

CURRICULUM VITAE

Karina Sheifer

Hanover, New Hampshire, USA
kar.sheifer@gmail.com



FIELD OF INTEREST

Low-resource NLP, field linguistics, morphological complexity, language revitalization, Northern Tungusic, Siberian Turkic, Chukotko-Kamchatkan, Finnic, Basque, Iberian Romance languages

LANGUAGE PROFICIENCY

	Listening	Reading	Speaking	Writing
Russian	C2	C2	C2	C2
English	C2	C2	C1	C2
Spanish	C2	C2	C2	C2
French	B2	C2	C1	C1
Basque	B2	C1	B2	B2
Catalan	B2	C1	B1	B2
Finnish	B1	B2	B1	B2
Dolgan	B1	B2	B1	B2
Evenki	B1	B2	B1	B2

PROGRAMMING SKILLS

Python, R, PostgreSQL, LaTeX (intermediate); Machine Learning (basic)

EDUCATION

2019 – 2021. Master studies in Computational linguistics, High School of Economics, Moscow; thesis: Automatic Analysis of Incorporation in Chukchi

2017 – 2018. Vocational studies in Programming (voluntary resignation), College of Automation and Information Technology, Moscow

2015 – 2018. PhD studies in Comparative, Typological and Contrastive Linguistics (all but dissertation), Lomonosov Moscow State University

2010 – 2015. Specialist degree program in Iberian Romance Philology, Lomonosov Moscow State University

2013. International mobility program, Faculty of Arts, University of the Basque Country, Spain

ADVANCED TRAINING COURSES

2022. School of Applied Big Data Analysis for Researchers of Indigenous Peoples of the North, Siberia and the Far East (online), University Consortium of Big Data Researchers

2019. Language Documentation Training Workshop (online), University of Hawai'i Language Documentation Training Center and Endangered Language Project

2018. 22nd EDUFI Winter School "Big Numbers, Small Numbers – from Data to Understanding" (Lammi, Finland), Finnish National Agency for Education

2018. Course on Phonetics and Graphics of the Altaic and Uralic languages (Tomsk, Russia), Laboratory of Linguistic Anthropology, Tomsk State University

EMPLOYMENT IN LINGUISTICS

2023 – present. Research Scholar, Department of Linguistics, Dartmouth College

2019 – present. Research Scholar, Laboratory of Studies and Preservation of Minority Languages, Institute of Linguistics, Russian Academy of Sciences

2015 – present. Support Staff, Russian Language Institute, Russian Academy of Sciences

2020 – 2022. Research Scholar, Laboratory of Linguistic Platforms, Institute of System Programming, Russian Academy of Sciences

2018 – 2019. Research Assistant, Department of the Uralic and Altaic languages, Institute of Linguistics, Russian Academy of Sciences

TEACHING

2020 – present. Associate Professor of [Spanish](#), School of Linguistics, High School of Economics, Moscow

2020 – 2022. Associate Professor of [Universal Language Data Annotation Systems](#), School of Linguistics, High School of Economics, Moscow

2015 – 2017. Teaching Assistant of Introduction to Linguistics, Lomonosov Moscow State University

DIGITAL PROJECTS

2022 – present. [Semantic Hub](#) (underdiagnosed patient journey modeling, data processing)

2022 – present. [Chukchi dictionary](#) (data processing)

2020 – present. [Diachronic Itelmen corpus](#) (research, development, data processing)

2019 – present. [Digital encyclopedia of the Minority Languages of Russia](#) (editing)

2021. [UniMorph Itelmen dataset](#) (research, development, data processing)

2021. [UniMorph Chukchi dataset](#) (research, development, data processing)

2021. [Indigenous Minority Peoples Atlas on Evenki](#) (editing)

2020. [Dynamics of language contact in the circumpolar region](#) (research)

2020. [Universal Dependencies Chukchi treebank](#) (research, development, data processing)

2018 – 2021. [Dynamics of the Language Situation in Siberia](#) (research)

2018 – 2019. [Multi-media collection in Evenki, Selkup and Ket](#) (research, data processing)

2018 – 2019. [Audial dictionaries in Minority languages of Russia](#) (research, data processing)

2016 – present. [Finnish-Russian parallel corpus](#) (data processing)

2015 – 2019. [Spanish-Russian parallel corpus](#) (data processing)

FIELD WORK

December 2021. Expedition on the Evenki language in the Tynda district, Amur province, Russia

August – September 2021. Expedition on the Evenki language in the North-Baikal district, Republic of Buryatia, Russia

July – August 2021. Expedition on the Evenki language in the Ayan-May district, Khabarovsk region, Russia

January 2020. Expedition on the Even language in the Bystraya district, Kamchatka region, Russia

October 2019. Expedition on the Evenki language in Krasnoyarsk, Krasnoyarsk region, Russia

July – August 2019. Expedition on the Evenki language in the Tugur-Chumikan district, Khabarovsk region, Russia

July – August 2018. Expedition on the Evenki language in the Katanga and Bodaybo districts, Irkutsk province, Russia

July 2016. Expedition on the Basque language in Amorebieta-Etxano, Bizkaia, Spain

July 2011. Expedition on the Basque language in Lazkano, Guipúzcoa, Spain

LANGUAGE ACTIVISM

2022 – present. Development of the language board game "[Nomad Camp](#)" in Forest Enets and Nganasan

2019 – present. Compiling a digital [encyclopedia](#) of the Minority Languages of Russia

2022. Supervising a "[master-apprentice](#)" project for Amur Evenki and Magadan Even

2022. Supervising an Evenki language nest project "[Evedy Umuk](#)" in Amur province

2022. Internship in a Karelian language nest "[Kielipezä](#)" in Republic of Karelia

2021, 2022. Organization and conduct of the [Festival of National Literatures of Russia's Peoples](#) on the Red Square

2022. Organization and conduct of the Program of National Literatures of Russia's Peoples on the Baku Book Fair and the St. Petersburg International Book Salon

2022. Supervising a project of the [Lullabies of the peoples of Russia](#)

2019 – 2022. Creation and administration of the community on the Evenki and Even languages and cultures "Tungusic Notes" in [Instagram](#) (3,521) and [Vkontakte](#) (1,673 subscribers); awarded [Ethnblog of the year-2021](#)

2022. Development and introduction of the language board game "Nomad Camp" in Kildin Saami and Mansi into the national communities

2021. Organization and conduct of the [International Olympiad "Turen"](#) on the language and culture of the Evenki people of Russia and the Oroqen people of China

2021. Development and introduction of the language board game "[Nomad Camp](#)" in Dolgan, Tundra Yukaghir and Forest Yukaghir into the national communities

2021. Organization and conduct of the Program of National Literatures of Russia's Peoples on the [Frankfurt Book Fair](#) and the Krasnoyarsk Book Culture Fair

2020, 2021. Organization and conduct of the Language Projects Workshop for the Indigenous Youth on the forum "Russian North"

2020. Development and introduction of the language board game "Nomad Camp" in Koryak and Alyutor into the national communities

2019 – 2020. Development and introduction of the language board game "Nomad Camp" in ten Evenki and four Even dialects into the national communities

2020. Longreads on [Itelmen](#) and [Even](#) for the popular science project "PostScience"

PUBLICATIONS

2022. Batsuren et al. [UniMorph 4.0: Universal Morphology](#). Proceedings of the 13th Language Resources and Evaluation Conference. PP. 840–855.

2022. Kazakevich O. et al. [Shkaly yazykovej vital'nosti i ih primenimost' k materialu konkretnyh yazykovyh situacij](#) {Scales of linguistic vitality and their applicability to specific linguistic situations}. Questions of Linguistics, 2022, № 4. PP. 7–47.

2022. Sheifer K. [Nastol'nye igry kak instrument izucheniya rodnogo yazyka](#) {Board games as a tool for learning the native language}. In proceedings of "Media and educational infrastructure for the languages of the peoples of Russia", Moscow.
2021. Sheifer K. and Ganieva S. Development of the Itelmen corpus: challenges, lacunae and solutions. In Proceedings of Electronic Writing of the Peoples of the Russian Federation – 2021 & IWCLUL2021.
2021. T. Pimentel et al. [SIGMORPHON 2021 Shared Task on Morphological Reinflection: Generalization Across Languages](#). 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology. Association for Computational Linguistics. Bangkok, Thailand. 2021. PP. 154–184.
2020. Tyers F. and Mishchenkova (Sheifer) K. 2020. [Dependency annotation of noun incorporation in polysynthetic languages](#). In Proceedings of the Fourth Workshop on Universal Dependencies (UDW 2020). Association for Computational Linguistics. Barcelona, Spain. PP. 195–204.
2020. Kazakevich O. et al. [The Website «Minority Languages of Russia»: Visualizing Language Data](#). Proceedings of the Linguistic Forum 2020: Language and Artificial Intelligence, Moscow.
2020. Mishchenkova (Sheifer) K. [Refleksy praevenkijskogo *s v govorah evenkijskogo yazyka vo vtoroj polovine XVIII v.](#) {Reflections of the Proto-Evenki *s in the Evenki dialects in the second half of the 18th century}. Ural-Altai Studies. № 2 (37). PP. 32-58.
2019. Mishchenkova (Sheifer) K. [Refleksy praevenkijskogo *s v govorah evenkijskogo yazyka v konce XVII v. i pervoj polovine XVIII v.](#) {Reflections of the Proto-Evenki *s in the Evenki dialects in the late 17th century and the first half of the 18th century}. Ural-Altai Studies. № 3 (34). P. 72-83.
2019. Kazakevich O. et al. (eds.) 'Evenkijskie rasskazy. Sbornik rasskazov evenkov Irkutskoj oblasti 1927–2018 gg.' {[Evenki stories. Collection of short stories by Evenki of Irkutsk region, 1927-2018](#)}. 'Prius' Publishing House, Irkutsk. P. 92. ISBN 978-5-6042268-5-8.
2018. Mishchenkova (Sheifer) K. 'Finsko-russkij i russko-finskij parallel'nyj korpus v sostave NKRYa: principy formirovaniya i perspektivy razvitiya' {Finnish-Russian and Russian-Finnish parallel corpus as part of the National Corpus of Russian: principles of formation and prospects for development}. Proceedings of All-Russia Conference 'From Language Machine Funds to Linguistic Corpora: in memory of V. M. Andryushchenko', Moscow.
2016. Joseba Sarrionandia. [Hamabi poema](#) transl. from Basque into Russian by Mishchenkova (Sheifer) K. EIZIE, Donostia/San Sebastián. 90 P. ISBN: 978-84-617-6543-0.

2015. Mishchenkova (Sheifer) K. Aspektual'nye glagol'nye perifrazy fazisnogo znacheniya v ispanskom i baskskom yazykah {Aspectual verbal periphrases in Spanish and Basque}. Proceedings of VIII International Scientific Conference "Romance Languages and Cultures: from Antiquity to Modernity", Moscow.