

# A keymorph analysis of Russian political news reporting

**Thomas Samuelsson**

Stockholm University

Stockholm, Sweden

thomas.samuelsson@slav.su.se

## Abstract

This paper presents a diachronic study of Russian prefixes in the political news reporting. The analysis examines derivational prefixes in the Russian media discourse for each year in the time period 2012–2020. The prefixes are analyzed using derivational keymorphs. The data material consists of a corpus of political texts from more than 60 Russian online media resources. Key-morphs have previously been used to investigate Czech presidential discourse (Fidler and Cvrček, 2019), the Russian media resource Sputnik Czech Republic (Fidler and Cvrček, 2018; Cvrček and Fidler, 2019) and Putin’s speeches (Janda et al., 2023). The use of keymorphs enables one to focus on morphological features and to capture general characteristics of the textual content in a language corpus. The work uses Corpus-assisted discourse studies (CADS) as the main framework and is a contribution to the understanding of Russian political discourse.

## 1 Introduction

This paper studies Russian morphological derivation in political discourse. More precisely, it studies prefixes in the Russian political news reporting from a chronological perspective in time period 2012–2020 by applying the concept of keymorphs. This concept has been introduced by Fidler and Cvrček (2019) as an extension of keyword techniques. Keywords are a useful part of corpus-based discourse studies (Partington and Duguid, 2020) and are often used in the study of political discourse (Ädel, 2010). While an investigation of keywords highlights major key topics and stylistic features in the discourse, keymorphs identify to a greater extent general characteristics of the discourse (Fidler and Cvrček, 2019). The extension of the concept of keyness beyond keywords to other key items have previously also been introduced for other linguistic units than morphemes. One major approach that uses key part of speech tags (pos-tags) and semantic categories in the analysis of linguistic characteristics has been developed by Rayson (2004, 2008). The method of identifying key pos-tags and key semantic categories has been exploited in a variety of studies. Culpeper (2009) utilized key part of speech and semantic field analysis to analyze the characters in Shakespeare’s *Romeo and Juliet*. Archer et al. (2009) explored Shakespeare’s plays in terms of key semantic fields and Afida (2007) analyzed business magazines using the same technique. Some recent studies on key parts of speech have been conducted by Breeze (2019) on legal genres and Smith and Waters (2019) on a British radio show.

The studies have not focused on morphosyntactic features since the opportunity to do that in English is limited due to a quite high degree of analyticity (Cvrček and Fidler, 2019). Czech and Russian are on the other hand typologically more synthetic and have inventories of identifiable inflectional markers that are richer. Fidler and Cvrček (2019) show that an investigation of inflectional keymorphs (case, number, person, finiteness, verb negation) and part of speech keymorphs revealed representations of situations and speaker images in Czech presidential speeches. Ideological differences between the presidents were associated with parts of speech and stylistic variations with inflectional features. They have also explored the Russian media portal Sputnik Czech Republic and showed that inflectional keymorphs are a tool that provide information about the structure of the discourse (Fidler and Cvrček, 2018; Cvrček and Fidler, 2019). As an illustration, the analyses show that Russia is likely to be portrayed as a victim and Putin to

be represented as an actor with agency in the Russian-controlled media outlet. Janda et al. (2023) tested for the first time keymorph analysis on Russian data in a study of Putin’s speeches delivered around the time of Russia’s full-scale invasion of Ukraine. The study revealed the roles of the actors Russia, NATO and Ukraine in the narrative by exploring key grammatical cases. To mention some roles, the different actors could be portrayed as agents, victims, dynamic, static, places and states.

This study, however, focuses on the derivational morphological system in Russian political news reporting within the framework of Corpus-assisted discourse studies (CADS) (Partington et al., 2013). In particular, it is built on the perspective that discourse studies target semantic issues and consequently ascribe meanings to all linguistic elements in the contexts, including the smallest meaningful constituents in the language (Spitzmüller and Warnke, 2011). The study is also motivated from the approach of Cognitive Grammar, which “treats all linguistic units and categories as meaning-bearing, in all contexts” (Divjak and Janda, 2008). From empiricism, researchers have noticed that the Russian morphological derivational system reacts to extra-linguistic factors such as socio-political processes (Ratsiburskaya et al., 2015). Based on the Russian socio-political development during the investigated time period, diachronic variations are to be expected. Characteristic of the contemporary Russian language is the rise of the derivational nominal prefixes whose use extensively are of non-Slavic origin (Zemskaya, 2006). The prefixes originate alongside English from Greek and Latin (Koriakowcewa, 2009).

There have been advances in the creation of Russian derivational resources in the recent years that enables one to investigate the Russian discourses by exploiting derivational morphemes.<sup>1</sup> This paper examines co-occurrences of lexical constituents and discursive functions in the Russian political media discourse that belong to other aspects of the discourse that are not otherwise investigated such as key topics or typical linguistic items identified through the use of keywords. Since a keymorph analysis uses the same principles as other keyness statistical approaches, it also has the same advantages compared to a qualitative analysis. A keyword analysis can reveal features that are not obvious to an unaided eye and are hard for an observer to detect (Culpeper, 2009). Another benefit according to the author is that it also uncovers patterns without the use of intuition of an observer. The aim of this preliminary study is to probe the Russian political media reporting by identifying changes in the salience of derivational prefixes. The keyness of the prefixes in the study is estimated by using the derivational keymorph technique. Against this background, the study will present preliminary results based on the use of the derivational morphology resource DeriNet.RU 0.5 for Russian (Kyjánek et al., 2022).

## 2 Methodology

The diachronic study approaches Russian political news reporting through the evolution of keyness of derivational prefixes. The prominent prefixes for each year are identified by comparing the relative frequencies of the prefixes in a target corpus composed of texts published in the same single year with the relative frequencies of the same prefixes in a reference corpus in the time frame 2012–2020. The target corpora are created by partitioning the language data according to year. The corpora of interest represent Russian political online reporting for each year in the studied time period. The comparisons are made with the entire undivided reference corpus that represents Russian language presented on the web. Since Russian has a comparatively high degree of inflection and the target is semantic constituents, lemmas from both corpora are used. Every lemma in the corpora is automatically looked-up in the morphological database DeriNet.RU 0.5 to get derivational information (Kyjánek et al., 2022). If the lexicon does not contain information about the subparts of a prefixed lexeme for some reason, it is not integrated into the results.

The keymorph analysis is conducted in the same way as a keyword analysis. In both approaches, a target corpus is contrasted to a reference corpus. The prominence of the morphemes is calculated by comparing the frequencies using two types of calculations (Fidler and Cvrček, 2015, 2019). The first calculation makes sure that there are enough data evidence. For this purpose, the study uses the statistical test log-likelihood ratio. The null hypothesis is that there is no difference in the frequencies of a prefix. The differences in the uses of each of the most prominent prefixes are statistically significant at a level

---

<sup>1</sup><https://ufal.mff.cuni.cz/universal-derivations>

$p < 0.05$ . The second calculation estimates the effect size through the metric Log Ratio introduced by [Hardie \(2014\)](#). The effect size estimator is calculated for each prefix according to equation 1, where  $f_{target}$  denotes the relative frequency of a prefix in a target corpus,  $f_{reference}$  designates the relative frequency of the same prefix in the reference corpus and  $\log_2$  is the binary logarithm:

$$\text{Log Ratio} = \log_2 \frac{f_{target}}{f_{reference}}. \quad (1)$$

Previously, the effect size of keymorphs has been measured using Difference Index (DIN) ([Fidler and Cvrček, 2019](#)). The DIN-index is essentially the difference between the relative frequency of an item in the target corpus and the relative frequency of the same item in the reference corpus divided by the mean of the relative frequencies; and then the index is normalized so that the range is  $\pm 100$ . Difference Index was introduced to handle words that are present in the target corpus but absent in the reference corpus ([Fidler and Cvrček, 2015](#)). However, the choice of Log Ratio appears more intuitive and less complicated to interpret. On the one hand the measures differ when it comes to magnitude, range and treatment of the absence of lexical items. On the other hand they both generate the same rank order ([Gabrielatos, 2018](#)), which is an important part of the analysis ([Fidler and Cvrček, 2015](#)). While Difference Index has a range between  $-100$  and  $100$ , the range of Log Ratio is unbounded for both positive and negative values. Difference Index handles absence in corpora while Log Ratio on the other hand does not allow frequencies to be equal to zero. But in large corpora, the absence of prefixes is unlikely to be an issue. The interpretation of the value of the Log Ratio is as follows. A Log Ratio value of zero shows that the relative frequencies of the prefixes are equal in both corpora. A Log Ratio value of 1 means that the relative frequency of an item in the corpus of interest is twice the relative frequency of the same item in the reference corpus. A Log Ratio value of 2 corresponds to a relative frequency that is 4 times larger in the target corpus compared to the reference corpus. If we continue the with the values 3, 4 and 5, the relative frequency will be 8, 16 and 32 times larger ([Hardie, 2014](#)).

### 3 Data

To be able to calculate the Log Ratio values for the Russian prefixes, both well annotated corpora and a high-quality derivational analysis tool are needed. Since Russian, as mentioned above, has rich morphology, compared to for example English, corpora annotated with lemmas are therefore required. This section presents the target corpus and the reference corpus used in the study, as well as the database containing derivational information about the Russian lemmas.

#### 3.1 Corpora

The study requires, as previously mentioned, target corpora and a reference corpus to calculate the Log Ratio values of the prefixes. The target corpora consist of a collection of Russian political texts published online between 2012 and 2020. The data material is sampled from the most influential Russian online media resources according to a citation index provided by the leading Russian media monitoring company [Medialogia \(2023\)](#)<sup>2</sup> and the selection of texts is made on the basis of political classification. In the linguistic processing, deduplication of the texts has been applied. The raw texts are tokenized and lemmatized using the natural language analysis tool Stanza ([Qi et al., 2020](#)) trained on the SynTagRus treebank ([Dyachenko et al., 2015](#); [Droganova et al., 2018](#)). The performance for Russian measured in F1 scores are for tokens 99.57 and for lemmatization 97.51.<sup>3</sup> The size is more than 500 million tokens from more than 60 outlets. The extracted texts originate from a diversity of journalistic resources of different genres, geographical cover and political orientations. The dataset contains Russian political media texts annotated with publication date.

The acquisition of relevant keyed items is connected to the relation between the target corpus and the reference corpus ([Culpeper and Demmen, 2015](#)). A relation that is close between the corpora increases the likelihood to obtain keywords that are specific to the target corpus ([Culpeper, 2009](#)). Since the interest

<sup>2</sup><https://www.mlg.ru/>

<sup>3</sup><https://stanfordnlp.github.io/stanza/v100performance.html>

here is to uncover keymorphs that are characteristic to Russian political journalism, a general web corpus is preferred to a balanced general corpus. The target corpora are contrasted with a reference corpus that is much larger in size than any of the target corpora and is assessed to have achieved representativeness of the Russian web language use during the given time period. In this way, the political aspects of the textual content will be highlighted. Used as a reference corpus in the study is the internet corpus *Ara-neum Russicum III Maximum 19.03* (Benko, 2014a,b; Benko and Zakharov, 2016, 2021; Rychlý, 2007). It is based on web-crawled Russian language data that have been acquired by applying the strategy of including everything that is possible to come across. The size of the reference corpus is almost 20 billion tokens from the actual time frame. The reference corpus is tokenized and lemmatized using TreeTagger (Benko, 2014a; Benko and Zakharov, 2021). In an evaluation of the performance in lemmatization tasks, TreeTagger trained on the disambiguated subcorpus of the Russian National Corpus (RNC) performed an accuracy of 97.0% on RNC<sup>4</sup> and 86.9% on RU-EVAL gold standard (Kuzmenko, 2016). Kotelnikov et al. (2017) tested the parser on three corpora. The accuracy is 95.21% on the RU-EVAL corpus, 97.31% on the disambiguated subcorpus of the Russian National Corpus (RNC) and 96.95% on the disambiguated subcorpus of OpenCorpora. The results were biased on the RNC since TreeTagger was trained on it. Compared to the Russian national corpus, the selected reference corpus contains more rare lexical items but may on the other hand be less balanced (Benko and Zakharov, 2016).

### 3.2 Database

The frequencies of the prefixes in each of the corpora are calculated using the derivational lexical resource DeriNet.RU 0.5 (Kyjánek et al., 2022). DeriNet.RU is an open license state-of-the-art derivational model that captures derivational processes for the Russian language and includes more than 300 thousand lexemes and 164 thousand binary derivational relations, including derivational prefixation. The database outperforms other Russian derivational resources. It is the resource that contains the most number of lexemes and derivational relations. Besides that, all the lexemes are corpus-attested. The database includes derivational relations within and between nouns, adjectives, verbs and adverbs, but no compounds. Of the lexemes in the database, a majority (58%) consists of nouns; verbs and adjectives have about the same share, 20% respective 19%. The most common derivational relations involve nouns. The far most common derivational relation is the one where both the base and the derivate are nouns (42%). The maximum oracle score for a set of derivational relations was calculated to be 87.3%.

## 4 Results

The morpheme-discursive probe of the Russian political media reporting for each year of the time period 2012–2020 is presented in table 1. Among the most prominent prefixes, one can observe some key prefixes that are consistently high-ranked as well as prefixes that show large diachronic changes. Three prefixes, *éks-* ‘ex-’, *vnutri-* ‘intra-’ and *vice-* ‘vice-’, have a Log Ratio value of more than 2 for every year. They can be referred to as the most prominent prefixes in the discourse for the whole time period. The international prefix *giper-* ‘hyper-’ shows the largest increase (Spirkin et al., 1982, 129; Ryazanova-Clarke and Wade, 1999, 197–198). Between 2017 and 2018, Log Ratio increased from 2.08 to 4.82 and the prefix took the top place for the years 2018 and 2019. Its native counterpart *sverch-* ‘over-, super-’ is also activated in the second half of the time period (Ryazanova-Clarke and Wade, 1999, 194–196). The prefix *ul’tra-* ‘ultra-’ displays a growing trend during the time period and lands on a Log Ratio value of 2.77 in 2020. The prefix *trans-* ‘trans-’ is another prefix that has gained an increase during the time period and peaks in 2016 with a Log Ratio of 3.33. The prefixes *anti-* ‘anti-’, *kontr-* ‘counter-’ and *mež-* ‘inter-’ decline throughout the investigated time frame.<sup>5</sup>

The study confirms several findings in the literature. The nominal prefixes are more important in the Russian political reporting than the verbal prefixes and the international prefixes are more prominent than the prefixes of Slavic origin, for instance the international derivational prefixes such as *anti-* ‘anti-’ (Spirkin et al., 1982, 41), *éks-* ‘ex-’ (Spirkin et al., 1982, 574), *giper-* ‘hyper-’ (Spirkin et al., 1982, 129),

<sup>4</sup><https://ruscorpora.ru/>

<sup>5</sup>Transliteration according to Scando-Slavica is used (<https://www.tandfonline.com/journals/ssl20>).

*kontr-* ‘counter-’ (Spirkin et al., 1982, 249), *sub-* ‘sub-’ (Spirkin et al., 1982, 477), *trans-* ‘trans-’ (Spirkin et al., 1982, 502), *ul’tra-* ‘ultra-’ (Spirkin et al., 1982, 513), *vice-* ‘vice-’ (Spirkin et al., 1982, 104) are more dominant than the inherited derivational prefixes like *mež-* ‘inter-’ (Vasmer, 1955, 112; Gerd, 2008, 10; Kuznecov, 1998, 529) and *vne-* ‘extra-’ (Vasmer, 1953, 210; Gorbačevič, 2005, 641; Kuznecov, 1998, 137) and those in turn are more salient than prefixes typically associated with verbs like *po-* ‘a little’, *s-* ‘together, down’, *pro-* ‘through’ (Janda and Lyashevskaya, 2013), *vy-* ‘out of’ (Ožegov and Švedova, 1999, 108–109), *voz-* ‘up’ (Ožegov and Švedova, 1999, 108–109) and so on. An exception from this pattern is the salient nominal prefix *vnutri-* ‘intra-’ of Slavic origin (Ožegov and Švedova, 1999, 88; Vasmer, 1953, 211).

Year	Prefix	LR	Year	Prefix	LR	Year	Prefix	LR
2012	<i>eks-</i> ‘ex-’	5.23	2013	<i>eks-</i> ‘ex-’	5.07	2014	<i>eks-</i> ‘ex-’	4.89
	<i>vnutri-</i> ‘intra-’	4.07		<i>vnutri-</i> ‘intra-’	3.76		<i>vnutri-</i> ‘intra-’	4.14
	<i>vice-</i> ‘vice-’	3.05		<i>vice-</i> ‘vice-’	2.64		<i>vice-</i> ‘vice-’	3.26
	<i>mež-</i> ‘inter-’	2.22		<i>mež-</i> ‘inter-’	2.30		<i>sub-</i> ‘sub-’	2.64
	<i>anti-</i> ‘anti-’	1.68		<i>sub-</i> ‘sub-’	1.96		<i>anti-</i> ‘anti-’	2.05
	<i>sub-</i> ‘sub-’	1.52		<i>vne-</i> ‘extra-’	1.87		<i>trans-</i> ‘trans-’	2.03
	<i>kontr-</i> ‘counter-’	1.44		<i>anti-</i> ‘anti-’	1.71		<i>vne-</i> ‘extra-’	2.02
	<i>protivo-</i> ‘counter-’	1.43		<i>nedo-</i> ‘under-’	1.69		<i>protivo-</i> ‘counter-’	1.94
	<i>nedo-</i> ‘under-’	1.38		<i>trans-</i> ‘trans-’	1.67		<i>mež-</i> ‘inter-’	1.70
	<i>vne-</i> ‘extra-’	1.35		<i>kontr-</i> ‘counter-’	1.62		<i>kontr-</i> ‘counter-’	1.67
2015	<i>eks-</i> ‘ex-’	4.58	2016	<i>eks-</i> ‘ex-’	5.17	2017	<i>eks-</i> ‘ex-’	5.41
	<i>vnutri-</i> ‘intra-’	4.23		<i>vnutri-</i> ‘intra-’	4.43		<i>vnutri-</i> ‘intra-’	4.69
	<i>vice-</i> ‘vice-’	3.10		<i>vice-</i> ‘vice-’	4.00		<i>vice-</i> ‘vice-’	3.42
	<i>trans-</i> ‘trans-’	2.46		<i>trans-</i> ‘trans-’	3.33		<i>trans-</i> ‘trans-’	2.80
	<i>sub-</i> ‘sub-’	2.40		<i>sub-</i> ‘sub-’	2.50		<i>sub-</i> ‘sub-’	2.67
	<i>anti-</i> ‘anti-’	1.98		<i>anti-</i> ‘anti-’	2.02		<i>giper-</i> ‘hyper-’	2.08
	<i>obez-</i> ‘dis-’	1.86		<i>giper-</i> ‘hyper-’	1.88		<i>anti-</i> ‘anti-’	2.04
	<i>mež-</i> ‘inter-’	1.80		<i>sverch-</i> ‘over-’	1.77		<i>ul’tra-</i> ‘ultra-’	2.01
	<i>vne-</i> ‘extra-’	1.75		<i>ul’tra-</i> ‘ultra-’	1.74		<i>mež-</i> ‘inter-’	1.84
	<i>kontr-</i> ‘counter-’	1.67		<i>mež-</i> ‘inter-’	1.59		<i>sverch-</i> ‘over-’	1.68
2018	<i>giper-</i> ‘hyper-’	4.82	2019	<i>giper-</i> ‘hyper-’	5.12	2020	<i>eks-</i> ‘ex-’	5.02
	<i>eks-</i> ‘ex-’	4.77		<i>eks-</i> ‘ex-’	4.71		<i>giper-</i> ‘hyper-’	4.83
	<i>vnutri-</i> ‘intra-’	4.39		<i>vice-</i> ‘vice-’	4.23		<i>vnutri-</i> ‘intra-’	3.98
	<i>vice-</i> ‘vice-’	3.54		<i>vnutri-</i> ‘intra-’	4.22		<i>ul’tra-</i> ‘ultra-’	2.77
	<i>trans-</i> ‘trans-’	2.93		<i>sub-</i> ‘sub-’	2.51		<i>trans-</i> ‘trans-’	2.49
	<i>ul’tra-</i> ‘ultra-’	2.41		<i>ul’tra-</i> ‘ultra-’	2.38		<i>vice-</i> ‘vice-’	2.42
	<i>sub-</i> ‘sub-’	2.34		<i>trans-</i> ‘trans-’	2.14		<i>sub-</i> ‘sub-’	1.90
	<i>sverch-</i> ‘over-’	1.91		<i>mež-</i> ‘inter-’	1.76		<i>sverch-</i> ‘over-’	1.73
	<i>mež-</i> ‘inter-’	1.53		<i>sverch-</i> ‘over-’	1.63		<i>protivo-</i> ‘counter-’	1.65
	<i>protivo-</i> ‘counter-’	1.50		<i>vne-</i> ‘extra-’	1.58		<i>vne-</i> ‘extra-’	1.40

Table 1: The top-10 Log Ratio (LR) values of the prefixes in the corpus of Russian political news for each year in the time period 2012–2020

The Log Ratio values of the derivational prefixes point to a number of discourse properties in the Russian political news reporting. The prefixes *eks-* ‘ex-’, *vice-* ‘vice-’ and *vnutri-* ‘intra-’ belong to the most prominent prefixes in Russian political reporting. The most distinguishing prefix in the reporting is *eks-* ‘ex-’ with the meaning of former, for example:

- (1) *S drugoj storony, Belyj dom ne spešit so vtorym paketom sankcij protiv Moskvy v svjazi s otravle-*

niem **éks-polkovnika** GRU Sergeja Skripalja i ego dočeri Julii v britanskem Solsberi. ([https://www.gazeta.ru/politics/2019/05/16\\_a\\_12357661.shtml](https://www.gazeta.ru/politics/2019/05/16_a_12357661.shtml))

‘On the other side, the White House is in no hurry with a second package of sanctions against Moscow due to the poisoning of the **ex-colonel** of GRU Sergei Skripal and his daughter Yulia in the British Salisbury.’

Also salient in the political media discourse is the prefix *vice-* ‘vice-’ with a most likely focus on political deputies:

- (2) *Otmetim, čto ranee **vice-prem'er** Rossii Dmitrij Rogozin zajavil o tom, čto serijnoe proizvodstvo novejšego rossijskogo tanka T-14 "Armata" možet načat'sja v 2019 godu.* (<https://rg.ru/2017/06/21/rossiia-ne-budet-postavliat-za-rubezh-tank-armata-i-sistemu-s-500.html>)

‘Let us note that the **Deputy Prime Minister** of Russia Dmitry Rogozin earlier said that the serial production of the newest Russian tank T-14 “Armata” could begin in 2019.’

The prominence of the prefix *vnutri-* ‘intra-’ suggests a focus on internal relations, for example when they correlate with Russia’s ambitions to weaken the Western countries from within:

- (3) *Pri étom Putin v ékskluzivnom interv'ju avstrijskomu telekanalu ORF zajavil, čto ego vstreča s Trampom do sich por ne sostojalas' iz-za ožestočennoj **vnutripolitičeskoj** bor'by v SŠA.* (<https://www.vesti.ru/doc.html?id=3024574>)

‘Putin said in an exclusive interview with the Austrian TV channel ORF that his meeting with Trump had not yet taken place due to the fierce **intra-political** struggle in the USA.’

The prefix *giper-* ‘hyper-’ shows the strongest increase of all the prefixes, especially from 2017 to 2018. In parallel, the prefix *sverch-* ‘over-, super-’ with a similar meaning also increased its activity. They are associated with the militarization of Russia reflected in the political discourse, especially the development and introduction of powerful weapons in the Russian military arsenal:

- (4) *Putin upomjanul o rakete s jadernoj énergoustanovkoj, okeanskoj sisteme s bespilotnymi podlodkami na jadernoj ustanovke i **giperzvukovyh** raketach "Kinžal".* (<https://www.newsru.com/russia/23mar2018/kremlinglad.html>)

‘Putin mentioned a nuclear-powered missile, an oceanic system with nuclear-powered unmanned underwater vehicles and the **hypersonic** missiles “Kinzhal”.’

- (5) *Uničtožat' protivotankovye orudija, bronetehniku i betonnye doty protivnika "Terminator" sposoben s pomošč'ju **sverchzvukovyh** raket "Ataka-T".* (<https://rg.ru/2017/09/07/rossijskie-terminatory-pokorili-voennyh-sirii-i-izrailia.html>)

‘The “Terminator” is capable of destroying anti-tank guns, armored vehicles and enemy concrete pillboxes with the help of the **supersonic** missiles “Ataka-T”.’

The prefix *ul'tra-* ‘ultra-’ is often used with bases denoting political orientations like left, right, liberal and nationalist, for instance referring to events in the West:

- (6) *V poslednie mesjacy v Germanii usililis' **ul'trapravye** nastroenija.* ([https://www.gazeta.ru/politics/2018/11/30\\_a\\_12078325.shtml](https://www.gazeta.ru/politics/2018/11/30_a_12078325.shtml))

‘**Ultra-right** sentiments have intensified in Germany in recent months.’

The prominence of the prefix *trans-* ‘trans-’ is connected to the crossing of geographical spaces, for example in contexts where Kremlin’s intention of splitting the West is expressed:

- (7) *I v Kieve, i v Brjussele, i v Vašingtone, i v absoljutnom bol'sinstve zapadnyh stolic net nikakich somnenij v podgotovke Kremlëm masštabnyh vmešatel'stv v izbiratel'nye processy, čtoby slomat'*

*opasnoe dlja agressora evropejskoe edinstvo i solidarnost' s Ukrainoj, vnesti raskol v transatlantskij al'jans i podderžat' populistskie, nacionalističeskie i evroskeptično-političeskie sily.* (<https://life.ru/p/1143537>)

'Both in Kiev and in Brussels and in Washington and in the vast majority of Western capitals, there is no doubt that the Kremlin is preparing large-scale interventions in the electoral processes in order to break the dangerous European unity and the solidarity with Ukraine, which is dangerous for the aggressor, to cause a split in the **transatlantic** alliance and to support populist, nationalist and Eurosceptic political forces.'

The prefixes *anti-* 'anti-' and *kontr-* 'counter-' with the meaning of opposition show a decreasing trend throughout the time period. It suggests less polarization in the reporting over time. In the beginning of the studied time period, Putin's return to the Kremlin caused street protests in Russia:

(8) *Aktivisty prokremlevskich dviženij razdavali vsem želajuščim georgievskie lenty i kričali "Putin ljubit vsech!", a protivniki izbrannogo prezidenta skandirovali antiputinskie lozungi.* (<https://utro.ru/articles/2012/05/07/1045258.shtml>)

'Activists of the pro-Kremlin movements handed out St. George ribbons to everyone who wished it and shouted "Putin loves everyone!", but the opponents of the elected president chanted **anti-Putin** slogans.'

The early 2010s is also the time of armed conflict in the Caucasus:

(9) *Obstrel školy proizošel v chode kontrterrorističeskoj operacii, kotoraja provoditsja v Bujnaskom rajone s 5 sentjabrja.* ([https://www.vedomosti.ru/politics/news/2012/09/14/skr\\_shkoluinternat\\_v\\_bujnakske\\_obstrelyali\\_iz\\_minometa\\_po](https://www.vedomosti.ru/politics/news/2012/09/14/skr_shkoluinternat_v_bujnakske_obstrelyali_iz_minometa_po))

'The shelling of the school occurred in the course of an **anti-terrorist** operation that has been underway in the Buynaksky district since September 5.'

The prefix *mež-* 'inter-' also displays a decreasing keyness trend during the actual time frame and that points to less interconnectedness between different political forces as in this example from the beginning of the time period:

(10) *Usilija meždunarodnogo soobščestva dolžny byt' napravleny prežde vsego na dostiženie mežsirijskogo primirenija.* (<https://rg.ru/2012/02/27/putin-politika.html>)

'The efforts of the international community should be directed first and foremost towards achieving **inter-Syrian** reconciliation.'

## 5 Discussion

In this preliminary work, an approach to probe a discourse by using prefixal keymorphs in combination with a derivational resource within the framework of Corpus-assisted discourse studies has been described. The results suggest that a derivational keymorph analysis has the potential to reveal general properties of a discourse. A refinement of the methods awaits further research.

## Acknowledgments

I thank Tora Hedin for proofreading this paper and the anonymous reviewers for their comments and suggestions on important aspects of the research.

## References

- Annelie Ädel. 2010. How to use corpus linguistics in the study of political discourse. In Anne O'Keeffe and Michael McCarthy, editors, *The Routledge Handbook of Corpus Linguistics*, Routledge, chapter 42, pages 591–604.
- Mohamad Ali Afida. 2007. *Semantic fields of problem in business English: Malaysian and British journalistic business texts.* *Corpora* 2(2):211–239. <https://doi.org/https://doi.org/10.3366/cor.2007.2.2.211>.

- Dawn Archer, Jonathan Culpeper, and Paul Rayson. 2009. Love – ‘a familiar or a devil’? an exploration of key domains in Shakespeare’s comedies and tragedies. In Dawn Archer, editor, *What’s in a Word-list?: Investigating Word Frequency and Keyword Extraction*, Routledge, London, pages 137–157. <https://doi.org/https://doi.org/10.4324/9781315547411>.
- Vladimír Benko. 2014a. Aranea: Yet another family of (comparable) web corpora. In P. Sojka, A. Horák, I. Kopeček, and K. Pala, editors, *Text, Speech and Dialogue: 17th international conference, TSD 2014, Brno, Czech Republic, September 8-12, 2014*. Springer, Cham, volume 8655 of *Lecture Notes in Computer Science*, pages 247–254. <https://doi.org/https://doi.org/10.1007/978-3-319-10816-2>.
- Vladimír Benko. 2014b. Compatible sketch grammars for comparable corpora. In Andrea Abel, Chiara Vettori, and Nataschia Ralli, editors, *Proceedings of the 16th EURALEX International Congress*. EURAC research, Bolzano, Italy, pages 417–430.
- Vladimír Benko and Victor P. Zakharov. 2016. Very large Russian corpora: new opportunities and new challenges. In *Computational linguistics and intellectual technologies*, Rossijskij gosudarstvennyj gumanitarnyj universitet, pages 79–93.
- Vladimír Benko and Victor P. Zakharov. 2021. Crowdsourcing for the Russian morphological lexicon. In Radomir V. Bolgov, Nikolay V. Borisov, Andrei V. Chugunov, Dmitry E. Prokudin, Alexander E. Voiskounsky, and Victor P. Zakharov, editors, *Proceedings of the International Conference “Internet and Modern Society” (IMS-2021)*. St. Petersburg, Russia, pages 111–119. <http://ceur-ws.org/Vol-3090/>.
- Ruth Breeze. 2019. Part-of-speech patterns in legal genres: Text-internal dynamics from a corpus-based perspective. In Teresa Fanego and Paula Rodríguez-Puente, editors, *Corpus-based research on variation in English legal discourse*. John Benjamins, volume 91 of *Studies in Corpus Linguistics*, pages 79–103. <https://doi.org/https://doi.org/10.1075/sc1.91>.
- Jonathan Culpeper. 2009. Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare’s Romeo and Juliet. *International Journal of Corpus Linguistics* 14(1):29–59. <https://doi.org/https://doi.org/10.1075/ijcl.14.1.03cul>.
- Jonathan Culpeper and Jane Demmen. 2015. Keywords. In Douglas Biber and Randi Reppen, editors, *The Cambridge Handbook of English Corpus Linguistics*, Cambridge University Press, Cambridge Handbooks in Language and Linguistics, pages 90–105. <https://doi.org/10.1017/CBO9781139764377.006>.
- Václav Cvrček and Masako Fidler. 2019. More than keywords: Discourse prominence analysis of the Russian web portal Sputnik Czech Republic. In Martina Berrocal and Aleksandra Salamurovič, editors, *Political Discourse in Central, Eastern and Balkan Europe*, John Benjamins, volume 84 of *Discourse Approaches to Politics, Society and Culture*. <https://doi.org/https://doi.org/10.1075/dapsac.84.05cvr>.
- Dagmar Divjak and Laura A. Janda. 2008. Ways of attenuating agency in Russian. *Transactions of the Philological Society* 106(2):138–179. <https://doi.org/https://doi.org/10.1111/j.1467-968X.2008.00207.x>.
- Kira Droganova, Olga Lyashevskaya, and Daniel Zeman. 2018. Data conversion and consistency of monolingual corpora: Russian UD treebanks. In *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018), December 13–14, 2018, Oslo University, Norway*. Linköping University Electronic Press, 155, pages 52–65.
- Pavel V. Dyachenko, Leonid L. Iomdin, Aleksandr V. Lazurskij, Leonid G. Mityushin, Olga Ju. Podlesskaya, Viktor G. Sizov, Tatyana I. Frolova, and Leonid L. Tsinman. 2015. Sovremennoe sostojanie gluboko annotirovannogo korpusa tekstov russkogo jazyka (sintagrus). *Trudy Instituta russkogo jazyka im. V.V. Vinogradova* 6:272–300.
- Masako Fidler and Václav Cvrček. 2015. A data-driven analysis of reader viewpoints: Reconstructing the historical reader using keyword analysis. *Journal of Slavic Linguistics* 23(2):197–239. <https://www.jstor.org/stable/24602151>.
- Masako Fidler and Václav Cvrček. 2018. Going beyond “aboutness”: A quantitative analysis of Sputnik Czech Republic. In Masako Fidler and Václav Cvrček, editors, *Taming the Corpus: From Inflection and Lexis to Interpretation*, Springer, Quantitative Methods in the Humanities and Social Sciences, chapter 10. <https://doi.org/https://doi.org/10.1007/978-3-319-98017-1>.
- Masako Fidler and Václav Cvrček. 2019. Keymorph analysis, or how morphosyntax informs discourse. *Corpus Linguistics and Linguistic Theory* 15(1):39–70. <https://doi.org/doi:10.1515/cllt-2016-0073>.



- Costas Gabrieliatos. 2018. *Keyness analysis: nature, metrics and techniques*. In Charlotte Taylor and Anna Marchi, editors, *Corpus Approaches to Discourse: A Critical Review*, Routledge, London, pages 225–258. 1 edition. <https://doi.org/https://doi-org.ezp.sub.su.se/10.4324/9781315179346>.
- Aleksandr Sergeevič Gerd, editor. 2008. *Bol'shoj akademičeskij slovar' russkogo jazyka*, volume 10. Nauka.
- Kirill Sergeevič Gorbačevič, editor. 2005. *Bol'shoj akademičeskij slovar' russkogo jazyka*, volume 2. Nauka.
- Andrew Hardie. 2014. *Log ratio: An informal introduction*. <http://cass.lancs.ac.uk/?p=1133>.
- Laura Alexis Janda, Masako Fidler, Václav Cvrček, and Anna Obukhova. 2023. *The case for case in Putin's speeches*. *Russian Linguistics* 47(1):15–40. <https://doi.org/10.1007/s11185-022-09269-2>.
- Laura Alexis Janda and Olga Lyashevskaya. 2013. *Semantic profiles of five Russian prefixes: po-, s-, za-, na-, pro-*. *Journal of Slavic Linguistics* 21(2):211–258. <https://doi.org/doi:10.1353/jsl.2013.0012>.
- Elena I. Koriakowcewa. 2009. *Internacional'noe vs. nacional'noe v slovoobrazovatel'noj sisteme: k postanovke voprosa*. In Elena Koriakowcewa, editor, *Przejawy internacjonalizacji w językach słowiańskich*, Wydawnictwo Akademii Podlaskiej, Siedlce, pages 179–198.
- Evgeny Kotelnikov, Elena Razova, and Irina Fishcheva. 2017. *A close look at Russian morphological parsers: Which one is the best?* In Andrey Filchenkov, Lidia Pivovarova, and Jan Žižka, editors, *Conference on Artificial Intelligence and Natural Language*. Springer International Publishing, Cham, pages 131–142. [https://doi.org/https://doi.org/10.1007/978-3-319-71746-3\\_12](https://doi.org/https://doi.org/10.1007/978-3-319-71746-3_12).
- Elizaveta Kuzmenko. 2016. *Morphological analysis for Russian: integration and comparison of taggers*. In Dmitry I. Ignatov, Mikhail Yu. Khachay, Valeri G. Labunets, Natalia Loukachevitch, Sergey I. Nikolenko, Alexander Panchenko, Andrey V. Savchenko, and Konstantin Vorontsov, editors, *Proceedings of 5th International Conference on Analysis of Images, Social Networks and Texts*. Springer Cham, Communications in Computer and Information Science, pages 162–171. <https://doi.org/https://doi.org/10.1007/978-3-319-52920-2>.
- Sergej Aleksandrovič Kuznecov. 1998. *Bol'shoj tolkovyj slovar' russkogo jazyka*. Norint, Saint Petersburg.
- Lukáš Kyjánek, Olga Lyashevskaya, Anna Nedoluzhko, Daniil Vodolazsky, and Zdeněk Žabokrtský. 2022. *Constructing a lexical resource of Russian derivational morphology*. In *Proceedings of the 13th Conference on Language Resources and Evaluation Conference (LREC)*. European Language Resources Association, Marseille, France, pages 2788–2797. <https://aclanthology.org/2022.lrec-1.298>.
- Medialogia. 2023. <https://www.mlg.ru>.
- Sergej Ivanovič Ožegov and Natalija Jul'evna Švedova. 1999. *Tolkovyj slovar' russkogo jazyka*. Azbukovnik, Moscow.
- Alan Partington and Alison Duguid. 2020. *Political media discourses*. In Eric Friginal and Jack A. Hardy, editors, *The Routledge Handbook of Corpus Approaches to Discourse Analysis*, Routledge, chapter 8, pages 116–135. 1 edition.
- Alan Partington, Alison Duguid, and Charlotte Taylor. 2013. *Patterns and meanings in discourse: theory and practice in corpus-assisted discourse studies (CADS)*. Number 55 in *Studies in corpus linguistics*. John Benjamins, Amsterdam/Philadelphia.
- Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. *Stanza: A Python natural language processing toolkit for many human languages*. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. pages 101–108. <https://doi.org/https://doi.org/10.18653/v1/2020.acl-demos.14>.
- Larisa V. Ratsiburskaya, Nadezhda A. Samylicheva, and Anna V. Shumilova. 2015. *Novye tendencii v sovremenom medijnom slovotvorčestve*. In Larisa V. Ratsiburskaya, editor, *Novye tendencii v russkom jazyke načala XXI veka*, FLINTA, Moscow, chapter 3, pages 134–221. 2 edition.
- Paul Rayson. 2004. *Keywords are not enough*. Invited talk for JAECS (Japan Association for English Corpus Studies) at Chuo University, Tokyo, Japan. [https://www.lancaster.ac.uk/staff/rayson/publications/jaecs\\_tokyo04.pdf](https://www.lancaster.ac.uk/staff/rayson/publications/jaecs_tokyo04.pdf).
- Paul Rayson. 2008. *From key words to key semantic domains*. *International journal of corpus linguistics* 13(4):519–549. <https://doi.org/https://doi.org/10.1075/ijcl.13.4.06ray>.
- Larissa Ryazanova-Clarke and Terence Wade. 1999. *The Russian language today*. Routledge.

- Pavel Rychlý. 2007. Manatee/Bonito – a modular corpus manager. In *1st Workshop on Recent Advances in Slavonic Natural Language Processing*. Masaryk University, Brno, pages 65–70.
- Nicholas Smith and Cathleen Waters. 2019. [Variation and change in a specialized register: a comparison of random and sociolinguistic sampling outcomes in desert island discs](https://doi.org/https://doi.org/10.1075/ijcl.17117.smi). *International Journal of Corpus Linguistics* 24(2):169–201. <https://doi.org/https://doi.org/10.1075/ijcl.17117.smi>.
- Aleksandr Georgievič Spirkin, Igor' Alekseevič Akčurin, and Regina Semenovna Karpinskaja. 1982. *Slovar' inostrannyh slov*. Russkij jazyk, Moscow, 9 edition.
- Jürgen Spitzmüller and Ingo H. Warnke. 2011. [Discourse as a 'linguistic object': methodical and methodological delimitations](https://doi.org/10.1080/17405904.2011.558680). *Critical Discourse Studies* 8(2):75–94. <https://doi.org/10.1080/17405904.2011.558680>.
- Max Vasmer. 1953. *Russisches etymologisches Wörterbuch*, volume 1. Carl Winter, Heidelberg.
- Max Vasmer. 1955. *Russisches etymologisches Wörterbuch*, volume 2. Carl Winter, Heidelberg.
- Elena Zemskaya. 2006. Aktivnye processy v russkom slovoobrazovanii našego vremeni. *Acta Neophilologica* VIII:9–21.