

The Uppsala Russian Corpus

SND-ID: ext0071-1.

Creator/Principal investigator(s)

Lennart Lönngren - University of Tromsø, Department of Language and Linguistics

Uppsala University, Department of Modern Languages

Research principal

[University of Tromsø](#) - Department of Language and Linguistics

Description

The Uppsala Corpus (Upsal'skij korpus russkix tekstov) consists of some 600 Russian texts with a total of one million running words (word tokens), equally divided between informative and literary prose. The informative texts are from between 1985 and 1989, while the literary texts, whose vocabulary does not date as quickly, cover a longer period, 1960-88. The corpus does not include poetry or drama.

Within the given framework, considerable effort has been made to ensure as representative and varied a corpus as possible. The informative texts are drawn from 25 different subject areas: economics, foreign affairs / foreign policy, ideology / domestic policy, party matters, Soviet society, social issues, defence, education, law, history, culture, linguistics, medicine / health care, psychology, environment / ecology, agriculture, engineering, information technology, space research, energy, biology, geology / geography, physics, chemistry and sport. Certain areas which were felt to be more important are represented by a larger volume of texts.

The literary half of the corpus comprises work by the following 40 authors: Abramov, Ajtmatov, Astaf'ev, Baklanov, Bek, Belov, Bitov, Bondarev, Dubov, Ganin, Gladyshev, Granin, Grekova, Goncharov, Iskander, Kaverin, Kazakov, Kochnev, Kozhevnikova, Nagibin, Lixanov, Lidin, Paustovskij, Pogodin, Pristavkin, Troepol'skij, Rasputin, Shcherbakova, Simonov, Solouxin, Shmelev, Tendrjakov, Tokareva, Tolstaja, Trifonov, Vasil'ev, Vorobl'ev, Zalygin and Zorin. Here, too, there is unequal representation, with a larger amount of writing by the better-known authors.

For further details about the corpus, see Lönngren, Lennart (ed.), 1993. Chastotnyj slovar' sovremennogo russkogo jazyka. (A Frequency Dictionary of Modern Russian. With a Summary in English.) Acta Universitatis Upsaliensis, Studia Slavica Upsaliensia 32. 188 pp. Uppsala. ISBN 91-554-3134-8.

Purpose:

The aim is to provide a corpus of Russian prose texts.

Data contains personal data

No

Time period(s) investigated

1960 - 1988

1985 - 1989

Data format / data structure

[Text](#)

Resource type

Corpus

Responsible department/unit

Department of Language and Linguistics

Research area

[Languages and literature](#) (Standard för svensk indelning av forskningsämnen 2011)

[Language and linguistics](#) (CESSDA Topic Classification)

Keywords

[Russian](#), [Corpus](#)

Publications

Lönngren, Lennart (ed.), 1993. Частотный словарь современного русского языка. [Chastotnyj slovar' sovremennogo russkogo jazyka.] (A Frequency Dictionary of Modern Russian. With a Summary in English.) Acta Universitatis Upsaliensis, Studia Slavica Upsaliensia 32. 188 pp. Uppsala

ISBN: 91-554-3134-8

If you have published anything based on these data, [please notify us](#) with a reference to your publication(s). If you are responsible for the catalogue entry, you can update the metadata/data description in DORIS.

Accessibility level

Access to data through an external actor

Data are freely accessible

Homepage

[Uppsala Russian Corpus](#)

CLARIN Virtual Collection Registry

[Add to collection](#)

A virtual collection is connected to a specific research purpose and contains links to data resources from various digital archives. It is easy to create, access, and cite the collection.

Read more about virtual collections on the [CLARIN website](#).

Download metadata

[DataCite](#)

[MetaShare](#)

[MetaShare-CMDI](#)

[DDI 2.5](#)

[DDI 3.3](#)

[DCAT-AP-SE 2.0](#)

[JSON-LD](#)

[PDF](#)

[Citation \(CLS\)](#)