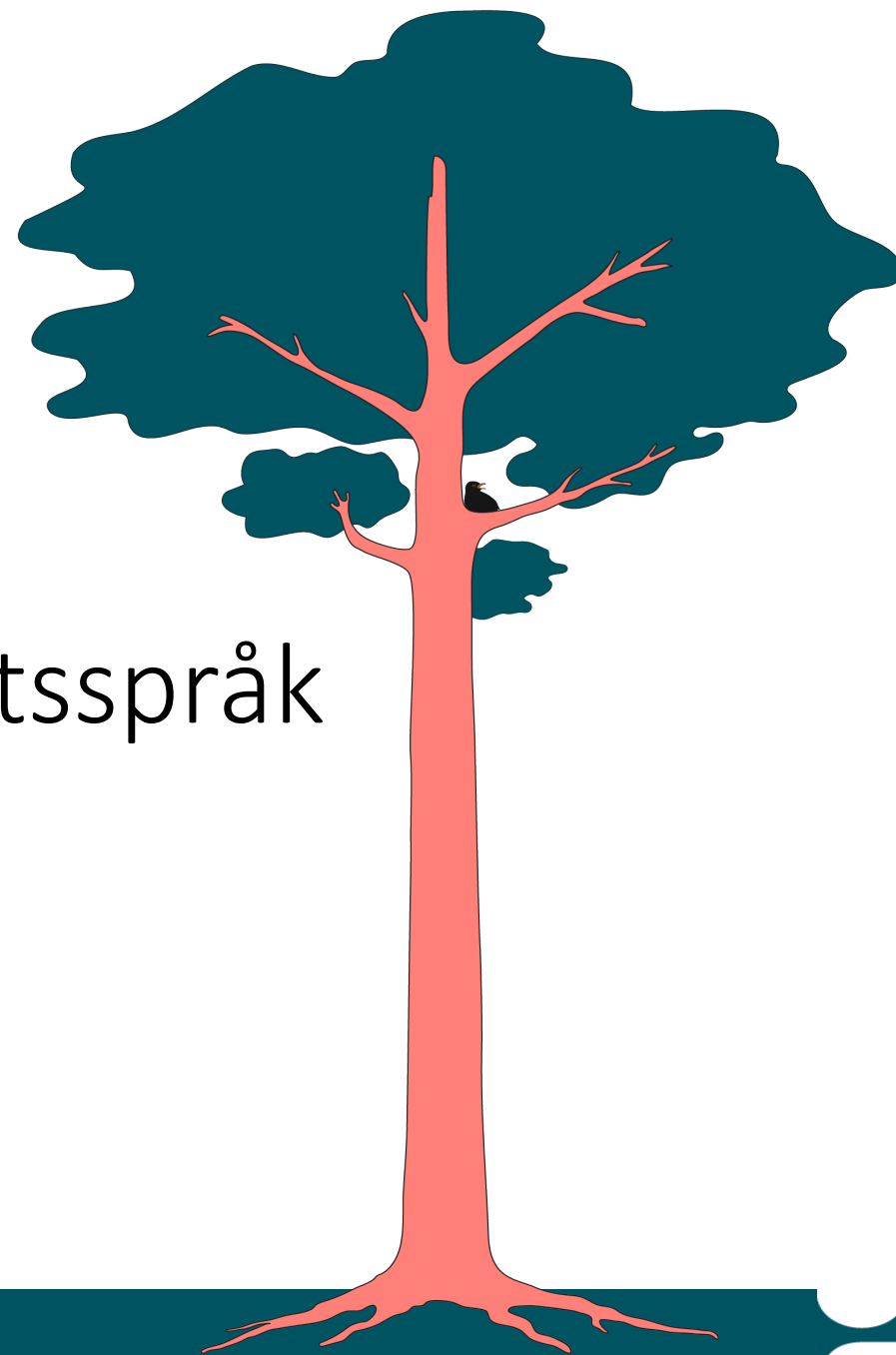


# Språkteknologi för meänkieli och andra nationella minoritetsspråk

Rickard Domeij, Elina Kangas & Jacob Larsson

Språkbanken Sam @isof.se



# Vi bygger infrastruktur för framtidens forskning

Nationella språkbanken arbetar för att bygga upp en svensk e-infrastruktur för forskning i språkteknologi, språkvetenskap och andra ämnesområden som bedriver forskning baserad på språklig data.



# Välkommen till Små språk i Norden!

Med projektet Små språk i Norden vill vi arbeta för en förändring så att de små språken kan leva vidare och berika den språkliga mångfalden i Norden.



# Språkteknologiskt initiativ

1. Uppmärksamma betydelsen av språkteknologi för de små språken i Norden
2. Inkludera de små språken i Norden i arbetet med European Language Equality
3. Arbeta för att öppna techjättarnas plattformar för språkteknologi för små språk

## Konkreta insatser:

- Rapport med kartläggning och strategi
- Utbildning av ambassadörer och slutanvändare
- Webbportalen Borealium
- Nordisk konferens med fokus på alla språk



**D1.38**

**Report on the Nordic  
Minority Languages**

Authors	Sjur Nørstebø Moshagen, Rickard Domeij, Kristine Eide, Peter Juel Henriksen, Per Langgård
Dissemination level	Public
Date	09-05-2022

# Nationella minoritetsspråk

Språkcentrum och språkvård för  
finska, jiddisch, meänkieli och romska

Meänkieli

Romani  
chib

יידיש

Suomi

**Finska** >

Språkcentrum och  
språkvård för finska.

**Jiddisch** >

Språkcentrum och  
språkvård för jiddisch

**Meänkieli** >

Språkcentrum och  
språkvård för meänkieli

**Romska** >

Språkcentrum och  
språkvård för romska

# Isofs uppdrag

1 §

Institutet för språk och folkminnen har till uppgift att bedriva språkvård och på vetenskaplig grund öka, levandegöra och sprida kunskaper om språk, dialekter, folkminnen, namn och immateriella kulturarv i Sverige.

# Vad är språkvård?

- Språkvård kan se mycket olika ut beroende på vilket språk man arbetar med, i vilket språksamhälle och på vilka nivåer.
- Det kan handla om att reglera i vilka situationer olika språk ska användas, att skapa inlärningsmöjligheter, att motarbeta negativa attityder mot enskilda språk, att verka för mer begripliga texter eller att sprida kunskap om ords böjning och betydelse.

Vi ansvarar för språkvården av meänkieli i Sverige. Vi ger språkliga råd och rekommendationer.



Institutet för  
språk och  
folkminnen

Kielen ja  
kansanperintheen  
instityytti

# Språkvård för meänkieli

- Olika aktörer inom språksamhället har jobbat med språkfrågor i meänkieli: aktiva författare, skribenter, lärare, översättare
- Första ordboken Kenttä & Wandé 1992






# Språkvård för meänkieli

- Olika aktörer inom språksamhället har jobbat med språkfrågor i meänkieli: aktiva författare, skribenter, lärare, översättare
- Första ordboken Kenttä & Wandé 1992



# Test our spellers

Type or paste your text in the editor window:



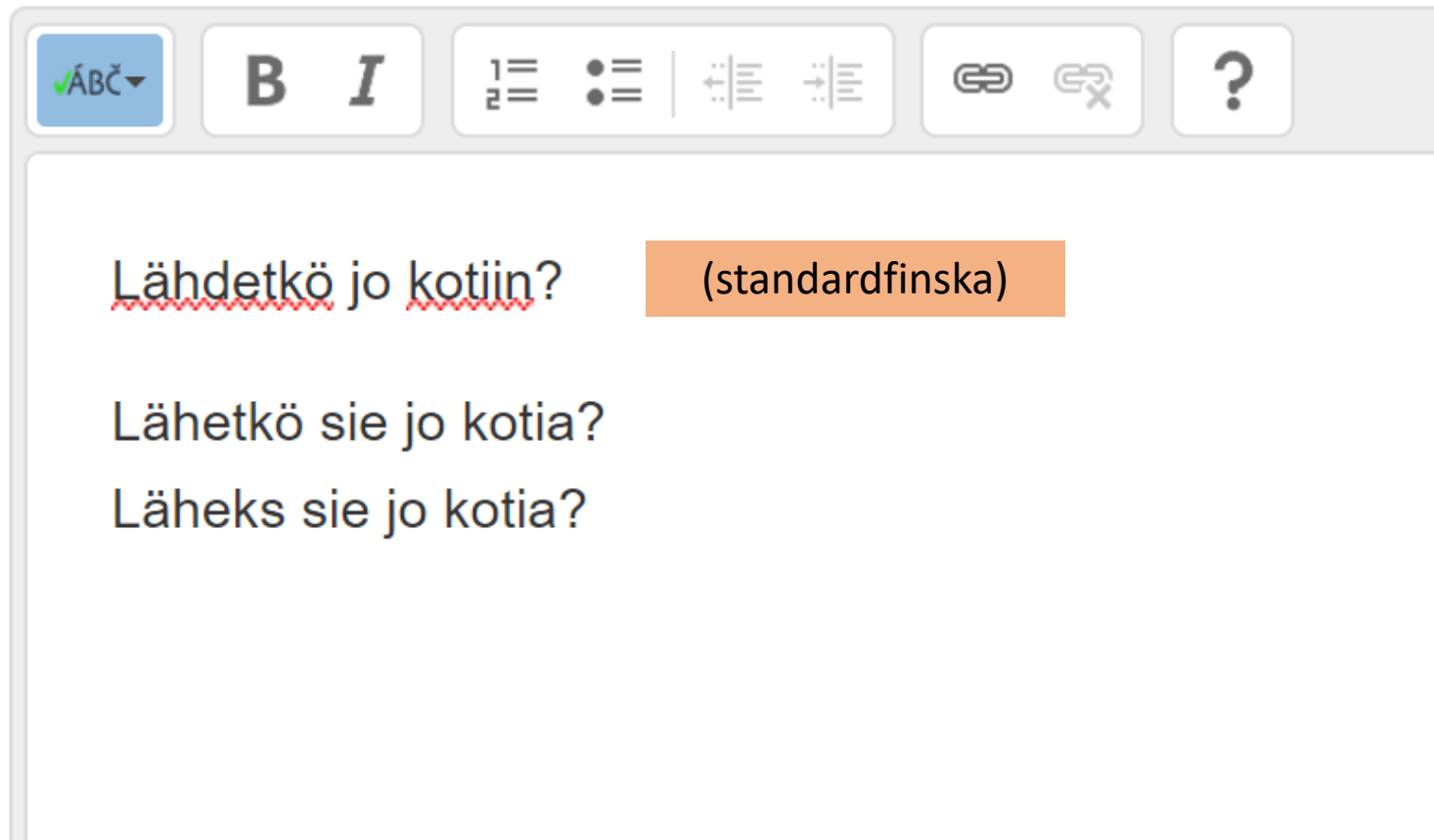
Lähdetkö jo kotiin?

Lähetkö sie jo kotia?

Läheks sie jo kotia?

# Test our spellers

Type or paste your text in the editor window:



The image shows a text editor window with a toolbar at the top. The toolbar contains several icons: a blue button with a green checkmark and 'ABC' (spell checker), buttons for bold (B) and italic (I), buttons for numbered and bulleted lists, buttons for indenting and outdenting, a link icon, an unlink icon, and a question mark icon. Below the toolbar, the text area contains three lines of text. The first line is 'Lähdetkö jo kotiin?' with red wavy underlines under 'Lähdetkö' and 'kotiin?'. To the right of this line is an orange box containing the text '(standardfinska)'. The second line is 'Lähetkö sie jo kotia?' and the third line is 'Läheks sie jo kotia?'.

✓ABC

**B** *I*

☰ ☱ | ☲ ☳

🔗 🔗✕

?

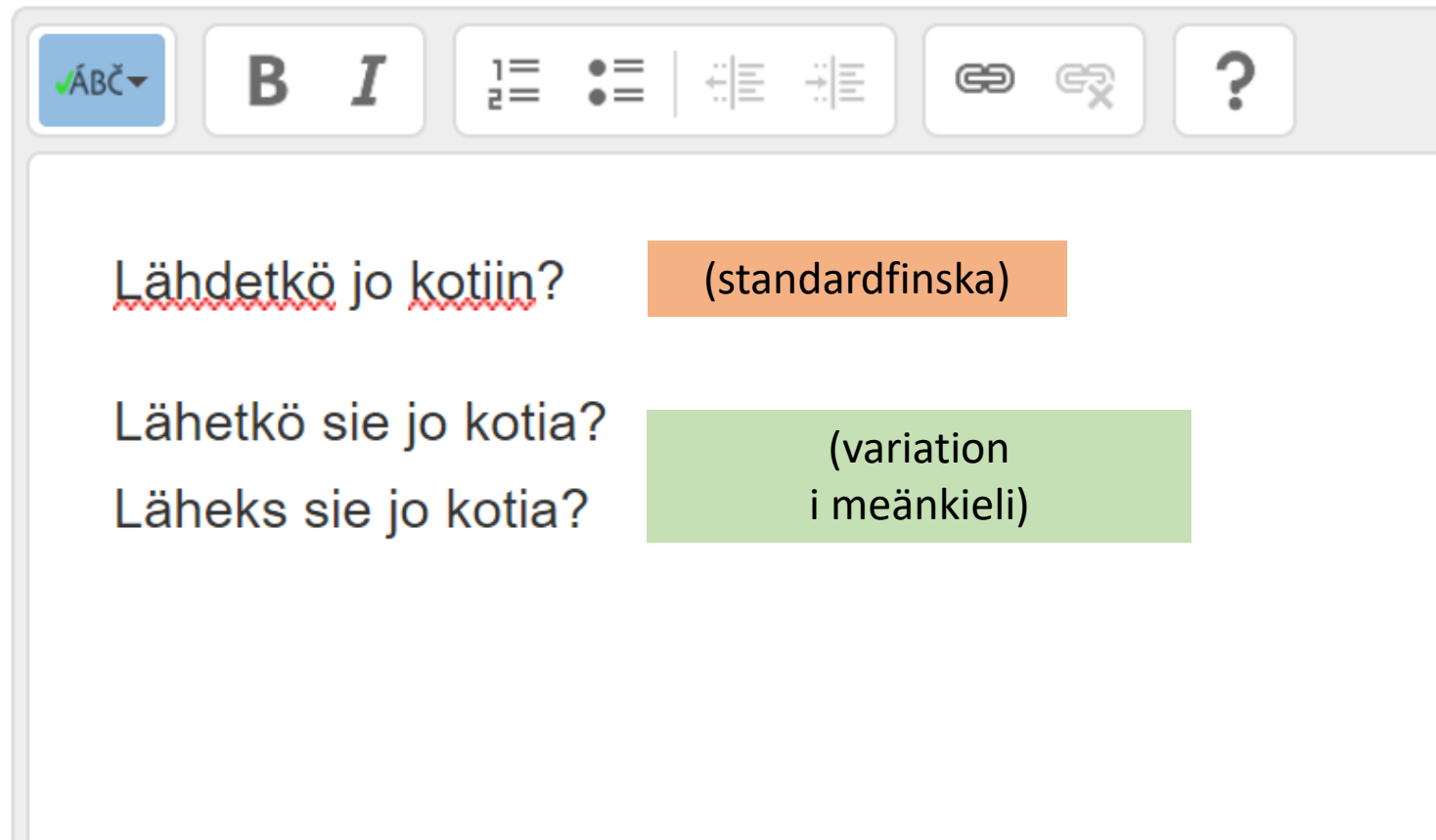
Lähdetkö jo kotiin? (standardfinska)

Lähetkö sie jo kotia?

Läheks sie jo kotia?

# Test our spellers

Type or paste your text in the editor window:



The image shows a text editor window with a toolbar at the top. The toolbar contains icons for spell checking (ABC), bold (B), italic (I), bulleted list, numbered list, indent left, indent right, link, unlink, and help (?). Below the toolbar, the text "Lähdetkö jo kotiin?" is displayed with red wavy lines under "Lähdetkö" and "kotiin?". To the right of this text is an orange box containing "(standardfinska)". Below this, the text "Lähetkö sie jo kotia?" is shown. Below that, the text "Läheks sie jo kotia?" is shown. To the right of these two lines is a green box containing "(variation i meänkieli)".

Stavningsprogram är ett extremt kraftigt verktyg för standardiseringsarbete. Om man vill normera ett skriftspråk är det väldigt svårt utan ett sånt program.

Ett bra stavningsprogram kan ge en ny generation av språkbrukare den tryggheten de behöver för att börja skriva på meänkieli.

(Trosterud 2020)



# Samarbeten



**UiT** Norges arktiske universitet



Trond Trosterud, UiT

# Arbete med teknik- och språkresurser för meänkieli på Språkrådet

- E-ordböcker
- Text- och talkorpusar
- Språkanalys
- Stavningskontroll
- Informationshantering
- Datorstödd språkinlärning
- Dialogsystem
- Taligenkänning
- Talsyntes
- Maskinöversättning

## Ordbok meänkieli-svenska

Meänkieli-svensk-meänkieli ordbok är en digital ordbok som består av cirka 33 000 ord med översättningar från meänkieli till svenska och från svenska till meänkieli. Orden är märkta med ordklass. I vissa fall finns exempel och geografisk märkning.

Till Ordbok meänkieli – svenska – meänkieli

### Om ordboken

Meänkieli är ett av fem nationella minoritetsspråk i Sverige. Meänkieli består framför allt av tre dominerande varieteter; tornedalsvarieteteten, gällivarevarieteteten och lannankieli.

Ordboksredaktionen har bestått av teol. kand. Anders Alapää, lärarna och kulturarbetarna Linnea Nylund, Märta Nylund och Birgitta Rantatalo samt prof. emer. och ordförande Erling Wande. Samtliga har sin främsta meänkielbakgrund i de centrala delarna av svenska Tornedalen. Några av redaktionsmedlemmarna medverkar även i Luleå stifts och Svenska bibelsällskapets arbete med översättning av hela Bibeln till meänkieli. Under en kortare period har även avlidne Hans Lindmark (Kieruna), Barbro Rantatalo (Korpilombolo) och Eva Kvist (Nuksjärvi, f.d. medarbetare vid Meänraatio) varit delaktiga i arbetet med material i ordboken. Harriet Kuoppa, tidigare forskningsarkivarie vid Dialekt-, ortnamns- och folkminnesarkivet i Umeå (DAUM), har också bistått i arbetet med att hitta material och förmedla kontakter främst med lannankieli-talande.

För den tekniska sidan av produktionen har vi fått ovärderlig hjälp av professor Trond Trosterud och chefstekniker Ciprian Gerstenberger vid Giellatekno, Tromsø universitets specialinstitut för språkteknologi gällande främst samiska och andra finsk-ugriska språk. Institutet har specialkompetens med avseende på produktion av datoriserade ordböcker och läromedel. Giellatekno har tidigare producerat ett antal ordböcker för samiska varieteter och en ordbok för kvänska, ett språk som är nära besläktat med meänkieli.

Ordboken fick ett ekonomiskt startbidrag av Statens Kulturråd, men större delen av arbetet har finansierats av Institutet för språk och folkminnen, Isof. En stor del av arbetet har även genomförts på ideell basis.

Ordboksredaktionen vill tacka alla som bidragit med att ge information om ord, varieteter och som varit delaktiga i arbetet med att skapa denna ordbok.

## **Textinsamling**

- Arkivmaterial
- Material från myndigheter och kommuner
- Eget publicerat material
- Litteratur
- Media

## **Språkanalys**

- Språkmodell
  - Morfologisk analys
  - Syntaktisk analys
  - Lexikal analys

## **Språkstöd**

- Ordböcker
- Tangentbord
- Textsökning
- Korpinstans
- Stavningskontroll
- Talteknologi



# Språkanalys

- Vad ska den åstadkomma?
  - Lexikal analys
  - Morfologisk analys
  - Syntaktisk analys
- Förutsättningar
  - Lite data
  - Saknar grammatikbeskrivning
  - Lexikon
- Egenskaper
  - Regelbaserad
    - Finite state
  - Analys, generering
- Nackdelar / fördelar



```
"o copying for h metathesis"
!! **RULE: @RULENAME@**
%V:o <=>
o: [%^HMETA:|^HMET2:] %> h: _ [:i|o] ;
o: [%^HMETA:|^HMET2:] %> h: :o _ ;
o %> (h) _ [n|t|s|l|%>|.#.] ;
[. #. |#] Cns: [a|e|i] (Vow:) (:h) Cns:* a: [%^HMETA:|^HMET2:] %> h: _ :i ; !saunhoin/saunhoin
[. #. |#] Cns: [a|e|i] Cns:* (:h) (Cns:) a: [%^HMETA:|^HMET2:] %> h: _ [:i|i2:] ; !kernas>kernhoin/kerhoin
[#|. #.] (Cns:*) Vow:+ Cns:+ Cns:+ Vow:+ Cns: Cns: (Cns:) a: [%^HMETA:|^HMET2:] %> (h:) _ [:i|i2:] ; !plural 3/fierstaviga nouns
! [. #. |#] Cns:* UnroundedVowel:+ u Cns:+ Vow: (%^HMETA:|^HMET2:) %> h _ [i2:] [i3:] [i8:] [j|i:] ;

!! **Tests:**

!!€ talo^HMETA>h^V^Vn
!!€ tal00>hoon

"i copying for h metathesis"
!! **RULE: @RULENAME@**
%V:i <=>
i: [%^HMETA:|^HMET2:] %> h: _ :i ;
i: [%^HMETA:|^HMET2:] %> h: :i _ ;
:i %> (h) _ [n|t|s|l|%>|.#.] ;
[. #. |#] Cns:* RoundedVowel+ Cns:+ (:Cns) a: (%^WG:|^HMETA:|^HMET2:) %> h: _ [i2:] ; !muna>munhiin Pl+I11
```

morphology test page

takki

Analyse Generate fit Reload Model

Results

Token	Lemma	Tags
takki	takki	N Sg Nom

morphology test page

takki

Analyse Generate fit Reload Model

Results

Class: nx  
Examples: maafiskaali, patruuni, tuska, liiteri, pilleri

	Sg	Pl
Nom	takki	takit
Gen	takin	takien,takitten
Ill	takkia	takhiin
Ine	takhiin	takissa
Ela	takissa	takista
All	takista	takile
Ade	takile	takila
Abl	takila	takilta
Tra	takilta	takiksi
Abe	takkina	takitta

# Textinsamling

- Källor
  - litteratur
  - media
  - myndighetstexter
  - arkivmaterial
- Manuell och automatisk insamling
- c.a 600 000 löpord idag
- Vägledning och testning för språkanalys
- Sökbar textsamling via Korp

**Meänkieli** Institutet för språk och folkminnen

Vårt uppdrag Stöd och språkråd Arkiv och insamling Nationella minoritetsspråk Lär dig mer På gång

Meänkieli  Meänkieli  Suomi  Romani arli  Romani kalderaä  Davvisámegiella  Bihánsámegiella  Aarjelsáemien  Julevsámegiella

Startsida / Meänkieli / Sinun oikeus meänkielien / Sinun oikeus kansalisiin vähemmistökiehliin

## Ette äänestää

Pyhänä 11 syyskuuta 2022 oon vaalit valtiopäivile, kunnan- ja rekiuunivaltuusthoon.

### Kukas saapi äänestää?

Ette olla äänestysoikeutettu sie piät olla täyträny 18 vuotta viimisthääns vaalipäivänä. Tiot kansankirjarekisterissä 30 päivää ennen vaalipäivää päättävät jos sulla oon äänestysoikeus ja mihinkä vaalistriktiin sie kuulut.

**Sulla oon äänestysoikeus valtiopäivile jos sie olet:**

- ruottalainen kansalainen ja olet elikkä olet ollu kansankirjoila Ruottissa.

**Sulla oon äänestysoikeus kunnan- ja rekiuunivaltuusthoon jos sie olet:**

- ruottalainen kansalainen ja kansankirjoila Ruottissa
- kansalainen jossaki Eurooppayniuuriin jäsenmaassa elikkä kansalainen Iislandissa elikkä Norjassa ja olet kansankirjoila Ruottissa.
- kansalainen jossaki muussa maassa ja olet ollu kansankirjoila Ruottissa kolme vuotta peräkkää ennen vaalipäivää.

## Sinun oikeus kansalisiin vähemmistökiehliin

Kaikila viranomaisilla Ruottissa oon pakolista antaa tietoa kaikile kansalisille vähemmistöile heän kielellisistä oikeuksista, niin kans jiddisch ja roominpuhujille.

Tämän tekstin kautta sie saat tietää mikkä sinun kieleliset oikeuet oon ja mitä sie saatat vaatia sinun kunnalta ja muilta viranomaisilta.

Kansalisiin vähemisthoin räknähään juutalaiset, roomit, saamet, ruottinsuomalaiset ja tornionlaaksoalaiset. Kansaliset vähemmistökielet oon suomi, jiddisch, meänkieli, roomi (romani chib) ja saame.

**RÖSTKORT**


Armonda Evelina Erlund  
GÄLLIVARE, Ullatti f. 1897

0 Vet bara om det sen jag började förstå  
"Varit friskt - roligt - har ej väsk - sores bra"

# Språkstöd

- Stavningskontroll
- Tangentbord
- Sökning texter – ord
  - Korp
  - Meänkieliordboken
- Ordböcker
- Textanalysverktyg
  - Lemmatisering
  - POS-tagging
- Talteknologi
  - Textuppläsning
  - Talinmatning

Kveenin tekstit | Meänkielen tekstit | Vepsän tekstit | Võron tekstit | Fäärin tekstit | Livvin tekstit | Rinnakkaistekstit

 **KORP**

7 korpusta valittuina (kaikki) — 447,46K/447,46K sanetta

Yksinkertainen Laajennettu Edistynyt compare

Etsi

Meän Akateemi  
Academia Tornedaliensis

Meänkieli ↔ ruotti sanakirja

Meänkieli ⇄ Svenska

Hae (vähintänsä kaks puustavia)

kieliteknologinen a språkteknologisk

# Kiitos! Tack!



Institutet för  
språk och  
folkminnen

Kielen ja  
kansanperintheen  
instityytti

[meankieli@isof.se](mailto:meankieli@isof.se)