

INVESTIGATING THE POTENTIAL FOR MERGER OF ICELANDIC ‘FLÁMÆLI’ VOWEL
PAIRS THROUGH FUNCTIONAL LOAD

by

ALEXANDER ANKIRSKIY

(Under the Direction of Margaret E. L. Renwick)

ABSTRACT

All languages exhibit the duality of patterning, where words are formed via discrete meaningful units called morphemes, which in turn are made up of lexically contrastive phonemes. Some contrasts have been found to do more ‘work’ in a language than others. This notion of ‘work’ done by a contrast is referred to as *functional load*, which can be computed using methods from information theory. This method has been found to accurately predict the potential for elimination of a phonological contrast, such as a merger. In this thesis, functional load and other quantitative measures are used to investigate the phonological relationships in a purported near-merger of two vowel pairs in Icelandic, termed *flámæli* (‘skewed speech’) vowels. The findings indicate the high propensity for elimination of one of the two contrasts, and suggest that this near-merger is an instance of an intermediate phonological relationship. The apparent nature of the merger is confirmed.

INDEX WORDS: Icelandic, vowel merger, functional load, phonology, information theory, quantitative phonology.

INVESTIGATING THE POTENTIAL FOR MERGER OF ICELANDIC ‘FLÁMÆLI’ VOWEL
PAIRS THROUGH FUNCTIONAL LOAD

by

ALEXANDER ANKIRSKIY

B.A., University of Iceland, Iceland, 2015

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment
of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2018

© 2018

Alexander Ankirskiy

All Rights Reserved

INVESTIGATING THE POTENTIAL FOR MERGER OF ICELANDIC ‘FLÁMÆLI’ VOWEL
PAIRS THROUGH FUNCTIONAL LOAD

by

ALEXANDER ANKIRSKIY

Major Professor: Margaret E. L. Renwick

Committee: Keith Langston
Mark R. Wenthe

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2018

ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. Margaret E. L. Renwick, for invaluable help and mentorship during my time in the graduate program; my committee members, Drs. Mark Wenthe and Keith Langston, for their help and support as instructors, committee members and fellow linguists; Jón Björn Ólafsson, Malo Adeux, Birgir Örn Sigurðsson, Jonathan Crum, and Nakita Barakadyn for their friendship, advice, and support; Becky Sexton, Yuki Sasaki Caldwell, and Kihoon Kim for being the best colleagues a graduate student could want; Dr. Masaki Mori and Mary Miller, for their guidance; Longlong Wang, Dot-Eum Kim, Rachel and Michael Olsen, Sofia Ivanova, Joey Stanley, Katy Dwyer, Ryan Dekker, and Trevor Ramsey, my fellow graduate students; Jill Talmadge and Martha Babendreier for their assistance and advice on all major administrative and life issues; my parents Elena and Viktor, who always believed in me, supported me and all my efforts, and continue to do so; Bonzo, for his feline companionship and outlook on life.

Most of all I am grateful to my beloved wife Rachel. I wouldn't have been able to do this without your relentless love and support, no matter what.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	iv
CHAPTER	
1 INTRODUCTION.....	1
Background.....	1
Research Objectives and Questions.....	3
Thesis Structure.....	4
2 BACKGROUND INFORMATION.....	6
Functional Load.....	6
The Icelandic Vowel System.....	11
Flámæli, or 'Skewed Speech'.....	14
Intermediate Phonological Relationships.....	20
3 METHODS AND PREDICTIONS.....	24
Predictions.....	24
Data.....	25
Corpus Processing.....	25
Calculations.....	27
4 RESULTS.....	32
Frequency Distribution of Vowels.....	32
Functional Load.....	33

	Kullback-Leibler Divergence.....	38
	Predictability of Distribution.....	40
5	DISCUSSION.....	42
	Flámæli and Functional Load.....	42
	Phonological Status of Flámæli Contrasts.....	45
	Directions for Further Study.....	48
6	CONCLUSION.....	50
	BIBLIOGRAPHY.....	52
	APPENDICES.....	60
A	Corpus Samples.....	60
B	R Script.....	65
C	Perl Script.....	67

CHAPTER 1

INTRODUCTION

1.1. Background

All human languages exhibit what is referred to as the “duality of patterning”, that is the ability to form discrete meaningful units, or morphemes, which in turn are made up of mostly meaningless, but contrastive elements that are phonemes (Hockett 1960; Sandler et al. 2011). Over the course of the life of a language, certain phonological changes occur: phonemes can merge, shift, split, and cease to exist. Some good examples are the merger of high and mid-high front vowels (i- and e-types) in a variety of British English (Norwich area), referred to as the *beer-bear merger* (Trudgill 1974); the Northern Cities Shift, a complex vowel shift, some of the characteristics of which include the raising, tensing and lengthening of /æ/, the fronting of /ɑ/, the lowering of /ɔ/, etc. (Labov 1994); an example of phoneme deletion would be the tendency of some American English speakers in the Northeastern states to pronounce the word *human* as [jumən] rather than [hjumən], thereby deleting [h] before the high front glide (Labov, Ash & Boberg 2006).

Observed changes like these caused some researchers to suggest (Mathesius 1929; Trubetzkoy 1939) that some phoneme contrasts might be doing more ‘work’ than others, and that this ‘work’ might indicate the degree of said contrasts’ stability, in particular concerning mergers. The notion of *functional load* was introduced (Martinet 1952), implying that every phonological contrast in a language carries a certain amount of yield and depending on whether the load is high or low, a contrast would be less or more likely to merge, respectively.

Currently, functional load is used in various fields in linguistics, such as automatic speech recognition (ASR), historical linguistics, child language acquisition, and linguistic typology (Surendran & Niyogi 2003). It can also be used to determine whether a contrast in a language is likely to be eliminated, and help investigate sociolinguistic implications in cases where only a certain group of speakers perceives and/or produces a phonological change. Until recently the functional load hypothesis was largely unsupported, with various calculation methods and inconclusive results failing to find proof that functional load accurately predicts the stability of a phonological contrast. However, in 2013 such evidence was found (Wedel, Kaplan & Jackson 2013), and thus the way was paved for the use of functional load to research phonological contrasts (Wedel, Jackson & Kaplan 2013; Renwick et al. 2016; Renwick & Ladd 2016; Nadeu & Renwick 2016; Renwick & Nadeu 2018).

This thesis uses functional load, as well as other quantitative measures to investigate the potential for merger of two mid vowel pairs in Icelandic, referred to here as the “*flámæli*” ([ˈflau:maiː, lɪ]) vowel pairs. This phenomenon emerged in the 19th century in Iceland and was viewed as something uncouth, to the point where the Icelandic government took steps to eradicate it by means of strict prescriptive instruction in schools. Had this merger occurred, words such as those in (1) would be perceived as homophonous:

(1) *Flámæli* examples

skyr [skɪːɾ] ‘Icelandic milk product’ vs. *sker* [skɛːɾ] ‘reef’
flug [flyːx] ‘flight’ vs. *flög* [flœːx] ‘patch of ground without vegetation-PL’

The findings of studies conducted in the late 20th – early 21st centuries (Björn Guðfinnsson 1981; Höskuldur Þráinsson & Kristján Árnason 1984; Birna Arnbjörnsdóttir 1990; Birna Arnbjörnsdóttir 2006) suggest that the merger did not, in fact, occur; most likely it was not a full, but rather an apparent merger, where two sounds do not merge, but rather assume

intermediate qualities (Labov, Yaeger & Steiner 1972; Labov 1994). This helps shed more light on the success of the government's effort to eradicate the phenomenon.

1.2. Research Objectives and Questions

Three main objectives are addressed in this thesis. The first objective is to investigate functional load values in relation to the perceived vowel merger in Icelandic spoken in Iceland (Modern Icelandic, or MI henceforth) and synchronically in the variety of the Icelandic language spoken in North America from the early 1900s to the present day (termed North American Icelandic, NAI henceforth).

The second objective is to determine whether the merger is likely to occur in the future in Iceland, in MI (Birna Arnbjörnsdóttir 1990)¹, and therefore to determine whether past prescriptive efforts to eradicate the merger as a stigmatized feature were substantiated, in other words, to determine whether there was an actual danger of lexical and morphological confusion. Prescriptivism itself is never justified; languages change, and stifling this change goes against the natural course of events.

The final objective is to determine whether the theory that *flámæli* is an apparent, rather than full, merger, which is further supported by quantitative evidence from functional load and other measures. The first objective is accomplished through a discussion of quantitative corpus data. The prediction for the second and the third objectives is that the functional load values will be lower for the *flámæli* vowels than for other contrasts in the language, thus indicating a higher likelihood for the elimination of these phonological contrasts in the future, and providing additional evidence to the apparent merger theory and substantiating the Icelandic government's

¹ Due to the Icelandic naming traditions, Icelandic authors are cited using their full names. Icelandic names consist of the first name, e.g. Jón, a middle name, most of the time omitted, and the patronymic, e.g. Ólafsson "son of Ólafur". In rare instances where an Icelandic individual does have a last name, e.g. Laxness, they will be cited using regular style guidelines.

concerns. Additional quantitative methods such as the predictability of the vowels' distribution and the Kullback-Leibler measure of dissimilarity between probability distributions are also involved in the determination of the phonological status of the mergers. The specific research questions posed are thus the following:

1. Is the apparent merger of the *flámæli* vowel pairs related to their functional load?
2. Are the *flámæli* contrasts likely to be lost in the future?
3. Had the *flámæli* contrasts been lost, would much confusion ensue?
4. Are the *flámæli* contrasts in a contrastive or allophonic relationship?
5. If the *flámæli* vowels are not in either contrastive or allophonic relationship, what factors play the biggest role in the difficulty in assigning one or the other type of relationship to these particular contrasts?

1.3. Thesis Structure

This thesis is comprised of five chapters, this introduction being the first. The second chapter presents background information on the main topics of the thesis: functional load, the vowel system of Icelandic, and the phenomenon of *flámæli*, as well as a review of the literature on these topics. In addition, part of the chapter is dedicated to the introduction of the concept of an intermediate phonological relationship. Chapter 3 discusses the predictions for the aforementioned research questions, introduces the data set and the experimental methods used to investigate the vowel contrasts, including functional load, the Kullback-Leibler Divergence, and Predictability of Distribution. The fourth chapter provides the results of the calculations presented in Chapter 3, and finally, Chapter 5 is devoted to the discussion of the results and the phonological status of the contrasts in question. The thesis is wrapped up with the conclusion,

followed by the bibliography and the Appendices, which present samples of the data set, scripts and programs used to obtain the quantitative results.

CHAPTER 2

BACKGROUND INFORMATION

2.1. Functional Load

The concept of functional load (FL) is not exactly new and is fairly well known among linguists, having been researched since the end of the 1920s. The first mentions of the term can be found in the works of the Czech structural linguist Vilém Mathesius (1929), and his colleague from the Prague Linguistic Circle, Nikolai Trubetzkoy (1939). The first detailed hypothesis was formulated by André Martinet (1952), who suggested that “*a phonological opposition that serves to keep distinct the phonemes in hundreds of most frequent and useful words might offer a more successful resistance to elimination than one which only serves a useful purpose in very few instances*”. Martinet went as far as was possible without offering any specific method of quantifying the functional load of a contrast. He defined functional load as the number of lexical pairs (minimal pairs) which would have been complete homophones if it were not the case that one word of the pair presents member A of the opposition where the other shows member B. The current understanding of functional load, which is discussed later in the chapter, is different from this.

The hunches that Martinet had about functional load were in some cases on the right track, and in other cases wrong: as an example of the former, he suggested that if the number of minimal pairs as mentioned above is high, for instance taking the /p~/b/ contrast in English, the functional load, or yield, would also be high. The opposite will hold true for a contrast with a small number of minimal pairs, such as /θ~/ð/ in English: a small minimal pair count equals low

functional load. This, as will be shown further, was a starting point of the current understanding of functional load. An example of Martinet's conclusions that turned out to be correct is the very definition of the term, proposed in his 1952 article.

In his later essays on the subject Martinet somewhat refined the definition of functional load, proposing that it referred to the amount of "work" done by a phonological contrast in a language (1955), but never suggested any method of quantification for it. Such suggestions were made by a number of other researchers (Kučera 1963; Hockett 1955; Hockett 1966; Wang 1967; King 1967). Henry Kučera proposed a method aimed at distinctive features, rather than phonological contrasts, and his data set included two corpora of 100,000 phonemes: one from Russian and one from Czech. Charles Hockett in 1955 proposed the first method for the quantification of functional load of a phonological opposition: his methodology was novel in that it took into account word frequency, something that hadn't been suggested in prior research, which focused on minimal pair count. Hockett's definition of FL involved some information theory-specific concepts introduced by Shannon (1948), including the concept of entropy, or the amount of uncertainty in a linguistic system. This definition treats language as a sequence of phonemes, whose entropy can be computed. The sequence in question is infinite, representing all possible utterances in the language, which is, of course, impossible in practice. Hockett's idea was that functional load can be expressed in the following equation:

Equation 2.1.1.

$$FL(x, y) = \frac{H(L) - H(L_{xy})}{H(L)}$$

In the above equation $FL(x, y)$ is the functional load of the contrast between phonemes x and y , $H(L)$ is the amount of information H transmitted by the language L . The denominator is included to make the equation interpretable as the fraction of information lost by L when the x/y

opposition is eliminated (Surendran & Niyogi 2003): the higher the uncertainty created by the elimination of a contrast, the higher the entropy.

Despite suggesting this methodology, Hockett did not, in actuality, perform any calculations with it. This was attempted (unsuccessfully) by William Wang (1967), and Robert D. King (1967). King found no evidence for Martinet's hypothesis, and came to the conclusion that it was false, perhaps overly harshly.

All of the aforementioned works focused on individual case studies, since the data and the computational power were extremely limited (the computations were usually done with the help of decimal computers such as the IBM 1410), and so for the time being the results remained inconclusive.

One of the main claims about functional load directly derives from the original assumption by Martinet: the greater the contribution that a phonological contrast brings to word differentiation, the less likely this contrast is to be eliminated over the course of language history. What it means for this thesis is that the higher the FL score, the less likely a phoneme pair is to merge, and the lower the FL score, the more it is likely that the merger will occur. As mentioned above, until recent time there was no evidence to support this claim. In their 2013 study Wedel, Kaplan, and Jackson presented the first analysis providing significant evidence that functional load scores are indeed indicative of the potential for elimination of phonological contrasts in that the probability of phoneme merger is inversely related to the functional load. In this study they used data from multiple languages in order to calculate the functional load scores for several phoneme mergers. They followed the work of Surendran and Niyogi (2006; 2003), using a combination of two methods of functional load calculations, discussed below.

The assessment of functional load in Wedel et al.'s (2013) study relies on two major

methods of calculation: the minimal pair count and the change in entropy count. Two levels of analysis were investigated for this study: the word level and the phoneme level. At the word level, the number of minimal pairs in the corpus defined by a phoneme pair was compared to the change in word-level entropy of the corpus upon merger of the phoneme pair. At the phoneme level, the phoneme probability in a corpus was compared to the change in the phoneme-level entropy upon merger of the phoneme pair. The individual phoneme probabilities are calculated according to Equation 2.1.2, where ϕ_i is a phoneme, $f(\phi_i)$ is the frequency-count in the corpus, C is the size of the corpus, and Φ is the set of all phonemes in that corpus.

Equation 2.1.2. (From Wedel, Kaplan & Jackson 2013:180)

$$p(\phi_i|C) = \frac{f(\phi_i)}{\sum_{\phi \in \Phi} f(\phi)}$$

Apart from the individual probabilities, the predictive values of the higher-probability member and (separately) the lower-probability member of the phoneme pair, as well as their sum, were also tested.

The second method of calculating the functional load of a phonological opposition, also used in that particular study involves the calculation of change in entropy, and is based on methods borrowed from information theory. In order to adequately explain this definition of functional load, a brief clarification of the concept of entropy needs to be provided.

As Surendran and Niyogi explain it,

Suppose there are two native speakers of L in a room. When one speaks, i.e. produces a sequence of phonemes, the other one listens. Suppose the listener fails to understand a phoneme and has to guess its identity based on her knowledge of

L. $H(L)$ refers to the uncertainty in guessing; the higher it is, the harder it is to guess the phoneme, and the less redundant L is. (2006:47)

This uncertainty, or $H(L)$, is the entropy that can be measured. Since it is impossible to obtain all of the phoneme sequences in a language, the entropy is calculated for a corpus C , representative for the given language.

Wedel et al. (2013) calculated the change in entropy for their corpora based on Equations 2.1.3 and 2.1.4, where x is an element from the set X , and $p(x)$ is the measured probability of the element x in the corpus C . The entropy was calculated at both phoneme and word level: for the phoneme level, X is the set of the phoneme types, while at the word level X is the set of word types in the corpus.

Equation 2.1.3. (From Wedel, Kaplan & Jackson 2013:181)

$$H(C_x) = - \sum_{x \in X} p(x) \log_2 p(x)$$

The change in entropy in a corpus upon a loss of contrast /a/~/b/ is calculated according to Equation 4. In this equation, $H(C_x)$ is the entropy of a corpus measured at level X , while $H(C_{X,a \sim b})$ is the entropy calculated with phonemes a and b merged by replacement with a single phoneme.

Equation 2.1.4. (From Wedel, Kaplan & Jackson 2013:181)

$$\Delta H(C_{X,a \sim b}) = H(C_x) - \frac{H(C_{X,a \sim b})}{H(C_x)}$$

The quantitative data obtained using the calculations described above was used in statistical analyses, the results of which represent the first statistical evidence that the probability of phoneme merger is inversely related to functional load.

The Wedel et al. (2013) study thus paved the way to using functional load in quantitative linguistic research, combined with other statistical and information theory-specific methods (Renwick et al. 2016). In this thesis I follow these and other studies to obtain functional load and other quantitative and statistical data and use the results to make generalizations about vowel contrasts in Icelandic.

2.2. The Icelandic vowel system

Before providing an overview of the Icelandic vowel system, it is necessary to note that in this thesis I discuss two varieties of the Icelandic language. The first variety, which is spoken currently in Iceland, I refer to as Modern Icelandic (MI). The second variety, an immigrant language spoken by communities of Icelanders and their descendants in North America (Manitoba in Canada and North Dakota in the United States), is referred to as North American Icelandic (NAI). The two varieties are mutually intelligible, displaying some minor morphological, lexical, and phonological divergence, and NAI shows traits consistent with trends observed for immigrant languages (Haugen 1953), such as morphological simplification, major influx of lexical borrowings, and phonological simplification (Birna Arnbjörnsdóttir 1990). The latter is of the most importance to this thesis, as it includes the perceived merger in the system of front unrounded (/i/~ε/) and rounded (/y/~œ/) vowels, also observed in the middle of the 20th century in Iceland (Kristján Árnason 2011), termed *flámæli*, or ‘skewed speech’. This merger and its nature are further discussed in Section 2.3.

The vowel system of Icelandic consists of 8 monophthongs and 5 true diphthongs (Kristján Árnason 2011; Birna Arnbjörnsdóttir 2006; Stefán Einarsson 1945; Gussmann 2011); it is a rather large vowel system, compared to others among the world’s languages (Maddieson 2013). The vowel system is presented in Table 2.2.1 below.

Table 2.2.1. The Icelandic vowel system

	front		back	
	-round	+round	-round	+round
high	/i/			/u/
high mid	/ɪ/	/ʏ/		
low mid	/ɛ/	/œ/		/ɔ/
low			/a/	

In Table 2.2.2, the phonemic units are represented in IPA notation, with the most common orthographic representation in parentheses:

Table 2.2.2. Phonemic and orthographic representations of Icelandic vowels

Phoneme	Orthography	Example	Gloss
/i/	(í, ý),	<i>bíll</i> [pitl̥]	‘car’
/ɪ/	(i, y)	<i>mynd</i> [mɪnt̥]	‘image’
/ɛ/	(e)	<i>fer</i> [fɛr]	‘to go-1.PRS’
/a/	(a)	<i>maður</i> [ˈma:ðʏr̥]	‘man’
/u/	(ú)	<i>úlfi</i> [ˈulvʏr̥]	‘wolf’
/ʏ/	(u)	<i>gulur</i> [ˈkʏ:lʏr̥]	‘yellow’
/ɔ/	(o)	<i>koma</i> [ˈkɔ:ma]	‘to come-INF’
/œ/	(ö)	<i>örn</i> [ˈœ:r̥t̥n̥]	‘eagle’

For the sake of consistency in the overall description of the vowel system, a short description of the diphthongs is necessary. The Icelandic diphthongs are combinations of one of the monophthongs and either /i/ or /u/. The /i/-diphthongs come in two varieties: rounded and unrounded. These diphthongs can be treated as combinations of /a/ and /o/ with /u/ (/u/-diphthongs) and /a/, /ɛ/, and /œ/ with /i/ (/i/-diphthongs), thus being concatenations of two monophthongs. Examples are presented in Table 2.2.3:

Table 2.2.3: MI diphthongs

Roundedness	Main vowel	Phoneme	Orthography	Example	Gloss
Unrounded	/i/-diphthongs	/ai/	(æ)	<i>dæmi</i> ['dai:.mɪ]	‘example’
		/ei/	(ei, ey)	<i>ey</i> [ei:]	‘island’
Rounded	/u/-diphthongs	/œi/	(au)	<i>auga</i> ['œy:.ga]	‘eye’
		/ou/	(ó)	<i>fólk</i> [fou:lk]	‘people’
		/au/	(á)	<i>máni</i> ['mau:.nɪ]	‘moon’

It is therefore evident that the /u/ diphthongs are combinations of /a/ and /o/ with /u/ as a second part, and that the /i/ diphthongs are concatenations of /i/ with /a/ and /ε/. Kristján Árnason (2011) argues that the apparent complication in the case of /œi/, being that the diphthong is normally represented phonetically as [œy], where the second constituent is rounded, can be explained by the rounding spread from [œ] to the original [i] due to coarticulation.

Each of the vowels in MI comes in short and long variants, as illustrated in Table 2.2.4:

Table 2.2.4: Long/short opposition in MI vowels and diphthongs

Long	Short
1. Monophthongs <i>kýr</i> [ki:r] ‘cow’ <i>skyr</i> [ski:r] ‘a variety of yogurt’ <i>hefur</i> ['hε:.vʏr] ‘to have-3.PRS’ <i>dalur</i> ['ta:.lʏr] ‘valley’ <i>búð</i> [pu:θ] ‘shop’ <i>boðið</i> ['bɔ:..ðiθ] ‘to offer-PST.PTCP’ <i>hugur</i> ['hy:..ʏʏr] ‘mind’ <i>öl</i> [œ:l] ‘beer’	<i>kýrnar</i> ['ki:r.naʁ] ‘cows-PL.DEF’ <i>synd</i> [sɪnt] ‘sin’ <i>fell</i> [fɛtl̥] ‘mountain’ <i>vatn</i> [va^ht̥n̥] ‘water’ <i>kúrnum</i> ['ku:r.nʏm] ‘diet-DEF.DAT’ <i>borða</i> ['pɔr.ða] ‘to eat-INF’ <i>gull</i> [kʏtl̥] ‘gold’ <i>björn</i> [bjœrt̥n̥] ‘bear’
2. Diphthongs <i>mæta</i> ['mai:.ta] ‘to appear-INF’ <i>auður</i> ['œy:..ðʏr] ‘wealth’ <i>eygja</i> ['ei:..gja] ‘to see-INF’ <i>mál</i> [mau:l̥] ‘matter’ <i>lón</i> [lou:n] ‘lagoon’	<i>sætt</i> [sai^ht] ‘settlement’ <i>auðs</i> [œyθs] ‘wealth-GEN’ <i>eygst</i> [eixst] ‘to see-PST.PTCP.MID’ <i>málfræði</i> ['maul.frai:..ði] ‘linguistics’ <i>lóns</i> [louns] ‘lagoon-GEN’

The distribution of long and short vowels appears to be generally regular and follow the so-called length rule (Gussmann 2001; Kristján Árnason 2011), whereby open syllables have long vowels, and closed syllables have short vowels, with several exceptions, such as when consonant clusters of /p, t, k, s/ followed by /v, j, r/ form well-formed onsets to the following syllable, or when the nucleus is formed by a diphthong. Vowel length plays a very important role in the distribution of the perceived vowel merger discussed in greater detail in the next section.

Vowel length in MI is determined by syllable weight, and by stress: it is necessary to mention that in Icelandic the primary stress always falls on the first syllable, while secondary stress falls on every second syllable after that. Icelandic morphemes are usually mono- or disyllabic; longer compounds are formed via combination of these:

- Monosyllabic: *mál* ['mau:l] 'language'
- Disyllabic: *fræði* ['frai:.ði] 'study'
- Combination of the two: *málfræði* ['mau:l.frai:.,ði] 'linguistics'
- Same with added suffix: *málfræðingur* ['mau:l.frai:.,ðɪŋ.kvɹ̥] 'linguist'

2.3. Flámæli, or 'Skewed speech'

While Icelandic is generally regarded as a homogeneous language with no significant dialectal variation dialectally and synchronically (barring, perhaps, the North American variety), some phonological features do exist that may identify a speaker from a certain geographical area of the country (Birna Arnbjörnsdóttir 1990). In the 1940s the first extensive study of dialectal variation in Icelandic was undertaken by Björn Guðfinsson (1946; 1964; 1981). The original study involved 10,000 informants, including 7,000 children, or 93% of all children in Iceland between the ages of 10 and 13. Eight variant groups of speakers were identified, however, only

one group is of interest to this thesis, namely the group that exhibits the perceived context-free merger of the high mid vowels /ɪ/ and /ʏ/ with the corresponding low mid vowels /ɛ/ and /œ/. This merger is almost exclusively restricted to the long system; however, there are still instances of *flámæli* in short vowels. Because of this, the length of *flámæli* contrasts is not specified.

Table 2.3.1. The perceived merger in the Icelandic vowel system

	front		back	
	-round	+round	-round	+round
high	/i/			/u/
	/ɪ/	/ʏ/		
mid	↕	↕		
	/ɛ/	/œ/		/ɔ/
low			/a/	

This feature emerged in the 19th century in three key areas of Iceland (see Figure 2.3.1), and quickly became stigmatized as ‘slack-jawed speech’ (Birna Arnbjörnsdóttir 2006).

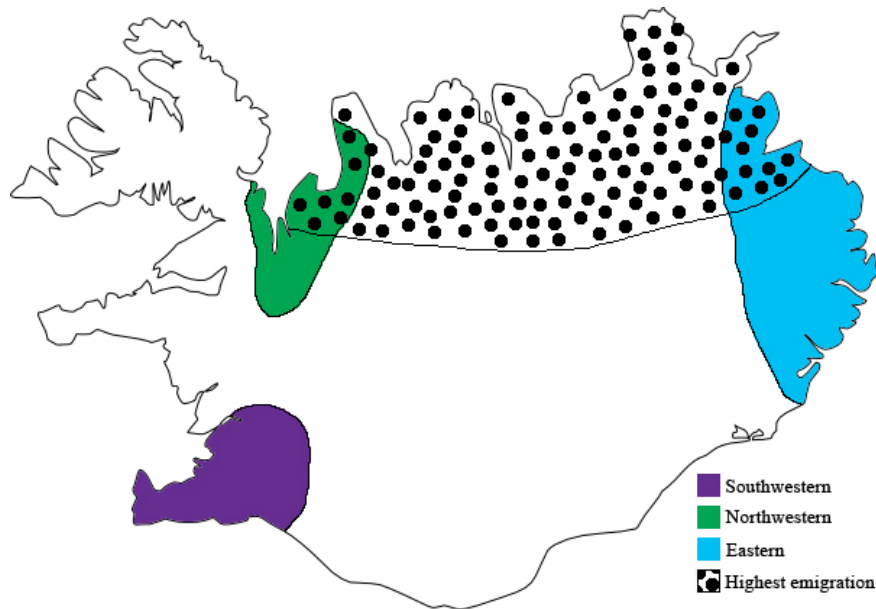


Figure 2.3.1. Geographical distribution of Flámæli in Iceland (Birna Arnbjörnsdóttir 2006)

For a better understanding of the social complications, a brief explanation of Icelandic language purism is in order. Icelandic is a conservative language (Vikør 2010), mainly because of the isolated nature of the area, tightly-knit circle of 321,040 speakers (Simons & Fennig 2018) and the identification of the language with the culture and literature (Birna Arnbjörnsdóttir 1997). Language contacts and active borrowing go as far back as the 9th century, when speakers of then Old Norse encountered Middle English in fellow settlers of Iceland. Contacts with Danish and German have also brought many loanwords into Icelandic (Eyvindur Eiríksson 1991). As a result, linguistic purism has been active in Iceland since at least 1609, when Arngrímur Jónsson the Learned called for the necessity to preserve the purity of Icelandic (Gísli Pálsson 1989). The main goal of the purism movement has been the eradication of foreign influences as well as internal linguistic changes, and one of the most successful efforts of such purification has been a widespread movement to battle *flámæli*, which has been viewed as ‘slack-jawed’ speech, thought to hinder communication because of confusion in pronunciation, which has also been transferred into orthography by schoolchildren (Birna Arnbjörnsdóttir 1990; Blöndal 1985).

In her extensive investigation of *flámæli* Þórunn Blöndal cites two examples of the criticism unleashed on this variation by the leading Icelandic language scholars of the day, Jakob Jóhannsson Smári and Jón Ófeigsson (1985):

Aftur er mesta nauðsyn, að uppræta algjörlega Nesjaframburðinn svonefnda...

Virðist hann heldur breiðast út og vita flestir íslenzkukennarar hér syðra, hver

vandræði eru að fást við þenna leiða málgalla.

[Again, it is of the greatest necessity to completely eradicate the Peninsula Pronunciation². ... It seems to be spreading out, and many teachers of Icelandic here, down South, know what sort of trouble goes with this diseased language.]
(Jakob Jóh. Smári 1917:112)³

Udtalen anses for plebejsk og undgaas af alle dannede Mennesker.

[This pronunciation is considered plebeian and is avoided by all well-rounded men.]

(Jón Ófeigsson 1980:XXVII)

Flámæli had its heyday in the middle of the 20th century (Kristján Árnason 2011), but the aforementioned calls for eradication were indeed answered by the government. The main onslaught was undertaken by teachers of Icelandic, who were given strict guidelines on teaching the pronunciation of *réttmæli*, or the “correct speech” (Björn Guðfinnsson 1981:55–63). As a result of this coordinated effort, *flámæli* was eradicated in Iceland and is not considered a feature of Modern Icelandic anymore.

Due to socio-economic circumstances, not the least being a near-famine due to volcanic eruptions, diseases and the death of thousands of sheep, on which Icelanders depended for food, the end of the 19th and the beginning of the 20th centuries saw mass migration to North America, mostly to Manitoba in Canada (so-called New Iceland), and North Dakota in the United States (Kristjanson 1965). The population of Iceland at the time was a little under 48,000, and the

² "Peninsula Pronunciation" (Icel. Nesjaframburður) refers to the southwestern variety of *flámæli*, geographically situated on the Reykjanes peninsula.

³ From here on, all quotes translated by the author.

numbers of emigrants were relatively small, compared to the numbers of emigrants from other countries: it is estimated that about 15,000 Icelanders settled in Canada and the United States between 1873 and 1914, when immigration ceased (Birna Arnbjörnsdóttir 1990).

The bulk of immigration occurred from the areas identified in Figure 2.3.1, the North and Northeast of Iceland, which coincided with the Northwest and East *flámæli* areas. Thus *flámæli*, not yet eradicated at that point, was brought to North America and allowed to develop unhindered. This provided a great opportunity for research, which was conducted in the 1980s by Birna Arnbjörnsdóttir (1987; 1990; 1997; 2006). The results of this study are discussed later in this chapter.

As shown in Table 2.3.1, the perceived merger involved the merger of /ɪ/ with /ɛ/, and of /ʏ/ with /œ/. More accurately, however, /ɪ/ and /ʏ/ were lowered, while /ɛ/ and /œ/ were raised to produce intermediate vowels, resulting in a merger by approximation, according to Labov's classification of mergers (Labov 1994). To an observer it appears that the minimal pairs are homophonous, while a more careful analysis reveals intermediate vowel qualities. This was found during Björn Guðfinnsson's study in the 1940s. He found that speakers from three areas identified in Figure 2.3.1 had a widespread tendency to merge the aforementioned vowels in the long system. He predicted that the apparent merger would spread to the short vowel system, the evidence of which he had (see the example in (2)), but admitted that it is not very frequent (1946).

(2)

<i>kind</i> [kɪnt] 'sheep'	becomes homophonous with	<i>kennd</i> [kɛnt] 'feeling'
<i>lund</i> [lynt] 'mood'	becomes homophonous with	<i>lönd</i> [lœnt] 'countries'

Prior to Birna Arnbjörnsdóttir's study in late 1980s, Kristján Árnason and Höskuldur Þráinsson conducted an investigation of Icelandic linguistic variation called RÍN (1984), during which the numbers of contemporary speakers exhibiting *flámæli* were compared to those made by Björn Guðfinnson almost forty years earlier. The numbers had decreased drastically in comparison to the previous research; in addition, it was also determined that the earlier variety of *flámæli* (the lowering of /ɪ/ and /ʏ/) has disappeared, but the 'new' variety (raising of /ɛ/ and /œ/ to intermediate qualities) was still present in the speech of teenagers in Reykjavík. The RÍN study also found that the raising of /ɛ/ and /œ/ were separately driven processes from those involved in the lowering of /ɪ/ and /ʏ/. Kristján Árnason notes in (2011) that currently *flámæli* only occurs in the speech of older people, and can be considered almost extinct.

Different linguistic factors were tested for their influence on *flámæli*: vowel length, preceding environment, and following environment. Of these three, only the vowel length had a significant impact; a very small effect was also found to be produced by the following environment. Þórunn Blöndal investigated three groups of sounds (stops, fricatives, and sonorants) following the *flámæli* vowels for their effect on the lowering of /ɪ/ and /ʏ/, and the raising of /ɛ/ and /œ/ (1985). Her findings were that 1) a following stop was the most constraining environment to the appearance of *flámæli* of /ɪ/ and /ʏ/ (56.05% of words where conditions for *flámæli* were met); 2) *flámæli* appeared in 75.22% of words where vowels were followed by a fricative; 3) it appeared in 77.65% of words where vowels were followed by sonorants. The small fluctuation between the highest and the lowest values here could indicate that the manner of articulation is not highly influential in affecting *flámæli*.

Birna Arnbjörnsdóttir in her research of *flámæli* in North American Icelandic in the 1980s and 1990s found that despite Björn Guðfinnsson's predictions it did not spread into the

short vowel system and is still confined to the long vowels, despite having spread diachronically (with a gradual increase in *flámæli* from the oldest to the youngest speakers). The second major finding was that the lowering of /ɪ/ and /ʏ/ and the raising of /ɛ/ and /œ/ are motivated by different factors, in that while the latter is confined almost exclusively to long vowels, the former can occur in short vowels, but there seems to be no other significant linguistic factor influencing it, and therefore only a few instances exhibit the raising. The frequency of occurrence is reported to be much lower for the raised /ɛ/ and /œ/ than for the lowered /ɪ/ and /ʏ/ (Birna Arnbjörnsdóttir 1987; 1990; 1997; 2006), which led to the conclusion that the short vowel system as a whole was not affected.

The main conclusion of the aforementioned study of the North American variety of Icelandic is that *flámæli* is not the case of a full merger, but rather an apparent merger, discussed by Labov et al. (1972), Trudgill (1983), and Harris (1985). In an apparent merger the vowels may approach each other without actually becoming identical and then splitting again. This explains why the phenomenon was successfully reversed in Iceland, where probably the same type of merger was underway. Another explanation for the success of the government's efforts in eradication of *flámæli* is that had the merger actually occurred, it would have been nearly impossible to reverse: once a merger – always a merger (Labov 1994).

2.4. Intermediate Phonological Relationships

The fact that *flámæli* was eradicated successfully in Iceland but lives on in North America, albeit assuming the intermediate qualities of vowels instead of a full-on merger may point at an instability of the phonological relationship between the two vowels. This suggests that *flámæli* may be a case of a so-called Intermediate Phonological Relationship.

Historically one of the central concepts of phonology has been that of contrast: two phonemes can be in either a contrastive or allophonic relationship, that is, they either can or cannot be used to distinguish between different words. In actuality, it is widely acknowledged that this categorical view is insufficient to adequately describe certain phonological relationships, particularly those that are neither contrastive, nor allophonic, such as, for instance, the German /ç/~/χ/ (Moulton 1947), the Spanish /ɾ/~/r/ (Hualde 2005), the Romanian /Λ/~/i/ (Renwick et al. 2016), or the Central Catalan and Italian mid vowels (Renwick & Nadeu 2018; Ladd 2006; Renwick & Ladd 2016). Some of the studies on marginal contrasts, as they are most often termed (Hall 2013), suggest that functional load can provide information on how phonological contrasts evolve and are perceived (Renwick et al. 2016; Wedel, Jackson & Kaplan 2013).

Hall (2013), in her typology of intermediate phonological relationships, discusses six major types of such relationships, with emphasis on the reasons for their intermediate nature. Three of these reasons are relevant to this thesis: problems with the concept of predictability of distribution, issues related to multiple strata, and issues of frequency. The discussion on how these reasons apply to the findings of this thesis follows in Section 5.2.

The first type of intermediate phonological relationships relevant here is where there exists a **problem with the concept of predictability of distribution**. Two subtypes can be distinguished: the choice between two sounds is either mostly unpredictable, but has some degree of predictability, such as in the /ɾ/~/r/ relationship in Spanish (Hualde 2005), or is mostly predictable, but has a few contrasts. An example of the latter is Canadian Raising, where the diphthongs [aɪ] and [Λɪ], or [ɑʊ] and [Λʊ] are generally predictably distributed, but still have surface near-minimal pairs that are distinguished by the two vowels, such as *writing* [ɾΛɪɾɪŋ] and *riding* [ɾaɪɾɪŋ] (Trudgill 1985). This is relevant to this thesis in that the distribution of the *flāmæli*

vowels is mostly unpredictable, but has some degree of predictability, as will be shown in Section 4.4.

The second type deals with cases where there are **multiple linguistic strata** in a language (polysystemicity). These strata may arise from historical foreign influences on the language, for instance, borrowings from other languages, where some of the source language's phonological principles also migrate into the target language. An example of this kind of borrowing is reported by Ladd (2006): several indigenous Mexican languages where voiced stops are normally allophonic are beginning to have contrastive voiced stops through contact with and subsequent borrowings from Spanish. This type of intermediate phonological relationships is of interest to this thesis because of the language contact-induced changes in North American Icelandic, imposed by the surrounding superstrate (English).

The third type of intermediate phonological relationships relevant to the present discussion is when there exist **issues of frequency**, that is, when the elements within the relationship occur with very low frequency. This can overlap with the cases of polysystemicity in that if a phone occurs only in loanwords, it is likely to be less frequent, and lexical exceptions tend to occur with lower frequencies than do regular patterns (Hall 2013). An example of such relationship is reported, among others, by Watson (2002), who claims that many of the “marginal or quasi-phonemes” that she lists for Cairene Arabic only occur in a few words. This is of interest to this thesis due to small number of minimal pairs found for the *flámæli* vowel pairs.

The other three reasons for intermediate phonological relationships are those where there are issues of gradience and variability, those where there exist issues of phonetic similarity, and those where subsets of natural classes are present. These, however, are not relevant to this thesis, and are therefore only mentioned for the sake of completeness.

While there is a multitude of different criteria by which a phonological relationship may be deemed either allophonic, or contrastive, or neither (intermediate), until recently there has not been suggested a way to include the latter into the phonological system, which is still based on the categorical two-way distinction. Hall (2009) proposed a model of phonological relationships that is based on a continuous scale of predictability rather than a binary distinction between ‘allophonic’ and ‘contrastive’. This model, like that of functional load, builds on methods from information theory, and provides a way of calculating the precise degree of distribution of two sounds. This calculation is employed in this thesis as the measure of Predictability of Distribution, and is formally introduced in the next chapter.

CHAPTER 3

PREDICTIONS AND METHODS

3.1. Predictions

As stated in Chapter 1, this thesis has five major research questions. The initial hypotheses for these questions are centered around the prediction that the functional load values for *flámæli* contrasts will be low, thus indicating their high propensity for merger, which would help explain the emergence of the phenomenon in the first place. However, there remains a question of why the Icelandic government was able to stop the merger, since mergers are generally irreversible by linguistic means (Labov 1994). The government's concern was that the merger would result in confusion; the fact that the morphology of Icelandic involves frequent use of two of the four vowels (/ɪ/ and /y/, Thomson 1987) suggests that their concerns, but not actions, were justified. The last two research questions are essentially one big question, dealing with the concept of an intermediate phonological relationship, outlined earlier in Chapter 2. The prediction for this is that the contrasts are indeed in such a relationship, even though the traditional prescriptive grammars of Icelandic treat these four sounds as being in a contrastive relationship (Stefán Einarsson 1945; Kristján Árnason 2011), which would help explain the problem with the secondary question of why the prescriptive eradication effort was successful. This also may point to the reason for the emergence of *flámæli* in the first place.

3.2 Data

The corpus used in this research is the Málrómur corpus (Jón Guðnason et al. 2012; Steinþór Steingrímsson et al. 2017), developed jointly by Reykjavík University and the Icelandic Centre for Language Technology. The data for the corpus were collected between 2011 and 2012 as part of Google’s effort to collect voice samples for different languages in order to develop databases of spoken sentences to aid the development of ASR (automatic speech recognition).

In this particular work the sound samples were not used; instead, the text version of the corpus is utilized. The raw version of the data set contained 55,000 sentences, including seven types of utterances, described in Table 3.2.1 below:

Table 3.2.1. Summary of sources⁴ for the Málrómur corpus

Online newspaper text	50%
Rare triphones	10%
Street names	10%
Personal names	10%
Miscellaneous	10%
Names of countries and capitals	5%
URLs ⁵	5%

3.3. Corpus Processing

In order to bring the data to a format suitable for performing calculations described in greater detail in Chapter IV, the corpus was processed using a number of software products.

⁴ Samples of the corpus data before and after processing are shown in Appendix A.

⁵ During the processing of the data for use in calculations the URLs were omitted as non-words.

The original corpus data was presented in the form of sentences in a comma-separated values (CSV) file, which was imported to Microsoft Excel. There the corpus was brought to a uniform standard: all of the sentences were split into 216,680 separate words using regular expressions.

Similarly to Renwick (2014), the data set for this thesis employs lexemes, rather than lemmas: a *lexeme* is a particular phonological word form (Levelt 1989), while a *lemma* is a semantically and syntactically defined lexical entry. Since the relationship between the two is not one to one (Jurafsky, Bell & Girand 2002), I have chosen to look at all possible word forms included in the corpus. The 216,680 tokens therefore were reduced to 28,346 unique tokens, one word per cell in a single column.

From there, the CSV file was imported into RStudio, a software bundle designed for working with the R programming language (R Core Team 2016). In RStudio the file was converted into a more machine-readable form: non-words were removed, columns renamed, encoding issues rectified. Corpus-internal word frequencies were also calculated and added to the data set, using the `plyr` package (Wickham 2011).

In order to calculate the necessary quantitative data such as functional load, all of the words in the corpus were converted into the International Phonetic Alphabet by means of a Perl script⁶. The transcription conventions are based on the works of Kristján Árnason (2011) and Stefán Einarsson (1945).

Due to time constraints, suprasegmental features such as quantity and stress were not included in the transcription. Since stress in Icelandic is rather predictable (primary stress on the

⁶ For the Perl script refer to Appendix C.

first syllable, secondary stress on every other syllable after that), the data obtained without its inclusion should still be representative of the language.

(1) Stress positioning

Aðalsteinn [ˈa.ðal.ˌsteiɲŋ] “Adalsteinn (personal name)”

3.4 Calculations

The quantitative methods employed in this thesis are mainly based on the methods used in Renwick et al. (2016), a discussion of the marginal contrasts between the vowels [ʌ] and [i] in Romanian.

The following calculations, discussed in detail below, were carried out on the main data set:

- Vowel frequency distribution
- Functional load
- Kullback-Leibler Divergence
- Predictability of Distribution

Most of these measures were obtained using Phonological CorpusTools (Hall et al. 2016), a computer program designed for performing calculations used in quantitative phonology on transcribed corpora.

The **vowel frequency distribution** was calculated using RStudio; the source code is provided in the Appendix B. The calculation itself is a simple count of vowel phonemes across the entire corpus with regard to token frequency, rather than type frequency. This is due to the fact that in Phonological CorpusTools every item in a corpus is assumed to represent one type. The difference between type and token frequencies here is that a *type* represents an abstract

category defined for the purpose of the analysis; within a type a *token* is a specific countable occurrence of that type (Renwick 2014).

The **functional load** of vowels was calculated using Phonological CorpusTools. Both minimal pair count-based functional load and change in entropy-based functional load values were obtained. The minimal pair count calculation produced the result relative to the entire corpus, which allows for a direct comparison between functional loads of different pairs in the same corpus. The calculation of functional load of phonemes x and y through relative minimal pair count is a simple count of minimal pairs $M_{(x,y)}$ divided by the length (number of words) of the corpus C :

Equation 3.4.1. Functional load through minimal pair count vs. corpus length

$$FL_{(x,y)} = \frac{M_{(x,y)}}{C}$$

The second method of functional load calculation, the change in entropy method as described in Surendran and Niyogi (2003), is calculated according to Equation 3.4.2, where for every symbol i in an inventory N (in this case, every phoneme in the phoneme inventory), the probability of i is multiplied by the binary logarithm of the probability of i . The entropy H is the sum of the products for all symbols in the inventory.

Equation 3.4.2. Calculation of entropy

$$H = - \sum_{i \in N} p_i * \log_2(p_i)$$

The functional load of a pair of sounds in a system is calculated by first determining the entropy of the system (corpus) with all sounds included, i.e. before the elimination of contrast, and then finding the entropy of the system after the elimination of contrast. The functional load is therefore the amount of entropy lost due to the elimination of contrast, that is, the higher the

number of confusions that will occur due to the merger, the higher the difference between entropy before the merger (H_1) and the entropy after the merger (H_2). In mathematical terms functional load is merely a difference ΔH between the two entropy values, as illustrated in Equation 3.4.3:

*Equation 3.4.3. Change in entropy*⁷

$$\Delta H = H_1 - H_2$$

In order to provide some context for the functional load scores of the *flámæli* vowels across the vowel system of Icelandic, the Kullback-Leibler Divergence (KLD) and the Predictability of Distribution (POD) measures are used in this thesis. While the functional load may help determine whether a contrast is likely to be eliminated, the KLD and POD measures help determine whether two sounds are in allophonic or contrastive relationship.

The measure of dissimilarity between probability distributions, or the **Kullback-Leibler Divergence** (Kullback & Leibler 1951), is also calculated using Phonological CorpusTools. The calculation is based on Peperkamp et al. (2006), a study on statistical learning algorithms applied to allophony, where the measure was included in order to test for robustness to statistical noise and non-systematic errors. Essentially the Kullback-Leibler Divergence measures the distribution for environments as a proxy for phonological relationships. The sounds that are in contrastive distribution appear in the same environments, producing minimal and near-minimal pairs. The sounds that are in complementary distribution never appear in the same environments. Therefore, the identical distribution would have a KLD score of zero, and the higher the probabilities that the two sounds are allophones, the higher the KLD score (Hall et al. 2016).

⁷ This method differs from that of Hockett (1966) in that Hockett's definition uses the whole language L instead of the corpus C ; the method proposed by Hockett treats language as a sequence of phonemes. This sequence is infinite, due to the nature of language, and thus impossible to use in practice.

The KLD scores are calculated as follows. The Kullback-Leibler measure of dissimilarity between the distributions of two segments is the sum for all contexts of the entropy of the contexts given the segments, as illustrated in the following equation:

Equation 3.4.4. The Kullback-Leibler Divergence

$$m_{KL}(s_1, s_2) = \sum_{c \in C} P(c|s_1) \log\left(\frac{P(c|s_1)}{P(c|s_2)}\right) + P(c|s_2) \log\left(\frac{P(c|s_2)}{P(c|s_1)}\right)$$

In Equation 3.4.4 s is a segment, c is a context, and C is the set of all contexts. The probability of context c given segment s is calculated according to Equation 3.3.5, where $n(c|s)$ is the number of occurrences of s in context c ; it is necessary to mention that in these equations the context that is being looked at is the right-hand context. In this thesis I looked at both the left-hand side and the right-hand side contexts.

Equation 3.4.5. Probability of context given the segment

$$P(c|s) = \frac{n(c, s) + 1}{n(s) + N}$$

Lastly, the measure of **Predictability of distribution**, mentioned earlier in Chapter 2 and also calculated in Phonological CorpusTools, was used to help determine the phonological relationships between the *flámæli* vowels. Similarly to the measure of functional load, the POD involves the calculation of entropy; it is using the same equation, however, the two measures differ in their input. Two sounds can be perfectly contrastive and still have either low or high functional load score, depending on the use of the contrast in minimal pairs. Given the fact that the predictability of distribution is measured between two sounds, the i in Equation 3.4.6 when applied to POD would have two values, for each of the two sounds, in contrast with that of functional load calculation (Equation 3.4.2), where i has only one input.

Equation 3.4.6. Predictability of distribution

$$H = - \sum_{i \in N} p_i * \log_2(p_i)$$

Because of this limitation, the entropy value will range between 0 and 1, where an entropy of 0 indicates perfect predictability (allophonic relationship), and an entropy of 1 indicates complete uncertainty about which of the two sounds will occur, that is, a perfect contrast (Hall 2009; 2012; Hall et al. 2016).

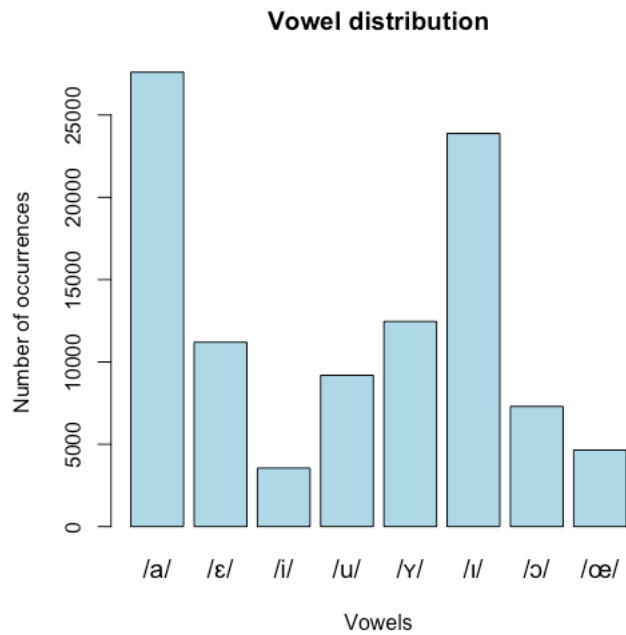
In order to produce the most accurate predictability of distribution results, environment exhaustivity and uniqueness were enforced during calculation in PCT. Exhaustivity in this case refers to the total list of environments encompassing all words in the corpus that contain either of the two sounds, without which the measure will be incomplete. Uniqueness is enforced to ensure that the set of environments do not overlap with each other. For this thesis, the POD values were calculated for all the environments in the corpus, by calculating the weighted average entropy across all environments where at least one of the two sounds in an opposition occurs (Hall et al. 2016).

CHAPTER 4

RESULTS

4.1. Frequency distribution of vowels

In order to determine the frequency distribution of vowels in the data set, an R script was used. The frequency distribution shows which vowels occur more frequently in the corpus, and therefore this information can be used to make assumptions about vowel distribution in the language. Figure 4.1.1 shows the distribution of eight vowel phonemes in the corpus, across the 28,346 unique words.



Phoneme	/a/	/ε/	/i/	/u/	/ʏ/	/ɪ/	/ɔ/	/œ/
Frequency	27600	11198	3554	9191	12463	23880	7298	4646
Percentage	28%	11%	4%	9%	12%	24%	7%	5%

Figure 4.1.1. Vowel distribution across the data set

As we can see from this bar plot, the *flámæli* vowels constitute a large portion of the vowel inventory in the corpus. /ɛ/ and /ɪ/ together constitute 35% of all 99,830 vowel tokens, /ʏ/ and /œ/ make up 17%. Together these four sounds make up 52% of all vowel tokens. Note that the diphthongs are not included in this chapter as they are treated as being composed of two monophthongs.

4.2. Functional load

As mentioned above, functional load was calculated using two methods: the minimal pair count and the change in entropy. Table 4.2.1 shows the scores for all pairs of vowels in the corpus using the minimal pair count, relative to the corpus (the *flámæli* vowel pairs' scores are shown in bold):

Table 4.2.1. Functional Load, minimal pair count⁸

	/ɛ/	/i/	/u/	/ʏ/	/ɪ/	/ɔ/	/œ/
/a/	21.8	3.9	29.3	21.2	51.2	31.4	8.1
/ɛ/		21.8	6.9	10.6	14.6	8.3	7.2
/i/			18.4	16.7	22.0	17.7	16.4
/u/				0.4	16.6	1.8	0.3
/ʏ/					19.3	1.2	0.9
/ɪ/						2.3	1.7
/ɔ/							1.1

As we can see from Table 4.2.1, the *flámæli* contrasts have quite different FL scores. This is surprising, as the prediction was that both of the contrasts would show (similarly) low scores, indicating higher potential for merger. The /ɛ/~ɪ/ opposition score is 14.6, while the /ʏ/~œ/ score is only 0.9. It should be added here that functional load values are not an absolute, but rather a relative measure: in order to gauge the scores of individual contrasts it is therefore

⁸ It is worth noting that for illustrative purposes all minimal pair count FL values were multiplied by 100 and rounded to one digit after the decimal point.

necessary to calculate FL scores for other vowel contrasts within a corpus to compare with those of interest to a particular study. The other vowel combinations in Table 4.2.1 show varying degrees of functional load, from 0.3 (/u/~/œ/) to 51.2 (/a/~/ɪ/). What these numbers indicate is that should, say, /a/ and /ɪ/ merge, a high percentage of minimal pairs would be homophonous, as opposed to the lowest value of 0.3 for /u/~/œ/, where the number of such collapsed minimal pairs would be much smaller.

The values from Table 4.2.1 are positively correlated to the combined vowel frequency for each pair. The Pearson's product-moment correlation test yielded the correlation coefficient of 0.5913, with a *p*-value of 0.0009188 (the null hypothesis is that the true correlation is equal to 0) (Baayen 2008). Note that even though the unit of measurement for the minimal pair count method is percent, the measure is intended to produce a relative output, rather than a straightforwardly interpretable value.

A more graphic look into the functional load distribution is given in Figure 4.2.1, which illustrates the scores of the *flámæli* vowels in comparison with all other vowel contrasts in the corpus:

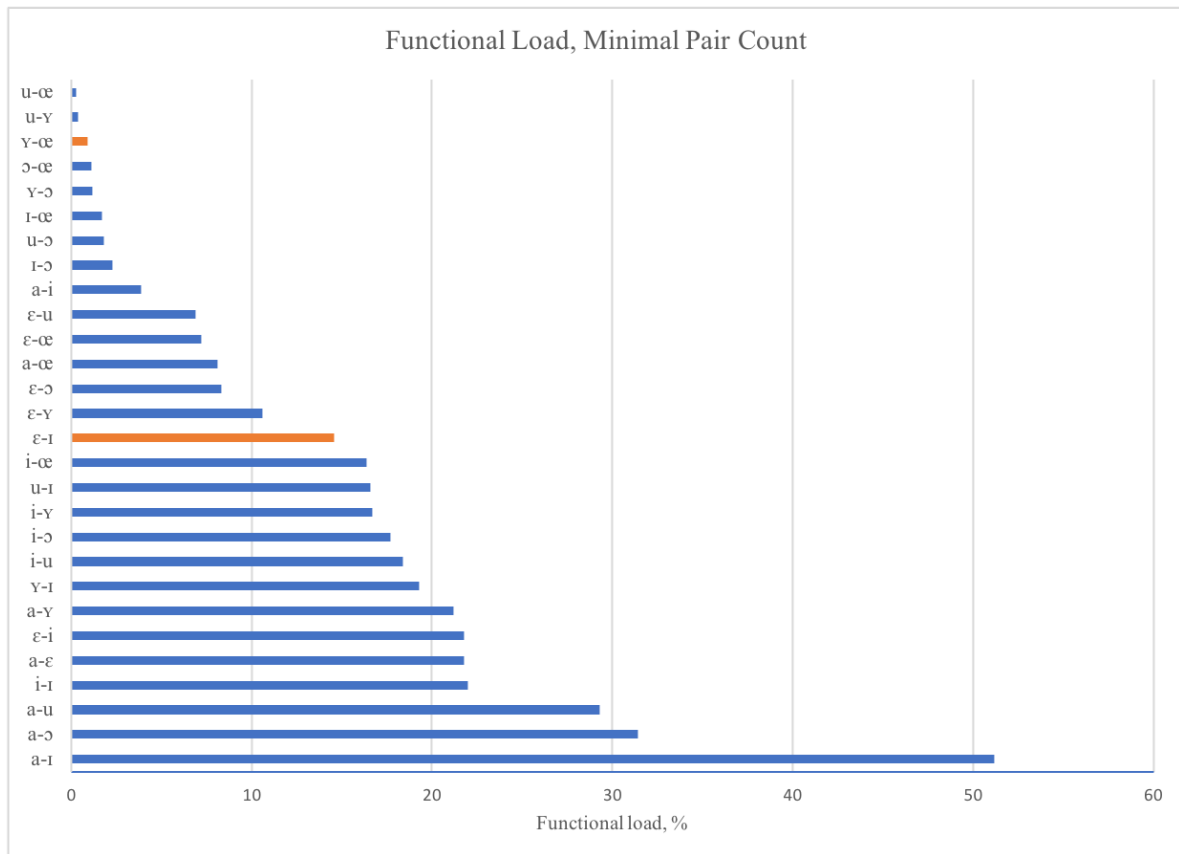


Figure 4.2.1. FL distribution, minimal pair count

Comparing the *flámæli* oppositions to all the others, the /ɛ/~i/ contrast FL score is quite high: it is higher than 54% of the other contrasts, and lower than 46%. The front rounded contrast's functional load is much smaller than the others: it is higher than just two of the other contrasts, and lower than 25.

The change in entropy method of functional load calculation produced the following scores:

Table 4.2.2. Functional Load, change in entropy⁹

	/ɛ/	/i/	/u/	/ʏ/	/ɪ/	/ɔ/	/œ/
/a/	21.2	4.1	3.2	32.1	62.6	11.4	6.0
/ɛ/		4.1	12.0	3.2	9.9	5.7	8.8
/i/			3.7	2.2	3.9	3.3	1.7
/u/				0.5	4.1	1.6	0.3
/ʏ/					27.6	1.3	1.1
/ɪ/						2.5	1.9
/ɔ/							1.6

Again, the scores are different for the two *flámæli* contrasts, which leads to partial rejection of the initial prediction: the unrounded /ɛ/~ɪ/ opposition produces the functional load of 9.9, whereas its rounded counterpart scores 1.1. Figure 4.2.2 provides a more illustrative look at the functional load distribution. Comparing to the other contrasts, the general tendency seems to hold: the /ɛ/~ɪ/ contrast is higher than 21 of all other oppositions, and consequently is lower than 6. The /ʏ/~œ/ contrast remains on the lower side, being higher than 2 of the other oppositions, and lower than 25.

Comparing the two measures of functional load, scores tend to vary for different contrasts, which is only to be expected as the calculation methods are quite different; the units of measurement for the two are also different: the output for the minimal pair count method is a percentage, while entropy from the second method is measured in bits, that is, basic units of information. When compared side by side the two measures show a moderate correlation of 0.7 (Pearson product-moment correlation coefficient; p-value of 0.00002681). This is illustrated in Figure 4.2.4. The regression line in the scatter plot indicates the general trend in the data (Levshina 2015). A combined graph showing both minimal pair count measure (in blue) and the change in entropy measure (in orange) is provided in Figure 4.2.3.

⁹ As the change in entropy method produces much smaller numbers (often multiples of 10^{-5}), for ease of reading these values have been multiplied by 10^4 and rounded to 1 digit after the decimal point.

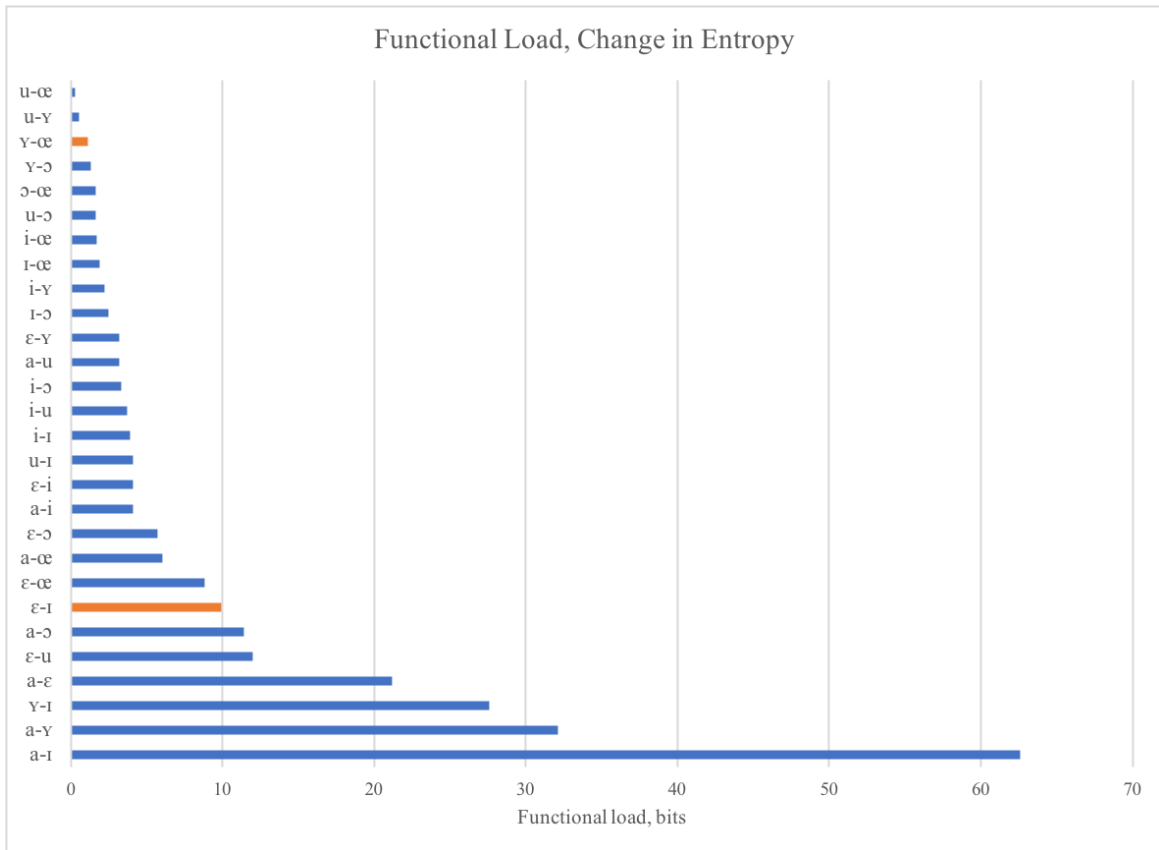


Figure 4.2.2. Functional load, Change in Entropy Distribution

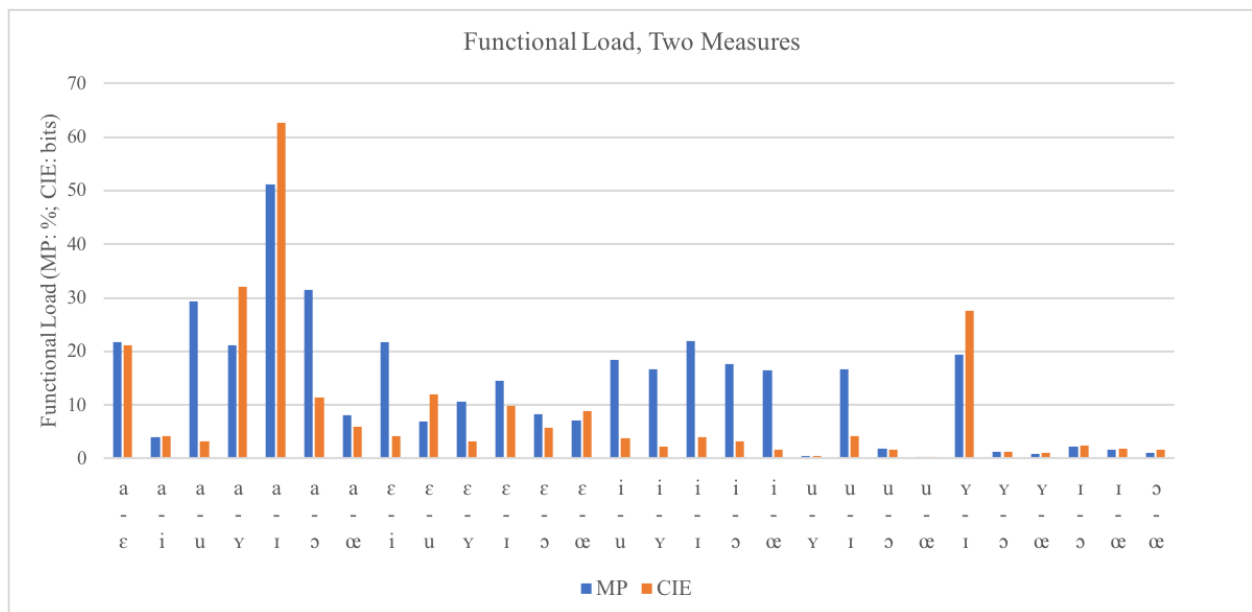


Figure 4.2.3. Functional load, two measures

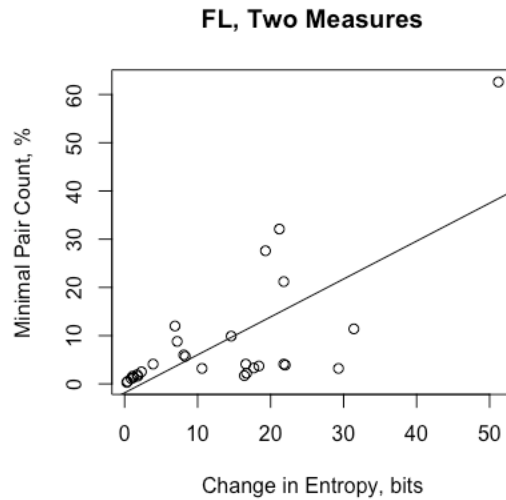


Figure 4.2.4. FL, Two measures, scatter plot

4.3 Kullback-Leibler Divergence

The Kullback-Leibler measure of dissimilarity between probability distributions was calculated for the *flámæli* vowel pairs, as well as for all other vowel contrasts in the corpus. This was done to provide room for comparison, since the KLD, similarly to functional load, is a relative rather than absolute measure. Tables 4.3.1 and 4.3.2 show, respectively, the KLD values for right-hand side and left-hand side environments. These are calculated in order to determine whether phonological conditioning by their environment is involved for the sounds in the contrast, on either side of the sound in question. As stated above, Kullback-Leibler Divergence has a lower bound of 0, indicating perfect contrast. The higher the value, the more the phoneme is affected by its environment.

Table 4.3.1. Kullback-Leibler Divergence, Right-hand side

	/ɛ/	/i/	/u/	/ʏ/	/ɪ/	/ɔ/	/œ/
/a/	0.55	1.33	7.12	0.62	2.17	0.61	0.56
/ɛ/		1.12	7.54	1.45	3.04	0.29	0.27
/i/			7.44	2.16	3.26	1.36	1.47
/u/				10.10	6.02	7.76	7.87
/ʏ/					1.69	1.44	1.26
/ɪ/						2.94	3.11
/ɔ/							0.29

Table 4.3.2. Kullback-Leibler Divergence, Left-hand side

	/ɛ/	/i/	/u/	/ʏ/	/ɪ/	/ɔ/	/œ/
/a/	2.43	2.44	1.84	2.52	2.17	1.51	2.55
/ɛ/		2.89	1.58	3.01	2.05	5.90	0.44
/i/			0.56	1.57	0.61	6.26	2.96
/u/				1.09	0.32	4.98	1.78
/ʏ/					0.87	5.73	3.69
/ɪ/						6.21	2.28
/ɔ/							5.62

The *flámæli* vowel pair KLD scores for both right- and left-hand environment show values that are not too high or too low. For the right-hand side, the score for the /ɛ/~ɪ/ contrast is 3.04, which is higher than 64% of all other contrasts; the score for the /ʏ/~œ/ contrast is 1.26, which is only higher than 26% of the other contrasts, therefore indicating that it is less potentially affected by its environment. As mentioned earlier, this contrast also has lower functional load values, which suggests that the phonological relationship here cannot be stated in terms of either allophony or contrast. This will be discussed in detail in Chapter 5. The *flámæli* vowel pairs' positions among the 28 total vowel contrasts are therefore the 10th for the unrounded opposition, and the 20th for the rounded one.

The left-hand side pattern is similar for the *flámæli* vowels, although closer to the higher bound¹⁰: the unrounded opposition's score is 2.05, being the 17th highest (higher than 40% of the other contrasts); the rounded contrast scored 3.69, being the 7th highest (higher than 75% of the others). Note that the left-hand side pattern is opposite to that of the right-hand side, where the unrounded contrast held a higher KLD score than the rounded one. In order to determine precise reasons for specific KLD scores, a more detailed look into co-occurrence of the *flámæli* vowel pairs with other vowels/consonants is necessary, as is done for Romanian /i ʌ/ in Renwick (2014). This should be addressed in further studies on the matter.

4.4. Predictability of Distribution

The Predictability of Distribution scores, calculated for all vowel contrasts in the corpus, are presented in the Table 4.4.1. The cells shaded dark grey indicate the highest and lowest values (/ɛ/~/ʏ/ as the highest and /a/~/œ/ as the lowest).

*Table 4.4.1. Predictability of Distribution (Entropy)*¹¹

	/ɛ/	/i/	/u/	/ʏ/	/ʊ/	/ɔ/	/œ/
/a/	0.906	0.593	0.823	0.874	0.994	0.720	0.521
/ɛ/		0.828	0.983	0.997	0.944	0.929	0.758
/i/			0.910	0.864	0.650	0.972	0.990
/u/				0.994	0.873	0.980	0.851
/ʏ/					0.917	0.954	0.798
/ʊ/						0.777	0.577
/ɔ/							0.932

It is evident that the Predictability of Distribution is spread rather uniformly across the vowel space in the corpus. As mentioned previously, the limits of relative entropy are {0,1}, where 0 indicates complete predictability (allophonic relationship), and 1 indicates perfect

¹⁰ By “bound” here I mean the highest value out of the 28 calculated values. The KLD is a relative, rather than absolute, measure, and therefore doesn't have a strictly defined upper bound (Hall et al. 2016).

¹¹ The entropy scores here are presented as is, with the exception of rounding to three digits after the decimal sign.

contrast (completely unpredictable distribution). In the case of /ε/~/ʏ/ we have an almost perfect contrast (which is consistent with the KLD results indicating that there is no major phonological conditioning involved), while on the other end is the /a/~/œ/, with the score almost half of the highest value. The *flámæli* vowel pairs are situated closer to the upper bound, especially the unrounded /ε/~/ɪ/ opposition with the score of 0.994. Its rounded counterpart has a lower POD value of 0.798, which is consistent with the other calculations done on this data set.

CHAPTER 5

DISCUSSION

5.1. Flámæli and Functional Load

Based on the quantitative evidence presented in the previous chapter, it is now necessary to relate the data to the original assumptions and predictions.

Before proceeding to the discussion of the functional load scores for the *flámæli* vowels it is worth pointing out that their frequency distribution across the corpus (see Figure 4.1.1) gives some idea about the amount of “work” that they do. The *flámæli* vowels comprise 52% of all the vowels in the corpus; the larger part of this 52% is taken up by the unrounded pair (67%), and the smaller part (33%) by the rounded pair. In fact, /ɪ/ turns out to be one of the most frequent vowels in the corpus at 23,880 occurrences, second only to /a/ with its 27,600 occurrences.

Such distribution is not entirely surprising: it is well-known that /ɪ/ and /ʏ/ occur in a multitude of case endings (Thomson 1987). In fact, in my opinion this was one of the factors that led to the ostracizing of *flámæli* as a regional variety: had a full merger occurred, the change would have been highly salient to native speakers of Icelandic who do not exhibit *flámæli*.

The data obtained by calculating the functional load of the contrasts is best viewed in comparison to the other vowel oppositions in the corpus (see Table 4.2.1).

The minimal pair count method yielded a rather surprising result for the unrounded pair, surprising in a way that it contradicts the initial prediction that the *flámæli* vowel pairs would exhibit lower functional load than the other contrasts, indicating their propensity for mergers. The quantitative evidence, however, shows a rather high FL score of 14.6 for the /ɛ/~ɪ/

opposition, a score that is higher than 54% of all other FL scores (see Figure 4.2.1). For the rounded pair, the score is more in line with the original prediction: it is on the lower side, larger than only two other contrasts and smaller than 89% of the others.

These scores suggest a higher propensity for merger in the rounded *flámæli* vowel pair, rather than the unrounded one. This was, however, in a way predicted by the vowel distribution across the corpus: the high percentage of /ɛ/ and /ɪ/ suggests a higher number of minimal pairs, and vice versa for the smaller percentage of /ʏ/ and /œ/.

In order to confirm or reject the tendency indicated by the minimal pair count method of functional load calculation, I used the change in entropy method. The scores for the *flámæli* vowels fall in the same pattern as in the case with the first method of FL calculation. The unrounded pair's score is 9.9, which is higher than 75% of all other contrasts, and the rounded pair scored 1.1, which is definitely on the lower side, being lower than 89% of the other contrasts.

This confirms the initial assumption based on just the minimal pair count method: the rounded *flámæli* contrast's functional load is low in comparison to all other vowel oppositions, while the unrounded one is rather high, on par with contrasts like /ɛ/~u/ and /a~/ɛ/.

5.1.1. Potential for future mergers

The evidence from functional load suggests that the unrounded opposition is much more stable and will not be eliminated in the present course of events. The same, however, cannot be said about the rounded opposition: low functional load scores suggest that the contrast is rather weak. This evidence provides a certain degree of explanation the resurgence of that particular variety of the *flámæli* phenomenon in teenagers in the 1980s (Höskuldur Þráinsson & Kristján Árnason 1984).

The functional load values presented above indicate that the stability of the *flámæli* contrasts is different: the /ɛ/~/ɪ/ contrast is more stable, with higher than average functional load values, while its rounded counterpart, the /ʏ/~/œ/ contrast, is much less stable and is more likely to be eliminated in the future.

As discussed in Chapter 2, the main factor influencing *flámæli* is vowel length; vowel length in Icelandic is largely determined by primary and secondary stress. Since due to technical and time limitations the data set for this thesis did not include such information, this would be a logical step for further research. Including vowel and consonant length, aspiration, stress and syllable information in the data set would greatly improve it and provide for much more informative quantitative measures. In this thesis the short and the long vowel systems were collapsed due to time constraints, but I believe that the results obtained using the present data set are still representative of the language in general and could be used as a first step in this direction.

5.1.2. Would the mergers have led to confusion?

The effort undertaken by the Icelandic government in the 20th century indicates a serious concern that the purism movement had regarding *flámæli*, more than any other regional variation in the language. One of the research questions of this thesis was to determine whether (had it actually occurred) the merger would have resulted in confusion on the part of the speakers, in other words, whether the government's concerns were justified. Based on the data obtained through functional load calculation, it seems possible to suggest that the eradication can only be justified partly: regarding the unrounded variation, the findings of this thesis seem to provide evidence of small likelihood of contrast elimination as indicated by its relatively high functional load, while for the rounded pair, the likelihood is much higher. As mentioned above, the vowels

/ɪ/ and /ʏ/ are used frequently in case endings and other morphological contexts. Had a merger of /ɪ/~/ɛ/ and /ʏ/~/œ/ occurred, the intelligibility rate in those speakers who do not exhibit *flámæli* would perhaps have suffered a slight decrease, perhaps increasing the time needed to process a word whose pronunciation is qualitatively different from the ‘standard’ variety. However, while /ɪ/ and /ʏ/ are fairly frequent in my data set (24% and 12% of all other vowels, respectively), the other pair, /ɛ/ and /œ/, has smaller distributions (11% and 5% of all other vowels), and is virtually absent from inflections, so the morphological confusion would probably not have been too severe.

5.2. Phonological Status of *Flámæli* Contrasts

While in Iceland the *flámæli* variation was successfully eradicated, it was allowed to develop unchecked in North America. According to studies conducted on North American Icelandic, the merger did not fully occur there, either, instead becoming an apparent merger, where the sounds in the contrast did not merge but rather change into intermediate qualities. In order to determine the status of this phonological relationship I used the two quantitative measures: the Kullback-Leibler Divergence and the Predictability of Distribution.

According to Peperkamp et al. (2006), the contrasts with higher KLD values are considered to be allophonic, while the closer to 0, the more contrastive an opposition will be. The values for the *flámæli* vowels were compared to all other vowel contrasts in the corpus (Tables 4.3.1. and 4.3.2). The left-hand side environment value for the /ʏ/~/œ/ contrast is 3.6, higher than 75% of other contrasts, including the unrounded one. The right-hand side value for the same contrast is lower, at 1.6 being higher than only 26% of other contrasts. Very similar, but opposite picture is with the /ɪ/~/ɛ/ contrast: the left-hand side value is 2.05, higher than 40% of

other contrasts; the right-hand side value is 3.09, higher than 64% of other contrasts, including the rounded one.

Such variation in KLD distribution indicates that neither of the contrasts involved is phonologically conditioned by their environment: the unrounded pair appears to be more affected by the following sound, while the rounded pair is more conditioned by the preceding sound. This points to a more allophonic rather than contrastive relationship between the *flámæli* vowels, which, in conjunction with low functional load of the rounded pair, suggests a higher chance for merger.

However, it should be borne in mind that KLD calculations performed on a corpus that is only transcribed phonemically are less informative than they would have been if a phonetic transcription was present as well. This, therefore, should be a direction for further study.

The difficulty of assigning one or the other type of phonological relationship here might be partly explained by the problem with the concept of predictability of distribution (Hall 2013). One of the most common reasons for the inability to determine the precise phonological relationship is a situation where a contrast is either mostly predictable but has a few contrasts, or vice versa: a contrast is mostly unpredictable, but has some degree of predictability. For this exact reason a calculation of the predictability of distribution was included in this thesis.

As evident from Table 4.4.1, all of the contrasts' values are between 0.52 and 0.99; both the *flámæli* vowel pairs' values are higher than the mean value of 0.85 (see Figure 4.4.1), indicating higher possibility of contrast. However, the entropy for the /ɣ/~ /œ/ contrast is much lower than that of its unrounded counterpart. This might be an indication that what we are dealing with here is an intermediate phonological relationship, more precisely, the kind where the environments are mostly unpredictable, but have some degree of predictability: in the *flámæli*

situation, this degree of predictability is obviously lower in the rounded contrast as opposed to the unrounded one.

Given the low minimal pair count functional load score for the /y/~œ/ contrast, low frequency of the contrasts' vowels across the corpus, and the low number of minimal pairs (24, to be precise, as determined by PCT as a side product of the Minimal Pair Count algorithm), it is possible that in this case the phonological relationship is also intermediate due to low frequency, as previously mentioned in Chapter 2.

The aforementioned results thus complement Birna Arnbjörnsdóttir's hypothesis that the case of *flámæli* in North American Icelandic is not a true merger, but rather an apparent merger, or merger by drift/approximation (in Labov's terms). It seems to be similar to a case study described by Hualde (2005) (cited in Hall (2013)), where the relationship between the trill [r] and the flap [ɾ] in Spanish is discussed as an example of a "quasi-phonemic" relationship. In that study the two sounds are found to be contrastive because of the present of minimal pairs where [r] and [ɾ] contrast intervocalically, but that the contrast is neutralized everywhere else. The rounded *flámæli* pair seems to be in a similar situation: mostly unpredictable, but with some degree of predictability.

Finally, it is possible that in the case of North American Icelandic the status of the *flámæli* relationship was influenced by contact with Canadian English and American English, therefore falling into the category of an intermediate phonological relationship conditioned by a foreign stratum in the language, that is, by lexical borrowings that also import some of the source language phonological principles.

5.3. Directions for Further Study

In order to determine whether the *flámæli* mergers are likely to occur in the future, a further study involving acoustic similarity may be useful. Audio recordings of the Málrómur corpus used in this study are available and through the use of software such as Praat, Phonological CorpusTools and RStudio it will be possible to make more accurate assumptions based on physical evidence.

As a direction for future work on determining the status of *flámæli* vowel pairs' phonological relationship, I suggest expanding the corpus to reflect not only phonemic representation, but also the phonetic representation, including suprasegmental features such as length, stress, syllable structure, among others. Having these levels of representation in the corpus will greatly enhance the accuracy of calculations, for instance, the Kullback-Leibler Divergence, which at the phonemic level only gives only a general idea of the tendencies of dissimilarity between probability distributions.

While corpora of Modern Icelandic do exist and are publicly available, the same cannot be said about the North American Icelandic. Due to the nature of NAI as an immigrant language, it is only a matter of time before it becomes extinct. It is therefore of utmost importance to compile a written and spoken corpus of North American Icelandic. This would be highly beneficial for the study of immigrant languages in general and Icelandic in particular. Having physical samples of North American Icelandic speech would also give an opportunity to conduct a comparative acoustical phonetic study of NAI and the varieties of English spoken in historically Icelandic communities. From a language contact point of view, this would give a valuable insight into the influence of the dominant language, that is English, on the subordinate language, which might in turn illustrate what might happen to Icelandic in particular and

European languages in general under constant pressure from English as a global language. While linguistic purism in Iceland is still strong, English definitely seems to be gaining more positions every year with increasing globalization. Using the example of North American Icelandic might prove a valuable insight into the future of Modern Icelandic and other European languages with a relatively small number of speakers.

CONCLUSION

In this thesis I investigated the phonological stability and the nature of phonological relationship between the so-called *flámæli* vowels, namely the two vowel pairs, /ɪ/~ /ɛ/ and /ʏ/~ /œ/, which were believed to have merged in two varieties of Icelandic: Modern Icelandic, as spoken in Iceland, and North American Icelandic, as spoken in the Icelandic immigrant communities in Canada and the United States. The term *flámæli*, meaning ‘skewed speech’, is used here to illustrate the stigma associated with this regional variety, characteristic of speakers from Southwestern, Northwestern, and Eastern Iceland, during the peak of the spread of the phenomenon in the middle of the 20th century.

The main research question was whether the apparent merger of the *flámæli* vowel pairs was influenced by the functional load of these contrasts, the functional load being the amount of work that a phonological contrast does in a linguistic system. The initial prediction that the contrasts in question would have low functional load values and thus would be more likely to merge proved to be only half accurate: the unrounded pair exhibited rather high functional load values, indicating its low propensity for elimination, whereas the rounded pair’s FL scores were relatively low, thus confirming the initial hypothesis, even if only by half.

Based on this data, I was able to only partially justify (on the basis of deliberate contrast maintenance and avoidance of homophony) the efforts of the Icelandic government in the beginning-middle of the twentieth century aimed at complete eradication of the phenomenon, which was viewed as uncouth and ‘diseased’.

Building on previous research on the matter I provided evidence that the phonological relationship between the rounded *flámæli* vowels is an intermediate one. Since the full merger never actually occurred, the general opinion was that the merger was not full but apparent, where a contrast between two sounds is not fully eliminated; the two sounds in these cases change into intermediate forms between the two qualities. Using quantitative methods such as the Kullback-Leibler measure of dissimilarity between probability distributions, and the Predictability of Distribution measure, I was able to determine that the amount of uncertainty between the vowels in the *flámæli* oppositions is rather high, indicating their contrastive status, but at the same time a certain degree of predictability exists, pointing at the potentially intermediate nature of these phonological contrasts.

I also suggested several directions for future study of *flámæli* in particular and North American Icelandic in general, such as compiling a written and speech corpus of the NAI, expanding the levels of representation for corpora to produce more accurate quantitative data, and investigating the influence of English phonology on North American Icelandic as a means to look in the future of Icelandic and European languages under the onslaught of English as rapidly spreading global language.

After investigating the relationships between *flámæli* vowel pairs and quantitative variables obtained through experimental work I came closer to better understanding the phenomenon in question, as well as the related concepts such as functional load and intermediate phonological relationships. The questions I have posed for this thesis have been answered to varying degrees, but there is still a lot to learn about apparent mergers. It is my hope that the directions for future studies on the matter that are outlined in this thesis will also be explored in time.

BIBLIOGRAPHY

- Baayen, R.H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge University Press.
- Birna Arnbjörnsdóttir. 1987. Flámæli í vesturíslensku. *Íslenskt Mál* 9. 23–40.
- Birna Arnbjörnsdóttir. 1990. North American Icelandic: The Linguistic and Social Context of Vowel Mergers. Austin, TX: University of Texas at Austin Dissertation.
- Birna Arnbjörnsdóttir. 1997. The life cycle of a language. *Canadian Ethnic Studies* 29(3). 31.
- Birna Arnbjörnsdóttir. 2006. *North American Icelandic: The Life of a Language*. Winnipeg, Manitoba R3T 2N2 Canada: University of Manitoba Press.
- Björn Guðfinnson. 1946. *Mállyzkur I*. Reykjavík: Ísafoldar-prentsmiðja.
- Björn Guðfinnsson. 1964. *Mállyzkur II*. Reykjavík: Heimspekideild Háskóla Íslands og Bókaútgafa Menningarsjóðs.
- Björn Guðfinnsson. 1981. Breytingar á Framburði og Stafsetningu. *Smárit Kennaraháskóla Íslands og Iðunnar*.
- Blöndal, Þórunn. 1985. Flámæli. Nokkrar athuganir á framburði Reykvíkinga fyrr og nú. Reykjavík: University of Iceland Unpublished BA-Thesis.
- Eyvindur Eiríksson. 1991. Some Remarks on Middle-English Influence on Icelandic. In Vladimir Ivir & Damir Kalogjera (eds.), *Languages in Contact and Contrast: Essays in Contact Linguistics*, vol. 54, 149–153. (Trends in Linguistics: Studies and Monographs). Berlin/New York: Mouton De Gruyter.

- Gísli Pálsson. 1989. Language and Society: The Ethnolinguistics of Icelanders. In Gísli Pálsson & Paul E. Durrenberger (eds.), *The Anthropology of Iceland*. Iowa City: University of Iowa Press.
- Gussmann, Edmund. 2001. Icelandic Vowel Quantity: Issues of Theory and Description. *Studia Neophilologica Posnaniensia* 3. 69–79.
- Gussmann, Edmund. 2011. Getting Your Head Around: The Vowel System of Modern Icelandic. *Folia Scandinavica* 12.
- Hall, Kathleen Currie. 2009. A Probabilistic Model of Phonological Relationships from Contrast to Allophony. Columbus, OH: The Ohio State University Dissertation.
- Hall, Kathleen Currie. 2012. Phonological relationships: A probabilistic model. *McGill Working Papers in Linguistics* 22.
- Hall, Kathleen Currie. 2013. A Typology of Intermediate Phonological Relationships. *The Linguistic Review* 30(2). 215–275. doi:10.1515/tlr-2013-0008.
- Hall, Kathleen Currie, Blake Allen, Michael Fry, Scott Mackie & Michael McAuliffe. 2016. *Phonological CorpusTools*. <http://phonologicalcorpustools.github.io/CorpusTools/>.
- Harris, J. 1985. *Phonological Variation and Change*. Cambridge: Cambridge University Press.
- Haugen, Einar. 1953. *The Norwegian Language in America*. . Vol. II. Philadelphia: University of Pennsylvania Press.
- Hockett, Charles F. 1955. *A Manual of Phonology*. Bloomington, IN: Indiana University.
- Hockett, Charles F. 1960. The Origin Of Speech. *Scientific American*.
- Hockett, Charles F. 1966. The Quantification of Functional Load - A Linguistic Problem. *Memorandum RM-5168-PR*. (U.S. Air Force Project Rand).
- Höskuldur Þráinsson & Kristján Árnason. 1984. Um Reykvísku. *Íslenskt Mál* 6. 113–134.

- Hualde, Jose Ignacio. 2005. Quasi-phonemic Contrasts in Spanish. In Vineeta Chand, Ann Kelleher, Angelo J. Rodriguez & Benjamin Schmeiser (eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics*, 374–398. Somerville, MA: Cascadia Press.
- Jakob Jóh. Smári. 1917. Um Framburð. *Landið*.
- Jón Guðnason, Oddur Kjartansson, Jökull Jóhannson, Elín Carstendóttir, Hannes Högni Vilhjálmsson, Hrafn Loftsson, Sigrún Helgadóttir, Kristinn M. Jóhannsdóttir & Eiríkur Rögnvaldsson. 2012. Almannarómur: An Open Icelandic Speech Corpus. *Proceedings of SLTU '12*. Cape Town, South Africa.
<http://www.malfong.is/index.php?lang=en&pg=malromur>.
- Jón Ófeigsson. 1980. Træk af moderne islandsk Lydlære [Features of Modern Icelandic Phonetics]. *Íslensk-Dönsk Orðabók*, XIV–XXVII. Reykjavík.
- Jurafsky, Daniel, Alan Bell & Cynthia Girand. 2002. The Role of the Lemma in Form Variation. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7*, vol. 1, 3–34. (Phonology and Phonetics 4). Berlin/New York: Mouton De Gruyter.
- King, Robert D. 1967. A Measure for Functional Load. *Studia Linguistica* 21(1). 1–14.
doi:10.1111/j.1467-9582.1967.tb00545.x.
- Kristján Árnason. 2011. *The Phonology of Icelandic and Faroese*. Oxford; New York: Oxford University Press.
- Kristjanson, Wilhelm. 1965. *The Icelandic People in Manitoba: A Manitoba Saga*. Winnipeg: R. W. Kristjanson.

- Kučera, Henry. 1963. Entropy, Redundancy, and Functional Load in Russian and Czech. *American Contributions to the Fifth International Congress of Slavists, Sofia 1: Linguistic Contributions*. 191–219.
- Kullback, S. & R.A. Leibler. 1951. On Information and Sufficiency. *Annals of Mathematical Statistics* 22(1). 79–86. doi:doi:10.1214/aoms/1177729694.
- Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford UK, Cambridge MA, USA: Blackwell.
- Labov, William, Sharon Ash & Charles Boberg. 2006. *Atlas of North American English: Phonetics, Phonology and Sound Change: a multimedia reference tool*. (Atlas of North American English). Berlin/New York: Mouton De Gruyter.
- Labov, William, Malcah Yaeger & Richard Steiner. 1972. *A Quantitative Study of Sound Change in Progress*. Philadelphia: US Regional Survey.
- Ladd, D. Robert. 2006. “Distinctive Phones” in Surface Representation. In Louis M. Goldstein, D. H. Whalen & Catherine T. Best (eds.), *Laboratory Phonology*, vol. 8, 3–26. Berlin: Mouton De Gruyter.
- Levelt, Willem J. M. 1989. *Speaking: From Intention to Articulation*. (Ed.) Aravind Joshi. (ACL-MIT Press Series in Natural Language Processing). Cambridge, MA / London, England: MIT Press.
- Levshina, Natalia. 2015. *How to do Linguistics with R*. (195). John Benjamins.
- Maddieson, Ian. 2013. Vowel Quality Inventories. In Matthew S. Dryer & Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/2> (20 February, 2018).
- Martinet, André. 1952. Function, structure, and sound change. *WORD* 8(1). 1–32.

- Martinet, André. 1955. *Économie des changements phonétiques. Traité de phonologie diachronique*. (Bibl. Romanica; X). Berne: Francke. <http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=lnp&AN=30-03635&site=eds-live> (1 January, 1955).
- Mathesius, Vilém. 1929. Funkční lingvistika [Functional Linguistics]. *Sborník přednášek pronesených na Prvém sjezdu československých profesorů filosofie, filologie a historie v Praze 3.-7. dubna 1929*. Prague.
- Moulton, W.G. 1947. Juncture in Modern Standard German. *Language* 23(3). 212–226. doi:<http://dx.doi.org/10.2307/409876>.
- Nadeu, Marianna & Margaret E.L. Renwick. 2016. Variation in the Lexical Distribution and Implementation of Phonetically Similar Phonemes in Catalan. *Journal of Phonetics* 58. 22–47. doi:<https://doi.org/10.1016/j.wocn.2016.05.003>.
- Peperkamp, Sharon, Rozenn Le Calvez, Jean-Pierre Nadal & Emmanuel Dupoux. 2006. The Acquisition of Allophonic Rules: Statistical Learning with Linguistic Constraints. *Cognition* 101. B31–B41.
- R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.r-project.org>.
- Renwick, Margaret E.L. 2014. *The Phonetics and Phonology of Contrast: The Case of the Romanian Vowel System*. . Vol. 19. (Phonology and Phonetics). Berlin/Boston: Mouton De Gruyter.
- Renwick, Margaret E.L. & D. Robert Ladd. 2016. Phonetic Distinctiveness vs. Lexical Contrastiveness in Non-Robust Phonemic Contrasts. *Laboratory Phonology: Journal of*

- the Association for Laboratory Phonology* 7(1). 19: 1-29.
doi:<https://doi.org/10.5334/labphon.17>.
- Renwick, Margaret E.L. & Marianna Nadeu. 2018. A Survey of Phonological Mid Vowel Intuitions in Central Catalan. *Language and Speech*. 1–41.
doi:<https://doi.org/10.1177/0023830917749275>.
- Renwick, Margaret E.L., Ioana Vasilescu, Camille Dutrey, Lori Lamel & Bianca Vieru. 2016. Marginal Contrast Among Romanian Vowels: Evidence from ASR and Functional Load. San Francisco, USA.
- Sandler, W., M. Aronoff, I. Meir & C. Padden. 2011. The Gradual Emergence of Phonological Form in a New Language. *Natural Language & Linguistic Theory* 29(2). 503–543.
- Shannon, Claude. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27.
- Simons, Gary F. & Charles D. Fennig. 2018. *Ethnologue: Languages of the World*. 21st ed. Dallas, Texas: SIL International. <http://www.ethnologue.com>.
- Stefán Einarsson. 1945. *Icelandic: Grammar, Texts, Glossary*. Baltimore: Johns Hopkins Press.
- Steinþór Steingrímsson, Jón Guðnason, Sigrún Helgadóttir & Eiríkur Rögnvaldsson. 2017. Málrómur: A Manually Verified Corpus of Recorded Icelandic Speech. *Proceedings of the 21st Nordic Conference of Computational Linguistics*. Gothenburg, Sweden: Linköping University Press.
- Surendran, Dinoj & Partha Niyogi. 2003. Measuring the Usefulness (Functional Load) of Phonological Contrasts. *Current Trends in the Theory of Linguistic Change. In commemoration of Eugenio Coseriu (1921-2002)*. Amsterdam & Philadelphia: Benjamins.

- Surendran, Dinoj & Partha Niyogi. 2006. Quantifying the Functional Load of Phonemic Oppositions, Distinctive Features, and Suprasegmentals. In O.N. Tomsen (ed.), *Competing Models of Linguistic Change: Evolution and beyond*, vol. 279, 43–58. (Current Issues in Linguistic Theory). Amsterdam: John Benjamins.
<https://benjamins.com/#catalog/books/cilt.279.05sur/details>.
- Thomson, Colin D. 1987. *Icelandic Inflections*. Hamburg: Helmut Buske Verlag.
- Trubetzkoy, N.S. 1939. *Grundzüge der Phonologie*.
- Trudgill, Peter. 1974. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Trudgill, Peter. 1983. *On Dialect*. Oxford: Basil Blackwell.
- Trudgill, Peter. 1985. New Dialect Formation and the Analysis of Colonial Dialects: the Case of Canadian Raising. In H. J. Warkentyne (ed.), *Papers from the 5th International Conference on Methods in Dialectology*, 35–45. Victoria: University of Victoria.
- Vikør, Lars S. 2010. Language purism in the Nordic countries. *International Journal of the Sociology of Language*(204). 9–30.
- Wang, William S.Y. 1967. The Measurement of Functional Load. *Phonetica* 16(1). 36–54.
- Watson, Janet C. E. 2002. *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.
- Wedel, Andrew, Scott Jackson & Abby Kaplan. 2013. Functional Load and the Lexicon: Evidence the Syntactic Category and Frequency Relationships in Minimal Lemma Pairs Predict the Loss of Phoneme Contrasts in Language Change. *Language and Speech* 56(3). 395–417. doi:10.1177/0023830913489096.

- Wedel, Andrew, Abby Kaplan & Scott Jackson. 2013. High Functional Load Inhibits Phonological Contrast Loss: A Corpus Study. *Cognition* 128. 179–186.
- Wickham, Hadley. 2011. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software* 40(1). 1–29.

APPENDICES

APPENDIX A

Corpus Samples

Sample of the corpus used in the study before processing: the file contained 11 columns, with the audio file name, environment, speaker's age group, the text read, audio file duration, type, and sample rate.

Name	Environment		Gender	Age	Text read	Duration	File type	Sample rate	Channels	Encoding
is_is-100-2011-11-25T13:23:17.452039.ogg	Indoors - Quiet	1	male	30_39	Tæplega yrði einangrun okkur til framdráttar	2.700188	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:23:21.824056.ogg	Indoors - Quiet	0.6	male	30_39	Gefjun	0.726313	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:23:24.170131.ogg	Indoors - Quiet	1	male	30_39	Lyngdal	0.924375	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:23:26.837428.ogg	Indoors - Quiet	1	male	30_39	Síðar í kvöld versnar veður Norðanlands	2.997125	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:23:29.396803.ogg	Indoors - Quiet	1	male	30_39	Kyrrsetning eigna verði felld úr gildi	2.562	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:23:37.208621.ogg	Indoors - Quiet	1	male	30_39	Einnig tóku þeir tölvuskjá	1.752	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:23:39.628718.ogg	Indoors - Quiet	1	male	30_39	Spáir lækkun verðbólgu	1.5965	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:23:41.900632.ogg	Indoors - Quiet	1	male	30_39	Miklaholti	1.067438	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:23:47.281334.ogg	Indoors - Quiet	0.7	male	30_39	Landsvirkjun fær leyfi til rannsókna	2.064625	vorbis	16000	1	Vorbis

is_is-100-2011-11-25T13:23:55.547178.ogg	Indoors - Quiet	1	male	30_39	Auglýsingin varð fræg á Íslandi af því Landsbankinn reyndi sömu aðferð	4.12175	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:23:58.124576.ogg	Indoors - Quiet	1	male	30_39	Vestlægari og el seinni partinn	2.326188	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:03.141063.ogg	Indoors - Quiet	1	male	30_39	Stúlkan kinkaði kolli og litla fólkið í stúdíóinu klappaði fyrir Sveinbirni	4.851063	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:07.673918.ogg	Indoors - Quiet	1	male	30_39	Við það sljákkaði í honum	1.638313	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:10.089070.ogg	Indoors - Quiet	1	male	30_39	Stormur á storm ofan	1.6455	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:12.428310.ogg	Indoors - Quiet	1	male	30_39	Málið hefur vakið mikinn óhug	2.015375	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:15.115902.ogg	Indoors - Quiet	1	male	30_39	Tungusiðu	0.923813	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:17.457427.ogg	Indoors - Quiet	1	male	30_39	Seljavöllum	1.132375	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:24.511758.ogg	Indoors - Quiet	1	male	30_39	Skerið var ekki merkt á sjókort	1.906125	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:26.781616.ogg	Indoors - Quiet	1	male	30_39	Sterkur straumur bar hann frá bátnum	2.307812	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:31.968825.ogg	Indoors - Quiet	1	male	30_39	Laugagerði	1.017437	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:34.191442.ogg	Indoors - Quiet	1	male	30_39	Mikil velta á gjaldeyrismarkaði	1.948875	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:44.288274.ogg	Indoors - Quiet	1	male	30_39	Barnið fæddist andvana	1.553937	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:46.784431.ogg	Indoors - Quiet	1	male	30_39	Ekkert kerfi stæðist hrún	1.824563	vorbis	16000	1	Vorbis
is_is-100-2011-11-25T13:24:49.575914.ogg	Indoors - Quiet	1	male	30_39	Skoða skatt á skuldir fjármálafyrirtækja	2.630125	vorbis	16000	1	Vorbis

All columns but the 'text read' were removed, leaving separate words and sentences. The sentences were separated into words, one per row; after that all the text in the corpus was brought to lower case (the R programming language is case-sensitive), word frequencies were calculated, and duplicates removed:

Corpus after the initial processing, sorted by word frequency:

orthography	frequency
í	8641
á	6442
og	4213
að	3895
er	2399
við	2163
ekki	2125
var	1978
um	1935
hann	1743
með	1712
til	1465
það	1392
fyrir	1389
af	1268
þetta	1213
ég	846
en	811
úr	755
hún	737
eru	726
eftir	699
upp	620

sem	617
hefur	564
frá	547
vegna	543
hafa	526
sér	519
mér	511
voru	492
hjá	485
yfir	452
enn	430
fram	407

After that, transcriptions were added, using a Perl script, provided in Appendix C:

orthography	frequency	transcription
í	8641	i
á	6442	au
og	4213	ɔx
að	3895	aθ
er	2399	ɛr
við	2163	viθ
ekki	2125	ɛhki
var	1978	var
um	1935	ym
hann	1743	hann
með	1712	mɛθ
til	1465	tɪl
það	1392	θaθ
fyrir	1389	fɪrɪr

af	1268	af
þetta	1213	þehta
ég	846	jex
en	811	en
úr	755	ur
hún	737	hun
eru	726	ery
eftir	699	eftir
upp	620	yhp
sem	617	sem
hefur	564	hevyr
frá	547	frau
vegna	543	vekna
hafa	526	hava
sér	519	sjer
mér	511	mjer
voru	492	vory
hjá	485	hjau
yfir	452	ivir
enn	430	enn
fram	407	fram

APPENDIX B

R Script

R source code for calculating vowel distributions in the data set

```
# Vowel frequency distribution calculator
# vowel-freq-distr.R by Alexander Ankirskiy

setwd("/Users/alex/Dropbox/UGA/MA Thesis/Thesis data/Final dataset")

library(stringr)
library(ggplot2)

# loading the dataset into R
corpus <- read.csv("dataset-full-revised.csv")

# converting the dataset into a data frame
corpus <- data.frame(corpus)

# isolating transcriptions for further analysis
corptrans <- data.frame(corpus$transcription)

# counting instances of vowel occurrence in all words in the corpus
corptrans$a <- str_count(corptrans$corpus.transcription, "a")
corptrans$e <- str_count(corptrans$corpus.transcription, "e")
corptrans$i <- str_count(corptrans$corpus.transcription, "i")
corptrans$u <- str_count(corptrans$corpus.transcription, "u")
corptrans$y <- str_count(corptrans$corpus.transcription, "y")
corptrans$r <- str_count(corptrans$corpus.transcription, "r")
```

```

corptrans$ɔ <- str_count(corptrans$corpus.transcription, "ɔ")
corptrans$æ <- str_count(corptrans$corpus.transcription, "æ")

# calculating vowel distributions from the data set

a.sum <- sum(corptrans$a, na.rm = TRUE)
e.sum <- sum(corptrans$ε, na.rm = TRUE)
i.sum <- sum(corptrans$i, na.rm = TRUE)
u.sum <- sum(corptrans$u, na.rm = TRUE)
y.sum <- sum(corptrans$ʏ, na.rm = TRUE)
i2.sum <- sum(corptrans$ɪ, na.rm = TRUE)
o.sum <- sum(corptrans$ɔ, na.rm = TRUE)
oe.sum <- sum(corptrans$œ, na.rm = TRUE)

# combining vowel distribution into new data frame
vowels <- data.frame(a.sum, e.sum, i.sum, u.sum, y.sum, i2.sum, o.sum,
oe.sum)

# assigning column and row names for ease of reading
colnames(vowels) <- c("[a]", "[ε]", "[i]", "[u]", "[ʏ]", "[ɪ]", "[ɔ]", "[œ]")
rownames(vowels) <- c("distribution")
vowels.matrix <- as.matrix(vowels)

# creating a bar plot of the vowel distribution
barplot(vowels.matrix, main = "Vowel distribution",
        col = "lightblue",
        cex.names = 1.2, xlab = "Vowel", ylab = "Frequency")

```

APPENDIX C

Perl Script

Perl script for adding transcriptions to the corpus

```
#!/usr/bin/perl

# corpus-transcription.pl by Alex Ankirskiy

use 5.18.0;
use warnings;
use strict;
use feature 'say';

open DATASET, "<dataset-test" or die "Cannot open file: $!";
open TRANSCR, ">transcriptions-test.txt";

while (<DATASET>) {

    for (<DATASET>) {

# DIPHTHONGS

        s/au/æi/gi;
        s/ei/ɛi/gi;
        s/ey/ei/gi;
        s/á/au/gi;
        s/ó/ou/gi;
```

s/æ/ai/gi;

MONOPHTHONGS

s/e/ε/gi;

s/i/ɪ/gi;

s/u/ʊ/gi;

s/y/ɪ/gi;

s/é/jε/gi;

s/í/i/gi;

s/ý/i/gi;

s/ö/œ/gi;

s/ú/u/gi;

CONSONANTS (ONE TO ONE)

s/p/θ/gi;

s/rl/rtl/gi;

s/rn/rdn/gi;

geminated voiceless stops are pre-aspirated

s/pp/hp/gi;

s/tt/ht/gi;

s/kk/hk/gi;

CONSONANT (COMPLEX)

/b/ -----

s/(`b)/p/gi; # word-initial /b/ is always voiceless

my \$dtsg = "d|t|s|g";

s/mb(\$dtsg)/m\$1/gi; # /b/ is deleted between /m/ and d, t, s, g

#s/b/p/gi; # in all other cases /b/ is voiceless

/d/ -----

my \$ln = "l|n";

my \$gnlks = "g|n|l|k|s";

s/(\$ln)d(\$gnlks)/\$1\$2/gi; #lost between l, n and g, l, n, k, s

/ð/ -----

s/ð\$/θ/gi; # voiceless word-finally

/f/ -----

my \$vowels = "a|e|i|o|u|á|é|ó|æ|ö|í|ý|ú";

my \$vocalic = "a|e|i|o|u|á|é|ó|æ|ö|í|ý|ú|r|l";

s/(\$vowels)f(\$vowels)/\$1v\$2/gi; #voiced intervocalically

s/(\$vocalic)f(\$vocalic)/\$1v\$2/gi;

```

s/fn/pn/gi;
s/fl/pl/gi;

my $nrst = "n|r|s|t";
s/lf($nrst)/lv$1/gi; #lost between /l/ and n, r, s, t

# /g/ -----

my $vowelsG = "e|i|í|y|ý|æ|ei|ey|j";
my $au = "a|u";
my $st = "s|t";

s/^g$vowelsG/k$1/gi;
s/g($au)$/ɣ($1)/gi; #lightly voiced velar fricative before a, u word-
finally
s/g($st)/x($1)/gi; #voiceless before s, t

# /l/ -----

my $vowelsrn = "a|e|i|o|u|á|é|ó|æ|ö|í|ý|ú";
my $consnorn = "b|p|t|d|k|g|s|m|j|h|f|v|ð|þ|x|z";

s/($vowelsrn)ll($vowelsrn)/dl/gi; # /dl/ between vowels or vowels and r,
n
s/ll($consnorn)/ll/gi; # voiceless half-long before consonants other than
r, n

# /n/ -----

```

```

my $vowelsnn = "á|é|í|ó|ú|ý|æ|au|ei|ey";
my $vowelsr = "a|e|i|o|u|á|é|ó|æ|ö|í|ý|ú|r";
my $rts = "r|t|s";
my $rsk = "r|sk";
my $fgs = "f|g|s";

s/($vowelsnn)nn($vowelsr)/($1)dn($2)/gi;
s/($rts)n($rts)/($1)($2)/gi;

# /p/ -----

my $stk = "s|t|k";

s/p($stk)/f($1)/gi;

# /x/ -----

s/x/xs/gi;

# /z/ -----

s/z/s/gi; # purely orthographic; Icelandic language does not officially
include z anymore

chomp; # remove the newline symbol from transcribed data

say TRANSCR; # write the transcriptions into new text file

```

```
}
```

```
}
```

```
close TRANSCR;
```

```
close DATASET;
```