CV_\mu$	$CV_{\mu\mu}.CV_\mu$	$CV_\mu.(C_\mu).CV_\mu$
MS	$\underline{CV_\mu C_\mu}$	$\underline{CV_{\mu\mu}C}$	$\underline{CV_\mu C_\mu}$	$\underline{CV_\mu}.CV_\mu$	$\underline{CV_{\mu\mu}}.CV_\mu$	$\underline{CV_\mu C_\mu}.CV_\mu$
SF	CVC	CV:C	CVC	CV.CV	CV:.CV	CVC̣:V

B1. During the quantity shift — before the loss of balance

UL	CVC	$CV_\mu C$	CVC_μ	CVCV	$CV_\mu CV$	$CVC_\mu V$
CS	$CV_\mu.[C]$	$CV_{\mu\mu}.[C]$	$CV_\mu.(C_\mu)$	$CV_\mu.CV_\mu$	$CV_{\mu\mu}.CV_\mu$	$CV_\mu.(C_\mu).CV_\mu$
MS	$\underline{CV_{\mu\mu}C}$	$\underline{CV_{\mu\mu}C}$	$\underline{CV_\mu C_\mu}$	$\underline{CV_\mu}.\underline{CV_\mu}$	$\underline{CV_{\mu\mu}}.CV_\mu$	$\underline{CV_\mu C_\mu}.CV_\mu$
SF	CV:C	CV:C	CVC:	CV.CV	CV:.CV	CVC̣:V

B2. After the quantity shift — lost balance

UL	CVC	CVC	CVC_μ	CVCV	CVCV	$CVC_\mu V$
CS	$CV_\mu.[C]$	$CV_\mu.[C]$	$CV_\mu.(C_\mu)$	$CV_\mu.CV_\mu$	$CV_{\mu\mu}.CV_\mu$	$CV_\mu.(C_\mu).CV_\mu$
MS	$\underline{CV_{\mu\mu}C}$	$\underline{CV_{\mu\mu}C}$	$\underline{CV_\mu C_\mu}$	$\underline{CV_{\mu\mu}}.CV_\mu$ or $\underline{CV_\mu C_\mu}.CV_\mu$	$\underline{CV_{\mu\mu}}.CV_\mu$	$\underline{CV_\mu C_\mu}.CV_\mu$
SF	CV:C	CV:C	CVC:	CV:.CV or CVC̣:V	CV:.CV	CVC̣:V

C. After the quantity shift — non-balance dialects

UL	CVC	CVC	CVC_μ	CVCV	CVCV	$CVC_\mu V$
CS	$CV_\mu.[C]$	$CV_\mu.[C]$	$CV_\mu.(C_\mu)$	$CV_\mu.CV_\mu$	$CV_\mu.CV_\mu$	$CV_\mu.(C_\mu).CV_\mu$
MS	$\underline{CV_{\mu\mu}C}$	$\underline{CV_{\mu\mu}C}$	$\underline{CV_\mu C_\mu}$	$\underline{CV_{\mu\mu}}.CV_\mu$	$\underline{CV_{\mu\mu}}.CV_\mu$	$\underline{CV_\mu C_\mu}.CV_\mu$
SF	CV:C	CV:C	CVC:	CV:.CV	CV:.CV	CVC̣:V

D. A “wrong” solution — underlying vowel quantity

UL	CVC	$CV_\mu C$	CVC	CVCV	$CV_\mu CV$	CVCV
CS	$CV_\mu.(C)$	$CV_{\mu\mu}.(C)$	$CV_\mu.(C)$	$CV_\mu.CV_\mu$	$CV_{\mu\mu}.CV_\mu$	$CV_\mu.CV_\mu$
MS	$\underline{CV_\mu C_\mu}$	$\underline{CV_{\mu\mu}C}$	$\underline{CV_\mu C_\mu}$	$\underline{CV_{\mu\mu}}.CV_\mu$	$\underline{CV_{\mu\mu}}.CV_\mu$	$\underline{CV_{\mu\mu}}.CV_\mu$
SF	CVC	CV:C	CVC	CV.CV	CV:.CV	CV:.CV

4.2.4.1 Underlying Consonant Quantity vs. Underlying Vowel Quality

The question why consonant quantity is suggested to be underlying in Riad's account is of primary importance for the present study, as phonetic evidence points towards the vowel as the main bearer of distinctive cues. Riad bases his argument on grammatical distinctness:

> *"Grammar requires that the underlying distinction remains expressed throughout the derivation, if it is expressed on the surface. If distinctness is lost at some point, it cannot be retrieved at a later point, and neutralisation is predicted. We will interpret grammatical distinctness in the most straightforward way, in that we will say that the distinction must be retrievable where it is made. That is, if the underlying quantity distinction is given to consonants, then it is the consonant that should express the distinction on the surface."* Riad (1992), pp. 294–295

In our interpretation, this would mean that the moraicity of the C alone after short vowels cannot trigger long consonants, as the distinction between non-moraic and moraic consonants in this position is lost during the derivation.

4.2.5 The Quantity Shift in Non-Balance Dialects

Stage C shows the situation in non-balance dialects after the quantity shift (see below). Here, the situation is less complicated, as the lengthening processes are not influenced by balance. Riad argues that there is no evidence for the delayed lengthening of bi-syllabic stems in non-balance dialects. Furthermore, non-balance dialects lengthen (almost) always the vowel. The re-interpretation of the bi-moraic condition leads directly to the bi-moraicity of the main-stressed syllable.

4.2.6 The Effect of the Quantity Change on Overlong Syllables

In the phonological system before the quantity shift, V:C: sequences are a logical consequence of independent vocalic and consonantal length distinctions. This "true" overlength is distinguished by Riad from "false" overlength. False overlength is the combination of a long vowel and a consonant cluster. True overlength combines three morae in a syllable, while false overlong syllables contain only two morae. In the following, only true overlength will be considered and the term "overlength" will only be applied to sequences of two long segments.

Riad assumes that overlong sequences are lost at a late stage of the quantity shift (see Riad, 1992, p. 244). This is due to the fact that overlong syllables do not violate Prokosch's law, unlike light syllables. However, the breakdown of segmental quantity also affects the overlong syllables. The loss of underlying vowel quantity causes the neutralisation of vocalic quantity in former V:C: sequences. This would predict a shortening of the vowels (and not of the consonant), which is, in fact, the usual change affecting overlong stems. Riad notes that overlong sequences did not only exist in certain stems before the quantity shift, but also arose in the regular derivations of certain verbforms (see Riad, 1992, p. 342).

While Riad's model correctly predicts that the loss of overlong sequences is mainly a consequence of vowel shortening, it is unclear, in our view, how the suggested process should explain the fact that many dialects abandoned the overlong sequences but kept the short sequences (see e.g. Ivars, 1988 for the situation in the Finland-Swedish dialects). Short sequences are generally more often reported than overlong sequences in the dialects. Thus, it is quite likely that overlong sequences are lost before short sequences. This corresponds also to the typological observation mentioned in Section 6.2.2, which reported an implicational correlation between the existence of long consonants after long vowels and short vowels. Long consonants after long vowels imply the existence of long consonants after short vowels, and in a range of languages, long consonants occur only after short vowels. This supports the assumption that the loss of V:C: before VC is the usual one.

4.3 Conclusion

In this chapter the processes occurring during the Swedish quantity shift were described, on the basis of an account in the framework of generative metrical theory. It was shown that the generative formalism, as developed by Riad (1992), claims underlying consonantal quantity for the modern Swedish quantity system, as only underlying consonant quantity can predict complementarity. Underlying vowel quantity would predict the neutralisation of consonant contrasts, resulting in V:C versus VC, and not V:C versus VC:.

In this context it is interesting to look at the parallels between Riad's account and the syllable-cut hypothesis by Vennemann (2000), which was described in Chapter 2. It is obvious that both accounts try to tackle similar problems. Both, Riad and Vennemann, are confronted with the problem that the quantity change did not affect only light syllables, although the outcome of the change results in obligatory heavy syllables in stressed position. Both accounts abandon the idea of underlying vocalic quantity, but in different ways. This difference is clearly motivated by a difference in the surface structure between modern Standard Swedish and modern Standard German (and other Germanic languages). Swedish shows a difference in consonant duration on the surface that accompanies differences in vocalic duration in a complementary manner. This provides the theoretical possibility to ascribe the underlying contrast to the consonant. For German, this possibility does not exist. But, on the other hand, a combination of a short vowel and a following consonant should be treated as heavy in German as well, otherwise the attraction of stress by these syllables would remain unexplained. CVCV is not syllabified CV.CV in German, but CVC̣V, with a short, but bi-syllabic consonant. The syllable is abruptly cut, according to Vennemann. This structure evolved from the same pre-quantity-shift structure as in Standard Swedish, that is from a prosodic structure with vowel and consonant quantity in stressed position. Vennemann's description of the change shows clearly the parallels to especially the balance dialects in Sweden. The phonetic contexts that favoured abrupt cut are very similar to the contexts where balance dialects favoured consonant lengthening. The clear difference between German and Swedish seems to lie in the fact that de-gemination has occurred in German, while Swedish has kept the long consonants.

Riad mentions de-gemination in connection with Danish. Danish, with the exception of East Danish, lengthened the vowels of *mostly* bi-syllabic stems during the quantity shift. As Danish does not have complementary quantity, Riad assumes that Danish retained underlying vowel quantity and abandoned consonant quantity with the effect that previously long consonants were shortened. Riad interprets the old CVC:V forms after de-gemination as forms with light syllables. Regardless of the situation in Danish, it is interesting to note that de-gemination does not necessarily lead to new short syllables. Vennemann's model suggests de-gemination in German, but not the development of light syllables by this process.

Chapter 5

Quantity in the Swedish Dialects

The historical section laid out the general characteristics of phonological quantity distinctions in Swedish and its dialects. The historical development can be viewed as leading from a system with four possibilities to combine long and short vowels and consonants (V:C:, VC, V:C and VC:), over a system that had three possibilities (VC, V:C and VC:) to a system with only two (V:C and VC:). As not all Swedish dialects have completed these changes, the three different systems still exist in some dialects.

In this chapter, the dialectological sources will be surveyed for evidence of the different systems. The main focus will lie on those dialects where recordings from the SweDia corpus are available. For an initial dialect classification, Wessén's dialectal areas (Wessén, 1969) will be used. For easier reference, systems with all four possible combinations of long and short vowels will be called 4-way systems, those with long consonants only after short vowels will be called 3-way systems and those with complementary quantity will be called 2-way systems.

5.1 Wessén's Dialect Areas

Wessén (1969) has divided the Swedish dialects into six regions: Sydsvenska mål (Southern Swedish varieties), Götamål (Göta varieties), Sveamål (Svea varieties), Norrländska mål (Northern Swedish varieties), Östsvenska mål (Finland-Swedish varieties) and Gotländska mål (Gotlandic varieties). This division is based on linguistic criteria. Figure 5.1 shows a map with the approximate geographical distribution of the six regions. In the following, mainly the differences with respect to phonological quantity will be discussed. Other dialectal differences will only be mentioned if they are relevant for the subject matter. Special attention will always be payed to those locations within a certain region that were also recording locations in the SweDia project (see Sections 7.2 and 7.3.3 for a detailed description of the SweDia corpus, including maps).

Additional subdivisions of the dialect areas will largely be based on the Swedish provinces ('landskap'). These provinces are traditional administrative and cultural regions of Sweden and are often used as a basis for dialectal divisions which are not purely based on linguistic categories (e.g. Pamp, 1978).

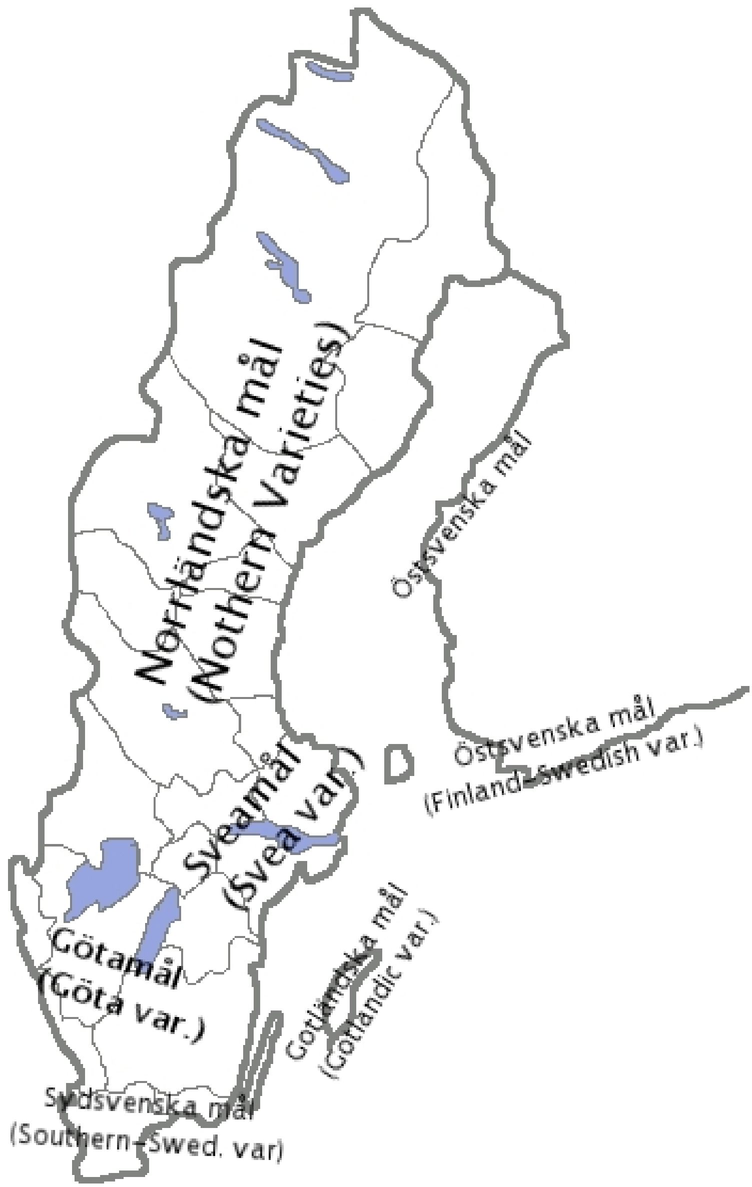


Figure 5.1: Approximate geographical positions of Wessén's dialectal regions. After Wessén (1969), p. 14.

5.2 Finland-Swedish Varieties

The Finland-Swedish varieties comprise two geographically distinct areas in Finland (see Figure 5.1). Ten SweDia recording locations belong to this region and nine of them were investigated in this thesis[1]. Dialectological descriptions of the dialects in this area show that dialects with 4-way distinctions still seem to be frequent.

Subdivisions of the dialects in this area are usually made by using traditional administrational regions, comparable to the Swedish provinces. These regions are "Österbotten", comprising the SweDia recording locations Vörå (vor) and Närpes (nap), Åland, comprising Saltvik (sal) and Brändö (bra), "Åboland", comprising Houtskär (hou) and Dragsfjärd (dra) and "Nyland", comprising Snappertuna (sna), Kyrkslätt (kyr) and Borgå (bor).

One recent work on the dialects in the Finland-Swedish region is Ivars (1988). Her work focusses on the dialect of Närpes, but also gives an overview of other dialects, especially with respect to their quantity system. Ivars (1988), p. 67, provides a map of those Finland-Swedish locations where the V:C: successions still exist. The map indicates that five of the nine investigated recording locations (Vörå, Dragsfjärd, Snappertuna, Kyrkslätt and Borgå) should show V:C: successions.

It is important to note that in these dialects the so-called overlength is not only found in some mono-morphemic words, but also results from regular inflections in at least two paradigms. The neutral form of adjectives ending in a long vowel or a long vowel + /t/ is marked by a geminated /tː/ while the long vowel is retained (see Freudenthal, 1889, p. 58 for the Vörå dialect). The neutral form of the word /viːt/ is thus /viːtː/, for example, and not /vɪtː/, as in Standard Swedish.

Overlength occurs as well in the third, weak conjugation of verbs. The perfect participle and even the supine are formed with -d, which leads to V:C: successions in forms like *bodd* (perfect participle of *bo*, 'to reside'). Again, the dialects keep the long vowel of the uninflected form, where Standard Swedish replaces the long vowel with a short vowel (see Ivars, 1988, p. 66 and Freudenthal, 1889, p. 48). Ivars (1988) suggests that the development of overlong syllables into long syllables is an innovation that reached southern Österbotten via Åland and surrounding areas, including Houtskär in Åboland (p. 64).

Ivars (1988) describes the dialect of Närpes in some detail. According to Ivars the Närpes dialect has lost the overlong successions, but this loss has happened quite recently in historical terms. Freudenthal (1878) reports that the dialect still has V:C: successions in some forms. Freudenthal mentions the word *flat*, which becomes /flaːtː/ in the neuter form, but refers to this process as "exceptional". This indicates that the gemination of the final /t/ without the shortening of the vowel is not the rule anymore towards the end of the 19th century, but has still survived in some word forms. According to the reports in Ivars (1988), the loss of the V:C: sequences has been completed by the end of the 20th century (see p. 66). This estimation is based on recordings of dialectal speakers. Furthermore, Ivars is herself a native speaker of the Närpes dialect (see pp. 44–45). VC successions, however, are a regular feature of the Närpes dialect. Ivars (1988), p. 60 provides several examples of minimal pairs where VC successions contrast with V:C or VC: successions.

Concerning the other Finland-Swedish dialects, the work of Hesselman (1902) has been mentioned before. Riad's summary of this work (1992, pp. 272–274) showed that the lengthen-

[1] The tenth, excluded recording location was Munsala in Northern Österbotten.

ing of the short syllables in the Finland-Swedish dialects[2] has progressed to different extents. Dialects in Nyland and Österbotten have both lengthened VC successions in some contexts, but kept the short syllables in many others. For Åland, however, Hesselman (1902) comes to the conclusion that the dialects in this region have fully implemented the quantity shift (see p. 98 and Riad (1992), p. 273). These dialects should thus have established a 2-way-system. This is also reported by Harling-Kranck (1988), who also mentions that the so-called “Skiftet”, the waterway between Åland and Åboland, marks the border for the total loss of VC sequences. Sundberg (1993) provides a detailed account of the dialects in a part of Åland, which also includes a recording location in the SweDia project, namely Brändö. Her reports on the quantity pattern in this region state on the one hand, that phonematic length is the same as in Standard Swedish (p. 132). On the other hand she reports that voiceless consonants after long vowels (not dipthongs) in medial and final position are longer than in Standard Swedish and chooses the transcription [V:KK] for these forms (p. 131). Her judgements were made by auditory inspection; acoustic measurements were not performed.

Concerning Åboland, Zilliacus (1992) reports the occurrence of V:C: successions for the Eastern parts, including Dragsfjärd (see p. 20). For Houtskär, V:C: is not reported, but VC successions in bi-syllabic words. Table 5.1 provides an overview over the systems in the respective Finland-Swedish recording locations, according to the literature.

Table 5.1: Occurrence of short and overlong sequences in Finland-Swedish dialects.

Recording location	VC	V:C:
Vörå	+	+
Närpes	+	–
Brändö	–	–
Saltvik	–	–
Houtskär	+	–
Dragsfjärd	+	+
Snappertuna	+	+
Kyrkslätt	+	+
Borgå	+	+

5.3 Northern Swedish Varieties

The Northern Swedish varieties cover an area of eight provinces, from the Northern parts of Hälsingland in the South[3] to Lappland and Norrbotten in the North, including the Eastern provinces of Härjedalen, Jämtland and Lappland. Nineteen recording locations within the Northern Swedish varieties have been investigated in the present study.

[2] Hesselman (1902) mainly describes dialects in Österbotten, Åland, Nyland and the Eastern parts of Åboland, which should include Dragsfjärd. The Houtskär dialect is not included.

[3] In accordance with Wessén (1969), the province Gästrikland and the Southern parts of Hälsingland were classified as “Sveamål” in the present study.

As mentioned before, the Northern Swedish dialects comprise an area where the existence of VC sequences is a widespread phenomenon. In the following, an attempt will be made to find evidence for the existence of VC sequences by looking for positive as well as negative reports in the literature. Pamp (1978), p. 29, reports “short syllables” for parts of Norrbotten, Ångermanland and Jämtland. Söderström (1972), p. 9, lists areas with VC successions, which should include the following SweDia recording locations: Nederkalix (nka), Nederluleå (nlu) and Piteå (pit) in Norrbotten, Åre (are) in Western Jämtland, Ragunda (rag) in Eastern Jämtland and Bjurholm (bju) in north-eastern Ångermanland. A similar list is given by an earlier source, Noreen (1907), pp. 118–119. For Bjurholm, VC is also reported in Wallström (1943) p. 43. The same source reports VC for Arjeplog (p. 41). Furthermore, the absence of VC is reported for Västerbotten, and for Ångermanland, with the mentioned exception of Bjurholm. However, Pamp (1978), p. 134, reports VC also for Anundsjö in Ångermanland, and Noreen (1907) reports the existence of VC for some locations in Västerbotten. The Swedia recording location Nysätra is mentioned explicitely. This observation is dated to the year 1860. Vestlund (1923), p. 32, reports the absence of VC in Medelpad and confirms the existence of VC in Eastern Jämtland (Ragunda). For Torp in Medelpad the absence of VC is also reported by Bogren (1921), p. 4. For Vilhelmina in Lappland, Dahlstedt (1962), p. 7, reports the absence of VC sequences.

Thus, for 13 recording locations in the Northern Swedish area explicit statements about the existence or the absence of VC sequences were found. In only one case the reports were inconsistent, which could be due to the great difference in time between the two reports. For the remaining six recording locations, no explicit statements about the occurrence of VC were found. From the general statements about the existence of VC in certain regions or provinces, some assumptions can however be made. According to the reports by Pamp (1978) and Söderström (1972), VC can be expected with a greater probability in Jämtland than in Hälsingland, for example. Table 5.2 provides a list of the 19 recording locations in the Northern region and indicates the existence of VC sequences. In cases where the possible existence was only derived from more general reports, and in cases were the reports were inconsistent, a question mark has been added.

Table 5.2: Occurrence of short sequences in Northern Swedish dialects.

Landscape	Recording location	VC
Norrbotten	Nederkalix	+
	Nederluleå	+
	Piteå	+
Lappland	Arjeplog	+
	Vilhelmina	–
Västerbotten	Burträsk	–
	Nysätra	+/–?
Ångermanland	Bjurholm	+
	Fjällsjö	+?
	Anundsjö	+
Jämtland	Åre	+
	Aspås	+?
	Ragunda	+
	Berg	+?
Medelpad	Indal	–
	Torp	–
Härjedalen	Vemdalen	–?
	Lillhärdal	–?
Hälsingland	Färila	–?

5.4 Svea Varieties

The Svea varieties cover the area of Uppland, Gästrikland, Southern Hälsingland, Dalarna, Södermanland, Närke and Västmanland. The dialects of Värmland have also been included in this group, although Wessén (1969) groups them together with the Göta varieties. For several locations in this area, especially Uppland, VC has been reported. Hesselman (1905), pp. 7–8, mentiones Western and South-western Uppland and the neighbouring areas of Västmanland. Furthermore, some smaller areas in Eastern Uppland are mentioned and Sorunda (soa) in Södermanland, which is a SweDia recording location.

5.4.1 The Dialects of Dalarna

Dalarna is known as a region with dialects that in many ways differ considerably from Standard Swedish, and that also show great variation among themselves. Pamp (1978) divides Dalarna into three different dialect areas: first, a Northern area with a strong Norwegian influence; second, the area around the lake Siljan, where the typical "Dalmål" is spoken, and third, a South-Eastern area, with dialects called "Dalabergslagsmål". As especially the Dalmål-dialects show great variation among themselves, they are further subdivided into "Ovansiljansmål" (Western Siljan area), "Nedansiljansmål" (Eastern Siljan area) and "Västerdalmål" (South-west of the other two areas, along the river "Västerdalälven").

A detailed description of the dialects called "Dalmål" can be found in Levander (1925) and Levander (1928). One interesting feature of these dialects is their conservation of V:C: and VC sequences, similar to the Finland-Swedish dialects. Even in these dialects, V:C: sequences are found in regular derivations, for example in some weak verb forms in the third conjugation (see Levander, 1925, p. 342). The existence of V:C: sequences is mainly described for three locations (see Levander, 1925, pp. 75–77): Älvdalen and Sollerö (Western Siljan area), and Öje (Västerdalmål). Furthermore, incidental appearances of V:C: sequences in certain words are reported for several other locations.

As with other dialect areas, the occurrence of VC successions is reported for a larger area. Levander (1925) reports them for the Western Siljan area and parts of the Eastern Siljan area. The VC forms are much less frequent in mono-syllabic words, although there are areas which have obviously retained those forms (see Levander, 1925, pp. 65–74). Outside the two mentioned areas, there are several areas which have, by and large, lost the VC forms. The descriptions in Levander (1925) indicate that dialects retaining V:C: have also retained VC, albeit to different extents.

For the Dalarna area, the match between the dialectological descriptions and the recording locations is quite difficult. Quantity data exists from four locations: Särna (sar), Malung (mal), Leksand (lek) and Husby (hus). Särna is situated in Pamp's "Norwegian region" (see above). Unfortunately, Pamp makes no comments about the quantity system of these dialects. Malung lies in the "Västerdalmål" area, close to Öje, for which V:C: structures were reported. Leksand lies in the Eastern Siljan area, but not within the main VC area reported by Levander (1925). Husby is situated in the "Dalabergslagsmål" area. Consequently, for none of these recording locations a definite source about the quantity structure has been found. V:C: structures are unlikely in all of these recording locations, with the possible exception of Malung. VC structures are more likely, but could, as mentioned, not be confirmed for any of these locations. For one dialect in the Western Siljan area, Älvdalen, there is, however, a recent survey of quantity data available (Nyström, 2000), which was made independently of the SweDia project. The results of this investigation are reported in the next section.

5.4.1.1 The Quantity System of the Älvdalen Dialect

Nyström (2000) presents a phonetic investigation of V:C:, V:C and VC: sequences in "Älvdalsmål", one of the mentioned dialects around the Western parts of lake Siljan. Nyström mentions the existence of VC sequences in this dialect, but they were not included in his material. His results regarding the segment durations are of great interest for the results of the present study, and are therefore reported here in some detail.

Nyström investigated five male speakers who were born between 1911 and 1929. The material consisted of several mono- and bi-syllabic test words, recorded in carrier sentences in focussed position, both in medial and final position. Comparable to the Finland-Swedish forms, the V:C: forms in Dalmål often occur in morphologically complex words, e.g. with certain case markings, which are retained in Dalmål but have been lost in Standard Swedish. Nyström presents summarised data for all test words in terms of mean values. The results of three test words are reported in more detail, though only for one speaker. For the latter data also the distribution of the durational values for about 50 repetitions is reported. These three words were mono-syllabic words, with a coda consisting of a high, back, rounded, slightly diphthongised vowel and an apical, dental or alveolar, nasal. Table 5.3 shows the mean values and ranges of values for the three words, as estimated from Figures 1a, 1b, 3a and 3b in Nyström (2000).

Table 5.3: Mean values of vowel and consonant durations in Nyström (2000).

Sequence	V duration (range)	C duration (range)
V:C	310 (260–390)	90 (70–115)
VC:	165 (130–190)	210 (160–280)
V:C:	260 (225–375)	170 (135–230)

Nyström emphasises the great durational differences between the consonants in V:C and VC: successions. C: can be more than twice as long than C, a considerably higher ratio than in Standard Swedish, for which Nyström reports an average ratio of 1.3 (derived from Elert, 1964). Another stable effect are the shorter V: and C: durations in V:C:, compared to the respective segments in V:C and VC:. Nyström interprets his data mainly in terms of relative durations. The relative length of vowels and consonants (as a percentage of the whole sequence) is kept apart by the speaker to mark the phonological identity for the listener. In Section 2.6.2, it was argued against similar accounts and hypotheses for complementary quantity, and similar arguments can be raised here against Nyström's assumption. First, Nyström's data clearly indicate (see e.g. Figure 4 in Nyström, 2000) that the long segments have a greater elasticity than the short segments. This points towards considerable changes in relative durations for different speech rates or stress conditions. Second, the Älvdalen dialect also shows VC successions. It is unclear how these sequences should be distinguished from V:C: sequences on the basis of relative durations within the sequence.

5.5 Göta, Southern Swedish and Gotlandic Varieties

For the dialects South of the Svea varieties, the dialectological literature reports only V:C and VC: sequences, that is 2-way distinctions (see Pamp, 1978; Wessén, 1969). Therefore, a detailed description of the dialects in this area is not given here. As mentioned before, the lengthening of VC sequences in these dialects took place in a different way than in the Northern and Svea varieties. In the Northern varieties, either the V or the C in VC forms was lengthened (if the VC sequence was not retained). In the Göta and Southern Swedish varieties, mostly the V was lengthened. Consequently, the frequency of VC: successions should be higher in the North than in the South, as the Northern and Middle-Swedish dialects gained VC: successions during the quantity shift, contrary to the Göta and Southern varieties. Some sources (e.g. Malmberg, 1944) have claimed that the dialects in Skåne use only vocalic quantity, that is only a V:C vs. VC contrast. The correctness of these statements is hard to estimate, as phonetic measurements are almost totally lacking. An exception is an investigation from Hansson (1969) who investigated durational patterns of speakers from a location in South-Western Skåne (Östra Ingelstad). His measurements show longer consonants after short vowels than after long vowels (C:/C ratio about 1.25), indicating complementary quantity. Thus, it is unclear whether there are dialects in Southern Sweden that do not have complementary quantity. The results of the present study should shed some light on this issue.

Another sound change with potential effects on the quantity patterns in Göta and Southern Swedish dialects is connected with voicing. In a South-West Swedish area, comprising parts of the landcapes of Blekinge, Skåne, Halland and Bohuslän, the voiceless plosives /p/, /t/ and /k/ have become /b/, /d/ and /g/ after long vowels (see e.g. Pamp, 1978; Pettersson, 1996). This change is generally viewed as an influence from Danish.

5.6 Preaspiration in Swedish Dialects

Preaspiration, that is the occurrence of glottal fricative noise between vowels and consonants — usually voiceless stops — has been described for some dialects in Sweden. Helgason (1999) provides a detailed survey of preaspiration in the Nordic languages, including the Swedish dialects. He reports preaspiration mainly for three geographically distant locations: some dialects in Härjedalen and Northern Dalarna, the dialect of Arjeplog in Lappland and the dialect of Gräsö, an island on the North Eastern coast of Uppland[4]. The three regions show differences in the distribution of preaspiration. In the Härjedalen/Northern Dalarna area, preaspiration is mainly found in long voiceless stops after short vowels. The dialect of Arjeplog shows stronger preaspiration after long than after short vowels (see Wretling et al. (2002); Helgason, 1999). In the Gräsö dialect, preaspiration is reported after long and after short vowels.

Helgason (1999) reports preaspiration to be a wide-spread phenomenon even in Standard Swedish and makes a distinction between normative and non-normative preaspiration. The term "normative" is defined in the following way:

> *"If the absence (or presence) of a particular phonetic trait leads to a pronunciation that is considered deviant by the speakers of a given dialect, that trait*

[4] For a fourth region near Gräsö, Åland and Åboland, a group of islands close to the South Western Finnish coast, preaspiration is reported from sources at the end of the 19th century, but has obviously ceased to exist (see Helgason, 1999, p. 86)

> *can be classified as normative (or normatively absent) in that dialect.*"Helgason, 1999, p. 21

SweDia recording locations with potentially normative preaspiration include Vemdalen (vem), Lillhärdal (lil) and Särna (sar) in Härjedalen and Northern Dalarna, Arjeplog (arj) and Gräsö (grs).

Non-normative preaspiration in Swedish dialects has already been investigated on the basis of the SweDia material. A study by Tronnier (2002) showed frequent preaspiration in Southern Swedish dialects. For Northern Sweden, similar observations have been made by Wretling et al. (2002).

5.7 Conclusion

The dialectological literature describes 4-way, 3-way and 2-way distinctions for the dialects of distinct geographical areas. In general, 4-way distinctions have been described for the Finland-Swedish area and Dalarna, 3-way distinctions for various locations in Northern Sweden, while the Southern Swedish dialects (Göta varieties and Southern Swedish varieties) have only 2-way distinctions. In the second part of this study, we will compare the geographical distribution derived from the dialectological literature with the distribution derived from a data-driven categorisation of durational measurements from various dialects across the Swedish-speaking area.

Chapter 6

Typological Aspects

6.1 Typologies for Phonological Quantity

An attempt to typologise the use of quantity in the world's languages has been made by Bannert, 1979 (see also Bannert, 1976), which is given here in the version of Bruce (1998).

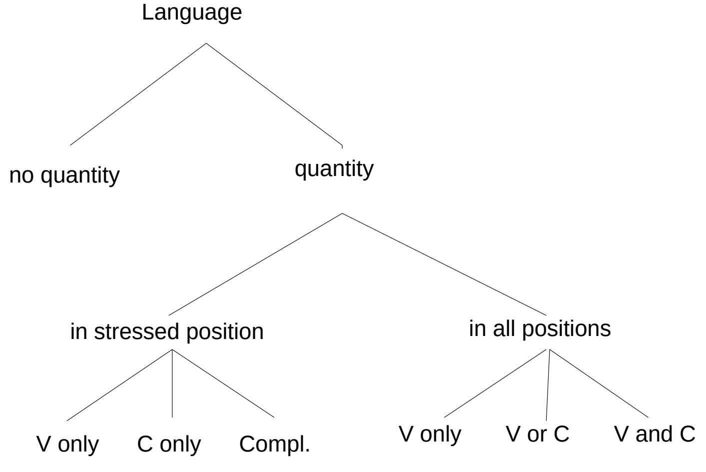


Figure 6.1: The quantity typology of Bruce (1998).

As the scheme in Figure 6.1 shows, seven different quantity types are suggested. The main distinction is made between languages with and without quantity. Quantity languages are then divided into languages that use quantity only in stressed position versus languages that use quantity in all positions. Languages using quantity in stressed position are subdivided into languages which use only vocalic quantity, languages which use only consonantal quantity and languages with complementary quantity. Languages with quantity in all positions are subdivided into languages using only vocalic quantity, languages using vocalic *or* consonantal quantity and languages using vocalic *and* consonantal quantity.

The typology excludes certain logically possible types: There are no languages using V and C, but only in stressed position, there are no languages using only consonantal quantity, but in all positions, and there are no languages with complementary quantity in all positions.

6.2 Quantity in the World's Languages

6.2.1 Vocalic Quantity

The most extensive database of sound inventories is probably the UCLA Phonological Segment Inventory Database, UPSID (see e.g. Maddieson, 1984). Unfortunately, the structure of this database is only partly suitable to analyse languages with respect to vocalic quantity, as vocalic quantity was marked in UPSID only under certain circumstances:

> *"Vowel length contrasts . . . are only recorded as phonemic in UPSID if they are linked to vowel quality differences. (. . .) There are three types of situations where vowel length is represented in the phoneme inventory. In some languages the long and short vowel sets do not overlap in quality; (. . .) More commonly, some vowels in each set have the same quality (. . .) The third situation in which length is represented is where the qualities of the longer vowels are a subset of the qualities of the shorter vowels or vice versa . . .* Maddieson (1984), p. 128–129

In other words, in all cases where languages show a quantity contrast without a quality contrast and the number of short and long sounds is equal, the information about the quantity contrast is missing in UPSID.

On the basis of the UPSID data, Maddieson (1984) concludes:

> *"The probability of length being part of the vowel system increases with the number of vowel quality contrasts. (. . .) We may speculate that there are two diachronic factors responsible for this trend; languages with an originally suprasegmental vowel length contrast may begin to add quality differences to the quantitative difference . . . , and languages with large numbers of qualitatively distinct vowels may begin to recruit length differences to additionally distinguish them . . . In either case the outcome is the same — the use of combinations of durational and qualitative differences to mutually reinforce the distinctiveness of vowel contrasts."* Maddieson (1984), p. 129

In the light of the recording convention of the UPSID data base, this conclusion has to be evaluated with some caution. It is based on a recording convention that excluded the most typical type of a quantity language (see below). The question is whether quantity contrasts just have a higher chance to be recorded in UPSID with larger vowel systems, as with a higher number also the probability increases that the short vowels are a subset of the long vowels or vice versa.

Crothers (1978) provides a typological analysis of 208 languages. Here, the quantity contrast was always recorded. Crothers' analysis focusses on vocalic quality, but the inventory listings are thorough enough to provide important information about quantity contrasts and accompanying tenseness contrasts as well. The material was collected within the 'Stanford

Phonology Archiving Project'[1]. The material is described as areally and genetically balanced, thus providing a certain degree of representativeness. In the following, some generalisations are presented that were derived from a re-analysis of the data published in Crothers (1978).

100 languages (around 48%) in the sample show a vocalic quantity distinction[2]. 42 of these languages are described with inventories that have durational distinctions without accompanying quality differences. The "non-quantity" languages are only slightly more frequent than the quantity languages (totally 108). There is no evidence in the sample that the use of quantity increases with the number of vowels in a system[3]. The material shows 17 quantity languages with three vowels, compared to six non-quantity languages. Of the 64 five vowel systems exactly half of the languages use quantity. In the seven vowel systems, 16 languages do not use quantity and 12 use it. In the systems with more than seven vowels (which are about 15% of all systems in Crothers' material), 12 languages are quantity languages and 19 are not. Thus, if there is any tendency, then the use of quantity seems to decrease with the number of vowels in a system.

Crother's material also allows estimations of the frequency of "pure quantity" systems and of systems that have combined quality and quantity contrasts. 42 languages are described as showing only quantity distinctions with no accompanying quality differences. 30 of these languages (71%) show entirely parallel systems with the same inventory for long and short vowels[4]. It is possible that quality differences between long and short vowels are sometimes simply overlooked in descriptions, so that the number of languages which fit into this category may be somewhat lower. Nevertheless, the number of languages in this category is too high to ascribe their existence solely to imprecise descriptions.

About 20–30 languages in Crother's material can be described as having a "mixed system", that is languages which are described as having both purely durational and combined durational-qualitative contrasts. The precise number is hard to estimate, because some languages show only a quantity contrast for one vowel, and others have long and short qualities that make it rather difficult to relate long-short pairs to each other. In spite of these difficulties it can be stated that a considerable number of languages use systems similar to Standard Swedish, where pure durational contrasts occur together with combined durational and quality contrasts. This is additional evidence for the observation made in Chapter 3: a strict separation between a tense-lax system and a long-short system is often impossible, making a description of languages using these two features problematic.

6.2.2 Consonantal Quantity

The data in Crothers (1978) show that vocalic quantity contrasts are quite common in the world's languages. The number of languages that show consonantal quantity contrasts seems to be considerably lower. Décsy (1988) (again cited in Holt, 1997) estimates that 14% of the world's languages have a consonant quantity contrast.

Thurgood (1993) gives an overview of some universal aspects of long consonants. His conclusions may be summarised in the following way:

[1]Much information in UPSID was also based on this material.

[2]A similar estimate is given in Décsy (1988), cited in Holt, 1997

[3]For this calculation, the concept is adopted that long and short vowels are variants of one vowel.

[4]In UPSID, these languages are not described as showing a quantity distinction, according to the definitions of UPSID described above.

- Long consonants occur more often after short vowels. This seems to be implicational: If there are geminates after long vowels in a language, there are also geminates after short vowels in a language. In a number of languages, long consonants occur only after short vowels (see Thurgood (1993), pp. 129–130).
- Long consonants tend to occur after stressed vowels.
- Long consonants tend to occur intervocalically.
- Long consonants are often ambi-syllabic.
- Universal voicing characteristics of sonorants and obstruents apply to consonant quantity, as well. Voiceless long obstruents are more frequent than voiced long obstruents, and voiced long sonorants are more frequent than voiceless long sonorants.
- Long stops are more frequent than long affricates and fricatives.
- Long consonants imply the existence of short consonants at the same place of articulation. If a language has consonants with and without quantity distinctions, the consonants without quantity seem to be short.
- Long fricatives are most often alveolar sounds, followed by palatal, labiodental, velar and pharyngeal, and glottal sounds. Long nasals are mostly alveolar, followed by bilabial sounds, then followed by palatals and velars, with a weak preference for palatals. Other sound types (semi-vowels, stops) do not show clear patterns with respect to place of articulation.

6.3 Quantity in Swedish and its Dialects — A First Typology

If the dialectological evidence presented in the Chapter 5 is combined with the phonological and typological considerations in this chapter and Chapter 2, the following categorisation of the dialects is possible: In parts of the Finland-Swedish area and in Dalarna a type is found that shows vocalic and consonantal quantity in stressed position (called “4-way distinction” in the present study). In other parts of the Finland-Swedish area and in parts of Northern Sweden a type is found that also has vocalic and consonantal quantity in stressed position, but consonantal quantity is phonotactically restricted: long consonants (and thus a long/short distinction) occur only after short vowels (called “3-way distinction” in the present study). These two types constitute an extension of the typology suggested by Bannert (see Bannert, 1976, Bannert, 1979) and Bruce (1998) (see Section 6.1), where “V and C” in stressed position was not included. The third type attested in the dialects and Standard Swedish is the complementary quantity type (called “2-way distinction” in the present study), and some authors (e.g. Malmberg, 1944) claim the existence of a fourth quantity type in Southern Sweden with only vocalic quantity.

The 4-way, 3-way and 2-way distinctions are usually also viewed as different stages in the historical development. The 4-way distinction preceded the 3-way distinction which preceded the 2-way distinction. In general, the diachronic analysis suggests at least two more types, an old Swedish type with vocalic and consonantal quantity distinctions in stressed and unstressed position (which is already included in the Bannert/Bruce typology), and a type with vocalic quantity in stressed position and consonantal quantity in all positions, which is not yet included in the Bannert/Bruce typology. These two types are however only attested from historical sources, not from dialect data. In the following, the discussion will therefore be restricted to the types attested in the dialects.

It is evident that linguistic typologies are dependent on the linguistic framework. The Chapters 2 and 3 described a variety of possibilities to analyse quantity distinctions. Trubetzkoy's analysis would for example lead to a typology with mora- and syllable-counting languages, and a sub-type of the syllable-counting languages would be syllable-cut languages. In Vennemann's system, the typological difference would mainly lie between the quantity and the syllable-cut languages. The difference between these two types is mainly constituted by the existence of light stressed syllables. Light stressed syllables occur if a language (or dialect) allows short vowels in open syllables. According to this definition, 4-way and 3-way distinctions would constitute quantity languages, while a language or variety with a 2-way distinction would constitute a syllable cut language. Vennemann's theory also includes a phonetic hypothesis: quantity languages show smooth cut after short vowels, while syllable-cut languages show abrupt cut after short vowels. Acoustic investigations that have tried to find evidence for this hypothesis have up to now been largely unsuccessful, but certain results of articulatory investigations have been interpreted as evidence for the syllable-cut hypothesis. There is, however, a general lack of articulatory investigations that directly compare quantity and syllable cut languages (in Vennemann's sense).

Part II

Empirical Studies of Quantity in Swedish Dialects

Chapter 7

Introduction, Material and Method

7.1 Aims and Structure of the Empirical Part

The first part of this thesis placed phonological quantity in a phonological, phonetic and diachronic context. It was shown how the Modern Standard Swedish complementary quantity system has developed from a vowel and consonant quantity system that was not restricted to stressed positions. Typologically, this system was similar to the quantity system of Modern Standard Finnish.

The changes that have lead to the present systems started with the loss of quantity distinctions in unstressed positions. This change has, as far as we know, been completed in all dialects. Later, V:C: and VC sequences were abandoned in stressed positions, but these changes have not been implemented in all dialects.

Chapter 5 provided a detailed account of the occurrence of V:C: and VC structures in the dialects represented in the SweDia database. As there exist at least three different types of quantity systems in the present Swedish dialects, it is to be expected that these differences should also be reflected in the duration data that can be extracted from the corpus. Explorative studies by Schaeffler and Wretling (2003) and Strangert and Wretling (2003) have found evidence in support of this assumption. The following part of the thesis will therefore investigate the relationship between duration data in the SweDia quantity corpus and descriptions of quantity found in traditional, dialectological studies. Possible parallels between a data-driven categorisation of the dialects based on duration data and traditional dialectological categorisations will be investigated.

In addition to duration, two other factors, vowel quality and preaspiration, which often appear as correlates to quantity will be studied. Vowel quality has played an important role in discussions of Swedish quantity. Phonetic investigations have tried to establish whether spectral or durational cues are primary for the perception of quantity distinctions. Phonological studies have investigated how vowel quality and duration should be treated in a phonological description of quantity. Historical studies, in addition, have tried to explain changes in the Swedish quantity system as caused by changes in vowel quality.

The notion of a trading relationship between vowel quality cues and durational cues has been brought forward in phonetic perception studies (Hadding-Koch and Abramson, 1964) and in diachronic studies (Widmark, 1998). A study of correlations between durations and spectral variables reflecting vowel quality has therefore been included.

Preaspiration has, until recently (see e.g. Helgason, 2002; Wretling et al., 2002;Tronnier, 2002), only played a minor role in the discussion of Swedish quantity. Its inclusion in this part will therefore mainly refer to the geographical distribution of preaspiration duration and frequency.

The empirical part consists of five chapters. The present chapter describes the SweDia corpus and the methods used for the analysis of the corpus, with an emphasis on cluster-analysis techniques. Chapter 8 provides the results of a cluster analysis of the quantity material of the SweDia corpus, based on durational variables from 86 recording locations in Sweden and Finland. In Chapter 9, preaspiration in the test words and spectral differences between the long and short qualities of the vowels are investigated and discussed. Chapter 10 presents two additional investigations of the quantity patterns and vocalic spectral differences. This chapter aims at a verification of the previous results and is based on additional material from the SweDia data base.

7.2 The SweDia Corpus

The SweDia corpus consists of recordings from 106 locations in Sweden and the Swedish speaking parts of Finland[1]. The corpus was recorded between 1998 and 2001 as a joint effort between the departments of linguistics at the universities of Umeå, Stockholm and Lund. By and large, the areas recorded by the three universities coincided with Wessén's dialect areas (see section 5.1): The department in Umeå was responsible for recording the dialects in the northern parts of Sweden and Österbotten in Finland. The department in Stockholm was responsible for recording the Svea varieties, the dialects in Åland and the Finland-Swedish dialects in Nyland and Åboland. The department in Lund carried out the recordnings of the Göta varieties, Southern Swedish and the Gotlandic varieties.

Usually 12 speakers were recorded for every location. The speakers were distinguished by age and sex, resulting in the recording of three elderly men, three elderly women, three younger men and three younger women per location.

The recordings were usually carried out in the homes of the participants or other familiar locations to make them feel comfortable and relaxed and to make the use of a local vernacular more natural. The technical equipment consisted of a lapel microphone (attached at an appropriate place of the participants' clothes), and a portable DAT-recorder. All recordings were done in 48 kHz sample rate and 16 bit sample size. For further processing, the data were stored on hard disk and downsampled to 16 kHz/16 bit.

The recordings in Sweden were carried out by research assistants from the universities in Umeå, Stockholm and Lund. Recordings in Finland were made by colleagues from the universities in Vasa and Helsinki or by locally recruited assistants with a keen interest in and a thorough knowledge of the local dialect. As a preparation for the field work, all assistants had completed a training program where they had received instructions about the handling of the technical equipment, interview procedures and the structure of the material to be recorded including word lists and procedures for eliciting these wordlist in a reasonably natural manner.

Four different corpora were recorded. Three corpora were non-spontaneous, elicited corpora focussing on specific linguistic and phonetic aspects of the dialects. The fourth corpus

[1] see *http://www.swedia.nu* for a description of the SweDia project and a complete list of the recording locations

was a corpus of spontaneous speech, consisting of monologues by the dialect speakers or dialogs between two dialect speakers. The three non-spontaneous corpora focussed on the sound system, the intonation and tone accent system and the quantity system of the dialects. For the present study, the quantity corpus was used. The quantity corpus consists of repeated utterances of certain target words in isolation. The intended number of repetitions was five times, usually indicated to the participants by hand signals to avoid final lengthening effects. However, informal inspection during the segmentation process revealed that for a considerable number of participants this procedure was not followed. In these cases, the participants were usually asked to say the respective word "five times in a row".

Three word pairs were recorded for all recording locations: *tak-tack*, *låt-lott* (or *gråt-grått*) and *väg-vägg*. The last word pair was abandoned from further investigation, as in large parts of the Southern Swedish region the pronunciation of *väg* does not include a final voiced velar plosive but a voiced palatal fricative or approximant (e.g. [vɛj]), which makes comparisons to the counterpart *vägg* (usually [vɛgː]) difficult.

For large parts of the Umeå recording area, an extended set of test words was recorded. All words, their Standard Swedish pronunciations and their English translations are given in Table 7.1. For the data-driven categorisation, only the word pair tak-tack was used. The other word pairs were used for more detailed investigations of the types derived from the initial categorisation, as described in the Chapter 10. For some contrasts, different pairs of target words were recorded at different recording locations (see alternatives in Table 7.1).Therefore, the word pairs are denoted as the A-class, O-class, E-class, I-class, U-class and Y-class words, respectively.

Table 7.1: Target words (Standard Swedish) of the quantity corpus in the SweDia database.

Identifier	Written form	Pronunciation	English translation
		All recording areas	
A-class	tak	[tɑːk]	roof
	tack	[takː]	thanks
O-class	låt, (alt. gråt)	[loːt] ([groːt])	song (crying)
	lott (alt. grått)	[lɔtː] ([grɔtː])	lottery lot (grey, neuter))
-	väg	[vɛːg]	way
-	vägg	[vɛgː]	wall
		Umeå recording area	
E-class	tät	[tɛːt], [tɛtː]	tight
	tätt	[tɛːt], [tɛtː]	tight, neuter
I-class	dit (alt. vit)	[diːt] ([viːt])	there (white)
	ditt (alt. vitt)	[dɪtː] ([vɪtː])	your (white), neuter
U-class	bot	[buːt]	remedy or fine
	bott	[bʊtː]	lived, verb perfect participle (conj. 3)
Y-class	byt	[byːt]	change, verb imperative
	bytt	[bʏtː]	change, verb perfect participle (conj. 2b)

7.3 Method

7.3.1 Software

For segmentation, either ESPS/xwaves or the WaveSurfer© program by K. Sjölander and J. Beskow[2] was used. Duration and formant values were extracted with the Emu speech database system by Steve Cassidy[3]. Statistical analysis was performed with R 1.01 (R Development Core Team, 2004) for Linux or Mac© OS X and SPSS© 11.0 for Mac© OS X.

7.3.2 Segmentation

The segmentation of the acoustic material was partly performed by trained research assistants of the SweDia project and partly by the author. The segmentation criteria were developed during the initial phases of the SweDia project and were available to all co-workers in form of a written manual. The conventions for the segmentation of the quantity material can be summarised as follows:

- Maximally three segments were delimited in the test words: the vocalic part of the word, an optional preaspiration phase between the vocalic and the final consonantal part and the closure phase of the final consonantal part. The initial consonant of the words was not segmented.
- The segmentation was always based on auditive as well as visual information in the form of oscillograms and spectrograms.
- The start of the vocalic phase was usually identified by the appearance of formants in the spectrogram. The appearance of the first clear glottal period in the sound wave oscillogram was used as an additional indicator.
- The end of the vocalic phase was marked when periodicity in the oscillogram ceased and formants, especially F2 and above, showed a major drop in intensity in the spectrogram.
- Preaspiration was marked when the vowel was followed by non-periodic noise. This noise could appear in two phases which were not distinguished in the segmentation: an aspiration noise with a formant structure, and a fricative-like noise without formant structure. The end of the preaspiration phase was marked when the frication noise ceased in the spectrogram or at the point where clear transients towards the consonantal part were visible.
- The closure phase of the consonantal part started with the end of the vocalic or preaspiration part. Its end was marked with the initiation of the closure burst.

Most recordings in the SweDia database were performed in the participants' homes, to promote a relaxed atmosphere and the choice of the local vernacular. This resulted in occasional echo effects which complicated the segmentation procedure, especially at the transition between the vocalic part and the preaspiration or the consonantal part. In most cases however, a distinction was possible on the basis of the spectral range of the noise. Echo effects appeared only

[2] see *http://www.speech.kth.se/wavesurfer/*
[3] see *http://emu.sourceforge.net/*

below 4000 Hz, while preaspiration noise often spanned the whole frequency range or had its emphasis in higher frequencies.

Another problem with a potential to have a greater effect on measured durations was the occurrence of unvoiced vowel onsets. For some utterances of some speakers, voice onset was delayed until later in the vowel, which could be told by auditive inspection and visually from a noise part with a clear formant structure in the spectrogram after the aspiration of the preceding consonant. In those cases, it was decided to depart from the rule to mark the beginning of the vowel at the first glottal impulse. Rather, vowel onset was marked at the point were the noise started to show a clear formant structure.

7.3.3 Graphical Displays

The plots used in the following sections are normally self-explanatory. The graphical display method of the "box plot" should be explained however, as different sources define it in different ways. The boxplots used in the present study were drawn with R (R Development Core Team, 2004). A boxplot consists of a "box" and so-called "whiskers" extending from the box (see e.g. Figure 8.10). The box of a boxplot displays the middle 50% of a distribution. The lower edge of the box thus marks the 25% quantile (lower quartile), the higher edge the 75% quantile (higher quartile). A line or a dot within the box marks the 50% quantile (the median).

The distance between the 25% quantile and the 75% quantile is usually called the "interquartile range". From the interquartile range, the length of the whiskers is calculated. By convention, the whiskers extend to the last value that is within a range of 1.5 times the interquartile range from the upper and lower quartiles. Values outside these ranges are treated as outliers and marked by small circles. Some boxplots show notches in the boxes (see e.g. Figure 10.1), representing the approximate 95% confidence intervals of the medians. Hence, non-overlapping notches between two boxplots provide strong evidence that the medians are significantly different at the 5% level (Chambers et al., 1983).

The present study frequently investigates geographical distributions of variables. Therefore, several maps are included. The maps cover the Swedish mainland and the Swedish speaking parts of Finland. Major inshore waters are included as well. Furthermore, the maps show the borders of the Swedish provinces. The names of the provinces are given in Figure 7.1.

The map in Figure 7.2 shows the three-letter short forms of the 86 recording locations investigated in the present study. A list with the full names of the recording locations is given in appendix A, and a map showing the full names of the recording locations and their geographical position is given in Appendix C.

7.3.4 Measurements

7.3.4.1 Segment Durations

For each word, four durational variables were measured: The duration of the vowel-consonant sequence (V:C or VC:), the duration of the vowel (V: or V), of the consonant[4] (C: or C) and of

[4]"Duration of the vowel" and "duration of the consonant" are actually imprecise expressions with respect to the level of description. However, as the term "duration" was defined as referring to a physical measure in Section 1.2, there is in my opinion no risk of confusion of levels. The terms "consonant" and "vowel" are therefore also used at the acoustic level.

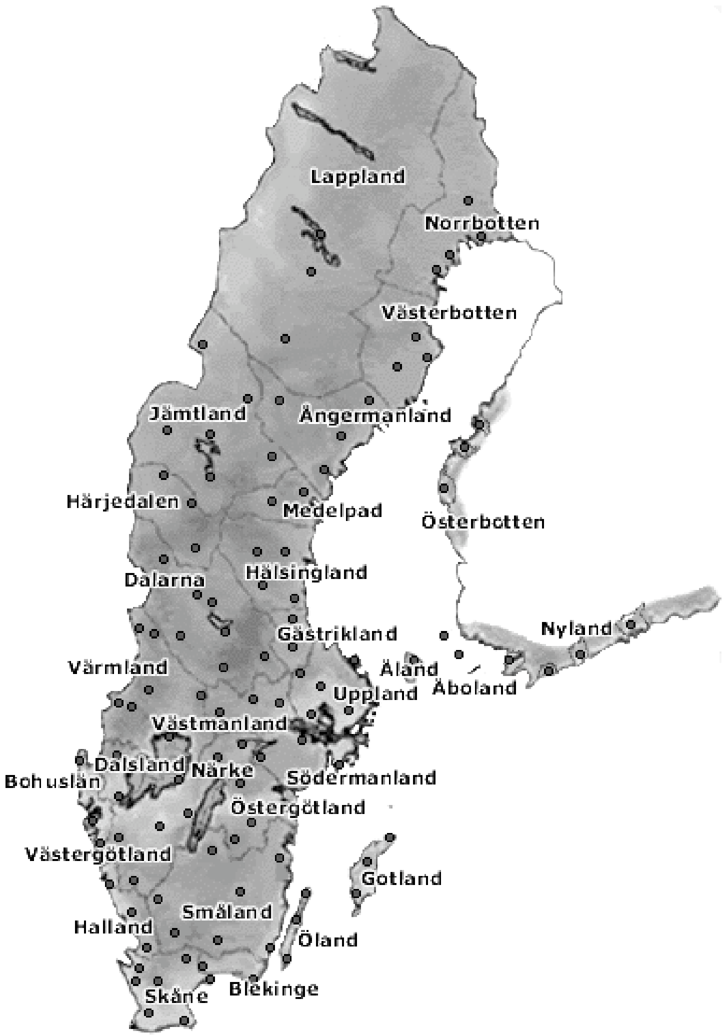


Figure 7.1: The provinces of Sweden and the Swedish speaking parts of Finland.

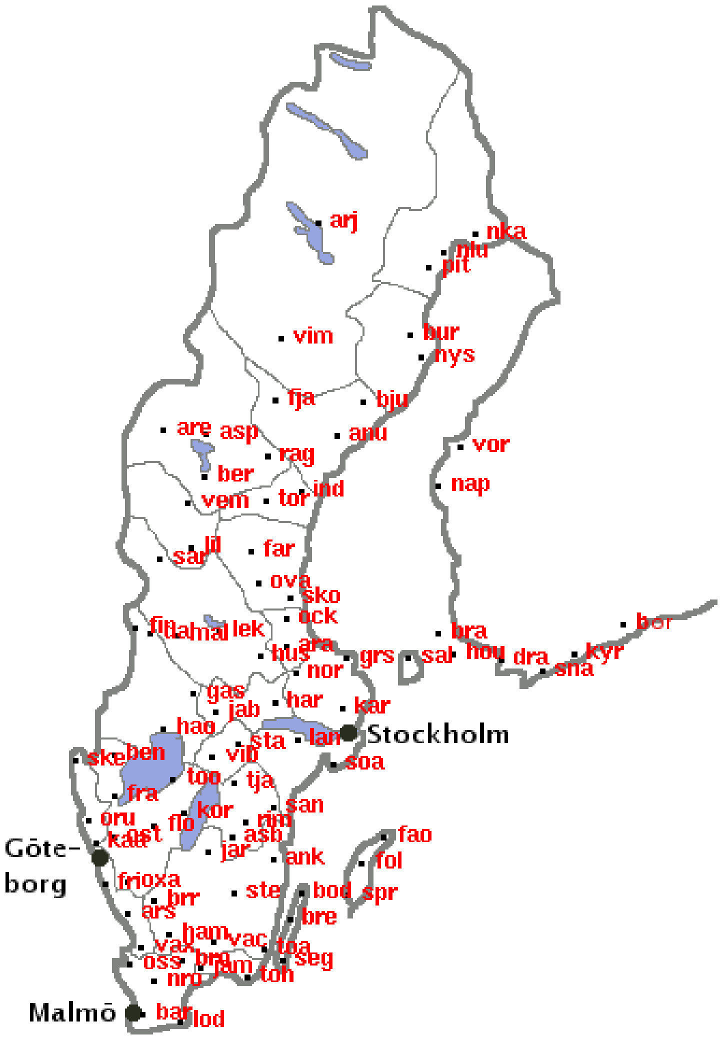


Figure 7.2: Three-letter short forms of the recording locations investigated in the present study.

the preaspiration phase (which could be 0). Consonant duration was always measured with the preaspiration phase included.

As the present study aimed at comparisons across recording locations, appropriate measures of central tendency had to be chosen. In a first step, the median for each variable of each speaker was calculated. The median was chosen instead of the arithmetic mean because it is less sensitive to outliers. If for example, final lengthening would influence the last utterance of some speakers, this would have an effect on the mean value, but rarely on the median. In a second step, the median of each recording location was calculated from the speakers' medians. Again, median calculation diminishes the influence of outliers, for example values from speakers that spoke extremely fast or slow. It should be pointed out however, that the differences between mean and median calculations are probably not too important. A comparison between the means and medians for each recording location revealed that the differences in most cases did not exceed 10 ms.

7.3.4.2 Vowel Formants and Acoustic Distance

Formant values were extracted at the midpoint of the vowels in all productions of the two words. The computer script used for this procedure was a tcl-tk script, which is part of the EMU-package. For every speaker the medians of F1 and F2 for the long and the short vowel were calculated. These values were bark-transformed using the formula given in Traunmüller (1990). The Euclidean distance between the bark-transformed F1 and F2 values in the F1-F2 vowel space was used as a measure for the quality difference between the long and the short vowel. In the following, this variable is called “acoustic distance” (AD).

The chosen method for formant analysis disregarded potential diphthongisation. An informal listening test revealed that this was hardly problematic in the case of the A-class words, as diphthongisation was rare for these words. Formant analyses for the other word classes were complemented by detailed auditory inspection (see Chapter 10).

7.4 Cluster Analysis

7.4.1 The Motivations for a Data-Driven Categorisation

A traditional account of the quantity patterns in Swedish dialects was presented in Chapter 5. In the following a "bottom-up" approach will be used in order to explore the possibility of defining meaningful dialect areas based on the clustering of acoustic properties derived from the data (see also Strangert and Wretling, 2003). This procedure was further motivated by three considerations: First, only an approach that does not assume already established dialect areas bears the possibility of discovering new ways of describing the geographical distribution of quantity patterns. Second, most descriptions of the quantity systems of Swedish dialects are based on dialectological sources often using auditive judgements but, in most of the cases, no acoustic measurements. Thus, with respect to Swedish dialects, there is not much knowledge about the distribution of the phonetic realisations of quantity. Third, the reports on quantity in Swedish dialects are incomplete. Some of the dialects in the SweDia database have been described extensively, others only scarcely or not at all. The geographical distributions generated by the data-driven approach will be compared with the traditional categorisations.

7.4.2 Approaches to Clustering

The term "cluster analysis" denotes a variety of methods for finding patterns or structures in a data-set. In this context, a data-set consists of a number of cases (usually called "objects"), which are described by a range of variables. Based on the values of these variables, a clustering algorithm tries to find groups that are similar to each other. A principal distinction is often made between "hierarchical" and "non-hierarchical" methods. The former ones can be "agglomerative" or "divisive" (Everitt, 1993). Agglomerative hierarchical clustering proceeds by treating each object as a single cluster and then combines the objects stepwise into larger clusters until all objects fall into one large cluster. Divisive methods proceed the other way round. Divisive methods are less popular than agglomerative methods. The relevant literature, therefore, often describes them only briefly. Consequently, these methods were not taken into account in the present study. For an agglomerative hierarchical method, usually a so-called "distance matrix" is needed. This distance matrix describes the distance between any two objects in the data-set. A variety of distance measures can be applied. The most usual one for interval-scaled variables is the Euclidean distance. Euclidean distances can be problematic if the different variables are expressed on different scales (Everitt, 1993). As this was not the case in the present study, alternative distance measures were not taken into account.

The distances in the distance matrix are used for the first step, where every object is joined with its most similar neighbour. From the second step on, a decision has to be made about which method should be used for the joining of clusters. Even here, the statistical literature provides a great variety of methods. Everitt (1993) comes to the conclusion that "Ward's method", "group average" and "complete linkage" are *"...the most useful in practice..."* (p. 89). Bortz (1993) recommends the Ward method. All authors stress that there are no really clear-cut criteria for the decision about which cluster-analysis method should be used. "Complete linkage" (also called "furthest neighbour") calculates the distances between the cluster members that are furthest away from each other and then joins those two clusters where this distance is lowest. "Group average" calculates the average distance between all members of two clusters and joins the two clusters where this value is minimal. The Ward method is quite different, although it is usually described in the same context as the other linkage methods. This method joins those two clusters where the within-class sum of squares — also called sum of error squares (SS_{Error}) — is minimal. SS_{Error} is calculated as the sum of the squared deviations of each value from the mean of the variable:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{7.1}$$

This value is calculated for every variable and the results for all variables are summed up. The algorithm then joins those two clusters where the increase in the total SS_{Error} is lowest. For the present study, the Ward method was used because of the recommendation in Bortz (1993), but the other two methods were tested as well. While "complete linkage" produced results very similar to the Ward method, "group average" produced different results. This stresses the fact that even a categorisation driven by computational means is not a completely objective method, and that its results have to be evaluated by some independent criterion.

Strictly speaking, the Ward method does not require a distance matrix. However, the algorithms implemented in SPSS and R still use a distance matrix. SPSS demands a squared Euclidean distance matrix for the calculation. R has two different algorithms (called "hclust" and “agnes”). Both proceed from a non-squared Euclidean distance matrix. For unknown reasons, “hclust” produces results slightly different from “agnes”. The results obtained with “agnes” are identical to those produced by the SPSS implementation of the algorithm. Therefore the agnes algorithm of R was used for further analysis[5].

After the application of a hierarchical method, it has to be decided how many clusters should be taken into consideration. A graphical tool to aid this decision is the so called “cluster dendrogram”. Such a dendrogram (see e.g. Figure 8.1) depicts the decisions made during the clustering process and shows the increase of the joining criterion as the length of vertical branches in the tree structure. Large vertical branches indicate the joining of rather dissimilar clusters, and thus point towards possible “break points”. Timm (2002) suggests the calculation of R^2 for different cluster splits to estimate the amount of explained variance for a certain cluster split. R^2 (or η^2, as it is often called in an ANOVA context) is calculated as:

$$1 - \frac{SS_{Error}}{SS_{Total}} \tag{7.2}$$

SS_{Total} can be interpreted as SS_{Error} of a 1-cluster split, i.e. the variation in the data if no grouping is assumed. The definition of R^2 implies that its value lies between 0 and 1: If all cases are in the same group, SS_{Error} is equal to SS_{Total}. If every case constitutes its own group (cluster), SS_{Error} is 0. The use of R^2 as a measure of the variation explained by an analysis of variance or a regression analysis is controversial. It is often considered as a “too optimistic” estimation. Usually, R^2 is weighted against the degrees of freedom for a certain number of groups (here, clusters) and the number of cases:

$$1 - \frac{SS_{Error}/df_{Error}}{SS_{Total}/df_{Total}} \tag{7.3}$$

df_{Error} denotes the difference between the number of cases and the number of groups (clusters), while df_{Total} denotes the number of cases minus 1. The calculation of the adjusted R^2 can lead to negative values in some cases. In the following, R^2 always refers to this adjusted R^2. For a range of different cluster splits, R^2 can be calculated for every variable involved. The increase of the term with increasing numbers of clusters can then be used for an estimation of the increase in explained variance associated with a certain number of clusters. With cluster numbers increasing from 1 to the number of objects, it can be expected that R^2 initially rises fast and then “flattens out” with higher cluster numbers. If the R^2 values of the different variables are compared, the importance of a certain variable for a certain cluster split can be estimated, and the “flattening” of the curves can be used as an indication of the point where additional clusters do not add much additional information.

As mentioned before, non-hierarchical clustering methods exist as well. Those methods require a predefined number of clusters (Everitt, 1993). As such a predefinition was neither desired nor possible in the present study, non-hierarchical methods were not taken into account for the initial clustering. Non-hierarchical methods are, however, often recommended to optimise a given hierarchical clustering, as hierarchical clusterings have the disadvantage of being “uncorrectable”; clustering decisions at a lower level cannot be revised at a higher level. This might lead to sub-optimal end results. To overcome this problem, it has often been

[5] In Schaeffler (2005), the “hclust” algorithm was used. Therefore, the results reported there are slightly different from those reported in the present study.

recommended (see e.g. Bortz, 1993) to verify the results of a hierarchical method with a non-hierarchical method. A common non-hierarchical method is the so-called “k-means” method. Like the hierarchical Ward method, it tries to minimise SS_{Error}, but it does so for a given number of clusters with given cluster centers. A “cluster center” is defined by the mean values of the variables from all objects in a cluster. Consequently, if a hierarchical method has been applied, a certain cluster split can be optimised and tested with the “k-means” algorithm. In the following, some further aspects of cluster analyses are discussed, before the results of the present analysis are presented in Chapter 8.

7.4.3 Standardisation

The statistical literature often recommends the standardisation of variables before applying a cluster analysis. In this context, “standardisation" usually means the transformation into z-scores. This is done by subtracting the variable’s mean from each value and then dividing the result by the standard deviation of the variable, resulting in transformed variables with a mean of 0 and a standard deviation of 1. The transformation makes immediate sense if variables with different scales and ranges are compared, as otherwise variables with larger scales would dominate the clustering. In the analysis performed here, the case is different, as all variables are measured on the same scale (milliseconds). Transformation to z-scores would here merely equalise the variance found in the four variables. It was decided therefore to use the raw values of the variables. Attaching minor importance to those variables that showed less variance seemed to be a sensible choice. For example, /a/ duration showed a standard deviation of 12.9 and a range of 66.9 ms, while /k:/ duration showed a standard deviation of 28.6 and a range of 143.6. It is thus likely that /k:/ duration has a greater potential than /a/ duration to distinguish between dialects. Transformation to z-scores would mean in this case to value the same amount of durational variation higher with /a/ than with /k:/.

7.4.4 Assets and Drawbacks of Cluster Analyses

Cluster analyses always arrive at some sort of categorisation and it is no trivial task to interpret such a categorisation. First, one has to decide how many clusters are relevant for the purpose. Second, one has to suggest causes for the observed patterns. This causal interpretations will always be given “after the fact”, thus the method is “hypothesis-creating” not “hypothesis-testing”.

An important aid in the interpretation of the results of the cluster analysis in the context of the current investigation is the geographical distribution. If a certain cluster shows a non-random geographical distribution, it becomes more likely that a significant dialectal feature has been observed. This is derived from the very fact that dialectal features are often geographically bound. The common historical model of language development in Scandinavia is one of diversification (Pettersson, 1996): A common variety splits up into several dialects. This split-up emanates from certain places, often cultural centres, where innovative features arise and then spread to surrounding areas, sometimes including jumps to other cultural centres while neglecting less central areas in between. This model predicts non-random geographical distributions, although it does not exclude the possibility that some features can be found isolated at different places, either because they constitute recent innovations that have not spread yet, have been kept from spreading for independent reasons, or constitute “anachronisms” that only survived at a few (possibly remote) places.

A further aspect that should be taken into consideration is the fact that cluster analyses always create categorisations, even if the data has no internal structure. Furthermore, the categorisation suggests sharp borders between categories, even if a gradient or scale-like description would be more appropriate. This has to be taken into consideration when interpreting the results of a cluster analysis. The data-driven categorisation is thus only a first step, which has to be followed by a sensible interpretation of the resulting categories.

Chapter 8

A Cluster Analysis of the A-class Words

In this chapter, the cluster analysis as described in Section 7.4.2 will be applied to a subset of the quantity material, the A-class words *tak* /tɑːk/ and *tack* /takː/ (see Table 7.1).

8.1 Determination of the Number of Clusters

As described before (see Section 7.4.2), the cluster analysis was performed using the "agnes" algorithm of the R software, and the "Ward-method". Every object (i.e. recording location) was described by four variables: the median durations of the four segments (Vː, V, Cː and C) measured in the A-class word pair. Figure 8.1 shows the resulting cluster dendrogram. The height of the vertical edges of the dendrogram represent the increase of the error term (here SS_{error}) when two clusters are joined. Visual inspection of the dendrogram suggests, that a large amount of variation of the four variables should be explainable by 2–4 clusters.

As an additional source of information, the adjusted R^2 was calculated for every variable and for the 2 to 15 and the 20 cluster splits. Figure 8.2 shows the increase of R^2 from 2 to 20 clusters (interpolated between 15 and 20). It can be seen that there is no major increase in the R^2 values between 14 and 20 clusters. Together, the dendrogram and the R^2 values suggest a main partitioning into 3 to 4 clusters, and possibly further interesting groups up to a 14 cluster split. Therefore, it was decided to analyse the data in two steps: First, the two to four-cluster splits were investigated. Second, further sub-clusters were inspected in order to determine whether the smaller clusters provide interesting partitionings in terms of geographical distribution and durational structure.

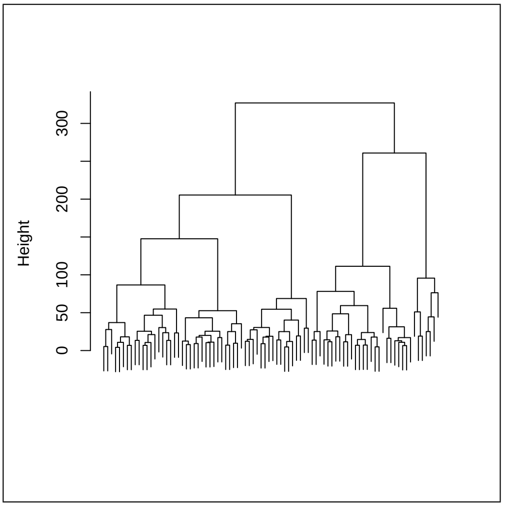


Figure 8.1: Cluster dendrogram for the hierarchical clustering of V:, C, V and C: duration.

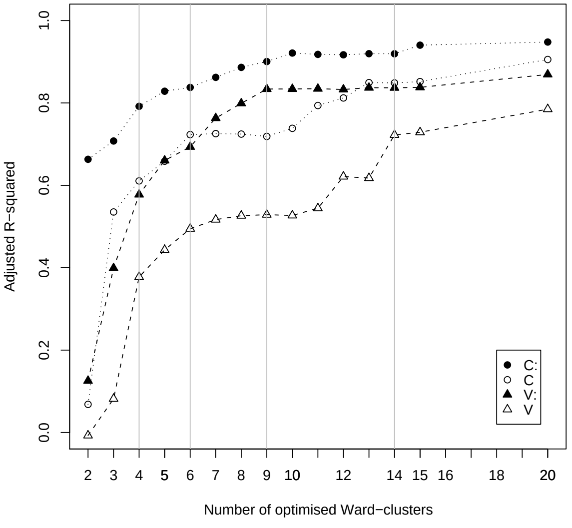


Figure 8.2: The increase of R^2 for V:, C, V and C: duration by cluster number.

8.2 Geographical Distribution

The geographical distribution of the two, the three and the four-cluster splits of the A-class words with the Ward method is shown in the Figures 8.4, 8.5 and 8.6. The two-cluster split shows a clear geographical distribution. In general, the dialects group into a Southern versus a Northern and Finland-Swedish cluster. The first cluster incorporates 52 recording locations, mainly in the Southern Swedish region up to Uppland and Middle Dalarna. The second cluster incorporates 34 recording locations in the remaining parts of Sweden and the Finland-Swedish parts. A few recording locations break the clear North-South partition: The three Northern locations Aspås (asp) in Jämtland, Anundsjö (anu) in Ångermanland and Torp (tor) in Medelpad group together with the Southern cluster. The three Southern locations Järnboås (jab) in Västmanland, Bengtsfors (ben) in Dalsland and Öxabäck (oxa) in Västergötland group together with the Northern cluster.

The three-cluster split shows a further partitioning of the Northern cluster into a Northern Swedish cluster and a mainland Finland-Swedish cluster. The four-cluster split partitions the Southern cluster. Here the geographical distribution is rather scattered, but one of the clusters tends to be more concentrated on the coastal regions in the West, South and East.

If the geographical information is combined with the information from the dendrogram in Figure 8.1 and from the R^2 analysis, it seems justified to assume three main clusters. These three main cluster are then separated into sub-clusters.

The five to 14 cluster splits have been combined in three maps, based on the three main clusters (Figure 8.7, Figure 8.8 and Figure 8.9). For easier reference, the 14 sub-clusters have been labelled. The labels partly maintain the hierarchical structure, so that more similar clusters get more similar labels[1]. Figure 8.3 shows a horizontal dendrogram for the different subgroups of the 14-cluster split and shows the chosen label for each cluster.

[1] A complete rendering of the hierarchical relationships would have led to rather clumsy labels, because of the high numbers of intermediate levels.

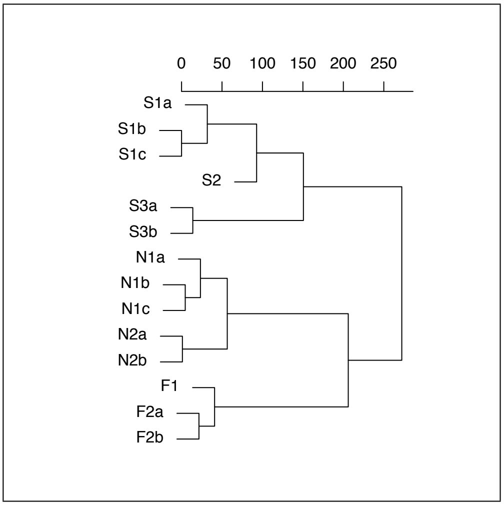


Figure 8.3: Horizontal cluster dendrogram for the hierarchical clustering of V:, C, V and C: duration, cut at the 14th level.

At the 14 cluster level, not all clusters show geographical concentrations. The mentioned "Southern coast" cluster (S3a, S3b) was split up once, with S3b including only two — geographically distant — places. S3a is mainly found in the West and along the Southern and Southern East coast, up to Southern Gotland. Around the Mälaren and Stockholm region the S2 cluster dominates. The S1 clusters do not show a clear geographical contiguity either, although there are some areas where members of these clusters concentrate (for example four places in Eastern Dalarna and Western Uppland). As a tendency, the S1 distribution could be interpreted as a "buffer zone" between S3 and S2. S3 dominates the Southern and South-Western coastal regions, and S2 dominates the North-East (of Southern Sweden), the S1 members follow the South-Western and Southern coastline as well, but lie more towards the inland. There are, however, many exceptions from this rule.

In the North, there is one cluster, N1b, that clearly shows a geographical concentration in the Jämtland, Härjedalen and North Western Dalarna region. One Northern cluster (N2a) consists of a single dialect, Arjeplog (arj) in Northern Lappland. The dialect of Arjeplog is well-known for its preaspiration (see Wallström, 1943;Helgason, 2002; Wretling et al., 2002), thus the separation of this dialect is hardly surprising (see also Section 9.1). The other clusters do not show a clear geographical contiguity, although there are some local concentrations, e.g. in Västerbotten.

The Finland-Swedish sub-clusters divide into the two dialects of Österbotten (Vörå (vor) and Närpes (nap)), and the southern Finland-Swedish dialects. Kyrkslätt (kyr) in Nyland (F2b) is separated from all other Finland-Swedish dialects.

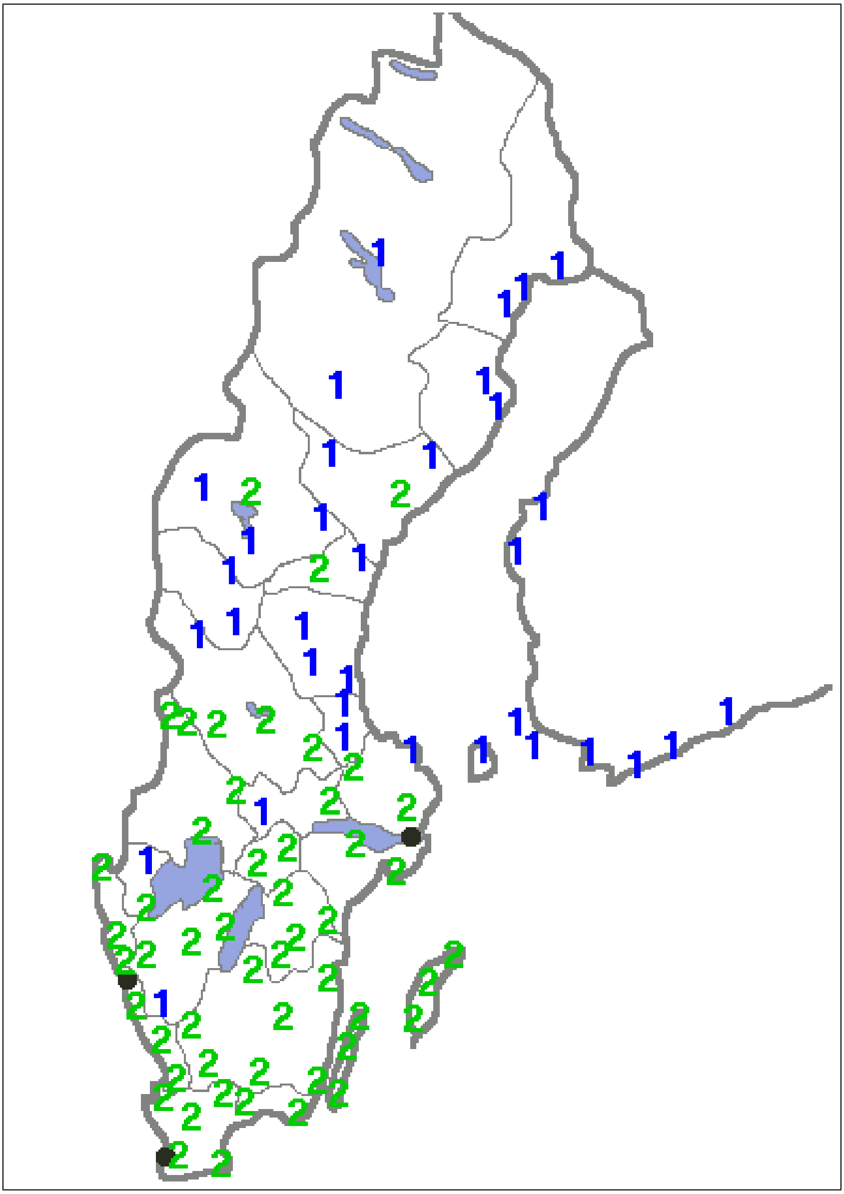


Figure 8.4: Geographical distribution of the clusters obtained with the Ward method. 2 cluster split.

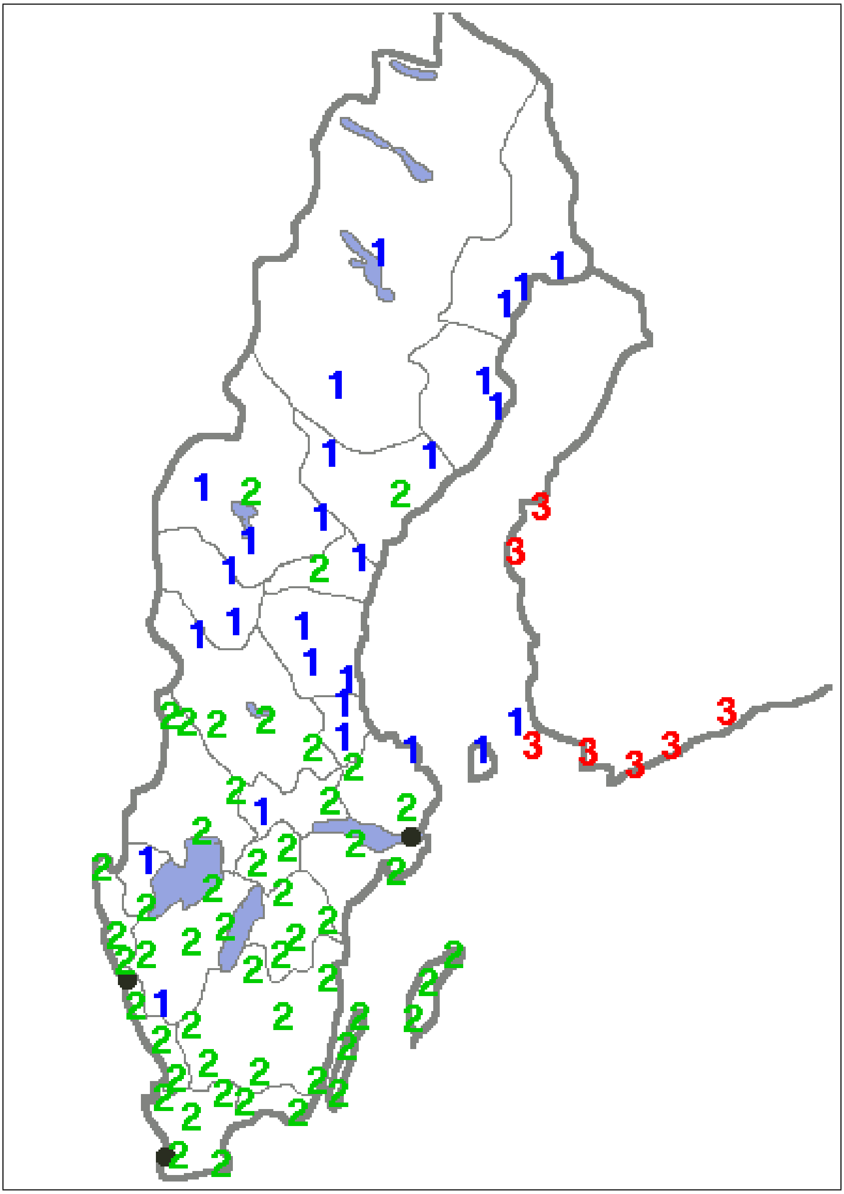


Figure 8.5: Geographical distribution of the clusters obtained with the Ward method. 3 cluster split.

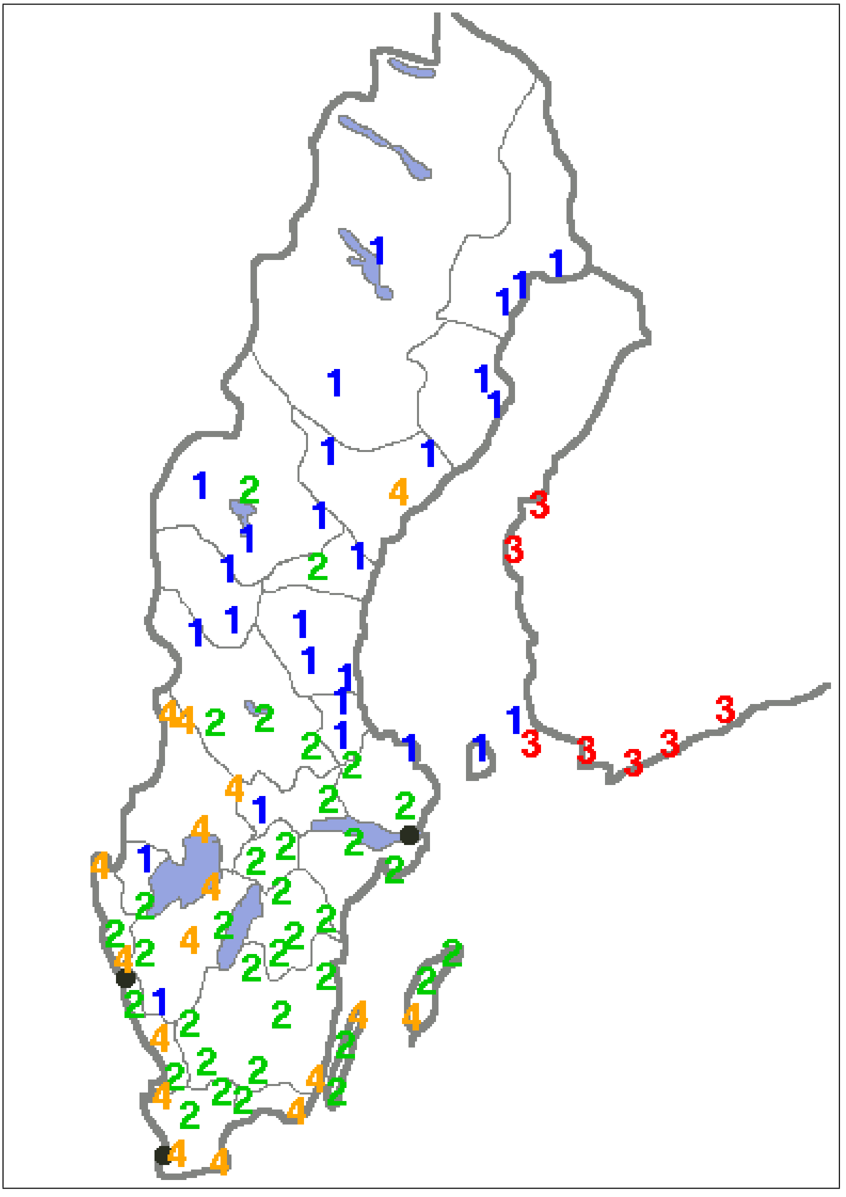


Figure 8.6: Geographical distribution of the clusters obtained with the Ward method. 4 cluster split.

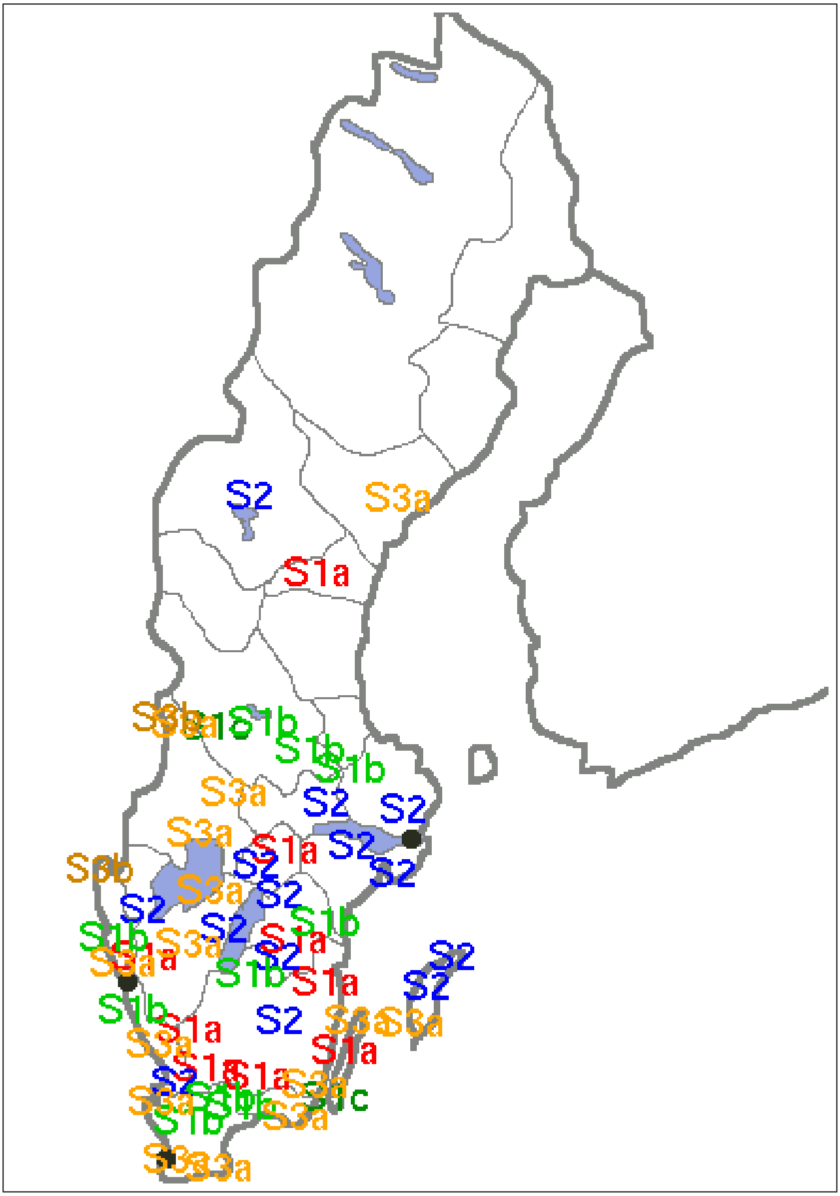


Figure 8.7: Geographical distribution of the clusters on the 14th level for the Southern cluster.

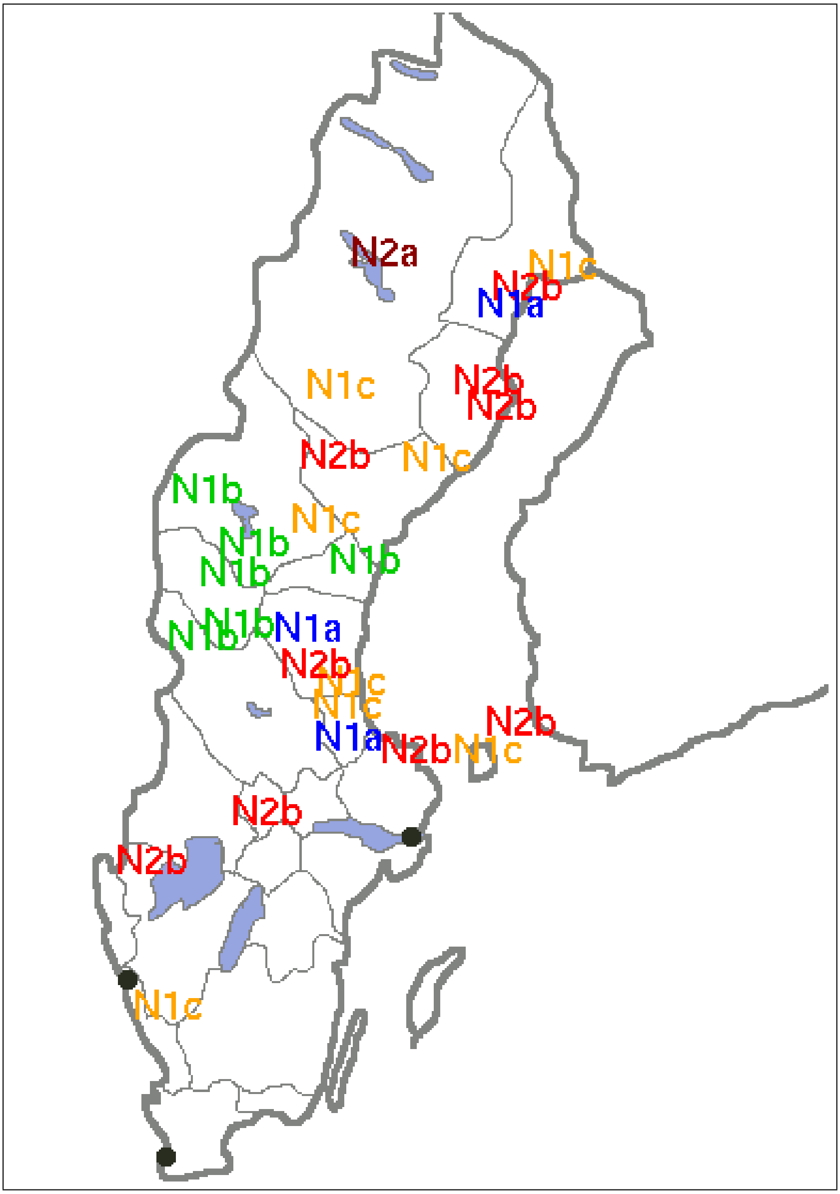


Figure 8.8: Geographical distribution of the clusters on the 14th level for the Northern cluster.

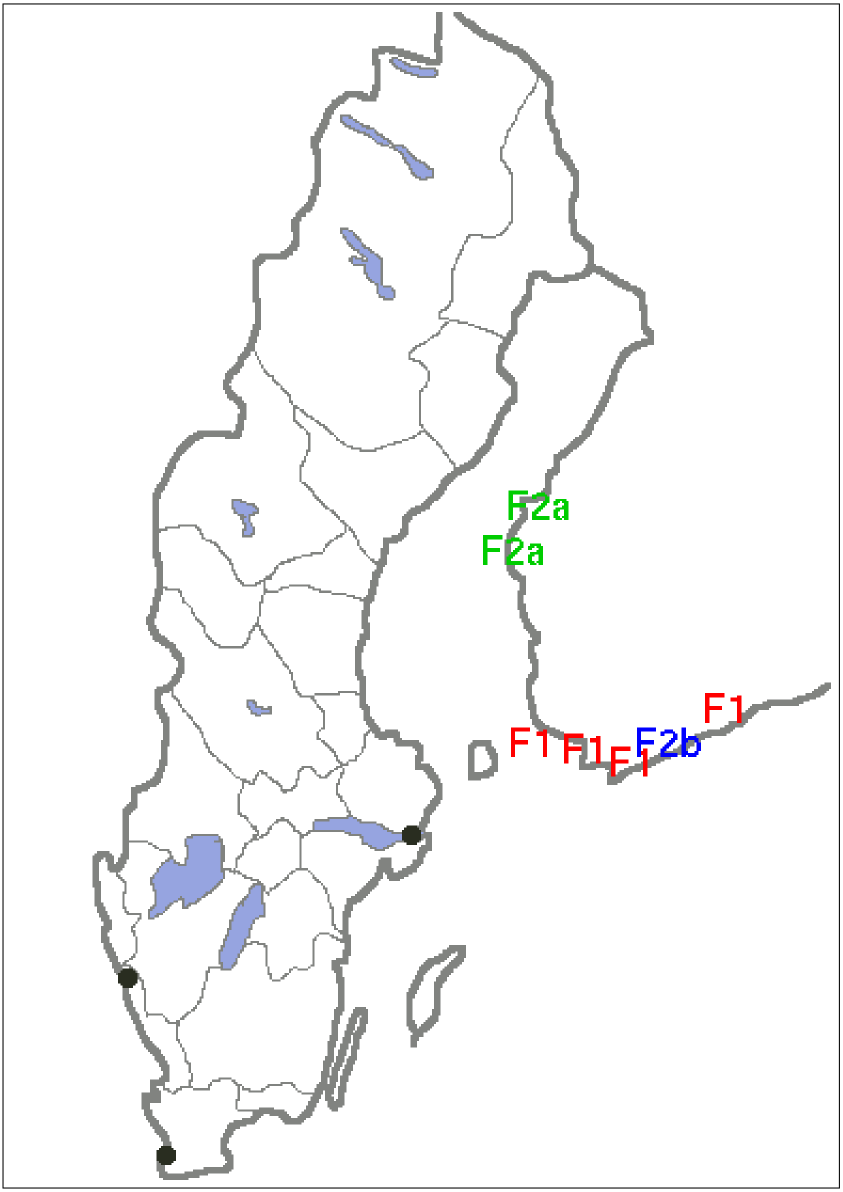


Figure 8.9: Geographical distribution of the clusters on the 14th level for the Finland-Swedish cluster.

8.3 Durational Characteristics of the Clusters

The cluster analysis separated the dialects into clusters based on the four variables median V: duration, V duration, C duration and C: duration. In the following it will be analysed how the different duration variables contributed to the separation into clusters. Besides the absolute duration values, it is also of interest how the segment durations relate to each other. Additionally, the relation of segments in sequence, together with the durational characteristics of the whole sequence, should allow judgements about differences in the composition of the vowel-consonant sequences across dialects. Therefore, six additional variables were taken into consideration: the total durations of the V:C and VC: sequences, the V:/V and C:/C ratios, and the V:/C and V/C: ratios. The results will be discussed in two steps. In the first step, the three main clusters will be adressed. In the second step, the sub-divisions within the three main clusters will be analysed.

The three main clusters will be referred to as F, N and S. The F cluster is the Finland-Swedish cluster (number 3 in Figure 8.5), the N cluster is the Northern Swedish cluster (number 1 in Figure 8.5) and the S cluster is the Southern Swedish cluster (number 2 in Figure 8.5).

8.3.1 Segment Durations in the Three Main Clusters

The median durations of the four segments (V:, V, C: and C) across the three main clusters are shown in Figure 8.10. V: durations tend to be longer in the F cluster, while V: durations in the N and S cluster are of a similar order of magnitude. However, there is considerable overlap in the distributions of V: durations in F and S. Only three of the seven Finland-Swedish locations show V: durations that are longer than those in S.

The V durations generally show less variation (range 77 ms) than the V: durations (range 128 ms). V durations tend to be shortest in N and longest in S, although the F median is higher than the S median. Generally, the V durations show considerable overlap across the three main clusters, which is in accordance with the results from the R^2 analysis.

The C durations show clear differences across the three clusters. C are shortest in the Finland-Swedish cluster and longest in the Northern cluster. The C durations in the S cluster lie in between those of F and N. However, there is considerable overlap between the N and the S cluster. S and F show only three recording locations where the median C durations overlap: One F location with exceptionally long C durations and two S locations with exceptionally short ones.

C: durations show a clear pattern as well. F and N have more or less equal C: durations, while S has clearly shorter C: durations. With only two exceptions, C: durations in the N and F clusters exceed those of the S clusters.

Figure 8.11 shows the sequence durations for the three clusters. Generally, V:C sequence durations are more similar than VC: sequence durations across the clusters. F and N have approximately equal V:C sequence durations, although the variation is higher in F (in spite of the fact that the F cluster is much smaller). This shows that the shorter C durations in F are compensated by longer V: durations (see Figure 8.10). S tends to have shorter V:C durations than F and N. As N and S show virtually equal V: durations, the shorter V:C sequences in S should be caused by the shorter C durations.

VC: durations tend to be somewhat longer in F than in N, and are clearly longer in N than in S. The corresponding segment durations suggest that the difference between F and N is a combined effect of somewhat longer V and C: in the F cluster. The durational difference between N and S, on the other hand, is clearly an effect of the longer C: in N, as S tends to show slightly longer V than N.

Figure 8.12 shows the ratios between the long and short vowels, long and short consonants and between the segments in succession. The F cluster tends to have the highest V:/V ratios, there is, however, considerable variation within this cluster. F and S show V:/V ratios in the same order of magnitude. As a whole, V:/V ratios are rarely below 1.4, indicating a stable durational distinction between long and short vowels.

As a consequence of the long C: and the short C durations, C:/C ratios are high in the F cluster and very different from the N and S clusters, without a single overlap. C:/C ratios also tend to be higher in the N than in the S cluster, although the distributions overlap considerably. Very low C:/C ratios are, however, mainly found in the South. In the whole material, there are 39 recording locations with C:/C ratios below 1.2. 36 of these recording locations are found in the Southern cluster. Figure 8.12 indicates that only the C:/C ratios in the F cluster reach values that are comparable to the V:/V ratios. In both N and S, C:/C ratios are much lower than V:/V ratios.

If V:/V and C:/C ratios are compared for all recording locations in the N and S cluster, it becomes clear that V:/V ratios are generally much higher. For V:/V ratios, the lowest observed value is 1.3. 75% of the dialects show V:/V ratios above 1.7. The lowest C:/C ratio observed is 1.0, and 75% of the C:/C values lie *below* 1.3.

Interestingly, the large difference in C:/C ratios between F and N is clearly not an effect of shorter C:, but of longer C. N and S however, are mainly distinguished by C: duration. The shorter C: durations in S also lead to lower C:/C ratios, although this effect is partly compensated by the shorter C durations in S.

Regarding the ratios between the segments in succession, the most clearly marked differences show up in the V:/C ratios. Here the longer V: and shorter C in F lead to a much higher ratio than in the N and the S cluster. V: durations in F are up to two times longer than the following C durations. N and S are much closer in their ratios, but S has generally somewhat higher vowel-to-consonant ratios than N, for both V:/C and V/C: ratios. This is mainly caused by shorter C and C: and not longer V: and V durations.

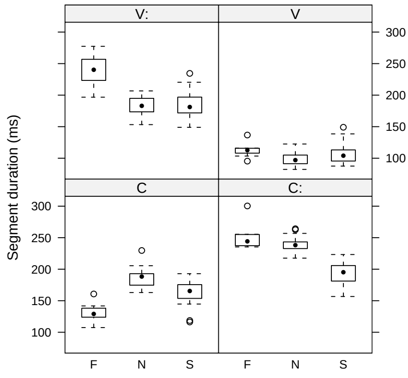


Figure 8.10: Duration of the four segments in the three Ward clusters.

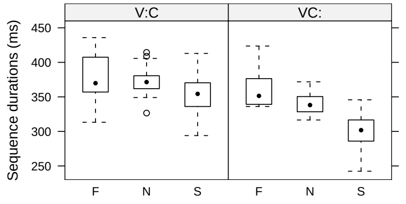


Figure 8.11: Duration of the two sequences in the three Ward clusters.

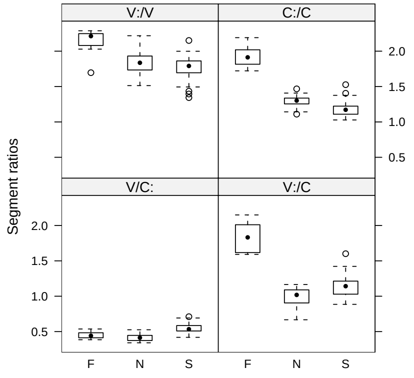


Figure 8.12: Ratios of the four segments in the three Ward clusters.

8.3.2 Subcategorisations in the South

At the 14th level, the S cluster consists of six sub-clusters. These have been named S1a, S1b, S1c, S2, S3a and S3b. The dendrogram in Figure 8.3 shows that the S cluster is first split into S1 and S2 versus S3. Then S2 and S1 are separated. S1 is split into three sub-clusters, and S3 in two. At the 14th level, the clusters differ considerably in size. There are two very small clusters with only two members each: S1c and S3b. The rest of the sub-clusters are larger in size.

Figure 8.13 shows the absolute segment durations in the six sub-clusters, Figure 8.14 shows the sequence durations and Figure 8.15 shows the respective ratios. The two dialects in S3b are separated from the others by very short C durations. The other durations are similar to the larger cluster S3a. Together, the S3 clusters tend to show shorter C: durations, and also V: and V durations are rather short in the S3 clusters. The S1 and S2 clusters share longer C: durations in comparison to the S3 clusters. V and V: durations tend to be longer in the S1 clusters than in the S2 cluster. S2 has longer C durations than the S1 clusters. The two dialects in S1c are separated from the rest of the S1 dialects mainly by longer V durations.

If the sequence durations are also taken into account, it can be seen that the S3 clusters have generally the shortest sequence durations. If the very small clusters S1c and S3b are ignored, there is not a clear effect of V: or C alone that causes these shorter sequences in S3a. For S3a versus S2 a longer C in S2 seems to play a role. For S3a versus S1b, mainly the V: seems to be longer in S1b, while the difference between S3a and S1a seems to be caused by both segments.

The difference between the two larger S1 clusters seems to mainly depend on V:, with S1b showing the longer V: and thus the longer sequence. Concerning VC:, the difference between S3a and S2 is again a consonant difference. S2 and S1b seem to be mainly distinguished by V duration, while S2 and S1a show opposite differences for V and C: duration (with V shorter and C: longer in S2), leading to similar VC: durations.

If the V:/V and C:/C ratios are considered, it can be seen that the four larger clusters show very similar V:/V and C:/C ratios, with a tendency for S1a to show somewhat lower V:/V ratios than the other three clusters. S2 has the lowest V/C: and V:/C ratios, which is a consequence of the comparatively long consonant in this cluster. S3a is distinguished from S2 mainly by shorter C: and C, and has thus somewhat higher V/C: and V:/C ratios. S1a shares with S3a similar V:/C ratios, because both, V: and C are somewhat longer in S1a. S1b has rather long V: and rather long C:, leading to low V/C: but comparatively high V:/C ratios.

Until now, the two small clusters S3b and S1c have been ignored. For S3b, the cause of the exceptional position is relatively straight-forward: C durations in the two recording locations in S3b are clearly shorter than in the other Southern clusters (see Figure fig:S3segbox). C durations under 150 ms are otherwise only found in the Finland-Swedish area. S1c is mainly distinguished from the other clusters by longer V durations.

While the reason for the longer V durations in the S1c cluster is still an open question, the very short C durations in S3b were further investigated through auditive analysis and inspection of the sound wave. For one of the S3b locations, Skee (ske) in northern Bohuslän, voicing of the consonant after the long vowel is probably the reason for the unusually short C durations, as voiced plosives are usually shorter than unvoiced ones.

Auditive and acoustic analysis of the dialects in the S cluster showed that voicing of consonants after long vowels occurred in seven additional locations: Orust (oru) and Kärna (kaa), also in Bohuslän, Årstad-Heberg (ars) in Halland, Össjö (oss), Norra Rörum (nro) and Löderup (lod) in Skåne and Jämshög (jam) in Blekinge. These recording locations lie in an area that is known to have voiced consonants after long vowels (see Section 5.5). However, the auditive analysis showed that not all speakers of these dialects produced voiced consonants. As medians were used in the present study for the description of the durational patterns in the dialects, occasional voicing of a few speakers had no effect on the values. In Skee, voicing was more frequent, and thus influenced the median.

The situation is different for the other recording location in S3b. This location, Södra Finnskoga (fin) in Värmland, did not have voicing of the consonant, but still very short C durations. For the moment, we have no explanation for this exceptional pattern.

Although the geographical distributions of the clusters on the 14th level in the South did not seem completely random, it is hard to estimate which significance should be attached to the observed durational differences. The segment durations and the relations of the durations to each other do not suggest fundamental differences in the quantity systems of these dialects, especially if only the larger clusters are taken into account.

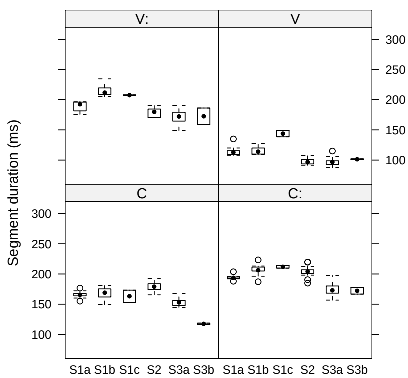


Figure 8.13: Duration of the four segments in the Southern sub-clusters.

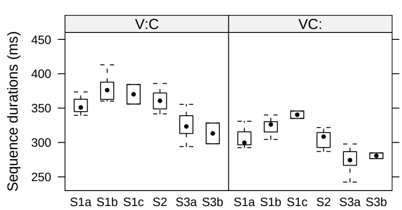


Figure 8.14: Duration of the two sequences in the Southern sub-clusters.

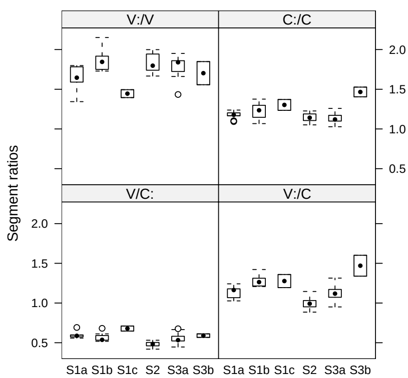


Figure 8.15: Ratios of the four segments in the Southern sub-clusters.

8.3.3 Subcategorisations in the North

At the 14th level, the N cluster was split up into five sub-clusters. Figure 8.16 shows the absolute segment durations in these sub-clusters, Figure 8.17 shows the sequence durations and Figure 8.18 shows the respective ratios. Two sub-clusters were very small; N2a consisted of a single recording location, Arjeplog (arj) in Lappland. N1a consisted of three recording locations with no obvious geographical relation. For the larger clusters, only N1b showed a clear geographical correlation. In the following, only the three larger clusters N1b, N1c and N2b are compared.

N1b, N1c and N2b show very similar C: durations. C durations are similar as well, although N1c shows a tendency towards somewhat shorter C. V: durations are shortest in N2b, while N1b and N1c show a small V: difference that matches the small C difference between these two clusters. Regarding V, N1b shows longer durations than the other clusters.

V:C is somewhat longer in N1b than in N1c, because of the longer V: and C. N2b has slightly shorter V: and slightly longer C than these two clusters, resulting in V:C durations of similar duration. VC: durations are longer in N1b, because of the longer V.

These longer V also result in lower V:/V ratios for N1b. Because of the slightly shorter V: and longer C, N2b has somewhat lower V:/C ratios. Otherwise there are no remarkable differences between the three clusters.

Again, the significance of the differences is hard to estimate. Only N1b has a clear geographical correlation. Thus, the subcategorisations in the North seem to reveal at least one stable effect for a certain area: V durations tend to be longer than in surrounding dialects.

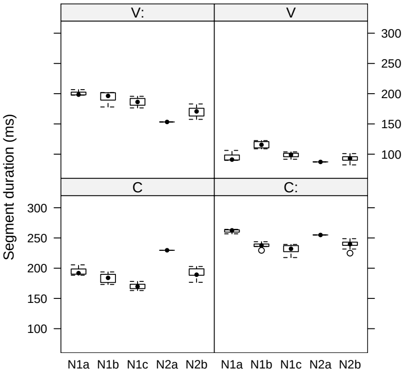


Figure 8.16: Duration of the four segments in the Northern sub-clusters.

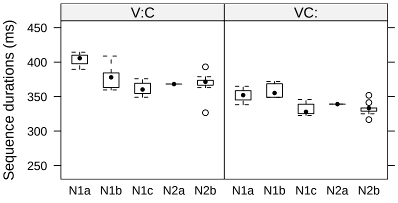


Figure 8.17: Duration of the two sequences in the Northern sub-clusters.

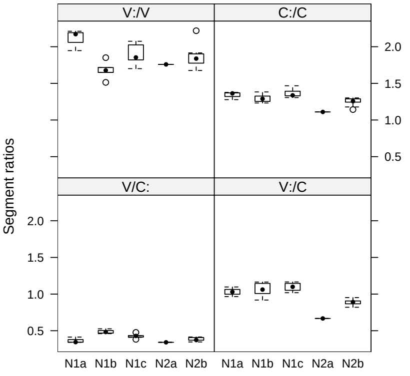


Figure 8.18: Ratios of the four segments in the Northern sub-clusters.

8.3.4 Subcategorisations in Finland

The rather small Finland-Swedish cluster has three sub-clusters at the 14th level. Figure 8.19 shows the absolute segment durations in these sub-clusters, Figure 8.20 shows the sequence durations and Figure 8.21 shows the respective ratios. One of the Finland-Swedish sub-clusters, F2b, included just a single recording location, Kyrkslätt (kyr) in Nyland. The other two clusters separated the Southern-Finland-Swedish dialects in Åboland and Nyland (F1) from the northern Finland-Swedish dialects in Österbotten (F2a). There is thus a geographical correlation in spite of the rather small cluster sizes.

In terms of segment durations, F2b shows the longest C: duration of all dialects, while F1 and F2a are rather similar in terms of C: as well as C durations. F2a and F2b have similar V: durations, while F1 has slightly shorter V: durations. V durations are similar in all three Finland-Swedish sub-clusters, although one of the F2a locations shows a higher value.

The three clusters differ considerably in V:C and VC: duration. F1 durations are always shortest and F2 durations always longest. In spite of these differences, all ratios are quite equal, with the exception of V:/C for F2a, which results from a combination of longer V: with respect to F1 and shorter C with respect to F2b.

As mentioned before, the Finland-Swedish clusters are the only ones where V:/V and C:/C ratios lie in the same order of magnitude. The two Österbotten dialects (F2a) are mainly separated from the four F1 dialects by V: duration.

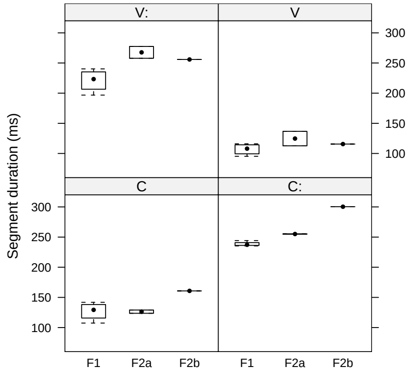


Figure 8.19: Duration of the four segments in the Finland-Swedish sub-clusters.

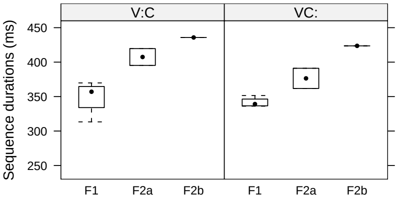


Figure 8.20: Duration of the two sequences in the Finland-Swedish sub-clusters.

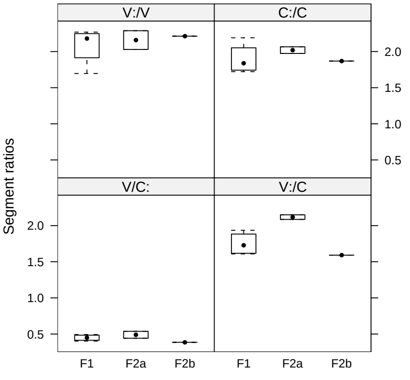


Figure 8.21: Ratios of the four segments in the Finland-Swedish sub-clusters.

8.4 Discussion of the Cluster Analysis

The cluster analysis resulted in a clear tripartitioning of the data, corresponding to easily interpretable geographical regions.

All further partitionings are subpartitions with a generally more scattered geographical distribution. According to their geographical distribution, the three main clusters were called the “Finland-Swedish cluster” (F), the “Northern cluster” (N) and the “Southern cluster” (S).

The F cluster included seven recording locations on the Finnish mainland. The N cluster includes 27 recording locations. With the exception of three recording locations, the Southern border of the N cluster runs from Northern Dalarna in the West via Hälsingland, Gästrikland to Northern Uppland in the East of the Swedish mainland. Two Finland-Swedish recording locations, Brändö (bra) and Saltvik (sal), located on the Åland islands off the South-Western Finnish coast, belong to the N cluster as well. The S cluster includes 52 recording locations. With the exception of three locations, Aspås (asp) in Jämtland, Anundsjö (anu) in Ångermanland and Torp (tor) in Medelpad, all of them are located South of the border of the N cluster described above, including also the islands Gotland and Öland.

The durational factors behind this tripartitioning are mainly the consonant durations. The F and the N cluster are separated from the S cluster by longer C: durations, while the F cluster is separated from the N cluster (and also from most of the dialects in the S cluster) by shorter C durations. Long vowel durations should also have a minor influence on the tripartitioning, as they tend to be longer in the Finland-Swedish region than in the Northern region and most of the Southern region. Short vowel duration, however, obviously had no great effect on the tripartitioning (see Figure 8.10 (b) and the R^2 value for V with the three-cluster split in Figure 8.2). As a consequence of the differences in segment duration, there is also variation in the sequence durations across the three main groups.

In general, Finland-Swedish dialects combine rather long V: with very short C, Northern dialects and many Southern dialects combine shorter V: with rather long C and some Southern dialects combine short V: with rather short C, resulting in the shortest V:C durations of all dialects. VC: sequence durations show greater differences between the three main clusters. This is a consequence of the shorter C: durations in the South, which is not generally combined with longer V durations.

The durational characteristics of the segments involved lead to great differences in C:/C ratios. The high C: durations, combined with low C durations in the Finland-Swedish dialects lead to high ratios. The Northern dialects have C: durations that are about as long as those of the Finland-Swedish regions, but longer C durations. This lowers the C:/C ratios quite considerably in comparison to the Finland-Swedish dialects. The functional implications of these relationships are discussed in Section 11.3.

Chapter 9

Preaspiration and Vocalic Quality

9.1 Preaspiration in the A-class Words

In Section 5.6, preaspiration in the Swedish dialects was described. It turned out that preaspiration is a normative feature in some Swedish dialects, but also a common non-normative phenomenon in Standard Swedish and presumably also in many dialects.

As described in the segmentation section, preaspiration is not always identified without problems, in particular in recordings of varying acoustic intensity and interfering echo effects. Furthermore, preaspiration often does not occur in every utterance of a speaker, making mean or median calculations cumbersome[1]. Therefore, it was decided to combine measurements of preaspiration duration with measurements of preaspiration frequency.

9.1.1 Measures

Preaspiration frequency was calculated as the ratio between preaspirated utterances and the total number of utterances per recording location. For preaspiration duration, first the median duration of preaspiration for preaspirated utterances of a speaker was calculated, and then the median of the speaker medians was used as the measure of central tendency for each recording location.

9.1.2 Results

Five of 86 recording locations had no preaspiration markings at all, and three recording locations showed preaspiration markings only after the short vowel. The preaspiration showed inverse tendencies for duration and frequency after long and short vowels, respectively. The median duration of preaspiration was longer after long than after short vowels (/ah/ 31 ms; /ɑːh/ 38 ms). This difference was significant, according to a Wilcoxon rank sum test[2] (W=1915,5,

[1] To give an example: a speaker might produce four utterances without preaspiration and one utterance with a preaspiration of 40 ms. In this case neither the mean (8 ms) nor the median (0 ms) is a good description of what is going on.

[2] As the distributions deviated significantly from a normal distribution according to a Shapiro-Wilks test, a non-parametric test was chosen for this comparison.

p-value $<$ 0.01). Preaspiration was however more frequent after short than after long vowels (/ah/ 34.7%; /ɑːh/ 18.3%). Again, a Wilcoxon test revealed significance for this difference (W=4561.5, p$<$0.01).

For locations with preaspiration, the relative frequency of preaspiration varied between 1.4% and 93.6% for the short vowel context, and between 1.0% and 71.7% for the long vowel context. The range of median preaspiration durations was 10.5 to 88.5 ms for the short vowel context and 22.2 to 103.4 for the long vowel context.

9.1.2.1 Preaspiration in the Ward Clusters

Preaspiration was first analysed with respect to the three main clusters resulting from the cluster analysis. The Figures 9.1 and 9.2 show the distribution of preaspiration duration and frequency in the V:C and VC: context. The Preaspiration durations do not show a clear cluster dependent distribution. Although the mean values vary somewhat from cluster to cluster, the spread within the clusters is considerable. The situation is different for the frequencies. The N cluster has higher preaspiration frequencies in the V:C as well as in the VC: context.

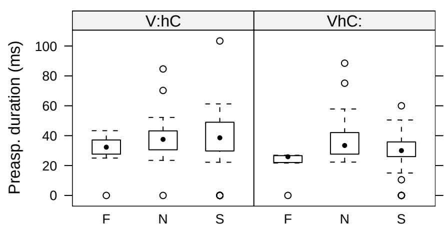


Figure 9.1: Preaspiration duration by Ward cluster.

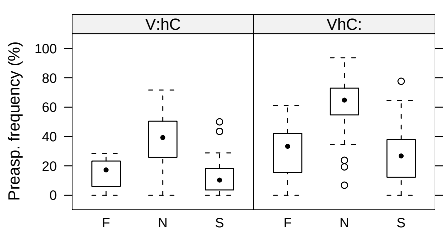


Figure 9.2: Preaspiration frequency by Ward cluster.

9.1.3 Geographical Distribution

In the previous section, the preaspiration measures were related to the cluster structure obtained from the segment duration values. It is quite possible, however, that the geographical distribution of preaspiration (in terms of frequency and duration) shows a geographical distribution independent of the quantity structure. Therefore, the geographical distribution of preaspiration was also analysed without a reference to the clusters.

First, it was analysed whether certain recording locations have deviating values with respect to preaspiration. Figure 9.3 shows preaspiration frequency as a function of preaspiration duration for V:C, and Figure 9.4 shows both variables for VC:. Recording locations with deviant values have been labelled with their three-letter short forms.

For the V:C context, the plot confirms the exceptional position of Arjeplog (arj), if duration and frequency are considered together. Although some recording locations have similar duration or frequency values, none shows similarly high values in combination. One location, Burseryd (brr) in Småland, has an exceptionally long preaspiration duration, but a low frequency (around 10%). Some locations, for example Ovanåker (ova) in Hälsingland, have high preaspiration frequencies, but shorter durations. Gräsö (grs) in Uppland, which is known for its normative preaspiration, is also separated in terms of higher duration values, while the frequency values lie well within the normal range.

For the VC: context, Vemdalen (vem) in Härjedalen shows a similar separation from the other dialects as Arjeplog in the V:C context. Exceptionally high duration values are combined with exceptionally high frequency values. Besides Vemdalen, Gräsö is another recording location with high frequency and duration values, which corresponds to reports in the literature as well. Five additional places are worth mentioning. Fjällsjö (fja, Ångermanland), Ovanåker (ova, Hälsingland) and Arjeplog (arj, Lappland) show very high frequencies, but durations are shorter than in Vemdalen and Gräsö. Piteå (pit, Norrbotten) comes close to Gräsö in terms of duration and frequency. Finally, Särna (sar, Dalarna) is exceptional in combining low frequency with extremely long duration values.

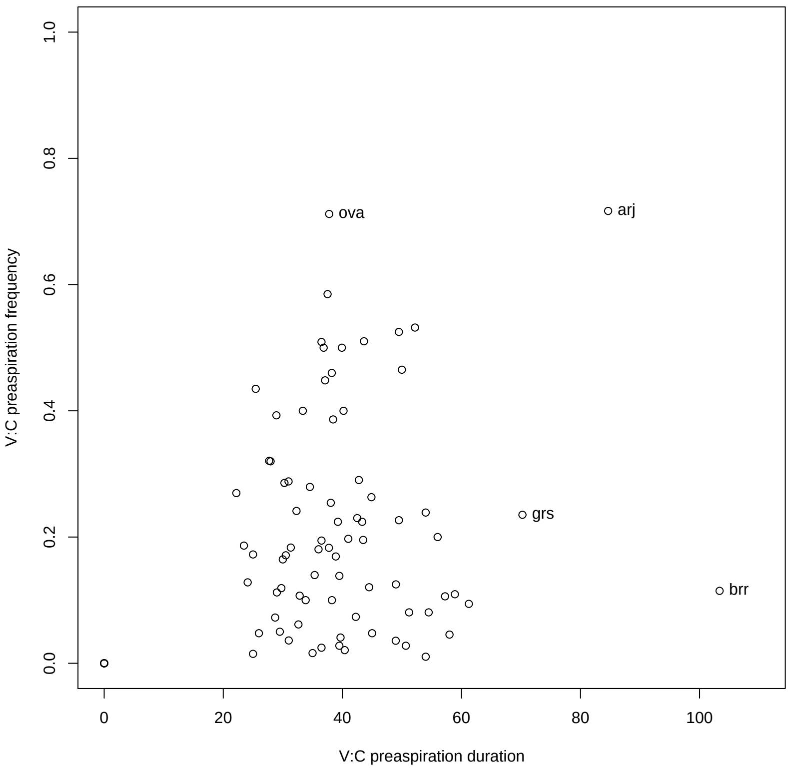


Figure 9.3: Preaspiration in *tak* per recording location. Preaspiration frequency as a function of preaspiration duration.

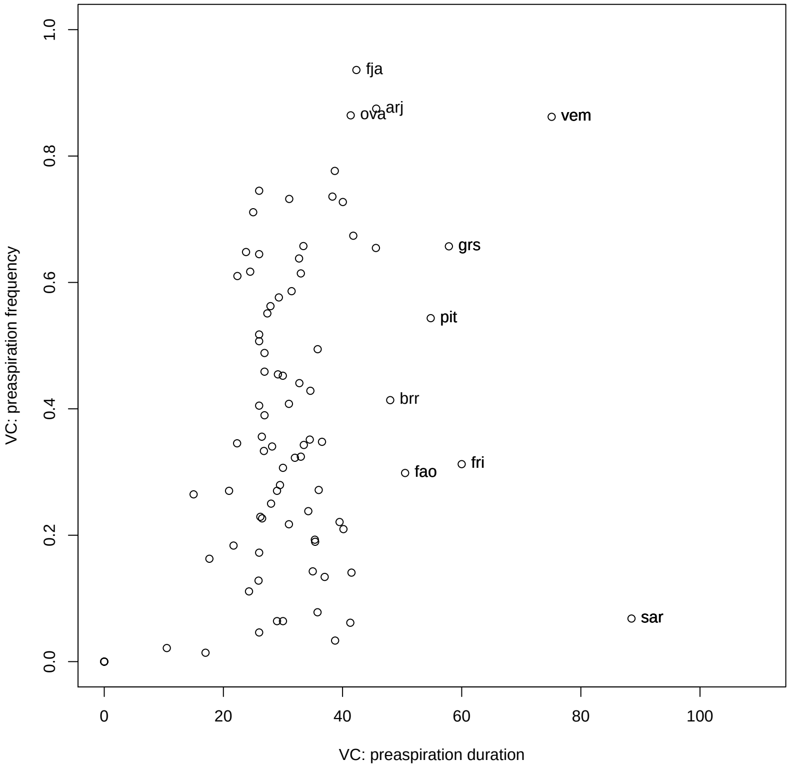
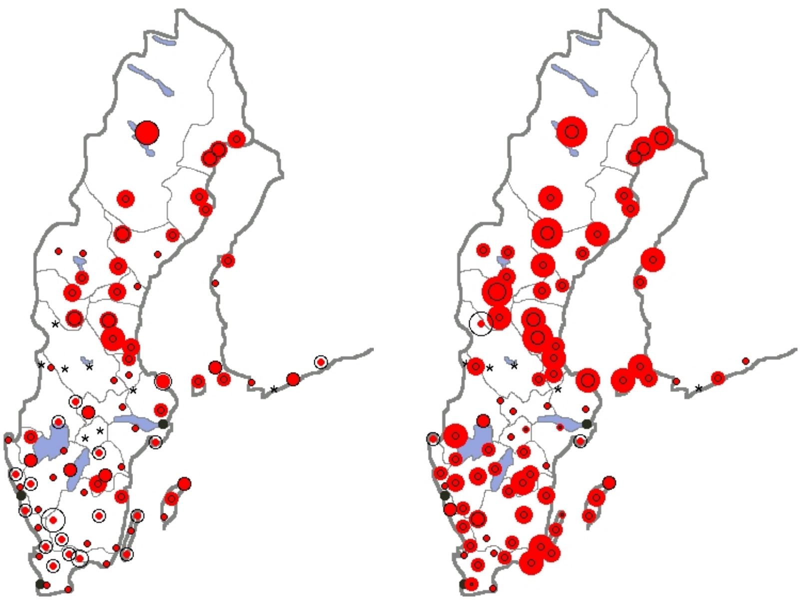


Figure 9.4: Preaspiration in *tack* per recording location. Preaspiration frequency as a function of preaspiration duration.

The geographical distribution of preaspiration frequency and median duration for both contexts is shown in Figure 9.5. In these maps, the size of the red dots reflects the relative frequency of preaspiration, while the size of the black circles reflects the median duration of preaspiration. Places without preaspiration are marked with stars. To create the map, the preaspiration measures were classified into six classes (relative frequencies: 0, 0-0.2, 0.2-04, 0.4-0.6, 0.6-0.8, 0.8-1; durations (ms): 0, 0-20, 20-40, 40-60, 60-80, >80). The affiliation of each recording location to these classes was then expressed in dot and circle size. Larger dots or circles reflect higher values. Both figures reflect the higher frequency values in the Northern areas already observed with the Ward clusters.



(a) Preaspiration in V:C

(b) Preaspiration in VC:

Figure 9.5: Preaspiration in *tak* and *tack*. The size of the red dots reflects the relative number of preaspirated utterances, the size of the black circles reflects the median duration of preaspiration.

9.1.4 Discussion of Preaspiration

In general, the preaspiration patterns strongly supported the suggestion of Helgason (1999) that preaspiration is a frequent non-normative feature in Swedish. Furthermore, it was found that most of the locations where preaspiration has been reported as a normative feature (see Section 5.6) tended to separate from the other dialects by preaspiration duration and/or frequency. This was clearly the case for Vemdalen (vem), Gräsö (grs) and Arjeplog (arj). Särna (sar) in Northern Dalarna showed a long preaspiration duration, but a low frequency. This might indicate that a part of the speakers showed the normative pattern. For Lillhärdal (lil), the values in both contexts did not clearly separate from the rest of the dialects, although even here comparatively high values were reached (V:C duration: 43.6 ms; V:C frequency 0.5; VC: duration 25 ms; VC: frequency 0.7). Some locations outside the typical “preaspiration regions” had very high preaspiration values as well, for example Ovanåker (ova) in Hälsingland. These locations and dialects await further investigation.

In general, the data showed more frequent preaspiration for the Northern area than for Southern Sweden. However, it has to be taken into account that preaspiration is a feature that is rather difficult to recognise in the speech signal, especially with recordings of varying quality and accompanying echo effects. This might even have caused different interpretations by different groups performing the segmentation. It is conceivable that preaspiration was just recognised more frequently by research assistants performing the segmentation in the Northern area.

Therefore, further investigations of preaspiration in the dialects should be performed with an extended data set under more controlled conditions. Though beyond the scope of this thesis, the data presented here can be used as a starting point for new investigations.

9.2 Vocalic Quality in the A-class Words

Spectral differences between the long and the short vowel in the A-class words were described by the “acoustic distance” (AD) measure (see section 7.3.4.2). The minimal median AD measured was 0.67, the maximum was 3.8. In the following, the vocalic quality differences in terms of the AD measure, will be discussed with respect to the three-cluster split from the quantity analysis. Figure 9.6 shows the distribution of the AD values by Ward cluster. The figure suggest a major difference between the F cluster and most other dialects. The AD values in the F cluster lie consistently below 1.5 bark. Similarly low values are also reached by some dialects in the other clusters, but the great majority of recording locations has values above ca 2.5.

The great spread of the values within the N and the S cluster indicates that the AD values might have a geographical distribution that is different from the durational values. To investigate this, the AD values were organised in four classes: values below 1.5 bark (class 1), values between 1.5 and 2.5 bark (class 2), values between 2.5 and 3.5 bark (class 3) and values above 3.5 bark (class 4).

Figure 9.7 shows the geographical distribution of these four classes. The map reveals some areal concentrations. The Finland-Swedish region has very low AD values, as was already derived from the cluster-specific analysis. The dialects of Brändö (bra) and Houtskär (hou) on the Finnish coast show the same low AD values as the dialects on the Finnish mainland. This contrasts with the results from the cluster analysis, where these two recording locations did not group together with the dialects on the Finnish mainland, but with the Northern clusters (cf. Figure 8.5).

A middle West Swedish area, comprising recording locations in Southern Jämtland, Härjedalen, Western Hälsingland, Western Dalarna, Värmland and even extending into Västergötland, shows lower quality differences (class 1 and 2) than the rest of the Swedish mainland dialects (mostly class 3). Low AD values are also found on the island of Gotland, in accordance with dialectological sources that describe the Gotland dialect as a dialect that has not performed the usual backing of the long a-sound (Pamp, 1978). Exceptionally high differences

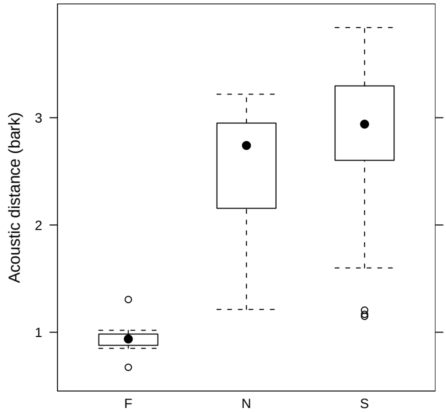


Figure 9.6: Acoustic distance by Ward cluster.

(class 4) are found in some locations of South Eastern Sweden.

9.3 The Relation Between Quality and Quantity

Studies like Hadding-Koch and Abramson (1964) and Engstrand and Krull (1994) suggest that there might be a trading relationship between the durational and the spectral distinction of vowels. To test this hypothesis, the V:/V ratios of the A-class material were correlated with the median AD values. As the distributions of both variables deviated significantly from a normal distribution (Shapiro-Wilks test, Vowel ratios: W=0.9713, p=0.05, AD: W=0.9062, $p < 0.001$), Spearman's ρ was calculated. The test revealed a weak negative correlation ($\rho = -0.21$) which was close to significant (p=0.058).

As it has often been suggested that complementarity might be a feature aiding quantity perception, the relation between C:/C ratios and AD was investigated as well. Consonant ratios and AD showed a ρ-value of -0.37 ($p<0.01$). Thus, there seems to be a stronger relationship between C:/C ratios and vowel quality than between vowel ratios and vowel quality. However, this might be an effect of the Finland-Swedish cluster alone. If this cluster is removed from the correlation calculation, ρ equals -0.19 and falls short of significance (p= 0.09). For the relation of the vowel ratio and AD, the exclusion of the Finland-Swedish clusters has a similar effect. Spearman's ρ drops to -0.06 and is far from significant (p=0.6).

9.4 Discussion of Vocalic Quality and its Relation to Quantity

Except for the Finland-Swedish dialects, the geographical distribution of the quality data did not reveal a systematic relationship to the distribution of the durational values. Overall, the

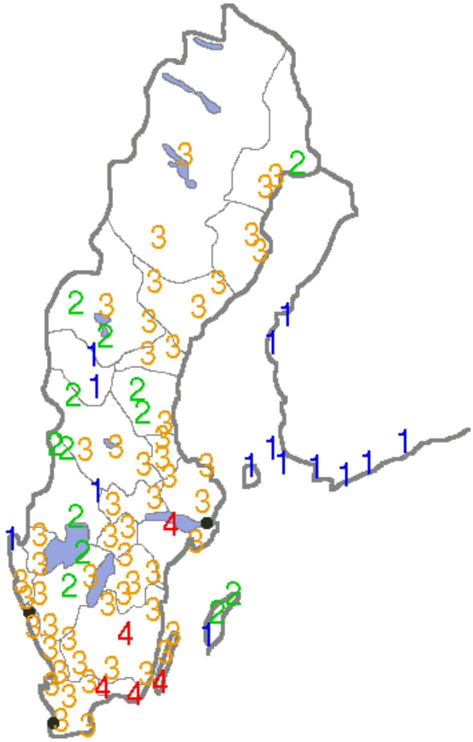


Figure 9.7: Geographic distribution of the acoustic distance (AD) between long and short vowels. Numbers 1 to 4 denote classes of increasing acoustic distance.

results suggest that there is only a weak — if any — relationship between acoustic distance and vowel ratios. At first sight, the relationship between acoustic distance and consonant ratios appears to be stronger. However, the observed effect is mainly caused by the Finland-Swedish dialects, which combine higher vowel and consonant ratios with lower acoustic distances. It remains to be explored whether the observed negative correlation is based on intrinsic connections between vowel quantity and quality, or whether this relation should rather be interpreted as a coincidence and vocalic quality and quantity follow more or less independent developments. This question will be discussed further in Section 10.2.8.

Chapter 10

Verification of the Durational Categorisation

The cluster analysis in Chapter 8 was based on a single word pair. It was therefore decided to verify the results of this analysis by the investigation of additional word pairs. O-class word pairs were recorded for the whole SweDia recording area. As mentioned before, the I-class, Y-class, U-class and E-class words were only recorded in parts of the Umeå recording area, including two locations in the Finland-Swedish area. The cluster analysis was therefore verified by two separate studies: the first one investigated the O-class patterns for the whole SweDia recording area. The second one was restricted to the Northern and Finland-Swedish region and was based on the entire set of word pairs.

10.1 The O-Class Words

The material included measurements from 210 speakers, originating from 75 different recording locations with usually 3 speakers from the older men group per recording location. The material was investigated with respect to the three main clusters. For every speaker, median V:, V, C: and C duration was measured and the V:/V and C:/C ratios were calculated from the absolute durations. The distributions of the speaker medians for the four segments are shown in Figure 10.1. The distributions of the ratios are shown in Figure 10.2.

With regard to the results from the investigation of the A-class, the following hypotheses were formulated: C: durations in the F and the N group should exceed those of the S group. C durations should be shorter in the F groups than in the other two groups. C:/C ratios should, as a consequence, be highest in the Finland-Swedish dialects and lowest in the Southern dialects. For the vowel durations, no significant effects were expected because of the great variation within the Southern, Northern and Finland-Swedish cluster. As only the three largest clusters were compared, the comparisons were done on the speaker level and not on the recording location level. Because of deviations from normality in the samples and the great differences in sample size, the non-parametric Mann-Whitney U-test was used to statistically test the hypotheses. The sample sizes of the three groups were 18 speakers (F), 65 speakers (N) and 124 speakers (S).

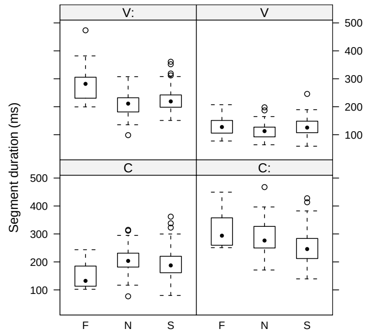


Figure 10.1: Absolute segment durations for the O-class words in the three main clusters.

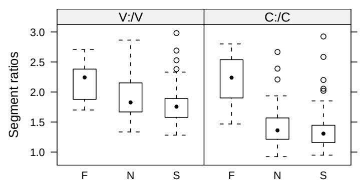


Figure 10.2: Ratios of vowels and consonants for the O-class words in the three main clusters.

10.1.1 Results

Table 10.1 shows the differences between the medians for each measure. The expected differences could be confirmed. Even on the speaker level, F dialects show much longer C: duration than S (difference around 50 ms), and much shorter C durations than both N and S (difference 50–70 ms). Consequently, the speakers of the F area show higher C:/C ratios. The differences between N and S are not as pronounced, however. C: are significantly longer in the North than in the South, as are C, albeit the absolute difference is about half as high as the C: difference. C:/C ratios are not higher in the North than in the South.

For the vowels, V: durations are considerably longer in F than in N and S (60–70 ms), confirming the trend observed with the A-class. V durations are longer in F than in N, with a median difference of ca. 15 ms. V durations in F and S show no significant difference. For N versus S, both V: and V durations are somewhat shorter in N. Thus, the differences between N and S are — at least partly — compensatory. Shorter vowel durations occur together with longer consonant durations. Overall, however, the differences between N and S were only minor, and C:/C and V:/V differences were always low, albeit sometimes significant.

Table 10.1: Median differences for six variables across the three main regions. Stars indicate significance levels (* p<.05, ** p<.01, *** p<.001).

	F - N	N - S	F - S
V:	70.6***	−7.85*	62.75***
C	−71.3***	16.15*	−55.15***
V	14.6*	−12.6**	2
C:	17.4	30.5***	47.9***
V:/V	0.42**	0.07*	0.49***
C:/C	0.88***	0.05	0.93***

10.2 A Closer Look at the Three Main Types

10.2.1 The Extended Data Set in the Northern Recording Area

For the Umeå recording area, an extended data set was recorded, as mentioned in section 7.2. In these parts of Northern Sweden, not only the A- and O-class words were recorded, but also the words of the E-, I-, U- and Y-class. The recording locations with extended data sets were initially evaluated through a listening test by the author. During this test it turned out that the speakers did not produce the expected target forms in all recording locations. In some cases, certain words were not used in a dialect. For example, the Standard Swedish word for 'crying', *gråta*, is not used in large parts of Northern Sweden. Other variations occurred due to the fact that the words in the extended data set were morphologically complex. This affected for example the *tät* versus *tätt* word pair. Standard Swedish has two genders, called "common gender" ("utrum" in Swedish) and "neuter". The common gender is a fusion of older masculine and feminine forms. In Standard Swedish, the adjective corresponds with the noun in both attributive and predicative position. For the neuter form, usually a /t/ is added to the stem of the adjective. If the adjective ends in a long vowel + /t/ in the common gender, the neutral form is usually expressed by shortening the vowel and lengthening the /t/.

During the listening test it turned out that not all Northern Swedish dialects showed this grammatical structure for the target words in question. Instead of varying the vowel-consonant sequence from V:C (common gender) to VC: (neuter), speakers of these dialect always used the VC: form, even in connection with common gender nouns. The assessment, that in fact identical forms were used here, was made by auditory judgement and visual inspection of spectrograms by the author. For selected cases, the durational measurements were inspected as well, as differences between long and short consonants are not always easily detectable. The assessment was also supported by the fact that several speakers in the recordings report on the fact that these two forms are actually "the same".

Another complication occurred with the perfect participle forms *bott* of the verb *bo*. Several speakers used here a form with a voiced final plosive instead of the Standard Swedish unvoiced one. The imperative *byt* of the verb *byta* showed also forms different from Standard Swedish in some recording locations. Table B.1 in the appendix provides a list of all Northern recording locations and shows whether the speakers of a certain location used the target V:C and VC: forms. It should be mentioned however, that this table presents a simplification. In many cases, not all speakers of a certain recording location produced identical word forms.

These differences to the Standard Swedish grammar posed a problem for the current investigation, as the expected V:C versus VC: structures with otherwise comparable sound qualities could not be investigated. This reduced the selection of possible recording locations for the following comparison, as it was intended to choose recording locations where the full set of target forms was available.

10.2.2 A Comparison of Four Recording Locations

Section 10.1 verified the differences between the Finland-Swedish and the mainland Swedish dialects, and — with some restrictions — also those between the Northern and the Southern Swedish types. In this section, the analysis was continued by looking at four selected recording locations in detail. This procedure was motivated by the following considerations: First, the data from the Umeå recording area provided an extended data set for several locations in the Northern Swedish and Northern Finland-Swedish area, allowing a more detailed investigation. Second, as some of the geographically Northern recording locations had clustered together with the Southern dialects, the extended data set allowed in principle the comparison of dialects from the three main clusters.

Third, the two recording locations Vörå (vor) and Närpes (nap) are of special interest: As already described (see Section 5.2), the dialectological literature reports different quantity structures for Närpes and Vörå: Närpes shows a 3-way system, Vörå a 4-way system, and previous analyses of the SweDia quantity corpus (see Strangert and Wretling, 2001 and Schaeffler et al., 2002) have already shown that some of the Vörå target words in the extended data set show V:C: sequences. Contrary to these differences in phonological structure, the cluster analysis — which was based on V:C and VC: structures — grouped these two recording locations in the same cluster, even at the 14th level.

From the Northern area, two recording locations were chosen where the extended data had been recorded and showed the expected structure. Furthermore, one of the recording locations should represent the Northern cluster, the other one the Southern cluster. This constrained the choice considerably (see section 10.2.1). The two places Indal (ind) and Torp (tor) lie in close vicinity in the province of Medelpad. The cluster analysis grouped Torp in the Southern cluster and Indal in the Northern cluster. Thus, these two recording locations appeared to be appropriate candidates for the detailed analysis.

For these four places (Vörå, Närpes, Indal and Torp), the E-class, O-class, I-class and Y-class words were segmented for all speakers, following the conventions outlined in Section 7.3.2. The word pair *bot* versus *bott* was excluded, as the Finland-Swedish dialects use a voiced long consonant for the form *bott*. The remaining data was available for 12 speakers from Vörå, Närpes and Indal, and for 10 speakers from Torp, where data from one elderly woman and one younger woman was missing. Besides durational variables (segment and sequence durations), F1 and F2 values were measured as well. The acoustic distance (AD) was calculated as described in section 7.3.4.2.

10.2.3 Results

Figure 10.3 (a–d) shows the distribution of the speaker medians for the four segment durations by word class and location. The figures do not indicate larger durational differences between different vowel or consonant qualities. All durations of a certain phonological class lie within similar ranges for a given recording location. Thus, intrinsic effects on vowel and consonant duration seem to have only minor influences in the current material. Remarkable differences within a vowel class are found for the V durations in Vörå. The V durations of the E-, I- and Y-class are much longer than those of the A- and O-class. This results from the V:C: structures found in this dialect (see Section 5.2). These vowels are thus phonologically long vowels. Further aspects, for example the durational differences to the long vowels in V:C in Vörå are described below.

Figure 10.3 (a-d) indicates that the differences observed for the A-class in Chapter 8 and for the O-class in Section 10.1 hold in similar ways for the rest of the word classes. The differences in segment duration across the four different places were tested with a parametric test (ANOVA). This procedure was chosen as the sample sizes were more similar in the current set, which should mitigate the effects of violations of the assumptions of an ANOVA (Howell, 2002). While the figures include the A-class words, they were excluded in the statistical tests, as they had been used for the generation of the hypotheses. For V and C: durations, the values from Vörå were excluded, thus only the values from Närpes, Indal and Torp were included in the test.

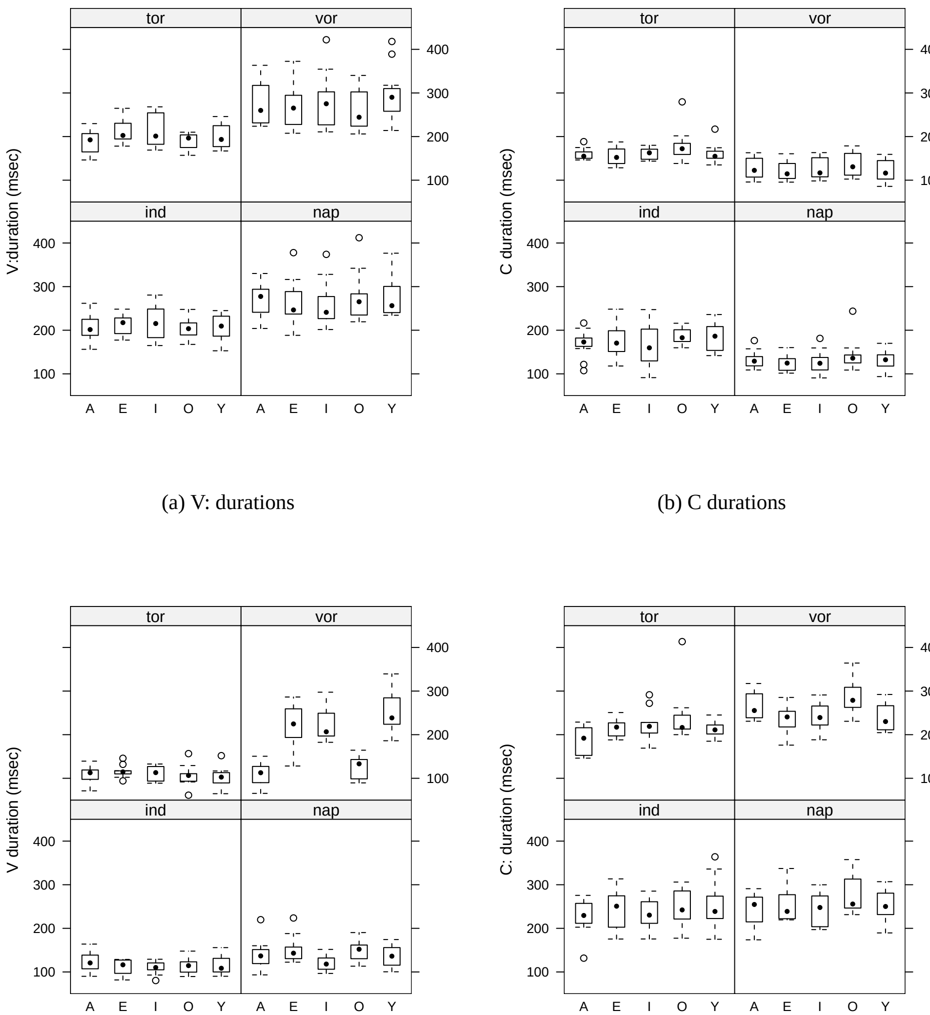
All four segment types showed significant differences across recording locations (V: duration: $F(3,177) = 34.5$, $p<.001$; V duration: $F(2,132) = 26.5$, $p<.001$; C: duration: $F(2,132) = 7.0$, $p<.01$; C duration: $F(3,177) = 37.3$, $p<.001$). Pairwise comparisons were done with a post-hoc Tukey test. The results of the pairwise comparison are shown in Table 10.2. The difference values in Table 10.2 for V and C: from Vörå (in parentheses) are based on the values of the A-class and O-class words in Vörå and were not tested for significance.

Table 10.2: Pairwise comparisons of segment durations for four recording locations. Stars indicate significance at the 5% level according to a post-hoc Tukey test.

Location pair	V: difference	C difference	V difference	C: difference
nap-ind	57.4*	−46.8*	26.4*	12.7
tor-ind	−5.7	−14.2	−3.4	−20.0
vor-ind	66.3*	−51.7*	(2.5)	(35.3)
tor-nap	−63.1*	32.6*	−29.8*	−32.7*
vor-nap	8.9	−4.9	(−21.8)	(22.4)
vor-tor	72.0*	−37.5*	(8.2)	(59.3)

Overall, Table 10.2 shows the following effects: Long vowels are generally much longer in Vörå and Närpes than in the other two places. The mean differences are around 60–70 ms. Differences within the Finland-Swedish dialects and the mainland Swedish dialects were minor and not significant. Similarly, C durations are much shorter in the Finland-Swedish dialects (30–50 ms). Again, differences between Vörå and Närpes were only minor and not significant, the same holds for the differences between Indal and Torp.

Regarding the short vowels, Närpes shows longer durations than the two mainland Swedish dialects (around 25 to 30 ms). The Vörå V durations are obviously closer to those of the mainland dialects. Long consonants in Närpes are not significantly longer than those of Indal, but are longer than those of Torp. Vörå C: (in VC:) seem to be even longer than those in



(a) V: durations

(b) C durations

(c) V durations

(d) C: durations

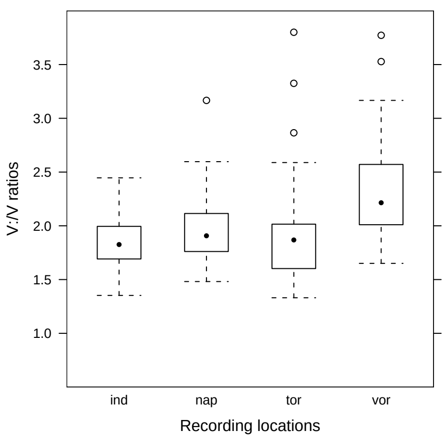
Figure 10.3: Segment durations in five different word classes for Torp (tor), Vörå (vor), Indal (ind) and Närpes (nap).

Närpes, but, as mentioned, this difference is only based on the values for the A- and O-class words. Indal showed longer C: than Torp, although this difference fell short of significance.

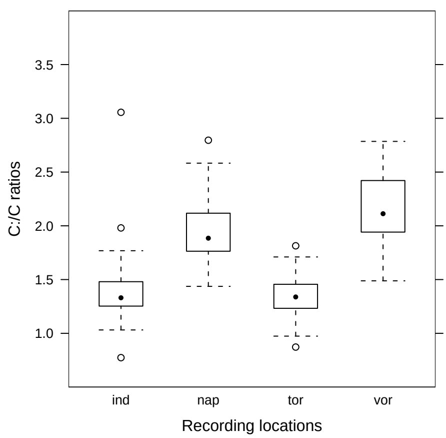
Generally, the durations of the four segments confirmed the earlier observations. Finland Swedish dialects show longer V: and shorter C than the mainland Swedish dialects. The places Torp and Indla are virtually equal in terms of long and short vowel duration, but Indal shows a tendency towards longer C:, although these differences were not significant in the current sample. The only unexpected finding were the longer V durations in Närpes.

10.2.4 Relations Between Segments

As the findings for absolute segment durations were similar to the earlier findings, it was expected that the resulting ratios were also similar. Figure 10.4 (a) and (b) shows the V:/V and C:/C ratios, respectively. V:/V ratios and C:/C ratios in Torp and Indal are very similar. V:/V ratios in Närpes are closer to those in Indal and Torp than to those in Vörå. This is clearly an effect of the longer V in Närpes, as V: durations in Vörå and Närpes were quite similar. C:/C ratios are higher in the two Finland-Swedish locations than in Torp and Indal. The absolute duration values for C: and C make clear that this is mainly an effect of a shorter C.



(a) V:/V ratios



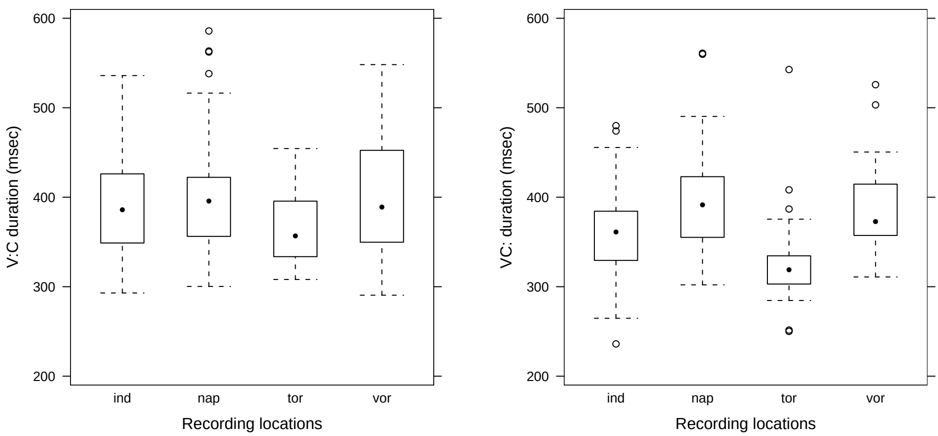
(b) C:/C ratios

Figure 10.4: Segment ratios for Torp (tor), Vörå (vor), Indal (ind) and Närpes (nap).

10.2.5 Sequence Durations

Figure 10.3 shows the V:C and VC: sequence durations The absolute V: and C durations for versus mainland Sweden within the V:C sequence are compensatory. This is confirmed by the median duration values for the V:C sequence duration (see Figure 10.5 (a)). Vörå, Indal and

Närpes show quite similar V:C duration values, in spite of the large differences in composition of these sequences. For the VC: sequences, the situation is different: VC: are longest in the two Finland-Swedish dialects, shorter in Indal and shortest in Torp (see Figure 10.5 (b)).



(a) V:C sequence duration

(b) VC: sequence duration

Figure 10.5: Sequence durations for Torp (tor), Vörå (vor), Indal (ind) and Närpes (nap).

10.2.6 The Case of Vörå

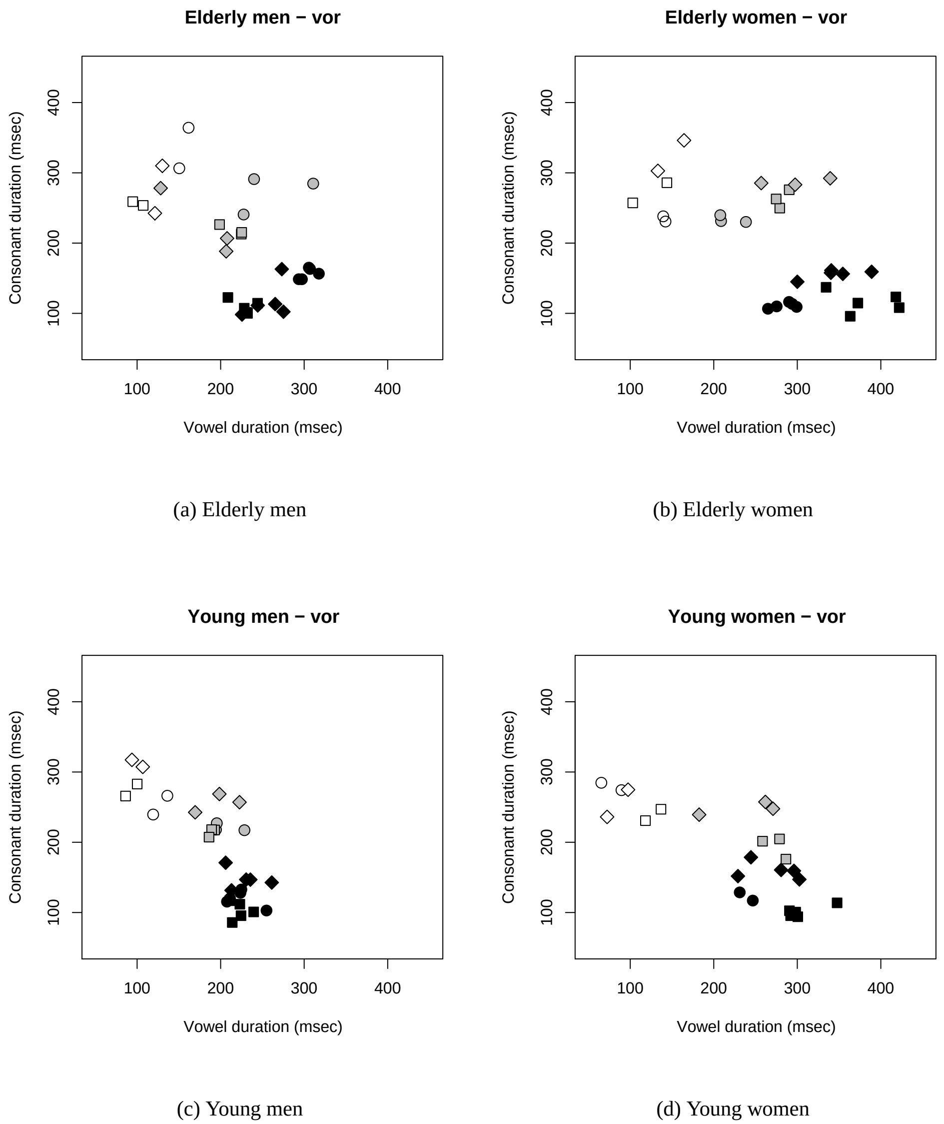
In section 5.2 it was described that V:C: structures are a regular feature of the Vörå dialect in grammatical derivations. Strangert and Wretling (2001) have provided durational measurements of these patterns, based on the “elderly men” data of the SweDia corpus, which was also a part of the material reported here. Strangert and Wretling (2001) found V:C: patterns within the Y-class, I-class and E-class words. This agrees with the reports of Ivars (1988) and Freudenthal (1889).

Figure 10.6 shows the C durations over V durations for all twelve speakers. With the exception of one female speaker from the younger generation, all speakers had produced the target words in question. The four figures show the four speaker groups separately, and the three speakers in every scatter plot are marked by different symbols. In spite of considerable variations across the twelve speakers the figures show clearly that the older as well as the younger generations use regularly the V:C: structures for the neuter of the adjectives *tät* (with the exception of one speaker) and *vit*, and for the perfect participle form of the verb *byta*.

Thus, the V:C: structure still seems to be a regular feature of the Vörå dialect, even within the younger generation. Strangert and Wretling (2001) reported that one speaker of the “elderly men” group produced a VC: structure for the word *tätt* (instead of V:C:). This is shown in Figure 10.6 (a). One of the symbols for the expected V:C: structures falls within the group of VC: patterns. For the other two words with expected V:C: structure, the values of this speaker are in congruence with those of the remaining speakers.

10.2.6.1 Segment and Sequence Durations of V:C:, VC: and V:C:

The three observed patterns in Vörå bring up the question how the sequence durations and the segments within the sequences relate to each other in durational terms. As a basis for a direct comparison, the median durations per speaker for the three sequences and the segments contained in them was used. This reduced the sample size to 12 (in one case 11) per sequence type, but provided the possibility to test the differences with repeated measurements tests. Figure 10.7 shows the three sequence durations and the durations of V: and C: in V:C:, V:C and VC:. The three figures clearly suggest that V:C: sequences are longer than V:C and VC: durations, and that V: and C: in V:C: are clearly shorter than their counterparts in V:C and VC:. This was confirmed by repeated measurements Wilcoxon tests. While the difference between V:C and VC: sequence duration was far from significant ($V = 44$, $p = 0.73$) V:C: was significantly longer than VC: ($V=0$, $p<0.001$) and V:C ($V=0$, $p<0.001$). V: in V:C: was significantly shorter than V: in V:C ($V=0$, $p<0.001$), and C: in V:C: was significantly shorter than C: in VC: ($V=0$, $p<0.001$).



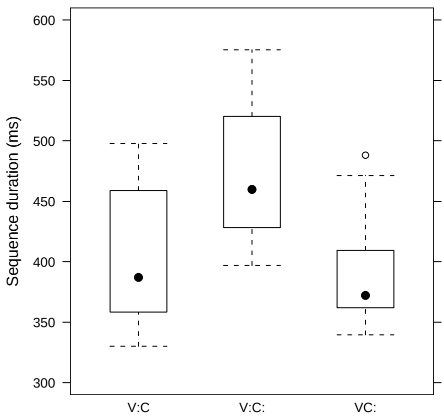
(a) Elderly men

(b) Elderly women

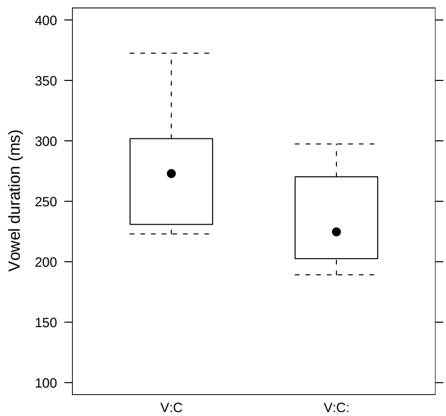
(c) Young men

(d) Young women

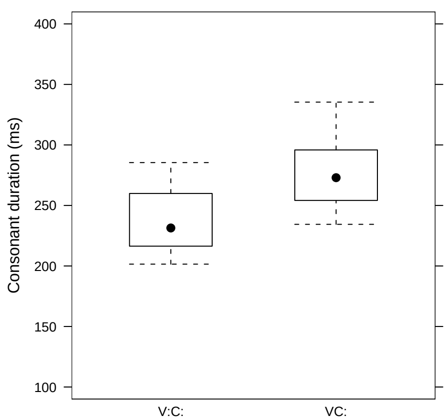
Figure 10.6: Consonant duration as a function of vowel duration for the four speaker groups of Vörå. Empty symbols indicate VC: forms (*tack* and *lått*), filled symbols in grey indicate V:C: forms (*tätt*, *vitt* and *bytt*) and filled symbols in black indicate V:C forms (*tät*, *vit* and *byt*). Different symbols distinguish between speakers.



(a) Sequence durations



(b) V: durations in V:C: and V:C



(c) C: durations in V:C: and VC:

Figure 10.7: Sequence durations and V: and C: durations in Vörå.

10.2.7 Discussion of the Durational Results

The results from the four selected recording locations confirmed in several aspects earlier observations. If all durational aspects are taken together, the following picture emerges: Vörå, Närpes and Indal show similar V:C durations, but the composition of these sequences is different: V: in Vörå and Närpes are much longer than those in Indal, and C are much shorter. V:C composition in Torp is similar to Indal, but V:C was somewhat shorter, which seems to be a consequence of shorter V: and C duration (cf. Table 10.2). VC: sequences showed shorter durations in Indal and Torp than in Närpes and Vörå. Thus, Indal shared V:C duration with the Finland-Swedish dialects, but not VC: duration. The shorter VC: sequence is a consequence of longer V and C: durations for Närpes versus Indal, but a consequence of mainly C: duration for Vörå versus Indal.

The different segment durations have also consequences for the relations between the segments. C:/C ratios in Närpes and Vörå are much higher than those in Indal and Torp. In accordance with the results from the cluster analysis, it can be observed that C:/C and V:/V ratios are of a similar order of magnitude in Närpes and Vörå, while V:/V ratios are much higher than C:/C ratios in Indal and Torp.

The V:C: structures in Vörå showed a very consistent pattern. With the exception of one speaker from the elderly men group, who realised the neuter form of *tät* as a VC: sequence, all speakers showed the V:C: structures in the expected environments. The durational composition of the V:C: sequence showed, on the one hand, that V:C: sequences are longer than both V:C and VC: sequences. On the other hand, V: and C: in V:C: are not as long as the respective long segments in V:C and VC:. The fact that long V: before C: (or CC) are not as long as before C has been observed for several languages (see Árnason, 1980; Lehtonen, 1970) and can also be found in Nyström's (2000) Älvdalen data.

10.2.8 Vocalic Qualities

In Chapter 9 the vowel qualities in the A-class were investigated and the acoustic distance between the long and the short vowel was compared to the durational ratios of the vowels and consonants. For V:/V and AD, the correlation was negative, but weak and not significant. For C:/C and AD, a negative correlation was found, which, however, was mainly an effect of the values of the Finland-Swedish region. Thus, the data of the A-class did not provide strong evidence for a trading relationship between durational and acoustic differences.

For the extended data set, the acoustic distance and the relationship between segment ratios and acoustic distance was investigated for all vowel pairs.

The acoustic distance measure was calculated as described before (see Section 7.3.4.2). The visual inspection of the formant values showed sometimes artefacts in the measurements, especially for the high front vowels of the female speakers. In those cases, the formant values were corrected by hand. The Figures 10.8 and 10.9 show formant charts for all four places, separately for men and women and the two generations. The labels indicate the respective vocalic quality, each label represents the median for the respective speaker group. The values of the older generation are always given in bold face. In the case of Vörå, the formant values of the vowels in the overlong sequences (V:C:) are given in italics.

The vowel charts indicate that the long and short high front vowels (I-class and Y-class) show similar qualities in all four locations. The A-class shows the same effect as observed earlier. The qualities of the long and short vowels in Närpes and Vörå lie clearly closer together than those in Indal and Torp. For the O-class, all dialects show a quality difference, which corresponds with the usual description for Standard Swedish. The short o-qualities are usually lower ([ɔ]) than the long qualities ([o]). The E-qualities show a difference between Närpes on the one hand and Torp and Indal on the other hand. In the latter locations, the long and short qualities in the E-class lie rather close together, while Närpes shows a clear quality difference. The long qualities in the E-class are clearly higher than the short ones.

The formants were only measured at the midpoints of the vowels. Potential formant changes during the vowel were therefore not derivable from the measurements. A listening test revealed clear diphthongisations for some vowels in Närpes. The quality of the vowel in *tät* was, in broad transcription, normally [eːə], the quality of the vowel in *lot* was normally [oːɐ], and the quality of the vowel in *byt* was normally [øy]. Besides, the vowel quality in *vit* was often [ej] or even [eː], the onset consonant in this word was realised as [kv]. In Vörå, a diphthongisation in the word *lot* occurred, similar to Närpes, but often less pronounced ([oːᵉ]).

The AD values (speaker medians) for the four locations are shown in Figure 10.10. Generally, the O-class shows relatively high AD values in all recording locations. The A-class shows, as described before, high AD values in Indal and Torp, and low values in Vörå and Närpes. The E-class, I-class and Y-class words show generally low values in Torp and Indal. In accordance with the formant charts, Närpes shows much higher AD differences than the two mainland Swedish places. The values for the E-class, I-class and Y-class are low in Vörå, but these measurements do not compare V: with V but V: in V:C with V: in V:C:. Therefore, these values will not be taken into account for the following comparisons.

Table 10.3 shows the correlation coefficients (Spearman's ρ)[1] between the AD-values and the V:/V and C:/C ratios. The first row shows the ρ values for all word classes together. The Figures 10.11 and 10.12 show the corresponding scatter plots, separately for each word-class. For the AD to V:/V values, overall correlation is not significant. The A-class showed a significant negative correlation, in accordance with the tendency observed in Section 9.3. The O-class showed also a significant correlation, but the correlation was positive in this case.

For the AD to C:/C values, overall correlation was weak and not significant. As in Section 9.3, there was a significant negative correlation for the A-class. The E-class showed a significant positive correlation, the Y-class showed a significant negative correlation and the I-class showed a negative correlation that was close to significant. The correlation for the O-class was low and not significant.

10.2.8.1 Discussion of the Vocalic Qualities

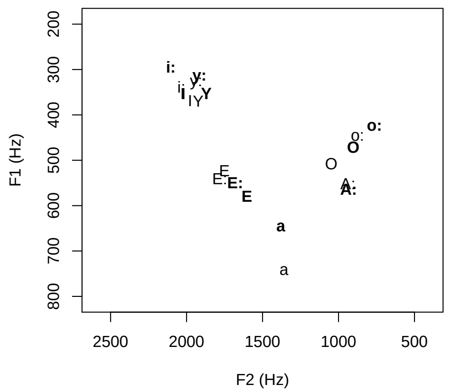
Overall, the results of the extended vowel set presented a heterogeneous picture with regard to the connection of durational and spectral distinction. No clear indication was found for a trading relationship between spectral and durational distinctions in the vowel. It has to be taken into account, however, that vowels in stressed position in isolated utterances with a probably rather slow speech rate are maybe not the most appropriate material for the observation of such effects. On the other hand, the data of Elert (1964) and Hadding-Koch and Abramson (1964) did show such an effect under quite similar conditions for the /ʉ/ versus /ɵ/ contrast. Hadding-Koch and Abramson (1964) suggested an effect of sound change for the /ʉ/ versus

[1] Spearman's ρ was used as the distributions of the AD values deviated significantly from a normal distribution, according to Shapiro-Wilk tests.

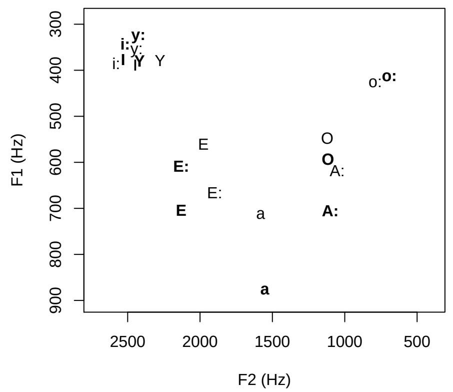
/ɵ/ contrast, and a similar idea was formulated by Delattre (1962) regarding an English tense-lax contrast. It is thus conceivable that one has to distinguish in these cases between long-term and short-term effects.

Table 10.3: Correlation coefficients (Spearman's ρ) for AD-values and segment ratios (* $p<.05$, ** $p<.01$, *** $p<.001$).

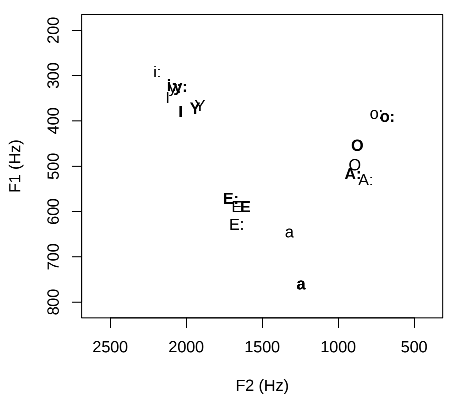
Word-class	ρ for AD–V:/V	ρ for AD–C:/C
Overall	−0.10	−0.09
A-class	−0.39**	−0.68***
E-class	−0.26	0.47**
O-class	0.36*	−0.15
I-class	0.05	−0.33
Y-class	0.09	−0.50**



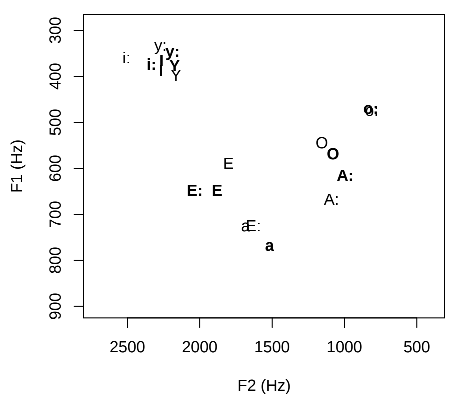
(a) Torp — Male speakers



(b) Torp — female speakers

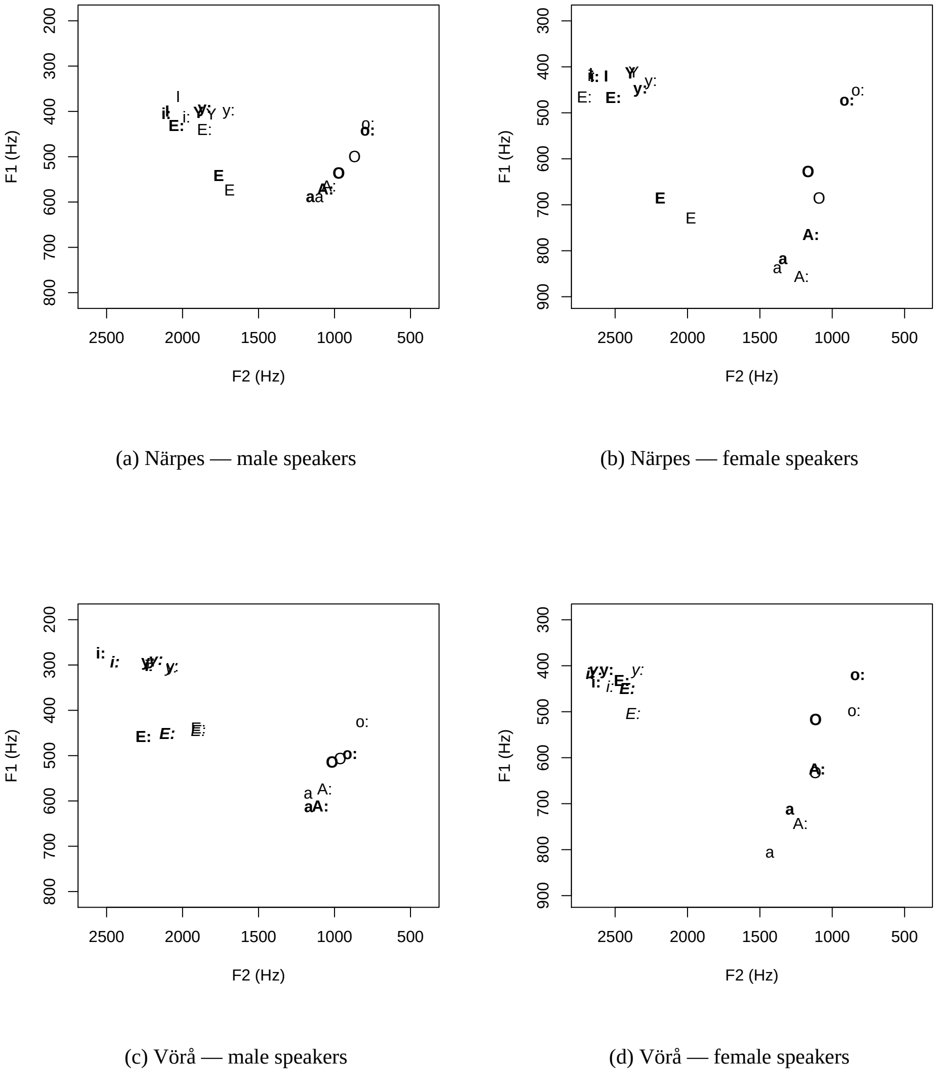


(c) Indal — male speakers



(d) Indal — female speakers

Figure 10.8: Formant charts for Torp and Indal, separately for men and women and the two generations. Labels indicate vocalic quality. Values of the older generation are given in bold face.



(a) Närpes — male speakers

(b) Närpes — female speakers

(c) Vörå — male speakers

(d) Vörå — female speakers

Figure 10.9: Formant charts for Närpes and Vörå, separately for men and women and the two generations. Labels indicate vocalic quality. Values of the older generation are given in bold face. For Vörå, formant values in overlong sequences (V:C:) are given italics.

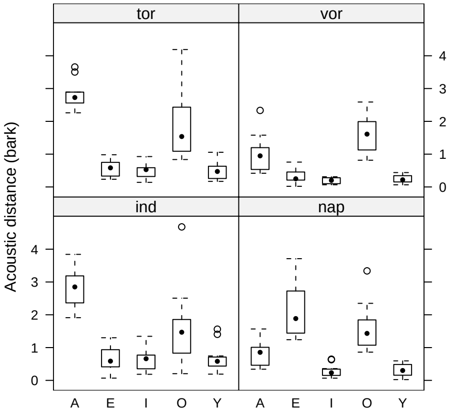


Figure 10.10: Acoustic distance values by word class in four recording locations.

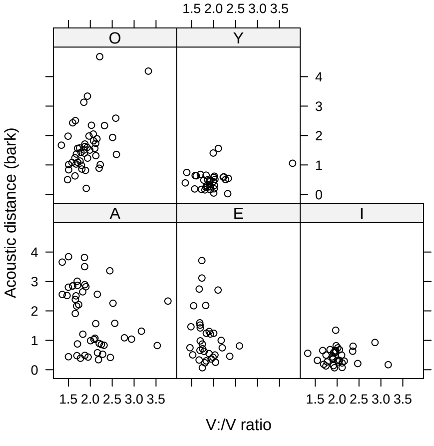


Figure 10.11: Acoustic distance over vowel ratio for five word classes.

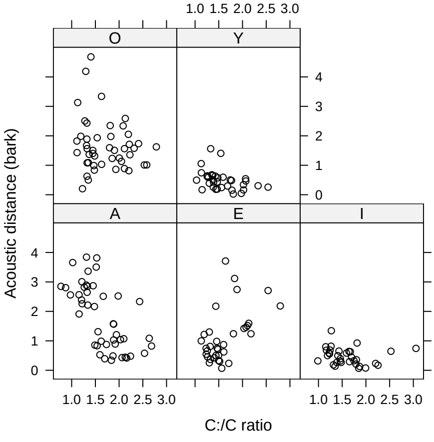


Figure 10.12: Acoustic distance over consonant ratio for five word classes.

Chapter 11

Variation, Typology and Diachrony — An Interpretation of the Data

11.1 Introduction

The data in the present study, while covering a wide range of the Swedish dialects, was rather restricted in terms of the phonetic material investigated. The material under study consisted of only mono-syllabic words, and the broadest study was performed with only one word pair. The studies in Chapter 10 brought a considerable expansion of the results in terms of vowel quality. However, regarding the consonants, only voiceless velar and alveo-dental plosives were studied, leaving out all other places of articulation. The role of voicing in consonants with respect to quantity could also not be considered. This makes it necessary to formulate typological generalisations with some caution.

11.2 A Tentative Typology

The cluster-analysis suggested a tripartitioning of the data with a clear geographical distribution. This geographical distribution showed a good agreement with the geographical distribution of the independently derived phonological types. In other words, the cluster analysis seemed to replicate the geographical distribution of dialects with 4-way, 3-way and 2-way distinctions. Any evidence for a "new" quantity type in a certain region did not emerge.

Table 11.1 shows a summary of the phonetic correlates for the three types. The durational differences between the types lie mainly in the consonants. Finland-Swedish and Northern Swedish are separated by the duration of the short consonant (which is longer in Northern Swedish), and Northern Swedish and Southern Swedish are separated by the duration of the long consonant, and partly also the short consonant (both segments tend to be shorter in the South).

Two other phonetic correlates — preaspiration and vocalic quality differences — coincided only partly with the durational characteristics. The Northern type showed higher preaspiration frequencies than the other two types. The general distribution of preaspiration complies with the results of Helgason (1999). Preaspiration is a wide-spread phenomenon as a non-distinctive feature. Only in a few rare cases, preaspiration becomes normative in Helgason's sense, that is it becomes a socio-phonetic marker or a phonological feature. For the Swedish mainland, normative preaspiration has been described for three areas, and these three areas also showed deviant values in our material. Only a few other locations had values that came close to these well-known locations. In typological terms, the role of preaspiration is hard to estimate on the basis of the present material. The observed duration and frequency values do not indicate that preaspiration should be included as a regular feature in the description of the investigated dialects, with the exception of some already well-known ones, and a few dialects that await further investigation. Therefore, preaspiration will not be considered any further here.

The vocalic quality distinctions showed no obvious connection to the durational distinctions. The coincidence of low vocalic distinctions and higher durational distinctions that turned out as a tendency with the A-class words, was not confirmed with the other word classes. The results of the present material suggest that duration and spectral difference are not closely connected. Vowel pairs with large spectral differences showed the same clear durational separation as those with low spectral differences. At first sight, this speaks against the trading relationship hypothesis, as for example formulated by Hadding-Koch and Abramson (1964). However, our observations do not exclude the lessening of durational constraints in a long-term perspective (see below). However, with reference to the typological observations made in section 6.2.1 the observations made in the present study confirm again that long-short vowel pairs sometimes show low spectral separation and sometimes high spectral separation within the same vowel system.

Table 11.1: The three main types and their qualities.

Region	Finland-Swedish type	Northern Swedish type	Southern Swedish type
V: duration	long – very long	long	long
V duration	short or medium	short	short
C: duration	long	long	short
C duration	very short	long	medium
V:/V	high – very high	high	high
C:/C	high	low	very low
AD	high and low	high and low	high?
Preaspiration	less frequent	frequent	less frequent

11.3 Phonetic Variation and Phonological System

The three quantity types described for Swedish include a type with vocalic and consonantal quantity distinctions, a type with consonantal quantity in restricted position (only after short vowels) and a type where long vowels in closed syllables are always followed by short consonants, while short vowels are followed by a long consonant. The occurrence of these three types has been described in Chapter 5. If the geographical distribution of these phonological types is compared to the geographical distribution obtained by the cluster analysis, clear congruences are observed. In general, 4-way distinctions have been described for areas in Finland (and Dalarna, see below), 3-way distinctions for many Northern dialects and 2-way distinctions for Southern Sweden. This tripartitioning is reflected in the main partitioning of the cluster analysis which resulted in a clear geographical tripartitioning: Northern Sweden, Southern Sweden and Finland Sweden.

The main durational differences between the dialects of these three regions are found on the consonantal side. Finland-Swedish dialects have long C: and short C, resulting in high C:/C ratios. For the A-class words, the ratios were always above 1.7. Northern Sweden has long C: as well, but also longer C, resulting in lower C:/C ratios. The ratios for the A-class words never exceeded 1.5 and were below 1.3 in 70% of the cases. Southern Sweden had shorter C:, and also somewhat shorter C, resulting in low C:/C ratios as well: 70% of the locations in the Southern cluster had a C:/C ratio below 1.2. Differences in vowel duration did not show a comparable geographical distribution. Furthermore, vowel ratios for the A-class words were always higher than 1.3 and in 75% of the cases higher than 1.7, which suggests that vowel duration can be used as a phonological cue.

If the differences in consonant duration are related to the geographical distribution of the phonological types, the congruences are obvious: In Finland, many dialects show 4-way distinctions. In Northern Sweden, many dialects show 3-way distinctions, and in Southern Sweden, mostly 2-way distinctions are reported. It is thus legitimate to ask whether the observed durational patterns are related to the phonological types, as the common geographical distribution would suggest.

Sixteen of the 19 recorded dialects in the Northern area (as defined by Wessén, 1969) fell into the Northern cluster. According to the literature, eight out of these 16 places show VC sequences. Three are explicitely reported not to have VC sequences. For the remaining five, the situation is unclear. Thus, it is likely that, for the Northern region as well, the durational values are connected to the phonological structure.

On the whole, if the clear geographical tripartitioning resulting from the cluster analysis — a Southern group, a Northern group and a Finnish group — is compared to the geographical distribution of the three phonological types as derived from the dialectological literature, it becomes clear that these areas largely coincide. This geographical match makes it very likely that the observed different durational patterns are actually *caused* by the different phonological systems, although it has to be taken into account that other causes might exist as well. If the causal connection is assumed, one may ask how the durational patterns may be connected to the phonological quantity structure.

11.3.1 Duration and Structure — a Constraint-based Account

In this context it is important to dicuss the structural position of the investigated sequences V:C and VC:. They do not constitute minimal pairs in the 4-way and the 3-way systems. In four way systems, V:C contrasts with VC (only vocalic duration) and V:C: (only consonantal duration). VC: contrasts with V:C: (only vocalic duration) and VC (only consonantal duration). Consequently, in dialects like the one of Vörå in Finland, V: and V, as well as C: and C are durationally well distinguished. The durational differences do not exist only to distinguish those two sequences, but also to tell them apart from the two other possible sequences (VC and V:C:).

In 3-way systems, V:C contrasts with VC (only vocalic duration), and VC: contrasts with VC (only consonantal duration). In terms of durational constraints, V:C has lost a consonantal distinction and VC: has lost a vocalic distinction. It is likely that such a loss has (long-term) consequences for the durational structure of the sounds involved. The loss of V:C: lessens the constraint on V in VC: and on C in V:C. The analysis of the data has shown that the main difference between the Finland-Swedish and the Northern group of the dialects lies in the shortness of the C. Finland-Swedish dialects have a shorter C than the Northern dialects, while C: durations are quite similar. As a consequence, C:/C ratios are much higher in Finland-Swedish dialects than in Northern Swedish dialects. For five of the seven dialects in the Finland-Swedish cluster there is clear evidence for a 4-way distinction (see section 6.2). Thus, it is justifiable to assume that the durational characteristics are connected with the phonological structure. A stable difference in vowel *and* consonant duration between V:C and VC: is in accordance with the existence of the additional VC and V:C: sequences.

In the 3-way systems, V:C: does not exist, but VC: is distinguished from VC, which could explain why the C: duration values in the North are as long as those in the Finland-Swedish region. The C in V:C, however, has lost a durational constraint, as a C: after a V: does not exist[1]. This would allow a longer C after V:, which is exactly what is observed in the Northern region in comparison to the Finland-Swedish region. This causes a major drop in the C:/C ratios. Consequently, for the difference between the V:C and the VC: sequence, the vowel duration seems to be the stronger cue (V:/V ratios are much higher than C:/C ratios in the Northern clusters).

In 2-way systems, VC does not exist. Thus, a long C: in VC: is not as important as in 3-way systems, which should allow shorter C: in VC:. Again, this difference was observed between the Northern and the Southern clusters: C: in the Southern clusters are clearly shorter than C: in the Northern clusters. At the same time, however, C in V:C show a tendency for shorter values in the Southern clusters as well. As a consequence, the amount of “complementarity” is only slightly higher in the North, and even the values in the South can be seen as complementary quantity: Long vowels are still followed by somewhat shorter consonants than short vowels.

The outlined constraint-based interpretation of the durational values provides no explanation for the shorter C values in the South. It is important to note, however, that a range of Southern dialects does show shorter C: without showing shorter C. Especially cluster S2 has C values well within the range of the Northern clusters, while C: duration values are clearly shorter. Some very short C in the South are clearly based on voicing of the C(c.f. cluster S3b).

[1]This idea implicitly assumes, that the phonotactics of a variety are used by listeners for estimations of the likelihood of upcoming sounds. If a listener hears a long vowel in a three way system, she knows that only a short consonant can follow. In a 4-way system, both C: and C can follow

For the other Southern dialects, occasional voicing is not enough as an explanation for the lower C duration values. It is conceivable, however, that the Southern dialects show a tendency towards generally “laxer” plosives, which should be the subject of further investigations under speech rate-controlled conditions. It is also possible to interpret the shorter C in the South as the maintenance of a certain amount of complementarity.

Another interesting area that has not yet produced unequivocal results is the perception of quantity distinctions with changes in speech rate and the normalisation procedures involved. Although it is evident that quantity perception must be “somehow relative” and not bound to specific absolute durations, there exists to our knowledge no general theory of how quantity-bearing segments are compared to their neighbouring environment, and whether these mechanisms are universal or language-specific. The answers to those questions will also help to tackle the problem of “quantity change”.

11.3.2 A comparison between Finland-Swedish and Dalarna

The recorded dialects in the Dalarna region did not show any separation from surrounding areas, although the dialectological literature suggests 4-way systems in this region. This is a possible objection to our argument that the observed durational values are reflections of the phonological system. However, the quantity material for Dalarna in the SweDia data base was rather scarce and no recordings were available for the Western Siljan area, where most of the 4-way systems are reported. It is thus possible that our data might just have missed the dialects with 4-way systems. The durational values reported by Nyström (2000) provide an opportunity to compare the durational structure of the Finland-Swedish dialects with those of the Dalarna region. As Strangert and Wretling (2001) have already pointed out, there are clear similarities between the durational values reported by Nyström and those found in the data from Vörå. Especially the C:/C ratios for VC: versus V:C are remarkable here. These ratios are high in Dalarna, Närpes and Vörå (around 2.0 or higher) and low (often around 1.3) for the rest of the investigated dialects and also Standard Swedish. This strengthens our assumption that the main difference between the three types lies in the consonantal part.

11.4 Quantity and Sound Change: An Interpretation of the Data

While the data of the present study showed an overall good match between phonological types and phonetic outcome, which also makes sense under the premises of a “phonological constraint” interpretation (see Section 11.3.1), there are some interesting mismatches between phonetic substance and phonological form for some recording locations. The dialectological and historical literature suggests that 4-way systems developed into 3-way systems which developed into 2-way systems. The different systems observed today are usually seen as reflections of a common Germanic proto-language, where the 4-way systems reflect the most conservative existing forms, and the 2-way systems the most recent development.

Ivars (1988) concludes that for the Finland-Swedish dialects, the loss of the V:C: sequence spreads as an innovation from the South-Western area. In Österbotten, the innovation has reached Närpes, but not Vörå. If Närpes has adopted the 3-way system which is found in many regions of Northern Sweden, it is worth asking why Närpes shows durational characteristics much closer to Vörå than to the dialects in Northern Sweden. In our view, a concept suggested by Labov (1994) could shed some light on this issue. In sound change research, it has been a matter of dispute whether sound change proceeds “sound by sound” or “word by word”. The so-called “Neogrammarians” proposed that sound changes proceed without exception and mechanically whenever their phonetic condition is met (e.g. Osthoff and Brugmann, 1878).

Another position was often argued for by dialectologists, expressed in the slogan “Every word has its own history” (c.f. Labov, 1994, p. 472). This position received major support through the work of Wang and colleagues (e.g. Chen and Wang, 1975; Wang, 1977, cited in Labov, 1994). They investigated an extensive database of Modern Chinese dialects and questioned the Neogrammarian principle that sound change is phonetically gradual, but lexically abrupt. Their hypothesis is exactly the opposite: Sound change is phonetically abrupt, but lexically gradual (see Labov, 1994, p. 424). Sound change operates, according to these authors, not on sounds, but on words. This concept of sound change is called “lexical diffusion”.

Labov (1994) presents an extensive discussion of the two different positions. His analysis shows that there is strong evidence for both positions. Labov concludes that there seem to exist two different types of sound change. For example, he shows that a sound change observed in the Philadelphian English vernacular in the US, namely the tensing and raising of short /æ/, obviously consists of two different processes. He develops eight criteria that distinguish these two types of change (p. 527). Table 11.2 shows these eight criteria.

Table 11.2: Two types of sound change (following Labov, 1994, p. 527)

	Raising of tense /æ/	*Tensing of lax /æ/*
Lexical diffusion	no	yes
Discrete	no	yes
Phonetic differentiation	single feature	many features
The phonetic conditioning is	precise	approximate
Grammatical conditioning	no	yes
Social affect	yes	no
Categorically perceived	no	yes
Learnable	easily	with great difficulty

The tensing of /æ/ shows lexical diffusion. On the one hand, this can be seen in phonetically irregular distributions. In Philadelphia, all vowels before voiced stops are lax, with the exception of three words: *mad*, *bad* and *glad*. Further evidence comes from Philadelphian regions where the change has not progressed as far as in the core dialect. Here, it was observed that the sound change proceeded by lexical diffusion. Words with identical phonetic environments showed different rates of tensing, suggesting that the change is connected with individual lexical items.

Furthermore, tensing is discrete and based on several phonetic features. The differences between the lax and tense realisations of /æ/ (before the raising) lie in duration, peripherality, height and the existence of an inglide. Thus there is no continuous development from one sound to the other. Rather, a speaker is either producing a lax sound or a tense sound where all the involved phonetic features are changed.

Labov proceeds by comparing the results from the Philadelphian dialect to the changes occurring during the English Great Vowel Shift. He comes to the general conclusion:

> *"... vowel lengthening and shortening in English are implemented by lexical diffusion, while raising, lowering, backing, and fronting proceed by regular sound change"*(Labov, 1994, p. 531

Labov's analysis of quantity changes might provide a possible answer to the present question. The phonetic abruptness is the more important aspect here: Phonetic abruptness does not mean that the change has always to be observed at a certain time, but that it is a kind of all-or-nothing process. A speaker that adopts the change for a certain word or form, replaces a long segment by a short segment, or vice versa. In the situation described here, V: in V:C: is replaced by V. If it is assumed that a speaker has certain durational targets when producing a sound, then it can be expected that this process does not change the durational targets of the sounds involved. The V: in V:C: is replaced by V, but the durational targets of the remaining sounds stay the same. They still reflect the well separated values of a 4-way distinction. Note that such a process is also indicated by the production of one older speaker in Vörå (see Section 10.2.6): Although he produces Standard Swedish *tätt* with a VC: structure, his remaining productions lie within the durational characteristics of the dialect. This, as well, supports the "lexical diffusion" hypothesis.

After the total loss of V:C:, three possible sequences in the dialect remain: V:C, VC: and VC, which still show the same durational composition as before. However, as V:C: has been lost, a consonant length distinction after long vowels does not exist any more. The length of the short consonant is the most striking difference between the dialects of the Finland-Swedish type and those of the Northern type. The consonant is much shorter in the Finland-Swedish type than in the Northern type, guaranteeing a stable durational distinction between C in V:C and C: in VC: in the Finland-Swedish type.

With the loss of V:C:, the shortness of the C in V:C is not as constrained as before, therefore giving room for a lengthening of this segment. In this case, the change might be phonetically gradual, as the duration of the segment is no longer determined by a phonological length distinction. We would thus assume a process that proceeds in two different phases. In the first phase, V:C: structures are replaced with VC: structures by lexical diffusion. The durational targets in a dialect are not altered. After the complete loss of V:C:, a second process begins. Gradually, the durational targets in the dialect adapt to the new situation. Thus, this hypothesis would predict that the durational realisations in a dialect are not only dependent on the phonological system, but also on the period of time that has passed since the loss of V:C:.

11.5 Phonological Theory and Phonetic Outcome

In Riad's diachronic system, the prosodic structure changes mainly in terms of the vowel. Vowels loose their underlying length distinction and vocalic duration is derived by rule. However, this system suggests no difference between rule-based and underlying vocalic quantity at the phonetic level. Short vowels are still mono-moraic, and long vowels bi-moraic. This seems to be adequate for the vowel distinctions observed in the present material. Although the 4-way distinctions show somewhat higher V:/V ratios, the general long-short distinction is maintained in all dialects. No dialect showed clear signs of a loss of a durational distinction on the part of the vowels. The situation is different for consonant quantity. Here, Riad's theory assumes no crucial changes before and after the quantity shift. Long consonants are underlyingly moraic, short consonants are not. The former surface as long, the latter as short. The only exception from this rule occurs shortly before the quantity shift, where the distinction is neutralised in mono-syllabic words.

The data investigated in the present study suggest, however, that there is, in terms of consonant distinctions, a crucial difference between 4-way distinctions on the one hand, and 3-way and 2-way distinctions, on the other. While 4-way distinctions show long consonants that are about twice as long as short consonants, the durational distinction is dramatically lower for 3-way and 2-way distinctions. This, in our view, is not predicted by the phonological account. There is now abundant evidence from several phonetic investigations, including the present one, that the vowel carries the main perceptual load of the opposition in 2-way systems. And it is argued that the present investigation extends this conclusion to 3-way systems, as far as the V:C versus VC: contrast is concerned. If this is restated as underlying vowel quantity, then the (small but usually perceptible) differences in consonant duration really "remain a mystery", as Riad (1992) has put it (see p. 295).

Why is there complementary quantity after all? We have argued earlier, that the loss of V:C: forms in Närpes showed signs of lexical diffusion. Thus, V: is replaced by V in V:C: sequences without an immediate effect on the durational targets for the other segments. If this is correct, the system ends up with a kind of "durational overspecification". The segment durations are those of a 4-way system, but the dialect has only a 3-way system. The next step would be, as outlined above, an adaptation to the new situation which proceeds gradually over maybe several generations. This process can be assigned to the transition from 3-way to 2-way distinctions. Also in this case, it is conceivable that lexical diffusion is at work. VC is replaced by V:C or VC:, but the original or acquired VC: keeps the durational targets that were necessary for the 3-way distinction. Then, again, a continuous process begins, where the durational constraints on C and C: are gradually lessened.

This assumption is admittedly highly speculative. There are, however, other sources that report the maintenance of "lost" distinctions. Labov (1994) reports spurious maintenance of vocalic quality differences in communities where mergers have occurred, and Svärd (2001) reports the spurious maintenance of the Swedish accent distinction in an area that has apparently abandoned this distinction in its phonological system. In the present case, the "inheritance" of a non-distinctive difference might be even less problematic, as the differences are clearly perceptible. In this respect the durational differences resemble allophonic variations, which have to be learned as well by a new generation, as far as they are not purely automatic effects of the phonetic realisation.

Naturally, there might be independent reasons for the maintenance of complementary quantity as well. Lengthening of segments is a well-known side-effect of focus. If one segment in a syllable resists lengthening (in Standard Swedish obviously the short vowel), it seems quite straight-forward that the following segment is lengthened. However, this explanation might be too simple as there are languages, for example Standard German, with quite similar stress realisation and quite similar constraints on vocalic duration, where this effect has not been observed. This speaks against the assumption that complementary quantity is just an automatic effect of stress as soon as certain basic conditions are met.

Some scholars (Bannert, 1976; Pind, 1986) view complementary quantity as a special case of quantity distinction where the quantity distinction is encoded as a ratio between the two segments in question and decoded by the listener via this ratio. Section 2.6.2 has argued against this assumption on the following grounds: First, empirical investigations of complementary quantity suggest an influence of consonant duration on the perception of vocalic quantity, but this influence is lower than the ratio calculation would predict. Second, the lengthening processes affecting vowels and consonants due to differences in stress or speech rate are non-linear, especially in the case of VC:. As a consequence, the V/C: ratio is variable. If the V/C: ratio would be the decisive cue, one would expect that the speaker keeps this ratio stable. Third, only the C: after V is obligatory in Swedish phonotactics. Consonants after long vowels are optional and CV: words are not uncommon in Swedish. If they occur pre-pausally, the quantity cannot be derived by the listener by a ratio calculation with the following consonant.

In accordance with Traunmüller and Krull (2003) it is assumed that the perception of the quantity of a certain segment is weighted against the sound durations of neighbouring segments. How wide the window for neighbouring segments is might be language-dependent, but it is in all likelihood wider than just the following segment. In our view, this model would also allow for certain durational thresholds, that is durational values that always result in "long" or "short" judgments, independent of the surrounding segment durations.

Thus, we argue here that the perception of complementary quantity, as realised in Standard Swedish, is not principally different from the perception of, for example, vocalic quantity, where V:C contrasts with VC. We assume that the additional change in consonant duration observed with complementary quantity could aid quantity perception but it is not as important as a "ratio account" of complementary quantity would suggest.

Consequently, there are no obvious articulatory or perceptual reasons for complementary quantity to occur. Therefore, we would like to adopt the "historical hypothesis" for the present study: it is assumed that complementary quantity exists because of the recent loss of a consonantal quantity distinction that left its traces in the durational realisation of consonants. This, however, does not predict that complementary quantity might cease to exist within a few generations. The number of variables in the socio-phonetic contexts are too many to allow certain predictions about the future of phonological systems.

11.6 Future Research

The present study outlined a quantity typology of Swedish dialects on the basis of a data driven analysis of segmental durations. Three main types were distinguished: 4-way-, 3-way and 2-way systems, a categorisation with close correspondence to categorisations derived from the dialectological literature. The study included data from a wide range of dialects. The linguistic-phonetic context, however, was restricted to monosyllabic words. Therefore, future investigations addressing these quantity types should use a more varied material including polysyllabic words, additional vocalic qualities and consonantal differences in terms of manner and place. It is quite conceivable, for example, that the "short short consonants" in the Finland-Swedish region are only observable with intrinsically rather long consonants like voiceless plosives.

Furthermore, it would be of great interest to investigate the influence of stress and speech rate changes on the realisation of the quantity structures in the three types, and to study the perceptual effects of those changes. On the one hand, this might help to understand the perceptual processing of durational distinctions in general, and on the other hand, acoustic and articulatory investigations of the three types should help to verify or reject the syllable-cut hypothesis. As the non-existence of stressed light syllables (i.e. short vowels in open syllables) is a characteristic of syllable-cut languages, the transition from a 3-way to a 2-way system should imply a typological change. In this case, one would expect smooth cut after short vowels in 3-way and 4-way systems, but abrupt cut after short vowels in 2-way systems. This leads to concrete articulatory hypotheses which could be tested experimentally.

A follow-up study of the quantity structures in Vörå and Närpes would be of interest as well. If our “historical model” is correct, one would predict changes in the durational realisation of quantity in the Närpes dialect, which should become more similar to the realisation in mainland Sweden. Analysis of data from a younger generation might shed light on this issue.

Overall, the results of the present study can be used as a starting point for more detailed investigations of various aspects, be it the realisation of quantity, the role of preaspiration or the connection of durational and spectral cues, as well as other issues in relation to quantity that have not been possible to cover in the present thesis.

Appendix A

Recording Locations

Table A.1: List of the SweDia recording locations investigated in the present study.

No	Landscape	Short form	Full name
1	Småland	ank	Ankarsrum
2	Ångermanland	anu	Anundsjö
3	Gästrikland	ara	Årsunda
4	Jämtland	are	Åre
5	Lappland	arj	Arjeplog
6	Halland	ars	Årstad
7	Östergötland	asb	Asby
8	Jämtland	asp	Aspås
9	Skåne	bar	Bara
10	Dalsland	ben	Bengtsfors
11	Jämtland	ber	Berg
12	Ångermanland	bju	Bjurholm
13	Öland	bod	Böda
14	Nyland	bor	Borgå
15	Åland	bra	Brändö
16	Öland	bre	Bredsätra
17	Skåne	bro	Broby
18	Småland	brr	Burseryd
19	Västerbotten	bur	Burträsk
20	Värmland	dal	Dalby
21	Åboland	dra	Dragsfjärd
22	Gotland	fao	Fårö
23	Hälsingland	far	Färila
24	Värmland	fin	Södra Finnskoga
25	Ångermanland	fja	Fjällsjö
26	Västergötland	flo	Floby

No	Landscape	Short form	Full name
27	Gotland	fol	Fole
28	Dalsland	fra	Frändefors
29	Halland	fri	Frillesås
30	Värmland	gas	Gåsborn
31	Uppland	grs	Gräsö
32	Småland	ham	Hamneda
33	Värmland	hao	Hammarö
34	Västmanland	har	Haraker
35	Åboland	hou	Houtskär
36	Dalarna	hus	Husby
37	Medelpad	ind	Indal
38	Västmanland	jab	Järnboås
39	Blekinge	jam	Jämshög
40	Småland	jar	Järsnäs
41	Bohuslän	kaa	Kärna
42	Uppland	kar	Kårsta
43	Västergötland	kor	Korsberga
44	Nyland	kyr	Kyrkslätt
45	Södermanland	lan	Länna
46	Dalarna	lek	Leksand
47	Härjedalen	lil	Lillhärdal
48	Skåne	lod	Löderup
49	Dalarna	mal	Malung
50	Österbotten	nap	Närpes
51	Norrbotten	nka	Kalix
52	Norrbotten	nlu	Luleå
53	Uppland	nor	Nora
54	Skåne	nro	Norra Rörum
55	Västerbotten	nys	Nysätra
56	Gästrikland	ock	Ockelbo
57	Bohuslän	oru	Orust
58	Skåne	oss	Össjö
59	Västergötland	ost	Östad
60	Hälsingland	ova	Ovanåker
61	Västergötland	oxa	Öxabäck
62	Norrbotten	pit	Piteå
63	Jämtland	rag	Ragunda
64	Östergötland	rim	Rimforsa
65	Åland	sal	Saltvik
66	Östergötland	san	S:t Anna
67	Dalarna	sar	Särna
68	Öland	seg	Segerstad
69	Bohuslän	ske	Skee
70	Hälsingland	sko	Skog
71	Nyland	sna	Snappertuna

No	Landscape	Short form	Full name
72	Södermanland	soa	Sorunda
73	Gotland	spr	Sproge
74	Närke	sta	Stora Mellösa
75	Småland	ste	Stenberga
76	Östergötland	tja	Tjällmo
77	Småland	toa	Torsås
78	Blekinge	toh	Torhamn
79	Västergötland	too	Torsö
80	Medelpad	tor	Torp
81	Småland	vac	Väckelsång
82	Halland	vax	Våxtorp
83	Härjedalen	vem	Vemdalen
84	Närke	vib	Viby
85	Lappland	vim	Vilhemina
86	Österbotten	vor	Vörå

Appendix B

Extended Material in the North

Table B.1: The extended material of the "Umeå corpus". *reg* refers to the regular pattern; *p* indicates preaspiration; *form* indicates the use of a different word form.

Landscape	Recording loc.	A-class	O-class	E-class	U-class	Y-class	I-class
Österbotten	Närpes (nap)	reg	reg	reg	voiced	reg	reg
	Vörå (vor)	reg	reg	V:C:	voiced	V:C:	V:C:
Norrbotten	Kalix (nka)	reg	reg	-	-	-	-
	Luleå (nlu)	reg	reg	VC:	voiced	reg	reg
	Piteå (pit)	reg	reg	-	reg	reg	reg
Lappland	Arjeplog (arj)	p	p	p	p	p	p
	Vilhemina (vim)	reg	reg	VC:	voiced	reg	reg
Västerbotten	Burträsk (bur)	reg	reg	-	-	-	-
	Nysätra (nys)	reg	reg	-	-	-	-
Jämtland	Åre (are)	reg	reg	VC:	voiced	VC:	reg
	Aspås (asp)	reg	form	VC:	voiced	reg	reg
	Berg (ber)	reg	form	V:C	voiced	VC:	reg
	Ragunda (rag)	reg	reg	VC:	voiced	VC:	reg
Ångermanland	Anundsjö (anu)	reg	form	reg	reg	reg	reg
	Bjurholm (bju)	reg	reg	-	-	-	-
	Fjällsjö (fja)	reg	reg	VC:	reg	reg	reg
Medelpad	Indal (ind)	reg	reg	reg	reg	reg	reg
	Torp (tor)	reg	reg	reg	reg	reg	reg
Härjedalen	Lillhärdal (lil)	reg	reg	VC:	reg	VC:	reg
	Vemdalen (vem)	p	p	p	p	p	p
Hälsingland	Färila (far)	reg	reg	VC:	reg	reg	reg
	Ovanåker (ova)	reg	reg	reg	reg	reg	reg
	Skog (sko)	reg	reg	reg	reg	reg	reg
Gästrikland	Årsunda (ara)	reg	reg	reg	reg	reg	reg
	Ockelbo (ock)	reg	reg	reg	reg	reg	reg

Appendix C

Geographical Position of the SweDia Recording Locations

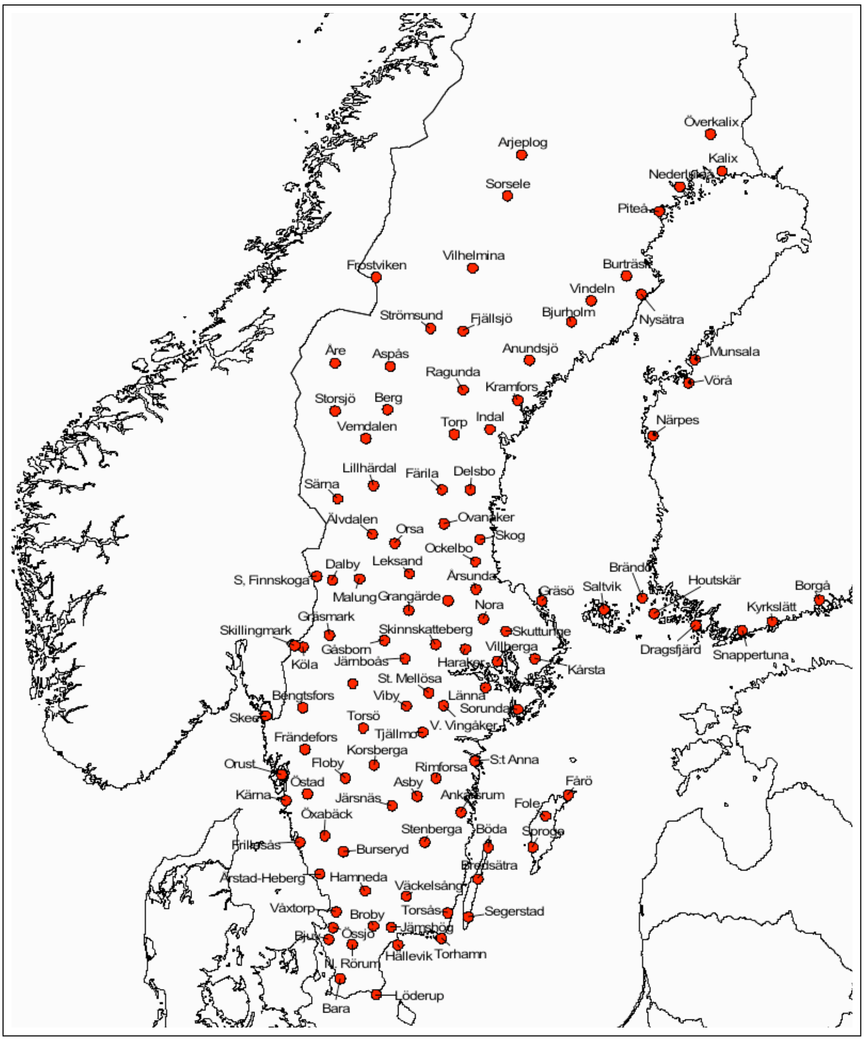


Figure C.1: Geographical Position of the SweDia Recording Locations.

Bibliography

Árnason, K. (1980). *Quantity in Historical Phonology. Icelandic and Related Cases.* Cambridge: Cambridge University Press.

Bannert, R. (1969). The Effect of Sentence Accent on Prosody. In *Proceedings of the Ninth International Congress of Phonetic Sciences*, pages 253–259. Copenhagen: Institute of Phonetics, University of Copenhagen.

Bannert, R. (1976). *Mittelbayerische Phonologie auf Akustischer und Perzeptorischer Grundlage*. Lund: Gleerup.

Bannert, R. (1979). Ordprosodi i invandrarundervisningen. In *Praktisk Lingvistik 3*. Institutionen för lingvistik, Lunds universitet.

Behne, D., Czigler, P., and Sullivan, K. (1996). Acoustic Characteristics of Perceived Quantity and Quality in Swedish Vowels. In *Speech Science and Technology 96, Adelaide*, pages 49–54.

Behne, D., Czigler, P., and Sullivan, K. (1998). Perceived Swedish Vowel Quantity: Effects of Postvocalic Consonant Duration. In *Phonum 6: Reports from the Department of Phonetics, Umeå University*, pages 91–98. Umeå University.

Bogren, P. (1921). *Torpmålets ljud- och formlära*. Stockholm: Norstedt & Söner.

Bortz, J. (1993). *Statistik für Sozialwissenschaftler*. Berlin: Springer.

Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Bruce, G. (1998). Allmän och svensk prosodi. In *Praktisk Lingvistik 16*. Institutionen för Lingvistik, Lunds universitet.

Bruce, G., Elert, C.-C., Engstrand, O., and Wretling, P. (1999). Phonetics and Phonology of the Swedish Dialects - A Project Presentation and a Database Demonstrator. In *Proceedings of the 14th ICPhS, San Francisco*, pages 321–324.

Bruce, G. and Gårding, E. (1978). A Prosodic Typology for Swedish Dialects. In Gårding, E., Bruce, G., and Bannert, R., editors, *Proceedings of Nordic Prosody I*, pages 219–228. Department of Linguistics, Lund University.

Chambers, J., Cleveland, W., Kleiner, B., and Tukey, P. (1983). *Graphical Methods for Data Analysis*. New York: Chapman & Hall.

Chen, M. and Wang, W. S.-Y. (1975). Sound Change: Actuation and Implementation. *Language*, 51:255–281.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Crothers, J. (1978). Typology and Universals of Vowel Systems. In Greenberg, J., editor, *Universals of Human Language 2: Phonology*, pages 93–152. Stanford: Stanford University Press.

Dahlstedt, K.-H. (1962). *Det svenska Vilhelminamålet. Språkgeografiska studier över ett norrländskt nybyggarmål och dess granndialekter. 2. Kvantitet, Apokope*. Uppsala: Almqvist & Wiksells Boktryckeri AB.

Décsy, G. (1988). *A Select Catalog of Language Universals*. Bloomington, Indiana: Eurolingua.

Delattre, P. (1962). Some Factors of Vowel Duration and their Cross-Linguistic Validity. *Journal of the Acoustical Society of America*, 34:1141–1144.

Dommelen, W. A. v. (1999). Auditory Accounts of Temporal Factors in the Perception of Norwegian Disyllables and Speech Analogs. *Journal of Phonetics*, 27:107–123.

Elert, C.-C. (1964). *Phonologic Studies of Quantity in Swedish*. Uppsala: Almqvist & Wiksell.

Eliasson, S. and La Pelle, N. (1973). Generativa regler för svenskans kvantitet. *Arkiv för nordisk filologi*, 88:133–148.

Engstrand, O. and Krull, D. (1994). Durational Correlates of Quantity in Swedish, Finnish and Estonian: Cross-Language Evidence for a Theory of Adaptive Dispersion. *Phonetica*, 51:80–91.

Engstrand, O. and Nordstrand, L. (1983). Acoustic Features Correlating with Tenseness, Laxness and Stress in Swedish: Preliminary Observations. In *Nordic Prosody III, Umeå Studies in the Humanities*, volume 59, pages 51–66. Umeå University.

Everitt, B. S. (1993). *Cluster Analysis*. London: Edward Arnold, 3rd edition.

Fant, G., Kruckenberg, A., and Nord, L. (1991). Durational Correlates of Stress in Swedish. *Journal of Phonetics*, 19:351–365.

Freudenthal, A. O. (1878). *Ueber den Närpesdialect*. Helsingfors: J. C. Frenkell & Sohn.

Freudenthal, A. O. (1889). Vöråmålet. In *Skrifter utgivna af Svenska Litteratursällskapet i Finland*, volume XII.

Gårding, E. (1974). *Kontrastiv prosodi*. Lund: Gleerup.

Greisbach, R. (2001). *Experimentelle Testmethodik in Phonetik und Phonologie*. Frankfurt am Main: Peter Lang.

Hadding-Koch, K. and Abramson, A. S. (1964). Duration Versus Spectrum in Swedish Vowels: Some Perceptual Experiments. *Studia Linguistica*, 18(2):94–107.

Hansson, Å. (1969). *Fonematiska studier i skånska dialekter*. Lund: Studentlitteratur.

Harling-Kranck, G. (1988). Dialektgränsen vid Skiftet. In *Åländskt språk mellan öst och väst*, pages 75–99. Ålands högskola, Mariehamn.

Haugen, E. (1976). *The Scandinavian Languages*. London: Faber and Faber.

Heldner, M. (2001). On the Non-Linear Lengthening of Focally Accented Swedish Words. In *Nordic Prosody: Proceedings of the VIIIth Conference, Trondheim 2000*, pages 103–112. Frankfurt am Main: Peter Lang.

Heldner, M. and Strangert, E. (2001). Temporal Effects of Focus in Swedish. *Journal of Phonetics*, 29:329–261.

Helgason, P. (1999). Phonetic Preconditions for the Development of Normative Preaspiration. In *Proceedings of the 14th ICPhS, San Francisco*, pages 1851–1854.

Helgason, P. (2002). *Preaspiration in the Nordic Languages - Synchronic and Diachronic Aspects.* Department of Linguistics, Stockholm University.

Hesselman, B. (1902). *Stafvelseförlängning och vokalkvalitet. Undersökningar i nordisk ljudhistoria*, volume 1. Uppsala: Uppsala Nya Tidnings AB.

Hesselman, B. (1905). *Sveamålen och de svenska dialekternas indelning*. K. W. Appelbergs Boktryckeri.

Holt, D. E. (1997). *The Role of the Listener in the Historical Phonology of Spanish and Portuguese: An Optimality-Theoretic Account*. Georgetown University, Washington D.C.

Hoole, P. and Mooshammer, C. (2002). Articulatory Analysis of the German Vowel System. In Auer, P., Gilles, P., and Spiekermann, H., editors, *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer.

Howell, D. C. (2002). *Statistical Methods for Psychology*. Pacific Grove, California: Duxbury Press.

Ivars, A.-M. (1988). Närpesdialekten på 1980-talet. In *Studier i nordisk fonologi*, volume 70.

Jakobson, R., Fant, C. G. M., and Halle, M. (1951). *Preliminaries to Speech Analysis*. Cambridge, Massachusetts: The MIT Press.

Jakobson, R. and Halle, M. (1961). Tenseness and Laxness. In *Preliminaries to Speech Analysis*. Cambridge, Massachusetts: The MIT Press.

Jessen, M. (1993). Stress Conditions on Vowel Quality and Quantity in German. In *Working Papers of the Cornell Phonetics Laboratory*, volume 8, pages 1–27.

Jessen, M., Marasek, K., Schneider, K., and Clahßen, K. (1995). Acoustic Correlates of Word Stress and the Tense/Lax Opposition in the Vowel System of German. In *Proceedings of the 13th ICPhS, Stockholm*, pages 428–431.

Kingston, J. and Diehl, R. (1994). Phonetic Knowledge. *Language*, 70:419–454.

Kluender, K., Diehl, R., and Wright, B. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, 16:153–169.

Krull, D., Traunmüller, H., and van Dommelen, W. A. (2003). The Effect of Local Speaking Rate on Perceived Quantity: a Comparison Between Three Languages. In *Proceedings of the 15th ICPhS, Barcelona*.

Labov, W. (1994). *Principles of Linguistic Change. Volume I: Internal Factors*. Cambridge, Massachusetts: Blackwell Publishers.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Massachusetts: The M.I.T. press.

Lehtonen, J. (1970). *Aspects of Quantity in Standard Finnish*. Jyväskylä University.

Levander, L. (1925). *Dalmålet. Beskrivning och historia*, volume I. Uppsala: Appelbergs boktryckeri.

Levander, L. (1928). *Dalmålet. Beskrivning och historia*, volume II. Uppsala: Appelbergs boktryckeri.

Liljencrants, J. and Lindblom, B. (1972). Numerical Simulation of Vowel Quality Systems: the Role of Perceptual Contrast. *Language*, 48:839–861.

Lindblom, B. (1963). Spectrographic Study of Vowel Reduction. *Journal of the Acoustical Society of America*, 35:1773–1781.

Lindblom, B. (1987). Adaptive Variability and Absolute Constancy in Speech Signals: Two Themes in the Quest for Phonetic Invariance. In *Proceedings of the 11th ICPhS, Tallinn*.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Malmberg, B. (1944). Die Quantität als phonetisch-phonologischer Begriff. In *Lunds universitets årsskrift N.F. Avd. I, 41:2*. Lund universitet.

Mooshammer, C. (1998). Experimentalphonetische Untersuchungen zur Modellierung der Gespanntheitsopposition im Deutschen. In *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, volume 36, pages 3–192.

Noreen, A. (1907). *Vårt språk. Nysvensk grammatik i utförlig framställning*. Lund: C. W. K. Gleerups Förlag.

Nyström, G. (2000). Språkhistorien och vår tids svenska dialekter. Nya rön om Älvdalsmålets fonologi och morfologi. In *Nordsvenska 11*. Institutionen för nordiska språk, Umeå universitet.

Osthoff, H. and Brugmann, K. (1878). *Morphologische Untersuchungen auf dem Gebiete der Indogermanischen Sprachen*. Leipzig.

Pamp, B. (1978). *Svenska dialekter*. Stockholm: Natur och Kultur.

Pettersson, G. (1996). *Svenska språket under sjuhundra år*. Lund: Studentlitteratur.

Pind, J. (1986). The Perception of Quantity in Icelandic. *Phonetica*, 43:116–139.

Pind, J. (1995). Speaking Rate, Voice-Onset Time, and Quantity: The Search for Higher-Order Invariants for Two Icelandic Speech Cues. *Perception & Psychophysics*, 57(3):291–304.

Pind, J. (1999). Speech Segment Duration and Quantity in Icelandic. *Journal of the Acoustical Society of America*, 106(2):1045–1053.

R Development Core Team (2004). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Riad, T. (1992). *Structures in Germanic Prosody*. Department of Scandinavian Languages, Stockholm University.

Schaeffler, F. (2005). Phonological Quantity in Swedish Dialects: A Data-Driven Categorization. In *Proceedings from FONETIK 2005*. Institutionen för lingvistik, Göteborgs universitet.

Schaeffler, F. and Wretling, P. (2003). Towards a Typology of Phonological Quantity in the Dialects of Modern Swedish. In *Proceedings of the 15th ICPhS, Barcelona*, pages 2697–2700.

Schaeffler, F., Wretling, P., and Strangert, E. (2002). On the Development of a Quantity Typology for Swedish Dialects. In Bel, B. and Marlien, I., editors, *Proceedings of Speech Prosody 2002, Aix en Provence*, pages 627–630.

Sievers, E. (1893). *Grundzüge der Phonetik*. Leipzig: Breitkopf & Härtel, 4th edition.

Sock, R., Löfqvist, A., and Perrier, P. (1996). Kinematic and Acoustic Correlates of Quantity in Swedish and Wolof: A Cross-Language Study. In *Proceedings of the 1st ESCA ETRW on Speech Production Modeling*, pages 81–84.

Söderström, S. (1972). *Om kvantitetsutvecklingen i norrländska folkmål*. Uppsala: Almkvist & Wiksell Informationsindustri AB.

Strangert, E. and Wretling, P. (2001). Kvantitet i svenska dialekter i Norr-, Väster- och Österbotten. In Sundman, M. and Londen, A., editors, *Svenskans Beskrivning 25. Åbo: Förhandlingar vid tjugofemte sammankomsten för svenskans beskrivning*, pages 239–250.

Strangert, E. and Wretling, P. (2003). Complementary Quantity in Swedish Dialects. In *PHONUM 9, Proceedings of Fonetik 2003*, pages 101–104. Umeå University.

Sundberg, E. (1993). *Dialekten i Ålands nordöstra skärgård*. Ålands högskola, Mariehamn.

Svärd, N. (2001). Word Accents in the Närpes Dialect: Is There Really Only One Accent? In *Proceedings of FONETIK 2001, Lund Working Papers 49*, pages 160–164. Department of Linguistics and Phonetics, Lund University.

Thurgood, G. (1993). Geminates: A Cross-Linguistic Examination. In Nevis, J. A., McMenamin, G., and Thurgood, G., editors, *Papers in Honor of Frederick H. Brengelman*, pages 129–139. Fresno: California State University.

Timm, N. H. (2002). *Applied Multivariate Analysis*. Springer Verlag, New York.

Traunmüller, H. (1990). Analytical Expressions for the Tonotopic Sensory Scale. *Journal of the Acoustical Society of America*, 88:97–100.

Traunmüller, H. and Krull, D. (2003). The Effect of Local Speaking Rate on the Perception of Quantity in Estonian. *Phonetica*, 60:187–207.

Tronnier, M. (2002). Preaspiration in Southern Swedish Dialects. In *Proceedings of Fonetik 2002. TMH-QPSR*, volume 44, pages 33–36.

Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie*. Göttingen: Vandenhoeck& Ruprecht.

van Leyden, K. (2002). The Relationship between Vowel and Consonant Duration in Orkney and Shetland Dialects. *Phonetica*, 59:1–19.

van Leyden, K. (2004). *Prosodic Characteristics of Orkney and Shetland Dialects: An Experimental Approach*. LOT Dissertation Series 92. Leiden University.

Vennemann, T. (1995). Der Zusammenbruch der Quantität im Spätmittelalter und sein Einfluß auf die Metrik. *Amsterdamer Beiträge zur älteren Germanistik (ABäG)*, 42:185–223.

Vennemann, T. (2000). From quantity to syllable cuts: On so-called lengthening in the Germanic languages. *Rivista di Linguistica*, 12.1:251–282.

Vestlund, A. (1923). Medelpads folkmål. *Svenska landsmål och svenskt folkliv*, 4:3–130.

Wallström, S. (1943). *Övre Norrlands språkgeografi med utgångspunkt från Arjeplogmålet*. Uppsala: Almqvist & Wiksells AB.

Wang, W. S.-Y., editor (1977). *The Lexicon in Phonological Change*. The Hague: Mouton.

Wessén, E. (1969). *Våra folkmål*. Lund: C E Fritzes Bokförlag.

Widmark, G. (1998). *Stora vokaldansen. Om kvantitativa och kvalitativa förändringar i fornsvenskans vokalsystem*. Kungl. Gustav Adolfs Akademien, Uppsala.

Wretling, P., Strangert, E., and Schaeffler, F. (2002). Quantity and Preaspiration in Northern Swedish Dialects. In Bel, B. and Marlien, I., editors, *Proceedings of Speech Prosody 2002, Aix en Provence*, pages 703–706.

Zilliacus, K. (1992). *Åboländska*. Helsingfors: Konstsamfundet.