Religious Language from a Corpus Perspective

Current Findings and Future Directions

Dr. Roman Fisun
Institute of Slavic Studies
Faculty of Languages, Literature and Culture



Universität Regensburg

Universität Regensburg

Contents

1. New religious corpora: OrthRus and IslamRus
2. Case studies:
   - Sins, corruption and other social ills
   - Ukraine and war in religious discourse
   - Patriarch Kirill: country references and linguistic choices
   - Church Slavonic elements in Russian Islamic discourserse
3. Future directions in corpus studies of religious language
   - Russian religiolect(s)
   - Multilingual Ukraine
   - Russian–Bosnian Islamic discourse

Religious language corpora

Universität Regensburg

What is a language corpus?

What is a corpus?

- A large, searchable collection of authentic texts
- Represents real language use
- Texts include metadata such as *year, source, genre* and *author*
- Words are linguistically annotated for *lemma*, *part of speech* and *grammatical features*

A corpus allows us to ask:

- How frequent is a word or expression?
- Where and when is it used?
- Which words typically occur with it?
- How does usage differ across genres, sources or groups of texts?

Corpus = texts + metadata + linguistic annotation

Universität Reg

Text metadata

того , чтобы переубедить этого одиночку , более чем лестные **посулы** самого императора показывают , насколько большое значение

type to search

Token number	**127628507**
Document number	**154075**
doc.author	**Алексей Иванович Сидоров**
doc.genre	**otechnik**
doc.language	**ru**
doc.portal	**azbyka.ru**
doc.pubdate	**1997**
doc.pubyear	**1997**
doc.rubric	**Библиотека**
doc.source	**Сидоров А.И. Преподобный Максим Исповедник и его эпоха в изображении Х. Урса фон Бальтазара // Альфа и Омега. 1997. № 3 (14). С.98-105.**
doc.status	**профессор**
doc.text_id	**154190**
doc.title	**Преподобный Максим Исповедник и его эпоха в изображении Х. Урса фон Бальтазара**
doc.url	**https://azbyka.ru/otechnik/Aleksej_Sidorov/prepodobnyj-maksim-ispovednik-i-ego-epoha-v-izobrazhenii-h-ursa-fon-baltazara/**

Why not work directly with web portals?

1. No control over website content or structure: a corpus provides a stable snapshot
2. Unreliable search functions: results may differ between Yandex and Google; lemmatisation is inconsistent or unavailable, and searches are difficult to reproduce
3. Limited analytical options without corpus tools:
   - no reliable relative frequencies
     10 occurrences in 1,000 words vs. 20 occurrences in 10,000 words
   - relative frequencies are necessary for comparing corpora, sources or semantic fields
   - no systematic trends over time showing when and how often a word or expression is used
   - limited analysis of collocations, grammatical patterns and differences between text groups

Corpora

OrthRus and IslamRus: first large webcorpora of contemporary religious Russian

Sources

- Major Orthodox and Islamic web portals
- Collected in 2025–2026
- Rich metadata: **year · source · genre · author**

Subcorpora

- **Core** — religious and social topics from a religious perspective
- **News** — news reports
- **Q&A** — questions from lay users and answers from clergy
- **Otechnik** (OrthRus only) — books and scholarly texts

Access

corpora.fisun.org

Login *test*, Passwort *test2026*

NoSketchEngine + tools for **syntax · collocations · graphs**

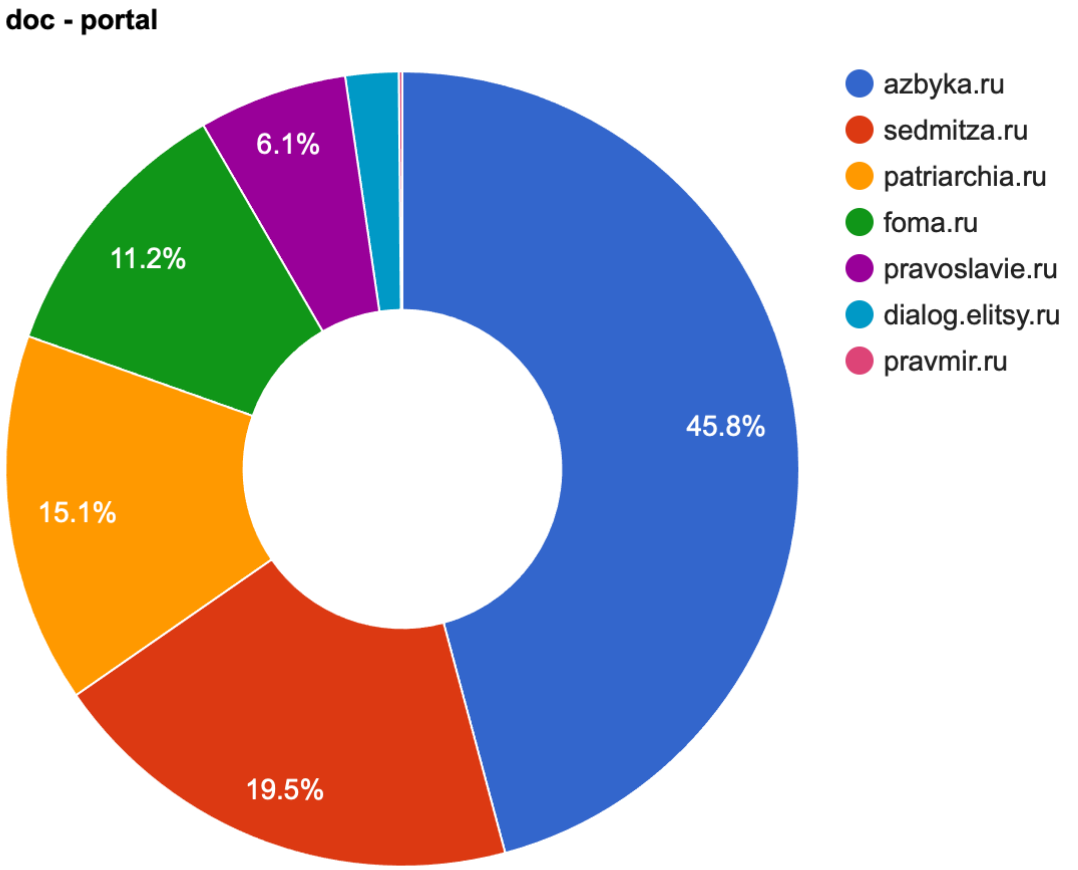
Corpus	OrthRus_v2	IslamRus	RNC Main	RNC Paper
Words	249,330,877	28,899,601	389,471,513	819,818,156

Universität Regensburg

OrthRus_v2: portals

Official ROC communication: patriarchia.ru, sedmitza.ru
Orthodox media: pravoslavie.ru, foma.ru
Pastoral and community communication: dialog.elitsy.ru, pravmir.ru
Religious knowledge and education: azbyka.ru

Church Slavonic liturgy, prayers and unrelated service sections are excluded



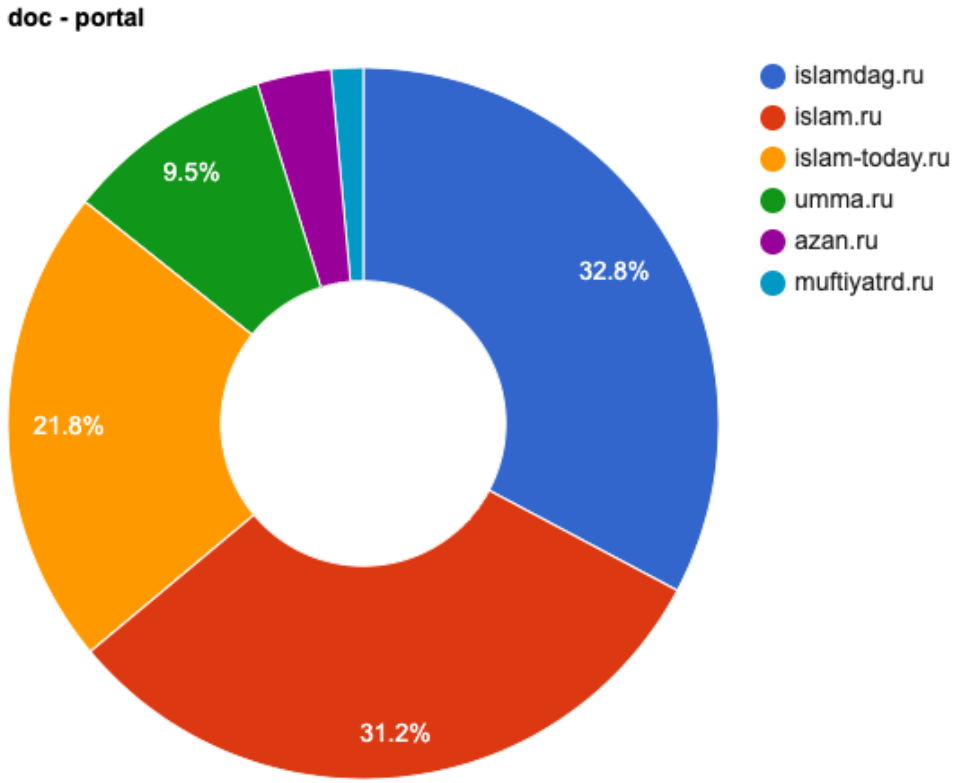
azbyka.ru: 27.8M visits
pravmir.ru: 4.0M visits
foma.ru: 3.4M visits
pravoslavie.ru: 1.96M visits
patriarchia.ru: 783K visits
sedmitza.ru: no data available

IslamRus: portals

Six sources with different functions

- Media, news and advice: islamdag.ru, islam.ru, islam-today.ru
- Religious education: umma.ru, azan.ru
- Institutional communication: muftiyatrd.ru

Regional profile	Portals
Dagestan / North Caucasus	islamdag.ru, muftiyatrd.ru
Tatarstan / Volga–Urals	islam-today.ru; partly azan.ru
Across regions; Volga Islamic influences	islam.ru, umma.ru



Audience context

islam-today.ru: 1.51M visits
umma.ru: ≈400K users; >1M views/month
islam.ru: ≈29K monthly organic search visits

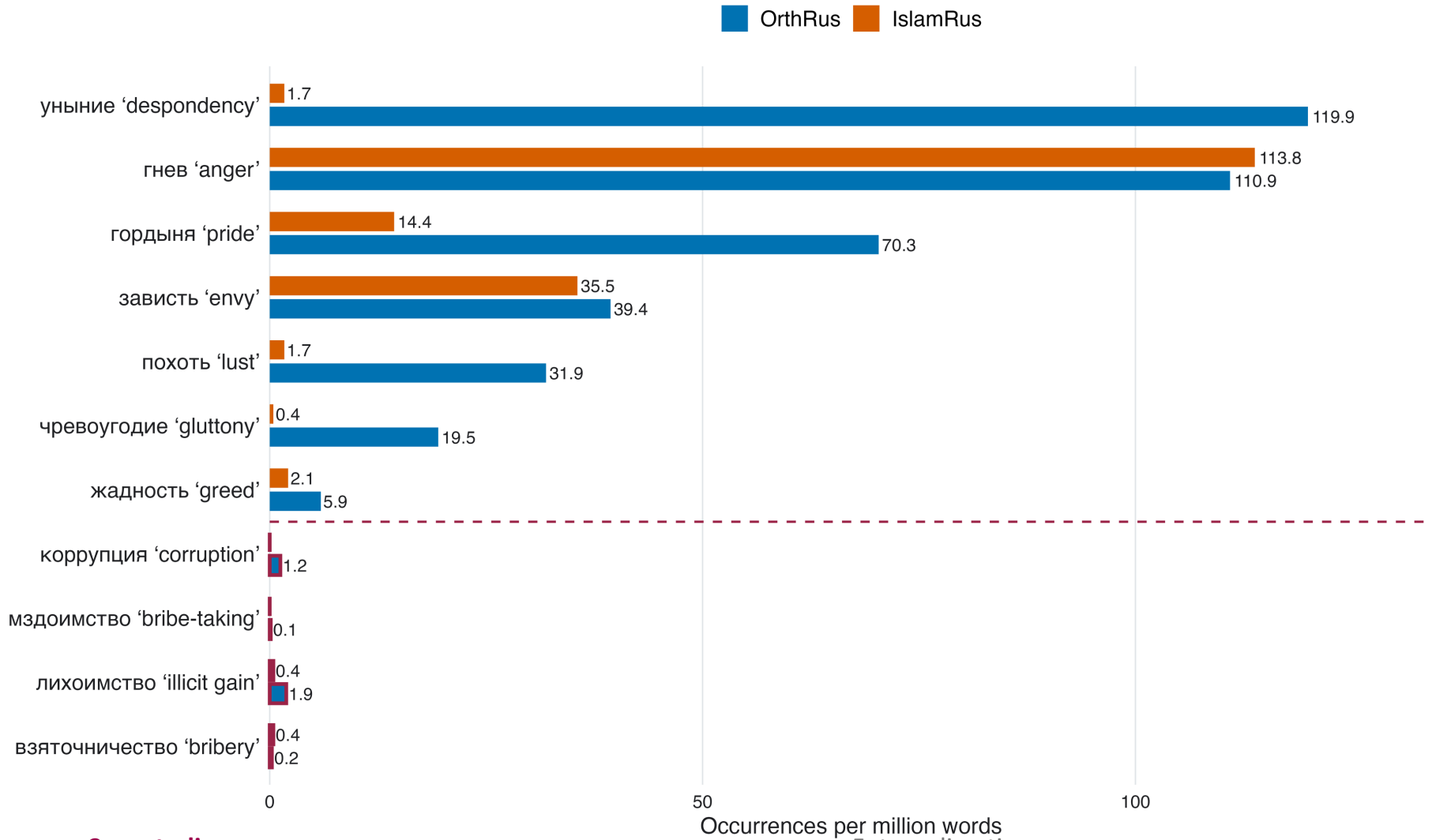
Case studies

What sins are Orthodox Christians and Muslims interested in?

Seven deadly sins and corruption-related vocabulary

Individual lexical labels · occurrences per million words

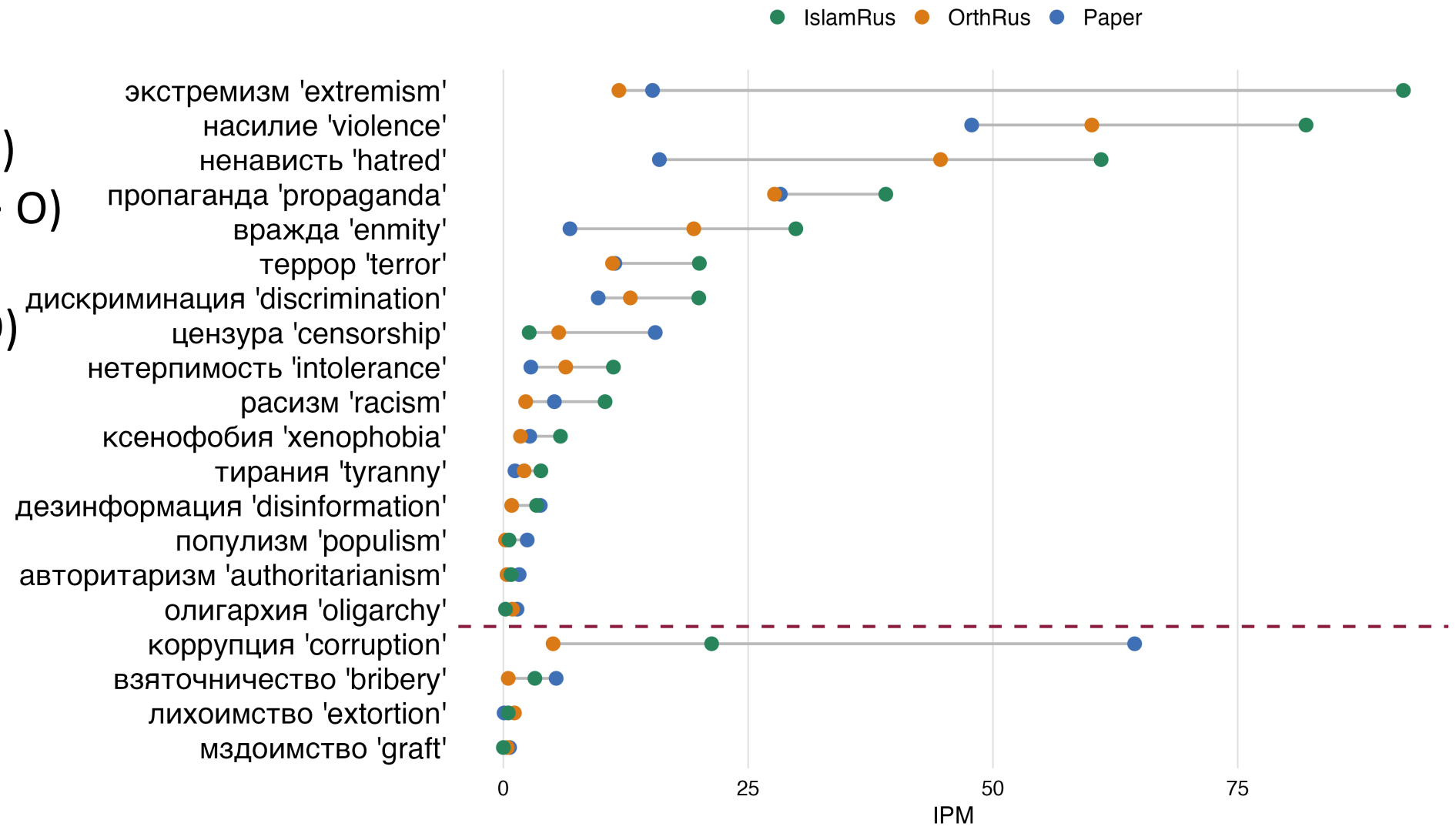
Deadly sins and corruption vocabulary in the Q&A subcorpora



Corruption among other social ills

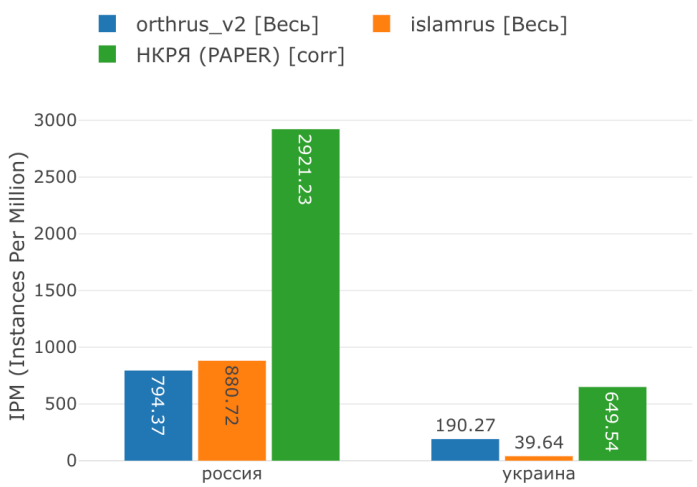
The farther right, the more frequent the term in the corpus

- extremism (I ≫ O/P)
- corruption (P ≫ I ≫ O)
- hatred (I > O ≫ P)
- censorship (P ≫ I/O)

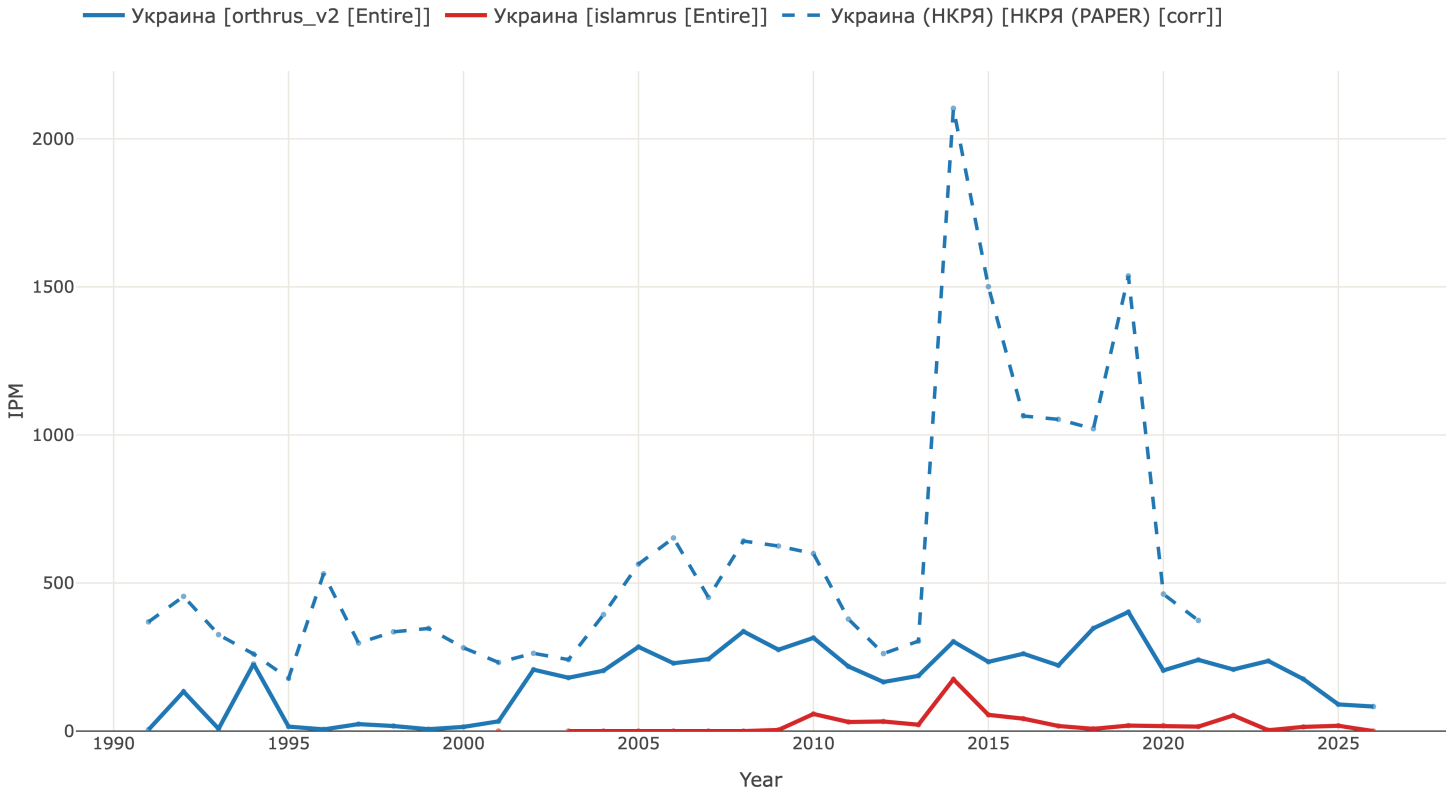


Universität Regensburg

How often is Ukraine mentioned?

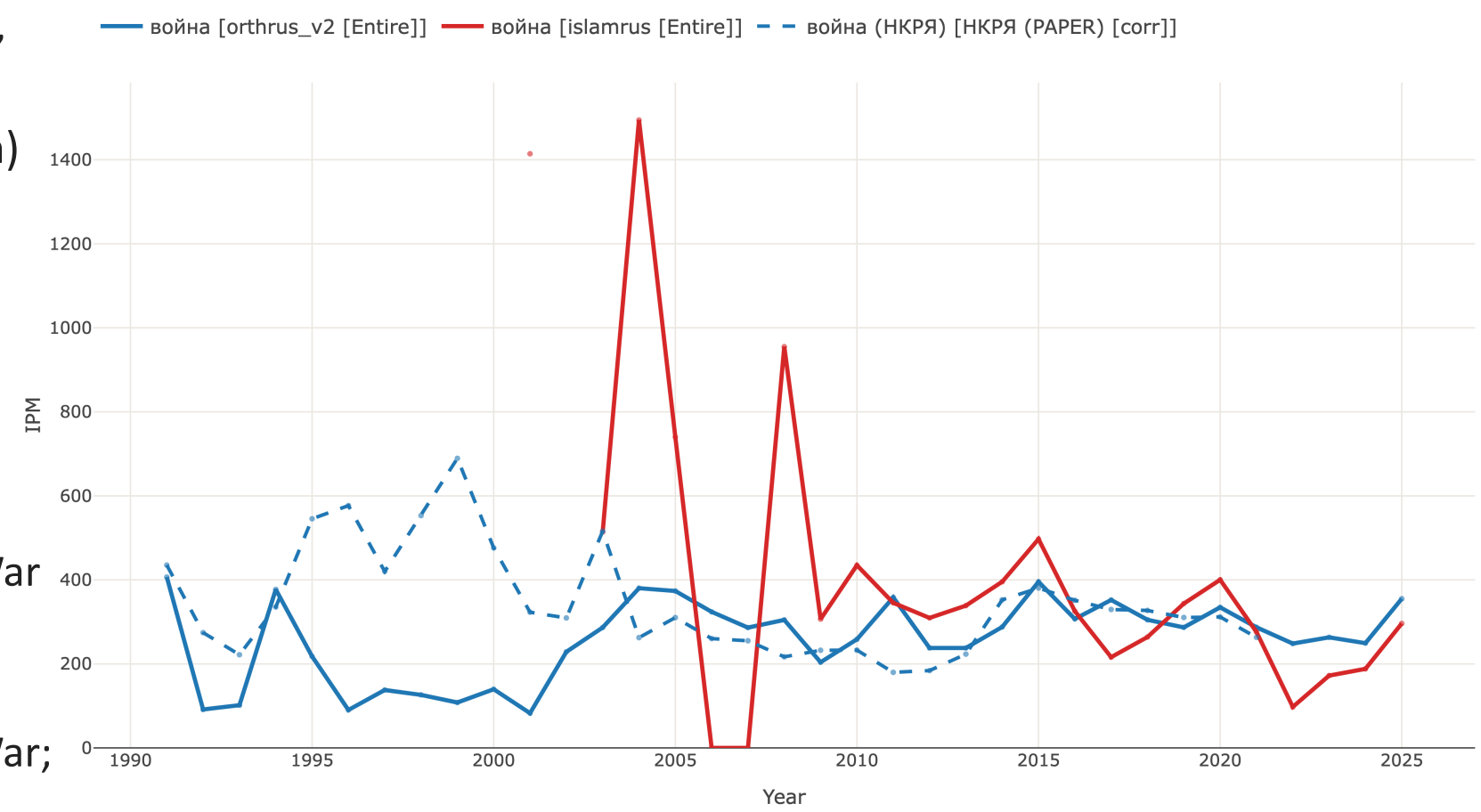


- *Украина* 'Ukraine': OrthRus > IslamRus; both below RNC Paper
- RNC Paper varies from 2000 onwards; the 2014 and 2019 peaks are clearest there
- 2019 Orthodox peak: coincides with recognition of the Ukrainian Orthodox Church of Kiev Patriarchate by Constantinople
- 2022: IslamRus rises; no similar Orthodox peak. No RNC data



And what about *war*?

- No religious peaks: *война* ‘war’, 2014/2022
- *war* falls in 2022 (legal situation)



Islamic peaks: possible contexts

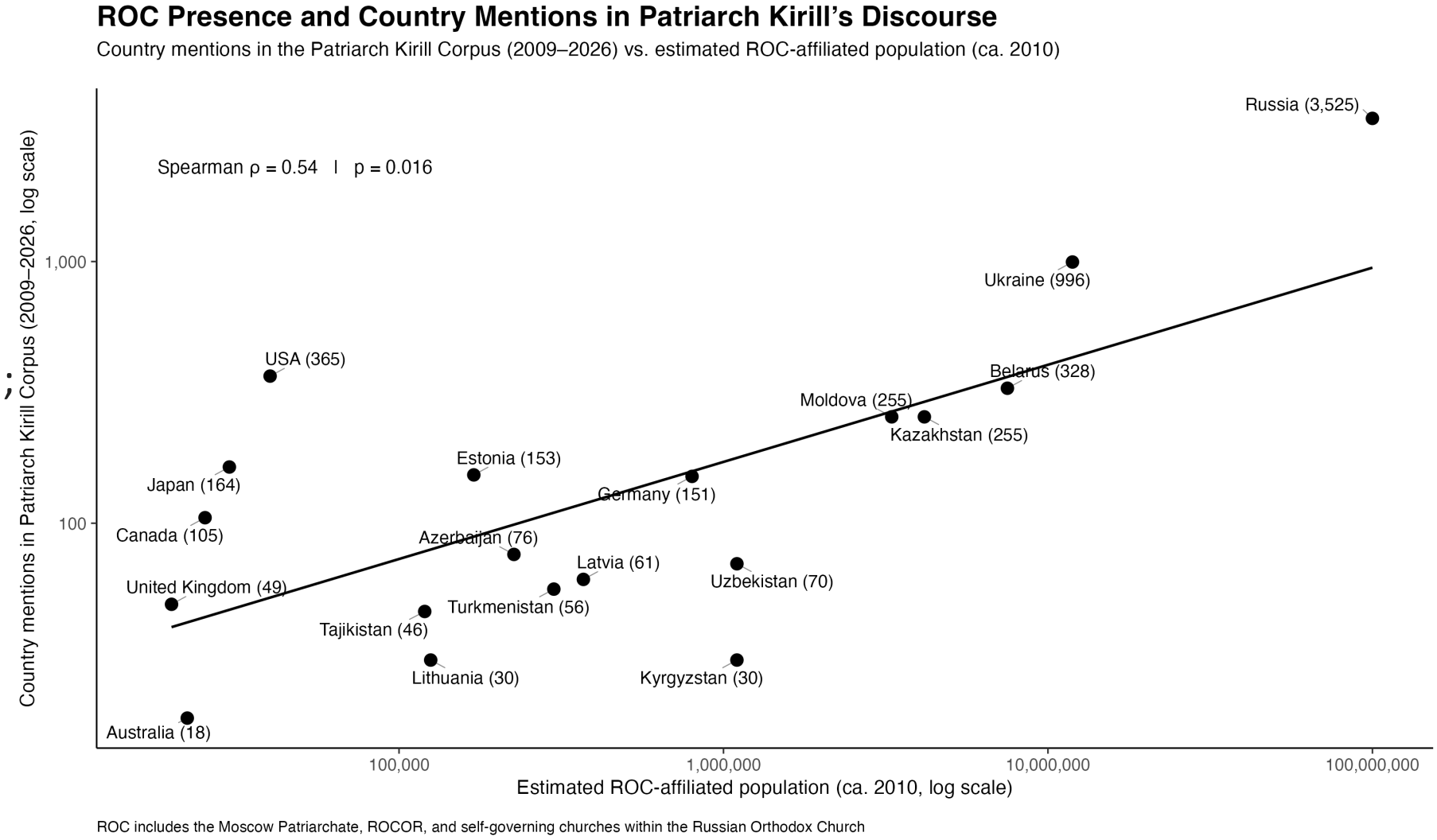
2004: Chechnya and Beslan
2008: Russo-Georgian War
2010: North Caucasus insurgency
2015: Syrian War and ISIS
2020: Second Nagorno-Karabakh War

OrthRus

peak 1994: First Chechen War
Rise from 1999: Second Chechen War; later peaks in 2011 and 2015

Country references in Patriarch Kirill's discourse

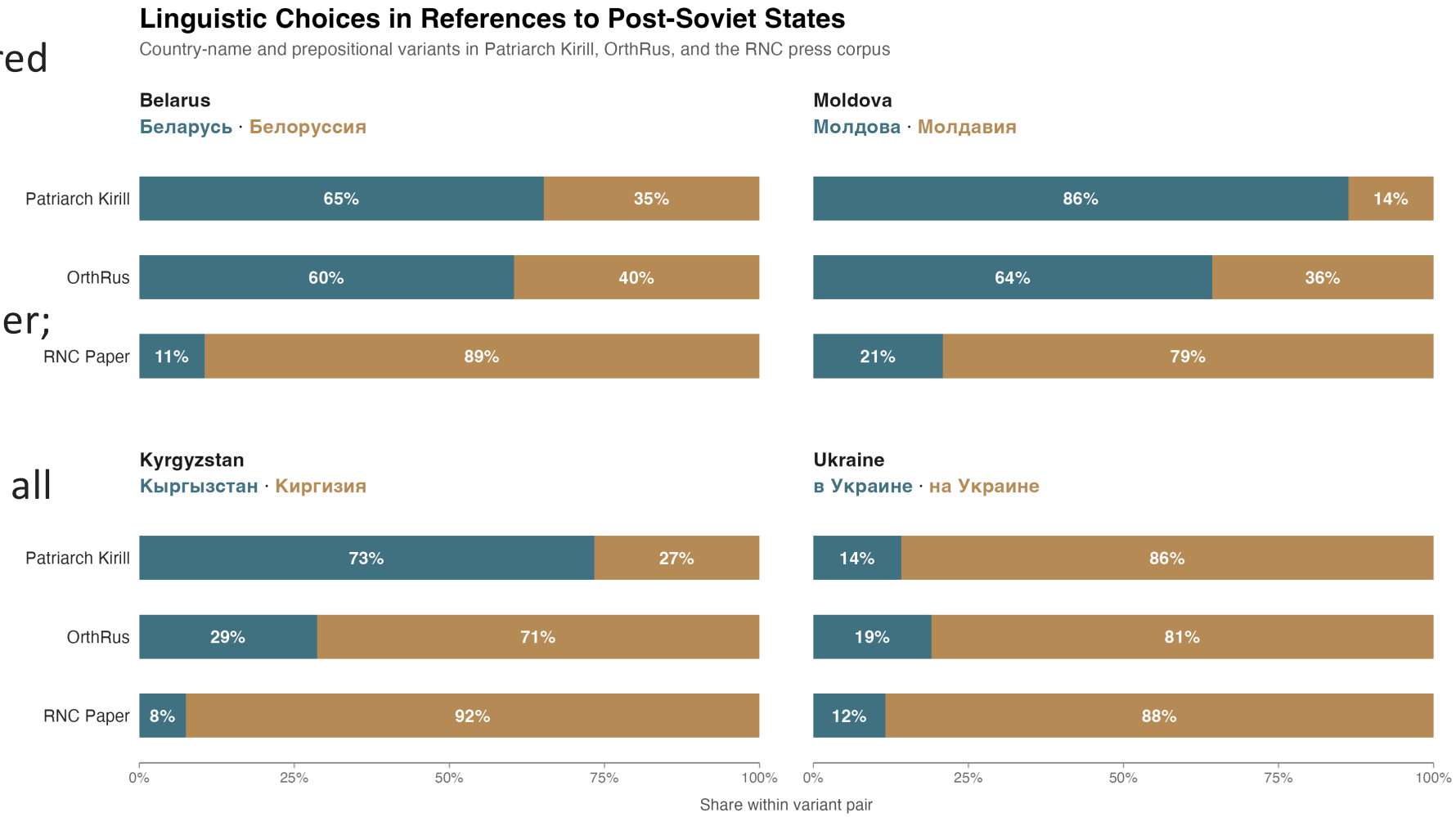
- Moderate association between ROC-affiliated population and country mentions: ρ = .54, p = .016
- Belarus is close to the trend; Ukraine lies above it
- USA, Japan and Canada: above the trend. Kyrgyzstan and Uzbekistan: below it



Universität Regensburg

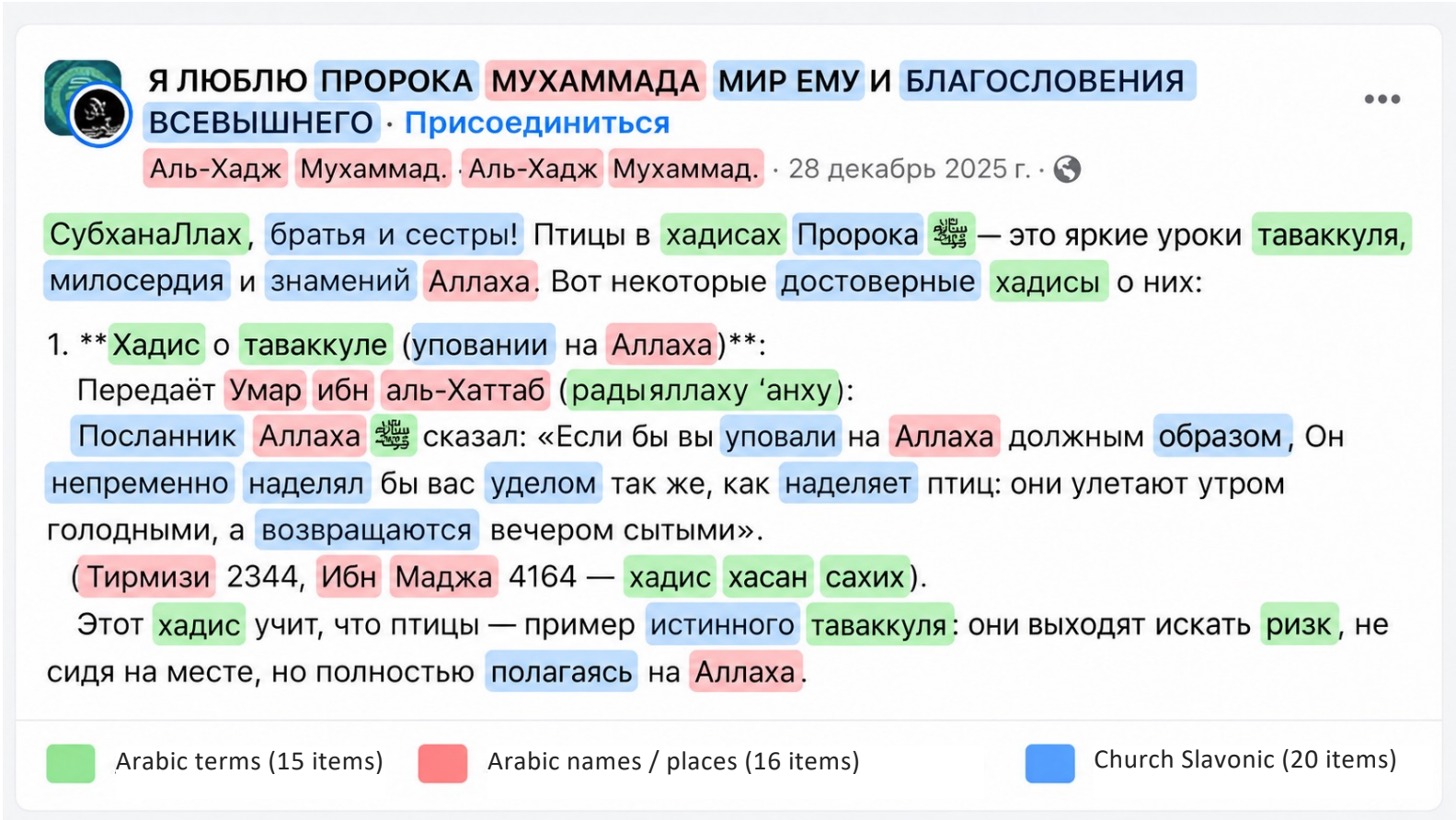
State-preferred country names and prepositions

- Kirill favours state-preferred names: *Беларусь* (65%), *Молдова* (86%), *Кыргызстан* (73%)
- These forms are more frequent than in RNC Paper; OrthRus generally lies between them
- *на Украине* dominates in all three corpora
- *в Украине*: 14% in Kirill, 12% in RNC Press, 19% in OrthRus



Church Slavonicisms in modern Russian religious discourse

Church Slavonic elements in Russian Islamic discourse



facebook.com/groups/226186135872513/posts/1378342843990164/

SubhanAllah, brothers and sisters!
Birds in the hadiths of the Prophet are vivid lessons in *tawakkul* ‘trust in Allah’, mercy, and the signs of Allah. Here are some authentic hadiths about them:

1. A hadith about *tawakkul* ‘trust in Allah’

Umar ibn al-Khattab (may Allah be pleased with him) reported:

The Messenger of Allah said: “If you were to put your trust in Allah as you should, He would surely provide for you just as He provides for the birds: they leave in the morning hungry and return in the evening full.

Material

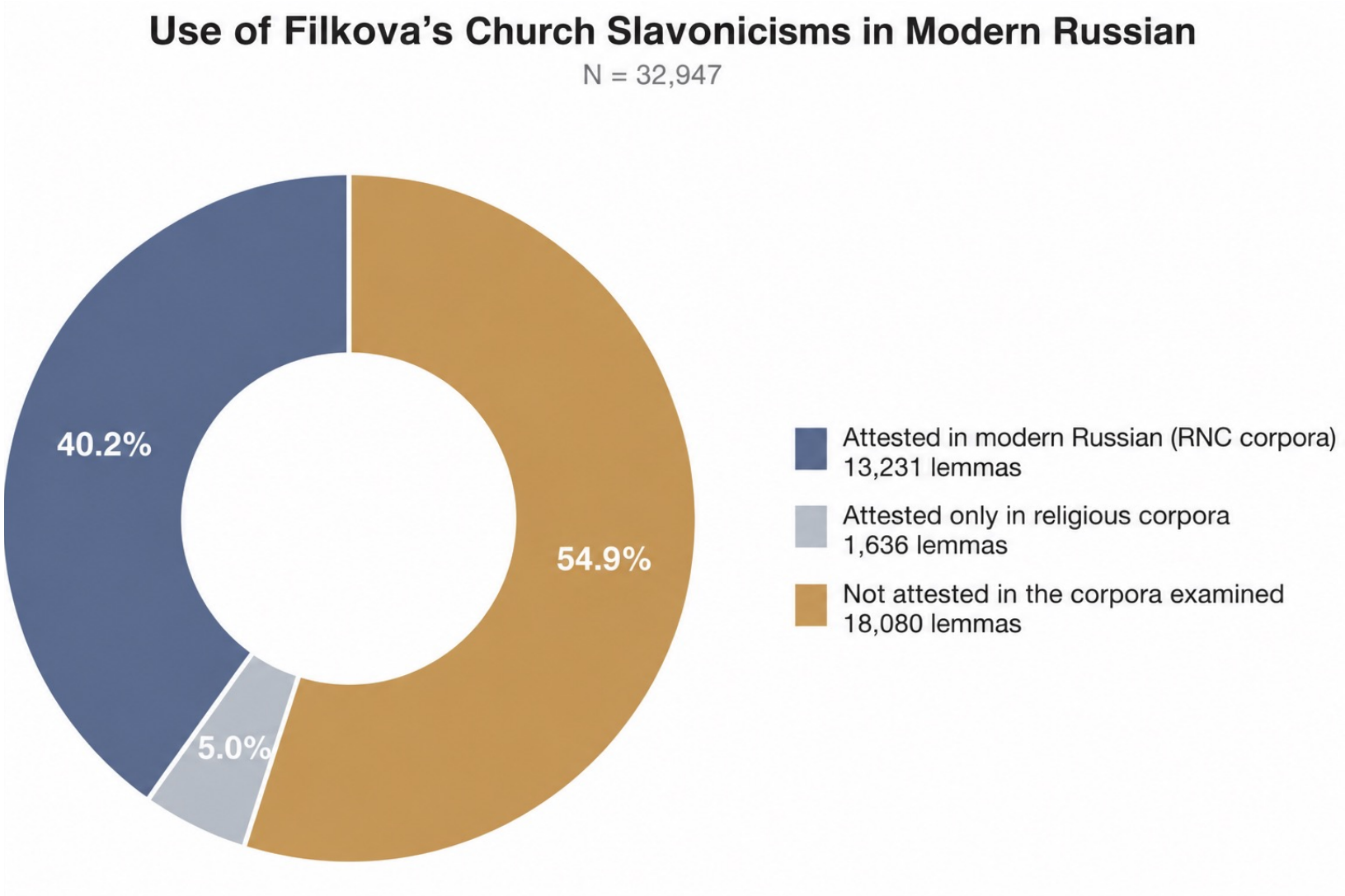
Genetic church slavonicisms: classification by historical origin, regardless of present-day stylistic marking

(1) *глад* – *голод* ‘hunger’

(2) *страна* – *сторона* ‘country’ – ‘side’

(3) *посланник* ‘messenger’

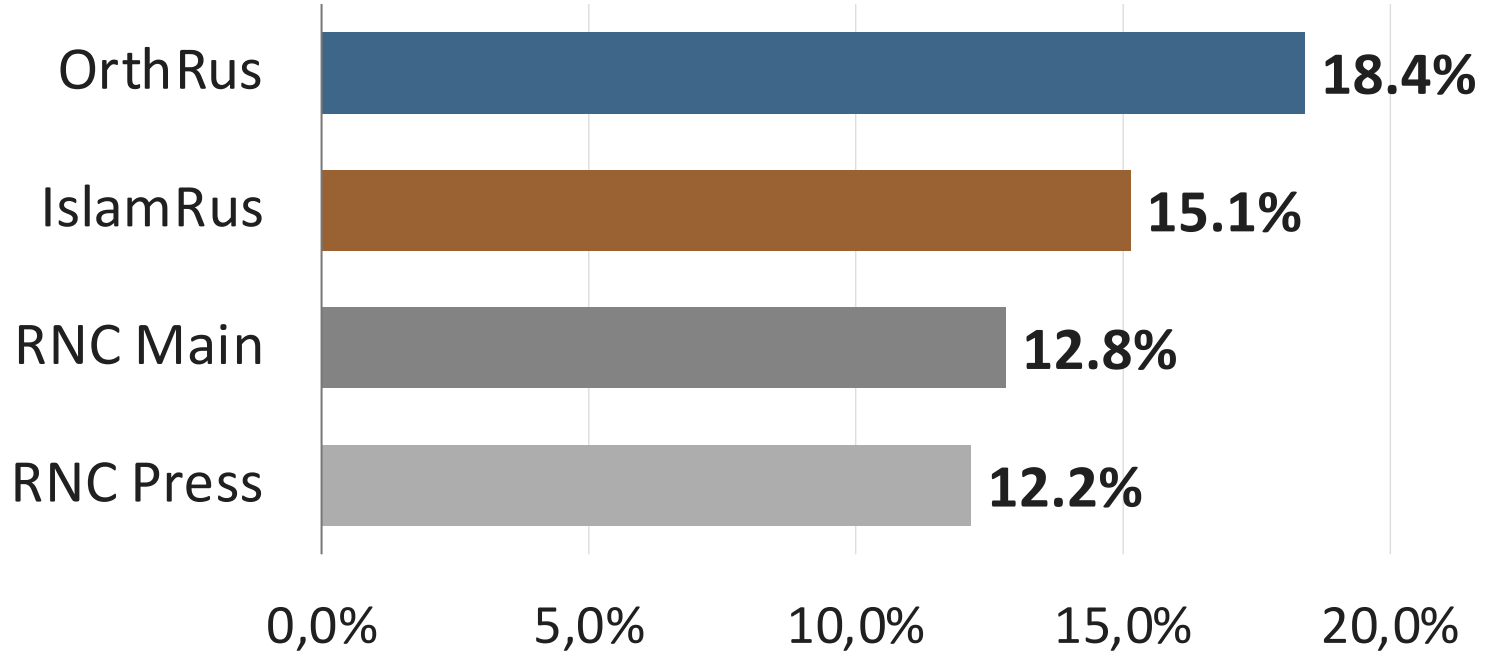
Source: Filkova (1986) *Starobolgarizmy i cerkovnoslavjanizmy v leksike russkogo literaturnogo jazyka*, about 32,000 lemmas



Universität Regensburg

RQ1. Frequency across corpora

How frequent is this vocabulary in religious and general Russian?



15.1%

in IslamRus

Between OrthRus and the reference corpora

IslamRus vs. RNC Paper: about 25% higher relative frequency

Universität Regensburg

RQ2. Shared religious vocabulary

Which words are especially frequent in both religious discourses?

651

Lemmas overrepresented in both religious corpora

r > 2 in IslamRus and OrthRus

пророк ‘prophet’
благословение ‘blessing’
молитва ‘prayer’
благодать ‘grace’

308

of total 13,231

Of these, below expectation in both reference corpora

r < 0 in RNC Main and RNC Press

Standardised residuals (r) show how observed counts differ from statistically expected counts.

RQ3. Differences between religions

How does the use of this vocabulary differ?

Lemma	IslamRus	OrthRus_v2
сподвижник 'companion'	+517.3	−29.0
православный 'Orthodox'	−116.1	+1,324.2
храм 'place of worship'	−100.7	+1,136.0

(4) *Лучший из всех – Пророк Мухаммад* (*мир ему и благословение*) *ушел, ушли и его* ***сподвижники***, *но Ислам остался и распространялся по всему миру*. (IslamRus)

'The best of all, the Prophet Muhammad (peace and blessings be upon him), passed away, as did his companions. Yet Islam remained and spread worldwide.'

(5) *Прожив в мечети около месяца, мужчина по собственному желанию покинул* ***храм***. (IslamRus)

'After living in the mosque for about a month, the man left the place of worship of his own accord.'

(6) *Тимофей был одним из семидесяти апостолов, учеником и ближайшим* ***сподвижником*** *апостола Павла*. (OrthRus v2)

'Timothy was one of the seventy apostles, a disciple and close companion of the Apostle Paul.'

(7) *На следующий день мы пришли в* ***храм*** *уже к концу праздничного богослужения*. (OrthRus v2)

'The next day, we arrived at the church towards the end of the festive service.'

Church Slavonic origin does not determine a word's religious profile. Table values: residuals.

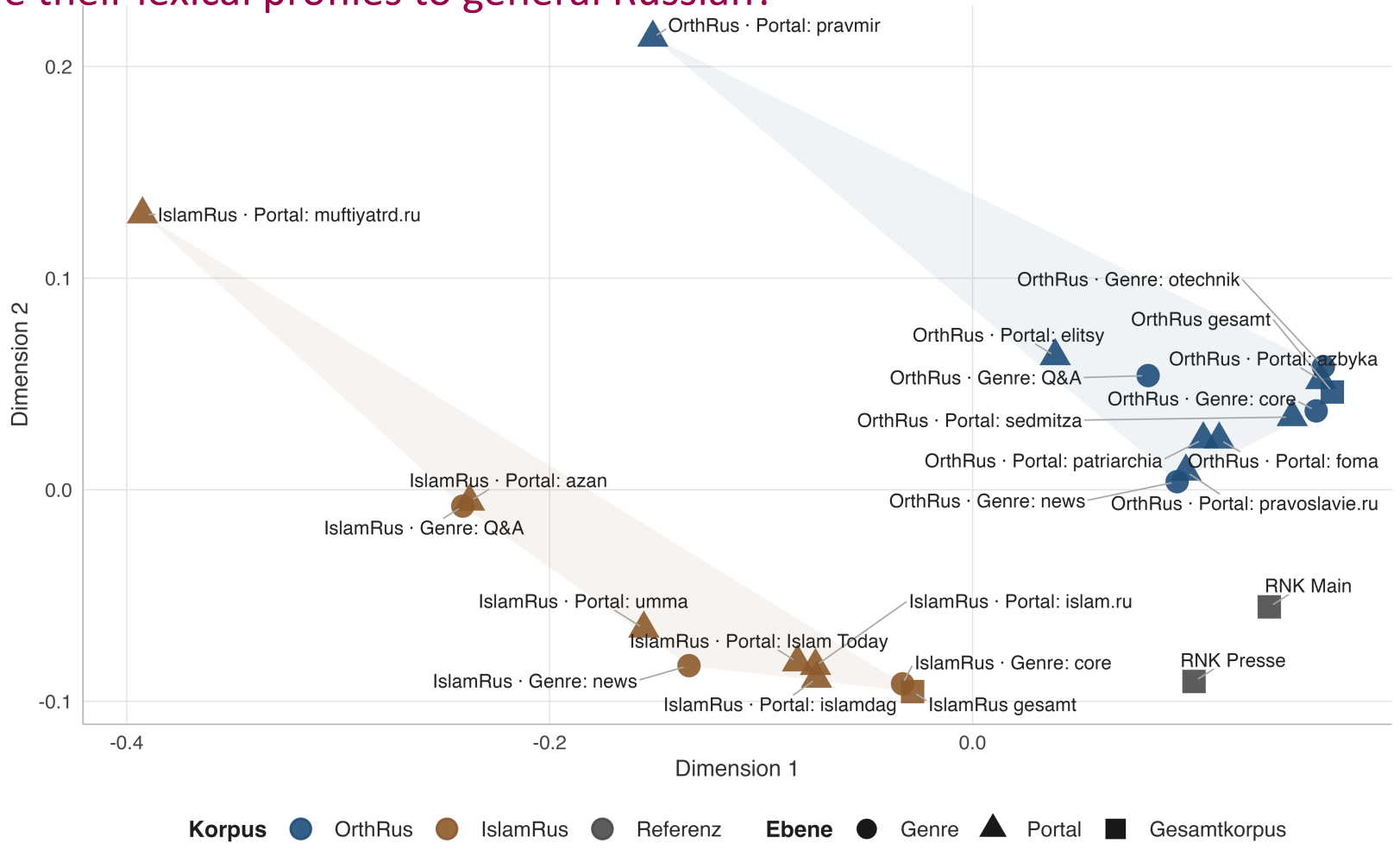
Universität Regensburg

RQ4. Variation by portal and genre

How similar are their lexical profiles to general Russian?

MDS map: Spearman correlations

- Closer points: similar lexical profiles
- Main groups follow religious affiliation
- No clear cross-religious clustering by genre



Future directions in corpus studies of religious language

Universität Regensburg

Habilitation project

Between Shared Linguistic Patterns and Religious Differentiation: Corpus-Based Studies of Slavic Religious Language

How do religious affiliation, multilingualism and historical language contact shape shared and distinctive linguistic patterns?

Three comparative perspectives

A shared language	**A multilingual society**	**One religion, different linguistic settings**
Across religions and Christian denominations in Russian	Across Christian denominations and languages in Ukraine	Across societies, community languages and language attitudes
Russian religiolect(s)	Religious language in Ukraine	Islamic discourses in Russia and Bosnia

Universität Regensburg

Subproject 1. Russian religious language(s)

Is there a shared religious variety?

Key questions

- Shared or community-specific patterns? Vocabulary, phraseology and grammar
- Where do they occur? Liturgical and clerical texts, institutional media, lay communication
- What shapes choices in Islamic discourse? Arabic-derived, Russian and Church Slavonic resources; genre, speaker role and multilingualism

Research programme

- Expand IslamRus; add Catholic, Protestant and Jewish material
- Match genres, institutions and speaker groups; use secular Russian as a reference
- Combine quantitative comparison with qualitative analysis of selected contexts

Aim: distinguish shared patterns from community-specific profiles

Universität Regensburg

Subproject 2. Religious language in multilingual Ukraine

Ukraine's Christian traditions use Ukrainian, Church Slavonic, Russian and Polish in different liturgical, institutional and community settings

Language choice cannot simply be equated with church affiliation.

Key questions

- Which languages dominate in each church, institution, portal and genre?
- Distribution of Church Slavonic, Russian-derived and Polish-derived elements?
- Denomination-specific profiles, or effects of language, genre and historical tradition?

Research programme

- Build comparable Orthodox, Greek Catholic and Roman Catholic corpora
- Record affiliation, language, source, genre and period
- Compare vocabulary, phraseology and grammar; add historical and philological interpretation

Aim: identify which features are confession-specific and wich are better explained by multilingualism or historical contact.

Universität Regensburg

Subproject 3. Islamic discourses in Russia and Bosnia

Background

- Russia: dominant public and interethnic Russian alongside community and heritage languages
- Bosnia: predominantly Slavic-speaking Muslims; different histories of religious contact and language politics

Same religion, different linguistic resources

Working hypothesis

Christian-associated linguistic elements are more integrated into Russian than into Bosnian Islamic discourse.

Key questions

- Arabic terms, local equivalents or Christian-associated expressions?
- Variation: genres, institutions, regions, religious authorities?
- Explanations: multilingualism, attitudes, institutional history?

Research programme

- Match genres, institutions and periods. Record available speaker and language metadata
- Compare frequencies, contexts, collocations and functions
- Cooperation with Klaus Buchenau

Aim: explain how linguistic environments and contact histories shape Islamic discourse

References

Buchenau, K. (2023). *Eine russisch-islamische Synthese? Interaktionsmuster zwischen Christen, Muslimen und Staat im Wolgaraum vom Mittelalter bis in die Gegenwart* (IOS Mitteilungen, Nr. 70). Leibniz-Institut für Ost- und Südosteuropaforschung (IOS). https://www.dokumente.ios-regensburg.de/publikationen/mitteilungen/mitt_70.pdf

Filkova, P. N. (1986) *Starobolgarizmy i cerkovnoslavjanizmy v leksike russkogo literaturnogo jazyka. Učebnyj slovar'. Bd. 1–3.* Sofijski universitet „Kliment Ochridski": Sofija

Kiynova, Ž. K. (2014) *Slavjanizmy kak istoriko-kul'turnyj fenomen: Sinchronija vs. diachronija*. Moskva: Gosudarstvennyj institut russkogo jazyka im. A. S. Puškina

Krysin, L. P. (1996) Religiozno-propovedničeskij stil' i ego mesto v funkcional'no-stilističeskoj sisteme sovremennogo russkogo literaturnogo jazyka. In *Poėtika. Stilistika. Jazyk i kul'tura: Pamjati T. G. Vinokur* (S. 135–138). Moskva: Nauka

Litvinceva, K. V. (2016) Cerkovnoslavjanizm kak leksikografičeskaja pometa. *Vestnik Pravoslavnogo Svjato-Tichonovskogo gumanitarnogo universiteta. Serija III: Filologija* 2(47), 26–44

Litvinceva, K. V., & Ljaševskaja, O. N. (2017) Primenenie kvantitativnych korpusnych metodik dlja vyjavlenija cerkovnoslavjanizmov v sovremennom russkom jazyke. *Vestnik Pravoslavnogo Svjato-Tichonovskogo gumanitarnogo universiteta. Serija III: Filologija* 53, 43–55

Sibgatullina, G. R. (2020). *Languages of Islam and Christianity in Post Soviet Russia*. Leiden & Boston: Brill Rodopi

Vinokur, G. O. (1947). *O slavjanizmach v sovremennom russkom literaturnom jazyke*. *Russkij jazyk v škole* 4

Thank you for your attention!