

# Niko Tapio Partanen

Haapaniemenkatu 11 B 75  
00530 Helsinki

Puhelinnumero: +358405068571

Sähköposti: [niko.partanen@helsinki.fi](mailto:niko.partanen@helsinki.fi)

Syntymäaika: 23. joulukuuta 1986 – Suomi

## Työpaikka

Tutkijakoulutettava, Helsingin yliopisto

## Erikoistumisala

Väitöskirjatutkimukseni käsittelee komisyrjäänin murteiden sosiolingvististä variaatiota. Työskentelen Koneen säätiön rahoittamassa tutkimushankkeessa 'Language Technology meets Language Technology: The Next Step in the Description of Komi'.

Avaintermi: Uralilaiset kielet, sosiolingvistiikka, tietokonelingvistiikka, arkistoaineistot, kielten dokumentoinnin metodologia, tutkimusetiikka

## Aiemmat työtehtävät

2018-2019	Erityisasiantuntija, Kotimaisten kielten keskus
2017-2018	Vieraileva tutkija, LATTICE-CNRS -laboratorio, Pariisi
2016-2017	Ohjelmistokehittäjä, INEL / Hamburger Zentrum für Sprachkorpora
2015	Ohjelmistokehittäjä, CLARIN / Hamburger Zentrum für Sprachkorpora
2014-2015	Työntekijä projektissa: Izhva Komi Language Documentation, Koneen säätiö
2013	Projektikoordinaattori hankkeessa Down River Vashka, Komi (Udoran murre), Koneen säätiö ja MinorEuRus
2012-2013	Tutkimusavustaja, Helsingin yliopisto

## Opinnot

2013	MA Suomalais-ugrilainen kielentutkimus, Helsingin yliopisto
2013	BA Suomalais-ugrilainen kielentutkimus, Helsingin yliopisto
syksy 2011	Vaihto-opiskelija, Syktyvkarin valtionyliopisto

## Palkinnot

2014 Gradupalkinto, Venäjän ja Itä-Euroopan tutkimuksen seura

## Valitut artikkelit ja opinnäytteet

- 2019 Rogier Blokland, Niko Partanen, Michael Rießler & Joshua Wilbur: Using computational approaches to integrate endangered language legacy data into documentation corpora: Past experiences and challenges ahead. Workshop on the Use of Computational Methods in the Study of Endangered Languages.  
URL: <https://journals.colorado.edu/index.php/computel/article/view/451/431>
- 2019 Jack Rueter & Niko Partanen: Survey of Uralic Universal Dependencies development. In: Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)  
URL: <https://www.aclweb.org/anthology/W19-8009.pdf>
- 2019 Niko Partanen, Mika Hämmäläinen & Khalid Alnajjar: Dialect Text Normalization to Normative Standard Finnish. The Fifth Workshop on Noisy User-generated Text (W-NUT 2019), s. 141–146 URL: <https://www.aclweb.org/anthology/D19-5519.pdf>
- 2019 Niko Partanen & Michael Rießler: An OCR system for the Unified Northern Alphabet. Proceedings of the Fifth Workshop on Computational Linguistics for Uralic Languages, s. 77–89 URL: <https://www.aclweb.org/anthology/W19-0307.pdf>
- 2018 Niko Partanen, Rogier Blokland, KyungTae Lim, Thierry Poibeau & Michael Rießler: The First Komi-Zyrian Universal Dependencies Treebanks. Universal Dependencies Workshop 2018 (2018 Conference on Empirical Methods in Natural Language Processing), 1.11.2018, Brysseli, Belgia URL: <http://www.diva-portal.org/smash/get/diva2:1313885/FULLTEXT01.pdf>
- 2018 Jeremy Bradley, Alexandra Kellner & Niko Partanen: Variation in word order in Permian and Mari varieties – a corpus-based investigation. Языковые контакты народов Поволжья и Урала. XI Международный симпозиум (Чебоксары, 21–24 мая 2018 г.) – Сборник статей, Г. Е. Корнилов, А. М. Иванова, Г. Н. Семенова (toim.), s. 238–244, Chuvash University Publishing House, Cheboksary
- 2018 KyungTae Lim, Niko Partanen & Thierry Poibeau: Multilingual Dependency Parsing for Low-Resource Languages: Case Studies on North Saami and Komi-Zyrian. Proceedings of the 11th Language Resources and Evaluation Conference, 7.–12.5.2018, Miyazaki, Japani URL: <https://hal.archives-ouvertes.fr/hal-01856178/document>
- 2018 Niko Partanen, KyungTae Lim, Michael Rießler & Thierry Poibeau: Dependency Parsing of Code-Switching Data with Cross-Lingual Feature Representations. Proceedings of the Fourth International Workshop on Computational Linguistics of Uralic Languages, s. 1–17 URL: <https://hal.archives-ouvertes.fr/hal-01722243/document>

- 2017 Niko Partanen: Challenges in OCR today. Report on experiences from INEL. Electronic Writing of RF Peoples: History, Issues, and Perspectives. 16.–17.3.2017, Syktyvkar  
URL: <https://www.elibrary.ru/item.asp?id=29055260>
- 2017 Ciprian Gerstenberger, Niko Partanen & Michael Rießler: Instant annotations in ELAN corpora of spoken and written Komi, an endangered language of the Barents Sea region. Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages, s. 57–66. URL: <https://www.aclweb.org/anthology/W17-0109.pdf>
- 2017 Ciprian Gerstenberger, Niko Partanen, Michael Rießler & Joshua Wilbur: Instant Annotations – Applying NLP Methods to the Annotation of Spoken Language Documentation Corpora. Proceedings of the Third Workshop on Computational Linguistics for Uralic Languages, s. 25–36. URL: <https://www.aclweb.org/anthology/W17-0109.pdf>
- 2016 Ciprian Gerstenberger, Niko Partanen, Michael Rießler & Joshua Wilbur: Utilizing language technology in the documentation of endangered Uralic languages. Northern European Journal of Language Technology, 4, 29–47  
URL: <https://journal.ep.liu.se/NEJLT/article/download/1660/1006>
- 2016 Partanen, Niko & Saarikivi, Janne: Fragmentation of Karelian language and its community: growing variation at the threshold of the language shift. Toivanen, Reetta & Saarikivi, Janne (toim.), New and Old Language Diversities, Multilingual Matters
- 2013 Partanen, Niko: Kahden suomalais-ugrilaisen yhteisön kielenvaihdon vertailevaa tarkastelua: Tuuksa Karjalassa ja Koigort Komissa, Julkaisematon pro gradu -tutkielma, Helsingin yliopisto
- 2013 Partanen, Niko: Kielenvaihdon tarkastelua Vuokkiniemellä ja Viteleellä: Vuosien 2004 ja 2010 kenttätutkimusten analyysi, Julkaisematon kandidaatintutkielma, Helsingin yliopisto

## Opetus

- 2019 Kieliteknologia ja tietosuoja -työpaja 13.8. (Giellagas-instituutti)
- 2018 ELAN-tiedostojen käyttö korpustutkimuksessa. 14.–15.5. (Giellagas-instituutti)
- 2017 Advanced analysis and manipulation of ELAN corpus data with R and Python, 16.–17.11. (Freiburg Institute for Advanced Studies)
- 2017 Using ELAN to create and analyse spoken corpus, 1.–8.6. (Syktyvkarin valtionyliopisto)
- 2015 ELAN-korpuksen käyttö tutkimuksessa, 19.–21.10. (Giellagas-instituutti)

## Luottamustehtävät

- 2020 Kirjavarainhoitaja, Suomalais-Ugrilainen Seura
- 2018–2020 Puheenjohtaja, Sukukansojen ystävät ry

## Kielitaito

Suomi: äidinkieli  
Komisyrjäni: erinomainen  
Englanti: erinomainen  
Italia: hyvä  
Venäjä: hyvä  
Udmurtti: hyvä keskustelutaito  
Komipermjakki: hyvä keskustelutaito  
Ruotsi: arkinen keskustelutaito

## IT-aidot

Hyvä osaaminen: R, Python, NLP-työkalut, LaTeX, XML, GIS-työkalut, OCR-työkalut, tietokannat, multimedian käsittely  
Aktiivisesti käytössä: puheen- ja tekstintunnistuksen työkalut

## Kenttätökokemus

2016	Kuolan niemimaa, Murmanskin alue, kaksi viikkoa Rahoitus: Koneen Säätiö
2015	Eri alueet Udmurtiassa, kymmenen päivää Ei ulkoista rahoitusta
2015	Izhman piiri, Komin tasavalta, viisi päivää Rahoitus: Развития фундаментальных лингвистических исследований
2015	Narjan-Mar, Nenetsian autonominen okrug, kaksi viikkoa Rahoitus: Koneen Säätiö
2014	Izhman piiri, Komin tasavalta, kaksi viikkoa Rahoitus: Koneen Säätiö
2013	Udoran piiri, Komin tasavalta, kuukausi Rahoitus: Koneen Säätiö ja MinorEuRus
2013	Aunuksen piiri, Karjalan tasavalta, viikko Rahoitus: Helsingin yliopisto
2012	Koigorodokin piiri, Komin tasavalta, kaksi viikkoa Rahoitus: MinorEuRus
2012	Aunuksen piiri, Karjalan tasavalta, kaksi viikkoa Rahoitus: Helsingin yliopisto
2011	Prääsän piiri, Karjalan tasavalta, kaksi viikkoa Rahoitus: Helsingin yliopisto
2010	Aunuksen piiri, Karjalan tasavalta, viikko Rahoitus: Helsingin yliopisto