

Timofey Arkhangelskiy, Ph. D.

Born 14 September 1987.

Current position: Alexander von Humboldt postdoctoral fellow, Institute for Finno-Ugrian and Uralic Studies, Universität Hamburg

Home address: Peter-Timm-Straße 26, 22457 Hamburg, Germany

e-mail: timarkh@gmail.com

Education

- 2004–2009: specialist degree, Lomonosov Moscow State University, faculty of Mechanics and Mathematics; magna cum laude
- 2009–2011: postgraduate studies (Aspirantur), Lomonosov Moscow State University, faculty of Mechanics and Mathematics
- 2012: Ph. D. (Candidate of Sciences), area 10.02.21 “Applied and Mathematical Linguistics”. Dissertation “Principles of development of a morphological analyzer for languages with different structures”, defended at Lomonosov Moscow State University, Faculty of Philology, Department of Theoretical and Applied Linguistics

Teaching

- 2011–2017: National Research University Higher School of Economics (Moscow, Russia), School of Linguistics (formerly Linguistic department of the Faculty of Philology)
- “Best Teacher” award in 2012, 2013, 2014, 2015 (based on students’ vote)

Previous work experience

- Senior lecturer, National Research University Higher School of Economics (Moscow, Russia), School of Linguistics (2014–2017)
- Lecturer, National Research University Higher School of Economics (Moscow, Russia), School of Linguistics (2011–2014)

Language skills: Russian (native); English (fluent); German, French, Modern Greek, Udmurt, Yiddish (reading); Erzya, Hungarian, Polish, Eastern Armenian (poor)

Technical skills: python (very good); C++, C# (good); perl, php, JavaScript (intermediate level); SQL, elasticsearch; HTML/CSS, T_EX; Linux server administration experience

Professional interests: corpus linguistics, natural language processing, language documentation, morphology, language contact, Uralic languages, Udmurt

Fieldwork

- August 2012; February, August 2013; January, August 2014; January, August 2015; January, July–August, October 2016; January, August 2017; August 2018: Shamardan village, Yukamenskoe district, Udmurtia, Russia (Beserman dialect of the Udmurt language)

- July 2016; July 2017: Parshakova, Antipina, Vankova villages, Krasnovishersk district, Perm krai, Russia (Yazva Komi language)
- July 2015: Moldova (Chişinău, Tiraspol, Rîbniţa), Bessarabian dialect of Yiddish
- July 2014: Kiangang municipality, Ifugao province, the Philippines (Tuwali Ifugao language)
- July 2013: Hatazhukay village, Shovgenovsk district, Adygea, Russia (Adyge language)

Committees and reviewing

Program committees:

- International Workshops for Computational Linguistics in 2015–2018
- “Language, culture and history of the Beserman people: state of the art and perspectives” conference (Glazov, Russia, April 20–21, 2018)
- 4th workshop on languages of the Volga-Kama Sprachbund (Moscow, Russia, November 16–18, 2018)

Reviewing done for *Linguistica Uralica*, Mandenkan, Annual conference “Dialogue” (Moscow, Russia; 2017, 2018)

Participation in projects

- 2006 – present: various projects within the framework of the Russian National Corpus (main, poetic, parallel, dialectological subcorpora)
- 2008–2009: research project of the Eastern Armenian National Corpus “Corpus studies of grammar”
- 2008–2011: projects aimed at development of the Database of Polysemous Qualitative Adjectives and Adverbs (Russian Foundation for Basic Research grants 08-06-00197 and 11-06-00385)
- 2011–2014: number of projects supported by the “Corpus Linguistics” program of the Russian Academy of Sciences, including projects aimed at development of Albanian, Armenian dialectological, Buryat, Modern Greek, Ifugao, Kalmyk, Lezgian, Luwian, Ossetic corpora, as well as historical subcorpora of the Russian National Corpus
- 2011–2012: “Comparative corpus study of Russian and Belarusian languages and development of parallel corpora” (Russian Foundation for Humanities grant 11-24-01004)
- 2012–2014: “Contrastive corpus study of Russian and French verbal categories in parallel texts with multiple translations” (Russian Foundation for Basic Research grant 12-06-33038)
- 2013: “System for automatic online publishing of grammatically annotated Old Russian texts” (Russian Foundation for Humanities grant 13-04-12006)
- 2014: Project aimed at creation of Kazakh language corpus (in cooperation with Al-Farabi Kazakh National University)
- 2014–2016: “Corpus of dialectal texts within the Russian National Corpus” (Russian Foundation for Humanities grant 14-04-12012)

- 2014–2015: “Development of a Russian Romani language corpus” (Russian Foundation for Basic Research grant 14-06-31038)
- 2014–2015: “Development of a series of tools for language documentation projects” (Russian Foundation for Humanities grant 14-04-12012)
- 2014–2015: “Developing and applying corpus tools for the study of Modern Yiddish”, research project supported by National Research University Higher School of Economics (head of the project)
- 2014–2017: Project aimed at development of a comparative dictionary of Anatolian languages “Digitales philologisch-etymologisches Wörterbuch der altanatolischen Kleinkorpussprachen”, supported by Deutsche Forschungsgemeinschaft
- 2015–2017: “Digital documentation of a polysynthetic language” (Russian Foundation for Basic Research grant 15-06-07434)
- 2015–2017: “Buryat-Russian parallel corpus” (Russian Foundation for Basic Research grant 15-46-04417)
- 2016–2018: “Integral analysis of NP in Finno-Ugric languages: Cohesion and information structure encoding” (Russian Foundation for Humanities grant 16-24-17003)
- 2017–2019: “Developing a corpus of Neo-Aramaic languages: Judaeo-Aramaic and Turoyo” (Russian Foundation for Humanities grant 17-04-00472)

Publications

1. T. Arkhangelskiy. 2019 (to appear). Russian verbal borrowings in Udmurt. In: *Folia Linguistica*.
2. T. Arkhangelskiy, M. Usacheva. 2018 (to appear). Case compounding in Beserman Udmurt. In: *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*.
3. N. Serdobolskaya, M. Usacheva, T. Arkhangelskiy. 2018 (to appear). Grammaticalization of possessive markers in the Beserman dialect of Udmurt. In: Johanson L., Nevskaya I., Mazzitelli L.F. (eds.). *Linguistic possession. New insights from the languages of Europe and North and Central Asia. (Studies in Language Companion Series)*. John Benjamins. P. 291–311 (preliminary).
4. T. Arkhangelskiy, E. Georgieva. 2018. Sound-aligned corpus of Udmurt dialectal texts. In: *Proceedings of the Fourth International Workshop on Computational Linguistics of Uralic Languages*. P. 26–38.
5. M. Usacheva (chief ed.); T. Arkhangelskiy, O. Biryuk, R. Idrisov, V. Ivanov (eds.). 2017. *Тезаурус бесермянского наречия. Имена и служебные части речи (говор деревни Шамардан) [Thesaurus of the Beserman dialect. Nouns and functional parts of speech (Shamardan village variety)]*. Moscow: Izdatelskie resheniya, 2017.
6. T. Arkhangelskiy, N. Serdobolskaya, M. Usacheva. 2017. Corpus-oriented lexicographic database for Beserman Udmurt. In: *Acta Linguistica Academica* 64(3), P. 397-415.
7. M. Usacheva, T. Arkhangelskiy. 2017. Grammaticalization of New Cases in Beserman Udmurt. In: *Linguistica Uralica* 53(1). P. 27–48.

8. T. Arkhangelskiy, Yu. Lander. 2016. Developing a polysynthetic language corpus: problems and solutions // *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2016"*. No. 15(22). P. 38–47.
9. T. Arkhangelskiy, M. Medvedeva. 2016. Developing morphologically annotated corpora for minority languages of Russia. In: *CEUR Workshop Proceedings*, Vol. 1607. P. 1–6.
10. T. Arkhangelskiy, M. Usacheva. 2015. Syntactic and Morphosyntactic Properties of Postpositional Phrases in Beserman Udmurt as Part-of-Speech Criteria // *SKY Journal of Linguistics*. No. 28. P. 103–137.
11. T. Arkhangelskiy, N. Tyshkevich. 2015. Gradual language death: The case of Bessarabian Yiddish. Working paper within the framework of the Basic research program, National Research University Higher School of Economics, series "Linguistics", No. WP BRP 36/LNG/2015
12. T. Arkhangelskiy, Yu. Lander. 2015. Some challenges of the West Circassian polysynthetic corpus. Working paper within the framework of the Basic research program, National Research University Higher School of Economics, series "Linguistics", No. WP BRP 37/LNG/2015
13. Yu. Lander, T. Arkhangelskiy. 2015. Producing polysynthetic verb forms in West Circassian (Adyghe): an experimental study. Working paper within the framework of the Basic research program, National Research University Higher School of Economics, series "Linguistics", No. WP BRP 23/LNG/2015
14. T. Arkhangelskiy, R. Idrisov, N. Serdobolskaya, M. Usacheva. 2015. Designing a Lexicographic Database for an Online Dictionary. Annual conference "Dialogue": Conference Proceedings "Computational Linguistics and Intellectual Technologies"
15. T. Arkhangelskiy, O. Sozinova. 2015. Мультимедийный корпус языка идиш [Multimedia corpus of the Yiddish language] // *Научно-техническая информация. Серия 2: Информационные процессы и системы*. No. 3. P. 18–25
16. T. Arkhangelskiy. 2014. Estimating Native Vocabulary Size in an Endangered Language // *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: Proceedings of the Workshop on the Use of Computational Methods in the Study of Endangered Languages*. P. 63–67
17. T. Arkhangelskiy. 2014. Clitics in the Beserman dialect of Udmurt. Working paper within the framework of the Basic research program, National Research University Higher School of Economics, series "Linguistics", No. WP BRP 10/LNG/2014
18. T. Reznikova, E. Rakhilina, O. Karpova, T. Arkhangelskiy, M. Kyuseva, D. Ryzhova. 2013. Polysemy Patterns in Russian Adjectives and Adverbs: a Corpus-Oriented Database // *Current Studies in Slavic Languages*. Philadelphia, Amsterdam: John Benjamins Publishing Company. P. 313–322
19. A. Kuznetsova, N. Serdobolskaya, M. Usacheva, O. Biryuk, R. Idrisov (eds.). 2013. Словарь бесермянского диалекта удмуртского языка [Dictionary of the Beserman dialect of the Udmurt language]. Moscow: Thesaurus.

20. T. Arkhangelskiy, A. Rusakov, M. Morozova, M. Daniel. 2012. Korpusi i gjuhës shqipe: drejtimet kryesore të punës [Albanian language corpus: principal directions of development] // Shqipja dhe gjuhët e Ballkanit – Albanian and Balkan Languages. Konferencë e mbajtur në 10–11 dhjetor 2011, Prishtinë. Prishtinë: ASHAK. P. 635–642
21. T. Arkhangelskiy, O. Belyaev, A. Vydrin. 2012. The creation of large-scaled annotated corpora of minority languages using UniParser and the EANC platform // Proceedings of COLING 2012: Posters. Mumbai: The COLING 2012 Organizing Committee. Ch. 9. P. 83–91
22. T. Arkhangelskiy, V. Panov. 2012. Аспект в греческом языке: проблемные зоны и типология [Aspect in the Greek language: problematic areas and typology] // V. Plungian (ed.). Acta linguistica Petropolitana. Transactions of the Institute for linguistic studies, Vol. VIII, part 2 / Studies in the theory of grammar. Issue 6: Typology of aspectual systems and categories. St. Petersburg: Nauka. P. 122–148
23. D. Sitchinava, T. Arkhangelskiy. 2012. Параллельные белорусско-русский и русско-белорусский корпусы: совместный проект Национального корпуса русского языка [Parallel Belarusian-Russian and Russian-Belarusian corpora: a joint project of the Russian National Corpus] // Корпусы национальных языков: модели и технологии. Труды Казанской школы по компьютерной и когнитивной лингвистике TEL-2012. Kazan: Publishing house of the Academy of Sciences of Tatarstan “Fen”. P. 54–60
24. T. Arkhangelskiy. 2012. Электронные корпуса албанского, калмыцкого, лезгинского и осетинского языков [Electronic corpora of the Albanian, Kalmyk, Lezgian, and Ossetic Languages] // Научно-техническая информация. Серия 2: Информационные процессы и системы. No. 4. P. 24–29
25. T. Arkhangelskiy, O. Belyaev. 2011. A Comparison of Eastern Armenian and Iron Ossetic Spatial Systems // Languages and Cultures in the Caucasus. Papers from the International Conference “Current Advances in Caucasian Studies”, Macerata, January 21–23, 2010. München – Berlin: Verlag Otto Sagner. P. 285–299
26. D. Ryzhova, T. Arkhangelskiy, O. Karpova, M. Kyuseva, T. Reznikova, E. Rakhilina. 2011. База данных по многозначным качественным прилагательным и наречиям русского языка: первые результаты работы [Database of polysemous qualitative adjectives and adverbs in Russian: first results] // Проблемы компьютерной лингвистики Вып. 5. Voronezh: Voronezh state university
27. T. Arkhangelskiy, M. Tagabileva, L. Kholkina. 2011. Качественные признаки «чистый», «грязный», «прозрачный», «мутный»: к построению семантической типологии [Properties “clean”, “dirty”, “transparent”, “opaque”: towards the development of the semantic typology] // N. Kazansky, D. Gerasimov (eds.). Acta Linguistica Petropolitana. Transactions of the Institute for linguistic studies, Vol. VII, part 3. St. Petersburg: Nauka. P. 257–261
28. O. Karpova, T. Arkhangelskiy, M. Kyuseva, E. Rakhilina, T. Reznikova, D. Ryzhova, M. Tagabileva. 2010. База данных по многозначным качественным прилагательным и наречиям русского языка [Database of polysemous qualitative adjectives and adverbs in Russian] // A. Kibrik (ed.). Computational Linguistics and Intellectual Technologies: papers from the Annual conference “Dialogue”. Vol. 9, Issue 16. Moscow: Russian State University for the Humanities. P. 163–168

Conference talks

- T. Arkhangelskiy. Конструкция с соположением в бесермянском [Juxtaposition construction in Beserman Udmurt] // The Conference on Uralic, Altaic and Paleo-Asiatic languages in the memory of Alexander P. Volodin. St. Petersburg, December 6–8, 2018
- T. Arkhangelskiy. Russian verbal loans in Udmurt // 15th Conference on Typology and Grammar for Young Scholars. St. Petersburg, November 22–24, 2018
- T. Arkhangelskiy. Beserman corpora and corpus studies // 4th workshop on languages of the Volga-Kama Sprachbund. Moscow, November 16–18, 2018
- T. Arkhangelskiy. Handling multilingual content in social media corpora in Uralic languages // Language diversity and digital infrastructures workshop. Hamburg, September 20–21, 2018
- M. Usacheva, T. Arkhangelskiy. Interface of morphology, syntax and information structure in NPs in Beserman Udmurt // 51st Annual Meeting of the Societas Linguistica Europaea. Tallinn, August 29 – September 1, 2018
- M. Usacheva, T. Arkhangelskiy. Experiments in linguistic fieldwork: the case for case compounding in Beserman Udmurt // Word-Formation Theories III and Typology and Universals in Word-Formation IV. Košice, June 27–30, 2018
- T. Arkhangelskiy, E. Sokur, M. Usacheva. Non-verbal nominalization in Udmurt // International Morphology Meeting 2018. Budapest, May 10–13, 2018
- T. Arkhangelskiy. Изучение бесермянского наречия при помощи корпуса [Researching the Beserman dialect with the help of a corpus] // Language, culture and history of the Beserman people: state of the art and perspectives. Glazov, April 20–21, 2018
- T. Arkhangelskiy. Identifying dialectal features of the Udmurt language with the help of an internet corpus // The 6th Conference on Text Interpretation. Tyumen, February 12–18, 2018
- T. Arkhangelskiy. Processing fieldwork data in the Beserman Udmurt documentation project // Computational Methods for Endangered Language Documentation and Description. Paris, February 1–2, 2018
- T. Arkhangelskiy, E. Georgieva. Sound-aligned corpus of Udmurt dialectal texts // Fourth International Workshop on Computational Linguistics for Uralic Languages. Helsinki, January 8–9, 2018
- T. Arkhangelskiy, M. Usacheva. Russian influence on NP structure in Yazva Komi // 14th Conference on Typology and Grammar for Young Scholars. St. Petersburg, November 23–25, 2017
- T. Arkhangelskiy, M. Usacheva. Complex adpositions in Permic languages // Workshop on Complex Adpositions in European Languages. Neuchâtel, October 6–7, 2017
- T. Arkhangelskiy, M. Usacheva. Morphological encoding of NPs and information structure in Beserman Udmurt // Syntax Of Uralic Languages. Budapest, June 27–28, 2017
- T. Arkhangelskiy, M. Usacheva. Spontaneous Russian verbal borrowings in Udmurt // Contact-driven Multilingual Practices. Helsinki, June 1–2, 2017

- T. Arkhangelskiy. Делиберативное будущее в бесермянском диалекте удмуртского языка [Deliberative future in Beserman Udmurt] // Workshop on typology of the future-prospectual semantic area. St. Petersburg, December 8–10, 2016
- T. Arkhangelskiy, M. Usacheva. Functions of the 3Sg possessive marker in Beserman Udmurt: Corpus analysis // 13th Conference on Typology and Grammar for Young Scholars. St. Petersburg, November 24–26, 2016
- T. Arkhangelskiy, M. Usacheva. Конструкции с внешним посессором в бесермянском удмуртском [External possessor constructions in Beserman Udmurt] // Typology of Morphosyntactic Parameters 2016. Moscow, October 12–14, 2016
- T. Arkhangelskiy. Digital corpora at web-corpora.net: Features and development // 8th Tsakonian conference. Leonidio, Greece, September 9–11, 2016
- T. Arkhangelskiy, M. Usacheva. Case compounding in Beserman Udmurt: Syntax and information structure // 49th Annual Meeting of the Societas Linguistica Europaea. Naples, Italy, August 31 – September 3, 2016
- T. Arkhangelskiy, M. Medvedeva. Developing Morphologically Annotated Corpora for Minority Languages of Russia // Corpus Linguistics Fest. Bloomington, IN, July 6–10, 2016
- T. Arkhangelskiy, Yu. Lander. Developing a Polysynthetic Language Corpus: Problems and Solutions // Annual conference “Dialogue” 2016. Moscow, Russia, June 1–4, 2015
- T. Arkhangelskiy, E. Kuzmenko. Composing the Corpus of Modern Greek: features and methods // 12th International Conference on Greek Linguistics. Berlin, Germany, September 16–19, 2015
- E. Kuzmenko, T. Arkhangelskiy. Automatic morphological disambiguation in the Corpus of Modern Greek // 12th International Conference on Greek Linguistics. Berlin, Germany, September 16–19, 2015
- T. Arkhangelskiy, N. Serdobolskaya, M. Usacheva, R. Idrisov. A multimedia lexicographic resource for the Besermyan dialect of the Udmurt language // XII International Congress for Finno-Ugric Studies. Oulu, Finland, August 17–21, 2015
- T. Arkhangelskiy, T. Panova. Мультимедийный корпус языка идиш [Yiddish Multimedia Corpus] // 20th International youth conference on Jewish Studies “Sefer”. Moscow, Russia, June 12, 2015
- T. Arkhangelskiy, R. Idrisov, N. Serdobolskaya, M. Usacheva. Designing a Lexicographic Database for an Online Dictionary // Annual conference “Dialogue” 2015. Moscow, Russia, May 27–30, 2015
- T. Arkhangelskiy, M. Medvedeva. An online annotated corpus of Udmurt language // 4th Mikola Conference on Lexicology and Lexicography of the Uralic and Siberian languages. Szeged, Hungary, November 14–15, 2014
- T. Arkhangelskiy, E. Mishina, A. Pichkhadze. Система электронной грамматической разметки древнерусских и церковнославянских текстов и её использование в веб-ресурсах [A System for Digital Morphological Tagging for Old Russian and Church Slavonic Texts and Its Use in Web Resources] // El'Manuscript-2014: Textual Heritage and Information Technologies. Varna, Bulgaria, September 15–20, 2014

- T. Arkhangelskiy. Estimating Native Vocabulary Size in an Endangered Language // 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, MA, June 22–27, 2014
- T. Arkhangelskiy. О создании электронного корпуса диалекта тували языка ифугао [About the development of a digital Ifugao language corpus] // Maclay readings – 2014. St. Petersburg, Russia, April 15–16, 2014
- T. Arkhangelskiy. Система морфологического анализа текстов UniParser [*UniParser* morphological analysis tool] // Conference on Computational and Cognitive Linguistics “TEL-2014”. Kazan, Russia, February 6–9, 2014
- T. Arkhangelskiy. Энклитики в бесермянском диалекте удмуртского языка [Enclitics in the Beserman dialect of Udmurt] // 10th Conference on Typology and Grammar for Young Scholars. St. Petersburg, Russia, November 21–23, 2013
- T. Arkhangelskiy, N. Zevakhina, A. Letuchiy, A. Plisetskaya. Корпус ошибок в письменных текстах носителей русского языка [A corpus of mistakes in texts written by native Russian speakers] // Corpus technologies: Digital humanities and contemporary knowledge. Nizhny Novgorod, Russia, October 17–19, 2013
- T. Arkhangelskiy, M. Daniel, M. Morozova, A. Rusakov. The morphology of Albanian nouns in the Albanian National Corpus: some issues of representation // 46th Annual Meeting of the Societas Linguistica Europaea. Split, Croatia, September 18–21, 2013
- E. Klyachko, T. Arkhangelskiy, O. Kisselev, E. Rakhilina. Automatic error detection in Russian learner language // Corpus Linguistics 2013. Lancaster, UK, July 22–26, 2013
- T. Arkhangelskiy. Групповая флексия в именной сочинительной конструкции восточноармянского языка [Suspended Affixation in Nominal Coordination in Modern Eastern Armenian] // IX International conference on Armenian linguistics. St. Petersburg, Russia, November 7–9, 2012
- T. Arkhangelskiy, M. Tagabileva, L. Kholkina. Качественные признаки «чистый», «грязный», «прозрачный», «мутный»: к построению семантической типологии [Properties “clean”, “dirty”, “transparent”, “opaque”: towards the development of the semantic typology] // 8th Conference on Typology and Grammar for Young Scholars. St. Petersburg, Russia, November 17–19, 2011
- T. Arkhangelskiy, V. Panov. Non-Prototypical Uses of Aspectual Markers in Greek // Verbal Aspect: Grammatical Meaning and Context. Third Conference of the International Commission on Aspectology of the International Committee of Slavists. Padova, Italy, September 30 – October 4, 2011
- T. Arkhangelskiy, V. Panov. On diachrony of the Greek aspect // Workshop on the typology of aspectual systems. St. Petersburg, Russia, September 16–18, 2010
- O. Karpova, T. Arkhangelskiy, M. Kyuseva, E. Rakhilina, T. Reznikova, D. Ryzhova, M. Tagabileva. База данных по многозначным качественным прилагательным и наречиям русского языка [Database of polysemous qualitative adjectives and adverbs in Russian] // Annual conference “Dialogue” 2010. Bekasovo, Russia, May 26–30, 2010
- T. Arkhangelskiy, O. Belyaev. A comparative study of Eastern Armenian and Iron Ossetic spatial systems // Current Advances in Caucasian Studies. Macerata, Italy, January 20–23, 2010