

Koneen
Säätio

Kieliohjelma
2012–2016

Koneen Säätiö
Kieliohjelma 2012–2016

I.2.2012

GRAAFINEN SUUNNITTELU
Patrik Söderlund

KIRJAPAINO
Mynäprint, Mynämäki 2012

Koneen Säätiön kieliohjelma 2012–2016

SISÄLLYS

Kieliohjelman taustaa	5
Suomen kieli ja Itämeren alueen pienet kielet Koneen Säätiön painopisteenä	5
Kieliohjelmalla halutaan edistää monikielisyttä	6
Eri toimijat suomalais-ugrilaisia vähemmistökansoja ja -kieliä tukemassa	7
Tehtävä	12
Visio	12
Tavoitteena pienten suomalais-ugrialaisten kielten, suomen sekä Suomen vähemmistökielten dokumentointi	14
Aineistojen saatavuuden, saavutettavuuden ja käytettävyyden parantaminen	15
Avoimuuden ja vuorovaikutuksen kulttuurin edistäminen muun muassa kansalaistieteen keinoin	17
■ Erialaisten kielenkäyttäjien huomioon ottaminen	17
Monitieteisyyden ja uudenlaisten rajojen ylitysten edistäminen	19
Miten tavoitteita toteutetaan?	20
Toimenpiteet 2012–2016	22
Hankkeiden välinen yhteydenpito ja yhteistyö	22
Vuositain toistuvien apurahahakujen kautta tuettavat hankkeet, suunnatut haut ja säätiön käynnistämät hankkeet	22
■ Esimerkki kieliteknologisten sovellusten käyttämisestä ja kehittämisestä	27
Ohjelman seuranta	30

Koneen Säätiön ovat perustaneet vuorineuvos Heikki H. Herlin ja ekonomi Pekka Herlin vuonna 1956. Perustajat toimivat samanaikaisesti sekä säätiön että Kone Osakeyhtiön johdossa, joten kahden organisaation yhteys oli läheinen. Nykyisin Koneen Säätiö on itsenäinen ja riippumaton yleishyödyllinen organisaatio, joka edistää tieteellistä tutkimustyötä, erityisesti humanistista, yhteiskuntatieteellistä ja ympäristöntutkimusta, sekä taidetta ja kulttuuria Suomessa.

Koneen Säätiön tavoitteena on olla tiede- ja taidepoliittisesti valpas suomalaisen tieteen ja taiteen rahoittaja, joka myös aktiivisesti etsii ja tarjoaa vaihtoehtoja valtavirralle. Tehtävänsä toteuttamiseksi säätiö muun muassa myöntää apurahoja, ylläpitää Saaren kartanon residenssiä ja järjestää toimialaansa liittyviä seminaareja ja tilaisuuksia.

Kieliohjelman taustaa

SUOMEN KIELI JA ITÄMEREN ALUEEN PIENET KIELET KONEEN SÄÄTIÖN PAINOPISTEENÄ

Koneen Säätiö on strategiassaan määritellyt yhdeksi toimintansa painopisteeksi suomen kielen ja Suomessa puhuttavien vähemmistökielten sekä Itämeren alueella puhuttavien pienten kielten aseman vahvistamisen. Varsinkin suomen kielen osalta tämä otetaan huomioon kaikessa toiminnassa. Suomen kielen aseman vahvistamiseen tähtäävät erityisesti suomenkielisen tietokirjoittamisen tukeminen ja *Vuoden Tiedekynä* -palkinnon jakaminen. Suomessa puhuttavien vähemmistökielten sekä Itämeren alueen pienten kielten asemaan kiinnitetään huomiota myös esimerkiksi tuettavista taide- ja kulttuurihankkeista päätettäessä. Lisäksi Saaren kartanoon rekrytoidaan pienillä vähemmistökielillä kirjoittavia kirjailijoita ja kääntäjiä.

Säätiö haluaa nyt erikseen laaditun kieliohjelman kautta entistä tehokkaammin vahvistaa etenkin suomalais-ugrialaisten kielten asemaa. Monien kielten puhujamäärät ovat pienentyneet ja useat kielet ovat jopa uhanalaisia. Syinä ovat muun muassa kieliyhteisön hajoaminen esimerkiksi kaupungistumisen vuoksi ja heikot kielelliset perusoikeudet.

Globaalistuminen on myös yksi syy monen kielen aseman heikkeneemiseen. Esimerkkinä tästä on suomen kielen käytön väheneminen tieteen kielenä. Ajattelun ylläpitämisen ja kehittämisen kannalta on tärkeää, että tieteessä tutkituista asioista voidaan keskustella myös suomeksi ja että eri tieteenalojen peruskäsitteet ja termistöt ovat olemassa omalla äidinkielellä. Vaikka suomen kielen asema on heikentynyt tieteen kentällä, on suomi kuitenkin käyttökelpoinen ja toimiva kieli kaikilla elämänalueilla. Tilanne on toinen monen pienen suomalais-ugrilaisen kielen kohdalla.

Suomen kielen asema¹ Suomessa on maailman mittakaavassakin vahva ja valtiolta huolehtii kielestä. Julkisen sektorin rahoituskriisillä on kuitenkin vaikutuksia myös suomen asemaan. Esimerkiksi valtion tuottavuusohjelman mukanaan tuomat leikkaukset vaikuttavat Kotimaisten kielten keskuksen² eli Kotuksen toimintaan ja julkisen sektorin rahoitusongelmat

1. Suomen kielen asemasta on kirjoitettu laajasti Kotuksen *Suomen kielen tulevaisuus – Kielipoliittinen toimintaohjelma* -julkaisussa (2009), jossa tehdään ehdotuksia suomen monipuolisen ja laajalajalaisen käytön turvaamiseksi myös tulevaisuudessa.

2. 31.12.2011 asti Kotimaisten kielten tutkimuskeskus.

heijastuvat myös äidinkielen opetuksen asemaan uudessa tuntijaossa sekä suomen kielen asemaan monissa valtion laitoksissa ja yliopistoissa.

Kieli on tärkeä maailman hahmottamisen väline ja identiteetin osa ja vahvistaja. Kielet ja maailman kielellinen moninaisuus ovat Koneen Säätiön näkökulmasta itseisarvoja, joita on syytä tukea. Kielten tutkimus, dokumentointi ja tukeminen auttavat yhteiskunnan, kulttuurin ja maailman kuvan ymmärtämistä, ja tuota tietoa voidaan myös hyödyntää eri tavoin.

Vuonna 2011 Koneen Säätiö palkkasi FT Jack Rueterin puoleksi vuodeksi valmistelemaan kieliohjelmaa ja keräämään tietoa sitä varten.

KIELIOHJELMALLA HALUTAAN EDISTÄÄ MONIKIELISYYTTÄ

Monikielisyydellä tarkoitetaan äidinkielen tasoista kielitaitoa useammassa kielessä; joskus se ymmärretään myös laajemmin kykyä kommunikoida sujuvasti useammalla kielellä. On enemmän sääntö kuin poikkeus, että ihminen käyttää aktiivisesti useampaa kieltä. Maailman lähes seitsemän miljardia ihmistä puhuu lähes 7000 kieltä. Myös Itämeren piiri, johon Suomi kuuluu, on monikielinen alue. Alueella puhutaan niin suomalais-ugrilaisia, germaanisista ja slaavilaisia kuin muita kauempaa tulleita tai vasta hiljakkoin kehittyneitä kielimuotoja. Suomessa useimmat ruotsia äidinkielenään puhuvat osaavat myös suomea äidinkieltä vastaavalla tasolla. Myös maahanmuuttajat ovat suuri monikielinen ryhmä nyky-Suomessa.

Toiset kielet ovat lakisääteisiä valtakielinä omissa maissaan, ja toisilta puuttuu virallinen yhteiskuntaa ylläpitävä asema. Koska valtakielen osaaminen on useimmiten edellytyksenä muun muassa opiskelu- ja työpaikan saamiseksi, jäävät vähemmistökielet usein alakynteen. Etenkin kaupunkeihin muuttavilla nuorilla on vaarana luopua äidinkielestään, mikäli sen käytöllä ei näy olevan perspektiivejä uudessa ympäristössä.

Yhä nykyäänkin on tavallista, että ihmisillä on vääriä käsityksiä useamman kielen taitamisesta. Vanhemmat saattavat lakata puhumasta omaa äidinkieltään lapsilleen siinä luulossa, että kahden kielen osaaminen häiritsee lapsen kielellistä kehitystä. Tällaiset väärinkäsitykset johtavat vähemmistökielten puhujamäärien vähenemiseen ja kielen aseman heikkenemiseen.

Vähemmistökielet edustavat omien yhteisöjensä, kulttuuriensa ja perinteidensä tärkeää voimavaraa ja maailmankuvaa, ja kulttuuri, perinteet ja

maailmankatsomus välittyvät parhaiten äidinkielen vapaassa kielenkäytössä. Kieliohjelmassa halutaan edistää ymmärrystä siitä, että on mahdollista elää käyttäen monia kieliä ilman että siitä on haittaa yksilölle. Päinvastoin: monikielisydestä on hyötyä. Kun ihminen osaa useita kieliä, hänelle hahmottuu tietoisuus maailman tulkitsemisen moninaisuudesta, ja tuo tietoisuus voi rikastaa koko yhteiskuntaa.

ERI TOIMIJAT SUOMALAIS-UGRILAISIA VÄHEMMISTÖKANSOJA JA -KIELIÄ TUKESSA

Monet sekä julkisen että kolmannen sektorin organisaatiot Suomessa tukevat pieniä suomalais-ugrilaisten kansojen ja kielten. Säätiön kieliohjelman tarkoituksena on löytää oma lokeronsa näiden toimijoiden joukosta ja välttää päällekkäisyyksiä toiminnassa. Tarkoituksena on kuitenkin tehdä yhteistyötä samalla kentällä työskentelevien kanssa ja tarvittaessa tukea niiden toimintaa.

Suomalais-ugrilaisten vähemmistökansojen ja -kielten tutkitaan Helsingin, Turun, Oulun ja Itä-Suomen (Joensuu) yliopistoissa. Lisäksi kansojen ja kielten tuetaan ja/tai tutkitaan merkittävästi ainakin seuraavissa organisaatioissa:

- »» **Kotimaisten kielten keskus Kotus** tutkii ja huoltaa suomea, ruotsia, saamea, suomalaista viittomakieltä ja romanikieltä. Kotuksen osuus suomalais-ugrilaisten vähemmistökielten tutkimuksessa ja sanakirjatyössä on sen sijaan nykyään vähäinen. Suurin osa Kotuksen tutkimustuloksista julkaistaan sanakirjoina. On huolestuttavaa, että resurssien puutteessa Vanhan kirjasuomen sanakirjan teko päätettiin lakkauttaa ja Suomen murteiden sanakirjan rahoitus on vähentynyt. Uusia sanakirjahankkeita Kotuksessa edustaa maahanmuuttajakielten sanakirjahanke.
- »» Kansainvälisen liikkuvuuden ja yhteistyön keskuksessa (CIMO) toimiva **Ulkomaisten yliopistojen Suomen kielen ja kulttuurin opintojen asiantuntijaryhmä UKAN** lähettää sukukansojen yliopistoihin kieliassistentteja opettamaan suomea ja suomalaisten yliopistojen opettajia noin kahden viikon ajaksi luennoimaan. Opetukseen sisältyy kielen lisäksi suomalaisen kulttuurin opetusta. Lisäksi UKAN järjestää kaunokirjallisuuden käännoseminareja.

»» **M. A. Castrénin seuran** tarkoituksena on luoda yhteyksiä ja edistää vuorovaikutusta Suomen sekä Venäjällä asuvien uralilaisia kieliä puhuvien kansojen välillä, tehdä tunnetuksi näiden kansojen kulttuuria Suomessa sekä tukea niiden kulttuuripyrkimyksiä ja oman kielen viljelyä. Seura tukee muun muassa kirjallisuuden kääntämistä.

»» Vuodesta 1994 lähtien M. A. Castrénin seura on toiminut opetusministeriön rahoittamaa **sukukansaohjelmaa** koordinoivan neuvottelukunnan koollekutsujana ja ohjelman vastuullisena toteuttajana. Ohjelman pääkohtia ovat opetusalan yhteistyö, henkilövaihto, tutkimusyhteistyö, museo- ja kirjastoalan yhteistyö sekä kulttuurivaihto. Sukukansaohjelmasta rahoitetaan muun muassa tutkijoiden vierailuja Suomeen CIMO-stipendiaatteina. Sukukansaohjelmaa ollaan uudistamassa lähiaikoina.

»» **Suomen Kulttuurirahasto** tuki 2007–2009 yhteensä yli 400 000 eurolla Venäjällä asuvia kielisukulaisia kielen ja kulttuurin säilyttämisessä. Hankkeet sisälsivät muun muassa vepsän ja udmurtin kielileirien järjestämistä, museotyön tukemista ja moniammatillista julkaisukoulutusta. Syksyllä 2006 käynnistynyt suuri rahoitushanke sai vuonna 2008 jatkoa yhteensä yli miljoonalla eurolla. Hankkeissa tuettiin muun muassa kielipesätoimintaa, kansalaisjärjestötoimintaan kouluttamista ja Haikolan kylän elvytystä. Kulttuurirahasto kutsuu koolle rahoittamiensa hankkeiden ja alan muita toimijoita pari kertaa vuodessa. Myös Koneen Säätiö on osallistunut näihin tilaisuuksiin kevästä 2011 lähtien.

»» **Suomi-Venäjä Seura** kehittää suomalais-venäläisiä kulttuuri- ja kansalaisyhteyksiä ja osallistuu suomalais-ugrilaiseen toimintaan esimerkiksi vastaamalla suomalais-ugrialaisten kansojen maailmankongressin valmisteluista Suomessa. Seura on toteuttanut Kulttuurirahaston ja ulkoasiainministeriön rahoituksella muun muassa kansalaisjärjestökoulutusta ja kielipesätoimintaa sekä ulkoasiainministeriön lähialueyhteistyön rahoituksella nuorison suomalais-ugrialaista kulttuuriyhteistyöhanketta Karjalan tasavallassa.

»» **Suomalais-Ugrilainen Seura** edistää uralilaisten ja altailaisten kielten ja näitä kieliä puhuvien kansojen kulttuurien, kansatieteen, kansanrunouden ja esihistorian tutkimusta julkaisutoiminnan ja apurahojen avulla.

»» **Sukukansojen ystävät ry** levittää sukukansatietoutta Suomessa, järjestää tapahtumia ja tilaisuuksia ja luo yhteyksiä suomalaisten ja sukulielisten alueiden kansalaisjärjestöjen ja ruohonjuuritason toimijoiden välille. Järjestö on myös Suomalais-ugrialaisten kansojen nuorten liiton (MAFUN) edustaja Suomessa.

»» **Helsingin yliopiston ylioppilaskunnan sukukansavaliokunta** tukee sukukansojen pyrkimyksiä vaalia omaa kieltään ja kulttuuriaan. Valiokunta suuntaa keräämiään varoja suomalais-ugrialaisten vähemmistöjen hankkeille. Keräysvaroista on tuettu useita kymmeniä hankkeita, joiden suunnittelemisessa ja toteuttamisessa erityisesti nuoret ovat olleet aktiivisia. Avustuksia on myönnetty varsinkin julkaisutoimintaan, yhteiskunnallisten ja kulttuuritapahtumien järjestämiseen sekä laitehankintoihin.

»» **Saamelaiskäräjät** on saamelaisten itsehallintoelin. Saamelaiskäräjien kielineuvosto muun muassa ylläpitää ja kehittää Suomessa käytettäviä saamen kieliä, antaa ohjeita ja suosituksia kielen- ja nimistönhuoltoon ja terminologiaan liittyvissä asioissa, edistää kielen tutkimusta, jakaa tietoa saamen kielistä ja osallistuu pohjoismaiseen kieliyhteistyöhön. Saamelaiskäräjät myös jakaa avustuksia saamelaistaiteen ja -kulttuurin eri aloille sekä saamelaisjärjestöille ja vastaa oppimateriaalityöstä.

»» **Suomen Akatemia** suunnittelee yhteistyöhanketta Unkarin Tiedeakatemian kanssa liittyen suomalais-ugrialaisten kielten ja kulttuurin tutkimukseen.

Lisäksi Koneen Säätiön kieliohjelman kannalta ovat tärkeitä erilaiset aineistojen digitointihankkeet, kieliteknologiset sovellukset ja muutamat muut kieleen liittyvät hankkeet. Tällaisia ovat muun muassa seuraavat:

»» **Tieteen kansallinen termipankki** (The Bank of Finnish Terminology in Arts and Sciences, BFT) on Suomen Akatemian rahoittama Helsingin yliopistossa toteutettava hanke. Helsingin yliopiston, Kotuksen, Tieteellisten seurain valtuuskunnan ja Sanastokeskus TSK:n yhteistyöhankkeessa tavoitteena on luoda perusta Suomessa harjoitettavien tieteenalojen yhteiselle monikieliselle termipankille, joka olisi vapaasti käytettävissä.

»» **Kansallinen digitaalinen kirjasto** (KDK) on opetusministeriön käynnistämä hanke, jonka tavoitteena on kirjastojen, arkistojen ja museoiden keskeisten kansallisten tietovarantojen saatavuuden ja käytettävyyden edistäminen sekä tietoverkoissa ja digitaalisessa muodossa olevien kulttuuriperintöaineistojen pitkäaikaissäilytysratkaisujen kehittäminen.

»» **Kirjallisuuspankki** on suunnitelma suomalaisen – suomen- ja ruotsinkielisen – tekijänoikeusvapaan kaunokirjallisuuden digitaalisen verkopalvelun perustamisesta kansainvälisten esikuvien tapaan. Palvelun pohjana ovat jo digitoidut ja oikoluettut tekstit. Kirjallisuuspankkia voidaan hyödyntää niin tutkimuksessa kuin kielen opetuksessa (esimerkiksi UKANin suomen opettajat ulkomailla).

»» **FIN-CLARIN** on suomalainen osapuoli **ESFRI**-projektissa, joka suunnittelee Euroopan laajuista kieliaineistojen ja -teknologian infrastruktuuria.

»» **Giellatekno** soveltaa avoimeen lähdekoodiin perustuvia kieliteknologisia menetelmiä saamenkielisen digitaalisen infrastruktuurin toteuttamiseen. Tämä tarkoittaa esimerkiksi eri saamelaiskielille ja pohjoisen alueen kielille rakennettavia oikeinkirjoitus-, tekstianalyysi- ja kielenoppimishohjelmia sekä digitaalisia sanakirjoja ja tekstistä ääneen -ohjelmia. Käytännössä ensisijaisia ovat saamelaiskielet ja kveeni, mutta Norjan kuuluminen Barentsin alueeseen mahdollistaa muidenkin kielten avoimen dokumentoinnin ja sovelluskäytänteiden vuorovaikutteisen hyödyntämisen.

»» **HFST** (Helsinki Finite-State Transducer Technology 'Helsingin äärellistilainen transduktoriteknologia') tuottaa avoimeen lähdekoodiin perustuvia automaatteja luonnollisten kielten morfologista tunnistamista varten. Automaatin rakentamiseen kuuluu olennaisina osina kuvaitavan kielimuodon paradigmaattisen morfologian kuvaus hakusana- muotoineen ja taivutusluokkineen. Käytännössä nämä hakusanalistat taivutustietoineen ovat samoja, joilla rakennetaan sanakirjoja. Samaa automaattityötä voidaan hyödyntää minimaalisin muunnoksia oikeinkirjoitussovelluksissa, tekstintunnistamisohjelmissa ja morfologisissa jäsentimisissä.

»» **META-NORD** -projekti pyrkii aikaansaamaan avoimen kielitiedeinfrastruktuurin Baltiassa ja Pohjoismaissa.

»» **Kielipankki** on CSC:n (opetus- ja kulttuuriministeriön hallinnoima tieteen tietotekniikan keskus) tarjoama kielentutkimuksen palvelu, joka koostuu kattavasta joukosta teksti- ja puheaineistoja sekä niiden tutkimiseen soveltuvia ohjelmistoja tehokkaassa laiteympäristössä.

»» **Suomalais-ugrilaiset kirjastot Venäjällä** -portaali on tietokanta, jossa on linkkejä Venäjän suomalais-ugrialaisten kirjastojen sivustoille. Lisäksi kirjastojen tavoitteena on luoda elektroninen kokoelma, johon kuuluu julkaisuja suomalais-ugrilaisilla kielillä. Venäjän suomalais-ugrilaiset kirjastot toteuttavat projektia Venäjän ja Itä-Euroopan instituutin, Helsingin yliopiston slaavilaisen kirjaston ja M. A. Castrénin seuran avulla.

Tehtävä

Koneen Säätiön kieliohjelman tarkoituksena on pienten suomalais-ugrialaisten kielten, suomen sekä Suomen vähemmistökielten dokumentointi ja niiden aseman vahvistaminen.

Visio

Koneen Säätiön kieliohjelman kautta tuotetaan tieteellisesti relevanttia tutkimustietoa ja dokumentointia pienistä suomalais-ugrilaisista kielistä, suomesta ja Suomen vähemmistökielistä. Sekä tiedeyhteisö että kielenkäyttäjät hyödyntävät tutkimustietoa ja dokumentoinnin tuloksia, ja vuorovaikutukseen perustuvat tutkimuskäytänteet vahvistuvat.

Suomen vähemmistökielet

ruotsi*, suomen- ja ruotsinkielinen viittomakieli, pohjoissaame, inarinsaame, koltansaame, karjala, romanikielet, tataari ja maahanmuuttajien kielet.

* Toisena Suomen kansalliskielenä ruotsilla on Suomessa vahva asema. Lisäksi ruotsi on valtakieli Ruotsissa, jossa sen tutkimuksella on vahva asema. Muut Suomen vähemmistökielet ovat kieliohjelmassa tärkeämmässä asemassa.

Suomalais-ugrilaiset kielet

- a) itämerensuomalaiset: suomi*, viro*, liivi, võru, seto, vatja, inkeroinen, inkeri, vepsä, lyydi, livvi (aunuksenkarjala), eteläkarjala, vienankarjala, kveeni ja meänkieli
- b) saamelaiskielet: eteläsaame, piitimensaame, luulajansaame, uumajansaame, pohjoissaame, inarinsaame, koltansaame, kildininsaame ja turjansaame
- c) mordvalaiskielet: ersä ja mokša
- d) marilaiskielet: vuorimari [länsimari] ja niittymari [itämari]
- e) permiläiskielet: udmurtti, komipermjakki ja (syrjäänin) komi
- f) ugrilaiset kielet: hanti, mansi ja unkari*
- g) samojedikielet: nganasani, nenetsi, enetsi, selkuppi, matori** ja kamassi**

* Suomen lisäksi viron ja unkarin kielet ovat valtakielinä valtioissaan ja niiden tutkimuksella on huomattavasti vahvempi asema kuin suomalais-ugrilaisilla vähemmistökielillä. Koneen Säätiön kieliohjelmassa keskitytään suomen lisäksi pieniin suomalais-ugrilaisiin kieliin.

** Matoria ja kamassia voidaan pitää jo kuolleina kielinä, mutta se ei vähennä niiden kiinnostavuutta kielitieteellisen tutkimuksen näkökulmasta.

Tavoitteena pienten suomalais-ugrialaisten kielten, suomen sekä Suomen vähemmistökielten dokumentointi

Koneen Säätiön kieliohjelman päätavoitteena on edistää pienten suomalais-ugrialaisten kielten, suomen sekä Suomen vähemmistökielten dokumentointia.

Kielen dokumentointi tarkoittaa puhutun, viitotun tai kirjoitetun kielen ilmiöiden havaitsemista, tallentamista ja tieteellistä analyysiä. Dokumentoinnissa hyödynnetään tallennettuja aineistoja ja tehdään niistä havaintoja. Dokumentointi on viime vuosikymmeninä ollut murroksessa, kun yhä enemmän on siirrytty käyttämään sähköisiä aineistoja ja kieliteknologisia sovelluksia.

Dokumentoinnin piiriin kuuluu niin kenttä- ja arkistotyö tallenteiden kartuttamisineen ja ylläpitämisineen kuin tallenteiden pohjalta tehtävä perus- ja jatkotutkimus. Tallenteisiin perustuviin tutkimuksen osa-alueisiin kuuluvat esimerkiksi sanasto-, muoto-oppi-, lauseoppi- sekä diskurssitutkimus. Perusteellisimpiin dokumentointilajeihin kuuluu sanakirjatyö. Laajasti katsottuna dokumentoinnilla voidaan tarkoittaa myös kielen tutkimista kielisosiologisesta näkökulmasta, jolloin etsitään vastauksia muun muassa siihen, millaisessa tilanteessa kielet ovat ja millaisten kielellisten identiteettien avulla kielet elävät.

Koska erilaiset aineistot ovat tärkeä osa kielten dokumentointia, kiinnitetään kieliohjelmassa aineistojen kartoittamiseen, kartuttamiseen ja ylläpitoon erityistä huomiota. Aineistojen hyödyntämisen edellytyksenä on myös niiden jalostaminen sekä saatavuuden ja käytettävyyden edistäminen: on tärkeää, että olennaisten aineistojen löytämisen lisäksi niitä on mahdollista hyödyntää ajanmukaisilla tutkimusvälineillä. (Ks. s. 15, *Aineistojen saatavuuden, saavutettavuuden ja käytettävyyden parantaminen.*)

Sekä vanhoja että uusia kieliaineistoja on tärkeää saada tiedeyhteisön ja yhteiskunnan avoimeen ja vuorovaikutteeseen käyttöön. Kieliohjelman avulla pyritään edistämään kielenkäyttäjien ja tutkijoiden avointa vuorovaikutusta dokumentointiprosessin eri vaiheissa esimerkiksi kansalaistieteen keinoja hyödyntäen. Kielenkäyttäjillä on oltava mahdollisuudet saada käsiinsä tietoa omasta kielestä: on tärkeää, että tutkimuksessa saatu tieto

palautuu kieliyhteisölle. (Ks. s. 17, *Avoimuuden ja vuorovaikutuksen kulttuurin edistäminen muun muassa kansalaistieteen keinoin.*)

Tieteidenvälisyyden edistäminen on ollut Koneen Säätiön tavoitteena jo pitkään; viime aikoina on keskusteltu myös tieteen ja taiteen näkökulmien yhdistämisestä. Myös kieliohjelmassa tuettavia hankkeita kannustetaan monitieteisyyteen esimerkiksi erilaisten aineistojen hyödyntämisen kautta. Samoin taiteen ja tieteen välistä vuorovaikutusta edistetään. (Ks. s. 19, *Monitieteisyyden ja uudenlaisten rajojen ylitysten edistäminen.*)

AINEISTOJEN SAATAVUUDEN, SAAVUTETTAVUUDEN JA KÄYTETTÄVYYDEN PARANTAMINEN

Kieliohjelman tavoitteena on saattaa sekä vanhoja että uusia kieliaineistoja tiedeyhteisön ja muun yhteiskunnan avoimeen käyttöön. Olemassa olevia aineistoja on kartoitettava, ja vanhojen vahakiekkujen, nauhojen, käsikirjoitusten ja painotuotteiden tuhoutumisen varalta on ryhdyttävä pelastustoimiin. Tämä tarkoittaa useimmiten vanhojen aineistojen mikrofilmaamisen ja digitoinnin edistämistä. Uusien aineistojen tallentamisessa pyritään käyttämään mieluiten avointa lähdekoodia ja nykyteknologiaa hyödyntäviä audio-video-laitteita.

Muun muassa Pietarissa sijaitsevassa Venäjän kansalliskirjastossa on runsaasti suomalais-ugrilaisilla pienillä kielillä kirjoitettuja aineistoja, kuten Neuvostoliiton alkuvuosina tehtyjä oppikirjoja, valistuskirjallisuutta, kaunokirjallisuutta, lehtiä ja aikakauskirjoja. Kieliohjelman tavoitteena on saada esimerkiksi näitä aineistoja avoimeen käyttöön. Myös Suomessa on paljon digitoimattomia kieliaineistoja, kuten vähemmistökielten vahakiekkoja ja tallennusohjelmistokehityksessä jälkeenjääneitä tallenteita (esimerkiksi Kuurojen liiton viittomakielten haastattelunauhoja).

Parin viime vuosisadan saatossa on kerätty paljon tietoa, joka tulisi saatua suuremman yhteisön ulottuville ja jonka voisi palauttaa oman kansan tietoisuuteen. Etenkin 1800- ja 1900-lukujen kansallisherätyksen ja valistuksen kaudella kerättiin paljon omaa kansaa ja kieltä kuvaavaa materiaalia. Suomalaiset tieteilijät, tutkimusmatkailijat ja maallikot keräsivät ja tutkivat runsaasti oman kielen lisäksi myös sukukieliä ja -kansoja. Kenttätyöllä hankittujen haastattelu- ja muiden kielenkäyttöaineistojen kartuttaminen

ja saatavuuden, saavutettavuuden sekä käytettävyyden varmistaminen on tärkeää. Näitä suuria aineistoja tulisi palauttaa kieliyhteisöjen hyödynnettäväksi ja vapaasti luettaviksi, saataviksi ja haettaviksi esimerkiksi kirjastojen yhteistyön avulla ja siten edistää tutkimuksen lisäksi muun muassa kielitietoisuutta ja kulttuurin ymmärrystä.

Jotta tallennetut aineistot olisivat käyttäjien löydettävissä, täytyy niistä tehdä riittäviä kuvailu- eli metadatatiedostoja hakutietokantoihin. Tällaiset kuvailutiedostot olisi hyvä tehdä heti aineistojen hankintavaiheessa ennen käytettävyyden lisäämistä.

Kieliohjelman tavoitteena on, että aineistot eivät olisi vain saatavilla vaan että erilaisten käyttäjien olisi helppo lähestyä ja käyttää niitä. Kieliaineistojen käytettävyys edellyttää kieliaineuksen tunnistavaa digitointia ja saattamista sellaiseen muotoon, jota voidaan hyödyntää yleisesti ja vapaasti saatavilla sovelluksilla. Aineistojen sisältämän kielellisen aineksen ymmärtäminen edellyttää sähköisten sanakirjojen ja koneellisten kääntäjien käyttöä. Tekstin tunnistamista edistää sanakirja- ja kielioppitiedon ja täydellisen morfologia-kuvauksen saatavuus. Aineiston ja sen sisältöjen laajempi tunnistaminen ja laajempien kuvailutiedostojen laatiminen kuuluu käytettävyyden edistämisen seuraavaan vaiheeseen, jossa voidaan hyödyntää muun muassa kansalaistiedettä automatisoidun tunnistamisen ja kuvailujen tarkistamisessa (ks. s. 17, *Avoimuuden ja vuorovaikutuksen kulttuurin edistäminen muun muassa kansalaistieteen keinoin*).

Myös aineistoinfrastruktuuriin on syytä kiinnittää huomiota. Mitä suurempi osa eri tahoilla koottavista aineistoista päätyy samaan ympäristöön, sitä suuremman merkityksen jokainen yksittäinen aineisto saa ison kokonaisuuden osana. Yhteisillä ratkaisuilla, verkostoitumisella ja eri aineistoille sopivien teknisten alustojen rakentamisella ja vakiinnuttamisella tehostetaan aineistonkeruu- ja jalostustyön päätymistä tiedeyhteisön ja muiden käyttäjien hyödyksi. Jatkossa on tärkeää rakentaa dynaamisia järjestelmiä, joita tutkijat ja muut käyttäjät voivat jatkuvasti täydentää ja rikastaa esimerkiksi jatkokehittämällä avoimeen lähdekoodiin perustuvia järjestelmiä.

Tekijänoikeudet suojaavat useimpia aineistoja. Yleensä tekijänoikeudet ovat voimassa 70 vuotta tekijän kuolemasta. Tämän vuoksi monien aineistojen kohdalla avoimen käytön toteuttaminen voi olla monimutkaista tai

Erilaisten kielenkäyttäjien huomioon ottaminen

Dokumentoinnissa ja kieliyhteisön osallistamisessa on otettava huomioon kielenkäyttäjien kirjo. Rivikielenkäyttäjien lisäksi on ammattimaisia kielenkäyttäjiä, kuten opettajat, kirjailijat, runoilijat, toimittajat sekä ohjelmien juontajat ja selostajat. Kaikilla kielenkäyttäjillä on käsitys käyttämästään kielestä ja omasta roolistaan kielenkäyttäjinä. Usein kielen ammattilaisilla on normittajan tehtävä. Kielissä, joissa ei ole olemassa yhtenäistä kirjakieltä, tällaisten normittajien rooli voi olla suurikin. Hankalaksi asia käy, jos normittajien näkemykset kirjakielen muodosta ovat erilaisia. Epäyhtenäinen kirjakieli antaa tilaa vaihtelulle, mikä voi olla niin hyvää kuin pahasta. Hyvää siinä on se, että kielenkäyttäjillä on enemmän vapauksia ilmaisukeinoissa; huonoina puolina voidaan mainita käsitteiden monimerkityksisyys ja kielen ytimekkyyden kärsiminen.

Yhtenäinen käyttökieli syntyy useamman kielenkäyttäjryhmän yhteistyöstä. Normina sitä arvostetaan, kun se tuntuu syntyvän terveille pohjalle ilman erityistä pakottamista. Sen aikaansaamiseksi tarvitaan siis pitkäaikaista yhteistyötä, jossa eri osapuolten arvokkuutta ei loukata ylemmydentunnolla. Myös asiantuntijuutta on arvostettava. Erilaiset muutokset kielessä tapahtuvat parhaimmillaan vähitellen ja ilman suuria mullistuksia, jolloin niiden omaksuminen on helpompaa eivätkä kielenkäyttäjät vieraannu omasta kielestään.

mahdotontakin. Uusien aineistojen keruussa tekijänoikeuksiin liittyviin sopimuksiin ja lupiin on syytä kiinnittää erityistä huomiota, jotta avoimuuden ja vapaan käytön ajatukset voivat toteutua mahdollisimman laajasti. Erilaisia mahdollisuuksia saada aineistoja epäkaupalliseen käyttöön on selvitettävä muun muassa Creative Commons -lisenssien avulla. Creative Commons -lisenssillä oikeudenhaltija voi luopua hallitusti joukosta oikeuksia, jotka tekijänoikeus lähtökohtaisesti antaa.

AVOIMUUDEN JA VUOROVAIKUTUKSEN KULTTUURIN EDISTÄMINEN MUUN MUASSA KANSALAISTIETEEN KEINAIN

Kieliohjelman tavoitteena on edistää avoimuuden ja vuorovaikutuksen kulttuuria tieteen tekemisessä. Tämä tarkoittaa esimerkiksi lähdeaineistojen ja tutkimustulosten vapaata saatavuutta, saavutettavuutta ja käytettävyyttä esimerkiksi virtuaalisen kirjaston kautta ja kieliyhteisön osallistumista dokumentoinnin ja tutkimustiedon hyödyntämisen eri vaiheisiin.

Myös aineistojen ja tiedon muokkaaminen ymmärrettävään ja kiinnostavaan muotoon on tärkeää, jotta dokumentoinnin ja tutkimuksen tulokset olisivat paremmin suuren yleisen saavutettavissa. Kieliohjelmassa tuetaan yhteisön osallistumista uusien aineistojen muodostamiseen ja vanhojen aineistojen käytettävyyden parantamiseen muun muassa kansalaistieteen keinoin. Myös taidetta ja tutkimusta yhdistävillä hankkeilla tuetaan yhteisön osallistumista ja vuorovaikutuksen syntymistä.

Kansalaistieteellä tarkoitetaan vuorovaikutteista tutkimustyötä, johon osallistuvat tutkijat, opiskelijat ja muuten asiasta kiinnostuneet kansalaiset. Kansalaistiede perustuu luotettavien ja vapaaehtoisten kansalaisten – kuten asiaan vihkiytyvien koulutettavien ja maallikoiden sekä niiden verkostojen – työhön, jota hyödynnetään havainto-, mittaus- ja laskentatyössä. Kansalaistiede on keino hankkia uusia aineistoja sekä suorittaa laajamittaista tarkistustyötä. Entisaikojen esimerkkinä tästä voi mainita Suomen murteiden sanakirjan sanastustyön, joka erinomaisesti kuvaa vastuullisten vapaaehtoisten maallikoiden tekemää hankintatyötä.

Vuorovaikutteisessa tutkimustyössä voidaan hyödyntää myös talkoistamista eli yleisöosallisuutta (engl. crowdsourcing). Sillä tarkoitetaan ennalta määrittelemättömän joukon yhteistyön käyttämistä esimerkiksi kehitys- tai tarkistustyössä. Hyvä esimerkki talkoistamisesta kielen kentällä on Kansalliskirjaston myyräpeli, jossa pelaajan vastauksia hyödynnetään automaattisen tekstintunnistamisen laadun arvioinnissa ja korjaamisessa.

Kansalaistieteen ja talkoistamisen ero ei välttämättä ole kovin suuri. Talkoistamisessa työtä tekevä joukko on yleensä kasvoton, kun taas kansalaistieteilijöillä viitataan tiettyihin tutkijoiden kanssa yhteistyötä tekeviin henkilöihin, jotka noudattavat tutkimuksellisia käytäntöjä.

Kansalaistieteeseen ja talkoistamiseen kuuluu oleellisena osana myös se, että vapaaehtoisten työpanostus on kaikkien hyödynnettävissä ja kehitettävissä. Monilla pienillä suomalais-ugrilaisilla kielillä onkin jo käytettävissä omakielinen Wikipediansa ja Skype-, Vkontakte-, sekä OpenOffice-käyttöliittymänsä, jotka ovat syntyneet talkoistamisen tuloksena. Pienten kielten sovelluksia kehitettäviä vapaaehtoisia on vähemmän kuin suurten kielten kohdalla, joten lisätuki pienten kielten sovellusten kehittämiseen voi olla tarpeen.

Vihkiytyneiden äidinkielisten käyttäminen kielentutkimuksen apuna on toimintamallina oivallinen. Äidinkielinen haastattelija-kerääjä toimii parhaiten oman sukunsa tai kylänsä piirissä. Hän on sekä sisäpiirin ihminen että tieteellisen yhteisön jäsen. Kansalaistieteilijä osaa itse arvioida materiaalin käyttökelpoisuuden ja hän nauttii haastateltavien luottamusta. Toisaalta sisäpiirin jäsenenä hänen läsnäolonsa ei saa haastateltavia käyttämään kieltä epäaidosti tai luonnottomasti. Äidinkielisten työpanosta voidaan hyödyntää kielitieteellisessä tutkimuksessa, mutta samalla se voi vahvistaa oman kielen asemaa ja käyttöä sekä kielellistä identiteettiä.

Kansalaistieteilijöiden ja talkoistajien rekrytoinnissa ja mobilisoinnissa voidaan hyödyntää esimerkiksi kansalaisjärjestöjä ja niiden verkostoja. Etenkin nuorten vapaaehtoisten perehdyttäminen kansalaistieteeseen ja talkoistamiseen voi olla hedelmällistä, sillä heillä on lähtökohtaisesti hyvät valmiudet sähköisten sovellusten käyttämiseen.

MONITIEETEISYYDEN JA UUDENLAISTEN RAJOJEN YLITYSTEN EDISTÄMINEN

Monitieteisyys, tieteidenvälisyys tai poikkitieteisyys voi tuoda tutkimukseen uusia näkökulmia, mahdollisuuksia ja ratkaisuja. Kielentutkimuksen aineistoja voidaan käyttää myös muilla aloilla ja kieliaineistoja voidaan pitää jopa monitieteisyyden arvokkaana solmuna. Kieliaineistot voivat toimia kielitieteen lisäksi myös muun muassa historia(llise)n- ja sosiologisen tutkimuksen aineistoina. Vastaavasti vaikkapa musiikkitieteellisen tai lääketieteellisen tutkimuksen kieliaineistot voivat toimia kielitieteellisen tutkimuksen aineistoina. Tavoitteena on, että tällaisia aineistojen käyttöön liittyviä monitieteisyyden mahdollisuuksia hyödynnetään.

Yhteiset aineistot saattavat olla yksi hedelmällinen eri alojen yhteistyön lähtökohta, mutta esimerkiksi kielitiede ja yhteiskuntatiede tai historian tutkimus voivat hyötyä yhteistyöstä myös muulla tavoin. Myös kieliteknologisia sovelluksia on kehitetty eri tieteenalojen käyttöön: muun muassa lääketieteessä ja kriminologiassa kieliteknologiaa käytetään äänentunnistamisessa.

Monitieteisyyteen kannustamisen lisäksi kieliohjelmassa edistetään hankkeita, joissa tieteen ja taiteen näkökulmat kohtaavat. Tieteen kentällä poikkitieteellisyys ja taiteessa poikkitaiteellisuus ovat jo pidemmän

aikaa nostaneet uudenlaisia kysymyksiä tutkimuksen ja taiteen aiheiksi ja monipuolistaneet aiheiden käsittelyä. Kieliohjelmassa halutaan kannustaa tarkastelemaan asioita ja käsittelemään aiheita myös taidetta ja tiedettä yhdistävistä näkökulmista ja rikkomaan eri alojen välisiä rajoja.

MITEN TAVOITTEITA TOTEUTETAAN?

Hankkeiden rahoittaminen ja alkuun paneminen

Kieliohjelma huomioidaan kaikessa säätiön toiminnassa. Tavoitteiden toteuttamiseksi järjestetään suunnattuja hakuja ja käynnistetään aktiivisesti hankkeita, joiden käytäntöönpanossa säätiö on eriaikaisesti mukana. Lisäksi kieliohjelmaa toteutetaan jo olemassa olevien säätiön apurahahakujen kautta painottamalla valinnoissa ohjelman tavoitteita toteuttavia hankkeita. Voidaan myös aktiivisesti kehottaa eri tahoja hakemaan apurahoja säätiön apurahahauissa.

Tavallisesti Koneen Säätiö ei ole ollut itse aktiivisena osapuolena rahoittamisessa hankkeissa. Säätiö on kyllä järjestänyt esimerkiksi seminaareja yhdessä rahoittamiensa hankkeiden kanssa, mutta hankkeiden sisältöön ei ole puututtu. Perinteisen apurahojen myöntämisen lisäksi kieliohjelma sisältää toimenpiteitä, joissa säätiö itse toimii aktiivisena osapuolena ja on esimerkiksi mukana suunnittelemassa tutkimushanketta ja määrittämässä sen sisältöä.

Yhteistyön edistäminen ja hyvien käytäntöjen levittäminen

Lisäksi kieliohjelman tavoitteiden saavuttamiseksi hyödynnetään olemassa olevia verkostoja, edistetään verkostoitumista erilaisin keinoin ja tehdään yhteistyötä samalla kentällä työskentelevien toimijoiden kanssa (ks. s. 7, *Eri toimijat suomalais-ugrilaisia kansoja ja kieliä tukemassa*).

Ohjelma sisältää paljon osioita, joita edistetään yhteistyössä jo kentällä toimivien organisaatioiden ja hankkeiden kanssa. Erityisesti tämä koskee erilaisia tekijänoikeuksiin ja infrastruktuurin kehittämiseen liittyviä toimenpiteitä, joita on jo tehty erilaisissa digitointihankkeissa ja kieliteknologisten sovellusten kehittämistyössä. Kieliohjelman puitteissa kehitetään myös erilaisia hankkeiden välisen yhteydenpidon ja yhteistyön välineitä.

Suomessa on toteutettu ja ollaan toteuttamassa erilaisia suomen kielen digitointihankkeita tai muita hankkeita, joissa hyödynnetään kieliteknologiaa. Monia tällaisia hankkeita voidaan kieliohjelman näkökulmasta pitää pilottihankkeina, joita voidaan mahdollisesti myöhemmin soveltaa muiden kielten tutkimuksessa ja dokumentoinnissa. Kieliohjelman tavoitteena on tukea tällaisia hankkeita ja sovelluksia. Erityisesti halutaan tukea avoimuutta ja vuorovaikutuksellisuutta edistäviä hyviä käytäntöjä.

Konkreettisista toimenpiteistä eli ohjelman tavoitteisiin perustuvista hankkeista ja suunnatuista hauista kerrotaan ohjelman luvussa *Toimenpiteet 2012-2016* (s. 22).

Toimenpiteet 2012–2016

Toimenpideosiota päivitetään ja tarkennetaan vuosittain ja otetaan huomioon mahdollisesti myöhemmin esille tulevat tarpeet hankkeiden tukemisessa. Kieliohjelmasta ja rahoitetuista hankkeista kerrotaan säätiön verkkosivuilla.

Hankkeisiin on vuosille 2012–2016 varattu alustavasti noin kuusi miljoonaa euroa.

HANKKEIDEN VÄLINEN YHTEYDENPITO JA YHTEISTYÖ

Säätiö järjestää vuosittain seminaareja ja tapaamisia hankkeiden välisen yhteydenpidon ja yhteistyön tukemiseksi. Myös erilaisia yhteydenpidon ja yhteistyön sujuvoittamisen verkkopohjaisia ratkaisuja kartoitetaan ja kehitetään.

Hankkeiden kanssa yhdessä ja/tai osana säätiön itse käynnistämiä hankkeita suunnitellaan kansalaistieteen ja talkoistamisen käyttämisen ja hyödyntämisen keinoja.

Kieliohjelman toimenpiteiden suunnittelussa mietitään, miten tekniset ratkaisut syntyvät oikeaan aikaan suhteessa ohjelmasta rahoitettavaan sisältötyöhön. Esimerkiksi aineistojen käyttöön saattaminen voi helposti muodostua pullonkaulaksi. Myös tällaisia kysymyksiä pohditaan hankkeiden yhteisissä tapaamisissa.

VUOSITTAIN TOISTUVIEN APURAHAHAKUJEN KAUTTA TUETTAVAT HANKKEET, SUUNNATUT HAUT JA SÄÄTIÖN KÄYNNISTÄMÄT HANKKEET

Säätiön vuosittain toistuvien apurahahakujen kautta tuettavia (2012–2016)

Säätiön vuosittain toistuvissa apurahahauissa otetaan huomioon kieliohjelma ja sen tavoitteet. Säätiö voi myös kehottaa eri tahoja hakemaan säätiön apurahaa hankkeisiinsa. Kieliohjelman painotuksia otettiin huomioon jo vuoden 2011 apurahapäätöksissä.

»» Sanakirjojen, oppikirjojen ja oppimateriaalien ja muiden kieliohjelmalla sopivien kirjojen kirjoittamista ja kääntämistä tuetaan säätiön

tietokirjahaun kautta. Rahoitusta myönnettiin 2011 muun muassa valtion tuottavuusohjelman leikkausten vuoksi keskeytetyille *Vanhan kirjasuomen sanakirjahankkeelle*. Työ jatkuu nyt Kotuksessa Koneen Säätiön tuella.

»» Vähemmistökansojen taiteilijoita tuetaan **Saaren kartanon residenssin ja taiteen ja kulttuurin apurahojen kautta.** Erityisesti halutaan tukea sukukansojen kirjailijoiden ja kääntäjien työtä ja rekrytoida heitä Saaren kartanon residenssiin. Myös muita kieliohjelman tavoitteita toteuttavia taiteen ja/tai kulttuurin hankkeita tuetaan. Esimerkiksi professori Jyrki Nummen hankkeelle *Suomalainen kirjallisuuspankki* myönnettiin Koneen Säätiön rahoitusta vuonna 2011. Tarkoituksena on suunnitella ja testata klassisen suomalaisen kaunokirjallisuuden digitaalinen kokoelma ja kokoelmaan liittyvän käyttäjälähtöisen verkkopalvelun toimintamalli.

»» Vuosittain myönnettävistä **tutkijastipendeistä ja hankerahoituksesta** päätettäessä otetaan huomioon kieliohjelman painotukset. Vuonna 2011 rahoitusta myönnettiin muun muassa yliopistonlehtori Tiina Onikki-Rantajääskön johtamalle hankkeelle *Suomen kognitiivista kielioppia sukukielten kehyksessä*.

Säätiön käynnistämä hanke: Pietarin aineistojen digitoinnin pilottihanke (2012–2013) ja mahdollinen jatkohanke (2013–2016)

Ks. s. 15, *Aineistojen saatavuuden, saavutettavuuden ja käytettävyyden parantaminen*.

Koneen Säätiö on käynyt neuvotteluja Kansalliskirjaston ja Helsingin yliopiston kirjaston kanssa Venäjän kansalliskirjastossa Pietarissa säilyttävien suomalais-ugrilaisten aineistojen digitoinnista. Alustavasti pilottiprojektiksi on valittu eräiden marilaisten ja mordvalaisten 1920- ja 1930-lukujen sanomalehtien ja vepsän- ja inkeroiskielisten kirjojen digitointi. Pietarissa selvitetään lehtien tekijänoikeudet ja tehdään perusdigitointi ja mahdollisesti metatietojen tallentaminen. Tätä seuraavat vaiheet tehdään luultavasti Suomessa: tekstin optimointi, lehtien siirtäminen jollekin käyttöalustalle jne. Hankkeessa on alusta saakka syytä ottaa huomioon tutkijoiden tarpeet ja ottaa heidät osalliseksi hankkeen suunnitteluun ja toteutukseen.

Jatkohanke voi sisältää yhteistyötä Venäjän kansalliskirjaston lisäksi muiden tasavaltojen kirjastojen kanssa. Kesällä 2012 Helsingissä järjestetään suuri kirjastoalan kokoontuminen, IFLAN (The International Federation of Library Associations and Institutions) vuotuinen konferenssi, jonka yhteydessä on mahdollisuus pitää hankesuunnittelupalaveri.

Säätiön käynnistämä hanke: Morfologia jäsentimiä suomalais-ugrilaisille vähemmistökielille (2012–2016)

Säätiö käynnistää hankkeen, jossa kieliteknologian ja kielitieteen yhteistyön tuloksena saadaan avoimeen lähdekoodiin perustuvia tiivistä ilmaisu hyödyntäviä tekstinjäsentimiä (transduktoreita). Jäsentimiä voidaan hyödyntää oikeinkirjoitusohjelmissa, tekstikorpusten tekstien analyysissa sekä skannattujen tekstien tunnistamisessa. Alustava suunnitelma on, että hankkeessa työskentelee yksi kieliteknologi, joka koordinoi ja ohjeistaa neljää muuta eri kieliin keskittyvää kielentutkijaa. Hankkeessa mukana olevat kielet voisivat olla esimerkiksi mari, mokša, nenetsi ja vepsä. Helsingin yliopiston nykykielten laitoksen kanssa on neuvoteltu hankkeen toteuttamisesta laitoksella. Hankkeen vaiheet ovat:

- a) morfofonologinen ja säännöllisen taivutuksen peruskuvauksen jäsentimien varten
- b) morfofonologinen ja säännöllisen taivutuksen kuvaus ynnä hakusanaalista taivutustietoineen (noin 20 000 sanaa)
- c) morfofonologinen ja säännöllisen taivutuksen kuvaus ynnä hakusanaalista taivutustietoineen sekä alustava glossaus ja frekvenssiin perustuva painotus
- d) sanastotutkimusta ja tekstintunnistamista edistävien sovellusten kehittäminen.

Osalla kielistä on olemassa ainakin alustavia taivutus- ja morfofonologisia kuvauksia ja/tai sanastoja, joilla voidaan päästä nopeammin alkuun. Yhdestä kielestä tehtyjä kuvauksia voidaan hyödyntää läheisen sukukielen kuvauksia tehtäessä. Kaikista suomalais-ugrilaisista kieliryhmistä paitsi samojedikielistä on olemassa vähintään yksi alustava morfofonologinen ja taivutuskuvaus.

Hankkeessa on tärkeää olla läheisessä yhteistyössä esimerkiksi CLARIN-hankkeen kanssa, jotta vältytään päällekkäisyyksiltä ja kirjavilta ratkaisuilta.

Säätiön käynnistämä hanke: Suomalais-ugrilaisten teatterintekijöiden ja kielentutkijoiden yhteishanke (suunnittelu 2012, toteutus 2013–2014)

Hankkeen valmistelu aloitetaan 2012, mutta varsinainen tuettava hanke toteutetaan 2013–2014.

Suunnattu haku: Aineistot (2013, alkuvuosi)

Säätiön käynnistämässä aineistohankkeessa keskitytään etenkin Pietarissa sijaitseviin suomalais-ugrilaisiin aineistoihin. Kieliohjelmassa aineistot ovat suuressa osassa, joten aineistoihin liittyen olisi syytä rahoittaa myös muita hankkeita. Suunnatussa haussa voidaan tukea sekä aineistojen pelastamista että niiden hyödyntämistä ja kansalaistieteen keinojen käyttämistä. Erityishuomiota voisivat saada suomalaisten tutkijoiden 1800-luvun lopussa ja 1900-luvun alussa keräämät aineistot. Aineistoja koskeva suunnattu haku voi olla myös osa Pietarin aineistojen digitoinnin jatkohanketta.

Suunnattu haku: Leksikologinen ja leksikografinen tutkimus ja sanastotyö (2013, loppuvuosi)

Leksikografisen ja sanastotutkimuksen olisi syytä toimia nykyistä tiiviimässä vuorovaikutuksessa keskenään. Yhteistyössä nämä tukevat parhaiten sanastojen tallennusta, kehitystyötä ja sanaston ilmentämän maailmankuvan tuomista laajempaan tietoisuuteen.

Leksikografiseen ja sanastotutkimukseen liittyvä suunnattu haku pyritään toteuttamaan ohjelman alkuvaiheessa, jotta sen kautta saadaan sanastolähteitä hakusanalistoisiin. Aiheita voivat olla esimerkiksi suomalais-ugrilaisten kielten terminologian muodostus ja suomalais-ugrilaisten kielten 1920- ja 1930-lukujen sanasto. Myös muun muassa omakielisen oppisanaston kehittäminen vähemmistökielten kohdalla on tärkeää kielten elinvoiman kannalta, ja erikoisalanastosta täytyy myös valtakielissä ylläpitää ja täydentää.

Tulokset hyödyttävät sekä *Tieteen kansallisen termipankin* kehitystä että pienten suomalais-ugrilaisten kielten sanastokehitystä.

Suunnattu haku: Kielioppien tekeminen ja kieliopin keskeisiä osa-alueita koskeva tutkimus (2014)

Niissäkin suomalais-ugrilaisissa vähemmistökielissä, joissa on jo olemassa kielioppeja, on paljon selvitettävää kieliopin osa-alueilla. Monet olemassa olevat kieliopit ovat varsin suppeita yleistyksiä ja saattavat perustua pieneen ja yksipuoliseen materiaaliin.

Vähemmistökielten kielioppeja voidaan toteuttaa esimerkiksi väitöskirjatyönä. Väitöskirja voi kuvata osittain morfologista jäsennintä varten tarvittavaa kielittäistä morfofonologiaa, säännöllistä taivutusta ja johto-oppia sekä sanamuotojen tunnistamista syntaksitietojen avulla.

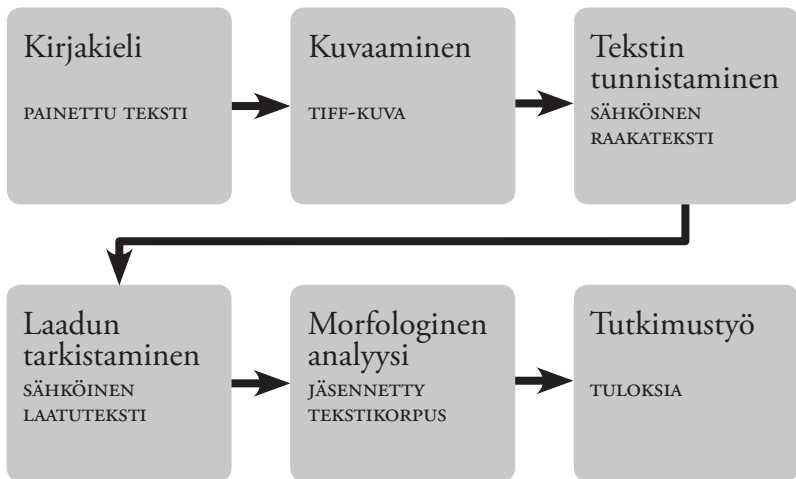
Suunnattu haku: Kielten elvytyksen ja kehityksen tutkimus (2015)

Kielten elvytystä ja kehitystä tutkivat hankkeet voivat liittyä muun muassa kaupungistumiseen liittyviin ongelmiin.

Esimerkki kieliteknologisten sovellusten käyttämisestä ja kehittämisestä

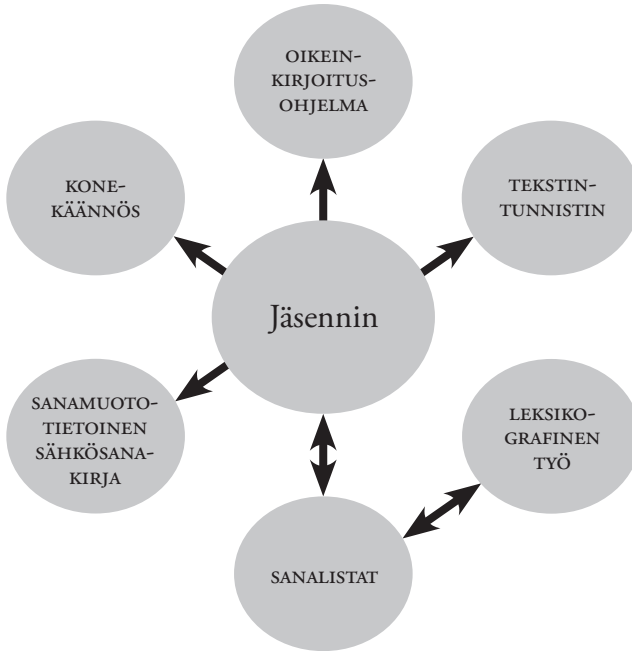
Tuettavia hankkeita ideoitaessa ja hankkeiden tukemisessa otetaan huomioon yhteistyön mahdollisuudet eri hankkeiden välillä. Esimerkkinä eri hankkeiden yhteistyöstä ja hankkeiden välisen synergian hyödyntämisestä voidaan pitää esimerkiksi erilaisten kieliteknologisten sovellusten käyttämistä ja kehittämistä eri hankkeissa ja saman hankkeen eri vaiheissa (ks. kuvat 1-3).

Kuva 1: Painetusta tekstistä jäsennetyn tekstin hyödyntämiseen



Kun dokumentoidaan kirjoitettua kieltä, lähtökohtana on painettu tai sähköinen teksti. Painettu teksti on ensin kuvattava ja tunnistettava tekstiksi. Tunnistetun tekstin laatuun vaikuttaa tunnistimen rakenne. Yksinkertainen tunnistin tunnistaa kirjaimet ja välimerkit. Painetun tekstin jälki ei ole kuitenkaan aina optimaalinen, esimerkiksi kirjaimet saattavat olla katkonaisia. Tällöin tekstintunnistamisen laatua kohentaa sanamuotojen tunnistaminen, joka on mahdollista tiivistä ilmaisua hyödyntävällä morfologisella jäsentimellä.

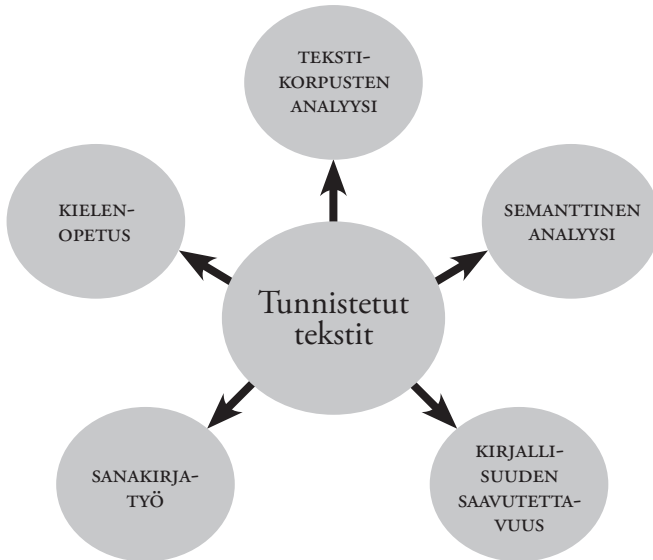
Kuva 2: Mihin jäsenintä sovelletaan?



Jäsenin on automaatti, joka yhdistää hakusanalistat ja taivutustiedot. Sitä voidaan hyödyntää tekstintunnistamisen lisäksi esimerkiksi oikeinkirjoitusohjelmissa, sanakirjoissa ja koneellisessa kääntämisessä.

Jäsentämisessä tarvittavat sanalistat ovat hyödynnettäviä myös leksikografiseen työhön. Yhteistä tietokantaa käyttämällä voidaan lisätä jäsentimiin semanttista tietoa, joka puolestaan helpottaa muulla kielellä kirjoitettujen tekstien ymmärtämistä. Sanalistoja suurennetaan tunnistamattomien sanojen osalta. Digitaaliset tekstit ja tarkistetut tekstikorpukset ovat testausmateriaalia, jota voidaan hyödyntää jäsentimen kehittämisessä. Tunnistuskyvyn kasvaessa jäsenintä käyttävien sovellusten tarkkuus myös kasvaa.

Kuva 3: Tunnistettujen tekstien käytettävyys



Tunnistettujen tekstien määrän kasvaessa myös kirjallisuuden saavutettavuus parantuu. Tekstejä voidaan hyödyntää tekstikorpuksina tai kielen ja kulttuurin opetuksessa ja niitä voidaan myös käyttää semanttisessa tutkimuksessa, esimerkiksi sanakirjatyössä.

Ohjelman seuranta

Ohjelmalle ei perusteta ohjausryhmää, mutta huolehditaan siitä, että suomalaiset ja ulkomaalaisetkin toimijat ovat tietoisia ohjelman etenemisestä ja Koneen Säätiössä tiedetään, mitä muut tekevät. Osallistutaan esimerkiksi Suomen Kulttuurirahaston sukukansatyöryhmän kokoontumisiin ja tiedotetaan aktiivisesti omasta ohjelmasta.

Ohjelman toteuttamisen ensimmäisen vuoden aikana mietitään, miten ohjelmaa ja sen vaikuttavuutta seurataan.

KONEEN SÄÄTIÖ
Tehtaankatu 21 B 49
00150 Helsinki
koneensaatio@koneensaatio.fi
Puh. (09) 260 0617
Fax (09) 260 0618
www.koneensaatio.fi