PHONETIC REDUCTION IN REGIONAL VARIANTS OF KAZAKH: A STUDY OF "ZH" OMISSION AND POSTVERBAL AFFIXATION ACROSS REGIONS

by

NAZERKE DOSKELDI

B.A., Academician Y. A. Buketov

Karaganda State University, 2016

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Master of Arts

Department of Linguistics

2025

Committee Members:

Hannah Haynie

Chase Wesley Raymond

Raichle Farrelly

Nazerke Doskeldi (MA, Linguistics)

Phonetic Reduction in Regional Variants of Kazakh: A Study of 'zh' Omission and Postverbal Affixation Across Regions

Thesis directed by Assistant Professor Hannah Haynie

Abstract

This study examines regional variation in the Kazakh language, specifically the phonetic reduction of the auxiliary verb "zhatyr" [ʒɑtɯr] in progressive constructions. While some people may assume minimal regional differences, this research questions that idea by studying how people in different parts of Kazakhstan pronounce and perceive "zhatyr."

The findings indicate that "zhatyr" reduction varies by region, though speaker perceptions of regional variation do not always align with actual linguistic data. The study explores sociolinguistic factors, specifically geographical location, age, and gender, that may influence phonetic reduction.

As one of the first studies to investigate this phenomenon in modern Kazakh, this research contributes to our understanding of grammaticalization, phonetic variation, and regional linguistic diversity.

Acknowledgments

In the name of Allah, the most gracious and the most merciful. First and foremost, I am thankful to Almighty Allah for giving me the strength, knowledge, and determination to be able to complete this study.

Secondly, I would like to express my deepest gratitude to my respected advisor, Dr. Hannah Haynie, for her invaluable guidance, support, and feedback throughout every stage of my research. I am especially thankful for the emotional encouragement she provided, which helped me believe in my ablity to accomplish something I once thought was impossible.

I thankfully acknowledge the support and encouragement of my committee members, Dr. Chase Wesley Raymond, and Dr. Raichle Farrelly. Also, I am grateful to Dr. Giorgia Troiani, postdoctoral scholar at Nazarbayev University, who helped me with corpus data.

I am also thankful to Nancy Capelle, who helped with editing and language support. Her careful attention to detail and thoughtful suggestions greatly contributed to the clarity of this thesis. I truly appreciate the time and care she dedicated to helping me to express my ideas more effectively.

On a personal note, I am very grateful to my parents and friends for their constant support and encouragement throughout this journey.

CONTENTS

CHAPTERS

TABLES

Table

FIGURES

Figure

Chapter I Introduction

This study investigates phonetic reduction in the Kazakh language, specifically in the progressive aspect auxiliary verb "zhatyr"[ʒɑtɯr]. The purpose of this study is to see if phonetic reduction of a grammaticalized form ("zh" [ʒ] omission in certain grammatical constructions) is becoming more common and whether there are any regional differences in adoption of this change.

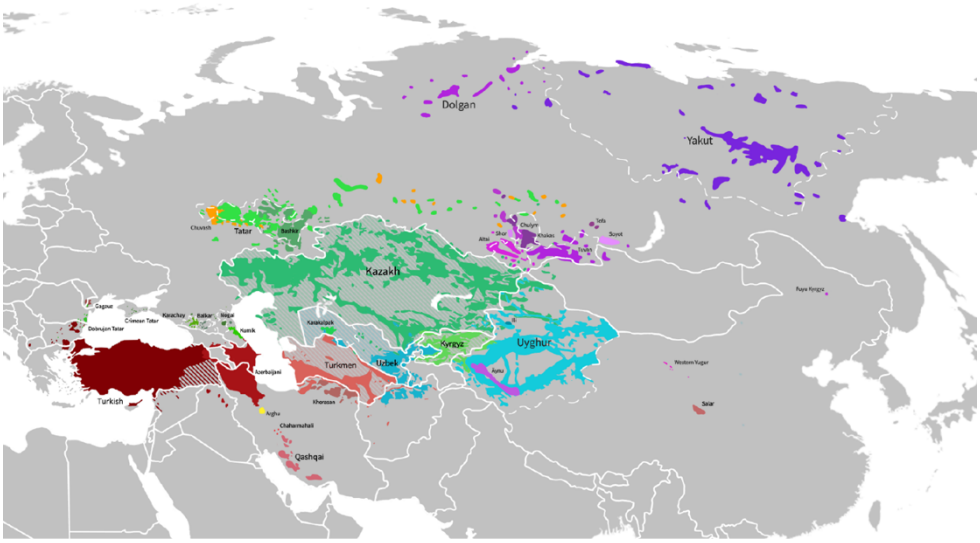
1.1 A Brief Overview of Kazakh and Its History

Kazakh is one of the members of the Turkic language family, belonging to the Kipchak branch, and serves as one of the official languages of Kazakhstan. It is also spoken in China (the Xinjiang Uyghur Autonomous Region and the Ili, Altay, and Tarbaghatay regions), Mongolia, and South Russia (Abish, 2021, Mukhamedova, 2015).

Turkic languages belong to the hypothesized Altaic language family, mainly spoken in Eastern Europe, the Middle East, Central Asia, and Siberia. The geography of Turkic languages is illustrated in the map below (Figure 1) (Turkic Languages, 2024). Turkey is the most populous country where a Turkic language is spoken, with around 82 million people (Alex, 2024). The most well-known Turkic languages are Turkish, Uzbek, Kazakh, Azerbaijani, and Turkmen. In addition to being spoken in independent nations and states, Turkic languages are also acknowledged in partially sovereign regions, such as Uyghur in China's Xinjiang Uyghur Autonomous Region and various republics within the Russian Federation, including the Chuvash Republic and the Tuva Republic (Györfi, 2022).

Kazakhstan's historical context within the Soviet Union (Union of Soviet Socialist Republics, or U.S.S.R.) has significantly influenced Kazakh language change and development. To avoid confusion with the Russian term “cossack,” the Kazakhs were referred to as Kara-Kirgiz, also Kirgiz-Kaisakskii, until the early twentieth century (Sauranbaev, 1955, Abish, 2021). During the Soviet period, Kazakh was not considered a state language. Instead, it remained the native language of Kazakhs. Due to inconsistent use of Russian terminology and a mixture of literary and colloquial Kazakh and Tatar, it failed to be codified at the beginning of the Soviet era. In addition, it is believed that Russian influenced Kazakh syntax, and extensive use of Russian calques and loanwords affected the Kazakh lexicon (Grenoble, 2003).

Figure 1. The geographic scope of Turkic languages[1]



[1] The map was taken from https://en.wikipedia.org/wiki/Turkic_languages#/media/File: Turkic_ Languages_distribution_map.png. The colors represent the different branches of the Turkic language family. Shades of green indicate Kipchak Turkic languages, such as Kazakh. Red signifies Oghuz Turkic, blue represents Turkestani (also known as Turki or Karluk), and purple Siberian Turkic (Györfi, 2022).

Most Kazakhs are fluent in at least two languages, a phenomenon that stems from Kazakhstan's political, economic, and social transformations during the Soviet era. These changes include the forced relocation of various groups (such as Germans and Koreans) during World War II, the Virgin Lands campaign in the postwar years, and the influx of industrial workers from different regions of the U.S.S.R., among other factors. Due to these influences, Russian affected the language development of Kazakh, which is still one of the consequences in the present day. The Soviet Russification language policy led to the marginalization of Kazakh. Russian has become widely spoken by nearly everyone, while Kazakh is often considered a "primitive rural language" reserved only for private, informal settings (Mukhamedova, 2015).

During the twentieth century, there were several orthographic changes in Kazakh. Until 1927, a modified Arabic script was used for Kazakh, which was then replaced by the Latin script. In 1940, the Cyrillic alphabet was adopted. After the dissolution of the Soviet Union, Kazakhstan's ethnic makeup shifted significantly, with many Russian speakers leaving the country. At the same time, the government launched a campaign to encourage Kazakhs from China, Mongolia, Turkey, Uzbekistan, and other nations to return to Kazakhstan (Mukhamedova, 2015).

Before the fall of the Soviet Union in 1991, most of the provinces of Kazakhstan were inhabited primarily by non-Kazakh speakers, specifically Russians and other Slavs, except for two provinces: Atyrau and Kyzylorda (Fierman, 2006). In the post-Soviet era, in accordance with Article 7 of the

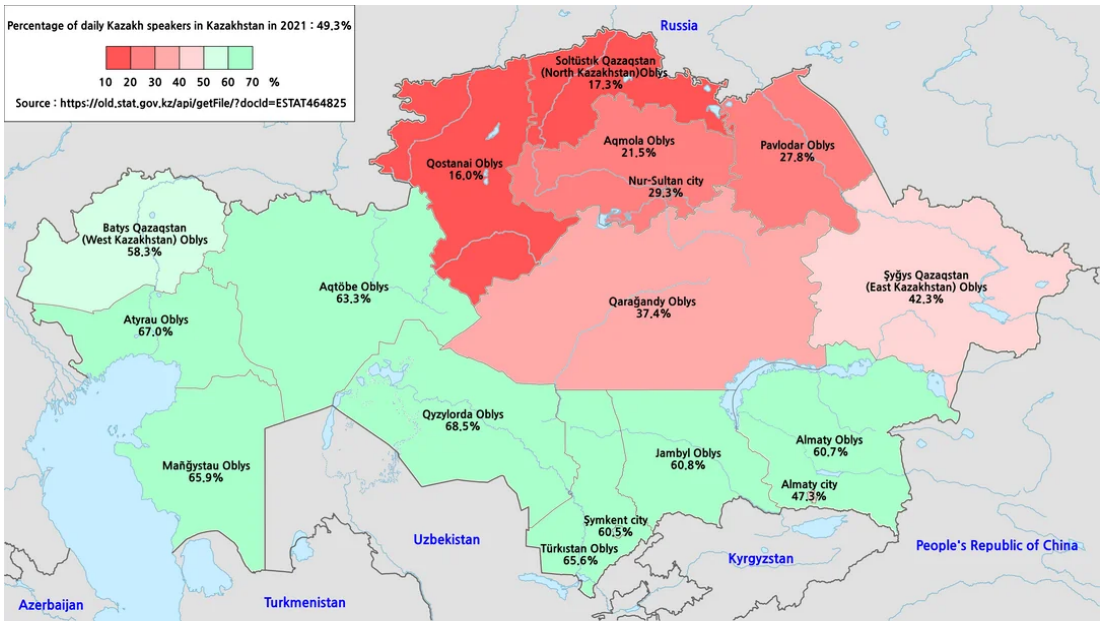
constitution of the Republic of Kazakhstan, the Kazakh language became the state language, and Russian became Kazakhstan's official language. Kaftan (2004) observes that the president of Kazakhstan traditionally addresses his population first in Kazakh and then in Russian, and this tradition is kept until now.

Kazakhstan is the ninth-largest country in the world in terms of land area, and it has a population of around 20 million, according to World Health Organization's 2023 data. In addition to the language differences, there are also geographic variations (Figure 2). The majority of ethnic Kazakhs live in the western and southern parts of Kazakhstan (more than 70%). Additionally, these regions share their borders with Turkmenistan, Uzbekistan, and Kyrgyzstan, and their languages also belong to the Turkic language family. In the Almaty and East Kazakhstan regions, this percentage seems to be lower but still shows a majority (50–70%) (Jarlhøj & Valijärvi, 2024, Smailov, 2011).

Figure 2. Percentages of ethnic Kazakhs by regions of Kazakhstan (Jarlhøj & Valijärvi, 2024)

Kazakhstan gained its independence in 1991; however, due to the influence of the Soviet Union, Russian remains the language of education, daily communication, and even international relations. It is mostly spoken in the northern part of Kazakhstan (Figure 3). In the North Kazakhstan and Kostanay (Qostanai) regions, Kazakh is spoken by less than 20% of the population. In contrast, many people in the western and southern parts of the country use Kazakh daily, with approximately 70% of the population speaking it. In the eastern part of the country, approximately 50% of the population speaks Kazakh. However, in the Akmola, Pavlodar and Karaganda (Qarağandy) regions, less than 40% of residents speak Kazakh.

Figure 3. Percentages of daily Kazakh speakers in Kazakhstan in 2021 ("Languages of Kazakhstan," 2025)



1.2 Motivation

Kazakhstan's membership in the U.S.S.R. and shared borders with neighboring countries have brought remarkable language development. In particular, there have been noticeable changes in spoken Kazakh. This research mainly studies an auxiliary verb change occurring in colloquial Kazakh, which is interesting as it is undergoing grammaticalization while also becoming a contemporary language trend in Kazakh language usage.

This is the first study written in English that demonstrates the perception and production of Kazakh phonological change, particularly in the progressive aspect auxiliary verb "zhatyr." In the early twentieth century, several Russian linguists discussed this change in their work. For instance, Katarinskii ([1906] 2013) mentioned that in the auxiliary verb "zhatyr," the sound "zh" [ʒ] drops before consonants. According to Sarybaev (2000), Russian scholar N. Ilminsky noted that the sound "zh" at the beginning of the auxiliary verb "zhatyr" may either be omitted or shift to a sound resembling "y" [j]. This change has also been mentioned in a recent Kazakh scholar's work. Mukhamedova (2015) stated that "Žat- is the most grammaticalized auxiliary: it fuses with the preceding verb in the converb form in spoken Kazakh and undergoes phonological reduction. For example, the sentence Ol žumïs istep žatïr ('S/he is working now') is uttered in spoken Kazakh as Ol žumïs istevatïr" (p.131). Mukhamedova (2015) also mentions that "žat- is the most semantically bleached; it loses its meaning ('lie') when used as an auxiliary" (p.131).

This phonetic reduction is discussed in a recent work by Györfi (2022), as illustrated in Example (1). He argues that while both forms are used, the shortened version is considered less formal and is typically employed in casual conversations.

(1) (Györfi, 2022)

a. ol žumïs iste-p žatïr

3SG work do-CVB.B AUX(IPFV 'lie').PRS[3]

'S/he is working.'

b. ol žumïs iste=vatïr

3SG work do=AUX(IPFV 'lie').PRS[3]

'S/he is working.'

As mentioned above, this phonetic reduction has been a central concern for several scholars. However, despite the extensive research on this topic, there remains a lack of studies addressing the social aspect of phonetic reduction and how speakers perceive it.

The dropping of the "zh" sound is anecdotally perceived as less polite, associated with street language, and considered a more uneducated way of speaking in contemporary Kazakh. Nevertheless, recently, a new song[2] was released in which the artist used colloquial Kazakh, incorporating more "zh"-dropping phrases in the chorus, garnering significant attention from listeners. Although some people still consider the use of the phonologically reduced form of this auxiliary verb

[2]dudeontheguitar, jeltoksan., Daiynball, Ruslan Tai — Obal oilar -ai (Musić Video) https://www.youtube.com/watch?v=djjl5dvLmS8

uneducated, it is increasingly becoming a widely used feature of spoken Kazakh grammar.

Given these ongoing social perceptions and the growing prominence of “zh” reduction in popular media, this study aims to explore several questions regarding the frequency and regional variations of this phonetic phenomenon and the extent to which speakers perceive it accurately. First, the study will investigate how often people reduce “zh” in the auxiliary verb “zhatyr,” specifically in the progressive aspect. Secondly, the research will examine whether the reduction is more common in certain regions by comparing data from participants in different parts of the Kazakh-speaking world. The study will identify patterns and variations in language use that may be influenced by geography or social factors. Lastly, the study aims to compare Kazakh speakers’ perceptions (regarding the reduction of the “zh” sound by Kazakh speakers across all regions of Kazakhstan) with actual linguistic data (collected by Nazarbayev University students) to investigate how speakers from different areas perceive this linguistic feature. By addressing these questions, this research aims to provide a clearer understanding of the role of phonetic reduction in a grammatical form in modern Kazakh and its usage and social implications.

Chapter II Review of the Literature

This study examines a process of grammaticalization in Kazakh, where a word or sound gradually evolves into a grammatical element. This phenomenon is common in other Turkic languages, though it progresses at different rates. A key focus is how phonological reduction develops within this grammaticalization pattern in Kazakh. Additionally, the study explores regional variations and how speakers perceive these ongoing changes.

2.1 Grammaticalization

One of the fundamental concepts of language change in linguistics is grammaticalization. It has been a heavily debated topic by scholars and can be found in well-known linguists' works, e.g. Antoine Meillet (1912), Traugott and Heine (1991), Hopper and Traugott (2003), and Lehmann (2015).

Grammaticalization was first introduced by well-known French Indo-Europeanist Antoine Meillet, who defined it as "the attribution of grammatical character to a formerly independent word" (1912, as cited in Campbell, 2013, p. 281), meaning that if the independent word becomes a grammatical element, which changes its original meaning, and if this process continues, it results in a bound grammatical marker. One of the widely referenced definitions of grammaticalization is by Kuryłowicz (1965). According to him (1965), "grammaticalization consists in the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical

status" (p. 69). This definition includes two phases of lexical induction and ongoing grammatical change (Aaron Smith, 2011).

As stated by Hopper and Traugott (2003), grammaticalization is "the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and once grammaticalized, continue to develop new grammatical functions" (p. 232). Grammaticalization is typically shown by gradual changes along a cline, which also can be thought of as a "continuum," which is demonstrated in Example (2) (Hopper & Traugott, 2003).

(2) content item > grammatical word > clitic > inflectional suffix (Hopper & Traugott, 2003)

Aaron Smith (2011) uses the terms primary and secondary grammaticalization to describe this cline. For primary grammaticalization, he (2011) covers the first step from content item to grammatical word, while secondary grammaticalization covers the following shifts.

The most commonly cited example of grammaticalization is the English auxiliary verb "will," which initially had the meaning of "want" in German. For example, phrases like "have the will" (meaning desire), "if you will" (meaning if you want to), and "good will" (meaning wishes, desire) will retain their older, more lexical meaning. However, it has semantically bleached (lost its original meaning) over time and grammaticalized into a future tense marker. Grammaticalized forms are often linked to "phonetic erosion," which implies that fuller forms are subject to phonological reduction. For instance, the morpheme "will," used as a future tense,

for example, “I will,” “she will,” can be constructed into forms like “I’ll,” “she’ll,” and so on (Campbell, 2013).

Campbell (2013) states that some scholars view grammaticalization not as a unique process but as involving other linguistic changes like sound change, semantic change, and reanalysis. However, most scholars agree that grammaticalization depends on mechanisms like reanalysis, which do not always encompass grammaticalization. Examples like changing the word order or affixes becoming independent words do not include the shift towards grammaticalization. Semantic bleaching is a standard semantic change, associated with grammaticalization, where lexical meaning is lost. On the other hand, new meanings can also emerge, such as the “future” meaning of “will” or “gonna.” For this reason, the semantic change in grammaticalization is a part of general semantic change and not independent of it (Campbell, 2013).

According to Early (2000):

> … languages variously encode a large range of human physical postures, such as ‘crouch,’ ‘lean,’ ‘hunch,’ ‘kneel,’ ‘tip-toe,’ ‘head-stand,’ ‘sit with both legs folded back to one side’ (e.g., Tongan fāite), ‘sit cross-legged’ (e.g., Tongan fakata ‘ane), etc. However, of all of the possible stances or body orientations that might be described, there are three that can be regarded as basic: namely, the concepts ‘sit,’ ‘stand’ and ‘lie.’ Some variation may lay behind these labels: in some cultures, the default ‘sitting’ posture is actually ‘squatting,’ and some languages may distinguish kinds of standing, e.g. of tall thin objects or of short squat objects (Early, 2000, p. 81).

Several studies have been done about postural verbs becoming grammaticalized. William, one of the scholars mentioned in Early (2000), surveyed

what happens to postural verbs when they are grammaticalized, and he noticed two different functions: they either become tense-aspect markers or copula verbs.

When postural verbs undergo grammaticalization, they often become tense-aspect markers, often expressing the progressive aspect. According to Early (2000), William cited some languages where this occurs; for example, in Hebrew, when the postural verb "stand" takes a verb, it denotes immediate future. In Dutch (Example (3)), postural verbs plus the main verb would give the progressive aspect. In languages like Siouan, Yuman, Uyghur (Turkic), Hindi, and Nepali, postural verbs act as aspect markers.

(3) (Early, 2000, p. 98: Dutch).

lk zit een boek te lezen

I sit a book to read

'I am reading a book.'

The second function that postural verbs take on is as a copula. According to Early (2000), William suggested that postural verbs may either "compete with the copula" (as seen in the Russian example (4) below) or gradually become a copula. In these examples, the postural verbs when functioning as a copula may either keep their original meaning related to physical orientation, or they may be diminished and the verbs instead convey ideas of existence, equality, or a location. In Early's work, one example that is cited by William is

(4) (Early, 2000, p. 98: Russian).

Butylka stoit na stole

bottle stand on table

'The bottle is on the table' (in the unmarked standing orientation).

2.1.1 Grammaticalization in Turkic Languages

Having established the general framework for grammaticalization above, this section illustrates some instances of grammaticalization in Turkic languages, specifically Turkish. Turkish is the most widely spoken language within the Oghuz sub-branch of the Turkic language family. Modern standard Turkish is derived from Ottoman Turkish, developed on the basis of the Istanbul dialect. Among all Turkic languages, Turkish is the most documented language (Csató & Johanson, 2021).

Schiering (2006) examined two cases of grammaticalization in Turkish, pronouns which resulted in subject agreement (Givon, 1976) and auxiliaries which resulted in aspect-tense-mood markers (Lehmann, 2015). The second case more closely resembles what this study is investigating in Kazakh, which is auxiliaries becoming grammaticalized. Schiering (2006) illustrated several examples (Examples (5), (6), and (7)) from Old Turkish to compare with modern Turkish in which converb constructions have been grammaticalized to convey actionality, intention, ability, and version.

(5) geli yür, 'to be coming' (yüri, 'to go')

(6) kör-ü bil, 'to know how to obey' (bil, 'to know')

(7) alta-yu tur, 'to keep cheating' (tur, 'to stand')

In modern Turkish, this construction became more grammaticalized, as demonstrated in Examples (8), (9), and (10). Schiering (2006) indicates that the "progressive" -Iyor (8) and "potential" -(y) Abil are considered as disyllabic suffixes (Schiering, 2006, pp. 7–8).

(8) gel-íyor-um, 'I am coming' (yor, 'to go')

(9) gel-ébil-ir-im, 'I can come' (bil, 'to know')

(10) söylen-é-dur-ur, 'he keeps grumbling' (dur, 'to stand')

As we can see from the information above, what is happening in Kazakh looks a lot like the process of grammaticalization in Turkish.

The Kazakh language is one of the most grammatically complex among Turkic languages. It is agglutinative like other Turkic languages; it adds suffixes to the roots of words to build a sentence. The primary phonological rules that contribute to forming these structures are vowel harmony and consonant assimilation (Batayeva 2013). Kazakh also uses the progressive aspect to talk about ongoing action. However, the way it forms this construction is slightly different. To form the progressive aspect, Kazakh adds one of the converbs (CVB) -yp, -ip, -p [-ɯp,-ɪp,-p] to the main verb, depending on vowel harmony, and one of four different postural verbs: zhatu [ʒɑtu] 'to lie down,' turu [tʊru] 'to stand,' otyru [otɯru] 'to sit,' and zhuru [ʒʏru] 'to walk.' Example (11) shows how it is shortened in spoken Kazakh.

(11) a. men kitap ok-yp zhatyrmyn

1SG book read-CVB.B AUX(IPFV 'lie').PRS[1]

'I'm reading a book.'

b. men kitap oky=vatyrm

1SG book read=AUX(IPFV 'lie').PRS[1]

'I'm reading a book.'

Among all these postural verbs, "zhatyr" is becoming the most suffix-like. As mentioned above, according to many scholars, functional words lose their meaning, and here, the Kazakh postural verb "zhatyr" has become a common pathway of grammaticalization. In another relevant study, scholars have identified a number of language areas such as semantics, syntax, phonology and morphosyntactic processes that interact in the grammaticalization of morphemes and entire constructions (Heine & Reh, 1984, Lehmann, 2015). What is happening in contemporary Kazakh is phonological reduction, which is also addressed by other scholars, like Campbell. He (2013) finds that phonological change refers to all types of change in pronunciation (Campbell, 2013).

Words that are undergoing grammaticalization tend to undergo phonological reduction (Aaron Smith, 2011). For this reason, this study analyzes the postural verb "zhatyr" in Kazakh as undergoing grammaticalization. Heine and Kuteva point out that "grammaticalization requires a linguistic system that is used regularly and frequently within a community of speakers and is passed on from one group of speakers to another" (Heine & Kuteva, 2007, p. 164). Based on the results of this

study, we can predict that phonetic reduction in Kazakh is becoming widespread in all parts of Kazakhstan.

2.2 Dialectology

2.2.1 Kazakh Dialectology

Kazakh dialectology has been extensively studied by both Russian and Kazakh scholars. Notable Russian researchers of the past include N. I. Ilminski, V. V. Katarinskii, and V. V. Radlov, while prominent Kazakh scholars such as G. Kaliev, S. Amanzholov, N. Sauranbaev, and S. S. Sarybaev also made significant contributions. In particular, S. S. Sarybaev played a crucial role in the development of the field (Amanzholov, 1953, Kaliev & Sarybaev, 1991, Medetova, 2016).

Although there are relatively few dialectal differences, it is commonly stated that there are greater variations in vocabulary, grammar, and phonetics across regions (Abish, 2021, Schwarz, 1984, Sauranbaev, 1955, Vajda,1994). According to Abish (2021), there are four major dialects, a central, a southern, a western and an eastern dialect. However, Jarlhøj and Valijärvi (2024) present a map that divides the whole country into five major dialect regions.

Amanzholov (1953) states that the northeastern dialect became the foundation of the literary language. He (1953) provides two reasons for this; first this happened mainly as Russians spread more quickly in the northeastern areas starting from the second half of the nineteenth century; second, the connection between Kazakh workers and Russian settlers, which began in the sixteenth century, became stronger during this time, and Kazakh youth became more

interested in Russian culture. Russian cities in Western Siberia played a big role in shaping the Kazakh intelligentsia (Amanzholov, 1953).

Schwarz (1984) indicates that Kazakh dialectal differences are based on old tribal divisions, and illustrates Kazakh dialects spoken in China. Sauranbaev (1955) also states that the establishment of dialects is related not only to defining the boundaries of their distributions, but also to understanding the very complex process of settlement of tribal groups. He (1955) explains that since ancient times, long before the seventeenth century, Kazakh tribal groups, due to their nomadic lifestyle, frequent raids, enslavement by the Zhungar, Kokan, and Khiva khanates, as well as a number of other reasons, did not have clearly defined settlement boundaries. Even in the nineteenth century, before Russia took control of Kazakhstan, different clans could live together in the same village or area, or within the same family. Due to this fact, it was historically difficult to match dialectal differences with specific settlement areas (Kenesbaev & Musabaev, 1975).

2.2.2 Influence of Other Languages on Kazakh

Barnes (1954) first introduced the concept of a social network as an analytical tool in the Norwegian village of Bremnes. According to him (1954), understanding social behavior requires more than concepts based on social status, territorial position, or economic status. He argues that interpersonal connection structures would describe social behavior better. Following this idea, Lesley Milroy (1980) states that "the term social network refers quite simply to the informal social relationships constructed by an individual" (Milroy, 1980, p. 178). Türker (1995)

says the main focus is individuals' relationships with their family, friends, and coworkers. This idea helps us to understand a person's social network and how they connect to their community. He also indicates that every individual is considered the core of their network, and the people who interact with them will also have relationships with other groups. This network thus becomes multiplex (high density). Although the people in the network use the same language to interact with each other, they may come from other language backgrounds, which can influence the way that language is used in the community. Considering this theory, diverse ethnicities in Kazakhstan have some influence in the development of the Kazakh language. An example would be the Uzbek people moving to some parts of Kazakhstan, like Shymkent, Almaty, which is in the southern part of the country, and the Zhambyl region. Additionally, there are many Russian minority groups, and also every year people from Mongolia and China migrate to Kazakhstan. Particularly, the southern and northern regions have a greater impact on social networks due to shared borders and frequent social interactions.

The Kazakh language differs from other Turkic languages in terms of phonetics, primarily due to the vowel harmony rules and the assimilation of consonants based on the voicing and voicelessness of neighboring consonants. For instance, kizdar (girls), tastar (stones); in the Nogai and Karakalpak languages, it would be kizlar, taslar. Kazakh also consistently uses "zh" at the beginning of words instead of "y," "s" at the end of the word instead of "sh," and "u" instead of "f"; for example, "zhol" (road) in Kazakh would be "yol" in Uzbek, bas (head) in Kazakh

would be “bash” in Uzbek, and “tau” (mountain) in Kazakh would be “taf” in Uyghur (Kenesbaev & Musabaev, 1975). People from the south part of Kazakhstan tend to use the “y” [j] sound at the beginning of the progressive aspect auxiliary verb “zhatyr,” as in Example (12), also in deixis, such as “anayak” instead of “anazhak” (there). This would indicate that the migration of Uzbek people to the southern region of Kazakhstan and shared southern border with other Turkic countries might have affected where these phenomena occur.

(12) a. men kel-e zhatyrmyn

1SG come-CVB.B AUX(IPFV ‘lie’).PRS[1]

‘I’m coming.’

b. men kele=yatyrm

1SG come=AUX(IPFV ‘lie’).PRS[1]

‘I’m coming.’

Originally, most words in Kazakh derived from Arabic, Persian, Chinese, or Mongolian; however, the most recent loanwords are chiefly from Russian. Mongolian and Chinese lexical borrowings remain limited. The Kazakh lexicon is more heavily influenced by Arabic and Persian (Vajda, 1994). Among all languages, Russian has had the greatest influence on the development of Kazakh. Khasanov (1987) mentioned that at the end of the nineteenth century and beginning of the twentieth century, one out of every 22 or 23 (4.4%) words was adopted by Kazakhs through the Russian language.

Since Kazakh literacy was low, Russian words were borrowed through spoken language. As these borrowings became subject to phonetic assimilation, they were adapted to the rule of Kazakh phonology. For instance, the sound "v" was absent in Kazakh. For this reason, the "v" in Russian loanwords became an "u" [ʊ], such that "krovat" [krɑvɑt] in Russian (bed) becomes "kereuet" [kereʊet] in Kazakh; "zavod" [zɑvɔd] in Russian (factory) becomes "zauyt" [zɑʊɯt] in Kazakh (Vajda, 1994).

2.2.3 Regional Speech Variations

There is a remarkable number of scholars who have done some research about regional lexical dialects. One of them is Amanzholov (1953); in his book, he illustrates the main features of the western Kazakh dialects. One of the examples he (1953) provides is the auxiliary verb derived from "zhatyr," which is pronounced as "zhatyrgan" [ʒɑtɯrɣɑn] in the western region of Kazakhstan. The literary standard recognizes the auxiliary "zhatyr."

Nakysbekov (1982) studied the southern dialects of Kazakh. He (1982) provided many examples of lexical, grammatical, and phonetic variations in Kazakh across the southern regions, namely Shymkent, Kyzylorda, and Zhetisu. One of the observed speech variations is "zh" becoming "y" [j] within a pair of words, in a process of deixis. For example, in "kay zhak" [qai ʒɑq] (where?), "bul zhak" [bʊl ʒɑq] (here), and "sol zher" [sɔl ʒɛʁ] (there), the "zh" sound at the beginning of the second word becomes "y" [j]. This is considered a distinctive feature of the southern Kazakh dialect. This is again, as mentioned above, because of Kazakh language contact with

some Turkic languages, especially Uzbek, spoken in countries sharing the southern border.

Vajda (1994) emphasizes regional phonological differences, such as that western and northern Kazakh speakers use the alveopalatal fricative [ʃ], while eastern and southern speakers replace it with the affricate [tʃ]. For example, “shalgy” [ʃɑlʁə] or “chalgy” [tʃɑlʁə] (braid) and “anshi” [ænʃɪ] or “anchi” [æntʃɪ] (singer). “The literary standard recognizes [ʃ]” (p. 610). He (1994) also mentions the different production of the non-initial [d] sound in the western and northern regions of Kazakhstan, that is pronounced like [l] in the southern and eastern regions. For example, “tiridei” [tɪɾɪdjɪ͡j] or [tɪɾɪ1jɪ͡j] (alive) and “mandai” [mandaj] or [manlaj] (forehead). “Once again, the northern and western variant [d] is considered standard literary pronunciation. A few other sound variations occur, but the non-literary variants are extremely localized” (Vajda, 1994, p. 610).

2.3 Perception and Production of Kazakh

The perception and production of linguistic input are fundamental aspects of language behavior. Language evolves over time, which can lead to variations across different regions. Perception and production contribute to these changes and continue to develop in response to them. To understand how people use language in various social settings and environments, it is essential to examine both how they speak and how they perceive language (Campbell-Kibler, 2010).

Lambert et al. (1960) state that spoken language is a key characteristic of a cultural group and that their feelings toward the language come from how they feel

about the members of the group (Lambert et al., 1960). When we hear a language, we often tend to associate it with certain groups, and our reactions are usually based on the general stereotypes about them. For instance, in Kazakhstan, if someone speaks only Kazakh without mixing in Russian or other languages, they are often perceived as being from rural areas or the southern regions of Kazakhstan. This is because people in cities tend to use more Russian and are generally considered more educated.

The phonological reduction in Kazakh mentioned above has often been associated with street language or a lack of education. However, people do not always perceive the linguistic changes they themselves produce. To explore this gap between perception and actual usage, spoken language data collected by Nazarbayev University students and obtained from the surveys conducted for this study are compared.

Chapter III Methodology

The data for this study come from two sources. The first source is a corpus collected by Nazarbayev University students (described in 3.1.1), and the second source is a survey which I carried out to collect information on Kazakh speakers' perceptions of phonetic reduction and on the sociolinguistic context in which it happens (described in 3.2). I analyzed trends in each data source to understand the production and perception of this change that is in progress in Kazakh.

3.1 Research Design

This study applied a mixed-methods approach to investigating phonetic reduction in a progressive aspect auxiliary verb in spoken Kazakh. The data-collection process is divided into two primary components: a corpus data analysis and a survey. The corpus provides a rich source of audio and video recordings of everyday conversations from diverse groups of speakers across regions, while the survey gathers different participants' perspectives on "zh" omission in the progressive aspect auxiliary verbs in spoken Kazakh across all regions of Kazakhstan (Figure 4). Together, these methods allow for a comprehensive understanding of zh-dropping in spoken Kazakh, and of whether or not this is common only in specific regions or now becoming widespread across all regions. Combining both qualitative and quantitative data allows this study to investigate actual usage as well as language attitudes and strengthens the validity of the findings.

Figure 4. Kazakhstan regions (provinces) and their capitals as of 8 June 2022

3.1.1 Corpus Data Description

To investigate the first two research questions – namely, the frequency of "zh" being dropped in daily life and regional variations in its omission, a corpus-based data-collection approach was implemented (Filchenko et al., 2023, 2024). The corpus data was filtered to isolate the progressive aspect constructions in the Kazakh language, with the main auxiliary verb studied being "zhatyr," which means "lying down." The majority of data was sourced from 2023, with one additional source from 2024 included due to a lack of participants in 2023 for the Almaty region. In total, 244 tokens were recorded on a spreadsheet, indicating whether the sound "zh" was dropped or not.

3.1.2 Participants

A total of 88 informants participated in this corpus data. The age range was from 16 to 80 years old, and speakers came from 12 provinces of Kazakhstan. Participants came from diverse socioeconomic classes and education levels. The corpus included Kazakh people who were born in China, Karakalpakstan, and Uzbekistan, but living in Kazakhstan. However, there was not any specific information about when they immigrated to Kazakhstan. Participants were from various provinces across Kazakhstan, with a particular predominance of individuals from East Kazakhstan. Table 1 shows the number of people who participated across different provinces of Kazakhstan, and Table 2 shows the number of participants in macro regions of Kazakhstan.

Table 1. Corpus data participants' regional affiliations

Province	Number of Participants
Abai	2
Akmola	6
Almaty	4
East Kazakhstan	20
Jetisu	1
Karaganda	5
Kyzylorda	5
Mangystau	3
North Kazakhstan	3
Pavlodar	3
Turkistan	5
West Kazakhstan	6

Table 2. Corpus data participant demographics in five macro regions

Regions	Participants
North Kazakhstan	12
South Kazakhstan	15
East Kazakhstan	22
West Kazakhstan	9
Central Kazakhstan	5

3.1.3 Data Collection

The corpus data for this study was obtained from Nazarbayev University students' Multimedia Corpus of Modern Spoken Kazakh Language project (Filchenko et al., 2023, 2024). These scholars collected audio and video recordings of everyday conversations with people from all across Kazakhstan. The corpus incorporates various provincial groups, maintaining a diverse representation of speech patterns and language usage. This dataset contained certain formal discussions, which were excluded for the purposes of this research.

The speech contexts which were natural were used for analysis to capture diverse language use; for example, participants were recorded during mealtimes, in casual conversations on the street, while hosting guests, and in interactions among friends at school. These settings were specifically chosen to reflect everyday communication and ensure various conversational environments. Because this corpus was explicitly designed to capture colloquial speech, rather than written or formal communication, it is a useful source of information about a grammaticalization pattern that has been noted to occur in colloquial or informal

speech. The length of the conversations varied, with some lasting ten minutes and others extending to an hour.

3.1.4 Corpus Data Processing

A spreadsheet containing participant information was provided by Nazarbayev University students, along with their files, which included audio and video recordings. These files were sorted into a new folder, with any files lacking audio or video excluded from the data. Additionally, the Google Earth Pro application was used to map the locations of participants across all regions. This is illustrated in Figure 5, which shows the geographical distribution of participants across Kazakhstan, with stars indicating the locations of data points from different provinces.

Figure 5. The geographical distribution of corpus participants across Kazakhstan

Each participant's conversation was carefully listened to, and all progressive aspect "zhatyr" constructions were transcribed in a data table, along with a field

representing information about phonological reduction of the initial “zh.” Each token was associated in this table with the name of the pertinent file, the points in the conversation when “zh” was dropped or retained, the participant’s age, and their location.

3.1.5 Corpus Data Limitations

As this corpus is one of the earliest collections compiled by students of Nazarbayev University, certain limitations are present. Notably, the dataset includes only the birthplaces of individuals, which does not necessarily indicate where they have resided for most of their lives—information that would provide a more accurate geographic representation for this study. Additionally, some regions are underrepresented due to an insufficient number of participants.

Another limitation pertains to formality; when speakers were aware that they were being recorded, they tended to adopt a more formal tone, which affected the number of participants to be included in the study. Furthermore, speaker identification posed a challenge while analyzing the recordings. In some cases, two speakers of the same age but from different regions were present, making it difficult to distinguish between them. Alternative methods for more precise speaker identification should be considered in future research.

3.2 Introduction to the Survey

The primary purpose of the survey was to investigate the perception of zh-dropping across different regions. This aimed to understand how speakers in various regions perceive this sound change. In Kazakh dialectology, scholars such

as Sarybaev (2000), Katarinskii ([1912], 2013), Mukhamedova (2015), and Györfi (2022) have discussed this sound change. However, there has been little research on regional differences and how people perceive it. Therefore, this survey aims to explore this topic in greater depth and provide more insight into this phonetic change.

3.2.1 Participants

A total of 214 informants participated in this survey. Participants were recruited from all provinces in Kazakhstan through social media platforms, including Facebook, Instagram, and WhatsApp. The participants' ages ranged from 18 to 70. A detailed number of the participants by province is provided in Table 3, and Table 4 shows the numbers of participants in macro regions. Because the survey was not administered in conjunction with corpus creation, different individuals responded to the survey than are represented in the corpus, though both datasets cover the same regions of Kazakhstan.

Table 3. Survey participant demographics

Province	Number of Participants
Abai	27
Akmola	28
Almaty	25
Aktobe	6
Atyrau	18
East Kazakhstan	9
Jambul	4
Jetisu	5
Karaganda	23
Kostanay	4
Kyzylorda	9
Mangystau	5
North Kazakhstan	12
Pavlodar	2
Turkistan	16
Ulytau	7
West Kazakhstan	8

Table 4. Survey participant demographics in five macro regions

Regions	Participants
North Kazakhstan	46
South Kazakhstan	59
East Kazakhstan	36
West Kazakhstan	37
Central Kazakhstan	30

3.2.2 Survey Design

This study applies a structured survey to investigate regional variations in sound change, dialectal differences, and perceptions of speech patterns in the Kazakh language. It was designed to gather both demographic information and respondents' perceptions regarding the phenomenon of zh-dropping in everyday spoken Kazakh. The data for this survey were collected through a 23-question survey administered via Qualtrics. It contains multiple-choice, open-ended, Likert-scale, scale-based, and mapping questions.

The survey was divided into the following four main sections, each one addressing a specific aspect of the study.

Demographic Information

The first section focused on gathering demographic information. Participants were asked to provide details about their place of birth, region, age, and gender. These questions were designed to help address the research question concerning regional variations in language use and to explore potential age- and gender-based differences in the perception and use of the "zh" sound.

Language Usage in Informal Settings

The second section of the survey presented participants with various phrases commonly used in casual conversation, in which participants were asked to select the phrase they would most likely use in everyday conversations with friends and family. These questions attempted to investigate participants' perceptions of their own use of these phrases in everyday life, focusing on their own self-reported usage of zh-dropping in informal settings.

Social and Regional Perceptions of Language Variation

In the third section, participants were asked to respond to multiple-choice questions regarding their opinions on the factors influencing the omission of the "zh" sound in the auxiliary verb "zhatyr." Specifically, they were asked whether they believed the phenomenon was related to regional variations, lack of education, the identity of the conversational partner, or the prevalence of specific stereotypes associated with the dropping of the "zh" sound. This section explored participants' understanding of the social and cultural factors contributing to the perceived linguistic shift.

Mapping Regional Speech Patterns

The final section of the survey included a map of Kazakhstan, and participants were asked to rate on a scale from 0 to 100 the frequency with which they believed zh-dropping occurred in different regions. Participants were also asked to indicate the regions they most closely associated with their own speech patterns, providing insight into their regional identity and linguistic affiliation.

This survey also included several optional open-ended questions, where participants were invited to reflect on their personal experiences with language judgments that had impacted their interactions. Furthermore, participants were asked whether regional dialects influenced their daily conversations and to elaborate on how such influences might manifest in different social contexts.

By incorporating both structured and open-ended questions, the survey aimed to provide a comprehensive understanding of the social perceptions and regional differences related to the dropping of the “zh” sound in Kazakh and the potential underlying sociolinguistic factors shaping this phenomenon.

3.2.3 Survey Limitations

This survey provides an in-depth investigation into the omission of “zh” in Kazakh; however, certain limitations must be acknowledged. A significant challenge was that the majority of participants struggled with the scale-based questions, in which they were required to rate zh-dropping for 17 separate regions, on a sliding 0–100 scale. This difficulty may be attributed to the change in the number of provinces from 14 to 17 in 2022, as well as to the possibility that many participants were not familiar with the speech patterns of individuals from different regions. Due to this limitation, fewer participants responded to the regional-prevalence questions compared to other sections of the survey.

3.3 Analysis

3.3.1 Corpus Data Analysis

This study examined the relative frequency of "zh" presence versus "zh" absence in the progressive aspect auxiliary verb "zhatyr" construction. To analyze this variation, the below graphs were created for each province, displaying the relative frequencies of "zh" omission. Additionally, overall statistics summarized on one graph show these differences across the entire corpus data.

All 17 provinces were categorized into five macro regions: North, South, East, West, and Central. Additionally, comparative analysis was conducted to demonstrate the average production level of zh-dropping across these five regions.

3.3.2 Survey Analysis

The survey analyzed the perceived relative frequency of self-reported use and perception of "zh" presence versus "zh" absence variants in the progressive aspect auxiliary verb "zhatyr" construction. To analyze the scale-based questions about perceived regional prevalence of zh-dropping, the average perceived level of zh-dropping was calculated for each province, and then aggregated into larger categories representing five macro geographic regions.

To understand how people perceive their own use of the progressive aspect auxiliary verb "zhatyr" in speech, specifically, whether they believe they omit "zh" or not, multiple-choice questions were analyzed. Participants were asked to select the variant of a phrase they believe they commonly use in daily life. A graph was created to summarize the responses, illustrating whether individuals perceive

themselves as dropping “zh” or preserving it. These individual patterns were aggregated by province to assess regional differences in self-reported phonological reduction in the progressive aspect auxiliary verb “zhatyr” constructions.

Likert-scale questions were used to capture information about what sociolinguistic variables Kazakh speakers associate with “zh”dropping. To analyze the Likert-scale questions, tables were created for each question to indicate how often participants perceived certain stereotypes about zh-dropping as aligning with their views. Open-ended responses were coded to capture the main ideas of each participant’s view.

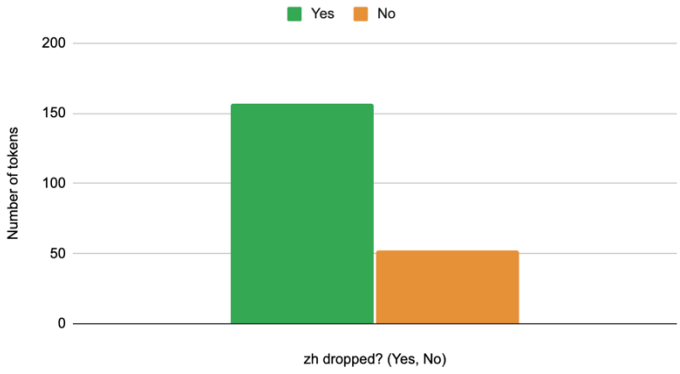
Chapter IV Results and Discussion

4.1 Corpus Data Results

In this section, the results of the analysis of zh-dropping in the corpus data are presented. The total number of tokens is 213, with some data excluded due to the evident use of a formal register. The first part shows the results of the full corpus, followed by regional breakdowns to highlight the differences across all regions.

Figure 6 shows the zh-dropping results across the full corpus, presenting the overall frequency of zh-dropping across the entire dataset. As seen in the graph, the number of tokens is 213 and zh-dropping occurs in the full corpus in 157 instances.

Figure 6. Zh-dropping across full corpus



The following figures are grouped by major geographic divisions (north, south, east, west, and central). Two regions of East Kazakhstan are presented in Figure 7 (Abai and East Kazakhstan). According to the results, participants from the Abai region (about 16 tokens out of 20) tend to drop the “zh” sound more than participants from the East Kazakhstan region (about 3 tokens out of 24).

Figure 7. Zh-dropping in East Kazakhstan regions

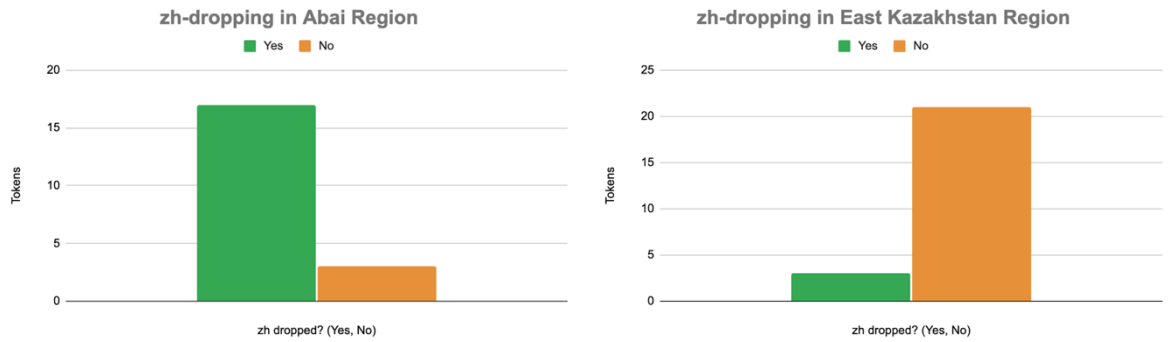


Figure 8 illustrates the results from the North Kazakhstan regions, specifically participants from the Akmola and Pavlodar regions. As shown in the graph, both regions frequently drop the “zh” sound. Since the number of tokens is not equal, the differences in proportions are difficult to compare, but the overall trends show high prevalence of zh-dropping.

Figure 8. Zh-dropping in North Kazakhstan regions

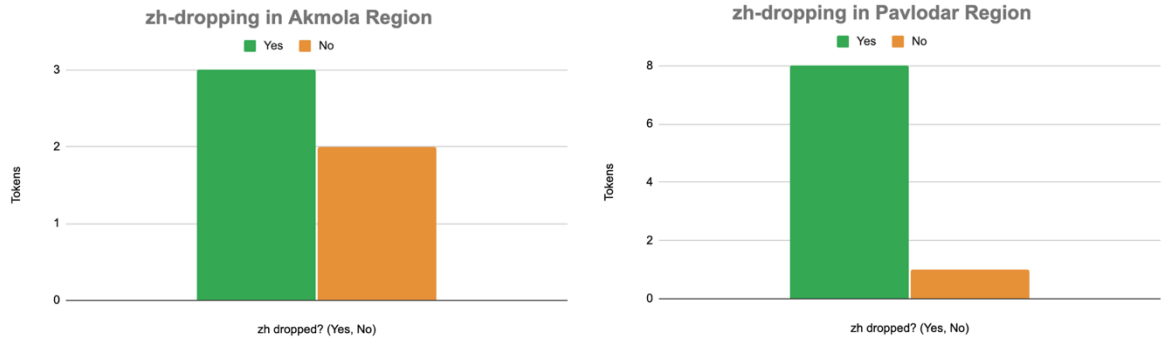


Figure 9 presents the results of the Karagandy region, one of the two regions (along with Ulytau) in Central Kazakhstan. Based on the results presented in the graph, all participants dropped the “zh” sound. There were 17 tokens in total, and none of them retained the “zh” sound.

Figure 9. Zh-dropping in Central Kazakhstan regions

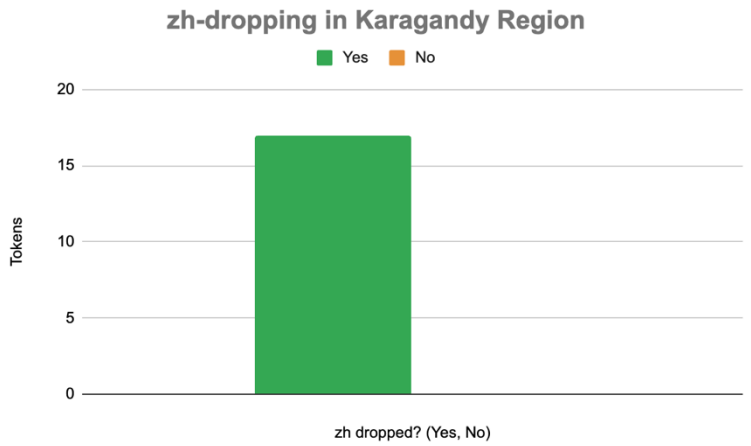


Figure 10 demonstrates the results from the South Kazakhstan regions; this corpus collection includes participants from the Turkistan, Kyzylorda, and Almaty regions. Among all three regions, participants from the Kyzylorda region dropped the “zh” sound most frequently. The total number of tokens was five, and none of them retained the “zh” sound. The zh-dropping was frequent in both the Almaty and Turkistan regions; however, the participants in the Turkistan region seemed to drop the “zh” sound less frequently (13 out of 27).

Figure 10. Zh-dropping in South Kazakhstan regions

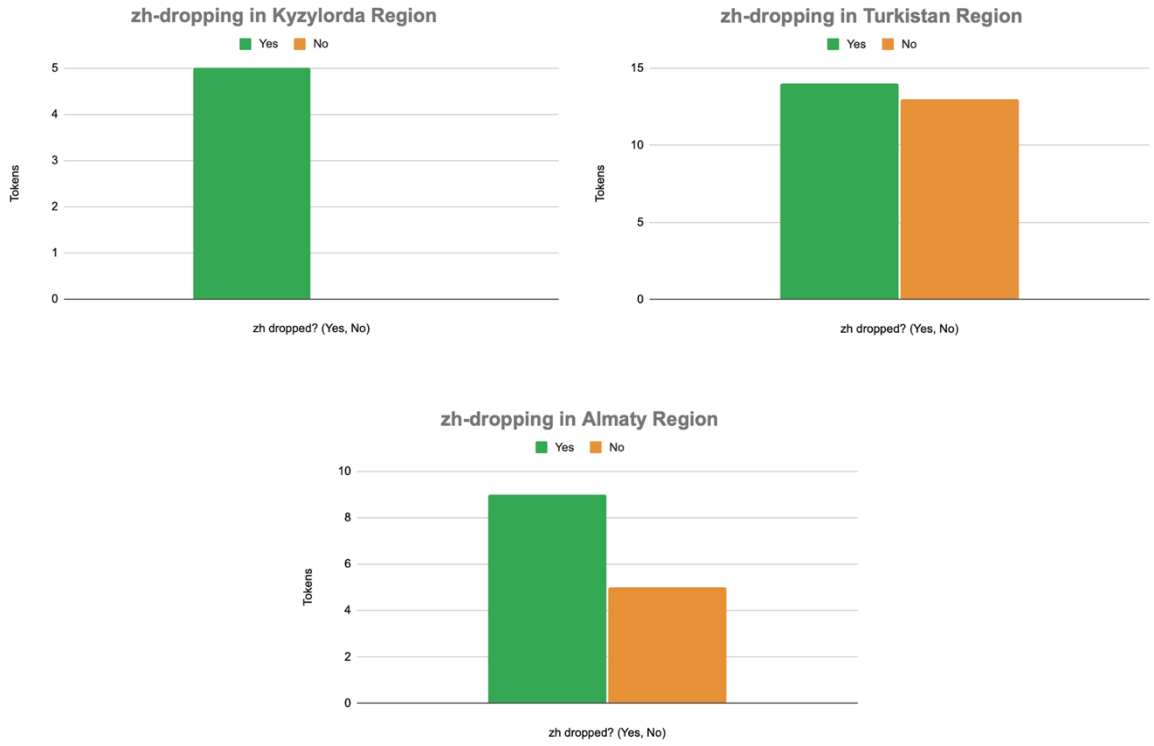
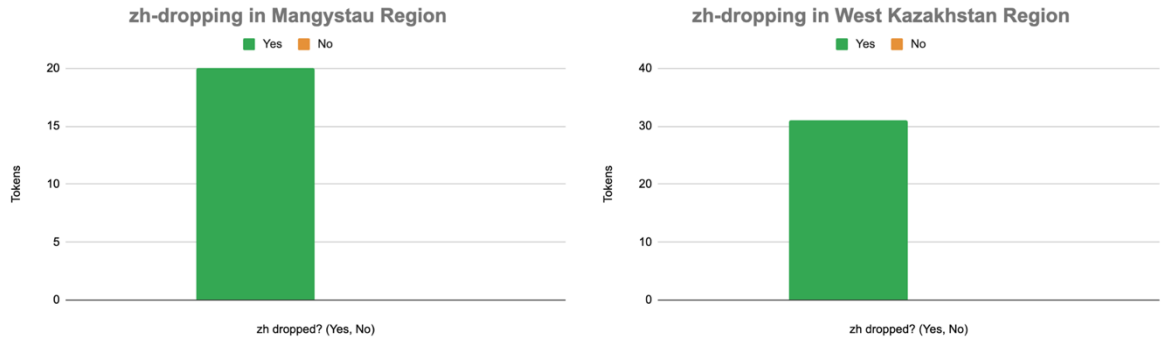


Figure 11 shows the results from the West Kazakhstan regions, particularly the Mangystau and West Kazakhstan regions. There were between 20 and 31 tokens presented and none of them retained the “zh” sound.

Figure 11. Zh-dropping in West Kazakhstan regions

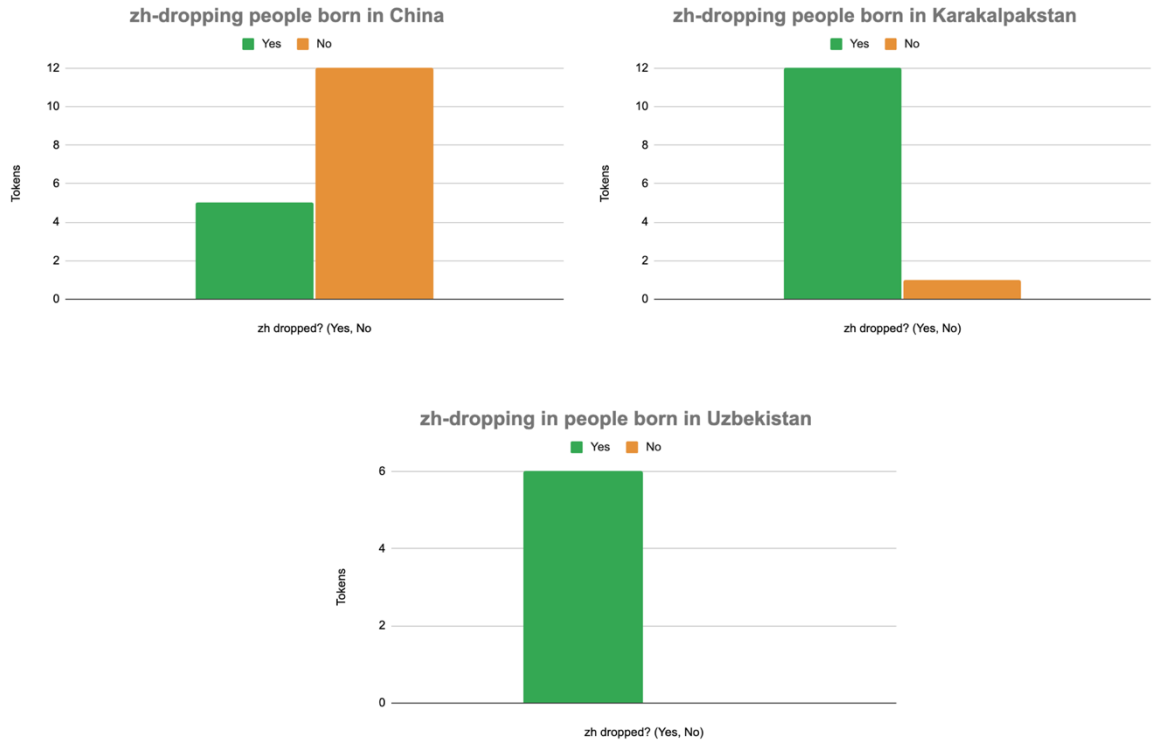


This corpus also contains Kazakh participants who were born in China, Uzbekistan, and Karakalpakstan, an autonomous republic within Uzbekistan. The

language of Karakalpakstan is Karakalpak; it is similar to Kazakh and is part of "a South Kipchak or Aralo Caspian Turkic language" (Abish, 2021, p. 352). "Karakalpak is so closely related to Kazakh that it can linguistically be considered a slightly Uzbekized variant of it" (Abish, 2021, p. 352).

Figure 12 presents the results of all participants who immigrated to Kazakhstan from China, Uzbekistan, and Karakalpakstan. Based on the results, people who were born in Uzbekistan and Karakalpakstan tend to drop "zh" more frequently compared to those who were born in China. 17 tokens were recorded, and participants who were born in China dropped the "zh" sound five times. There were six tokens in total among participants who were born in Uzbekistan, and none of the tokens retained the "zh" sound. Among participants who were born in Karakalpakstan, a single token represents the "zh" omission.

Figure 12. Zh-dropping in Kazakh speakers who were born outside of Kazakhstan



The following Figure 13 illustrates the prevalence of zh-dropping among individual participants across the entire corpus data. Notably, some participants from Almaty, China, East Kazakhstan, and Turkistan appear to have consistently retained the “zh” sound. On the other hand, participants from the rest of the regions shown dropped the “zh” sound more frequently.

Figure 13. Prevalence of zh-dropping by individual participants

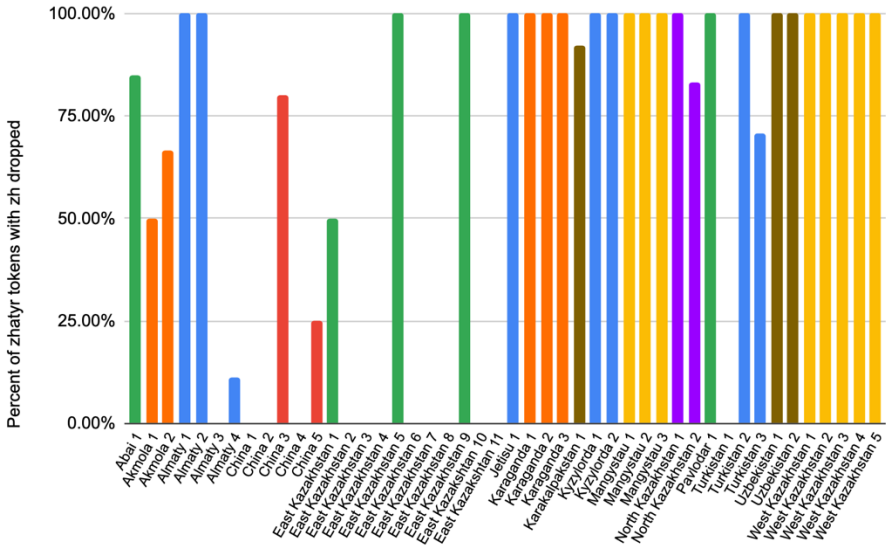
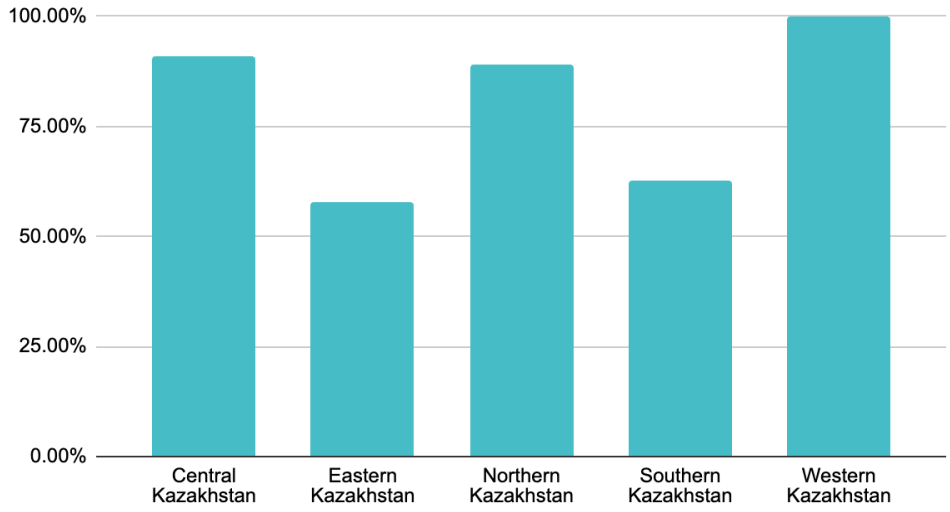


Figure 14 shows the results of the entire corpus's zh-dropping in macro regions. Based on the results, most participants from the western regions tended to drop the "zh" sound more frequently than did the particpants in the other regions.

Figure 14. Zh-dropping in macro regions



4.2 Survey Results

In this section, the results of the analysis of zh-dropping in the survey data are presented. The survey includes a total of 214 participants, including Kazakh speakers who immigrated from a number of countries: Turkmenistan, Mongolia, Uzbekistan (including the autonomous Karakalpakstan republic), Russia, and China. As demonstrated in the chart below (Figure 15), the majority of participants were born in Kazakhstan (179 participants), with a single participant each from Turkmenistan and Karakalpakstan. The second-largest group consists of Mongolian Kazakhs, totaling 26 participants. The remaining participants were born in Russia (2), Uzbekistan (2), and China (3), constituting a considerable portion of the data.

Figure 15. Survey participants across countries

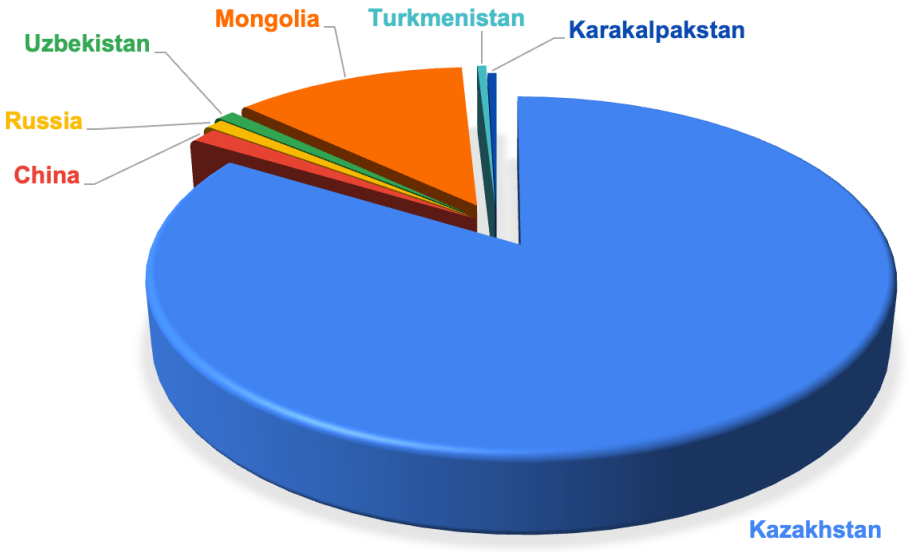


Figure 16 shows the proportion of participants across all regions of Kazakhstan. As illustrated below, the Abai (13%), Almaty (12%), Akmola (13%), and Karaganda (11%) regions account for the largest proportion of participants

compared to the other regions. The next largest group of participants is from the following regions: Atyrau (9%), Turkistan (8%), and North Kazakhstan (6%), with these next regions showing less than 5%: West Kazakhstan (4%), Kyzylorda (4%), East Kazakhstan (4%), Aktobe (3%), Ulytau (3%), Mangystau (2%), Kostanay (2%), Jambyl (2%), Jetisu (2%), and Pavlodar (1%).

Figure 16. Survey participants across all Kazakhstan regions

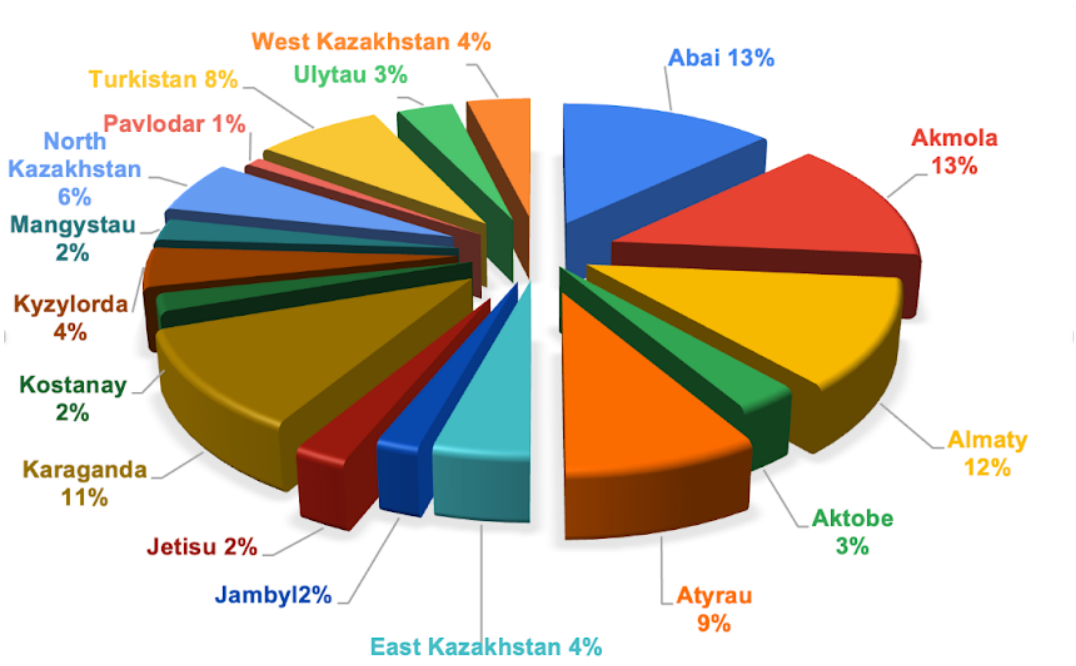


Table 5 presents the gender distribution of the participants, and Table 6 shows the number of participants across different countries. The majority are female (172), while male participants total 42.

Table 5. Gender distribution of the participants

Gender	Number of Participants
female	172
male	42

Table 6. Number of participants across different countries

Countries	Number of Participants
China	3
Karakalpakstan	1
Kazakhstan	179
Mongolia	26
Russia	2
Turkmenistan	1
Uzbekistan	2
Grand Total	214

Language Usage in Informal Settings

This section examines the self-reported usage of the Kazakh progressive aspect auxiliary verb "zhatyr" in informal settings. The analysis aims to explore the perception of phonetic reduction patterns across different regions.

Table 7 presents the self-reported use of zh-dropping across all relevant auxiliary verb constructions. It illustrates the standard form of all progressive aspect constructions and shows whether people perceive dropping of "zh" in them. The percentage of zh-dropping for each construction, along with the average across all constructions for each region, is shown in the table below. According to the survey results, the sole participants from Konyrat (average 83.3%) and Uygur (average 100%) reported that they do drop the "zh" sound in most constructions. In contrast, participants from Jetisu (average 0.0%) reported that they do not drop the "zh" sound in all constructions. The percentage for each construction varies in each

region; however, based on survey results, participants from West Kazakhstan—Aktobe (average 41.7%), Atyrau (average 44.4%), Mangystau (average 73.3%), and West Kazakhstan (average 77.1%)—tended to report that they drop "zh" frequently compared to other regions.

To zoom in on the constructions for a particular region, participants from Kostanay reported that they do not drop the "zh" sound in the constructions "bara zhatyrmyn" [bɑrɑ ʒɑtɯrmɯn] (I am going), "qoldanyp zhatqan" [qoldɑnɯp ʒɑtqɑn] (currently using), "zhep zhatyr" [ʒɛp ʒɑtɯr] (he is eating), and "aityp zhatqanym" [ɑitɯp ʒɑtqɑnɯm] (what I am saying) (0.0%); however, some participants reported that they do drop the "zh" sound in some constructions, for example in "ne istep zhatyr" [nɛ ɪstɛp ʒɑtɯr] (What are you doing?) (25.0%) and "ana zhaq" [ɑnɑ ʒɑq] (there) (25.0%). Participants from Jambyl reported that they drop the sound "zh" in all constructions less frequently (25.0%), except for the construction "bara zhatyrmyn" [bɑrɑ ʒɑtɯrmɯn] (I am going) (0.0%). Only a few participants from the Abai and Almaty regions reported dropping the "zh" sound in all constructions. The averages across all relevant constructions for these two regions were 8.1% for Abai and 16.0% for Almaty.

Table 7. Self-reported use of zh-dropping across all relevant auxiliary verb constructions

Participant Region of Origin	bara zhatyrmyn	qoldanyp zhatqan	aityp zhatqanym	ne istep zhatyr	zhep zhatyr	ana zhaq	Average across all relevant constructions	Sample Size
Abai	3.7%	3.7%	11.1%	11.1%	11.5%	7.7%	8.1%	27
Akmola	10.7%	14.3%	25.0%	28.6%	25.0%	10.7%	19.0%	28
Aktobe	50.0%	33.3%	50.0%	33.3%	33.3%	50.0%	41.7%	6
Almaty	12.0%	16.0%	12.0%	28.0%	12.0%	16.0%	16.0%	25
Atyrau	61.1%	38.9%	38.9%	38.9%	44.4%	44.4%	44.4%	18
East Kazakhstan	22.2%	22.2%	22.2%	22.2%	22.2%	0.0%	18.5%	9
Jambyl	0.0%	25.0%	25.0%	25.0%	25.0%	25.0%	20.8%	4
Jetisu	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5
Karaganda	13.0%	21.7%	26.1%	47.8%	30.4%	13.0%	25.4%	23
Konyrat	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%	83.3%	1
Kostanay	0.0%	0.0%	0.0%	25.0%	0.0%	25.0%	8.3%	4
Kyzylorda	33.3%	50.0%	25.0%	37.5%	37.5%	62.5%	41.0%	9
Mangystau	60.0%	80.0%	60.0%	80.0%	80.0%	80.0%	73.3%	5
North Kazakhstan	16.7%	16.7%	25.0%	25.0%	33.3%	8.3%	20.8%	12
Pavlodar	0.0%	0.0%	50.0%	50.0%	50.0%	0.0%	25.0%	2
Turkistan	18.8%	37.5%	31.3%	43.8%	37.5%	43.8%	35.4%	16
Ulytau	42.9%	28.6%	42.9%	57.1%	42.9%	57.1%	45.2%	7
West Kazakhstan	75.0%	62.5%	75.0%	87.5%	87.5%	75.0%	77.1%	8
Uygur	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	1
Mongolia	25.0%	25.0%	25.0%	25.0%	25.0%	25.0%	25.0%	4
Full Sample	22.9%	23.9%	26.8%	34.3%	29.7%	25.0%	27.1%	214

Amanzholov (1953) mentioned the use of “zhatyrgan” [ʒɑtɯrʁɑn] in western regions in his work. In standard grammar, when the auxiliary verb “zhatyr” takes the suffix -qan [qɑn], the suffix -yr [ɯp] is supposed to drop, as in one of the constructions above “aityp zhatqanym” [ɑitɯp ʒɑtqɑnɯm] (what I am saying). However, people from the west part of Kazakhstan keep the suffix -yr [ɯp] and make it “zhatyrgan” [ʒɑtɯrʁɑn].

Table 8 demonstrates the self-reported use of western Kazakh variants in auxiliary verb constructions. Based on the survey results, participants from West Kazakhstan—Aktobe (50.0%), Atyrau (38.9%), Mangystau (20.0%), and West Kazakhstan (37.5%)—appear to use the construction “qoldanyp zhatyrgan” [qoldɑnɯp ʒɑtɯrʁɑn] more frequently compared to the remaing regions. However, none of the participants from western regions except Atyrau (5.6%) and Kyzylorda (12.5%) reported that they employed the construction “aityp zhatyrganym” [ɑitɯp ʒɑtɯrʁɑnɯm]. According to the participants’ perceptions, the first construction, “qoldanyp zhatyrgan” [qoldɑnɯp ʒɑtɯrʁɑn], is also adopted by participants from the Abai (7.4%), Akmola (7.1%), Jambyl (25.0%), Karaganda (4.3%), Kyzylorda (12.5%), and Turkistan (12.5%) regions.

Table 8. Self-reported use of zh-dropping western Kazakh variants in auxiliary verb constructions

Participant Region of Origin	qoldanyp zhatyrgan	aityp zhatyrqanym	Average across all relevant constructions	Sample Size
Abai	7.4%	0.0%	3.7%	27
Akmola	7.1%	0.0%	3.6%	28
Aktobe	50.0%	0.0%	25.0%	6
Almaty	0.0%	0.0%	0.0%	25
Atyrau	38.9%	5.6%	22.2%	18
East Kazakhstan	0.0%	0.0%	0.0%	9
Jambyl	25.0%	0.0%	12.5%	4
Jetisu	0.0%	0.0%	0.0%	5
Karaganda	4.3%	0.0%	2.2%	23
Konyrat	100.0%	0.0%	50.0%	1
Kostanay	0.0%	0.0%	0.0%	4
Kyzylorda	12.5%	12.5%	12.5%	9
Mangystau	20.0%	0.0%	10.0%	5
North Kazakhstan	0.0%	0.0%	0.0%	12
Pavlodar	0.0%	0.0%	0.0%	2
Turkistan	12.5%	0.0%	6.3%	16
Ulytau	0.0%	0.0%	0.0%	7
West Kazakhstan	37.5%	0.0%	18.8%	8
Uygur	0.0%	0.0%	0.0%	1
Mongolia	0.0%	0.0%	0.0%	4
Full Sample	11.3%	0.9%	6.1%	214

Table 9 shows the self-reported lenition “zh” to “y” [ʒ > j] across all regions. Based on the results, participants from Ulytau (28.6%) and Mangystau (20.0%) seemed to shift the “zh” sound to “y” [j] more frequently than did those from other regions. The second region that shifts this sound to “y” [j] is Kyzylorda (12.5%), followed by Turkistan (6.3%), Atyrau (5.6%), and Almaty (4.0%). The remaining regions reported that they do not shift the “zh” sound into “y” [j] at all.

Table 9. Self-reported use of [ojerde] (zh > j) lenition

Participant Region of Origin	ojerde	Sample Size
Abai	0.0%	27
Akmola	0.0%	28
Aktobe	0.0%	6
Almaty	4.0%	25
Atyrau	5.6%	18
East Kazakhstan	0.0%	9
Jambyl	0.0%	4
Jetisu	0.0%	5
Karaganda	0.0%	23
Konyrat	0.0%	1
Kostanay	0.0%	4
Kyzylorda	12.5%	9
Mangystau	20.0%	5
North Kazakhstan	0.0%	12
Pavlodar	0.0%	2
Turkistan	6.3%	16
Ulytau	28.6%	7
West Kazakhstan	0.0%	8
Uygur	0.0%	1
Mongolia	0.0%	4
Full Sample	3.3%	214

Table 10 presents self-reported phonological reductions affecting sounds other than "zh." Two sample constructions are provided, (ol zherde > ozherde [ol ʒɛrdɛ > oʒɛrdɛ] and (ana zhaq > anyaq [ɑnɑ ʒɑq > ɑnjɑq]). According to the survey results, across all regions, participants from Mangystau (average 60.0%) appear to drop the last sounds of the first morpheme most often. In contrast, participants from Jetisu (average 0.0%), Kostanay (average 0.0%), and Pavlodar (average 0.0%) reported that this sound reduction does not occur. A few participants from remaining regions reported that this sound reduction occurs.

Table 10. Self-reported phonological reductions affecting sounds other than "zh"

Participant Region of Origin	ol zherde > ozherde	ana zhaq > anjaq	Average across relevant constructions	Sample Size
Abai	15.4%	3.8%	9.6%	27
Akmola	14.3%	3.6%	8.9%	28
Aktobe	16.7%	16.7%	16.7%	6
Almaty	24.0%	4.0%	14.0%	25
Atyrau	5.6%	0.0%	2.8%	18
East Kazakhstan	22.2%	0.0%	11.1%	9
Jambyl	50.0%	25.0%	37.5%	4
Jetisu	0.0%	0.0%	0.0%	5
Karaganda	30.4%	8.7%	19.6%	23
Konyrat	0.0%	0.0%	0.0%	1
Kostanay	0.0%	0.0%	0.0%	4
Kyzylorda	12.5%	25.0%	18.8%	9
Mangystau	60.0%	60.0%	60.0%	5
North Kazakhstan	8.3%	8.3%	8.3%	12
Pavlodar	0.0%	0.0%	0.0%	2
Turkistan	25.0%	25.0%	25.0%	16
Ulytau	14.3%	28.6%	21.4%	7
West Kazakhstan	25.0%	50.0%	37.5%	8
Uygur	100.0%	0.0%	50.0%	1
Mongolia	25.0%	0.0%	12.5%	4
Full Sample	18.9%	10.8%	14.9%	214

Figure 17 shows the self-reported zh-dropping in auxiliary verb constructions across different age groups. It appears that participants between the ages of 26 and 35 seem to drop the "zh" sounds more frequently than other age groups do. In comparison, participants between the ages of 51 and 70 drop the "zh" sound less frequently.

Figure 17. Self-reported zh-dropping in auxiliary verb constructions across different age groups

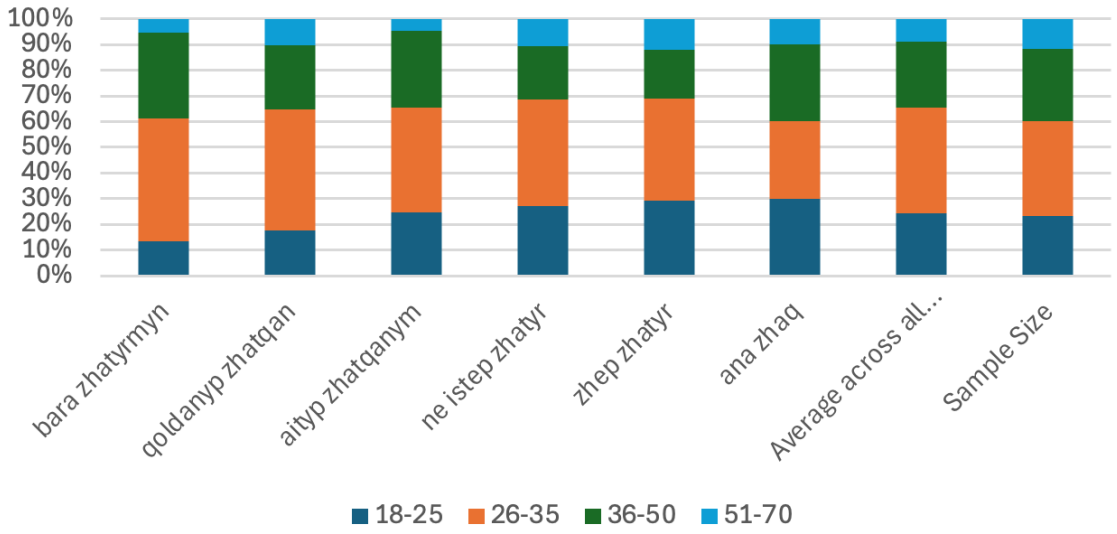
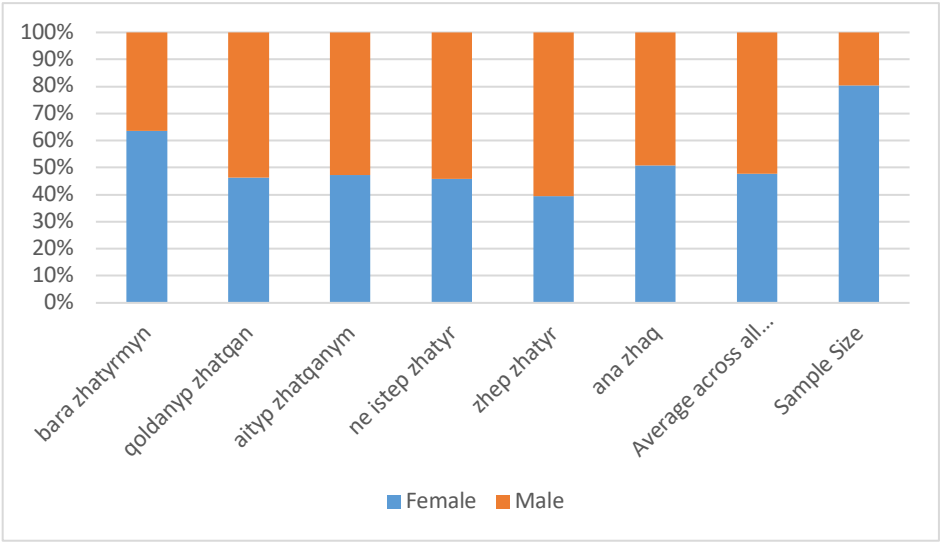


Figure 18 demonstrates the self-reported zh-dropping in auxiliary verb constructions across different genders. As mentioned above, the number of female participants exceeds the number of male participants. Based on the results, both groups show similar outcomes, except that the zh-dropping in some constructions is slightly more prevalent for male participants.

Figure 18. Self-reported zh-dropping in auxiliary verb constructions across different genders



Social and Regional Perceptions of Language Variation

This section describes the social and regional perceptions of language variation. To investigate these perceptions, several hypothetical stereotypical questions were asked in the survey.

Table 11 shows the perceived explanatory power of conditioning factors for regional variation, lack of education, and familiarity or formality of interlocutor (1 = certainly that factor is unimportant; 5 = certainly that factor is important). In total, 207 participants answered these questions.

According to the survey results, participants from East Kazakhstan (4.22), Kostanay (4.50), North Kazakhstan (4.08), and Mongolia (4.00) highly agree that zh-dropping is associated with regional variation; remaining participants somewhat agree or are not sure about this association.

Participants from the Jetisu (4.00) and Kostanay (4.00) regions most highly agree that zh-dropping is due to a lack of education. However, participants from Mangystau (1.80) and Ulytau (1.71) are less likely to agree with this statement.

Most participants seem to highly agree that being in a formal or informal social setting will affect their dropping of the “zh” sound (full sample average 3.61).

Table 11. Perceived explanatory power of conditioning factors (1 = certainly that factor is unimportant; 5 = certainly that factor is important)

Participants' Region of Origin	Regional variation	Lack of education	Familiarity or formality of interlocutor	Sample Size
Abai	3.76	3.13	3.79	25
Akmola	3.52	3.00	3.68	27
Aktobe	3.33	2.17	3.50	6
Almaty	3.17	2.92	3.75	24
Atyrau	3.12	2.35	3.59	17
East Kazakhstan	4.22	2.56	3.78	9
Jambyl	2.25	3.00	2.25	4
Jetisu	2.80	4.00	2.80	5
Karaganda	3.57	3.14	3.32	23
Konyrat	3.00	3.00	5.00	1
Kostanay	4.50	4.00	3.75	4
Kyzylorda	3.50	3.14	3.88	8
Mangystau	2.80	1.80	3.20	5
North Kazakhstan	4.08	3.50	3.92	12
Pavlodar	2.50	2.50	3.00	2
Turkistan	3.60	2.86	3.36	15
Ulytau	2.86	1.71	4.17	7
West Kazakhstan	3.63	2.63	3.75	8
Uygur	4.00	1.00	2.00	1
Mongolia	4.00	3.50	3.50	4
Full Sample	3.48	2.92	3.61	207

Table 12 presents the results of perceived association of zh-dropping with social stereotypes (1 = no perceived stereotyping; 5 = strong stereotyping). In total 200 participants answered this question. It appears that the number of respondents from the Karaganda (1) and Pavlodar (2) regions and Uygur (1) is very low compared with other regions. The average results for the sample show that most

participants somewhat agree that prevalence of zh-dropping is associated with social stereotypes.

Table 12. Perceived association of zh-dropping with social stereotypes (1 = no perceived stereotyping; 5 = strong stereotyping)

Participants' Region of Origin	How strongly do they associate social stereotypes with zh-dropping?	Sample Size
Abai	4.04	24
Akmola	3.60	25
Aktobe	3.50	6
Almaty	3.78	23
Atyrau	3.53	17
East Kazakhstan	3.56	9
Jambyl	3.75	4
Jetisu	4.40	5
Konyrat	3.55	22
Karaganda	4.00	1
Kostanay	4.00	4
Kyzylorda	3.25	8
Mangystau	2.80	5
North Kazakhstan	3.92	12
Pavlodar	4.50	2
Turkistan	3.57	14
Ulytau	2.67	6
West Kazakhstan	4.13	8
Uygur	1.00	1
Mongolia	3.50	4
Full Sample	3.68	200

Table 13 demonstrates the perceived impact of variation and language attitudes about one's own speech. The total number of participants for these questions is 194, as shown in the table below; however, the number of participants from Karaganda (1) and Pavlodar (2) is very low in comparison with the other

regions. The participants from the Kyzylorda (4.25), North Kazakhstan (4.08), Ulytau (3.67), and West Kazakhstan (3.63) regions seem to agree that regional stereotypes influence their own speech. Only a few people seem to have felt judged based on pronunciation.

Table 13. Perceived impact of variation and language attitudes on own speech

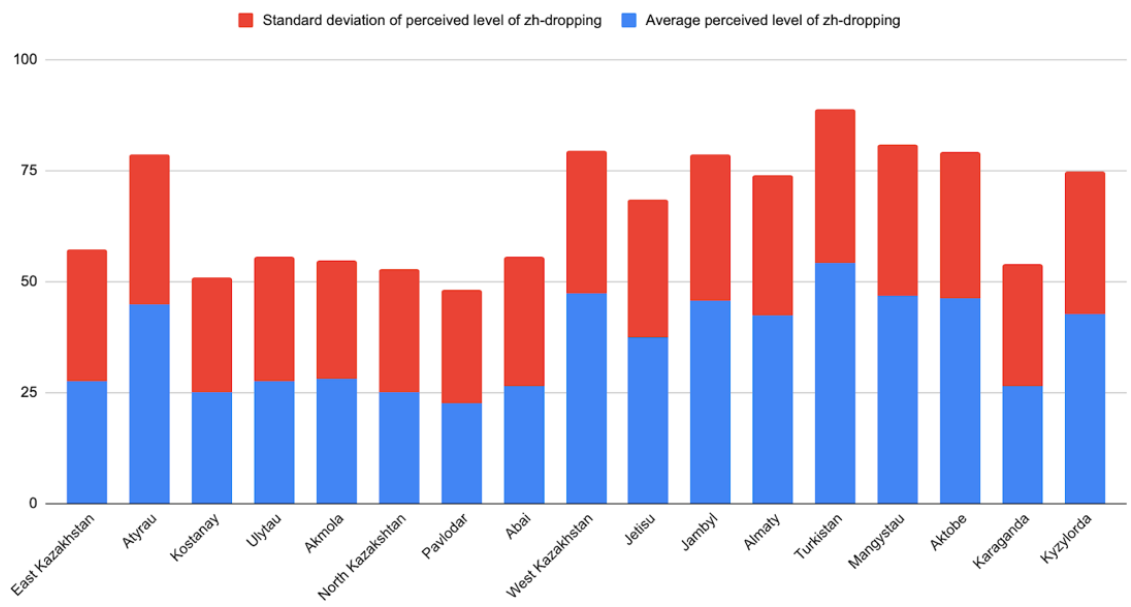
Participants' Region of Origin	How often do regional stereotypes influence own speech	Have felt judged based on pronunciation	Sample size
Abai	3.13	2.17	24
Akmola	3.36	2.18	22
Aktobe	2.67	2.00	6
Almaty	3.00	2.19	22
Atyrau	3.06	2.59	17
East Kazakhstan	2.78	3.11	9
Jambyl	2.75	1.50	4
Jetisu	2.80	1.20	5
Konyrat	3.64	2.23	22
Karaganda	4.00	1.00	1
Kostanay	2.75	1.75	4
Kyzylorda	4.25	3.13	8
Mangystau	3.75	2.00	4
North Kazakhstan	4.08	2.58	12
Pavlodar	3.50	2.50	2
Turkistan	3.15	3.00	13
Ulytau	3.67	1.67	6
West Kazakhstan	3.63	2.88	8
Uygur	5.00	1.00	1
Mongolia	4.50	2.00	4
Full Sample	3.32	2.33	194

Mapping Regional Speech Patterns

According to mapping speech patterns (Figure 19), more participants believe zh-dropping is frequent in the Turkistan region (34.6) than in other regions, such as

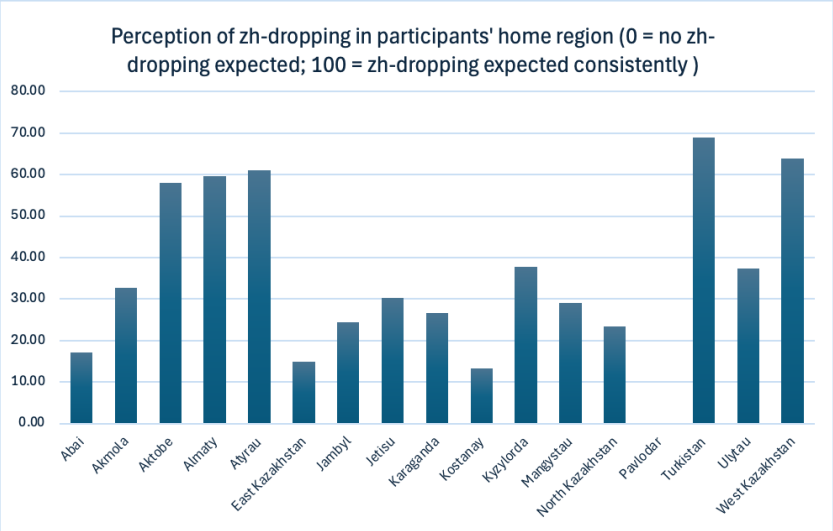
West Kazakhstan (32.3), Atyrau (33.8), Jetisu (30.8), Jambyl (32.8), Almaty (31.5), Mangystau (33.9), Aktobe (33.0), and Kyzylorda (32.1). The remaining regions seem not to drop the "zh" sounds, according to the participants' perceptions.

Figure 19. Average perceived level of zh-dropping by all regions



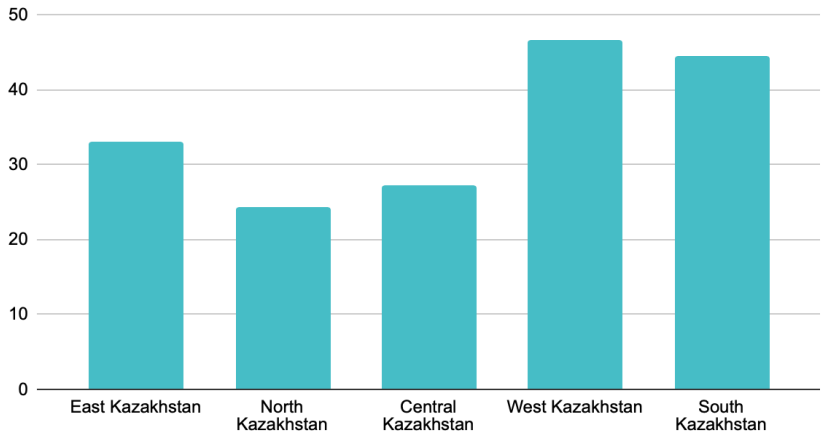
It was also interesting to see how participants from each region perceive zh-dropping in their home region. Figure 20 illustrates that participants from the Turkistan (68.9) and West Kazakhstan (63.83) regions perceive zh-dropping in their speech more than in any other regions. In addition, the Aktobe (58.00), Almaty (59.67), and Atyrau (61.07) regions' participants believe the zh-sound dropping occurs frequently in their speech as well. The participants from the remaing regions believe that they do not drop the "zh" sound very often. A single participant from the Pavlodar region believes they never drop the "zh" sound.

Figure 20. Average perceived level of zh-dropping by participants' home regions



According to Figure 21, the macro regions in which participants perceive that they drop the "zh" sound more frequently are West (46.7) and South (44.4) Kazakhstan. However, East (32.9), North (24.2), and Central (27.2) Kazakhstan participants believe that they drop the "zh" sound less frequently.

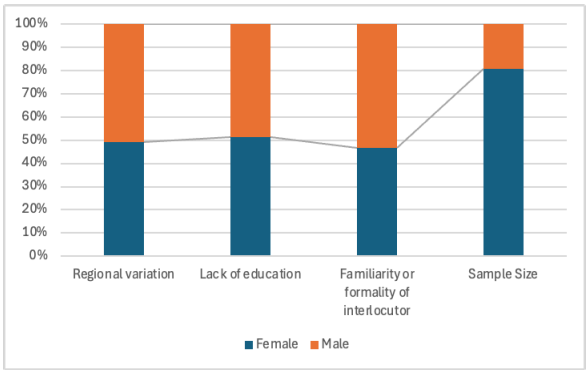
Figure 21. Average perceived level of zh-dropping by macro regions

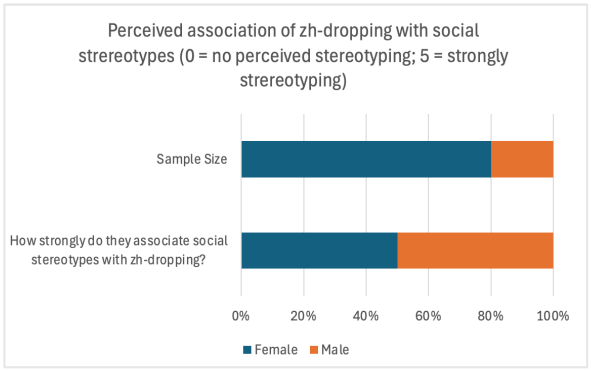


Based on the survey results in Figure 22, male and female participants' responses for social and regional variation of language perception show similar

results. There are no significant gender differences to specify. However, there is a slight difference among age groups. The number of participants between 51 and 70 is 24 in total, which is the lowest number in any age group. The majority of participants are between 26 and 35 (numbering 73). Figure 23 illustrates that almost all age groups show similar results, except that the numbers of participants are not equal.

Figure 22. Perceived explanations by gender





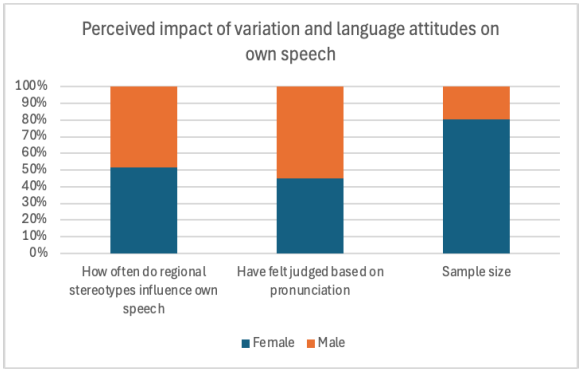


Figure 23. Perceived explanations by age

Chapter V Overall Conclusions

This study investigated the phenomenon of zh-dropping in Kazakh progressive aspect across all regions of Kazakhstan. The results show somewhat different regional variations. According to regional differences, the corpus data found that almost all regions except East Kazakhstan tend to drop the "zh" sound frequently. However, the survey data findings found that people perceived this phenomenon differently. Most participants perceived that people from Turkistan drop the "zh" sound consistently, but their perceptions do not match what is observed in the corpus data. Based on the results in the corpus data, 51.85% percent of the time "zh" was dropped by the participants. According to the macro regional differences, both data show that West Kazakhstan people drop the "zh" sound consistently.

The survey study found that there were several regional variations, such as the non-standard auxiliary verb "zhatyrgan," which is used by not only western Kazakhstan people but also in other regions, such as Kyzylorda, Abai, Akmola, Jambyl, Karaganda, and Turkistan. Most scholars mentioned in their studies that "zh" shifting to "y" sound is only used in the southern part of Kazakhstan; however, survey results show that people from Atyrau, Mangystau, and Ulytau appear to use it as well, and they are in the western and central parts of Kazakhstan.

Social and regional perceptions of language varied across different regions. In accordance with open-ended questions, most participants believe that zh-dropping is associated with lack of education, fast speech, environmental influences, different

dialects, informal conversation, and connection with people from neighboring countries. Since most participants linked this phenomenon to lack of education, this opens up more questions for further research. In the future, it would be interesting to look into the background of the participants to see if level of formal education plays a role in the behaviors or attitudes that were observed.

As mentioned in the introduction, several factors have influenced the development of the Kazakh language. One significant factor is Kazakhstan's history as part of the U.S.S.R., which contributed to the ruralization of Kazakh. This historical context has led to the emergence of a widespread stereotype: individuals who speak only Kazakh are often perceived as less educated and from rural areas.

Another major factor is the migration of ethnic Kazakhs from other countries, which has introduced variation and shaped language development. Additionally, neighboring languages, such as Uzbek, have influenced language change, particularly through language contact. Some participants even recognized that specific sound changes, such as lenition, are occurring due to these contact situations.

Social influences have also contributed to the formation of such stereotypes. Whether a speaker is from a rural or urban background, and whether or not they are fluent in Russian, appears to affect how sound changes are perceived and interpreted. These factors seem to reinforce the association between certain phonetic shifts and lack of education. Therefore, further research should explore

participants' educational backgrounds to better understand how these perceptions are formed.

This study shows that the regional variation in the "zh" sound is not a fixed feature of certain dialects but an ongoing process of change. The data show it is spreading but more so with certain social and geographic factors than others. Speakers perceive the "zh" dropping as being highly restricted (geographically and socially), but the actual production shows a more complex reality. This study found that reduction of the "zh" sound in the progressive aspect of a postural verb is a regular part of informal speech, even by speakers who do not consciously realize it is in their language pattern.

These findings contribute to broader discussions on phonological change by showing how perception and production can diverge in meaningful ways. The perceived reduction of "zh" shows that social and regional factors shape not just how sounds spread but how speakers think about linguistic variation. This has implications for how we understand sound change, particularly how features can spread despite being perceived as limited. Future research could investigate if similar mismatches between perception and production exist in other phonetic changes and what social forces drive linguistic evolution.

While this study provides valuable insights into zh-dropping in progressive aspect in auxiliary verb "zhatyr" construction, it is important to acknowledge certain constraints. One is the sample size of the corpus data, and another might be participant demographics, which may affect the generalizability of the findings.

Future research could address this by finding sufficient data from each region, also adding more information about participants' locations, such as where the participants spend most of their lives. Additionally, for survey data, it would be beneficial to have another small survey to investigate macro regions according to participants' views; it may help to group all 17 regions into fewer, more manageable regions. Finally, focusing on progressive aspect auxiliary verb "zh" omission may not fully capture the broader patterns of phonetic reduction in Kazakh. Expanding the study to different linguistic contexts would help to add further detail to these findings. Despite the parameters of this research, this study contributes to the understanding of different dialectal variations in Kazakhstan, and perception of these variations by different age, gender, and regional groups and lays the groundwork for further sociophonetic research.

References

Aaron Smith, K. (2011). Grammaticalization. Language and Linguistics Compass, 5(6), 367-380.

Abish, A. (2021). Kazakh and Karakalpak. In The Turkic Languages (pp. 337-353). Routledge. https://vividmaps.com/turkic-languages/

Alex. (2022, February 12). Where are Turkic languages used? - Vivid Maps. Vivid Maps.

Amanzholov, S. A. (1953). O dialect kazahskogo yazyka. Voprosy iyazykozna.

Barnes, J. A. (1954). Class and committees in a Norwegian island parish. Human relations, 7(1), 39-58.

Batayeva, Z. (2015). Colloquial Kazakh (eBook And MP3 Pack). Routledge.

Campbell-Kibler, K. (2010). Sociolinguistics and perception. Language and linguistics compass, 4(6), 377-389.

Campbell, L. (2013). Historical linguistics. Edinburgh University Press.

Csató, É. Į., & Johanson, L. (2021). Turkish. In The Turkic languages (pp. 195-223). Routledge.

Early, R. (2000). Sit, stand, lie: posture verbs and imperfectives. In SICOL Proceedings of the Second International conference on Oceanic Linguistics: Vol. 2, Historical and descriptive studies. Pacific Linguistics.

Fierman, W. (2006). Language and education in post-Soviet Kazakhstan: Kazakh-medium instruction in urban schools. The Russian Review, 65(1), 98-116.

Filchenko A., Troiani G., Du Bois J.W., Sarseke G., Akanov A., Bizhanova M., Mikhailov N., Temirbekova T., Seitak B., Turaliyeva Zh.,Dinmukhamed N. 2023/2024. Multimedia Corpus of Spoken Kazakh Language Nazarbayev University

Givon, T. (1976). Topic, pronoun and grammatical agreement. Subject and Topic/Academic Press Inc

Grenoble, L. A. (2003). Language Policy in the Soviet Union. Kluwer Academic Publishers.

Györfi, D. (2022). Auxiliary verb constructions in Modern Spoken Kazakh (Doctoral dissertation, University of Surrey).

Heine, B., & Kuteva, T. (2007). The genesis of grammar: A reconstruction (Vol. 9). Oxford University Press.

Heine, B., & Reh, M. (1984). Grammaticalization and reanalysis in African languages. (No Title).

Hopper, P. J., & Traugott, E. C. (2003). Grammaticalization. Cambridge University Press.

Jarlhøj, M. E., & Valijärvi, R. L. (2024). Changing language policies and ideologies in Kazakhstan: an interview-based study. SN Social Sciences, 4(2), 30.

Kaftan, L. (2004). Kak euvstvuet sebja "velikij i moguzij"russkij jazyk na prostorax byvsego SSSR? Fergana.ru. http://www.ferghana.ru (3-16-2007). June 17, 2004.

Katarinskii, V.V.C (2013). Grammatika kirgizskogo iyazyka. Fonetika, etimologia i sintaksis. Directmedia. *(Original work published 1906).*

Kenesbaev, I., & Musabaev, G. (1975). Sovremennyi kazakski yazyk. Almaty. (na kaz.iyz.).

Khasanov, B.H. (1987). Kazahsko-russkoe dvuyzychie: Socialno-lingvisticheskii aspect. Izd-vo “Nauka” Kazakhskoi SSR.

Kuryłowicz, J. (1965). The evolution of grammatical categories. Diogenes, 13(51), 55-71.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. The journal of abnormal and social psychology, 60(1), 44.

Languages of Kazakhstan. (2025, February 3). In Wikipedia. https://en.m.wikipedia.org/wiki/Languages_of_Kazakhstan

Lehmann, C. (2015). Thoughts on grammaticalization (p. 214). Language Science Press.

Medetova, N.K. (2016). Batys Kazakhstan oblysy Oral qalasy qazaqtaryynyn soileu tilindegi dialektilik erekshelikter. Redakcia alqasy, 97.

Meillet, Antoine. 1912. L’e´volution des formes grammaticales. Scientia (Rivista di Scienza) 12(26). 6.

Milroy, L. 1980. Language and Social Networks, in Language in Society 2. Oxford: Basil Blackwell.

Milroy, L., & Gordon, M. (2008). Sociolinguistics: Method and interpretation. John Wiley & Sons.

Mukhamedova, R. (2015). Kazakh: A comprehensive grammar. Routledge.

Nakysbekov, O. (1982). Qazaq tilining ontustik govorlar toby. Uchebnik/O. Nakysbekov-Almaty: Gylym, 167.

Pavlenko, A. (2009). Language conflict in post-Soviet linguistic landscapes. Journal of Slavic Linguistics, 247-274.

Qaliev, G., & Sarybaev, Sh. (1991). Qazaq dialectologiasy. Almaty: Ana tili

Sarybaev, Sh (2000). Qazaq til bilimi maseleleri. Voprosy kazakhskogo iyazykoznaniya. Almaty:«Arys» baspasy, 624.

Sauranbaev, N.T. (1955). Dialekty v sovremennom kazahskom yazyke. Voprosy iyazykoznaniya, 5, 43-51.

Schiering, R. (2006, August). Morphologization in turkish: Implications for phonology in grammaticalization. In Proceedings of the 13th international conference on Turkish linguistics.

Schwarz, H. G. (1984). The minorities of northern China: A survey. Center for East Asian Studies, Western Washington University.

Smailov, A. A. (2011). Results of the 2009 national population census of The Republic of Kazakhstan: Analytical report.

Traugott, E. C., & Heine, B. (Eds.). (1991). Approaches to Grammaticalization: Volume II. Types of grammatical markers(Vol. 19, No. 2). John Benjamins Publishing.

Turkic Languages. (2025, March 29). In Wikipedia. https://en.wikipedia.org/wiki/Turkic_languages#

Türker, E. (1995). Milroy's Social Network Theory-A Critical Account. Dilbilim Araştırmaları Dergisi, 6, 56-66.

Vajda, E. (1994). Kazakh phonology. Opuscula altaica, 603-650.

Appendix A

Survey Questions:

1. **Қайды тудыңыз?** (қала, мемлекет)

   Where were you born? (city, country)

2. **Қай өңірденсіз?**

   Which region are you from?

3. **Жасыңыз қаншада?**

   How old are you?

4. **Жынысыңыз? (ер/әйел)**

   Gender (male/female)

5. **Отбасыңызбен және/немесе достарыңызбен әңгімелескен кезде төмендегі тіркестердің қайсысын жиі қолданасыз?**

   Which of the following phrases do you find yourself using in casual conversations with family and/or friends?

   бара жатырмын (bara zhatyrmyn), баратырмын (baratyrmyn)

6. **Отбасыңызбен және/немесе достарыңызбен әңгімелескен кезде төмендегі тіркестердің қайсысын жиі қолданасыз?**

   Which of the following phrases do you find yourself using in casual conversations with family and/or friends?

   қолданып жатқан (qoldanyp zhatqan), қолданыватқан (qoldanyvatqan), қолданып атырған (qoldanyp atyrgan), қолданып жатырған (qoldanyp zhatyrgan)

7. **Отбасыңызбен және/немесе достарыңызбен әңгімелескен кезде төмендегі тіркестердің қайсысын жиі қолданасыз?**

   Which of the following phrases do you find yourself using in casual conversations with family and/or friends?

   (айтып жатқаным (aityp zhatqanym), айтып жатырғаным (aitypzhatyrganym), айтыватқаным (aityvatqanym)

8. **Отбасыңызбен және/немесе достарыңызбен әңгімелескен кезде төмендегі тіркестердің қайсысын жиі қолданасыз?**

   Which of the following phrases do you find yourself using in casual conversations with family and/or friends?

   Не істеп жатыр (ne istep zhatyr), не істеватыр (ne istevatyr)

9. **Отбасыңызбен және/немесе достарыңызбен әңгімелескен кезде төмендегі тіркестердің қайсысын жиі қолданасыз?**

   Which of the following phrases do you find yourself using in casual conversations with family and/or friends?

   жеп жатыр (zhep zhatyr), жеватыр (zhevatr)

10. **Отбасыңызбен және/немесе достарыңызбен әңгімелескен кезде төмендегі тіркестердің қайсысын жиі қолданасыз?**

    Which of the following phrases do you find yourself using in casual conversations with family and/or friends?

    ойерде (oyerde), ожерде (ozherde), ол жерде (ol zherde)

11. **Отбасыңызбен және/немесе достарыңызбен әңгімелескен кезде төмендегі тіркестердің қайсысын жиі қолданасыз?**

    Which of the following phrases do you find yourself using in casual conversations with family and/or friends?

    ана жақ (ana zhaq), аняқ (anyaq), анаяқ (anajaq)

12. **Сіз қандай сөйлеу ерекшеліктерін өңірлік деп ойлайсыз және оларды қай өңірлермен байланыстырасыз? (жауап беру міндетті емес)**

    What features of speech do you think are regional, which regions of you associate with? (optional)

13. **Қазақ тілінде кейбір грамматикалық формаларда, мысалы, созылыңқы «ж» дыбысының түсіп қалуын (мысалы, «Мен жұмыс істеп жатырмын» орнына «Мен жұмыс істеватырмын») Қазақстанның белгілі бір өңірлерімен байланысты деп ойлайсыз ба?**

    Do you think the omission of the 'zh' sound in certain grammatical forms, such as the progressive form using the auxiliary verb 'zhatyr' in Kazakh (e.g., 'Men zhumys istep zhatyrmyn' for 'I am working'), is associated with specific regions in Kazakhstan?

    - Иә, қатты байланысты (Yes, strongly associated)
    - Иә, аздап байланысты Yes, somewhat associated
    - Жоқ, онша байланысты емес (No, not really associated)
    - Жоқ, мүлдем байланысты емес (No, not at all associated)

- Сенімді емеспін (Unsure)

14. «Ж» дыбысының кейбір грамматикалық формаларда, әсіресе «жатыр» сөзінде түсіп қалуы білім деңгейінің жетіспеушілігімен байланысты деп ойлайсыз ба?

Do you think omitting 'zh' sound in certain grammatical forms, especially in 'zhatyr', is associated with a lack of education?

- Иә, толық келісемін (Yes, strongly agree)
- Иә, аздап келісемін (Yes, somewhat agree)
- Жоқ, аздап келіспеймін (No, somewhat disagree)
- Жоқ, мүлдем келіспеймін (No, strongly disagree)
- Сенімді емеспін (Unsure)

15. Адамдардың «ж» дыбысын түсіруі сөйлесіп отырған адамдарына (мысалы, достар, отбасы немесе ресми жағдайда) байланысты ма деп ойлайсыз?

Do you think the way people drop 'zh' sound is based on who they are speaking to (e.g., friends, family, or in formal settings)?

- Иә, өте көп (Yes, very much so)
- Иә, аздап (Yes, somewhat)
- Жоқ, байланысты емес (No, not really)
- Жоқ, мүлдем байланысты емес (No, not at all)
- Сенімді емеспін (Unsure)

16. **Қазақ тіліндегі «ж» дыбысының түсіп қалуымен байланысты әлеуметтік стереотиптер бар деп ойлайсыз ба?**

    Do you think there are any social stereotypes related to dropping the ‘zh’ sound in Kazakh?

    - Иә, толық келісемін (Yes, strongly agree)
    - Иә, аздап келісемін (Yes, somewhat agree)
    - Жоқ, аздап келіспеймін (No, somewhat disagree)
    - Жоқ, мүлдем келіспеймін (No, strongly disagree)
    - Сенімді емеспін (Unsure)

17. **Қазақ тіліндегі «ж» дыбысының түсіп қалуымен нақты қандай стереотиптер байланысты деп ойлайсыз? (жауап беру міндетті емес)**

    What specific stereotypes do you think are associated with dropping the 'zh' sound in Kazakh? (optional)

18. **Өңірлік диалектілер күнделікті әңгімелесуіңізге қаншалықты жиі әсер етеді деп ойлайсыз?**

    How often do you think regional dialects influence your daily conversations?

    - Өте жиі (Very often)
    - Біршама жиі (Somewhat often)
    - Сирек (Rarely)
    - Ешқашан (Never)
    - Сенімді емеспін (Unsure)

19. **Сіз тіл қолдануыңыз иә болмаса айтуыңыз бойынша сынауларға ұшыралыңыз ба?**

    Have you ever felt judged based on your language use or pronunciation?

    - Өте жиі (Yes, often)
    - Иә, кейде (Yes, sometimes)
    - Жоқ, өте сирек (No, rarely)
    - Жоқ, мүлдем (No, never)
    - Сенімді емеспін (Unsure)

20. **Сізге қандай сынаулар берілгенін сипаттай аласыз ба? (жауап беру міндетті емес)**

    Can you please describe how you were judged? (optional)

21. **Қазақ тіліндегі сөйлеу ерекшеліктеріне қатысты біз ескеруіміз керек басқа пікірлер бар ма? (жауап беру міндетті емес)**

    Are there any other comments regarding speech variation in Kazakh that we should take into consideration? (optional)

22. **Адамдардың ауызекі тілде "ж" әріпін түсіруі қай өңірлерге тән? Төмендегі шкала арқылы көрсеңіз. 0 - "ж" әріпін мүлдем түсірмейді 100 - үнемі "ж" әріпін түсіреді.**

    How likely are people to drop the "zh" sound in different regions? Please indicate in scales 0 - always keep the "zh," 100 - always drop the "zh".

23. Сізге ұқсас сөйлейтін адамдар қай өңірде тұрады? Картаға қарап сол аймақтарды белгілеңіз.

In which region do people speak in a way that is similar to you? Please circle the area on the map.