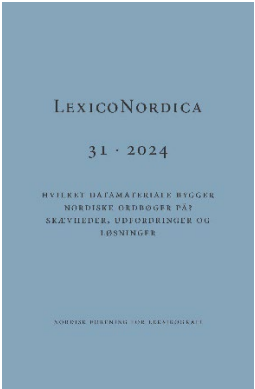


LexicoNordica

Titel:	Språkteknologi för att samla in texter och analysera språket i korpusverktyg – hur gör man på meänkieli?	
Forfatter:	Magnus Ahltop, Lina Lejdebros Enwald, Elina Kangas, Jacob Larsson, Rickard Domeij & Gunnar Eriksson	
Kilde:	LexicoNordica 31, 2024, s. 21-37	
URL:	https://tidsskrift.dk/lexn/issue/archive	

© 2024 LexicoNordica og forfatterne

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Språkteknologi för att samla in texter och analysera språket i korpusverktyg – hur gör man på meänkieli?

Magnus Ahltop, Lina Lejdebros Enwald, Elina Kangas, Jacob Larsson, Rickard Domeij & Gunnar Eriksson

The Language Council of Sweden at the Institute for Language and Folklore has a responsibility to support the development of language technology for the languages of Sweden. Together with the Department of National Minorities and Swedish Sign Language at the same institute, we are developing tools to collect and analyse texts in the national minority languages of Sweden. These texts will be available for use in research, for example in lexicographic work. We here present the specific challenges associated with building these tools for the Swedish national minority language Meänkieli.

1. Inledning

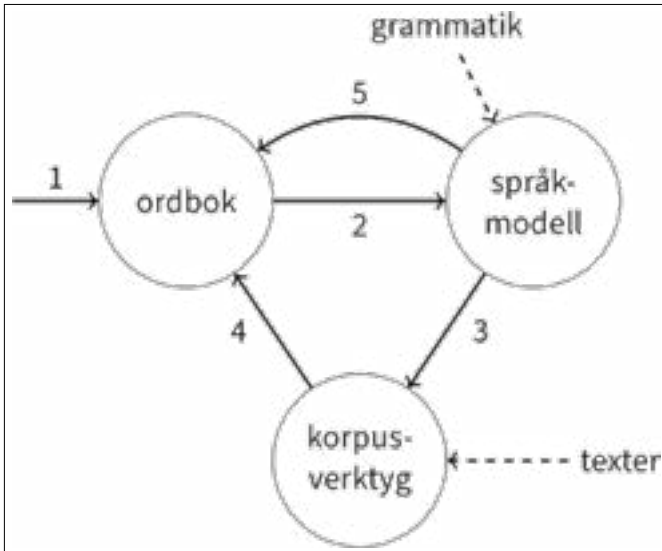
Institutet för språk och folkminnen (Isof) är en statlig myndighet vars uppgift är att samla in, bygga upp och sprida kunskap om språk och kultur i Sverige. Myndigheten ska bedriva språkvård och på vetenskaplig grund öka, levandegöra och sprida kunskaper om språk, dialekter, folkminnen, namn och immateriella kulturarv i Sverige. Språkvård för det svenska språket har bedrivits i semi-statlig regi sedan Nämnden för svensk språkvård blev Svenska språknämnden i mitten av 1900-talet. För ett nationellt minoritetsspråk som meänkieli, som har sämre förutsättningar och betydligt mindre resurser än svenskan, har ett sådant uppdrag endast funnits på Isof sedan 2006.

År 2000 ratificerade Sverige två av Europarådets konventioner, *Ramkonventionen om skydd för de nationella minoriteterna* och *Den europeiska stadgan om landsdels- eller minoritetsspråk*. Där-

med fick finska, jiddisch, meänkieli, romska och samiska status som nationella minoritetsspråk i Sverige. I och med undertecknandet har Sverige lovat att trygga rättigheter kopplade till nationella minoritetsspråk.

År 2009 fick Sverige en språklag som slår fast att svenska är huvudspråk i Sverige, men som också säger att det allmänna har ett särskilt ansvar för att skydda och främja de nationella minoritetsspråken. Isf har därför i uppgift att stötta meänkieli, liksom övriga nationella minoritetsspråk, och som ett led i detta att bidra till att skapa en infrastruktur för språkteknologi för språket.

Med ett litet antal språkbrukare, se avsnitt 3 nedan, följer problemet att det saknas kommersiellt intresse för att konstruera språkteknologiska verktyg som stavningskontroll, korpusverktyg eller talsyntes, vilket gör Isofs uppdrag än viktigare (Mattson & Ahltop 2023; Trosterud 2022).



Figur 1: Översikt över språkteknologisk infrastruktur för meänkieli. Siffrorna visar de olika relationerna komponenterna har till varandra.

För att kunna använda korpusar för att förbättra ordböcker måste det först finnas korpusverktyg. Ett korpusverktyg som är lämpligt för lexikografiska undersökningar förutsätter språkmodeller, som i sin tur förutsätter ordböcker (se figur 1). Med språkmodell menas här ett sätt att automatiskt analysera morfologi, syntax, etc. för språket. Denna artikel beskriver Isofs arbete med korpusverktyg, språkmodeller och språkvård för meänkieli, och hur dessa är beroende av varandra.

2. Meänkieli som språk

Meänkieli tillhör den östersjöfinska grenen av de uraliska språken. De närmaste besläktade språken är finska, där de största likheterna finns i de nordfinska dialekterna, och kvänska, som har minoritetsspråksstatus i Norge. Meänkieli är också besläktat med exempelvis samiska, estniska och på längre håll ungerska.

Meänkieli är sällan inkluderat i exempelvis språkträd och kartor som illustrerar de uraliska språkens släktskap och utbredning, något som delvis beror på att meänkieli inte fick språkstatus förrän år 2000. Enligt Blokland (2024) saknas dock ofta meänkieli även i nyare akademiska källor, alternativt klassificeras som en dialekt av finskan, se t.ex. Unescos *Atlas of the world's languages in danger* (Moseley & Nicolas 2010).

Meänkieli delar många språkdrag med övriga uraliska språk, exempelvis ett rikt kasussystem, nekandeverb, avsaknad av genus, stadieväxling (kvantitativ eller kvalitativ förändring hos stamkonsonanter) och vokalharmoni (främre och bakre vokaler förekommer inte i samma ord). Vad gäller det sistnämnda skiljer det sig emellertid åt mellan olika geografiska varieteter – i den meänkieli som talas kring Malmfälten frångås ibland principen med vokalharmoni.

Allra flest drag delar meänkieli med kvänskan och nordfinska dialekter, och meänkieli, kvänska och finska är sinsemellan hu-

vudsakligen ömsesidigt förståeliga. Meänkieli har emellertid stått under stark påverkan av svenskan, medan kvänskan istället påverkats av norskan. Varken meänkieli eller kvänskan har varit en del av standardiseringsprocessen för finska i Finland, och båda språken har bevarat en del språkdrag som tidigare funnits eller finns i finska dialekter men som inte tagits upp i det finska skriftspråket.

Skillnaderna mellan meänkieli och standardfinska är tydliga exempelvis vad gäller personliga pronomen, verbböjningar, inskott av h-ljud i efterstavelser samt i meänkielis många moderna låneord från svenska.

Exempel på skillnader mot kvänska är vissa kasusändelser, verbböjning och kvänskans dentala tonande frikativa /ð/, ett ljud som tidigare förekom även i meänkieli men som saknas i dagens språk (Paunonen 1987). Både kvänska och meänkieli har generell stadieväxling, men kvänskan har dessutom så kallad specialstadieväxling (t ex *poika* 'pojke' i nominativ singular blir *pojan* 'pojken's' i genitiv singular men *poikkaa* i partitiv singular, en form som saknas i meänkieli).

Precis som många andra uraliska språk har också meänkieli omfattande morfofonologisk växling, vilket nedanstående exempel visar. Denna egenskap komplicerar konstruktionen av språkmodellen.

- (1) a. kanala
hönsgård.SG.NOM
'hönsgård'
- b. kanalo-i-ta
hönsgård-PL-PAR
'hönsgårdar'
- (2) a. katto
tak.SG.NOM
'tak'

- b. kato-le
tak-ALL
'till tak(et)'
- (3) a. mettä
skog.SG.NOM
'skog'
- b. mettä-le
skog-ALL
'till skog(en)'

Exempel (1) visar hur vokalförändring äger rum som en följd av pluralsuffix; det avslutande a:et i *kanala* blir *-o-* på grund av pluralmarkören *-i-*. Exempel (2) visar klassisk stadieväxling där dubbelklusilen *-tt-* blir enkel på grund av allativ-suffixet. Exempel (3) utgör ett undantag från denna regel, vilket beskrivs vidare i sektion 6.

3. Språkbrukare och revitalisering

Traditionellt språkområde för meänkieli är Tornedalen och Malmfälten, områden som idag tillhör de svenska kommunerna Gällivare, Haparanda, Kiruna, Pajala och Övertorneå, men på senare tid har många talare flyttat till andra delar av landet, i synnerhet till storstäderna, som en del av en generell flyttvåg från Norrland. Antalet meänkielitalare är svårbedömt eftersom Sverige inte samlar in sådan statistik – uppskattningarna varierar mellan 20 000 och 75 000, beroende på källa och vilka man valt att inkludera i definitionen (Parkvall 2015; Valijärvi et al. 2022).

Meänkielitalare har under lång tid tvingats gå igenom en språkbytesprocess vilket lett till att meänkieli ofta inte har förts vidare till nästa generation, och språket klassas idag som allvarligt

hotat (Laakso et al. 2016). De flesta talare är äldre och flerspråkiga; alla meänkielitalande är också svensktalande.

Mobiliseringen för ett erkännande av meänkieli som ett eget språk inleddes under tidigt 1980-tal, när meänkielitalande började använda språket i kultursammanhang som teater, musik och litteratur. Intresseföreningen Svenska Tornedalingars Riksförbund – Tornionlaaksolaiset bildades 1981. Den första meänkielispråkiga romanen kom 1985 (*Lyykeri* av Bengt Pohjanen), de första teaterpjäserna på meänkieli kom snart efter det.

Idag görs många åtgärdsförsök för att återta (revitalisera) språket, allt från mindre projekt framtagna av privatpersoner till strukturerade myndighetsuppdrag. Det kan handla om exempelvis språkläger, universitetskurser eller andra språkkurser, särskilda förskolor med inriktning mot meänkieli, språkmyndigheten Isofs revitaliseringsuppdrag, etc.

Isof hade under åren 2006–2013 ett arkivinriktat arbete för meänkieli, och en språkvårdstjänst tillsattes år 2018. Isofs arbete bygger vidare på det arbete som redan gjorts av talarna, exempelvis i form av den digitala ordbok (*Meänkieli–ruotti sanakirja*) som tagits fram av Meän Akateemi-Academia Tornedaliensis (*Meän Akateemis Meänkieli-svensk/svensk-meänkieli digital ordbok* 2024) med delfinansiering från Isof. Detta arbete utgör steg 1 i figur 1.

År 2022 fick Isof därutöver ett treårigt regeringsuppdrag att inrätta ett språkcentrum för meänkieli, finska, jiddisch och romska, i syfte att ge stöd och kunskap som underlättar för språkbärare i hela landet att behålla, ta tillbaka och utveckla sitt språk.

Allt detta innebär att det idag också tillkommit nya talare som inte har förkunskaper i språket hemifrån. Fortfarande råder dock stor brist på läromedel, utbildningar etc. – det hänger till stor del på den enskilda individen att hitta inlärningsmöjligheter.

Språkverktyg som stavnings- eller grammatikkontroll är under utveckling på Isof men finns ännu inte tillgängliga för allmänheten.

4. Särskilda utmaningar för meänkieli

En särskilt komplicerande faktor för språkvården för meänkieli är att det saknas omfattande deskriptiv grammatik. Språkvårdsarbetet på Isof får därför utgå från publikationer inom finsk dialektologiforskning (Mantila 1992; *Suomen murteiden sanakirja* 2020), finsk grammatik (Karlsson 2012), kvänsk grammatik (Söderholm 2017) samt modernare meänkielimaterial i form av exempelvis normativ grammatik (Pohjanen 2022) och meänkielikorpus (Korp för Meänkieli 2024). Vid sidan av detta är utgångspunkten lingvistiska observationer av vardagligt språkbruk i media, i Facebookgrupper, i kontakten med talare (exempelvis vid språkrådgivningssituationer) osv.

Meänkieli är till stor del ett talat språk; många har inte fått lära sig läsa och skriva på meänkieli (Valijärvi et al. 2022). Det finns en förväntan från talare på att få hjälpmedel för att skriva på meänkieli, i synnerhet bland dem som inte fått språket hemifrån, samtidigt som de normativa val som gjorts eller föreslås ofta kan ifrågasättas – som talare kanske man inte alltid identifierar sig eller sitt språk med en viss stavning eller ett visst ordval.

De normeringsförslag som finns sedan tidigare kommer i första hand via aktiva språkaktörer, Facebook-grupper, författare som Bengt Pohjanen som publicerat såväl romaner som språkmaterial i form av ordböcker och skrivregler (Pohjanen 2022), osv. Många av meänkielitalarna utgår inte från dessa normeringsförslag i skrivprocessen, även om man känner till att de finns.

Det mesta som är skrivet på meänkieli är baserat på varieteten som talas i Tornedalen (Kuoppa 2015), trots att exempelvis vokalharmoni, verbböjning och många andra drag uppvisar regional variation. Skälet till detta är att antalet talare från andra områden är få, relativt sett, och att skribenter från dessa geografiska områden är ännu färre – de aktiva skribenterna kommer kort sagt från Tornedalen.

Meänkieli uppvisar dessutom andra typer av variation som språkvården måste hantera än den geografiska; till exempel den mellan äldre/yngre talare, förstaspråks-/andraspråkstalare, uttalskillnader och skillnaden mellan svenska och finska låneord.

Dessa komplicerande faktorer, i kombination med det faktum att meänkielikorpusen är mycket begränsad (se vidare under avsnitt 7), gör att arbetet med att stödja språket måste bedrivas på två fronter samtidigt; språknormer tas fram samtidigt som språkteknologin utvecklas, och i någon mån tvingar arbetet med språkverktyg också normeringen vidare – för att kunna göra exempelvis en stavningskontroll måste ortografin klargöras, och i och med det upptäcks fenomen som sedan behöver få en lingvistisk förklaring.

5. Språkmodellens betydelse för korpus-sökningar

Till skillnad från meänkieli har de stora nordiska språken haft både lexikografiska och språkteknologiska resurser under lång tid. Svenska har exempelvis haft ordböcker sedan åtminstone 1712 (Ralph 2009). Språkteknologiska verktyg för svenska var själva grunden till arbetet med *Nusvensk frekvensordbok* (Allén 1970), ett arbete som påbörjades på 1960-talet. Arbetet med ordböcker, språkmodeller och korpusverktyg har utvecklats parallellt under lång tid. För de stora nordiska språken har verktyg för stavningskontroll funnits brett tillgängliga sedan åtminstone tidigt 1990-tal. Dessa språk har en väl utvecklad språkvetenskaplig och språkteknologisk infrastruktur där insamlade texter och språkmodeller kan kombineras i verktyg som analysverktyget Sparv (Hammarstedt et al. 2022) och korpusverktyget Korp (Borin et al. 2012).

Enkla sökningar kan göras i korpusverktyg utan språkmodeller, men är begränsade till exakta strängar eller trunkering

(sökning efter delsträngar). Exempelvis hittar en exakt sökning efter strängen ”bära” i en svensk korpus bara den exakta formen. Vanligtvis kan en sådan sökning även hitta formerna ”Bära” och ”BÄRA”. Även om sökning oberoende av versaler och gemener kan tyckas trivialt är redan det en form av språkmodell, något som är långt ifrån språkoberoende. Enkla exempel är tyskans <ß> som behöver transformeras till <ss>, och turkiskans ortografi där den gemena formen av <I> inte är <i>, utan <ı>. Ännu mer komplicerat blir det i språk som har stor stavningsvariation. Ett extremt exempel är japanska, där den som skriver i de flesta fall kan välja vilka delar av ett ord som ska skrivas med logografiska tecken och uttalstecken, något som förekommer i t.ex. svenskan, men i mycket begränsad omfattning, som i <8:e> och <åttonde>.

Med trunkering kan ytterligare former hittas. Vill någon söka efter olika former av verbet ”bära” kan en sökning efter ”bär*” korrekt hitta formerna ”bära” och ”bär”. Denna sökning kommer dock inte hitta formerna ”bar”, ”burit”, ”buren” (undergenerering), men de orelaterade strängarna ”bärplockning” och ”bärsärk” hittas också (övergenerering). Beroende på vilka former av verbet som är intressanta kan de hittade strängarna ”bäras” och ”bärande” vara önskade eller oönskade.

En mycket enkel sorts språkmodell är s.k. stemming, som till viss del hanterar böjningsformer utan att användaren själv behöver söka med trunkering eller själv mata in alla böjningsformer. Stemming använder mycket enkla regler för att ”skära av” ordet efter dess stam. Även här är undergenerering och övergenerering vanliga problem, och stemming är mer lämpat för språk med okomplicerad morfologi.

Även för språk där en stemmer fungerar någorlunda bra är den oftast bara lämplig för enklare sökningar, som i sökmotorer för webbplatser eller hela webben. Sökningar i korpusverktyg behöver däremot normalt utgå från en fullständig morfologisk analys. En

anledning, förutom att få en ordentlig lemmatisering, är att det även är intressant hur ordet är böjt. Sökningar kan då göras av typen ”alla ord som börjar på ’a’ och står i bestämd form”.

Lemmatisering som tar hänsyn till kontext ger även möjlighet att disambiguera strängar som ”bär” (antingen ’liten frukt’ eller imperativ av *bära*).

För att korpusar ska kunna användas för lexikografiskt arbete enligt steg 4 i figur 1 krävs alltså att en språkmodell för språket är tillgänglig för korpusverktyget (steg 3).

6. En språkmodell för meänkieli

Flera av de stora nordiska språken har omfattande resurser, i form av grammatiska beskrivningar, ordböcker som täcker en stor del av språket och lättillgänglig text författad direkt på språket. Stora mängder text, särskilt annoterad sådan, kan då användas för att träna maskininlärningsmodeller.

När det gäller lågresursspråk, där det inte finns tillräckliga mängder språkdata att träna maskininlärningsmodeller på, måste arbetet med att ta fram en språkmodell i stället göras manuellt och explicit (Lejdebros Enwald 2024; Trosterud 2022): grammatik och morfofonologiska växlingar måste formuleras som maskinläsbara regler. För detta brukar man använda sig av så kallade ändliga tillståndsautomater, Finite State Transducers (FST) med tvånivåmorfologi. Forskningsgrupperna GiellaTekno och Divvun vid Universitetet i Tromsø Norges arktiske universitet (UiT) har under de senaste decennierna tagit fram just sådana modeller för ett stort antal lågresursspråk med rik morfologi, många av dem minoritetsspråk, och samarbetar nu med Isuf för att utveckla språkteknologi för meänkieli (Pirinen et al. 2023).

Ett exempel på en sådan regelbaserad språkmodell är den kväniska, som GiellaTekno/Divvun tagit fram i samarbete med

Kvensk Institut (nationellt center för kvänskt språk och kvänsk kultur) och som också används som en stavningskontroll.

År 2020 tog man vid UiT även fram en grund för en meänkieli-modell, baserad på den finska och kvänska språkmodellen men anpassad till meänkieli utifrån en grammatikbok (Kenttä & Pohjanen 1996) och den digitala ordboken (*Meän Akateemis Meänkieli-svensk/svensk-meänkieli digital ordbok 2024*; Trosterud 2020). Ordboken är helt nödvändig indata till språkmodellen (steg 2 i figur 1).

Syftet med språkmodellen är att kunna analysera och generera språkets ordformer. För språkmodellen input i form av lemma och önskad grammatisk form, exempelvis *pilvi* 'moln', plus den grammatiska taggen "singular illativ", ska den generera korrekt böjd ordform: *pilvheen*, och omvänt; utifrån en böjd ordform ska språkmodellen kunna generera lemma och grammatisk information. Förutom att språkmodellen behövs för korpusverktyget är det viktigt även direkt i ordboken där den kan användas både för sökning och presentation (steg 5 i figur 1).

Isofs språkvårdare och språkteknologer arbetar nu med att vidareutveckla denna språkmodell för meänkieli. Exempel på specifika anpassningar som behöver göras för meänkieli, vid sidan av att korrigera böjningsparadigm och utöka modellens lexikala delar, är de h-metateser som frekvent uppstår i suffix med tonande konsonanter – det som ibland kallats "varumärket för meänkieli" (Kuoppa 2015). Ovan nämnda *pilvheen* kan på grund av detta metatesfenomen också bli *pilhveen*. De kontexter där detta sker har tidigare undersökts för finska dialekter och för ett antal meänkieliter (Mantila 1992; Kuoppa 2015).

Andra exempel på meänkielianpassningar är de ord som utgör undantag från reglerna om stadväxling; jämför exemplen (2) och (3) ovan. Normalt uppstår stadväxling i samband med exempelvis allativ-suffix, men för ord vars motsvarighet i finskt skriftspråk har *-ts-*, exempelvis 'skog' som på finska heter *metsä*

men på meänkieli *mettä*, äger ingen stadieväxling rum. Här styrs ljudförändringen således inte av kontexten utan av det enskilda ordet, vilket kräver en annan typ av regler i språkmodellen än de som är kontextberoende.

Eftersom Isofs uppdrag är att bedriva språkarbetet på vetenskaplig grund, samtidigt som det saknas en deskriptiv grammatik för meänkieli, måste reglerna i språkmodellen tills vidare vara så generösa att en framtida stavningskontroll som bygger på modellen tillåter flera stavningsvarianter.

7. Insamling av texter

Som tidigare nämnts har Isof ett ansvar att bedriva verksamheten på *vetenskaplig grund*. För att kunna göra detta behöver det finnas ett språkmaterial med hög kvalitet som verksamheten kan grunda sig på. Det vore önskvärt med en korpus som är så representativ som möjligt, som täcker flera genrer som skönlitteratur, tidningsprosa, sociala medier m.m. Ett problem Isof dock har mött rätt tidigt i framtagningen av en sådan korpus är att antalet skribenter av publicerad text på meänkieli är lågt.

Arbetet har därför varit tvunget att bli mer pragmatiskt, och fokusera på så stora textmängder som möjligt istället för att fokusera på en bred representation. Det innebär också att större flexibilitet behövs när det gäller källorna till korpusarna och vilka metoder som används i insamlingen av detta material.

Den låga mängden aktiva skribenter innebär att flera vanliga källor för textinsamling, såsom sociala medier och forum, inte är tillgängliga för meänkieli. Dessutom finns det inte en lika lång tradition av att skriva skönlitteratur på meänkieli som det gör för exempelvis svenska, vilket ytterligare begränsar tillgången på textmaterial från denna domän.

En av de största producenterna av texter på meänkieli idag är den offentliga förvaltningen, inklusive kommuner, regioner och myndigheter. En betydande del av denna textmängd är publicerad digitalt och finns spridd över flera olika hemsidor med varierande strukturer och format. Ofta finns det endast enstaka sidor skrivna på meänkieli (Meänkielelä – kiruna.se 2024), medan mer omfattande översättningar förekommer hos andra organisationer (Meänkieli (Tornedalsfinska) – arbetsformedlingen.se 2024).

Man behöver även ha i åtanke att i många fall förekommer dessa texter parallellt med texter på övriga minoritetsspråk. Detta är särskilt besvärande när det gäller finska, som oftast är svårt att automatiskt avskilja från meänkieli (Larsson 2024) och där det är viktigt att undvika att ta med finska texter i korpusen. Att få med finska texter i korpusen riskerar att skapa ett normativt tryck på meänkieli när denna korpus senare används i språkvårdsyften.

Denna fragmentering och variation av texterna på meänkieli utgör ytterligare en utmaning för insamling av material. I andra språk där det hade gått att samla in en stor och representativ korpus med hjälp av texter från några enstaka källor (t.ex. Familjeliv eller Flashback som hos Språkbanken Text) med samma struktur behöver Isof istället ta fram verktyg som kan både samla in brett (över flera olika hemsidor), djupt (texterna ligger långt inbäddade i hemsidorna) och med hög precision så att inte material på andra språk inkluderas.

Allt detta kräver att Isof utvecklar nya anpassningsbara metoder för textinsamling som tar med dessa förutsättningar i ett tidigt stadium av dess utveckling. Ett sådant arbete pågår idag, men innebär även att det inte alltid går att nyttja befintliga verktyg och metoder för textinsamling som är mer anpassade för större språk med en mer etablerad digital närvaro.

8. Avslutande ord

Utan språkteknologisk infrastruktur kan Isof inte utföra sitt uppdrag att bedriva språkvård på vetenskaplig grund för meänkieli, eller för den delen de andra nationella minoritetsspråken. Att kunna använda korpusar för att förbättra ordböcker kräver ändamålsenliga korpusverktyg. Detta förutsätter språkmodeller, som i sin tur förutsätter lexikografiska resurser. Detta kan ses som en grundläggande del av arbetet med standardisering och revitalisering av de nationella minoritetsspråken.

Litteratur

Ordböcker och digitala resurser

Korp för meänkieli – GiellaTekno/Språkbanken Text. <gtweb.uit.no/f_korp> (juni 2024).

Meän akateemis meänkieli-svensk/svensk-meänkieli digital ordbok (2024). <språk.isof.se/meänkieli/> (april 2024).

Meänkielelä – kiruna.se (2024). <kiruna.se/kommun--demokrati/vara -minoritetssprak/meankielela.html> (april 2024).

Meänkieli (Tornedalsfinska) – arbetsformedlingen.se (2024). <arbetsformedlingen.se/other-languages/meankieli-tornedalsfinska> (april 2024).

Suomen murteiden sanakirja (2020): *Kotimaisten kielten keskuksen verkkojulkaisuja*. <kaino.kotus.fi/sms> (april 2024).

Annan litteratur

Allén, Sture (1970): *Nusvensk frekvensordbok baserad på tidningstext 1 Graford Homografkomponenter*. Stockholm: Almqvist & Wiksell.

- Blokland, Rogier (2024): *Meänkieli som ett uraliskt språk. I: Föreläsning med anledning av meänkielidagen 27 februari*. Isaf.
- Borin, Lars, Markus Forsberg, Johan Roxendal (2012): *Korp – the corpus infrastructure of Språkbanken. I: Proceedings of LREC 2012*. Istanbul: ELRA. 474–478.
- Hammarstedt, Martin, Anne Schumacher, Lars Borin & Markus Forsberg (2022): *Sparv 5 user manual*. Göteborg: Göteborgs universitet.
- Karlsson, Fred (2012): *Finsk grammatik. Nionde, utökade och reviderade upplagan*. Helsingfors: Suomalaisen Kirjallisuuden Seura.
- Kenttä, Matti & Bengt Pohjanen (1996): *Meänkielen kramatiikki*. Övertorneå: Kaamos.
- Kuoppa, Harriet (2015): *Varumärket för meänkieli – användningen av h i efterstavelserna i skrift*. Umeå: Umeå universitet.
- Laakso, Johanna, Anneli Sarhimaa, Athanasia Spiliopoulou Åkermark & Reetta Toivanen (2016): *Towards openly multilingual policies and practices: Assessing minority language maintenance across Europe*. Bristol: Multilingual Matters.
- Larsson, Jacob (2024): *Language identification for typologically similar low-resource languages – case study of Meänkieli, Kven and Finnish*. Stockholm: Stockholms universitet.
- Lejdebros Enwald, Lina (2024): *En morfologisk modell för meänkieli*. Stockholm: Stockholms universitet.
- Mantila, Harri (1992): *Ei tääläkhän senthän jokhaishen sanhan hootakhan panna*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Mattson, Marie & Magnus Ahlthorp (2023): *Lexikografiska resurser betydelse i utvecklingen av språkteknologiska verktyg för minoritetsspråk. I: LexicoNordica 30, 75–94*.
- Moseley, Christopher & Alexandre Nicolas (2010): *Atlas of the world's languages in danger*. Paris: Unesco.
- Parkvall, Mikael (2016): *Sveriges språk i siffror: vilka språk talas och av hur många?* Stockholm: Morfem.

- Paunonen, Heikki (1987): De finska dialekterna på Nordkalotten och deras förhållande till de andra finska dialekterna. I: *Studia Historica Septentrionalia nr 14: Nordkalotten i en skiftande värld*. 211–237.
- Pirinen, Flammie, Sjur Moshagen & Katri Hiovain-Asikainen (2023): GiellaLT — a stable infrastructure for Nordic minority languages and beyond. I: *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*. Tórshavn: University of Tartu Library. 643–649.
- Pohjanen, Bengt (2022): *Meänkieli: grammatik, lärobok, historik, texter*. Överkalix: Barents.
- Ralph, Bo (2009): När ordboken blev en ordlista. I: Martin Gellerstam (red.): *SAOL och tidens flykt: några nedslag i ordlistans historia*. Stockholm: Norstedts.
- Söderholm, Eira (2017): *Kvensk grammatikk*. Oslo: Cappelen Damm Akademisk.
- Trosterud, Trond (2020): *Språkteknologi för meänkieli*. <giellalt.github.io/lang-fit/ rapport.pdf> (april 2024).
- Trosterud, Trond (2022): Normative Language Work in the Age of Machine Learning. I: Željko Jozić & Sabine Kirchmeier (red.): *The role of national language institutions in the digital age*. Budapest: Hungarian Research centre for Linguistics. 61–70.
- Valijärvi, Riitta-Liisa, Rogier Blokland, Elina Kangas, Constanze Ackermann-Boström & Harriet Kuoppa (2022): Meänkieli. *Linguistic Minorities in Europe Online*. doi:10.1515/lme.18469972.

Magnus Ahltop
 Språkteknolog
 Språkrådet, Institutet för språk och
 folkminnen
 Alsnögatan 7
 SE-116 41 Stockholm
 magnus.ahltop@isof.se

Lina Lejdebros Enwald
 Språkteknolog
 Språkrådet, Institutet för språk och
 folkminnen
 Alsnögatan 7
 SE-116 41 Stockholm
 lina.lejdebros.enwald@isof.se

Elina Kangas
Språkvårdare i meänkieli
Avdelningen för nationella
minoriteter och svenskt
teckenspråk, Institutet för språk
och folkminnen
Alsnögatan 7
SE-116 41 Stockholm
elina.kangas@isof.se

Rickard Domeij
Språkteknolog
Språkrådet, Institutet för språk och
folkminnen
Alsnögatan 7
SE-116 41 Stockholm
rickard.domeij@isof.se

Jacob Larsson
Språkteknolog
Avdelningen för nationella
minoriteter och svenskt
teckenspråk, Institutet för språk
och folkminnen
Alsnögatan 7
SE-116 41 Stockholm
jacob.larsson@isof.se

Gunnar Eriksson
Språkteknolog
Språkrådet, Institutet för språk och
folkminnen
Alsnögatan 7
SE-116 41 Stockholm
gunnar.eriksson@isof.se