

УДК 811;51

DOI 10.25587/2310-5453-2025-39-1-6-18

*Original article***Identity-based language shift modeling****Dmitry A. Ammosov<sup>1</sup>✉, Yalchin Efendiev<sup>2</sup>, Lenore Grenoble<sup>3</sup>**<sup>1</sup>Khalifa University of Science and Technology, Abu Dhabi, UAE<sup>2</sup>Texas A&M University, Texas, USA<sup>3</sup>University of Chicago, Chicago, USA

✉ dmitrii.ammosov@ku.ac.ae

**Abstract**

The preservation of endangered languages is a widely discussed issue nowadays. Languages represent essential cultural heritage and can provide valuable botanical, biological, and geographical information. Therefore, it is necessary to develop efficient measures to preserve and revitalize endangered languages. However, the language shift process is complex and requires an interdisciplinary approach, including mathematical modeling techniques. This paper develops a new mathematical model that extends previous works on this topic. We introduce the factor of ethnic identity, which is a proxy for a more complex nexus of variables involved in an individual's self-identity and/or a group's identity. This proxy is socially constructed rather than solely inherited, shaped by community-determined factors, with language both indexing and creating the identity. In our model, we divide speakers into groups depending on with which language they identify themselves with. Moreover, every group includes monolinguals and bilinguals. The proposed model naturally allows us to consider cases of language coexistence and describe a broader class of linguistic situations. For example, the simulation results show that our model can result in cyclic language dynamics, drawing a parallel to cell population models. In this way, the proposed mathematical model can serve as a useful tool for developing efficient measures for language preservation and revitalization.

**Keywords:** mathematical modeling, language shift, language competition, language coexistence, endangered languages, language revitalization, ethnic identity, self-identity, group identity, cyclic dynamics

**For citation:** Ammosov D.A., Efendiev Ya., Grenoble L. Identity-Based Language Shift Modeling. *Arctic XXI Century*. 2025, Vol. 39, No 1. P. 6-18. DOI: 10.25587/2310-5453-2025-39-1-6-18

© Ammosov D. A., Efendiev Ya., Grenoble L., 2025

Оригинальная научная статья

## Моделирование языкового сдвига на основе идентичности

Д. А. Аммосов<sup>1</sup>✉, Я. Эфендиев<sup>2</sup>, Л. Гренобль<sup>3</sup>

<sup>1</sup> Халифский университет науки и технологий, Абу-Даби, ОАЭ

<sup>2</sup> Техасский университет A&M, Техас, США

<sup>3</sup> Чикагский университет, Чикаго, США

✉ dmitrii.ammosov@ku.ac.ae

### Аннотация

Сохранение исчезающих языков является широко обсуждаемой проблемой в настоящее время. Языки представляют собой важное культурное наследие и могут предоставить ценную ботаническую, биологическую и географическую информацию. Поэтому необходимо разработать эффективные меры по сохранению и возрождению исчезающих языков. Однако процесс смены языка сложен и требует междисциплинарного подхода, включая методы математического моделирования. В этой статье разрабатывается новая математическая модель, которая расширяет предыдущие работы по этой теме. Мы вводим фактор этнической идентичности, который является прокси для более сложной связи переменных, вовлеченных в самоидентификацию человека и/или идентичность группы. Этот прокси социально сконструирован, а не только унаследован, сформирован факторами, определяемыми сообществом, при этом язык как индексирует, так и создает идентичность. В нашей модели мы делим говорящих на группы в зависимости от того, с каким языком они себя идентифицируют. Более того, каждая группа включает монолингвов и билингвов. Предлагаемая модель естественным образом позволяет нам рассматривать случаи сосуществования языков и описывать более широкий класс языковых ситуаций. Например, результаты моделирования показывают, что наша модель может привести к циклической языковой динамике, проводя параллель с моделями клеточной популяции. Таким образом, предлагаемая математическая модель может служить полезным инструментом для разработки эффективных мер по сохранению и возрождению языка.

**Ключевые слова:** математическое моделирование, языковой сдвиг, языковая конкуренция, языковое сосуществование, исчезающие языки, возрождение языка, этническая идентичность, самоидентичность, групповая идентичность, циклическая динамика

**Для цитирования:** Аммосов Д.А., Эфендиев Я., Гренобль Л. Моделирование языкового сдвига на основе идентичности. *Арктика XXI век*. 2025, Т. 39, № 1. С. 6-18. DOI: 10.25587/2310-5453-2025-39-1-6-18

## Introduction

The preservation of endangered languages is a highly relevant topic in the modern world. Languages are not only an important cultural heritage [1] but also a source of botanical, biological, and geographical information [2, 3]. Therefore, the development of effective measures for their preservation and revitalization plays an essential role. However, language shift is a complex process that requires careful study with an interdisciplinary approach. Understanding the influence of various factors on the extinction of languages is the key to their preservation and revitalization [4, 5].

Mathematical modeling can be a helpful tool in studying the process of language shift. It allows us to identify the key factors influencing language extinction and to predict language dynamics. There are many works devoted to the development of mathematical models of language shift. However, the model proposed by Abrams and Strogatz is fundamental [6]. The main idea of this model is to determine the probability of transition from one language to another depending on the status of the language and the number of speakers. Subsequent works are mostly devoted to extending this mathematical model to account for various factors such as geography, population growth, and bilingualism [7, 8, 9, 10, 11, 12, 13].

However, the mathematical models noted above predict the disappearance of bilinguals when one of the monolingual populations disappears. Therefore, they predict that only one language survives over the course of a significant time period. At the same time, it is believed that coexistence is possible [14]. To account for such a case, Kandler et al. introduced the “Diglossia model” [15, 16]. In this model, there are additional linear decay terms for each language and linear growth terms for the bilingual population. Such a modification is motivated by social pressure.

In this paper, we propose a new mathematical model that extends previous works and allows us to consider a broader class of language scenarios, including language coexistence. We introduce the factor of ethnic identity, a proxy for more complex variables shaping an individual’s self-identity and/or a group’s identity [17]. Ethnic identity is socially constructed rather than solely inherited, shaped by community-determined factors, with language both indexing and creating the identity [18, 19]. In our model, we have two groups of speakers identifying themselves with one or another language. In each of these groups, there are both monolingual and bilingual populations. Transitions can occur from one language to another (through the bilingual stage) and from one identity to another. The obtained mathematical models can naturally consider language coexistence scenarios. Moreover, our model can describe cyclic language dynamics, which makes a connection with cell population models.

The paper has the following structure. In Section 2, we provide a brief review of previous mathematical models of language shift. Section 3 describes our proposed mathematical model. In Section 4, we present results and discussions. Finally, Section 5 presents conclusions.

### **Review of previous works**

Mathematical modeling of language shift has gained popularity with the seminal paper by Abrams and Strogatz [6]. In their work, the authors consider two languages (A and B) and denote the population percentages with and, where . The dynamics are governed by the equations

$$\begin{aligned}\frac{da_1}{dt} &= a_2 P_{12}(a_1, s) - a_1 P_{21}(a_2, 1 - s), \\ \frac{da_2}{dt} &= a_1 P_{21}(a_2, 1 - s) - a_2 P_{12}(a_1, s).\end{aligned}\tag{1}$$

Here, refers to the probability that an individual speaking language B converts to language A, and is the opposite. The parameter refers to the status of the language A. Typical expressions for and are and, .

The next generalization of this model includes adding various spatial dynamics, as proposed by Patriarca et al. in [7, 8]. Note Equation (1) includes the dynamics in time only. This includes adding various diffusion and convection terms that disperse the population over a region. The latter is important, but we will not dwell on this. For simplicity, we denote them by .

$$\begin{aligned}\frac{da_1}{dt} &= a_2 P_{12}(a_1, s) - a_1 P_{21}(a_2, 1 - s) + \mathcal{F}_1(a_1, a_2), \\ \frac{da_2}{dt} &= a_1 P_{21}(a_2, 1 - s) - a_2 P_{12}(a_1, s) + \mathcal{F}_2(a_1, a_2).\end{aligned}\tag{2}$$

The next generalization includes adding population growth, as proposed by Kandler and Steele in [9], modeled by logistic growth functions with certain carrying capabilities (maximum allowed growth). These results are further generalized by varying capacities and making probabilities to be nonlinear functions, as demonstrated by Isern and Fort in [10].

The next set of major generalizations came with the introduction of a model of bilingual population by Mira and Paredes in [11]. The bilingual population is denoted by, and the model problem is the following (we factor in spatial dynamics terms in these equations, which were contributed by other authors)

$$\begin{aligned}
 \frac{da_1}{dt} &= (a_2 + a_3)P_{12}(a_1, s) - a_1 \left( P_{21}(a_2, 1 - s) + P_{31}(a_2, 1 - s) \right) \\
 &\quad + \mathcal{F}_1(a_1, a_2, a_3), \\
 \frac{da_2}{dt} &= (a_1 + a_3)P_{21}(a_2, s) - a_2 \left( P_{12}(a_1, s) + P_{32}(a_1, s) \right) \\
 &\quad + \mathcal{F}_2(a_1, a_2, a_3), \\
 \frac{da_3}{dt} &= a_1P_{31}(a_2, 1 - s) + a_2P_{32}(a_1, s) - a_3P_{12}(a_1, s) \\
 &\quad - a_3P_{21}(a_2, s) + \mathcal{F}_3(a_1, a_2, a_3).
 \end{aligned} \tag{3}$$

Here, are probabilities going from to . Note that the authors assume that speakers can switch between languages A to B and also between monolingual to bilingual. Later, Minett and Wang assumed that one can switch between monolingual to bilingual only and modified this model [12].

Later generalizations involve various spatial dynamics terms for bilingual population, addition of logistic growth models, and introduction of spatially dependent coefficients that can change in some regions (see the work of Castelló et al. in [13]).

These models led to the fact that only one language survives over the course of a significant time period. It is believed that coexistence is possible [20]. Kandler et al. introduced the “Diglossia Model” by adding linear decay terms for each language and, as a compensation, linear growth for the bilingual population (the terms with coefficients) [15, 16]. The latter is argued using social pressure. The model equation has the following form. We repeat this model here, which assumes some simplifications for transition probabilities (we ignore logistic growth terms for simplicity).

$$\begin{aligned}
\frac{da_1}{dt} &= -\gamma_{21}a_2a_1 + \gamma_{13}a_3a_1 - w_1a_1 + \mathcal{F}_1(a_1, a_2, a_3), \\
\frac{da_2}{dt} &= -\gamma_{12}a_1a_2 + \gamma_{23}a_3a_2 - w_2a_2 + \mathcal{F}_2(a_1, a_2, a_3), \\
\frac{da_3}{dt} &= (\gamma_{21} + \gamma_{12})a_1a_2 - (\gamma_{13}a_1 + \gamma_{23}a_2)a_3 + w_1a_1 \\
&\quad + w_2a_2 + \mathcal{F}_3(a_1, a_2, a_3).
\end{aligned} \tag{4}$$

These models may lead to the coexistence of mono- and bilingual speakers via the terms in .

### Proposed mathematical model

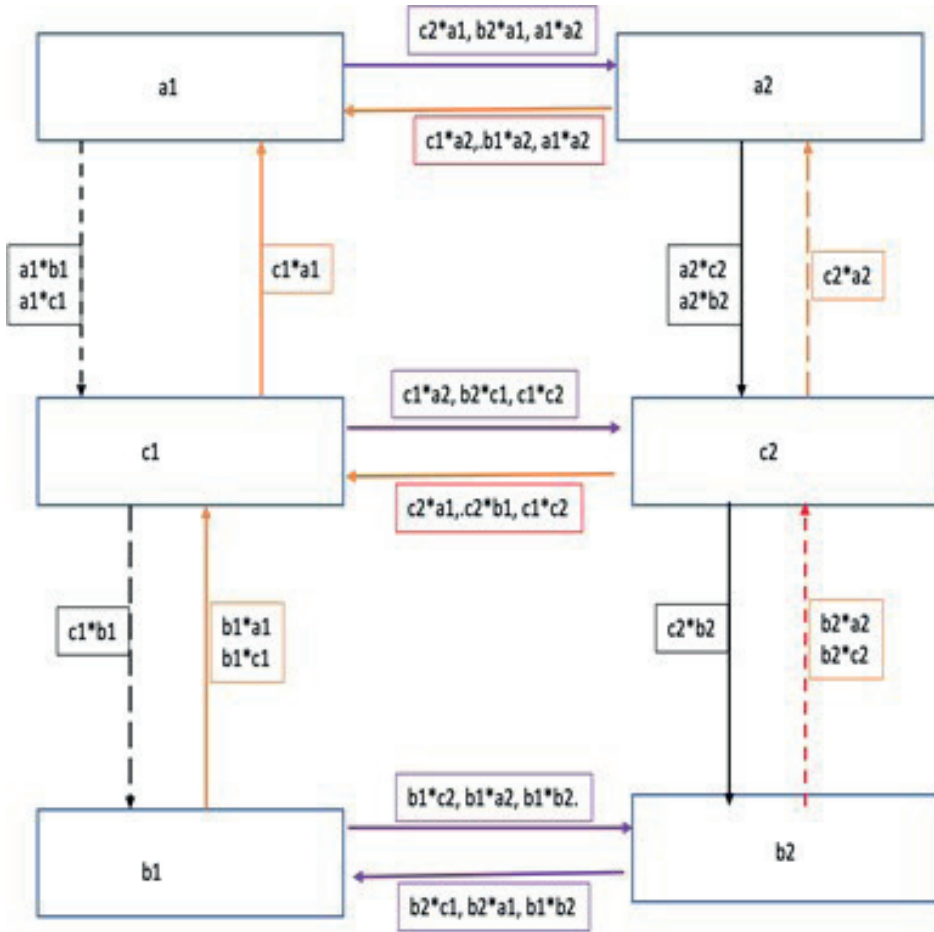
In our model, we divide the population into two groups. The members of Group 1 ethnically identify with language A, and the members of Group 2 ethnically identify with language B. We introduce the factor of ethnic identity, which serves as a proxy for a more complicated nexus of variables involved in an individual's self-identity and/or a group's identity. Ethnic identity is not simply determined by inheritance but is a socially constructed category involving other social variables including race, gender, social class, and status; [17] provides an overview. Which variables are relevant is largely determined by the community or communities involved in the identification. This proxy is used to account for not only inherited (genetic) identity but also ties that can be created by an individual who self identifies with a given group.

There are known ties between language and identity, with language at once used to index identity and also to be a marker of that identity. Language can index not only ethnic identity in the strict sense, but also social class or group membership [21, 22]. Moreover, identity can be created through language [23].

Note that members of each group may speak the other language or be bilingual. Thus, members of Group A may speak Language B and vice versa. We introduce the following notations:

- – the percentage of language A speakers identified as Group 1;
- – the percentage of language A speakers identified as Group 2;
- – the percentage of language B speakers identified as Group 1;
- – the percentage of language B speakers identified as Group 2;
- – the percentage of bilinguals identified as Group 1;
- – the percentage of bilinguals identified as Group 2.

Next, we illustrate the transition between different categories in Fig. 1. For simplicity, we ignore some terms and denote by dashed lines the weak transitions, which will also be ignored. The corresponding equations are (5). This model can easily be enhanced by adding spatial dispersion terms, logistic growth terms, and diglossia terms.



**Fig. 1.** Illustration of main transitions between different categories. The dashed line refers to a weak transition and will be ignored. Some connections depend on several factors that are separated by a comma. The model is simplified

**Рис. 1.** Иллюстрация основных переходов между различными категориями. Пунктирная линия относится к слабому переходу и будет игнорироваться. Некоторые связи зависят от нескольких факторов, которые разделены запятой. Модель упрощена

$$\begin{aligned}
 \frac{da_1}{dt} &= Z_{a_1c_1}a_1c_1 + Z_{a_2c_1}^a a_2c_1 + Z_{a_2b_1}^a a_2b_1 - Z_{a_1b_2}^a a_1b_2 \\
 &\quad - Z_{a_1c_2}^a a_1c_2 + X_{a_1a_2}a_1a_2, \\
 \frac{da_2}{dt} &= Z_{a_1c_2}^a a_1c_2 + Z_{a_1b_2}^a a_1b_2 - Z_{a_2c_1}^a a_2c_1 - Z_{a_2b_1}^a a_2b_1 \\
 &\quad - Z_{a_2b_2}a_2b_2 - Z_{a_2c_2}a_2c_2 - X_{a_1a_2}a_1a_2 \\
 \frac{db_1}{dt} &= Z_{b_2c_1}^b b_2c_1 + Z_{a_1b_2}^b a_1b_2 - Z_{a_1b_1}a_1b_1 - Z_{b_1c_1}b_1c_1 \\
 &\quad - Z_{b_1c_2}^b b_1c_2 - Z_{a_2b_1}^b a_2b_1 + X_{b_1b_2}b_1b_2, \\
 \frac{db_2}{dt} &= Z_{b_1c_2}^b b_1c_2 + Z_{a_2b_1}^b a_2b_1 + Z_{b_2c_2}b_2c_2 - Z_{b_2c_1}^b b_2c_1 \\
 &\quad - Z_{a_1b_2}^b a_1b_2 - X_{b_1b_2}b_1b_2, \\
 \frac{dc_1}{dt} &= Z_{a_1b_1}a_1b_1 + Z_{b_1c_1}b_1c_1 - Z_{a_1c_1}a_1c_1 + Z_{a_1c_2}^c a_1c_2 \\
 &\quad + Z_{b_1c_2}^c b_1c_2 - Z_{a_2c_1}^c a_2c_1 - Z_{b_2c_1}^c b_2c_1 + X_{c_1c_2}c_1c_2, \\
 \frac{dc_2}{dt} &= Z_{a_2b_2}a_2b_2 + Z_{c_2a_2}c_2a_2 - Z_{b_2c_2}b_2c_2 - Z_{a_1c_2}^c a_1c_2 \\
 &\quad - Z_{b_1c_2}^c b_1c_2 + Z_{a_2c_1}^c a_2c_1 + Z_{b_2c_1}^c b_2c_1 - X_{c_1c_2}c_1c_2.
 \end{aligned} \tag{5}$$

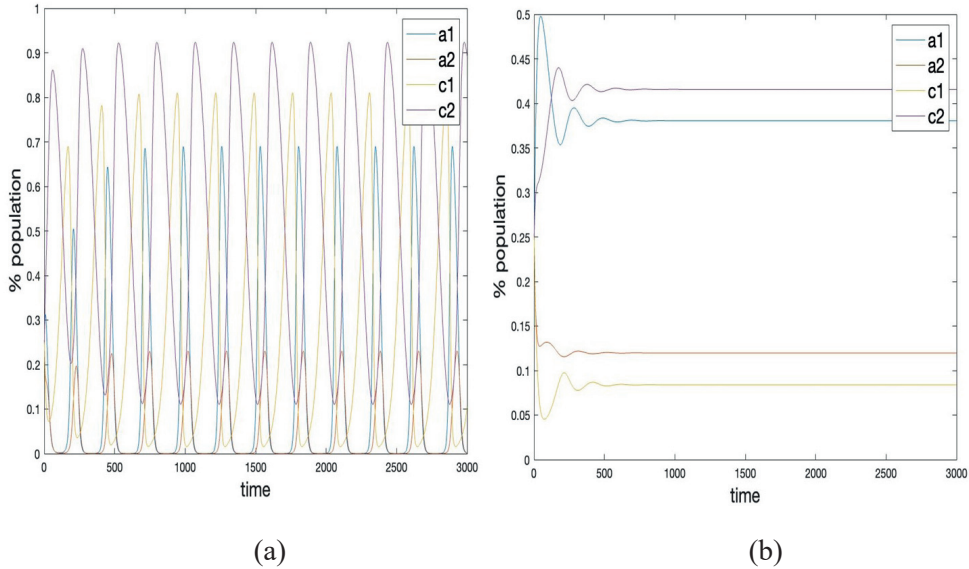
In the next section, we will present numerical results and show that our proposed model can capture various language scenarios.

### Results and discussion

In this section, we present some numerical results obtained by simulations of the proposed model (5). Note that a comprehensive numerical analysis of the model is beyond the scope of this paper. We consider the



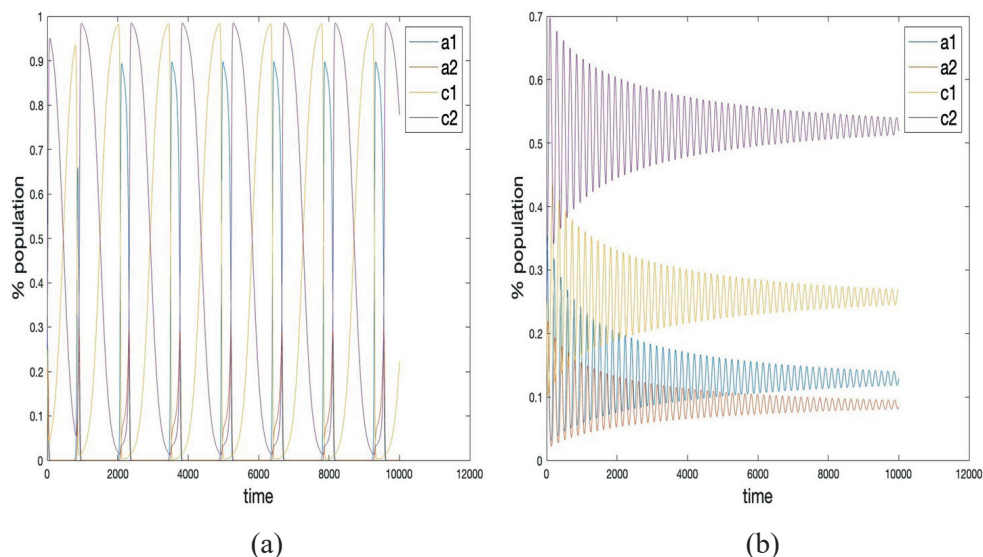
interaction between a bilingual population and a monolingual population of language A (with the disappearance of a monolingual population of language B). As we previously noted, most models assume the disappearance of the bilingual population in such cases.



**Fig. 2.** The dynamics of language A and bilinguals over time.  
The figure shows cyclic motion

**Рис. 2.** Динамика языка А и билингвов с течением времени.  
На рисунке показано циклическое движение

We present the results for different sets of coefficients of the proposed model in Fig. 2 and 3. One can see that our model can predict different scenarios for the coexistence of the bilingual population and the monolingual population of language A. For example, our model can result in cyclic dynamics, i.e., periodically varying populations in each category, as is shown in Fig. 2a, 3a, and 3b. Also, it can predict constant populations in each category (Fig. 2b). Previous approaches can not produce cyclic motions. We note that cyclic dynamics appear in cell population models, which is an indication of cells being alive. In general, it is difficult (mathematically) to understand parameter ranges when cyclic dynamics can occur.



**Fig. 3.** The dynamics of language A and bilinguals over time.

The figure shows cyclic motion. Different coefficients

**Рис. 3.** Динамика языка А и билингвов с течением времени.

Рисунок показывает циклическое движение. Различные коэффициенты

### Conclusions

In this paper, we have proposed a new mathematical model of language shift. We have introduced the ethnic identity factor, a proxy for more complex variables involved in self-identity and group identity. The resulting mathematical model is rich and naturally allows for the coexistence of all languages without adding diglossia terms. Moreover, our equations can result in cyclic dynamics (periodically varying populations in each category), which makes a connection with cell population models. The resulting mathematical model can serve as a useful tool for studying language situations and developing measures to preserve and revitalize endangered languages.

### References / Литература

1. Hale K. Language endangerment and the human value of linguistic diversity. *Language* (Baltimore). 1992;68(1):35-42.
2. Harrison KD. *When languages die: The extinction of the world's languages and the erosion of human knowledge*. Oxford University Press. 2008.
3. Nettle D, Romaine S, et al. *Vanishing voices: The extinction of the world's languages*. Oxford University Press on Demand. 2000.

4. Grenoble L, Whaley L. *Saving languages: An introduction to language revitalization*. Cambridge University Press. 2005.
5. Grenoble L. Language contact in the east slavic contact zone. *Balkanistica*. 2015;28:225-250.
6. Abrams D, Strogatz S. Modelling the dynamics of language death. *Nature*. 2003;424(6951):900-900.
7. Patriarca M, Leppänen T. Modeling language competition. *Physica A: Statistical Mechanics and its Applications*. 2004;338(1-2):296-299.
8. Patriarca M, Heinsalu E. Influence of geography on language competition. *Physica A: Statistical Mechanics and its Applications*. 2009;388(23):174-186.
9. Kandler A, Steele J. Ecological models of language competition. *Biological Theory*. 2008;3:164-173.
10. Isern N, Fort J. Language extinction and linguistic fronts. *Journal of the Royal Society Interface*. 2014;11(94):20140028.
11. Mira J, Paredes A. Interlinguistic similarity and language death dynamics. *Europhysics letters*. 2005;69(6):1031.
12. Minett J, SY Wang W. Modelling endangered languages: The effects of bilingualism and social structure. *Lingua*. 2008;118(1):19-45.
13. Castelló X, Loureiro-Porto L, Eguíluz V, Miguel MS. The fate of bilingualism in a model of language competition. *Advancing social simulation: the first World Congress*. Springer. 2007.
14. Edwards J. Bilingualism and multilingualism: Some central concepts. *The Handbook of Bilingualism and Multilingualism*, Second Edition. Chichester: Wiley-Blackwell. 2012.
15. Kandler A, Unger R, Steele J. Language shift, bilingualism and the future of britain's celtic languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2010;365(1559):3855-3864.
16. Kandler A, Unger R. Modeling language shift. *Diffusive spreading in nature, technology and society*. Springer. 2023.
17. Fought C. Language and ethnicity. In *The Cambridge Handbook of Sociolinguistics*. Cambridge University Press. 2011.
18. Goffman E. *The Presentation of Self in Everyday Life*. London: Allen Lane. 1969.
19. Waters MC. The social construction of race and ethnicity: Some examples from demography. *American diversity: A demographic challenge for the twenty-first century*. 2002.
20. Fishman JA. Bilingualism and biculturism as individual and as societal phenomena. *Journal of Multilingual and Multicultural Development*. 1980;1(1):3-15.
21. Eckert P. Where do ethnolects stop? *International Journal of Bilingualism*. 2008;12(1/2):25-42.

22. Eckert P. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*. 2012;41:87-100.

23. Bucholtz M, Hall K. Language and identity. In Alessandro Duranti, editor, *A Companion to Linguistic Anthropology*. Blackwell, Malden, MA/Oxford. 2003.

### ***About the authors***

*Dmitry A. AMMOSOV* – Cand. Sci. (Physics and Mathematics), Post Doctoral Fellow, Khalifa University of Science and Technology, ORCID: 0000-0002-8616-9397, ResearcherID: AAY-5903-2021, Scopus Author ID: 57208263321, SPIN: 3408-3737, e-mail: [dmitrii.amosov@ku.ac.ae](mailto:dmitrii.amosov@ku.ac.ae)

*Yalchin EFENDIEV* – Ph.D. (Applied Mathematics), Mobil Chair Professor, Texas A&M University, ORCID: 0000-0001-9626-303X, ResearcherID: F-2777-2015, Scopus Author ID: 12039192600, e-mail: [yalchinrefendiev@gmail.com](mailto:yalchinrefendiev@gmail.com)

*Lenore GRENoble* – Ph.D. (Slavic Linguistics), John Matthews Manly Distinguished Service Professor, University of Chicago, ORCID: 0000-0001-8810-7395, Researcher ID: H-5469-2017, Scopus Author ID: 6506649857, SPIN 3405-6850, e-mail: [grenoble@uchicago.edu](mailto:grenoble@uchicago.edu)

### ***Об авторах***

*АММОСОВ Дмитрий Андреевич* – кандидат физико-математических наук, научный сотрудник (постдокторант), Университет науки и технологий Халифы, ORCID: 0000-0002-8616-9397, ResearcherID: AAY-5903-2021, Scopus Author ID: 57208263321, SPIN: 3408-3737, e-mail: [dmitrii.amosov@ku.ac.ae](mailto:dmitrii.amosov@ku.ac.ae)

*ЭФЕНДИЕВ Ялчин* – Ph.D. (прикладная математика), профессор кафедры Mobil, Техасский университет А&М, ORCID: 0000-0001-9626-303X, ResearcherID: F-2777-2015, Scopus Author ID: 12039192600, e-mail: [yalchinrefendiev@gmail.com](mailto:yalchinrefendiev@gmail.com)

*ГРЕНОБЛЬ А. Ленора* – Ph.D. (славянское языкознание), заслуженный профессор имени Джона Мэтьюза Мэнли, Чикагский университет, ORCID: 0000-0001-8810-7395, Researcher ID: H-5469-2017, Scopus Author ID: 6506649857, SPIN 3405-6850, e-mail: [grenoble@uchicago.edu](mailto:grenoble@uchicago.edu)

### ***Authors' contribution***

*Ammosov D.A.* – conceptualization, methodology, software, validation, investigation, data curation, writing – original draft, writing – review & editing, visualization

*Efendiev Ya.* – conceptualization, methodology, software, validation, investigation, formal analysis, data curation, writing – original draft, writing – review & editing, visualization, supervision, project administration

*Grenoble L.* – conceptualization, methodology, validation, investigation, data curation, writing – original draft, writing – review & editing, supervision, project administration

### ***Вклад авторов***

*Аммосов Д.А.* – разработка концепции, методология, программное обеспечение, верификация данных, проведение исследования, администрирование данных, создание черновика рукописи, редактирование рукописи, визуализация

*Эфендиев Я.* – разработка концепции, методология, программное обеспечение, верификация данных, проведение исследования, проведение статистического анализа, администрирование данных, создание черновика рукописи, редактирование рукописи, визуализация, руководство исследованием, администрирование проекта

*Гренобль Л.* – разработка концепции, методология, верификация данных, проведение исследования, администрирование данных, создание черновика рукописи, редактирование рукописи, руководство исследованием, администрирование проекта

### ***Conflict of interests***

The authors Efendiev Ya. and Grenoble L. are members of editorial board of the journal “Arctic XXI Century”. The authors are not aware of any other potential conflict of interest relating to this article

### ***Конфликт интересов***

Авторы Эфендиев Я. и Гренобль Л. являются членами редакционной коллегии журнала «Арктика XXI век». Авторам неизвестно о каком-либо другом потенциальном конфликте интересов, связанном с этой рукописью

*Submitted / Поступила в редакцию 25.02.2025*

*Accepted / Принята к публикации 14.03.2025*