

**ЛИНГВИСТИЧЕСКАЯ СЕГМЕНТАЦИЯ И
ФОРМИРОВАНИЕ АЛГОРИТМИЧЕСКИХ КОКОНОВ:
ПРЯМЫЕ И МОДЕРАЦИОННЫЕ ЭФФЕКТЫ В
МНОГОЯЗЫЧНОМ ЦИФРОВОМ КОНТЕКСТЕ**

Аннотация. В статье рассматривается влияние языковых экосистем на формирование алгоритмических «коконов» в многоязычной цифровой среде Казахстана. Цель исследования состоит в выявлении прямого и модерационного воздействия языкового контекста на степень информационной замкнутости пользователей. Эмпирическую основу работы составили данные опроса 200 пользователей цифровых платформ, а также качественные ответы респондентов, раскрывающие особенности их взаимодействия с алгоритмическими рекомендациями. Результаты показывают, что пользователи казахоязычного цифрового пространства чаще сталкиваются с более ограниченной и повторяющейся информационной средой, тогда как русскоязычные пользователи имеют доступ к более разнообразному и транснационально связанному контенту. Установлено также, что цифровые навыки и активные информационные практики сильнее снижают степень алгоритмической замкнутости именно в казахоязычной среде. Сделан вывод о том, что язык выступает важным структурным фактором

алгоритмического неравенства и должен учитываться при анализе цифровой коммуникации в многоязычных обществах.

Ключевые слова: Языковая Сегментация, Алгоритмический Кокон, Цифровая Коммуникация, Алгоритмическое Неравенство, Многоязычная Среда.

Ge Ange

Master

Al-Farabi Kazakh National University

Almaty, Kazakhstan

**LINGUISTIC SEGMENTATION AND ALGORITHMIC
COCOON FORMATION:
DIRECT AND MODERATING EFFECTS IN A MULTILINGUAL
DIGITAL CONTEXT**

Abstract. The article examines the influence of linguistic ecosystems on the formation of algorithmic cocoons in Kazakhstan's multilingual digital environment. The purpose of the study is to identify the direct and moderating effects of linguistic context on the degree of users' informational enclosure. The empirical basis of the study consists of survey data from 200 digital platform users, as well as qualitative responses that reveal specific features of their interaction with algorithmic recommendations. The results show that users of the Kazakh-language digital space more often encounter a limited and repetitive information environment, whereas Russian-language users have access to more diverse and transnationally connected content. The study also finds that digital skills and active information practices have a stronger effect in reducing algorithmic enclosure in the Kazakh-language environment. The article concludes that language acts as an important structural factor of algorithmic inequality and should be taken into account when analyzing digital communication in multilingual societies.

Keywords: Linguistic Segmentation, Algorithmic Cocoon, Digital Communication, Algorithmic Inequality, Multilingual Environment.

Introduction

Digital platforms have become one of the main infrastructures through which contemporary users receive public information, follow social events, and participate in everyday communication. In this process, algorithmic recommendation systems play an increasingly important role, because they select, rank, and personalize content according to users' previous behavior and platform interactions. Although these systems help reduce information overload and make digital communication more efficient, they may also narrow the range of information available to users and repeatedly expose them to similar viewpoints. Such a tendency contributes to the formation of bounded informational environments, commonly described as algorithmic cocoons or filter bubbles [5,6].

The issue of algorithmic cocoon formation is closely related to the broader problem of digital inequality. Existing research shows that unequal digital experiences are shaped not only by access to technologies, but also by differences in skills, resources, and users' ability to navigate online environments [2,7]. However, in multilingual societies, these factors cannot be fully understood without considering the linguistic structure of digital communication. Language influences what content is available, how widely it circulates, and which information flows become visible to different groups of users. Therefore, linguistic segmentation may become an important condition shaping algorithmically mediated information exposure.

Kazakhstan provides a relevant context for examining this problem. The country's digital environment is characterized by the coexistence of Kazakh- and Russian-language informational spaces. These linguistic ecosystems function within the same national media environment, but they differ in content volume, thematic diversity, and degree of connection with wider information networks. Russian-language users may have access to a broader and more transnational information space, while Kazakh-language users may rely more strongly on localized content. As a result, the same algorithmic mechanisms can produce different informational outcomes depending on the dominant language of platform use.

Despite the growing importance of digital communication in Kazakhstan and

Central Asia, the role of linguistic segmentation in the formation of algorithmic cocoons remains insufficiently studied. Existing research has paid considerable attention to digital inequality, recommendation systems, and media agenda formation, but language is often treated as a secondary factor rather than as a structural condition of digital information exposure [3]. This creates a research gap concerning how linguistic ecosystems influence the degree of algorithmic enclosure and how they shape the relationship between users' digital capacities, information practices, and cocoon tightness.

The purpose of this study is to analyze the role of linguistic ecosystems in the formation of algorithmic cocoons in Kazakhstan's multilingual digital environment. The study examines both the direct effect of linguistic context on algorithmic cocoon tightness and its moderating role in the relationship between users' digital capacities, behavioral information practices, and informational enclosure.

The scientific novelty of the study lies in considering language as a structural factor of algorithmic inequality. Unlike approaches that focus mainly on individual preferences, platform mechanisms, or general digital skills, this article emphasizes that linguistic ecosystems create different conditions for the operation of algorithmic recommendation systems. This makes it possible to explain how parallel informational environments may emerge within one society and why users with different dominant languages may experience different levels of information diversity.

The theoretical significance of the study is connected with the extension of digital inequality and agenda-setting perspectives to a multilingual Central Asian context [4]. The practical significance lies in identifying the importance of expanding and diversifying Kazakh-language digital content and developing information literacy measures that take into account the specific constraints of different linguistic ecosystems.

THEORETICAL AND EMPIRICAL ANALYSIS

The concept of the algorithmic cocoon is used here to describe a fluid and continuously varying degree of informational enclosure, shaped through the

interaction of user behavior, platform mechanisms, and broader contextual constraints [5]. Rather than assuming complete isolation—as implied in more rigid notions such as filter bubbles—cocoon tightness captures the extent to which individuals are repeatedly exposed to preference-aligned content. This perspective makes it possible to examine how different factors gradually open up or narrow users' informational horizons.

Digital capacities are understood as a combination of algorithmic awareness, platform literacy, and the practical ability to navigate personalized systems [1]. Algorithmic awareness refers to how well users understand the logic behind content recommendations, while platform literacy involves more hands-on competencies, such as adjusting settings, using advanced search tools, or accessing alternative information channels. Individuals with higher levels of these capacities tend to engage more actively with information environments for example, by diversifying their subscriptions, deliberately seeking out contrasting viewpoints, or taking steps to limit algorithmic tracking. Such practices can reduce dependence on default recommendation systems and, in turn, mitigate cocoon-like effects [8].

In multilingual contexts, linguistic ecosystems can be seen as partially independent content environments. The amount and diversity of content available within a given language shape the range of perspectives users are likely to encounter. More expansive and transnationally connected ecosystems—such as Russian-language digital spaces—draw on a wider pool of content sources, which can dilute algorithmic concentration effects. By contrast, more localized ecosystems—such as those centered on the Kazakh language—tend to have fewer content producers and a narrower thematic range, which may reinforce informational homogeneity even when users possess comparable digital skills.

Language may also condition how individual-level factors translate into algorithmic outcomes. Where the overall diversity of available content is limited, users' own capacities and behaviors play a more decisive role in broadening exposure. In contrast, in environments where diverse content is already abundant, the influence of individual agency may be less pronounced, since even passive

engagement can lead to relatively varied information exposure.

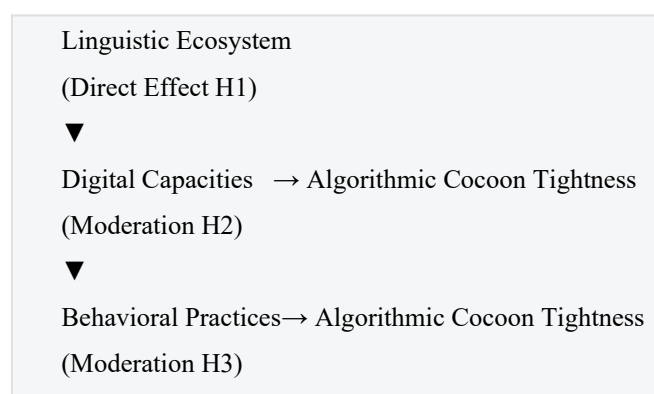
Building on these considerations, three hypotheses are proposed:

H1: Linguistic ecosystem is significantly associated with algorithmic cocoon tightness; Russian-dominant users exhibit lower levels of cocooning than Kazakh-dominant users.

H2: Linguistic ecosystem moderates the relationship between digital capacities and algorithmic cocoon tightness, with a stronger negative association observed in the Kazakh-language environment.

H3: Linguistic ecosystem moderates the relationship between behavioral practices and algorithmic cocoon tightness, with a stronger negative association in the Kazakh-language environment.

Figure 1. Conceptual model



Note. Solid arrows represent direct effects; dashed arrows represent moderating effects.

A sequential explanatory mixed-methods design was adopted to both test the proposed hypotheses and provide contextual interpretation of the results. This design combines quantitative analysis, used to identify relationships between key variables, with qualitative evidence that helps unpack the mechanisms behind these patterns.

The study draws on a sample of 200 digital platform users aged between 18 and 45 in Kazakhstan. Participants were selected through stratified purposive sampling to ensure variation across key demographic characteristics. Stratification criteria included place of residence (urban versus rural), educational background, and primary language use. Respondents were recruited from major urban centers—such as Almaty, Nur-Sultan, Shymkent, and Karaganda—as well as from smaller towns and rural areas, allowing the sample to reflect regional diversity.

Data were collected between March and April 2024 through an online survey administered in either Kazakh or Russian, depending on participants' preferences. In addition to standard survey items, the questionnaire included several open-ended questions inviting respondents to describe their experiences with algorithmic recommendations in their own words.

The main variables were operationalized as follows. Algorithmic cocoon tightness was measured using a five-item Likert scale (ranging from strongly disagree to strongly agree), capturing perceived repetition of content, similarity of viewpoints, and limited exposure to alternative perspectives. Linguistic ecosystem was coded as a binary variable based on respondents' self-reported primary language of platform use and the dominant language of the content they consume. Digital capacities were assessed through a Likert scale covering algorithmic awareness, platform literacy, and users' ability to navigate personalized systems. Behavioral practices were measured through items reflecting active information engagement, such as diversifying sources, intentionally seeking contrasting viewpoints, and adjusting platform settings. Age, gender, and daily internet use were included as control variables to account for potential confounding influences.

The quantitative analysis relied on hierarchical regression models, including interaction terms, to examine both direct effects and moderating relationships. To complement this, qualitative responses were analyzed using thematic analysis, with particular attention to how users describe the role of language in shaping their interaction with algorithmic systems and their broader information environments.

Table 1 reports the descriptive statistics and correlations among the main variables. A clear pattern emerges with respect to linguistic context: the linguistic ecosystem is positively associated with algorithmic cocoon tightness ($r = 0.31$, $p < 0.01$), suggesting that users operating primarily in the Kazakh-language environment tend to experience more constrained information exposure.

At the same time, both digital capacities and behavioral practices show moderate negative correlations with cocoon tightness ($r = -0.56$ and $r = -0.52$, respectively, $p < 0.01$). In practical terms, this indicates that users who possess stronger digital skills

and engage more actively with information—such as seeking diverse sources or adjusting platform settings—are less likely to be confined within narrow informational environments.

Table 1. Descriptive Statistics and Correlation Matrix (N = 200)

Variable	M	SD	1	2	3	4
1. Linguistic Ecosystem	–	–	1			
2. Digital Capacities	3.14	0.62	0.27**	1		
3. Behavioral Practices	2.81	0.69	0.23**	0.60**	1	
4. Algorithmic Cocoon Tightness	3.37	0.65	0.31**	-0.56**	-0.52**	1

$p < 0.01$. Linguistic Ecosystem: 0 = Kazakh-dominant, 1 = Russian-dominant.

The regression results, including both direct and interaction effects, are presented in Table 2. Linguistic ecosystem emerges as a significant predictor of algorithmic cocoon tightness ($\beta = 0.312$, $p < 0.001$), reinforcing the earlier finding that users in the Kazakh-language environment tend to experience more constrained information exposure.

More importantly, the interaction terms reveal that the effects of individual-level factors are not uniform across linguistic contexts. The negative association between digital capacities and cocoon tightness becomes stronger within the Kazakh-language ecosystem ($\beta = -0.185$, $p < 0.05$). A similar pattern is observed for behavioral practices ($\beta = -0.143$, $p < 0.05$), indicating that active information engagement plays a more pronounced role in reducing informational enclosure in this context.

Taken together, these results suggest that while linguistic environment directly shapes the degree of algorithmic cocooning, it also conditions how effectively users' skills and behaviors translate into more diverse information exposure.

Table 2. Direct and Moderating Effects Summary (H1–H3)

Hypothesis	Path	Effect	SE	95% CI	Result
H1	Language $\rightarrow$ CT	0.312	0.076	[0.162, 0.462]	Supported
H2	DC $\times$ Language $\rightarrow$ CT	-0.185	0.071	[-0.325, -0.045]	Supported

H3	BP × Language → CT	- 0.143	0.065	[-0.271, -0.015]	Supported
----	--------------------	------------	-------	------------------	-----------

CT = Algorithmic Cocoon Tightness

The simple slope analysis (Table 3) provides further support for the moderating role of linguistic ecosystem. The negative relationships between both digital capacities and cocoon tightness, and behavioral practices and cocoon tightness, are noticeably stronger among Kazakh-dominant users. Specifically, for this group, higher levels of digital capacities ($\beta = -0.42$, $p < 0.001$) and more active information practices ($\beta = -0.35$, $p < 0.001$) are associated with a substantial reduction in informational enclosure.

By comparison, the same relationships are weaker among Russian-dominant users, although they remain statistically significant ($\beta = -0.23$, $p < 0.01$ for digital capacities; $\beta = -0.21$, $p < 0.05$ for behavioral practices). This contrast suggests that individual agency plays a more decisive role in shaping information exposure within the relatively constrained Kazakh-language environment.

Table 3. Simple Slope Analysis by Linguistic Group

Group	Digital Capacities → CT	Behavioral Practices → CT
Kazakh-dominant	$\beta = -0.42^{***}$	$\beta = -0.35^{***}$
Russian-dominant	$\beta = -0.23^{**}$	$\beta = -0.21^{*}$

* $p < .05$, ** $p < .01$, *** $p < .001$

The qualitative evidence aligns closely with the quantitative results and helps illustrate how these patterns are experienced in everyday platform use. Many Kazakh-language users described their feeds as relatively narrow, often characterized by repeated topics and limited variation. Several noted that encountering new or different information required conscious effort—for instance, actively using search functions, following a broader range of accounts, or even switching to Russian-language content to access more diverse perspectives.

By contrast, Russian-language users tended to describe a more varied information environment. Without making particular adjustments, they reported being exposed to content from different regions and a wider range of viewpoints, including diverse political perspectives. For these users, informational diversity appeared to be more

readily embedded in the platform experience itself.

Taken together, these accounts point to the existence of parallel, language-based informational environments. Although users rely on the same underlying algorithmic systems, the outcomes they experience differ substantially depending on the linguistic context in which those systems operate.

RESULTS AND DISCUSSION

The findings point to the linguistic ecosystem as a key structural layer shaping how algorithmic filtering operates in Kazakhstan. The direct relationship between language context and cocoon tightness suggests that differences in content environments matter well before algorithmic personalization comes into play. The more expansive and transnationally connected Russian-language space appears to dilute informational enclosure, whereas the comparatively localized Kazakh-language environment tends to reinforce it. In this sense, linguistic segmentation does not simply accompany algorithmic stratification—it helps produce it, resulting in distinct informational experiences within the same national setting.

The moderating effects offer further insight into how individual agency operates under different structural conditions. The results suggest that digital capacities and active information practices do not have uniform effects across contexts. In more constrained environments, such as the Kazakh-language ecosystem, these individual factors become more consequential. Users with stronger skills or more proactive habits are better able to push against structural limits, and their efforts translate more clearly into broader exposure. By contrast, in environments where content diversity is already relatively high, such as the Russian-language space, structural conditions partially offset differences in individual behavior. As a result, even less active users may still encounter a relatively diverse range of information.

Taken together, these patterns highlight the importance of treating language as more than a background variable. While prior research on algorithmic inequality has emphasized factors such as access, skills, and platform design, language has often remained analytically peripheral. This study shows that linguistic segmentation can function as a foundational dimension of algorithmic stratification. Bringing language

into frameworks such as the digital divide and agenda-setting helps explain how parallel informational agendas can emerge in multilingual societies, shaping not only what people see but also how public discourse develops.

The findings also carry several practical implications. For platform providers, increasing the volume and diversity of Kazakh-language content could help mitigate structurally driven informational constraints. For policymakers, supporting local content production in Kazakh may contribute to a more balanced information environment. At the same time, digital literacy initiatives could benefit from being more context-sensitive, placing greater emphasis on active information strategies in environments where structural diversity is limited.

Several limitations should be noted. The cross-sectional nature of the data does not allow for strong causal claims about the observed relationships. Future research could adopt longitudinal designs to track how shifts in language use or platform engagement influence cocoon formation over time. In addition, as the analysis focuses on Kazakhstan, it remains an open question to what extent these findings apply to other multilingual contexts. Comparative studies would be particularly valuable in assessing the broader relevance of the mechanisms identified here.

Conclusion

The study confirms that linguistic ecosystems play a significant role in shaping algorithmic cocoon formation in Kazakhstan's multilingual digital environment. The findings show that users operating mainly within the Kazakh-language digital space are more likely to experience a narrower and more repetitive information environment, while Russian-dominant users tend to encounter more diverse and transnationally connected content. This demonstrates that algorithmic enclosure is not determined only by platform mechanisms or individual preferences, but is also strongly influenced by the linguistic structure of available digital content. In this regard, language should be understood as a structural factor of algorithmic inequality, since different linguistic ecosystems create unequal conditions for users' access to diverse information.

The results also show that linguistic context moderates the influence of digital

capacities and behavioral information practices on cocoon tightness. In the more constrained Kazakh-language environment, users' digital skills and proactive information-seeking behaviors have a stronger effect in reducing informational enclosure, whereas in the Russian-language environment broader content diversity partially compensates for differences in individual user behavior. These findings highlight the need to expand and diversify Kazakh-language digital content, support local content production, and develop digital literacy programs adapted to the specific conditions of multilingual information spaces. Future research may further examine this issue through longitudinal and comparative studies in other multilingual societies.

References

1. Bozdag E. Algorithmic literacy and the role of user profiling in filter bubbles // *Social Media + Society*. 2020. Vol. 6, No. 3.
2. Hargittai E. Second-level digital divide: Differences in people's online skills // *First Monday*. 2002. Vol. 7, No. 4.
3. Kabdullina G., Seydehanov S. A. Algorithmic recommendation and news consumption patterns in Almaty: An empirical survey // *Journalism and Communication (KazNU)*. 2024. Vol. 72, No. 1. P. 15–28.
4. McCombs M. E., Shaw D. L. The agenda-setting function of mass media // *Public Opinion Quarterly*. 1972. Vol. 36, No. 2. P. 176–187.
5. Möller J., Helberger N., de Vreese C. H. Measuring the information cocoon: A multi-platform empirical study // *New Media & Society*. 2024. Vol. 26, No. 2. P. 884–902.
6. Pariser E. *The filter bubble: What the internet is hiding from you*. New York: Penguin Press, 2011. 304 p.
7. Van Deursen A. J., van Dijk J. A. The first-level digital divide shifts from inequalities in physical access to inequalities in material access // *New Media & Society*. 2019. Vol. 21, No. 2. P. 354–375.
8. Shin D., Park Y. J. Role of algorithm awareness in user-platform interaction // *Media and Communication*. 2023. Vol. 11, No. 3. P. 210–222.