A keymorph analysis of Russian political news reporting

Thomas Samuelsson

Department of Slavic Languages
Stockholm University

DeriMo 2023
The Fourth International Workshop on Resources and Tools for
Derivational Morphology
5–6 October 2023, Dubrovnik, Croatia

Stockholm
University

Overview

Stockholm
University

- Prefixes, compounds
- Russian online media
- Discourse studies
- Diachrony, 2012–2020
- Corpus linguistics
  - keywords, keymorphs, productivity
- Distributional semantics
  - prefixes anti-, pro-
  - semantic disambiguation, semantic similarity
- Concordance lines

- Russian prefixes
  [Ryazanova-Clarke & Wade 2002, Raciburskaja et al. 2015]
- Russian online media
- Discourse analysis – Corpus-assisted discourse studies
  [Partington et al. 2013, Gillings & Mautner and Baker 2023]
- Keymorphs [Fidler & Cvrček 2019]

Key items

- Keywords [Fidler & Cvrček 2019]
  - content
  - stylistic
- Keymorphs [Fidler & Cvrček 2019]
  - general properties
- Effect size [Hardie 2014]

$$\mathrm{Log\ Rate} = \log_2 \frac{NF_{target}}{NF_{reference}}, \tag{1}$$

where NF is the normalized frequency

Material

Corpus

- Politics,
- 2012–2020
- Citation index [Medialogiya 2020]
- Business, tabloid, regional, independent, state-controlled, state-owned
- automatically tokenized and lemmatized with Stanza [Qi et al. 2020]
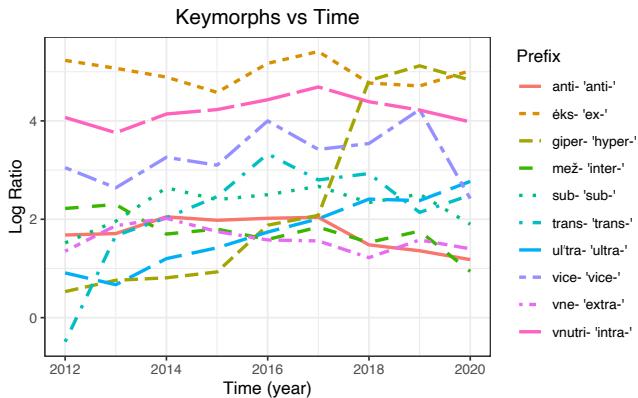- Size is $\approx$ 500 million words from more than 60 resources

Material

Corpus

- Reference corpus
  - Araneum Russicum Maximus
    [Benko 2014, Benko & Zakharov 2016]
  - Russian web language
  - 2010s
  - $20 \cdot 10^9$ tokens

Material

Derivational Resource

- DeriNet.RU 0.5 [Kyjánek et al.]
  - $3 \cdot 10^5$ lexemes
  - $1.63 \cdot 10^5$ binary relations

Stockholm
University

Results

Keymorphs vs Time

Prefix:
- anti- 'anti-'
- éks- 'ex-'
- giper- 'hyper-'
- mež- 'inter-'
- sub- 'sub-'
- trans- 'trans-'
- ul'tra- 'ultra-'
- vice- 'vice-'
- vne- 'extra-'
- vnutri- 'intra-'

Examples

*S drugoj storony, Belyj dom ne spešit so vtorym paketom sankcij protiv Moskvy v svjazi s otravleniem **ėks-polkovnika** GRU Sergeja Skripalja i ego dočeri Julii v britanskom Solsberi.* (Gazeta, 20190516)

'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.'

Stockholm
University

Examples

*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.* (RG, 20170621)

'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.'

Examples

Pri ėtom Putin v ėkskljuzivnom interv'ju avstrijskomu telekanalu ORF zajavil, čto ego vstreča s Trampom do sich por ne sostojalas' iz-za ožestočennoj **vnutripolitičeskoj** *bor'by v SŠA.* (Vesti, 20230604)

'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.'

Examples

Putin upomjanul o rakete s jadernoj ėnergoustanovkoj, okeanskoj sisteme s bespilotnymi podlodkami na jadernoj ustanovke i **giperzvukovych** *raketach "Kinžal".* (NewsRu, 20180323)

'Putin mentioned a nuclear-powered missile, an oceanic system with nuclear-powered unmanned underwater vehicles and the **hypersonic** missiles "Kinzhal".'

Stockholm
University

Examples

*Uničtožat' protivotankovye orudija, bronetechniku i betonnye doty protivnika "Terminator" sposoben s pomošč'ju **sverchzvukovych raket "Ataka-T".*** (RG, 20170907)

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

Examples

V poslednie mesjacy v Germanii usililis' **ul'trapravye** *nastroenija.*
(Gazeta, 20181130)

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

Examples

*I v Kieve, i v Brjussele, i v Vašingtone, i v absoljutnom bol'šinstve zapadnych stolic net nikakich somnenij v podgotovke Kremlëm masštabnych vmešatel'stv v izbiratel'nye processy, čtoby slomat' opasnoe dlja agressora evropejskoe edinstvo i solidarnost' s Ukrainoj, vnesti raskol v **transatlantičeskij** al'jans i podderžat' populistskie, nacionalističeskie i evroskeptično-političeskie sily.* (Life, 20230815)

'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.'

Examples

*Obstrel školy proizošel v chode **kontrterrorističeskoj** operacii, kotoraja provoditsja v Bujnakskom rajone s 5 sentjabrja.* (Vedomosti, 20120914)

'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.'

Examples

*Usilija meždunarodnogo soobščestva dolžny byt' napravleny prežde vsego na dostiženie **mežsirijskogo** primirenija.* (RG, 20120227)

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

Discussion

- Refinement of the method
- Russian derivational prefixes changes with time

References I

Benko V.

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, 2014.

Benko V. and Zakharov P.

Very large Russian corpora: new opportunities and new challenges.
In Computational linguistics and intellectual technologies, Rossijskij gosudarstvennyj gumanitarnyj universitet, pages 79–93, 2016.

Fidler M., and Cvrček V.

Keymorph analysis, or how morphosyntax informs discourse.
Corpus Linguistics and Linguistic Theory, 15(1), pp. 39–70
De Gruyter Mouton, 2019.

Gillings, M., and Mautner, G., and Baker, P.

Corpus-Assisted Discourse Studies.
Cambridge University Press.

Hardie A.

Log Ratio: An informal introduction. https://cass.lancs.ac.uk/?p=1133

References II

Kyjánek L., Lyashevskaya O., Nedoluzhko A., Vodolazsky D. and Žabokrtský Z.
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, 2022.

Medialogiya.
Available at: https://www.mlg.ru/ratings/media/federal/ (accessed April 13, 2022).

Partington A., Duguid A. and Taylor C.
Patterns and meanings in discourse: theory and practice in corpus-assisted discourse studies (CADS).
Amsterdam/Philadelphia: John Benjamins Publishing Company, 2013.

Qi, Peng and Zhang, Yuhao and Zhang, Yuhui and Bolton, Jason and Manning, Christopher D
Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In Association for
Computational Linguistics (ACL) System Demonstrations. 2020.

Raciburskaja L. V. and Samyličeva N. A. and Šumilova A. V.
Novye tendencii v sovremennom medijnom slovotvorčestve. In Novye tendencii v russkom jazyke načala XXI veka
FLINTA, 134–221, 2015.

Ryazanova-Clarke L and Wade T
The Russian language today
Routledge, 2002.

References III

📄 Universal Dependencies

CoNLL-U Format
https://universaldependencies.org/format.html

Thank you!