



NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Sanzhar Aitzhanov

**LANGUAGE IN THE CONSTRUCTION
OF ETHNICITY AMONG KOREANS IN
KAZAKHSTAN: A CASE STUDY OF
THE KOREAN COMMUNITY IN
KARAGANDA**

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: LINGUISTICS

WP BRP 116/LNG/2024

This Working Paper is an output of a research project implemented) within NRU HSE's Annual Thematic Plan for Basic and Applied Research.. Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

**LANGUAGE IN THE CONSTRUCTION OF ETHNICITY
AMONG KOREANS IN KAZAKHSTAN: A CASE STUDY OF
THE KOREAN COMMUNITY IN KARAGANDA²**

This study focuses on examining the role of language as an attribute in the construction of ethnicity within the Korean community in Kazakhstan. The research examines how language functions as an attribute in the categorization and identification processes, and how it interacts with other ethnic attributes such as descent and appearance.

Drawing on qualitative methods, including interviews and focus groups, the study reveals that language plays a central role in the identification of individuals as “Russian” or “Hanguk” (Korean nationals from South Korea), while local Koreans don't use it to differentiate themselves from other ethnic groups within Kazakhstan.

The findings suggest that language, particularly when other visible attributes are absent, can serve as an independent attribute of ethnic belonging. The study contributes to a broader understanding of ethnicity and identity construction.

JEL Classification: Z.

Keywords: ethnicity, constructivism, Koreans, Kazakhstan, South Korea, identification.

¹Laboratory for Comprehensive Interdisciplinary Projects, HSE University, Research Assistant.
E-mail: saitzhanov@hse.ru

² This work/article is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University).

Introduction

Ethnicity and nationalism play a crucial role in shaping modern societies. As Eriksen (2002: 2) highlights, these phenomena are impossible to ignore due to their visibility and impact.

The position of language as a core attribute of ethnicity is closely tied to German philosophy of nationalism, which viewed it in a highly essentialist manner (Herder in Fishman 1989: 15).

The relevance of this study lies in its aim to fill existing gaps in knowledge within the general theory of ethnicity and in the study of the Korean community in Kazakhstan. From the perspective of general theory, it is important to clarify a fundamental question in the study of language, ethnicity, and identity: why is language a significant attribute of ethnicity in some communities but absent in others?

In studies of the ethnicity of the Korean community in Kazakhstan, both methodological and theoretical issues and gaps exist. Methodological problems are mainly related to the reliance on quantitative data collection methods, which fail to capture the connections and causes behind the observed phenomena (Efremov, 2022; 2024)). There are also studies that examine the formation of ethnicity from a 'top-down' approach, while the 'bottom-up' perspective on ethnicity remains underexplored (Lee, 2013; Markova, 2021).

Theoretical problems in research on the Koreans in Kazakhstan are primarily associated with primordialist concepts³ (Khan and Sim, 2014; Khan, 2016) and groupism⁴ (Efremov, 2022; 2024).

This study examines the role of language in the construction and formation of ethnicity. The primary objective is to determine the role of language in constructing ethnicity.

This study addresses three research questions:

1. What ethnic categories exist within the Korean community of Karaganda?
2. What is the role of language in the categorization and identification of various categories within the Korean community in Karaganda?
3. Why does language play a critical role in certain ethnic boundaries but not in others?

Theoretical background

The key concepts of this study are categorization, identification, categories, and attributes. The concept of ethnic boundary is not used in this work; instead, the definitions of categorization

³ Primordialism or essentialism is a theory that views ethnicity as an innate characteristic of individuals. In this perspective, ethnicity is perceived as a static and unchanging variable. In contemporary studies of ethnicity, this theoretical concept is considered outdated (Brubaker, 2006; Wimmer, 2012; Varshaver, 2022).

⁴ Groupism is a concept introduced and critiqued by Brubaker (2006). He identified its central issue as the tendency in ethnicity and sociology studies to treat groups as the primary units of analysis, often perceiving them as homogeneous and bounded entities. This approach, according to Brubaker, overlooks the fluidity and internal diversity of such groups.

and identification are employed. Barth's (1969) concept of ethnic boundary, a metaphor, has exhausted its usefulness in contexts limited to only two categories⁵ (Varshaver, 2024).

A shift in ethnic studies from boundaries to categorization occurred with Brubaker's (2006) work, which changed the focus of research. Instead of focusing on specific groups, which do not exist in reality⁶, the subject of study shifted to the realm of cognitive science. However, Brubaker did not develop a full concept; rather, he highlighted the issues and set the direction for research on ethnicity.

These concepts were proposed by two major theorists in contemporary ethnicity research, Wimmer and Chandra. Wimmer (2013) proposed studies on the making of boundaries. Important analytical terms in this study are categorization and identification. Wimmer (2013: 64) defines “categorization as practices that define relevant groups” and “identification as who belongs to which groups”. Chandra (2012) proposed an alternative approach in which the primary focus is on the relationships between categories and attributes, which allow ethnicity to be identified based on identities. It is also important to clarify another set of analytical terms – nominal and activated identities or categories. Nominal ethnic identities “are those ethnic identity categories in which an individual’s descent-based attributes make her eligible for membership. Activated ethnic identities are those ethnic categories in which she actually professes membership or to which she is assigned membership by others.” (Chandra, 2012, p. 9).

Regarding identification, Brubaker (2006) suggested distinguishing between self-understanding, self-identification, identification by others, and self-representation.

An important theoretical framework is provided by Varshaver (2024), whose work offers a new perspective on the study of ethnicity. Varshaver introduces an analytical language, defining a range of concepts, including categorization, classification, taxonomy, and vernacular categories. According to him, categorization represents a loosely structured set of categories, classification is a structured but non-hierarchical set, and taxonomy is a structured, hierarchical set of categories (Varshaver, 2024, p. 9). He also introduces the notion of first- and second-order attributes, which can be characteristics, indicators, or norms. Generalized representations are also a key analytical term, encompassing a broad range of information.

In this work, I attempt to combine the theoretical concepts of various researchers into my analytical framework. Ultimately, I understand ethnicity as a phenomenon expressed through categories and attributes. Categories are the terms expressed in words, while attributes are the characteristics that essentially provide the possibility of claiming membership within a category.

⁵An excellent example is the article devoted to the construction of ethnicity and the attribute of language in the ethnic boundary between Chinese and Australians (Yang, 2023). However, due to the limited semantics of the metaphor, it rests on two categories.

⁶It is also important to recall Anderson and his concept of imagined communities (Anderson, 2006).

Methods and Data

This study employs qualitative data collection methods, specifically semi-structured interviews. The interview consisted of three sections: basic demographic information, a sociolinguistic section, and a section on categorization and identification. The basic demographic characteristics include sex, age (date of birth), marital status, occupation, education, and parents' nationality. The sociolinguistic section contained questions about language proficiency and attitudes toward languages. The final section was the core component, aimed at obtaining categories, attributes, discourses, markers, and social practices.

Qualitative data collection methods were chosen to capture the everyday discourses of ordinary people, rather than those of ethnic entrepreneurs. Qualitative methods reflect important causal relationships and are better suited to answering the research questions. It was crucial for me to uncover a broad network of categorization and identification, for which qualitative methods are more appropriate.

The total number of informants was 12 individuals, including 8 individual interviews and 2 group interviews. The interviews were conducted during fieldwork in August 2024 in the city of Karaganda, Kazakhstan. The age structure is somewhat incomplete, with 7 out of 12 informants being young individuals under 25 years old, and the remaining 5 informants being between the ages of 40 and 65. The study essentially lacks data collection on categories and attributes of individuals aged 65 and above. There is also a gender imbalance, with 3 out of 12 informants (25%) being male.

The focus group consisted of individuals with an activated Korean category, as the selection of informants was conducted using a snowball sampling method.

The methods used for data analysis include narrative analysis, Membership Categorization Analysis (MCA), and axial coding.

Results

Categories

The analysis revealed that informants from the Korean community in Karaganda categorize into four main ethnic categories: “Koreans”, “Koreans from South Korea”, “Kazakhs”, and “Russians”.

These are associated with other vernacular macro-categories, such as “Asians” and “Slavs”. This phenomenon is thoroughly described by Varshaver (2024) in his fieldwork in Moscow. Although we were unable to fully identify the informants' vernacular macro-categories, they are frequently categorized and identified by others as “Asians”. This category takes on the label of

“Kazakhs” due to their status as the state-forming ethnic community in Kazakhstan and as the most common variant within the Asian macro-category. This category was identified among all 12 informants:

- (1) *rt5501: They all say, everyone thinks, that I'm Kazakh.*
- (2) *rt5504: I realized why everyone mistakes me for a Kazakh.*
- (3) *rt5508: Yes, I'm often mistaken for a Kazakh.*
- (4) *rt5510: Yes, for instance, here, I'm very often perceived as Kazakh.*

The primary indicator of the macro-category was the informants' appearance. Appearance serves as an umbrella indicator, encompassing specific indicators such as “narrow eyes”, “cheekbones”, and “hair”.

Language and categorization and identification

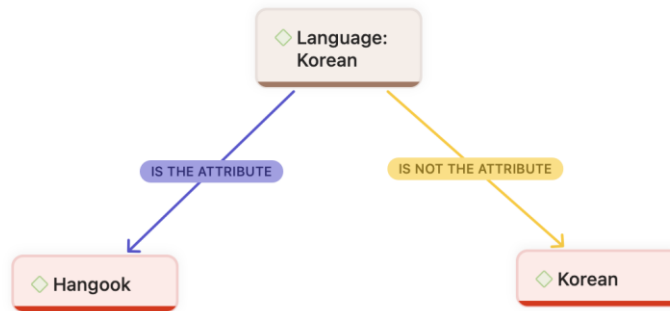
Membership in the aforementioned primary ethnic categories is based on the following most common attributes: descent, traditions, culture, mentality, and language. Language holds one of the most important and key attributes in the construction of ethnicity within the Korean community of Kazakhstan. Language possesses a unique characteristic as an attribute: it serves both as an indicator and as an attribute. Language is one of the primary elements highlighting differences in categorization between local Koreans and Koreans from South Korea.

As shown in Figure 1, the Korean language is both a crucial attribute and indicator for the categorization and identification of “Hangooks”, yet, on the other hand, it does not serve as an attribute for the categorization and identification of “local Koreans”. This appears to be related to the fact that categorization and identification of “local Koreans” are associated with other attributes, such as descent, appearance, and cultural factors.

- (1) *rt5501: Well, why? Haha. You can be Korean even without the language... For them (hangooks), first and foremost, it's a different language.*
int: But at the same time, to be a Hangook Korean in Korea, language is essential?
rt5501: Absolutely.
- (2) *rt5505: Well, considering that we live here and not in Korea—in Korea, of course, it's (Korean language) important... It's important for them to feel that you're one of their own, that you're Korean, that you can speak Korean... And here (in Kazakhstan), it's like, well, we don't know Korean, and that's okay.*
- (3) *rt5511: In Kazakhstan, it doesn't seem as important. I mean, there are many Koreans who don't know the Korean language—it really depends on the person, whether they want to learn it or not... It's really important to you to know Korean (in Korea). Important.*
int: So, if you don't know Korean, they might not consider you Korean, right?
rt5511: Well, basically, yes.

Figure 1

Attribute of Korean language in categorization and identification



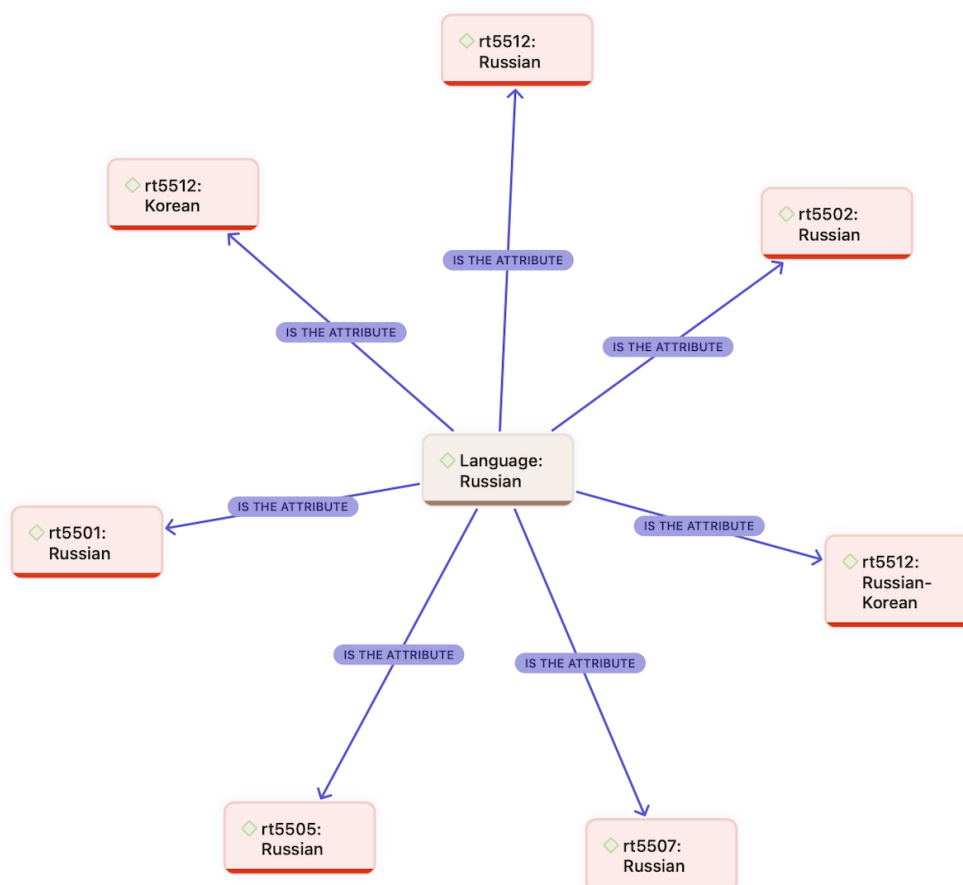
For all 12 informants, the Korean language plays a central role in categorization. However, beyond language, other significant components in categorizing Hangooks and local Koreans include attributes such as appearance, traditions, culture, citizenship, mentality, and mixed heritage—even among those Koreans who consider themselves "pure-blooded" Koreans. Among these attributes, only two are considered stickiness and visibility of attributes: appearance and language. Stickiness and visibility attributes are those that are difficult to conceal, and therefore, they play an active role in identification (Chandra, 2012). This generates interesting processes where language proficiency conceals identification attributes, leading the “hangook” to identify the person as one of their own.

- (1) *rt5512: My brother lives in South Korea; he has a South Korean passport and speaks Korean fluently, as well as English. They (hangooks) don't even suspect that he's not, how to say, a local—they consider him one of their own.*
- (2) *rt5506: Since there are many Koreans from the CIS working there, they can recognize us by our accent, and perhaps even by our appearance (laughs) and behavior. They identify us right away, so they don't really need to ask where we're from.*
- (3) *rt5506: Precisely because, for them in Korea, we're seen as migrant workers—not “real” Koreans, but outsiders who don't know our own culture, language, or anything else.*

Moreover, it is not only the Korean language that influences categorization and identification. As shown in Figure 2, the Russian language also serves as an important attribute that shapes various categories.

Figure 2

Attribute of Russian language in categorization and identification



Among the 12 informants, 5 revealed various connections between the attribute of the Russian language and the category of “Russians”. In addition to these 5, one more informant identified a connection between the attribute of the Russian language and the category of “Russian-Koreans”. Among the 6 informants, 3 have a descent-based attribute associated with the category of Russians. Thus, the attribute of the Russian language among Koreans can form identification with this category without the descent-based attribute, allowing it to function both as an independent attribute and in combination with the descent-based attribute. Even among informants with a descent-based attribute, the Russian language attribute is mentioned first as the primary identifying feature.

- (1) *rt5501: And undoubtedly, the fact that Russian... the Russian language, it's as if there's something Russian inside me.*
- (2) *rt5502: In terms of comparison, if someone asks me whether I'm more Korean or Russian, I'd probably say Korean. But the rest, the rest of me would be Russian.
int: And why is that? Why do you consider that part Russian?
rt5502: Well, again, probably because I speak Russian. I... (Pause for 5 seconds). Why? Well, somehow, plus still... yeah, and... Well, I think the main thing is that we speak Russian, and if you compare it with living in Kazakhstan, it's probably because we don't speak Kazakh and we're not Kazakhs, so we're more Russian than not.*

(3) *int: Can you explain why you consider yourself Russian?*
rt5505: Well, how? I don't know. My whole family speaks Russian.

(4) *int: Did you consider yourself Russian? Or do you consider yourself Russian?*
rt5507: Yes.

int: Why?

rt5507: Probably, I'm more Russian than Korean. Well, because, in general, first of all, I speak Russian.

(5) *rt5512: I don't know, it's just that with Koreans, it's already clear that if you speak the same language (Russian) as them, you're considered some kind of local Korean, definitely not a Han-guk (Korean from South Korea).*

The last quote is particularly noteworthy because it highlights the language attribute as one of the key elements in the categorization and identification of “local Koreans” as distinct from “Hangook” (South Koreans). This seems to be a fundamental attribute that differentiates “locals” from other Koreans around the world. However, during the study, none of the 12 informants interacted with categories of Koreans from large Korean communities outside of Korea (such as Koreans from the USA, Canada, New Zealand, and China). There is an assumption that language will play a primary role in the identification and categorization when local Koreans interact with these communities. However, this pattern of categorization and identification does not apply to different categories of Koreans from the CIS, as their language attribute is generally the same. Instead, attributes of mentality and culture come to the forefront, as only these can express the differences between the categories.

(1) *rt5512: For example, Central Asian Koreans, those from Uzbekistan and southern Kazakhstan, are very similar. However, those from the central regions are more like Russians, more calm, and among them, there are a lot of intellectuals, quite a few. ...And Sakhalin Koreans generally consider themselves a bit chosen as well, thinking they are closer to Korea. Often, their patronymics are Korean—like, for example, mine is Yuryevna, but theirs might be something like San Bonovna, more traditionally Korean. So, they see themselves as more purely Korean.*

(2) *rt5510: Yes, they are a little different... Our relatives used to live in Uzbekistan back in the Soviet times, and they are a bit different. They have more of the customs and habits, even the traditions, that they adopted from the Uzbeks, which we don't have in Kazakhstan. Also, in Russia, we have some relatives there as well, and they are a bit different too. There are two or three generations that stayed in Russia, and they are also a bit different.*

It is worth clarifying that categorization and identification takes place within the framework of the process of determining the relevant categories. For example, many informants do not have the category of Uzbek or other Koreans due to the lack of interaction with them.

Self-understanding, self-identification, identification by others and self-representation

The analysis also revealed the complex identification processes among the Korean community in Karaganda. I draw on Brubaker's (2006) theoretical interpretation of "identity," which he replaced with four analytical categories. Two of these present no issues: identification by others occurs explicitly and can be observed in informants' discourse. The most common way Koreans in Kazakhstan are identified by others is through the category "Kazakh," which is based on appearance. Self-representation is also clear, as it manifests through informants and reflects the way they present their ethnicity depending on the context. For example, the term "CIS Koreans" is not used in Korea, while "Koryo-in" is not used in Kazakhstan.

The main challenge lies in the remaining two analytical categories: self-understanding and self-identification, as the boundary between these categories is not clear. Brubaker described this term as "designates what might be called 'situated subjectivity': one's sense of who one is, of one's social location, and of how (given the first two) one is prepared to act." (Brubaker, 2006: 44). I interpret self-understanding as possessing an attribute for a category but rejecting self-identification with that category, whereas self-identification involves acceptance of identification with the category. A good example of self-understanding and self-identification is this fragment:

- (1) *rt5512: First, my appearance. Second, society—especially in Russia. That's one reason we left our daughter in Kazakhstan. Society there doesn't really see me as Russian, unless I start talking, for example, and begin interacting. I can't pound my chest and say, "I'm Russian too, here, I'll show you my passport"—no. I can't say I'm 100% Russian. It's mostly because of my appearance, I think. And maybe a bit of my own self-identification too. But there's a sense of belonging, like the concept of being a "Rossiyanin" (a citizen of Russia). To me, Rossiyanin isn't a nationality; it's a mix of many ethnicities. But to say I'm truly Russian—no, of course not.*

In this regard, the Koreans of Karaganda face different levels of identification. The language in these levels of identification is associated with various categories, however, consider the three main "hangoon", "local Koreans" and "Russians". The language attribute is one of the key factors in the self-understanding process. The ideology of "one language – one ethnicity" has an impact, leading informants to often perceive the Russian language as a crucial attribute in their self-understanding as Russians. However, self-understanding interacts with the identification and categorization by others, in which informants do not fall into Russian categories due to the lack of

appearance attributes. Even individuals with descent-based attributes do not identify as Russian. This creates a situation of dual identity, which has been well-documented in studies of Korean communities in Canada (Kim and Duff, 2012), the United States (Kang, 2013; Choi, 2015), China (Kang, 2008; Lim, 2009; Gao, 2012; Jiang, 2018), and New Zealand (Park, 2022). In the Korean community of Kazakhstan, this duality manifests in various hybrid categories such as "Russian Korean" or "Kazakhstani Korean." For 6 out of 12 informants, the dual category is largely explained by the language attribute, which effectively serves as a compromise in negotiating between self-identification and external identification.

Conclusion

The focus group analysis showed that there are stable categories formed in the discourses and narratives of all focus group members. These categories are formed through different configurations of attributes, and while different informants have varying representations of the attributes of each category, language plays one of the key roles in these representations. Language as an attribute for the Korean community in Karaganda grants access to certain categories, but membership in these categories is not solely based on the language element. Language does not play a role in the category of "local Koreans", but it serves as an indicator distinguishing "local Koreans" from "Hangooks." On the other hand, language is a primary attribute and indicator when categorizing and interacting with "Hangooks", suggesting that language can become an attribute and indicator when it reflects differences and when no other more convenient markers or attributes exist to distinguish between categories.

Identification is also influenced by language. Because the language attribute provides partial access to certain categories, people begin to self-identify with them. However, this does not align with how they are identified by others. This results in dual identities, which exist among half of the informants in this study.

Research question responses:

1. Within the community, there exists a vast array of different categories; however, these can essentially be generalized into four main categories: Kazakhs, Russians, local "Koreans", and "Hanguk". Additionally, there is one overarching macrocategory – "Kazakhs".
2. Language plays a central role in categorization related to access to membership within a category. Language is not an attribute of the category "local Koreans"; however, it becomes a central component in the categories "Russians" and "Hanguks". Language can function both as an independent attribute that shapes identification with a category, as well as an additional attribute complementing a descent-based one.
3. Language is one of the most crucial attributes of a category, particularly in cases where it socially organizes distinctions. Its significance is heightened when there are no more

visible or salient attributes. For instance, in interactions with “Hangooks”, who share the same descent-based attributes and physical appearance as “local Koreans”, language becomes the primary marker of difference.

References

1. Anderson, B. (2006). *Imagined Communities Reflections on the Origin and Spread of Nationalism*. Verso.
2. Barth, F. (1969). *Ethnic Groups and Boundaries: The Social Organization of Culture Difference. Series in Anthropology*. Boston: Little, Brown.
3. Brubaker, R. (2006). *Ethnicity without groups*. Harvard university press.
4. Chandra, K. (Ed.). (2012). *Constructivist theories of ethnic politics*. Oxford University Press.
5. Choi, J. K. (2015). Identity and language: Korean speaking Korean, Korean-American speaking Korean and English? *Language and Intercultural Communication*, 15(2), 240–266. <https://doi.org/10.1080/14708477.2014.993648>
6. Efremov, E. A. (2022). The ethnocultural identity of the Koreans of Kazakhstan. *Korean Studies in Russia: Direction and development*, 3(1), 32-38.
7. Efremov, E. A. (2024). *The diaspora identity of Koreans of Kazakhstan: conservation factors*. Modern problems of the Korean Peninsula.
8. Eriksen, T. H. (2002). *Ethnicity and nationalism: Anthropological perspectives*. Pluto press.
9. Fishman, J. A. (1989). *Language and ethnicity in minority sociolinguistic perspective*. Multilingual Matters.
10. Gao, F. (2012). Imagined identity of ethnic Koreans and its implication for bilingual education in China. *International Journal of Bilingual Education and Bilingualism*, 15(3), 343–353.
11. Han, W. S., & Sim, H. H. (2014). *Koreans of Central Asia: Past and present*.
12. Jiang, X. (2018). *Voices from In-Between: Korean Chinese Identity under the Gazes of China and South Korea*. [Master's thesis, University of Alberta]. https://era.library.ualberta.ca/items/b4e9f0d7-244a-4080-b376-451cb30b042a/view/b6b6d6ff-7637-40d3-a447-1453fa9c53be/Jiang_Xuxiang_201805_MA.pdf
13. Kang, H. S. (2013). Korean American college students' language practices and identity positioning: "Not Korean, but not American". *Journal of Language, Identity & Education*, 12(4), 248–261.
14. Kang, J. W. (2008). The dual national identity of the Korean minority in China: The politics of nation and race and the imagination of ethnicity. *Studies in Ethnicity and Nationalism*, 8(1), 101–119.

15. Kim, J., & Duff, P. A. (2012). The language socialization and identity negotiations of generation 1.5 Korean-Canadian university students. *TESL Canada Journal*, 81–81.
16. Lim, T. (2009). Who is Korean? Migration, immigration, and the challenge of multiculturalism in homogeneous societies. *Asia-Pacific Journal*, 30(1).
17. Park, M. Y. (2022). Language ideologies, heritage language use, and identity construction among 1.5-generation Korean immigrants in New Zealand. *International Journal of Bilingual Education and Bilingualism*, 25(7), 2469–2481.
18. Varshaver, E. (2022). «Perestat' pinat' mertvuyu loshad' primordializma»: aktual'nye povestki dnya v konstruktivistskikh issledovaniyakh etnichnosti [“Stop kick on the dead horse of Primordialism”]: on the current agendas in the Constructivist studies of Ethnicity]. *The Russian Sociological Review*, 21(3), 31-58.
19. Varshaver, E., Gupalova, J., Orlova, A. (2024). *Are there only two ethnic groups in Moscow: Slavs and Southerners? Research on vernacular categorization with elicitation methods used.*
20. Varshaver, E. (2024). What Exactly is Studied When Ethnicity is Researched? A Descriptive Model for Constructivist Studies of Ethnicity in the Context of the Cognitive Turn. *Preprint*
21. Wimmer, A. (2013). *Ethnic boundary making: Institutions, power, networks.* Oxford University Press.
22. Yang, Y. (2023). The role of heritage language in multiple dimensions of ethnic identity: a case study of Chinese-Australian adolescents. *International Journal of Multilingualism*, 21(4), 1775–1798. <https://doi.org/10.1080/14790718.2023.2210284>

Sanzhar Aitzhanov

National Research University Higher School of Economics (Moscow, Russia). Laboratory for Comprehensive Interdisciplinary Projects, Research Assistant.

E-mail: saitzhanov@hse.ru

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

© Aitzhanov, 2024