Vowel System of North-Eastern Slovene

Diversity and perception of the vowels of Prekmurje Slovene

Katja Štrakl

MA thesis General Linguistics

Supervisor: dr. B. M. Van 't Veer
University of Amsterdam
June 2024

Abstract

This MA General Linguistics thesis explores the Prekmurje dialect of Slovene. Prekmurje, a dialect of the Pannonian dialectal group exhibits some key differences from the Contemporary Standard Slovene, apart from lexicon most notably in its phonology. While standard Slovene has eight vowel phonemes, the Prekmurje dialect includes an additional vowel, a fronted version of /u/. This thesis maps the vowels of both standard and Prekmurje Slovene using the Artur1.0 corpus of spoken Slovene, using Praat to annotate the recordings of elicited and spontaneous speech. The results show evidence of Prekmurje dialect containing a 9th vowel, namely ʉ. The values for back mid-vowels further suggest a possible future merger. Other vowels show no significant difference from the previously established vowel space of Standard Slovene. For the second part of the thesis, a perception experiment was conducted, where Slovenes were tested on their comprehension of the vowels from the first part. The results showed no significant difference in identifying standard Slovene vowels and Prekmurje vowels. Additional analyses were done using demographic data on participants' region of origin, which also showed no significant advantages for Prekmurje or other regions.

Contents

1 Introduction

While only having about two million speakers, Slovene, a southern Slavic language, contains a rich dialectal variety. This can be linked to geological barriers, such as dense forests and mountains on one hand, and the political history of the area of what is today Slovenia on the other (Smole, 2009). Presently, Slovenia's dialects can be split into seven major dialectal regions, largely based on the geographical region they belong to; Panonska/Pannonian (the far North-East, more often also referred to as "Prekmurje", although a later section addresses the fact those two are not the same and should not be used interchangeably without care), Štajerska (North-east), Koroška (North), Dolenjska (South-East), Primorska (South-West and far West), Rovtarska (West), and Gorenjska (North-West and Central Slovenia) (Lundberg, 2007), however, variation even within regions is also often present, as evident by a common Slovene saying: "Every village has its own dialect".

An overview of the geographical distribution of Slovene dialects is given in the map in Figure 1, where the Panonska dialectal region is marked in yellow in the top right corner. The map also indicates the overlap of the regions as well as smaller dialectal sub-groups by using different patterns with the same main color.
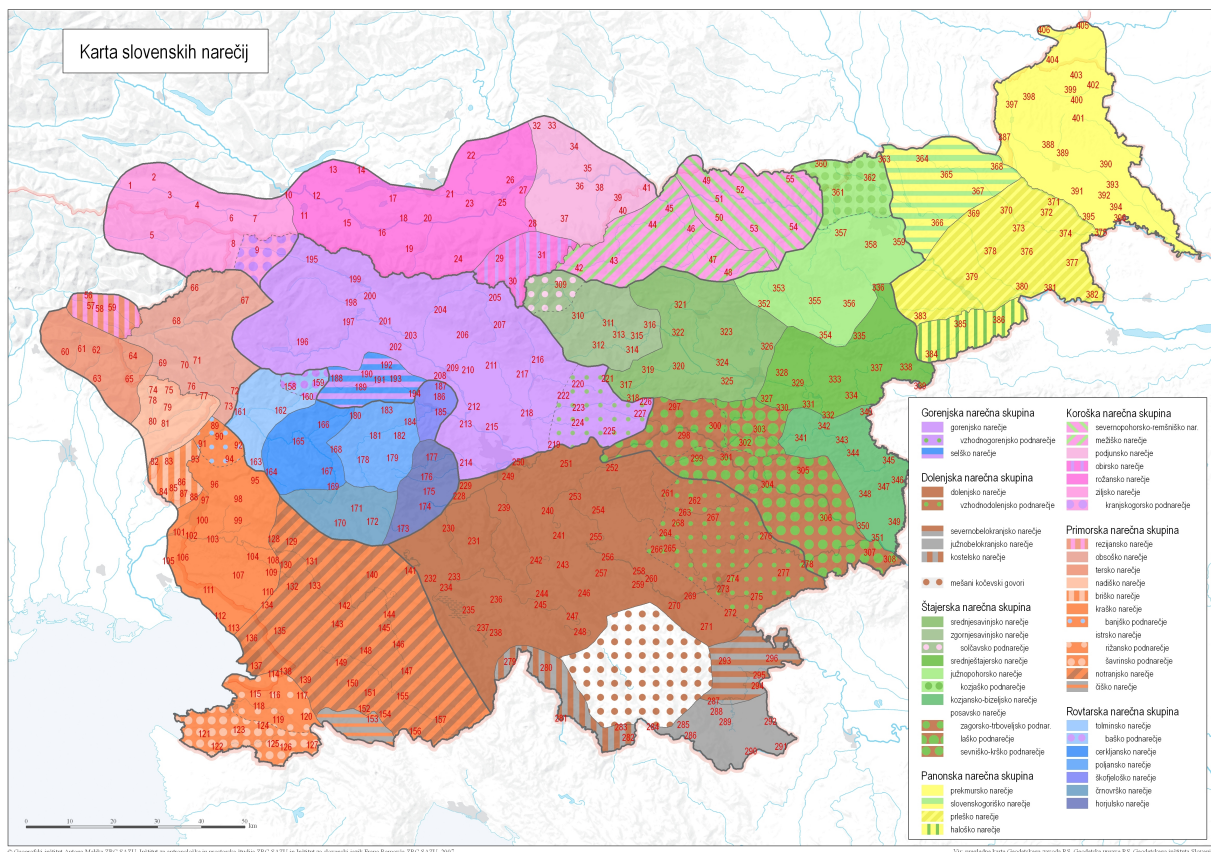

Figure 1: Map of Slovene dialects[1]

---

[1]Accessed at: http://bos.zrc-sazu.si/c/Dial/PonovneSLA/P/022OsnovnakartaJPG.jpg on 2024-06-16

2 Theoretical Background and Literature Review

2.1 Slovene

Slovene is a southern Slavic language, the official language of the Republic of Slovenia, and spoken by roughly two million people. Nested between four languages from four different language groups (Italian, German, Hungarian, and Croatian), it showcases a unique mix of linguistic features not commonly found in other Slavic languages.

Alongside Sorbian, a Western Slavic language, Slovene is one of the few (Indo-)European languages which have retained the dual number, once a feature of Proto-Indo-European. They are the only two Slavic languages to still distinguish three numbers (singular, dual, plural) (Gvozdanović, 2012). Similarly, they are the only two Slavic languages to count in the "ones-tens" format (two-and-twenty) instead of the more common "tens-ones" format (twenty-two). In both cases, it is likely that the physical proximity to German, a language which also employs this counting system, resulted in the deviation from the other Slavic languages. Lastly, the two languages are also the only ones to have kept the Slavic supine verbal form (Grošelj, 2023).

As described in the Slovene grammar by Toporišič (1976), Slovene has three noun genders (masculine, feminine, neuter) and three numbers; singular, dual, and plural. It further also has three grammatical tenses (future, present, past), with a fourth, perfect past nowadays being obsolete. Slovene also has a rich case system with six cases; nominative, genitive, dative, accusative, locative, and instrumental.

2.1.1 Slovene Phonology

The phonology of Slovene is relatively ordinary for a European setting. The standard Slovene vowel inventory contains eight vowels, namely /a, e, ɛ, i, o, ɔ, u/ and /ə/, while the standard consonant inventory has 21 consonant phonemes (Toporišič, 1976). While some sources claim standard Slovene has nine vowels, claiming /ʌ/ as the ninth (Jurgec, 2011), Kocjančič Antolík and Tivadar (2019) used ultrasound to show that speakers distinguish eight distinct tongue positions when articulating vowels, thus confirming the traditional distinction.

(Tivadar, 2004a)

Some allophonic variation is permitted in Slovene under certain conditions, for example [v]~[u] for word-final instances of /v/. The full consonant inventory can be seen in Table 1 and the distribution of the vowels can be seen in Figure 2. The values used were sourced from Petek, Šuštaršič, and Komar (1996).

Table 1: Consonant inventory of Slovene

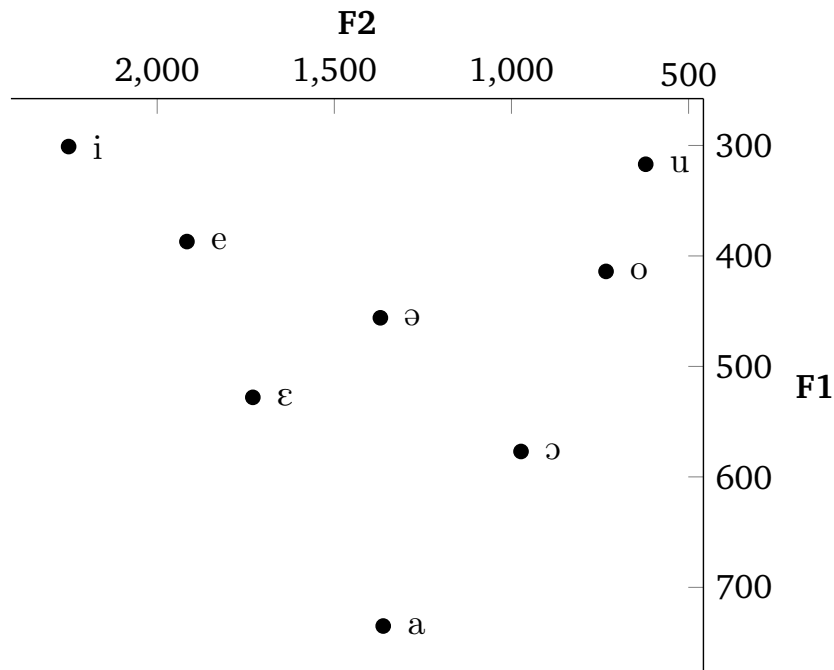| position of articulation	labial/labiodental	dental/alveolar	postalveolar/palatal	velar
voiceless stops	/p/	/t/		/k/
voiced stops	/b/	/d/		/g/
nasals	/m/	/n/		
voiceless fricatives	/f/	/s/	/ʃ/	/x/
voiced fricatives	/ʋ/	/z/	/ʒ/	
laterals		/l/	/ʎ/	
trills		/r/		
approximants			/j/	
voiceless affricates		/t͡s/	/t͡ʃ/	
voiced affricates			/d͡ʒ/	

Figure 2: Vowels of Standard Slovene

As mentioned, Slovene vowels can historically be divided into long and short, however, this distinction seems to be almost if not completely gone in present times. (Huber, 2006)

2.2 Contemporary Standard Slovene

When we speak of Slovene in a linguistic context, we refer to what is known as Contemporary Standard Slovene (CSS), a prescribed standard variation of the language. It exists mostly as a construct, with no true "native speakers", although it is largely based on the variety spoken in Ljubljana, the capital (border part between the Gorenjska and Dolenjska dialectal groups) (Jakop, 2012). It is the language that is used in official settings, such as news broadcasting, books, newspapers etc. There is also a less strict form, the *colloquial standard* which is used in semi-official settings, such as radio, television, theatre etc., situations where people are speaking spontaneously (Jakop, 2012). It has slightly less rigidity compared to CSS, allowing, for example, the verb-final /i/ in the infinitive form to be dropped in speech, which is not permitted in the CSS. The colloquial languages will also vary slightly, as they are affected and influenced by the dialectal regions in which they are spoken.

Due to this, it is reasonable to assume that (spoken) CSS is rarely used, as most people do not have the need for it in everyday life. As a consequence, available recordings are scarce and hard to find, even more so recordings, where the recording quality is sufficient for analysis. Therefore, for the purpose of this thesis, the central Slovene, specifically Ljubljana colloquial standard, will be used as the model for Contemporary Standard Slovene in all phonetic analyses.

2.3 Dialectal Variation

The dialectal diversity of Slovene is most prominent in the vocabulary and minor grammatical features specific to the region/dialect. The historical and geographical factors of Slovenia resulted in a lot of separation between regions, which allowed the language to develop more independently in each region, resulting in such a high level of dialectal variety. The lexical differences among the different dialects can sometimes be great enough that speakers from two very different regions will not be able to communicate effectively, and will thus have to resort to

the (colloquial) standard to understand each other. There are even grammars and dictionaries written about certain dialects, for example, Kramarič et al. (2014) is a dictionary about the Črnomelj sub-dialect of the Bela krajina dialect (South-Eastern Slovenia, Dolenjska dialectal region), and Greenberg (2020), a grammar about Prekmurje Slovene (A Pannonian dialect, north-eastern Slovenia).

When it comes to grammatical differences, Jakop (2012) proposes the dual number as one of the "most interesting areas of difference (and similarity) between the different codes of Slovene". The dual is mandatory in CSS and is required to agree across pronouns, nouns, verbs, and adjectives. The masculine dual ends in -a uniformly across all morpheme types (given in Example 1), while the feminine dual employs a mixture of -i, -a, and -e to mark the number (Example 2) and is thus considered weaker.

(1) Dv**a** stol**a** st**a** pripravljen**a**. (*Two chairs are ready.*)

(2) Dv**e** miz**i** stojit**a** v sobi. (*Two tables are (standing) in a/the room*)

The dual seems to be disappearing in some dialects - feminine nouns in central dialects show pluralisation (eg. *dve punce* "two girls" instead of *dve punci*), and some dialects only keep the dual in the nominative and accusative case forms, but replace it with the plural forms in other cases. The ongoing loss of dual is more strongly noticeable in the dialects closer to the national (and language) borders, notably the Croatian and Italian borders (south-east and south-west), and therefore the Italian, Friulian, and Čakavian and Kajkavian Croatian. The dual appears more intact in dialects close to the Austrian and Hungarian borders (German and Hungarian influence, respectively). The dialects of Pannonian dialectal group, specifically more north-eastern ones appear to have fully preserved the dual, much like the standard form (Jakop, 2012).

Some dialectal varieties also show variation in their phonological systems. These differences tend to be limited to the vowel system and phonological rules. The Inner Carniolan dialects, part of the Vipava valley dialects, for example, do not include the phonemes /ɛ and ɔ, but have the diphthongs /eːi̯/ and /oːu̯/ in their places (Stegovec, 2012). They further include additional diphthongs that do not appear in CSS, such as /uːə/

Another example of a prominent difference is the fronting of the /u/ vowel in the Panonska region dialects, which, anecdotally, sounds similar to the German umlauted ü, however, not much research has been done to explore this phenomenon. Greenberg (2020) does note it as a distinct vowel, but it does not show any phonological evidence or any support for it being distinct rather than an allophonic variation of the standard /u/. This is probably partly due to the fact the Grammar was written in 1942, when modern methods of measuring phonetic characteristics did not yet exist. This additional vowel, and the Prekmurje vowel system in general, will be the focus of this thesis.

2.4 Slovene Phonology Over Time

Language does not only change on a dialectal level, with region, but also with time. In Slovene, literature states that change often starts in a central dialect, which with time spreads out to the surrounding dialects (Greenberg, 2006). The peripheral dialects sometimes do not internalize this change, which means they are in some regards closer to the "ancient" standard language. Greenberg provides an example of such change in Slovene as "the ictus corresponding to Proto-Slavic circumflex", which he claims was caused by a general tendency of Proto-Slavic languages (especially Polish and Czech) of shortening the circumflex. This shift is not observed in at least two peripheral dialects of Slovene; Rezija in the west, and Prekmurje in the north-east. Bethin (1998) also mentions the circumflex, but notes that it could be a phonetically motivated change. In general, she states that Slovene tends to prefer syllabic-based prosody, which is then

often a motivator of language change. The important part of this section is the awareness that what is considered Standard Slovene is based on the central dialect, which in turn is usually the "trendsetter", with neighbouring dialectal varieties, although it may never reach the most distant ones. This is yet another reason why dialectal variety can sometimes be so big in Slovene.

2.5 Dialectal Attitudes in Slovenia

The dialect a person speaks can often be associated with either positive or negative prejudice. A negative attitude towards a dialect can discourage its inter-generational transmission and as a consequence, dialects can die. In Slovenia, the trend seems to be that dialects closer to the central one/CSS tend to be perceived more highly than those more remote (and thus usually more different and distinct from CSS) as was shown in (Lundberg, 2007) and (Lundberg, 2015). Lundberg (2007) conducted a study where participants were asked a series of questions about the various dialects/dialectal groups of Slovene to determine their attitude towards these dialects. The study also included questions about participants' own dialectal use. A total of 490 Slovene students participated in the research. 85% of them indicated that they speak a dialect at home, which was, according to the author, a reason for a positive outlook on the future of Slovene dialects. A slightly smaller number (about 70%) said they would speak in their dialect to their child.

Perhaps the most controversial question in the entire study was Question 7: "In what city or region is the ugliest Slovene spoken?". This did not target any specific linguistic feature, simply addressing the participants' internal view of the dialects. A good number of participants refused to answer the question, while some commented that "no Slovene dialect should be considered ugly." However, the respondees agreed that Prekmurje was the "ugliest". Even when dividing participants by their region, every region (apart from Prekmurje) put Prekmurje as the ugliest. The second place was assigned to Ljubljana (central Slovene). Prekmurje was furthermore also judged as the least prestigious of all the dialects. This suggests a possibly negative outlook on the Prekmurje dialect and could indicate a negative judgment of the people from that region solely based on their dialect. At the same time, however, people from Prekmurje were across regions selected to be the most loyal to their dialect, and they scored the highest when indicating the importance of their dialect to their local identity. We can conclude that this shows a strong sense of belonging, which is encouraging for the perseverance and longevity of dialects in general. It further demonstrates that the Prekmurje participants' attitudes towards their own dialect are positive, despite the associated prejudice.

2.6 Pannonian Region and Dialectal Group

The notions "Panonska region", "Panonska dialectal region/group", "Prekmurje" and "Pomurska/Mura region" are often used interchangeably in texts and colloquial Slovene. Regrettably, that can lead to some confusion, as they do not refer to the exact same things. To visualize it better, an overview of all four classifications is presented in Figure 3.[2]

The Panonska/Pannonian region (Figure 3a), is the geographical region of Slovenia in the far north-east of the country, and should not be confused with the dialectal region (Figure 3b), even though they do overlap to a good extent. The former includes parts of Slovenia which fall under a completely different dialectal group. A closer geographical boundary of the dialectal region would be the Mura statistical region ("Pomurska statistična regija" in Slovene), bounded by municipal borders in the (South) West, Croatia in the South, Hungary in the (North) East, and Austria in the (North) West, as seen in Figure 3c. According to Statistični Urad

---

[2]Base image accessed at: https://commons.wikimedia.org/wiki/File:SloveniaBaseMap.png on 2024-06-08

(a) Panonska geographical region

(b) Panonska dialectal region

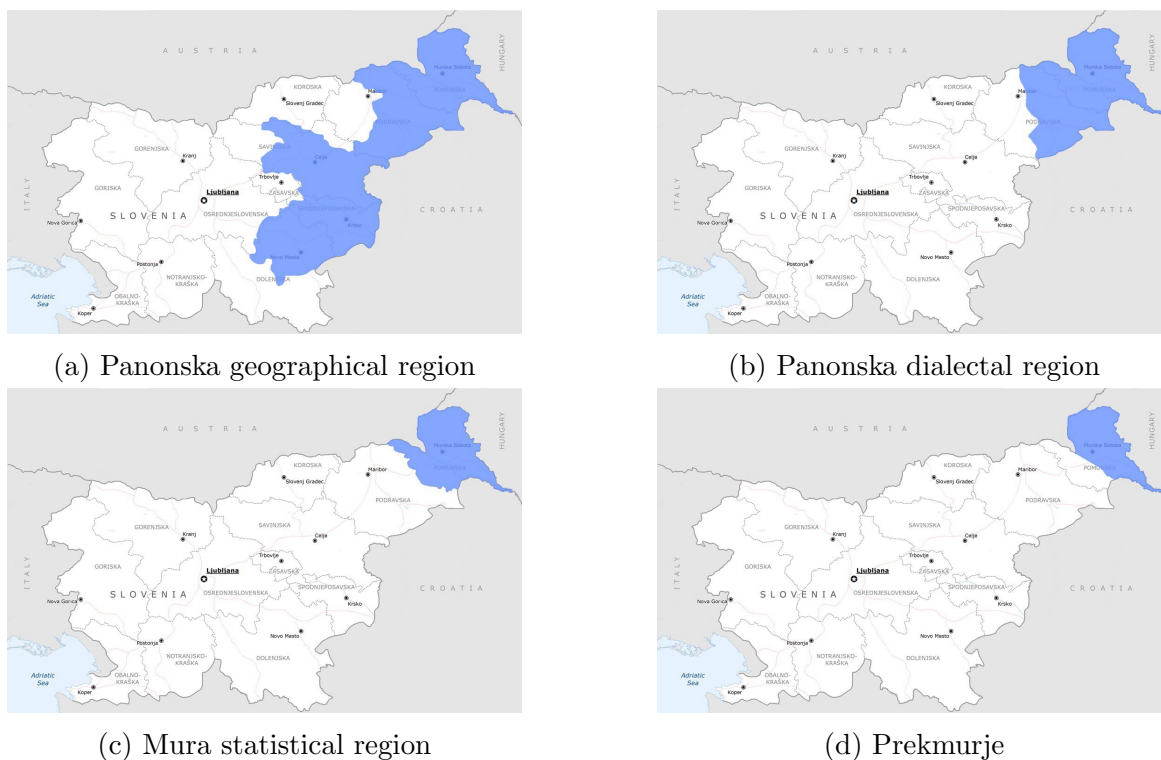(c) Mura statistical region

(d) Prekmurje

Figure 3: Comparison of Panonska geographical and dialectal regions with the Mura statistical region and Prekmurje

Republike Slovenije (The Statistical Office of the Republic of Slovenia), 114,163 people lived in this region on July 1st 2022, which could be a fair assumption for the number of speakers of the associated Pannonian dialect region, although the dialectal region is not a perfect overlap with the statistical region. Rather, the Pomurska region is a (large) subset of the area where the Panonska dialect is present. Since the missing area is very small, it suffices to say, for the purpose of this research, that the difference is negligible. Finally, Prekmurje (Figure 3d) is one of the dialects of the Pannonian/Panonska dialectal region as well as a sub-region of the Pannonian one, bounded by the river Mura on the southwest.

2.6.1 Pannonian dialects

As a dialectal group, Pannonian includes four distinct dialects; Slovenske Gorice, Prlekija, Haloze, and Prekmurje dialect (Jakop, 2012). Since, geographically, the dialects lie on the very edge of Slovene, this is reflected in their structural and lexical differences from other dialects. In the 2015 study by Lundberg, 169 out of 265 participants (64%) from the Koroška (Carintia) region indicated Panonian Slovene as the least comprehensible of all Slovene dialects. He had a similar finding in his 2007 study, where again, this time by participants from all over Slovenia, the Prekmurje dialect comprehensibility was ranked an average of 2.68 on a scale from 1 to 7, with a higher number indicating higher comprehensibility.

2.7 Prekmurje Dialect

Prekmurje is the most north-eastern of all Pannonian dialects, with the river Mura separating it from the other three. It is also the only dialect of Slovene that directly borders three other languages; Hungarian, as well as (Austrian) German and Croatian. It is the only Slovene dialect that borders Hungarian.

2.7.1 Grammar of Prekmurje

There is a comprehensive grammar of the Prekmurje dialect, originally written by Avgust Pavel and published in 1942 in Hungarian under the title "Vend nyelvtan". The grammar was translated into English in 2020 by Marc Greenberg. The grammar is an extensive and comprehensive account of the Prekmurje dialect, covering approximately 450 different features of the dialect, ranging from a phonological description, to morphosyntax and lexical peculiarities.

The most relevant for this thesis is the section on Prekmurje phonology. The grammar discusses the vowels on the basis of the prescribed orthography. Based on this, Prekmurje has the following vowels: a, á, e, ê, é, i, í, o, ô, ó, ö, ő, u, ú, ü, ű. Since the grammar is based on the orthography, the schwa is omitted, although it should be safe to assume it is present in the dialectal vowel inventory. All the vowels with the acute accent are long vowels, while the ones without are short. The circumflexed vowels indicate diphthongs; ê a 'short, closed e followed by a shorted, less fully articulated i", and ô "a short o and a short, less fully articulated u" (Greenberg, 2020). By this orthographical account, the Prekmurje dialect does not seem to contain the vowels /ɛ/ and /ɔ/. Overall the orthography itself is heavily based on the Hungarian orthography, as is the explanation of the pronunciation of each vowel. This makes sense when the original intention of the grammar is taken into account; it was commissioned during the second world war by the Hungarian Educational Society of Prekmurje as a tool of integrating the local Prekmurje population into the Hungarian education system and society (Greenberg, 2020). As no formant values of the vowels are given, this remains an open question of the dialect's phonology.

There has been some research that focuses on the vowel system of Prekmurje and the perception of Slovene vowels in speakers from the Prekmurje region Huber (2006) researched the perception and identification of Slovene vowels in high school students from the Mura statistical region. He was specifically interested in finding out the importance of segmental and suprasegmental features in differentiating the meaning of Slovene words. He tested the participants with minimal phonological pairs, where the most interesting results came from the /e/ /ɛ/ and the /o/ /ɔ/ pairs. The participants' scores ranged from around 65% to 90%, which was less than expected on a minimal pair discriminatory test. The participants furtermore scored significantly worse than university students in central Slovenia, which had been tested on the same set of stimuli in Tivadar (2004b). This can be an indication, that speakers of the Prekmurje dialect rely less on the segmental features, such as vowel hight, and therefore differentiate mid-vowels worse than speakers of the central Slovene dialect.

2.7.2 Subdialects of Prekmurje

According to Greenberg (1993), the Prekmurje dialect, already a sub-dialect of the Panonska region can be subdivided into three traditional subdivisions; goričko (northern Prekmurje, the highland dialect), ravensko (southwestern Prekmurje, the plains dialect), and polansko (southeastern Prekmurje, the lowland dialect). The article presents three village dialects, each corresponding to one of the three subdivisions. Even at this scale, the differences among the village dialects can be observed. Greenberg (1993) gives a three-way distinction between these village dialects, shown in Table 2.

The first feature is preferring a /j/ over a /d/ in front of a stressed vowel, or immediately following a consonant, the second feature is the roundedness of the /a/ phoneme, and the final feature the word-final consonant preference. This just further showcases the rich dialectal diversity of the Slovene language, and even more particularly the Pannonian dialectal group.

Table 2: Village dialects' characteristic distributions

Subdialect	j>d	ă>å	(-l >) -u >-o
goričko	+	+	+
ravensko	+	+	±
dolinsko	-	-	-

History of Prekmurje dialect

The development of the Prekmurje standard dialect is thoroughly described in Jesenšek (2005). The first printed instances of the dialect date back to the 18th century; a protestant religious text, and an alphabet booklet with examples for teaching literacy, however, the dialect itself is much older. According to Jesenšek (2005), already in the 9th century, religious rituals were conducted in Slovene. In the following centuries, the church language varied from German, to Kajkavian, to Hungarian, based on the jurisdiction of the regional areas. By the 18th century, the diocese unified Prekmurje, which significantly decreased the influence of Kajkavian. Also in the 18th century, the "Prekmurje standard language" (also "pure or old slovene language") was formed on the basis of the ravensko and goričko speech. It took until the late 19th/early 20th century for the polansko subdialect features to integrate in the Prekmurje standard. Throughout this time, the people of the region also kept awareness (and, to an extent, mastery) of the standard Slovene, which prevented the language from assimilating with the Kajkavian language. By the end of the first world war, the differences between the written and spoken Prekmurje were becoming more and more noticeable, while the influence of the neighboring languages was at an all-time low. Unfortunately, the new millennium did not bring the "language rebirth" some had been hoping for. Nowadays, schools focus more on teaching the contemporary standard Slovene, and then English and other foreign languages, instead of preserving the local dialect.

3 Research Question and Hypotheses

Since Slovene has such a rich dialectal diversity, and not much work has been done on the vowels of the Prekmurje dialect, that will be the focus of my research. For this thesis, the main aim is to map the vowels of the Panonska region speech, specifically the Prekmurje dialect, and compare them to the standard Slovene vowel inventory. Is the fronting of the /u/ vowel consistently present? If the fronting is not consistent, what determines if the vowel is fronted or not? If the fronting of the /u/ vowel is consistent, does it then also result in the raising of the /o/ and /ɔ/ vowels? Do all the other vowels align with the standard variety, or are there further deviations?

As a secondary part to this thesis, my interest is in the vowel diversity itself, does it exist solely in production, and can thus only be noticed in the acoustic recording, or is it noticeable in perception as well, where Slovene speakers from other regions can distinguish the Prekmurje vowels from the standard ones?

I expect the F2 values for some of the /u/ vowels to be significantly higher in the Panonska region speech than they are in Standard Slovene and furthermore I expect there to be two distinct closed (mid) back phonemes thus a confirmation of the existence of a fronted /u/ in coexistence with the standard corner vowel. Similarly, I expect the F1 values of the Panonska region /o/ vowel to be lower than those of the standard variety, as a possible compensation for the fronting of the corner vowel. The other vowels are expected to have similar F1 and F2 values to the standard vowels. Lastly, Slovene speakers are expected to be able to discriminate between

vowels produced by a central Slovene speaker better than vowels produced by a Prekmurje speaker, on the account of the predicted differences in phonology.

4 Research Part 1 - Formant Analysis

For the first part of this thesis, a formant analysis was carried out on the vowels of the Prekmurje dialect as well as the central Slovene dialect, intended to substitute the Contemporary Standard.

4.1 Methodology

To facilitate data collection in a limited time frame, an online corpus of recorded spoken Slovene, Artur 1.0 (Verdonik et al., 2023) was used to obtain formant values of the vowels for both the standard version and the prekmurje dialect.

The corpus contains multiple subsections, which include recordings of participants reading or spelling, studio recordings, public speech recordings, and private elicited speech recordings among others. As this research is focused on the natural, spontaneous version of the Northeastern dialect, the Artur-N sub-corpus was chosen, as this is the subcorpus with private speech. It includes three sub-sections. The first is the Artur-N-Obrazi, where speakers were presented with picture of faces and asked to describe them. The second is the Artur-N-PDom, which contains recordings of people reading a smart-home system set of commands. Both of these provided data for standard Slovene. For Prekmurje Slovene, the Artur-N-Prosti sub-section was used, which is comprised of monologues and dialogues of people conversing on everyday topics, which proved to be the most suitable option for obtaining naturalistic Prekmurje speech.

The corpus further included demographic information on all the participants in the recordings, where their age, sex, level of education, municipality of residence presently and growing up, style of speech (standard language or dialect), and native language were recorded. There was also an indication whether the speaker had a speech impediment or not. This demographic data made it easy to select recordings applicable to the research purpose of this thesis.

4.1.1 Participants

There were recordings of eleven speakers from the Panonska region included in the database, although one was eliminated immediately, as the description of the recording indicated they no longer lived in the region at the time of the recording. Another speaker had a speech impediment and their recordings were therefore also not included in the research. The remaining nine speakers had a total of thirteen recordings associated with them.

Out of the participants from the Panonska region, six were female and three were male. One participant was aged 12-17, five were in the 30-49 years range, and three speakers were older than 60 years. Four participants had higher education (Bachelor's, Master's or PhD degrees), three completed vocational secondary education, and two speakers had completed basic education (primary school).

To generate data for the modern standard Slovene, recordings of speakers from the central Slovene region were selected, as that is the closest natural speech to the standard variety. Originally, there were 226 recordings. After eliminating those where the native language was not Slovene and the style of language not indicated as "standard", 75 recordings were left. In the second step of elimination, everyone who was not born or did not grow up in central Slovenia was also removed. This left a remaining pool of 62 recordings, 9 of which were then matched to the Panonska region participants by gender, age, and level of education when possible. The order of importance of the match was gender > age > education. All matches had at least the first two criteria fulfilled, with the exception of the Pannonian male aged 12-17, who was

matched with a central Slovene male 18-29 years old. This match, however, still had two common criteria fulfilled, as the match had the same level of education.

The central Slovene participant group then also consisted of six female and three male speakers, with the age distribution being the same as the Pomurska group, apart from the switch of the 12-17 year-old participant for the 18-29 year-old one. Three speakers from central Slovenia had completed higher education and one had completed a professional study programme. Four had general, vocational, or technical secondary education and one speaker had basic education (primary school).

4.1.2   Annotation of Recordings

The recordings were imported into Praat (Boersma and Weenink, 2024), where a TextGrid object was created for annotation purposes. In the TextGrid object, a tier was created where each instance of a vowel was marked at its beginning and end, thus containing the vowel's full duration. The marking also included if the vowel was in a stressed or an unstressed syllable ("uu" for unstressed /u/, "us" for stressed /u/, etc.). The beginning and endpoints of the vowels were determined by observing the spectrogram, the wave form, and through auditory cues. The focus of this research was on monophtongs, thus diphthongs were not annotated, as the boundary between the two vowels would be impossible to unambiguously determine and the formant values are likely to differ from those of the corresponding monophtongs. A screenshot of an example TextGrid is given in Figure 4.
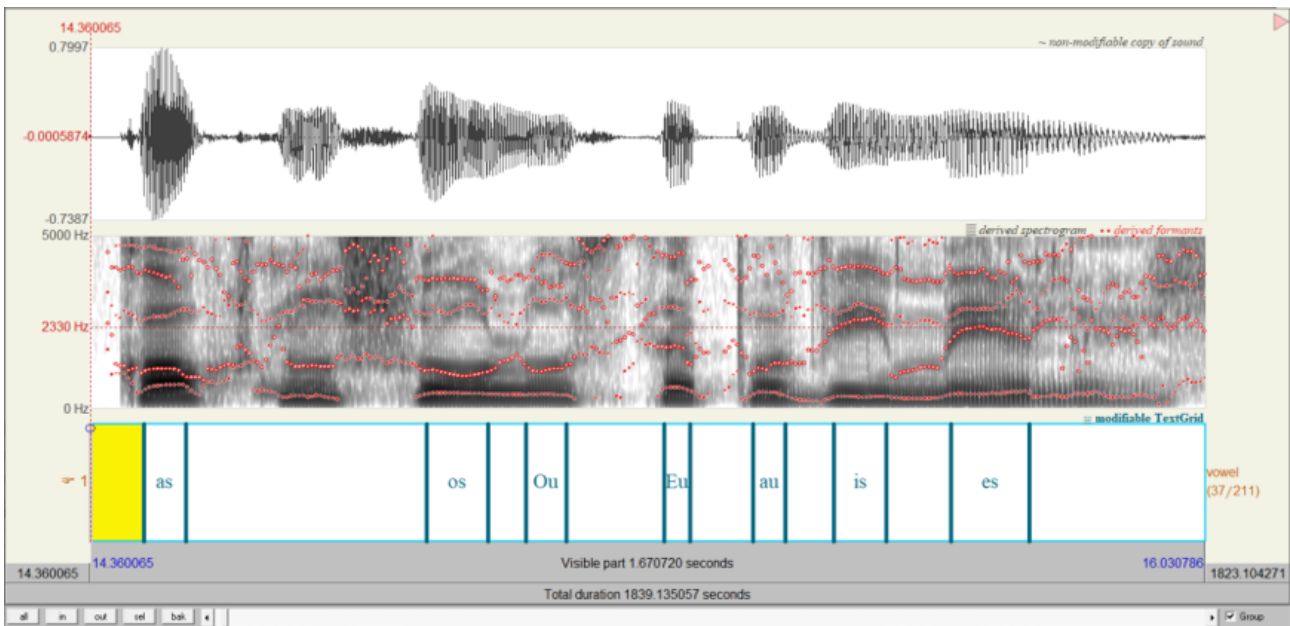

Figure 4: Example of TextGrid annotations

Not all of the recordings were of equal length, some were as short as 90 seconds, while others exceeded 30 minutes. In an attempt to extract as much speaker variety as possible and at the same time limit the influence of a single speaker as much as possible, the recordings were not annotated in their entirety.

The recordings shorter than 120 seconds (two minutes) were annotated in their entirety. For the longer recordings, two possible criteria for stopping an annotation of a recording were implemented;

The first ten instances of each of the eight vowels were marked, regardless of their stress. As an alternative condition, even if not all vowels had reached the ten annotations threshold, the recordings were annotated for two minutes of their duration. Annotations for a specific

recording were completed when either of the two criteria was met. Very rarely, two speakers would talk at the same time. Those instances were also left out from the annotation, as the frequencies of their individual speech could have skewed each other's formant values.

Due to the relative frequencies of vowels, not all of them were recorded the same amount of times. Marvin et al. (2019) found that /a/ is the most common Slovene vowel phoneme, and /ə/ the least. The difference between their distributions is quite big - /a/ occurs approximately eleven times more often than /ə/ in CSS.

After the annotations were complete, a PraatScript file was written, which sifted through all TextGrid files in the designated folder and extracted the first and second formant (F1 and F2) values for each vowel that had been marked on the TextGrid. The extraction was done at the midpoint of the vowel duration, where the vowel was assumed to be the most stable and clear in its pronunciation. The script than saved the values in a table, which was later imported into the statistical software used. The PraatScript code is given in Appendix 1.

4.2   Results

The following sections present the results of the Prekmurje and standard Slovene vowel formant analysis.

4.2.1   Prekmurje Vowel System

A total of 2504 vowel phonemes were marked in the annotation process across all the recordings. The average values of F1 and F2 were calculated for each vowel and those values are presented in Table 3, differentiating also stressed and unstressed instances of the vowels. A vowel chart of the average values was created to visualize the vowel space of the Prekmurje dialect and can be found in Figure 5.

Table 3: Average F1 and F2 frequency scores of Prekmurje vowels in Hz, rounded to a full number

vowel	avg-stressed		avg-unstressed		overall average	
	F1	F2	F1	F2	F1	F2
a	750	1294	637	1361	681	1335
e	460	2084	466	1873	461	2056
ɛ	698	1808	568	1740	596	1755
i	344	2299	398	2153	379	2203
o	476	950	471	962	474	954
ɔ	485	982	498	1152	497	1136
u	394	1090	431	1190	412	1138
ʉ	330	1730	434	1377	336	1712
ə	/	/	602	1476	602	1476

4.2.2   Central Slovene Vowel System

The same procedure was applied with the central Slovene standard. It yielded 1507 annotations across eight separate vowel phonemes. The vowel chart based on the average formant values is given in Figure 6.
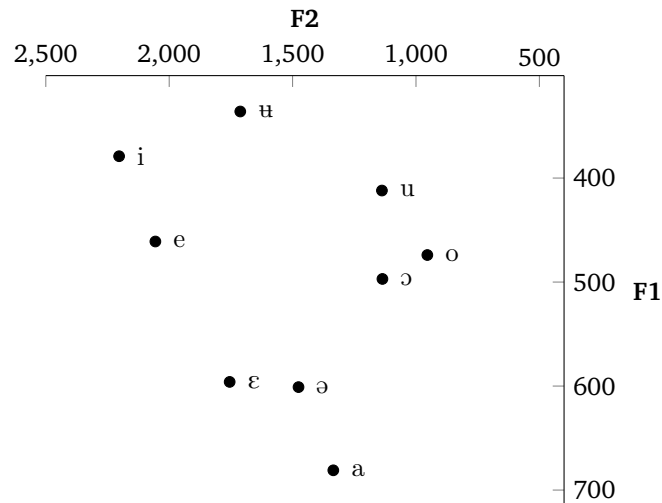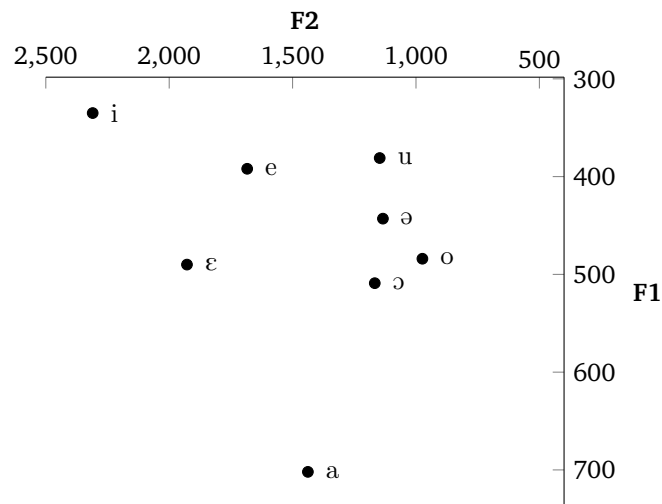
Figure 5: Vowel chart of the Prekmurje dialect

Figure 6: Vowel chart of the central Slovene dialect

4.3 Analysis

The results obtained in this part of the study showed clear support for the existence of a fronted /u/ vowel as a distinct phoneme in the Prekmurje dialect. It has a significantly higher F2 value than the standard /u/ variant. Based on its formant values, the phoneme was determined to be closest to the rounded central high vowel, /ʉ/, which is the notation that was used in the chart. This confirms the first hypothesis that there is phonetic evidence of a distinct fronted /u/ phoneme.

On the other hand, the remaining back vowels do not show any significant signs of raising, when comparing them to the vowels of the central standard. What can be noted in both graphs, however, is that the /o/ and /ɔ/ morphemes appear much closer to each other than they did in the Petek, Šuštaršič, and Komar (1996) study (Figure 2). A statistical analysis was done using R[3] and RStudio[4] on the data for Prekmurje back vowels which showed a significant difference ($p < 0.05$) for both formant values. Thus even if slight, there is still a clear difference between the two vowel in both formant measures. The difference is bigger in F2 values (F2 of /ɔ/ is on average 182Hz higher than that of /o/) and very slight in F1 values, where /ɔ/ is on average

---

[3]https://www.r-project.org/
[4]https://posit.co/download/rstudio-desktop/

22Hz higher. On the other hand, this difference (while significant) is small, and already much smaller than in the past. With fewer than 30 years since the publishing of the reference data, this difference could point towards an early stage of a merger between the two vowels.

One striking result of the central Slovene vowel map is the low F2 value for the /e/ phoneme, as well as a low F1 value for the /ə/. As the vowels of standard Slovene were sampled from the Artur-N-PDom sub-corpus, the recordings did not include many instances of the latter, and no instances of the former. In general, the recordings of the central dialect produced significantly fewer annotated vowels than those of the Prekmurje dialect. This is at least partially due to the fact the recordings of the Artur-N-PDom were only 90 seconds in length, and thus much shorter than the ones in the Artur-N-Prosti section from where the Prekmurje data was sourced. All the samples of the /e/ vowel in standard Slovene came from a single recording (from the Artur-N-Prosti corpus). His frontal vowels in general had lower formant values than the average, which is likely why the /e/ vowel ended up looking misplaced among the rest. A similar situation applies to the schwa.

Importantly, the corner vowels show no significant deviation from one another, nor do their formant values differ majorly from the ones established by Petek, Šuštaršič, and Komar (1996), which shows their stability.

5 Research Part 2 - Perception Experiment

The second part of this thesis addresses the perception of Prekmurje vowels compared to those of standard Slovene.

5.1 Participants

The participants were recruited on social media[5,6] and through personal connections. They were required to be adult native speakers of Slovene from any region of Slovenia. A further requirement was that they had not lived abroad for a longer time period (over a year). There were no other requirements concerning their linguistic background.

5.2 Methodology

Snippets of vowels were extracted from the recordings in the Artur 1.0 corpus (Verdonik et al., 2023). All the vowel snippets used came from the recordings that had been used in the first part of the research. Here again, the speakers from central Slovenia were presumed to be speaking in what is the closest approximation to modern standard Slovene. An effort was made to have an even distribution of genders of the speakers of the stimuli, to hopefully avoid any bias.

5.2.1 Stimuli

The sections with standard Slovene had seven recordings each (one of each vowel, excluding the schwa), and the sections with the Prekmurje dialect had eight recordings each, the seven standard vowels (again excluding schwa), as well as the fronted /u/ from Part 1 (/ʉ/), following both the Greenberg (2020)'s definition of the dialect differentiating between the two forms of u and the results of Part 1. The schwa was omitted from the testing as it was found to be inconsistent in pronunciation and highly speaker dependent, and, furthermore, rarely appears in a stressed position. This resulted in a total of 30 stimuli, four of each vowel (except /ʉ/ which only appeared twice as it was only relevant for the Prekmurje sections).

---

[5]https://www.instagram.com/
[6]https://www.facebook.com/

5.2.2 Procedure

A questionnaire was created using Qualtrics[7], where participants were asked to listen to these vowel snippets and decide which vowel they had heard. There were four sections of vowel recordings; two of the standard Slovene and two of the Prekmurje dialect, each split into two blocks for better transparency and easier responding. The sections were presented to the participants interchangeably, the first and the third were the Prekmurje dialect, while the second and fourth the standard Slovene vowels. The vowels within each section were randomized for order. For each vowel, the participant selected their response vowel from a multiple-choice option where all eight possible responses were listed. For the vowels /e/ and /ɛ/, and /o/ and /ɔ/, (in the questionnaire denoted as é, ê and ó, ô, respectively) the instructions on each page contained example words where the designated vowels appeared, as it was presumed most participants would not be familiar with linguistic notation. Slovene orthography uses one grapheme, "e", to denote both front mid vowels, and "o" for both back mid vowels, so this was done to aid the participants in their selection of the correct vowel.

The participants were not timed, nor did they have to answer the questions in a set order within a block, however, they were not able to progress to the next section without answering all the questions (selecting all the vowels) in the previous one. Similarly, they were not able to return to a previous block once they had progressed to the next one.

The questionnaire also included some basic demographic questions, such as age and sex, and, more importantly, two questions about the participants' region. The first asked the participants which region of Slovenia they grew up in, and the second asked them to indicate which region they now reside in.

5.3 Results

A total of 135 participants attempted the questionnaire, out of which 110 provided at least partial responses. 100 participants completed the entire questionnaire, which will serve as the basis for the analysis. Out of those 38 were male and 62 were female. 31 participants were 18 to 29 years old, another 31 between 30 and 49, another 31 were aged 50-69, and 7 were 70 or older. The overview of the distribution of gender and age is given in Table 4 and the overview of the regions the participants come from is given in Table 5.

Table 4: Distribution of Gender and Age (in years)

Gender	18-29	30-49	50-69	70+
Male	7	11	17	3
Female	24	20	14	4

No participant selected all the vowels correctly. One participant had one mistake, and eight participants had two mistakes. The participant with the most mistakes responded erroneously to 22 stimuli (out of 30, thus in 73.3% of the time). The average participant responded correctly to 22.87 stimuli. There was one stimulus where all the participants responded correctly; The first /a/ of the central Slovene vowels. An additional two stimuli only had two wrong answers, both the first and the second /i/ of central Slovene.

The best-performing section was Section 3 (second round of Prekmurje vowels), where the average correct responses per vowel was 83.50. However, Section 1, the other section with Prekmurje vowels had the lowest score of correctly identified vowels, with an average of 67.88

---

[7]https://www.uva.eu.qualtrics.com

Table 5: Distribution of Regions of participants' origin

Region	Resides presently	Resided in childhood
Central Slovenia	67	53
Southeast Slovenia	6	8
Coastal-Karst	2	0
Gorizia	0	1
Upper Carniola	2	5
Littoral-Inner Carniola	1	3
Drava	9	10
Mura	7	9
Savinja	1	2
Carinthia	0	2
Lower Sava	0	1
Central Sava	5	6

Table 6: Scores per Vowel and average scores per Section

Section	Score for vowel in section								Average score
	a	e	ɛ	i	o	ɔ	u	fronted u	
1 (Prekmurje)	92	76	76	65	47	65	52	70	67.88
2 (Cenral)	100	57	73	98	62	64	89	/	77.57
3 (Prekmurje)	98	79	84	84	71	84	88	80	83.50
4 (Central)	88	77	69	98	53	54	94	/	76.14
Overall average	94.50	72.25	75.50	86.25	58.25	66.75	80.75	75.00	76.23

participants correctly identifying the vowels. The vowel with the lowest amount of correct responses is also from this section, the vowel /u/. The sections with central Slovene vowels had an average correct response rate of 77.57 for Section 2, and 76.14 for section 4. The overall average score for central Slovene vowels was therefore 76.85 and 75.69 for the Prekmurje vowels. The overview of scores per vowel and average scores per section is given in Table 6.

5.4 Analysis

It was surprising to see the average vowel be identified correctly in only 76.23% of the participants. It was the expectation that participants would perform significantly better, as they were tested on the vowels of their native language and should therefore be familiar with them and adapt at distinguishing them correctly. There are a couple of factors, which could have contributed to this result.

Firstly, a lot of participants seemed to make mistakes of confusing two adjacent mid vowels, either mixing up /e/ and /ɛ/ or /o/ and /ɔ/. That could be due to Slovene orthography not having a separate character for the two vowels in each pair, but rather marking both with the glyphs e and o, respectively. Consequently, the storage of the two phonemes in the mind is likely closer to eachother and Slovene speakers might not consciously process them as two separate

vowels, but rather (allophonic) variations of each other. Alternatively, it is also possible that the participants confused the meaning of the symbols used in the questionnaire and therefore selected the wrong answer(s). The participant with the lowest score, for example, had no instances of the mid vowels correct, but often responded with the corresponding pair mid vowel instead of the target - thus indicating $é$ (/e/) in place of $ê$ (/ɛ/) when the latter was the target morpheme and vice-versa, with the same mistake present in their responses for the back mid vowels. This reversal mistake was also observed in other participants, although not to the same extent.

Another factor could be that speakers in general have a harder time distinguishing non-corner vowels and thus mistakes are more likely to occur. It is likely not coincidental that the vowels with the highest number of correct responses were the corner vowel stimuli (a, i, u) as was seen in Table 6. The overall average scores for all three of them are above 80%, with /a/ being the most distinctively recognizable with a 94.50% correct response rate. This is aligned with the prediction that participants will perform better on the corner vowels and struggle more with the mid vowels.

When it came to comparing the two stimuli sections of the test (Prekmurje and standard), the two groups received very similar mean scores, with the participants on average performing slightly better when presented with stimuli from the central/standard variant as opposed to the Prekmurje dialect. The difference in favor of the central vowels was 1.16 points. A statistical analysis was carried out in RStudio again to compare the two dialectal stimuli sections. The analysis did not show any significant difference between participants' responses to the Prekmurje or Standard Slovene sections. As a result, the hypothesis about Slovenes understanding central Slovene vowels better than Prekmurje vowels cannot be confirmed.

A linear model was also computed on participant's demographic data on their region, both the region they currently reside in and the region they grew up in. Looking first at the current place of residence, participants from the Mura (Prekmurje) statistical region on average scored 1.24 points worse than participants from other regions of Slovenia. The participants who presently live in Prekmurje had on average 1.37 mistakes more in the entire questionnaire than participants from other regions. This difference was not significant (p=0.34), thus we cannot speak of any effects of current region of residence. The same test was done on the data about the participants' childhood region, where again, participants from outside of Prekmurje scored slightly better, making on average 1.24 mistakes fewer than those who grew up in the region. This difference was also not significant (p=0.33), thus here as well, we cannot claim the region of their upbringing aided the participants in the task.

As a final statistical analysis, a test was run using the same regional demographic data, but this time focusing only on the stimuli from the Prekmurje region dialect. Once again, a slight advantage was seen in participants not from Prekmurje. When it comes to current region, Prekmurje participants scored on average 0.45 points fewer on the Prekmurje sections, compared to other participants. This difference was, however, not significant (p=0-58). The difference when taking into account the childhood region was also not significant (p=0.16), although the average Prekmurje participant performed 1.00 points worse than their peers from the rest of Slovenia.

6   Discussion

Many mistakes in the questionnaire point towards a confusion when selecting the appropriate symbol for the mid vowels. While example words were provided at the start of every session, participants often indicated the wrong vowel. While this could be a general indication of poor distinction of those vowels, the mistake was usually reversed with the corresponding mid-vowel pair. This suggests the mistake emerged from a misunderstanding of the symbols used in the

multiple choice selections. To avoid such problem in the future, the answers themselves could be monosyllabic words containing the target vowel, at least for the vowels which share a grapheme in the written system of the language.

Another limitation of the design came from the recordings themselves; as the vowels were sampled from spontaneous speech recordings in the Artur1.0 corpus (Verdonik et al., 2023), there was not as much consistency among them as one would ideally want. This meant that some recordings were clearer than other, and some were slightly longer or slightly shorter. In future research, it would make sense to recruit speakers to record prepared stimuli and extract the vowels from those recordings, although a formal setting like that could inhibit the strength of the dialect. Speakers would be likely to use a more standard or formal language, and would have to be explicitly told to pronounce the stimuli as they would in their dialect. This would, however, likely still result in less naturalistic production of the vowels compared to spontaneous speech.

As a final comment on the experimental design, the stimuli themselves could be automated to play multiple times. Thus instead of having to click on the play button multiple times to hear it, upon clicking, the vowel would play, followed by a second of silence, and then the vowel played again. This would make it easier for the participants to hear the stimulus multiple times, and it would let them focus on hearing the stimulus instead of clicking the play button.

The statistical analyses showed the Prekmurje participants scored slightly worse than participants from other regions. Furthermore, this was shown not only globally, but also when taking only the Prekmurje stimuli into account, which perhaps is even more interesting. It would be natural to assume that participants from that region should be at least as good at discerning the dialectal vowels, if not better than other participants. However, it should be noted again, that these differences were not statistically significant, and thus could be just an anomaly of the sample. It should also be pointed out that there were far more participants from other regions than there were from Prekmurje. Only seven out of the 100 participants currently reside there, and nine out of 100 lived there as children, thus we are comparing two groups where one is almost ten times bigger in size. Should any further research be done, effort should be made to recruit more participants from the target region. It is also possible that participants from Prekmurje are in general worse at differentiating nearby vowels (such as the front mid-vowels, or even more so the back mid-vowels), as was already indicated in Huber (2006). Since in this experiment, the participants only heard the vowels themselves, and not a word or sentence segment, they had only segmental clues to rely on. The lack of suprasegmental clues like intonation and stress could have been a factor in their performance. A future test could be done in the form of AA/AB testing, where participants hear two consecutive vowel snippets and must decide if they are the same vowel or not. Variation for speaker would, of course, have to be included as well. This would be especially relevant for the mid vowels with Prekmurje participants, and, on the other hand, to test of Slovene participants would classify /u/ and /ʉ/ as two different vowels or not. This research has shown that they can hear and differentiate it, but the AA/AB test would determine if they also perceive it as a separate vowel rather than an allophonic variation of the standard.

7 Conclusion

This thesis has built up on previous work on the Prekmurje dialect and has provided a comprehensive analysis of the vowel phonemes. The formant analysis confirmed the presence of a ninth vowel, unique to the dialect. Based on its characteristics, it was most similar to the high mid closed vowel, represented in the IPA by /ʉ/. This confirmed the first hypothesis. The second hypothesis could not be confirmed, as the /o/ vowel in Prekmurje did not show signs of being significantly higher (lower F1 values) than the /o/ in standard Slovene. Apart from that,

however, the vowel analysis showed the back vowels are significantly closer to each other in their formant values as they were in source literature. A similar phenomenon was observed when analyzing the vowel system of modern central Slovene. While the vowels are still statistically significantly different from one another, their increased proximity could be the early stages of a merger process. The experimental part of the thesis showed that Prekmurje participants had a harder time distinguishing mid phonemes, as had been previously recorded in Huber (2006), for example. This meant that the final hypothesis could also not be confirmed, as the Prekmurje participants did not differentiate their dialectal vowels better than participants from other regions.

Various potential future studies were proposed in the discussion section. Perhaps the most intriguing one would be the AA/AB experiment, which would shed some light on the processing of vowel phonemes in Slovene speakers, both those from Prekmurje, and those from other regions.

To conclude, there is still a lot of research that needs to be done in regards to dialectal diversity in Slovene, and on Prekmurje as a standalone dialect. While this thesis had provided some new insight and definitively showed the presence of the /ʉ/, further phonological (and linguistic in general) research could be done on the dialect to better understand the speakers' processing of it.

References

Bethin, Christina Yurkiw (1998). *Slavic Prosody: Language change and phonological theory*. Cambridge University Press.

Boersma, Paul and David Weenink (2024). *Praat: doing phonetics by computer [Computer program]*. URL: http://www.praat.org/ (visited on May 6, 2024).

Greenberg, Marc L. (1993). "Glasoslovni opis treh prekmurskih govorov in komentar k zgodovinskemu glasoslovju in oblikoglasju prekmurskega narečja". In: *Slavistična revija* 41, pp. 465–487.

Greenberg, Marc L. (2006). "The Slovene Sound System Through Time". In: *Slavistična revija*.

Greenberg, Marc L. (2020). *Prekmurje Slovene Grammar: Avgust Pavel's Vend nyelvtan (1942)*. Brill.

Grošelj, Robert (2023). "Namenilnik v panonski in štajerski narečni skupini slovenskega jezika: pregled na osnovi narečnih opisov". In: *Slovenski jezik / Slovene Linguistic Studies*, pp. 89–106.

Gvozdanović, Jadranka (2012). "Number in Sorbian with special reference to the dual". In: *STUF - Language Typology and Universals*, pp. 280–295.

Huber, Damjan (2006). "Percepcija samoglasniških opozicij pri dijakih pomurskih srednjih šol". In: *Jezik in slovstvo* 51.

Jakop, Tjaša (2012). "Use of dual in standard Slovene, colloquial Slovene and Slovene dialects". In: *Linguistica*, pp. 349–362.

Jesenšek, Marko (2005). "Nastanek in razvoj prekmurskega knjižnega jezika". In: *Slavistična revija*, pp. 1–12.

Jurgec, Peter (2011). "Slovenščina ima 9 samoglasnikov". In: *Slavistična revija* 59 (3), pp. 243–268.

Kocjančič Antolík, Tanja and Hotimir Tivadar (2019). "Ultrazvočni pogled na artikulacijo slovenskih samoglasnikov". In: *Slovenski javni govor in jezikovno-kulturna (samo)zavest. Obdobja 38*, pp. 71–79.

Kramarič, Janez et al. (2014). *Slovar črnomaljskega narečnega govora*. Knjižnica Črnomelj.

Lundberg, Grant H. (2007). "Perceptual Dialectology and the Future of Slovene Dialects". In: *Slovene Linguistic Studies* 6, pp. 97–109.

Lundberg, Grant H. (2015). "Preliminary Report on Dialect Attitudes in Austrian and Slovene Carinthia". In: *Slovenski jezik / Slovene Linguistic Studies*, pp. 167–182.

Marvin, Tatjana et al. (2019). "Porazdelitev fonemov v slovenščini in izdelava matričnega testa za govorno avdiometrijo". In: *Slavistična revija* 67, pp. 537–550.

Petek, Bojan, Ratislav Šuštaršič, and Smiljana Komar (1996). "An acoustic analysis of contemporary vowels of the standard Slovenian language". In: *Proceeding of Fourth International Conference on Spoken Language Processing*.

Smole, Vera (2009). "Pomen in vloga (slovenskih) narečij danes". In: *Slovenska narečja med sistemom in rabo*, pp. 557–563.

Statistični Urad Republike Slovenije (2022). *Slovenske statistične regije in občine v številkah*. URL: https://www.stat.si/obcine/sl/Region/Index/1 (visited on May 16, 2024).

Stegovec, Adrian (2012). *The Vowel System of the Dialects of the Vipava Valley in Government Phonology*. URL: https://www.academia.edu/download/36845846/stegovec_2012_vipavavowels_MS.pdf (visited on June 16, 2024).

Tivadar, Hotimir (2004a). "Fonetično-fonološke lastnosti samoglasnikov v sodobnem knjižnem jeziku". In: *Slavistična revija* 52, pp. 31–48.

Tivadar, Hotimir (2004b). "Priprava, izvedba in pomen perceptivnih testov za fonetično-fonološke raziskave (na primeru analize fonoloških parov)". In: *Jezik in slovstvo* XLX/2, pp. 17–36.

Toporišič, Jože (1976). *Slovenska slovnica*. Maribor: Založba Obzorja.

Verdonik, Darinka et al. (2023). *ASR database ARTUR 1.0 (audio)*. Slovenian language resource repository CLARIN.SI. URL: `http://hdl.handle.net/11356/1776`.

Appendix 1: Praatscript code used

```
clearinfo

list = Create Strings as file list: "List", "*.wav"

numberOfFiles = Get number of strings


appendInfoLine: "vowel 'tab$' F1 'tab$' F2"

for i from 1 to numberOfFiles
        selectObject: list
        fileName$ = Get string: i
sound = Read from file: fileName$
formant = To Formant (burg): 0, 5, 5500, 0.025, 50

        tfileName$ = replace$ (fileName$, ".wav", ".TextGrid",1)

textgrid = Read from file: tfileName$


        numberOfIntervals = Get number of intervals: 1 ; on the first tier


        for interval to numberOfIntervals
                selectObject: textgrid
                word$ = Get label of interval: 1, interval

                if word$ <> "" and word$ <> " "
                        startTime = Get start time of interval: 1, interval

                        endTime = Get end time of interval: 1, interval

                        middlePoint = startTime + ((endTime - startTime)/2)
                        selectObject:formant
                        fOne = Get value at time: 1, middlePoint, "Hertz", "Linear"
                        fTwo = Get value at time: 2, middlePoint, "Hertz", "Linear"
                        nfileName$ = replace$ (tfileName$, ".TextGrid", "", 1)
                        appendInfoLine: "'word$' 'tab$' 'fOne' 'tab$' 'fTwo'"
                endif
        endfor
endfor

select all
Remove
```