Russische orthodoxe und islamische Religionssprache digital untersuchen

Korpusaufbau, Analysewerkzeuge und Anwendungen

Dr. Roman Fisun
Institut für Slavistik
Fakultät Sprach-, Kultur- und Literaturwissenschaften

Digitale Slawistik · 50. ÖLT
Universität Salzburg · 20.–21. November 2026

Universität Regensburg

Universität Regensburg

Aufbau des Vortrags

01 Ausgangspunkt und Forschungslücke

02 Korpusaufbau und Korpora

03 Fallstudie I: Korruptionslexik

04 Fallstudie II: Kirchenslavismen

05 Fazit

Universität Regensburg

Forschungslücke: Korpusbasierte Erforschung slavischer Religiolekte

- Ausgangspunkt: Projekt „Korruption und Religion: Das orthodoxe Christentum in Russland und Serbien“ (DFG-HA 2659/11-2; Leitung: Björn Hansen)

Fragestellung: Spricht die russisch-orthodoxe Kirche über Korruption anders als säkulare Akteure?

- Forschungsstand: Bisher überwiegend qualitative Studien; geeignete umfangreiche Korpora zur russischen Religionssprache fehlen (Sibgatullina 2020; Litvinceva & Ljaševskaja 2017)

- Eigene Infrastruktur: Aufbau eines orthodoxen Korpus und Entwicklung zusätzlicher Analysewerkzeuge, da bestehende Korpussysteme nicht alle benötigten Funktionen bieten

- Quelle: meistbesuchte religiöse Web-Portale

- Erweiterung: Vergleich orthodoxer, islamischer und säkularer Diskurse im Russischen

Universität Regensburg

Von Webseiten zu Korpora: Pipeline

1	Rohtexte gewinnen	Webportale auslesen; Metadaten erfassen (Python; HTML-Parsing/APIs)
2	Texte aufbereiten	Bereinigung (Python; XML)
3	Linguistisch annotieren	Tokenisierung, Lemmatisierung, morphologische und syntaktische Annotationen (UDPipe 1.2 – SynTagRus)
4	Korpus kompilieren	.vert-Datei + Registry-Datei (Manatee)
5	Korpus veröffentlichen	NoSketch Engine (Korpussuche); eigene Tools: Graphs, SyntSearch, LemmaSketches (Python/Flask)

Wie kann KI einem Linguisten ohne Informatikausbildung dabei helfen?

Universität Regensburg

KI-Unterstützung in allen Arbeitsschritten

✓ KI	1	Rohtexte gewinnen	Webportale auslesen; Metadaten erfassen (Python; HTML-Parsing/APIs)
✓ KI	2	Texte aufbereiten	Bereinigung (Python; XML)
✓ KI	3	Linguistisch annotieren	Tokenisierung, Lemmatisierung, morphologische und syntaktische Annotationen (UDPipe 1.2 – SynTagRus)
✓ KI	4	Korpus kompilieren	.vert-Datei + Registry-Datei (Manatee)
✓ KI	5	Korpus veröffentlichen	NoSketch Engine (Korpussuche); eigene Tools: Graphs, SyntSearch, LemmaSketches (Python/Flask)

KI kann in jedem Arbeitsschritt beraten und bei der Umsetzung unterstützen.

Universität Regensburg

KI-Werkzeuge in der Korpuslinguistik: Erfahrungen aus der Praxis

- ChatGPT, Codex, DeepSeek: Hilfe beim Programmieren, bei der Fehlersuche und bei technischen Problemen
- KI als Informatikberater: Besonders nützlich bei schlecht dokumentierten Werkzeugen wie Manatee/NoSketch Engine
- Ohne Grundkenntnisse geht es nicht: Programmierung, Serveradministration, Datenformate, IT-Sicherheit
- Vor allem bleibt korpuslinguistische Kompetenz unverzichtbar: Was lässt sich wie untersuchen? Wie sind die Ergebnisse zu interpretieren?

- Meine Erfahrung: Auch ohne Informatikausbildung lassen sich heute in wenigen Monaten eigene Korpora und Analysewerkzeuge entwickeln

Universität Regensburg

Korpora: OrthRus und IslamRus

Große Webkorpora des modernen religiösen Russischen

- **Quellen**: große russischsprachige orthodoxe und islamische Webportale
- **Erhebung**: 2025–2026; Metadaten zu Jahr, Autor, Quelle und Genre
- **Subkorpora**:
  - *Core*: Artikel zu religiösen und gesellschaftlichen Themen aus religiöser Perspektive.
  - *News*: kurze Nachrichten
  - *Q&A*: Fragen von Laien an Geistliche und Antworten darauf
  - *Otechnik* (nur für OrthRus): Größere Texte ganze Bücher, wiss. Arbeiten usw.

Für Zugang über NoSketchEngine: kurze Mail an corpora@fisun.org

Korpora	OrthRus_v2	IslamRus	RNK cerkovno-bogoslovskij	RNK main	RNK paper
Größe, Wörter	249.330.877	28.899.601	5.290.038	389.471.513	819.818.156

Fallstudie I: Korruptionslexik im russischen orthodoxen Religionsdiskurs

- Vorarbeiten deuteten auf Unterschiede zwischen religiösem und säkularem Sprachgebrauch hin

Religiöser Diskurs
srebroljubie ‘Geldgier’
mzdoimstvo ‘Bestechlichkeit’
lichoimstvo ‘Korruption’
koryst’ ‘Eigennutz’

Säkularer Diskurs
korrupcija ‘Korruption’
vzjatočničestvo ‘Bestechung’
kaznokradstvo ‘Veruntreuung’

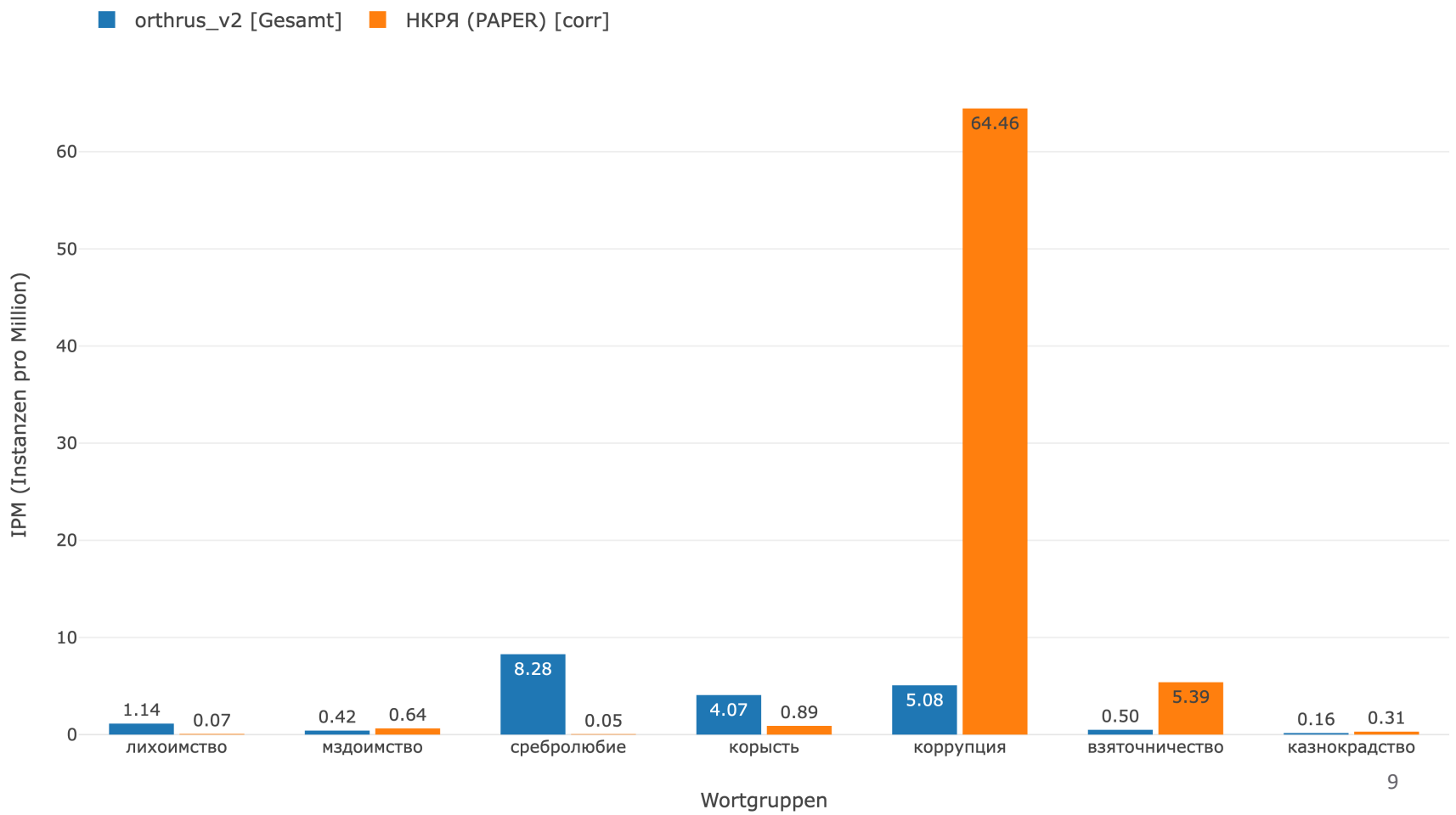
Ziel: Vergleich ihrer Verteilung in OrthRus und säkularem Referenzkorpus.

Universität Regensburg

Graphs: Frequenzen in verschiedenen (Teil-)Korpora

OrthRus_v2 vs. säkulare Presse (RNK)

- Deutliche Unterschiede zwischen religiösem und säkularem Sprachgebrauch: *srebroljubie* vs. *korrupcija*.

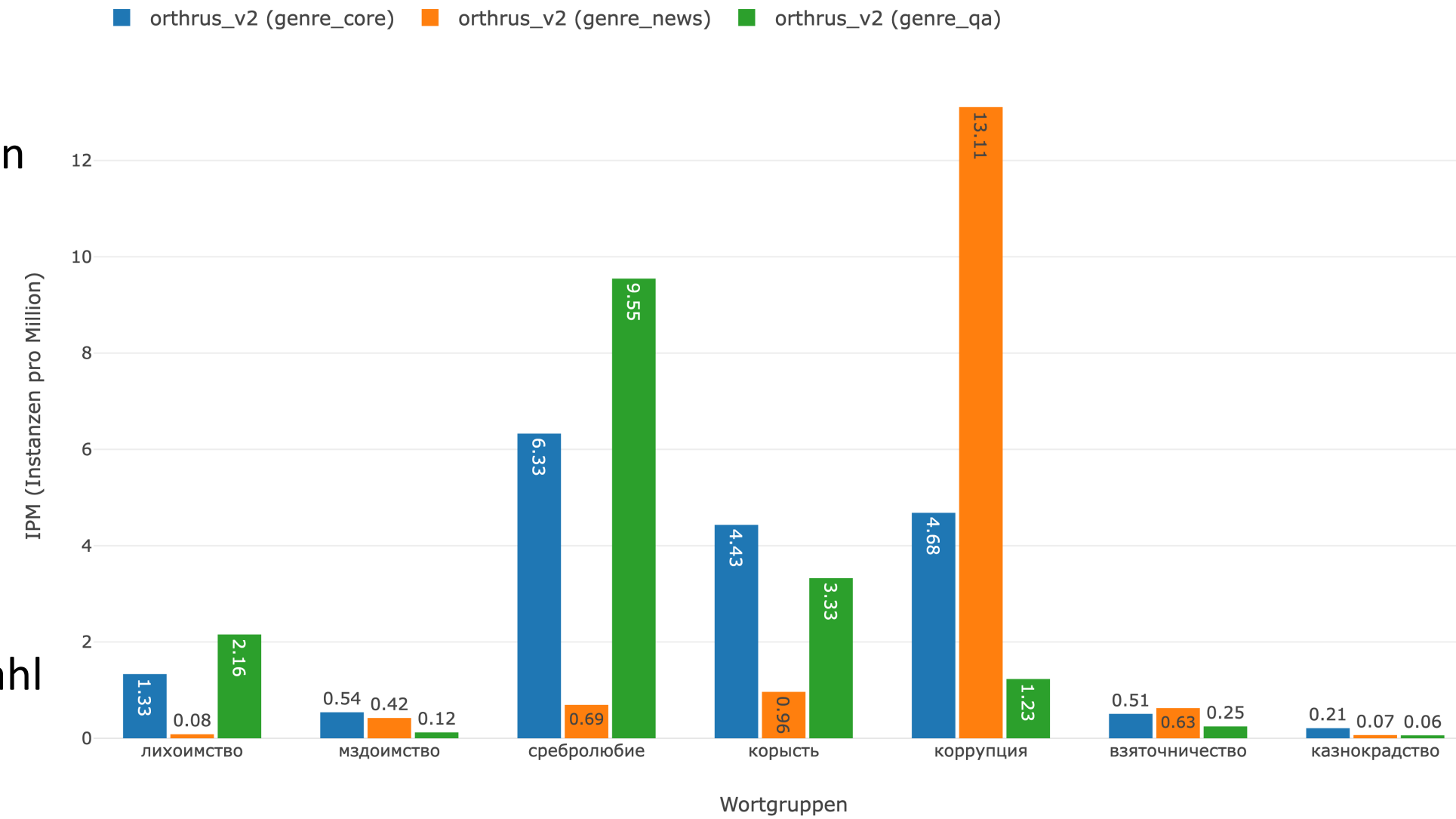


Textsorten innerhalb von OrthRus_v2

Traditionelle Lexik:
srebroljubie vor allem in Artikeln und Frage-Antwort-Texten

Moderne Lexik:
korrupcija besonders häufig in Nachrichten

Auch die Textsorte beeinflusst die Wortwahl



Universität Regensburg

SyntSearch: Syntaktische Suche

Funktionen:

- Suche nach syntaktischen Dependenzrelationen
- Analyse von abhängigen Wörtern und
-

Be
ko

ich			
ros			
sre			
vo			
p'j			
ob"jadenie 'Völlerei'	2	*nespravedlivost'* 'Ungerechtigkeit'	7
vzjatočničestvo 'Bestechlichkeit'	2	*vzjatočničestvo* 'Bestechlichkeit'	7

SyntSearch

Suche nach syntaktischen Strukturen in den Korpora von corpora.fisun.org

Korpus: orthrus_v2 | Deprel des Treffers bezüglich des Kopfes: any

Abfrage | Sortierung

Median 3

Beispiel 1: *лихоимство*

9. Составитель: протоиерей М.А. Поторжинский История русской церковной проповеди в биографиях и образцах пастырей-проповедников с IX–XIX вв. | Portal: **azbyka.ru**

Если столько зол телесных происходит от пьянства, то сколько зол душевных рождается от него: лихоимство, воровство, убийство, драка, ярость, несытое сладострастие, скверные и нелепые разговоры!

Gesuchtes Wort	лихоимство	parataxis	lemma=лихоимство	id=17	head=13	depth=1
Kopf	рождается parataxis рождаться depth=0					
Abhängige Wörter	воровство conj воровство	убийство conj убийство	драка conj драка	ярость conj ярость		
	сладострастие conj сладострастие	скверные conj скверный	разговоры conj разговор			
Anzahl abhängiger Wörter	gezeigt: 7	insgesamt: 7				

'Wenn schon so viele körperliche Übel aus der Trunksucht hervorgehen, wie viele seelische Übel entstehen dann erst aus ihr: Korruption, Diebstahl, Mord, Schlägerei, Wut, unersättliche Wollust, anstößige und unsinnige Gespräche!'

Beispiel 2: *коррупция*

13. Патриарх Кирилл: Приобщаясь к православной культуре мы впитываем нравственные идеи, автором которых является Сам Бог | Portal: **foma.ru**

- Мы хотим построить справедливое процветающее общество, из которого бы навсегда ушли коррупция, криминал, жуткие разборки между людьми, - заметил патриарх Кирилл.

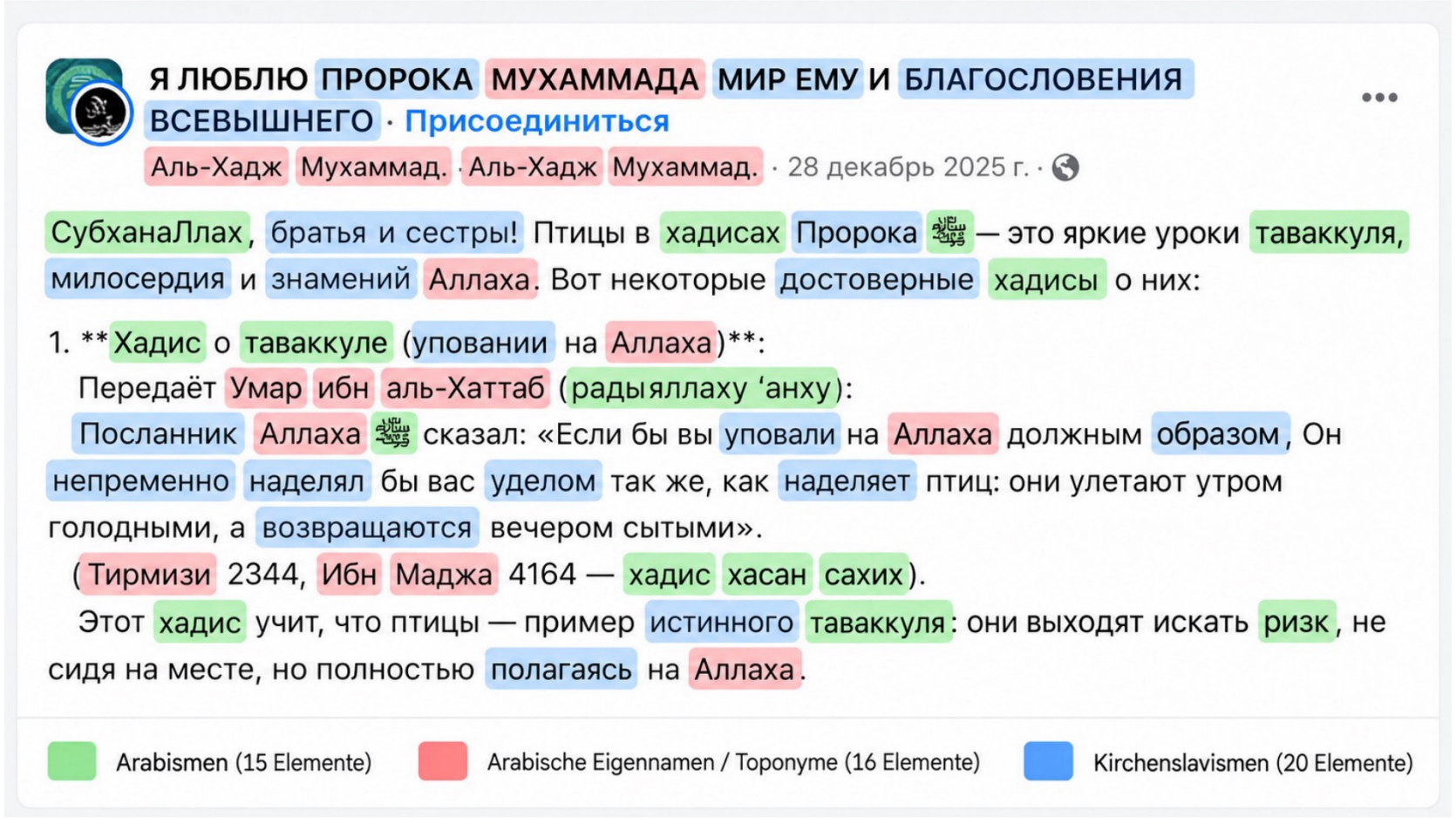
Gesuchtes Wort	коррупция	nsubj	lemma=коррупция	id=14	head=13	depth=4
Kopf	ушли nsubj уйти depth=3					
Abhängige Wörter	криминал conj криминал	разборки conj разборка				
Anzahl abhängiger Wörter	gezeigt: 2	insgesamt: 2				

'Wir wollen eine gerechte und prosperierende Gesellschaft aufbauen, aus der Korruption, Kriminalität und schreckliche Auseinandersetzungen zwischen den Menschen für immer verschwinden, bemerkte Patriarch Kirill.'

lichoimstvo erscheint häufig in Aufzählungen moralischer und religiöser Verfehlungen, *korrupcija* eher zusammen mit gesellschaftlichen Problemen und Kriminalität

Universität Regensburg

Fallstudie II: Kirchenslavismen in russischer Islamsprache



Arabismen (grün) · arabische Namen (rot) · Kirchenslavismen (blau)

Universität Regensburg

Material: genetische Kirchenslavismen

Genetische Zuordnung, unabhängig von heutiger stilistischer Markierung

(1) *glad* – *golod* ‘Hunger’
(2) *strana* – *storona* ‘Land’ – ‘Seite’
(3) *poslannik* ‘Bote’

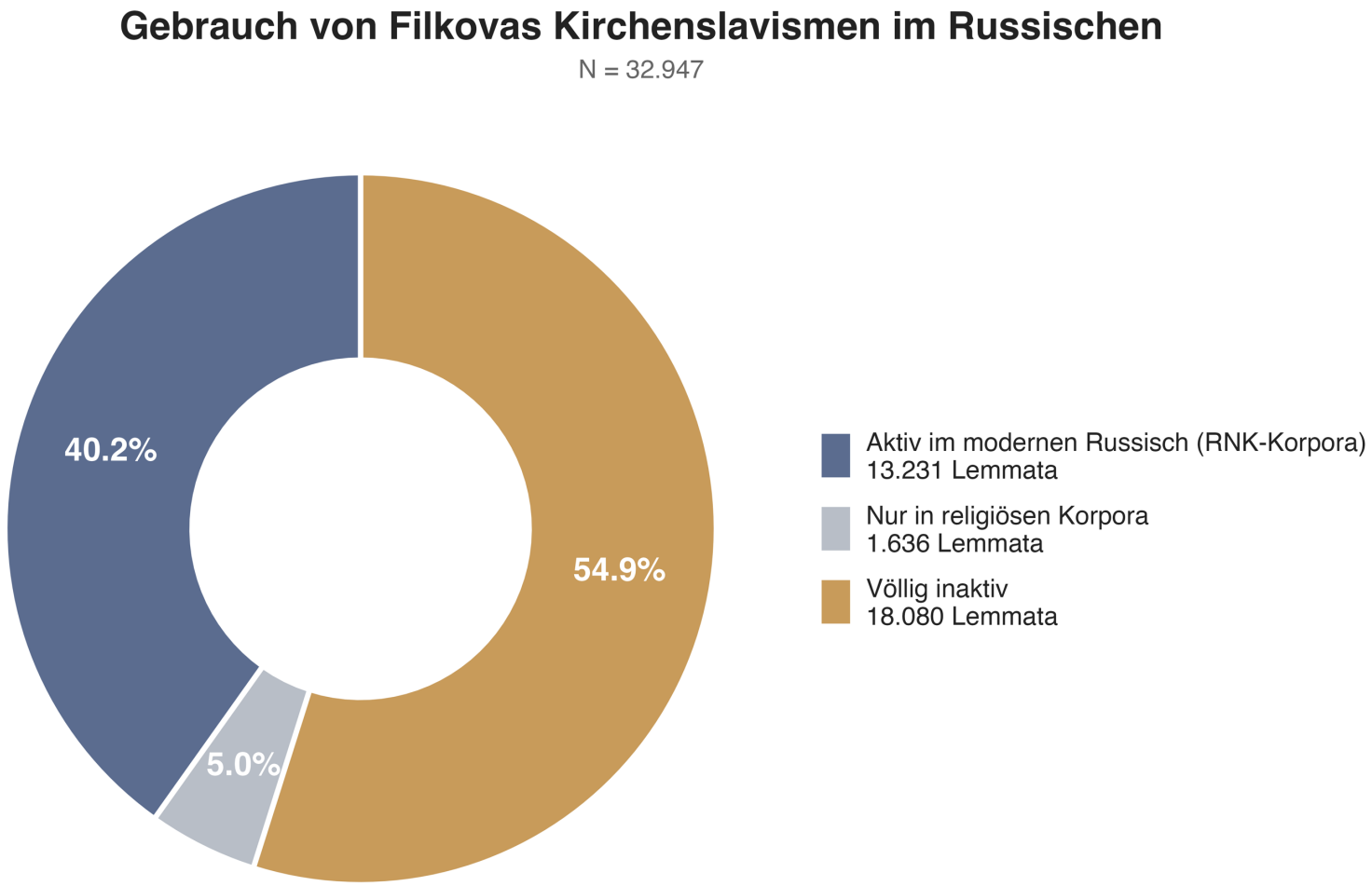
Quelle: Filkova (1986) *Starobolgarizmy i cerkovnoslavjanizmy v leksike russkogo literaturnogo jazyka*; ca. 33.000 Lemmata.

Abgleich mit vier Korpora: OrthRus_v2, IslamRus, RNK Paper und RNK Main

Universität Regensburg

Wie viele Kirchenslavismen sind im heutigen Russisch belegt?

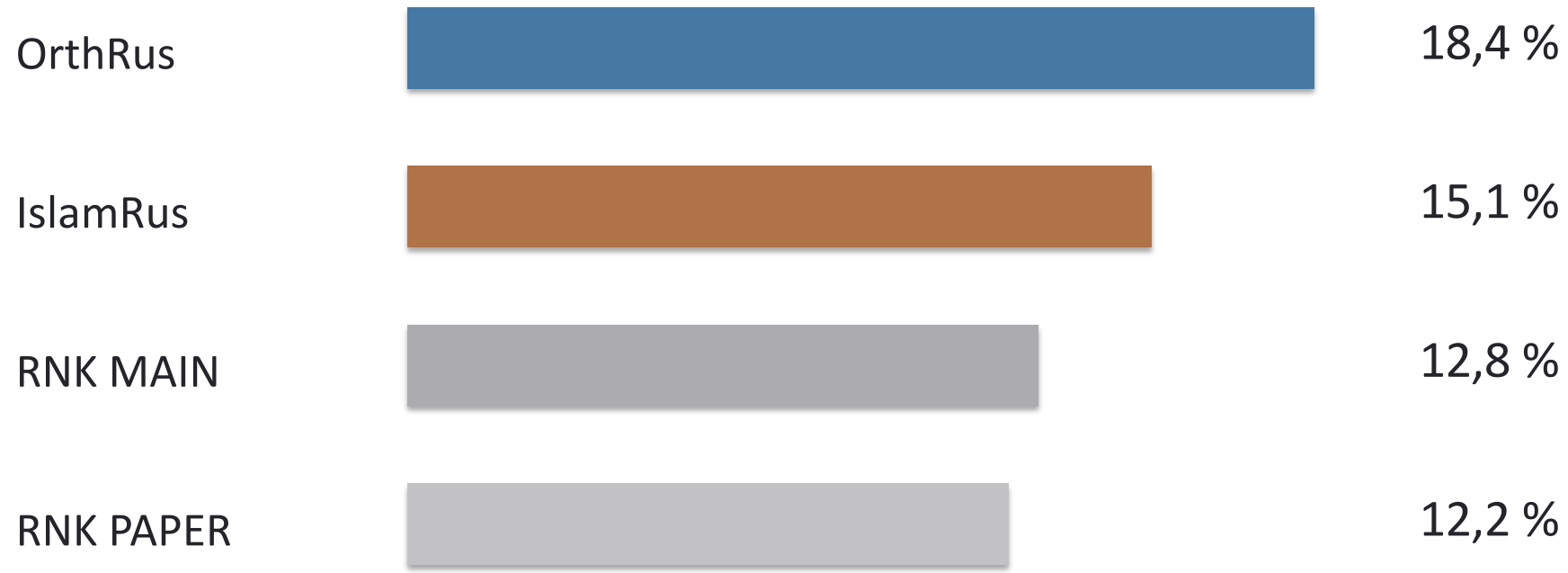
	Lemmata
Filkovas Wortliste	32.947
Aktiv im modernen Russischen	13.231
Ausschließlich religiös	1.636
Vollständig inaktiv	18.080



Universität Regensburg

Häufigkeit genetischer Kirchenslavismen

Anteil der erfassten lexikalischen Kirchenslavismen an allen Wörtern



IslamRus vs. RNK PAPER: rund 25 % höhere relative Häufigkeit.

Universität Regensburg

Gemeinsam religiös profilierte Lexik

651 **Lemmata in beiden Religionskorpora überrepräsentiert** — r > 2 in IslamRus und OrthRus

308 **davon zugleich in beiden Referenzkorpora unter Erwartung** — r < 0 in RNK MAIN und RNK PAPER

Lemma	IslamRus	OrthRus_v2	RNK MAIN	RNK PAPER
пророк ‘Prophet’	+1.276	+120	−157	−339
благословение ‘Segen’	+696	+316	−158	−332
молитва ‘Gebet’	+151	+806	−217	−534
благодать ‘Gnade’	+5	+434	−105	−275

Standardisierte Residuen r zeigen, wo eine Häufigkeit vom statistischen Erwartungswert abweicht.

Universität Regensburg

LemmaSketches: Kollokationen

- Syntaktische Kollokationsprofile auf Grundlage der UD-Annotation
- Häufigkeiten und Assoziationsmaße (logDice), direkter Korpusvergleich

konfessionelle Unterschiede zeigen sich auch im syntaktischen Gebrauch.

Beispiel: *повиноваться* ‘gehorchen’
IslamRus OrthRus_v2

Indirekte Objekte

Collocate	Freq	logDice
наставник	71/–	9.266/–
закон	110/–	8.904/–
христос	162/–	8.356/–
господ	39/–	8.348/–
голова	33/–	8.295/–
слово	77/–	8.233/–
голос	32/–	8.158/–
власть	61/–	8.02/–
муж	317/34	10.255/9.093
повеление	77/14	9.466/9.114
воля	114/12	8.662/8.716
веление	39/11	8.526/8.762
шайтан	–/3	–/7.123
аллаh	–/4	–/7.563
всевышний	–/24	–/7.685
родитель	–/18	–/8.173
посланника	–/22	–/8.65
правитель	–/14	–/9.051
аллах	–/153	–/9.41
приказ	–/17	–/9.657

Universität Regensburg

Fazit und Ausblick

- **Forschungsinfrastruktur:** OrthRus, IslamRus und die eigens entwickelten Analysewerkzeuge stehen für weitere Untersuchungen zur Verfügung
- **Forschungsperspektive:** Die Korpora ermöglichen vergleichende Untersuchungen religiöser und säkularer Diskurse auf lexikalischer und syntaktischer Ebene
- **Geplante Erweiterungen:** Aufbau eines Korpus russischsprachiger katholischer Diskurse und orthodoxer Diskurse in Polen; vergleichbare Korpora verschiedener Religionsgemeinschaften in der Ukraine

corpora.fisun.org

Universität Regensburg

Literatur

Buchenau, Klaus. 2023. *Eine russisch-islamische Synthese? Interaktionsmuster zwischen Christen, Muslimen und Staat im Wolgaraum vom Mittelalter bis in die Gegenwart*. IOS Mitteilungen 70. Regensburg: Leibniz-Institut für Ost- und Südosteuropaforschung.

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

Lyashevskaya, Olga, Kira Droganova, Daniel Zeman, Maria Alexeeva, Tatiana Gavrilova, Nina Mustafina & Elena Shakurova. 2016. *Universal Dependencies for Russian: A new syntactic dependencies tagset*. Higher School of Economics Research Paper WP BRP 44/LNG/2016.

Rychlý, Pavel. 2007. Manatee/Bonito – A modular corpus manager. In Petr Sojka & Aleš Horák (eds.), *1st Workshop on Recent Advances in Slavonic Natural Language Processing*, 65–70. Brno: Masaryk University.

Sibgatullina, Gulnaz R. 2020. *Languages of Islam and Christianity in post-Soviet Russia*. Leiden & Boston: Brill Rodopi.

Straka, Milan & Jana Straková. 2017. Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In Jan Hajič & Dan Zeman (eds.), *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 88–99. Vancouver: Association for Computational Linguistics.

Universität Regensburg

Danke für die Aufmerksamkeit!

Kontakt: roman@fisun.org

Folien sowie Links zu den Korpora:
https://fisun.org/files/conf/digitaleslawistik/

Universität Regensburg

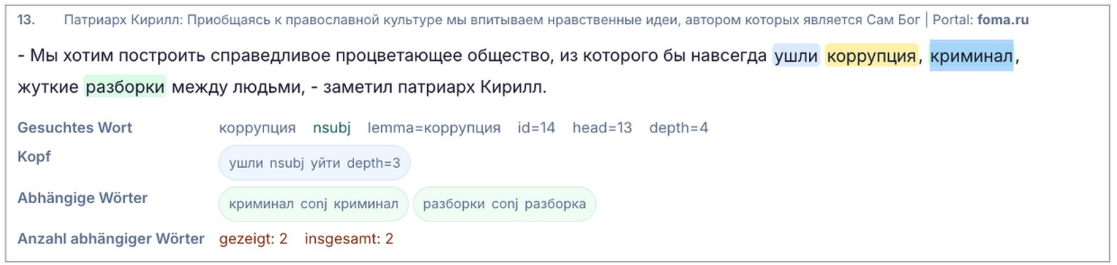
SyntSearch: Zwei Korpusbeispiele

Beispiel 1: *лихоимство*



‘Wenn schon so viele körperliche Übel aus der Trunksucht hervorgehen,
wie viele seelische Übel entstehen dann erst aus ihr: Korruption, Diebstahl,
Mord, Schlägerei, Wut, unersättliche Wollust, anstößige und unsinnige Gespräche!’

Beispiel 2: *коррупция*



‘Wir wollen eine gerechte und prosperierende Gesellschaft aufbauen,
aus der Korruption, Kriminalität und schreckliche Auseinandersetzungen
zwischen den Menschen für immer verschwinden, bemerkte Patriarch Kirill.’

Beispielansichten: *лихоимство* und *коррупция* – Koordinationen (conj)

Vom annotierten Text zum durchsuchbaren Korpus

.vert-Datei
Korpusinhalt

- Dokumente und Metadaten (<doc>)
- Sätze (<s>)
- Eine Zeile pro Token

```
id · word · lemma · tag · feat · head · deprel
```

→

Registry-Datei
Korpuskonfiguration

- Name und Pfade
- Tokenattribute
- Metadaten-Strukturen
- Einstellungen für Indexierung und Suche

→

Manatee / NoSketch Engine
Kompilierung und Indexierung

→ **durchsuchbarer Korpus**

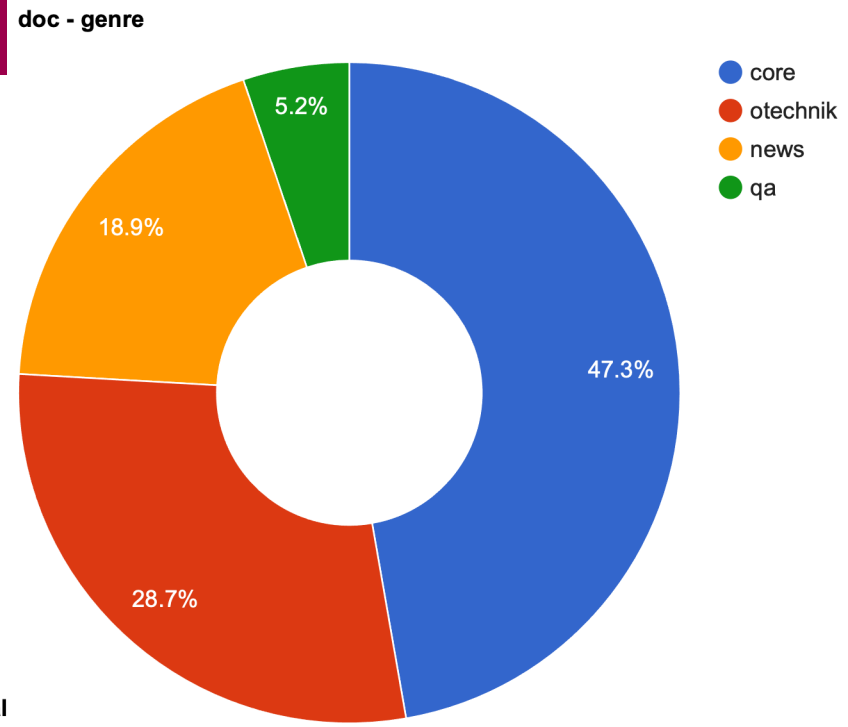
Syntaktische Attribute wie *head* und *deprel* bilden die Grundlage für Dependenzabfragen.

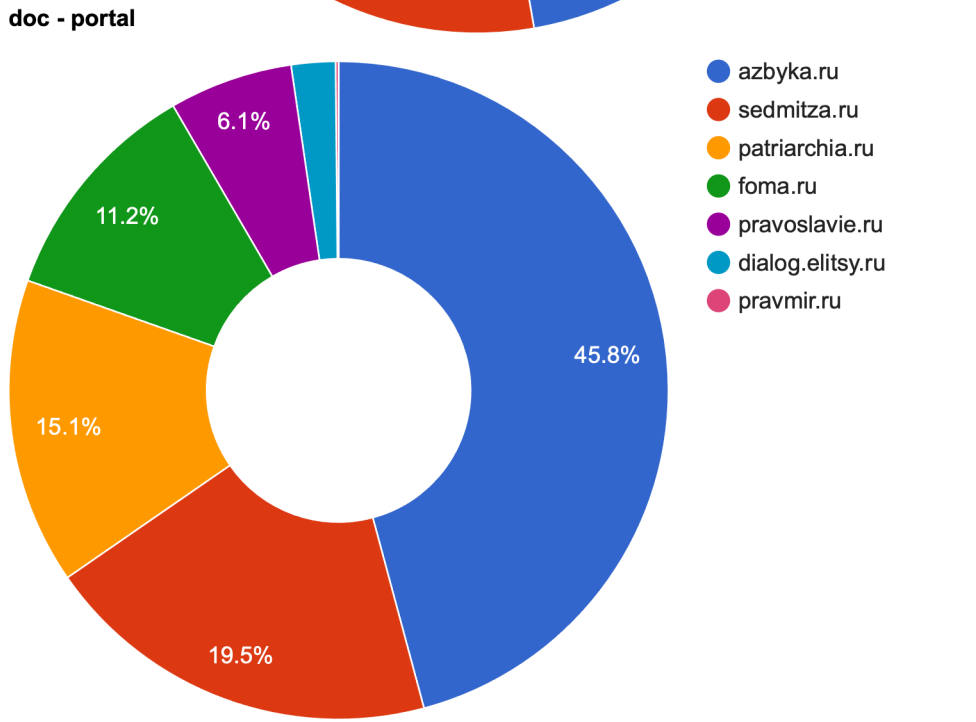
Universität Regensburg

OrthRus_v2: Korpusprofil

- Großes Korpus orthodoxer russischer religiöser Sprache; Daten erhoben von Dezember 2025 bis Mai 2026
- öffentliche Subkorpora nach Genre und Portal
- Besonders wichtig sind **azbyka.ru** als größtes mehrgenres Portal und **sedmitza.ru** / **patriarchia.ru** für Nachrichten bzw. institutionellen Diskurs

Audience reach
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





Universität Regensburg

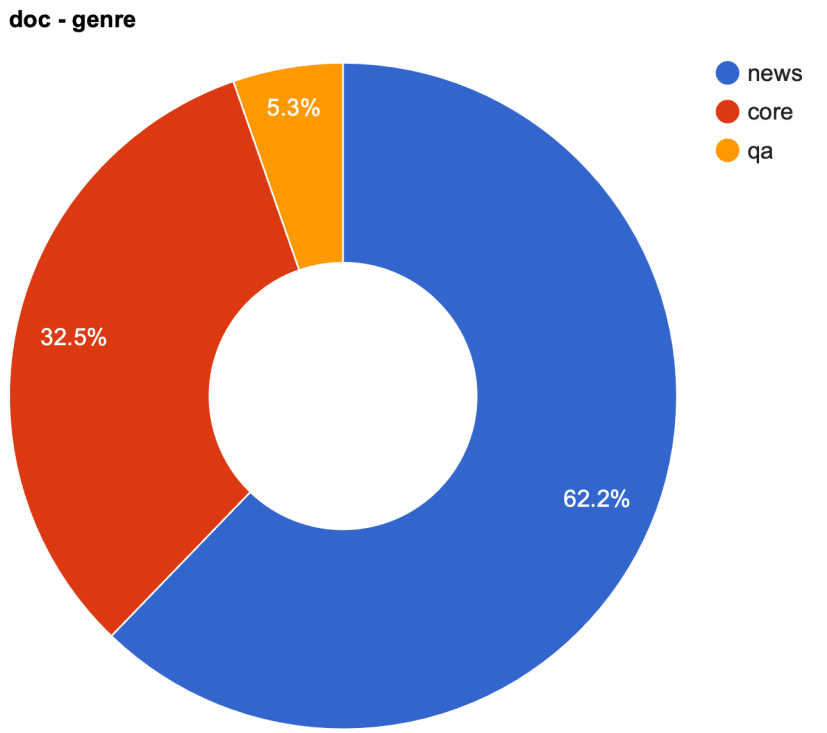
IslamRus: Korpusprofil

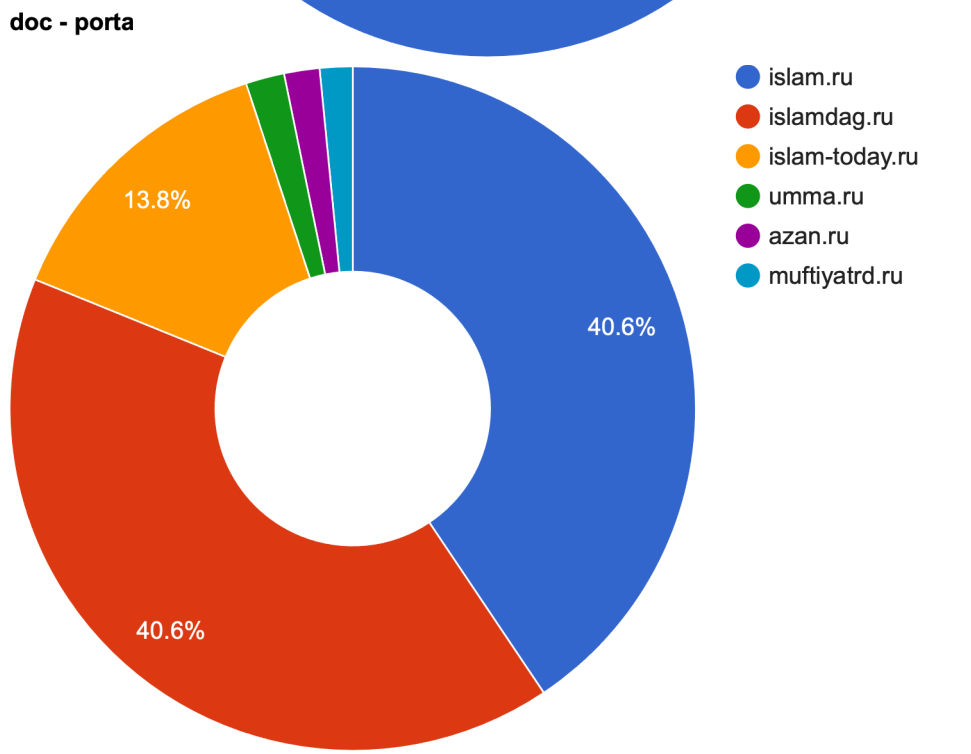
- Großes Korpus islamischer russischer religiöser Sprache; Datenerhebung: Dezember 2025 – Januar 2026
- Religiös-publizistische Texte, Nachrichten, Q&A / fatwa-artige Formate
- Besonders groß **islam.ru** und **islamdag.ru**; Dominant ist **genre_core** mit redaktionellen und didaktischen Texten

Audience context

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

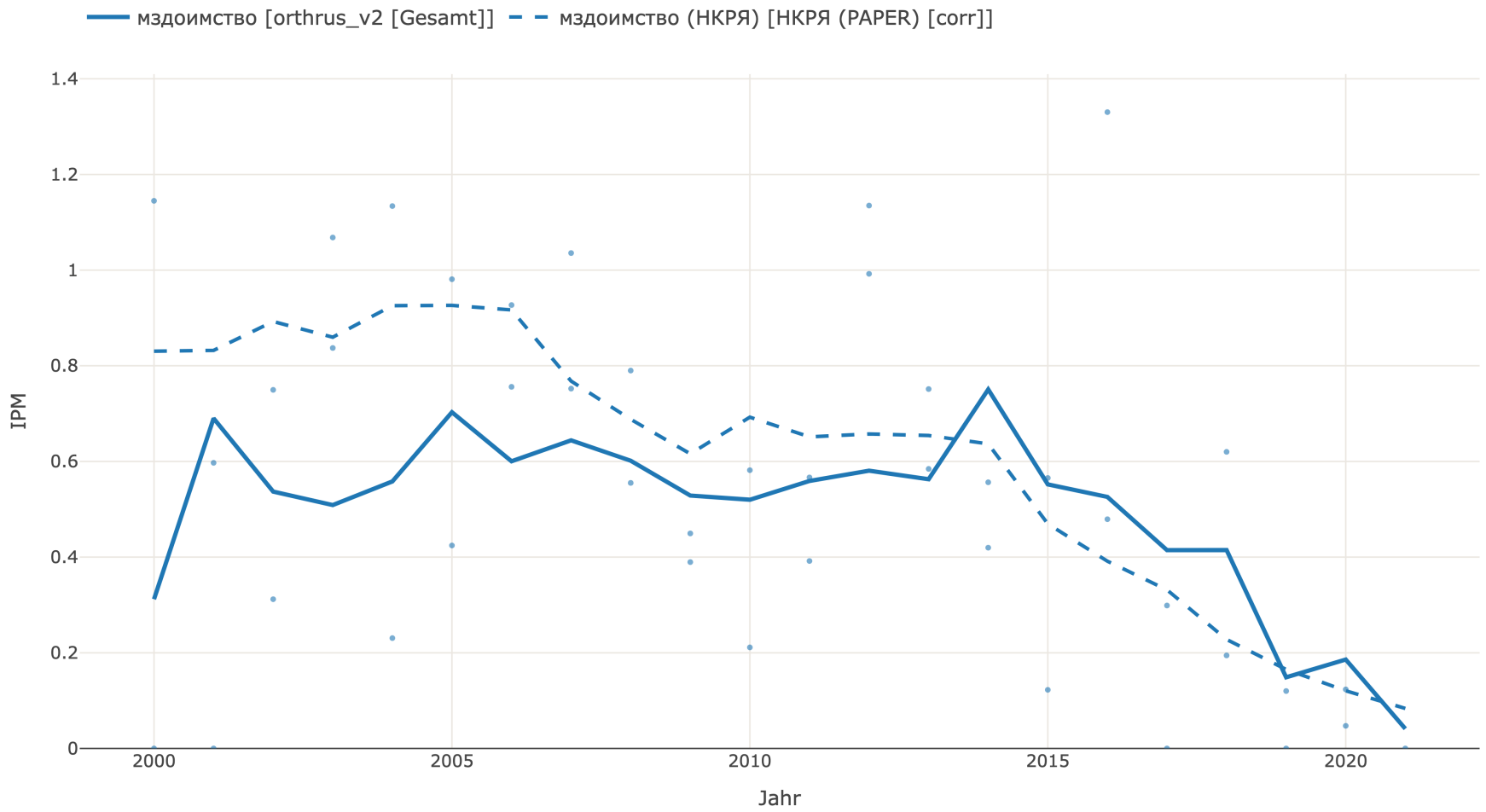
Zentrum	Portale
Dagestan / Nordkaukasus	islamdag.ru, muftiyatrd.ru
Tatarstan / Wolga-Ural	islam-today.ru, teilweise azan.ru
Überregional / nicht eindeutig regional (beinflusst vom Wolga-Islam)	islam.ru, umma.ru





Universität Regensburg

Graphs: Diachronischer Vergleich – мздоимство ‘Bestechlichkeit’



- Keine eindeutige Präferenz für den religiösen Diskurs.
- Seit etwa 2015 rückläufige Häufigkeit in beiden Korpora.

Universität Regensburg

Konfessionelle Unterschiede

Wie unterscheidet sich ihr Gebrauch zwischen den Religionen?

Lemma	IslamRus	OrthRus v2
сподвижник ‘Gefährte’	+517,3	−29,0
православный ‘orthodox’	−116,1	+1.324,2
храм ‘Tempel’	−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)
‘Der Beste von allen, der Prophet Muhammad (Friede und Segen seien mit ihm), ging dahin, ebenso seine **Gefährten**. Doch der Islam blieb und verbreitete sich weltweit.’

(5) *Прожив в мечети около месяца, мужчина по собственному желанию покинул* ***храм***. (IslamRus)
‘Nachdem der Mann etwa einen Monat in der Moschee gelebt hatte, verließ er **sie** aus eigenem Entschluss.’

(6) *Тимофей был одним из семидесяти апостолов, учеником и ближайшим* ***сподвижником*** *апостола Павла*. (OrthRus v2)
‘Timotheus war einer der siebzig Apostel, Schüler und engster **Gefährte** des Apostels Paulus.’

(7) *На следующий день мы пришли в* ***храм*** *уже к концу праздничного богослужения*. (OrthRus v2)
‘Am nächsten Tag kamen wir erst gegen Ende des Festgottesdienstes in die **Kirche**.’

Auch Lexik kirchenslavischer Herkunft kann ein islamisches Profil aufweisen. Werte: Residuen.