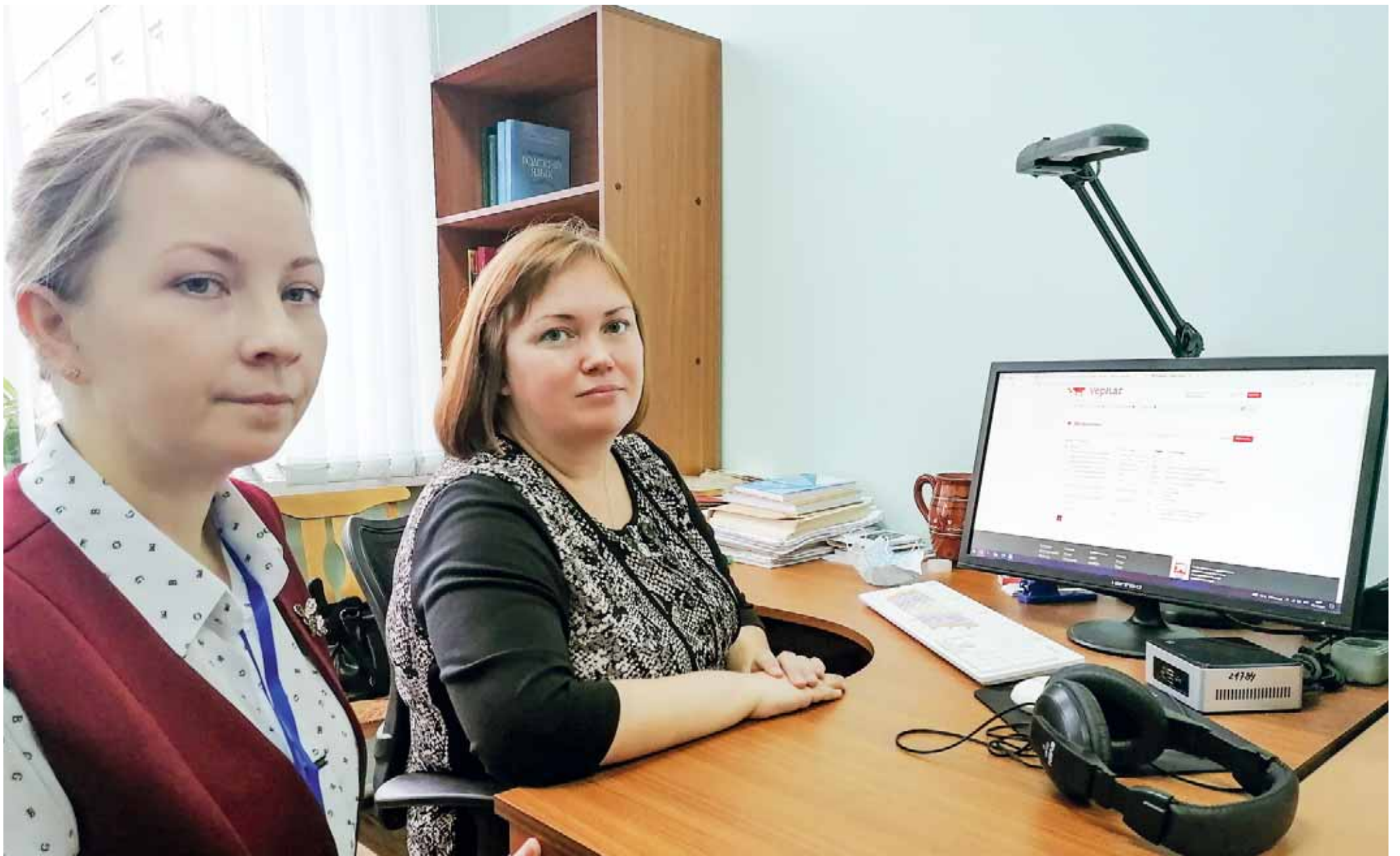


Kielet



Karjalan tiedekeskuksen vepsän ja karjalan kielen tutkijat Irina Novak (edessä) ja Aleksandra Rodionova ovat kiinnostuneita siitä, että VepKar-korpuksesta saisivat tietoa mahdollisimman monet vepsän ja karjalan opiskelijat. KUVA: MARINA TOLSTYH

Kielten mallit säilyvät korpuksessa

Karjalan tiedekeskuksen vepsän ja karjalan murteiden tutkijat keräävät korpukseseen erityyppisiä tekstejä ja tutkivat sanojen merkityksiä ja kielioppia.

Tutkimus

Marina Tolstyh

Karjalan Sanomat
Petroskoi

karjalansanomat@yandex.ru



Venäjäntiedeakatemian Karjalan tiedekeskuksen kielen, kirjallisuuden ja historian instituutin vepsän ja karjalan kielen tutkijat esittelivät vepsän ja karjalan kielen VepKar-korpuksen välituloksia Petroskoissa pidetyssä kansainvälisessä konferenssissa.

– Korpuksen päätarkoituksena on kerätä erityyppisiä vepsän- ja karjalankielisiä tekstejä, tutkia tekstisanoja, digitalisoida nämä tekstit ja esitellä ne käyttäjille sopivassa muodossa, Karjalan tiedekeskuksen kielentutkija **Irina Novak** sanoo.

Kielentutkijoiden käytössä on ollut paljon eri vuosikymmeninä tallennettuja puhemalleja vepsän ja karjalan kielillä. Kuitenkaan kielten opiskelijoilla ei ollut mahdollisuutta tutustua näihin puhemalleihin ennen korpuksen tekemistä.

– Käytännössä korpus tarkoittaa erityyppisten vepsän- ja karjalankielisten tekstien sähköistä kirjastoa, Novak täsmentää.

VepKar-korpus otti alkunsa vuonna 2009, jolloin vepsän kielen tutkijan **Nina Zaitsevan** aloitteesta ja johdolla perustettiin vepsän kielen korpus. Vuodesta 2016 alkaen korpus on sisältänyt myös varsinaiskarjalan, livvin-

karjalan ja lyydin murteiden kielenaineiksia. Tämä yhteinen kielellinen kenttä on saanut sitten julkisen vepsän ja karjalan kielen eli VepKar-korpuksen nimen.

Tekstit ja sanakirjat ovat yhdessä

VepKar-korpus sisältää runsaasti niin tekstejä kielillä sekä vepsän ja karjalan sanakirjoja kuin myös tekstien hakua ja sanamuotojen analysointia helpottavan tietokoneohjelman.

Nyt korpuksen teksteistä löytyy yhteensä yli kaksimiljoonaa sanaa, joiden merkitykset voi löytää korpuksen laitetuista sanakirjoista. Taas sanakirjojen avulla voi saada tietoa yli miljoonan sanan rakenteesta ja muista kieliopillisista ominaisuuksista.

– Korpus on tarkoitettu niin kielten säilyttämiseksi kuin myös kielten tutkimiseksi. Tietysti vepsän ja karjalan kielen opiskelijat voivat saada korpukselta paljon apua opinnoissaan. Kieliä voi oppia hyvin, kun lukee korpuksen tekstejä ja analysoi sitten tekstien sanoja, Novak havaitsee.

” VepKar-korpus tarkoittaa erityyppisten vepsän- ja karjalankielisten tekstien sähköistä kirjastoa.

IRINA NOVAK

Tällä hetkellä korpuksessa on esillä 3 138 tekstiä, jotka on saatu 350 tekstintekijältä ja 53 tekstejä keränneeltä tutkijalta. Esimerkiksi kansanperinnetekstien joukosta voi lukea 1918 vuodesta peräisin olevaa virsiä.

Korpus-projektiin on osallistunut Karjalan tiedekeskuksen seitsemän kielentutkijaa ja neljä IT-alan tutkijaa sekä Tverin valtion yliopiston karjalan kielen tutkija. IT-tutkijat auttavat ratkaisemaan korpuksen teknisiä ongelmia.

Edessä työ sovellusten kanssa

Lyydin murteen tutkija **Aleksandra Rodionova** on sitä mieltä, että täytyy järjestää korpuksen käyttäjien kysely. Se auttaa tutkijoita ymmärtämään, kuinka usein kielten opiskelijat käyttävät korpusta ja kuinka helppo korpuksen käyttö on.

– Korpus sisältää hyvin paljon tekstejä ja kielioppia. Olemme alkaneet ymmärtää, että korpus voi olla tavallisille kielenkäyttäjille vaikea. Nyt pitää miettiä, millä tavalla voimme helpottaa korpuksen käyttöä, Rodionova sanoo.

Kielentutkijat haluavat esittää VepKar-korpusta laajasti kaikkialla, jossa opetetaan ja opiskellaan karjalan ja vepsän kieltä.

Tietysti tutkijat haluavat kehittää korpusta eteenpäin. Esimerkiksi suunnitelmassa on laatia karjalan kielen murteiden ja vepsän kielen audiokartta.

– Laitamme korpukseseen uusia tekstejä myös jatkossa. Ajatuksissa on luoda kor-

VepKar-korpus

- Korpus sisältää yli 3000 tekstiä.
- Vepsänkielisiä aineistoa on kerätty vuodesta 2009.
- Tekstejä karjalan murteilla on lisätty viiden vuoden ajan.
- Korpuksen sanakirjoista löytyy tietoa yli kahdestamiljoonasta sanasta.
- Projektiin osallistuu 12 kielentutkijaa ja IT-alan asiantuntijaa.
- Korpus on Karjalan tiedekeskuksen tuote.

puksen pohjalta erilaisia karjalan- ja vepsänkielisiä sovelluksia tavallisille käyttäjille, Rodionova kertoo.

Tänä vuonna korpuksen pohjalta on tehty digitaalinen sanakirja Sahahelmi, jota voi käyttää Android-puhelimessa.

Korpuksen sivuilta löytyy tietoa niistä laetoksista ja fyysisistä henkilöistä, joiden tekstit on otettu tekijänoikeussopimuksen nojalla mukaan kielenkorpukseseen. Osa karjalan- ja vepsänkielisistä teksteistä on peräisin Periodika-kustantamon lehdistä.

– Artikkeleita lyydin murteella on kirjoitettu hyvin vähän. Tämän takia olen alkanut itse kirjoittaa kansanperinteistä kertovia juttuja Oma Mua -sanomalehteen. Ne ovat esillä myös korpuksessa, Rodionova esittää. **KS**