

Digital humanities for minority languages – against a couple of odds

Chiara Argese and Trond Trosterud
Giellatekno, Centre for Saami Language Technology
<http://giellatekno.uit.no/>



November 30, 2017

Contents

Introduction

Digital humanities

 The sparseness of data

 The complexity of the signal

Language technology infrastructure in Tromsø

 Some case studies

 Language technology for whom

Conclusion

Introduction

- ▶ We will
 - ▶ reflect over digital humanities
 - ▶ discuss some of the problems faced by minority languages
 - ▶ and give some examples of how they may be overcome

Digital humanities is collective work

- ▶ The **Giellatekno** research group at UiT
 - ▶ Chiara Argese, Lene Antonsen, Ciprian Gerstenberger, Trond Trosterud,

Digital humanities is collective work

- ▶ The **Giellatekno** research group at UiT
 - ▶ Chiara Argese, Lene Antonsen, Ciprian Gerstenberger, Trond Trosterud,
- ▶ The **Divvun** group
 - ▶ funded by the Ministry of Local Government and Modernisation
 - ▶ Børre Gaup, Maja Kappfjell, Sjur Moshagen, Thomas Omma, Elena Paulsen, Tomi Pieski, Kevin Unhammer, Linda Wiechetek,

Digital humanities

Digital humanities

- ▶ Was sind eigentlich *digitale Geisteswissenschaften*?

Digital humanities

- ▶ Was sind eigentlich *digitale Geisteswissenschaften*?
 - ▶ the use of computer-supported methods,
 - ▶ the systematic use of digital resources
 - ▶ and the reflection over their use

Digital humanities

- ▶ Was sind eigentlich *digitale Geisteswissenschaften*?
 - ▶ the use of computer-supported methods,
 - ▶ the systematic use of digital resources
 - ▶ and the reflection over their use
- ▶ Practitioners are researchers within
 - ▶ computer philology
 - ▶ digital historiography
 - ▶ information science
 - ▶ computational linguistics

Problem 1: The sparseness of data

Problem 1: The sparseness of data

- ▶ Norwegian
 - ▶ Every published book and newspaper is electronically available
 - ▶ Historical data are available in machine-readable form

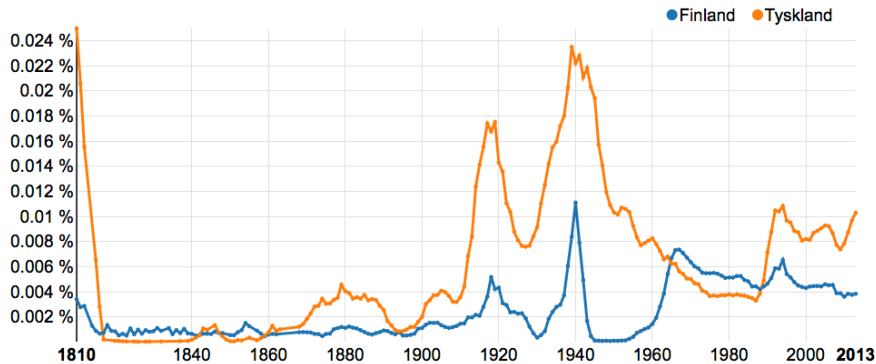
Problem 1: The sparseness of data

- ▶ Norwegian
 - ▶ Every published book and newspaper is electronically available
 - ▶ Historical data are available in machine-readable form
- ▶ Swedish
 - ▶ 11 billion words of running text is available, with a grammatical analysis
 - ▶ The text is classified according to source type, topic and year

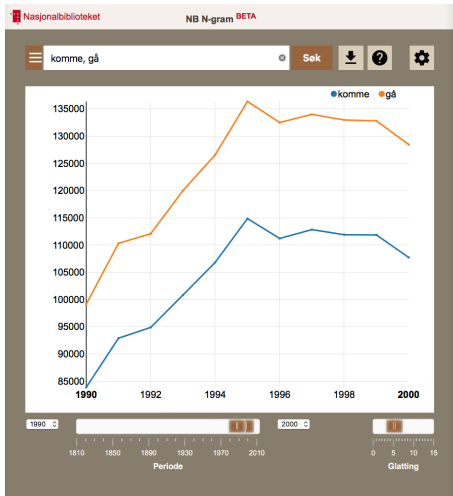
Problem 1: The sparseness of data

- ▶ Norwegian
 - ▶ Every published book and newspaper is electronically available
 - ▶ Historical data are available in machine-readable form
- ▶ Swedish
 - ▶ 11 billion words of running text is available, with a grammatical analysis
 - ▶ The text is classified according to source type, topic and year
- ▶ North Saami
 - ▶ UiT: **SIKOR** Sámi Internationála KORpus
 - ▶ North Saami 25 mill words (almost no fiction), other Saami languages appr 1 mill words each)

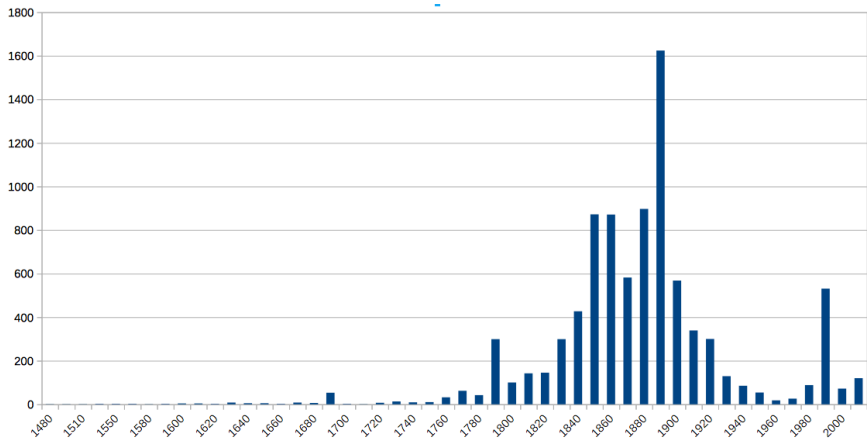
Nasjonalbiblioteket gives us a temporal dimension



... but not for the Saami languages



Not all is freely available, though





125 av 232 korpusar valda — 2,08G av 11,65G token

Enkel Utökad Av

även som förled efterled

KWIC: träffar per sida: 25

Akademiska texter (2)
 August Strindberg (2)
 Finlandssvenska texter (56)
 Medicinska texter (13)
 Myndighetstexter (24)
 Skyddade korpusar (12)
 Skönlitteratur (6)
 Sociala medier (64)

- Bloggmix (21)
- Diskussionsforum (39)
 - Familjeliv (23)
 - Familjeliv: Adoption
 - Familjeliv: Allmänna rubriker – Ekonomi & juridik
 - Familjeliv: Allmänna rubriker – Familjeliv.se
 - Familjeliv: Allmänna rubriker – Fritid & hobby
 - Familjeliv: Allmänna rubriker – Hus & hem
 - Familjeliv: Allmänna rubriker – Husdjur
 - Familjeliv: Allmänna rubriker – Kropp & själ

Markera alla
 Avmarkera

Visa statistik Visa ordbild

Familjeliv: Allmänna rubriker – Familjeliv.se
 Antal token: 62 459 449
 Antal meningar: 5 195 388
 Senast uppdaterad: 2017-09-19

Korp for North Saami

KWIC
Statistics
Word picture
map

Results: 60,644

Dependency Tree

AdvL → ADVL → V → ADVL → ADV → N → N → N → P → CLB

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale
bohccot
ráji
badjel
.

Dákkko
bohtet
álohii
ruotabeale

Problem 2: The complexity of the signal

The English verb: walk – 4 word forms: add -s, -ed, -ing
walk, walks, walked, walking

The English verb: walk – 4 word forms: add -s, -ed, -ing
walk, walks, walked, walking

**North Sámi verb: vázzit – 59 word forms,
with complex stem changes:**

váccán, váccát, vázzá, vázze, vázzibeahhti, vázziba, vázzit,
vázzibehtet, vázzet, vácce, vázzen, vázzet, váccii, vácciime,
vácciide, vácciiga, vácciimet, vácciidet, váccášin, váccášit, váccášii,
váccášeimme, váccášeidde, váccášeigga, váccášeimmet,
váccášeiddet, váccášedje, váccáše, vázzon, vázzos, vázzu, vázzi,
vázzoska, vázzot, vázzoset, váccežan, váccežat, vácceža,
váccežetne, váccežeahppi, váccežeaba, váccežit, váccežehpet,
vácceš, vázzán, váccážit, vázzime, vácci, váccikeahtá, váccidettiin,
váccidettiinan, váccidettiinat, váccidettiinis, váccidettiineame,
váccidettiineatte, váccidettiineaskka, váccidettiineamet,
váccidettiineattet, váccidettiineaset

... and then come the derivations

Grammar modelling

- ▶ To analyse circumpolar languages we must be able to analyse their word forms

Grammar modelling

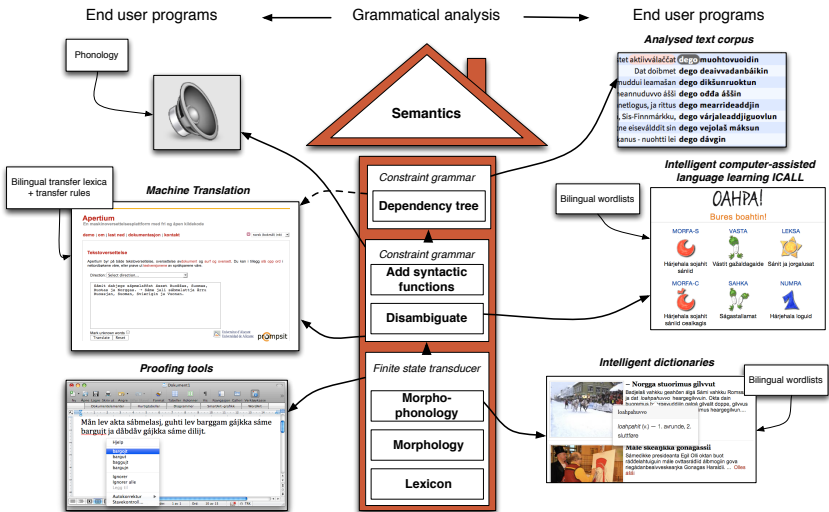
- ▶ To analyse circumpolar languages we must be able to analyse their word forms
- ▶ The methods commonly in use for the majority languages do not work
 - ▶ A solution must be based upon the grammatical structure of the languages themselves

Generating paradigms

lávka	lávka+N+Sg+Nom: lávka	eadni+N+Sg+Nom: eadni	...
eadni	lávka+N+Ess: lávka n	eadni+N+Ess: eadni n	
...	lávka+N+Sg+Ill: lávka ii	eadni+N+Sg+Ill: eadni ái	
	lávka+N+Sg+Acc: lávka ka	eadni+N+Sg+Acc: ead ni	
	lávka+N+Sg+Loc: lávka kas	eadni+N+Sg+Loc: ead ni s	
	lávka+N+Pl+Nom: lávka kat	eadni+N+Pl+Nom: ead ni t	
	lávka+N+Pl+Ill: lávka kaide	eadni+N+Pl+Ill: ead ni ide	
	lávka+N+Sg+Com: lávka ka in	eadni+N+Sg+Com: ead ni in	

Saami language technology at the University of Tromsø – Giellatekno and Divvun:

Linguistics in practice



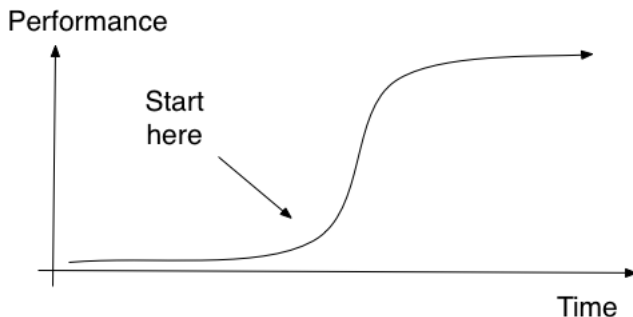
Language technology infrastructure in Tromsø

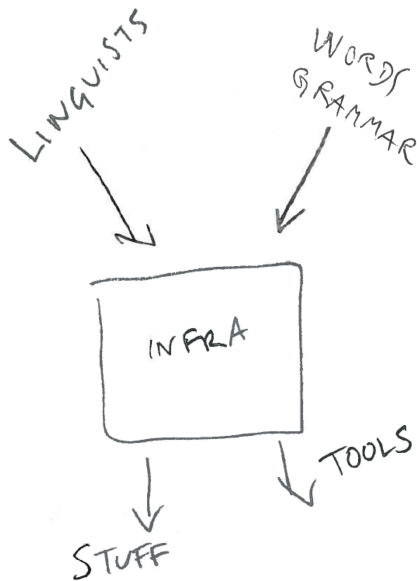
Language technology infrastructure in Tromsø

- ▶ The infrastructure is built as open source code
- ▶ It is language-independent
- ▶ The language model is converted into user programs

Language technology infrastructure in Tromsø

- ▶ The infrastructure is built as open source code
- ▶ It is language-independent
- ▶ The language model is converted into user programs





International cooperation

Research group	Language	(Language)
Giellatekno + Divvun (UiT)	8	9
Research groups using our servers:		
Børselv ▶ Kvena instituhtta	1	
Helsinki ▶ AKU	8	6
Tartu, Estonia ▶ SAMEST	2	1
Syktyvkar, Russia ▶ FU-lab	1	
Ulan-Ude, Russia ▶ BSU	1	5
UAlberta, Canada ▶ ALTLab	2	4
Nuuk ▶ Oq	1	
All in all appr. 50 languages:	24	25

Language technology

Language technology

Keyboards, hyphenation, spellchecking, grammar checking,
e-learning, e-lexicography, terminology work,
machine translation, information retrieval,
synthetic speech, speech recognition,
data and opinion mining, information retrieval, ...

Language technology

Keyboards, hyphenation, spellchecking, grammar checking,
e-learning, e-lexicography, terminology work,
machine translation, information retrieval,
synthetic speech, speech recognition,
data and opinion mining, information retrieval, ...

We are increasingly dependent upon language technology

Language technology

Keyboards, hyphenation, spellchecking, grammar checking,
e-learning, e-lexicography, terminology work,
machine translation, information retrieval,
synthetic speech, speech recognition,
data and opinion mining, information retrieval, ...

*We are increasingly dependent upon language technology
especially in multilingual contexts*

Why especially in multilingual contexts?

Why especially in multilingual contexts?

- ▶ Publishing in multiple languages is cumbersome

Why especially in multilingual contexts?

- ▶ Publishing in multiple languages is cumbersome
we need the computer to help us keeping track of the
language versions and their content

Language technology for whom

- ▶ Linguists?
- ▶ Translators?
- ▶ Terminologists?
- ▶ Students?
- ▶ Ordinary language users (both L1 and L2 users)?

Translation to Norwegian

"mun"
Pron Pers Sg1 Loc <hab> @ADVL>

"leat"
V IV Ind Prs Sg3 @+FMAINV

"bárdni" N Sem/Hum Sg Nom <ext> @<SUBJ

<ext> : indef. + artihkal
<ext> : det

<hab> : har, - det

loc : på, i, av, fra

mun > jeg leat > være/har

bárdni > sønn/gutt

Mus lea bárdni.
Šiljus lea bárdni.

Jeg har en sønn.
På gårdsplassen er det en gutt.

The rules govern lexical selection, word order, addition of pro form, etc.

Grammar-based machine translation that actually works

http://samas.no/se/oktavuohta

originalside Maskinoversatt av Gielletekno

På samisk Norsk (beta) English

HIEM STUDIER FOR STUDENTER BIBLIOTEK FORSKNING **OM SH** Q

Sámi allaskuvla
Sámi University of Applied Sciences

Om høyskolen

INSTITUTT FOR SPRÅKFAG
Institutt for språk gjennomfører samisk språket og fremgangen til samisk litteratur i forskning og undervisning.

INSTITUTT FOR SAMFUNNSFAG
Institutt for samfunn gjennomfører forskning og undervisning i reindrift, journalistikk og urfolkvitenskap.

INSTITUTT FOR DUODJI- OG LÆRERUDANNINGER
Institutt for duodji- og lærerutdanninger gjennomfører undervisningen og forskningen i duodji og lærerutdanninger.

ORGANISASJON +

STYRER OG KOMITEER +

FAGFORENINGER +

ANSATTE

ROMKART

DOKUMENT +

om SH

Samisk undervisning – Samisk forskning – Samisk institusjon 1989 - 2017

Samisk høyskoles offentlig åpning skjedde den 1. november 1989. Slik gjennomførtes det mellom samer lengre tid ønske å tilby samisk lærerutdanning - og etter dette annen også høyere nivå studier – på samisk i et samiskspråklig miljø. Særskilte samisk høyskole gav også muligheten å utvikle tilbudene som bedre svare til behov i samisk samfunn. Nå har 28 år sist og høsten 2017 er de på Samisk høyskole 270 studenter.

Neahttadigisánit for 27 languages

- ▶ The Nordic countries
 - ▶ North Saami: <http://sanit.oahpa.no>
 - ▶ South Saami: <http://baakoeh.oahpa.no>
 - ▶ Inari Saami: <http://saanih.oahpa.no>
 - ▶ Skolt Saami: <http://saan.oahpa.no>
 - ▶ Northern Baltic Finnic: <http://sanat.oahpa.no>
 - ▶ Southern Baltic Finnic: <http://sonad.oahpa.no>
- ▶ Russia
 - ▶ Mordvin languages: <http://valks.oahpa.no>
 - ▶ Nenets: <http://vada.oahpa.no>
 - ▶ Mari languages: <http://muter.oahpa.no>
 - ▶ Komi and Udmurt: <http://kyv.oahpa.no>
- ▶ Canada
 - ▶ Plains Cree: <http://itwewina.oahpa.no>
 - ▶ North Haida: <http://guusaaw.oahpa.no>



About the dictionaries

- ▶ Free: all dictionaries and other analytical tools are free
- ▶ Open source: everything is based on open-source tools
- ▶ Linguistic basis: the dictionary analyze the word you write to break apart compound words and find base forms

The dictionary "understands" inflected forms

South Sámi → Norwegian (⇌ Swap)

iehtsedh (v., i)

1. elske, være glad i

Manne datnem eahtsam.
Jeg elsker deg.

2. engste seg for noen

Datneste eahtsam.
Jeg er engstelig for deg.

Neahttadigisánit

Nordsamisk (#SoMe) ↔ Norsk (⇌ Snu)

čáhci (s.) –

1. vann, vatn

cazi er en mulig form av ...

čáhci

subst. ent. akk.

subst. ent. gen.

It (will) show you derived words

Neahttdigisániit

[Home](#)[Plugins](#)[About](#)

Dictionaries

[North Sámi → Finnish](#)[North Sámi → Norwegian](#)

WRITTEN VARIANT

[Standard \(áčđŋšž\)](#)[Tolerant \(incl. acđnstž\)](#)[Norwegian → North Sámi](#)[Finnish → North Sámi](#)[Other North Saami dictionaries](#)[Other dictionaries](#)

North Sámi (Standard) → Norwegian (⇌ Swap)

á č đ ŋ š ž

á ohcansátni

Search

Search texts

geavahuvvogoahhtá is a possible form of ...

geavahuvvogoahhtá**geavahit (verb)**

- (verb) bruke

Galggašii geavahit daid riggodagaid mat mis leat, dan sajis go viežžat importagullid Lulli-Amerihkás.

Vi bær hæste av de ressurser som vi har, istedenfor å importere fisk fra Sør-Amerika.

geavahit [Texts →](#)

geavahit verb passiv verb begynne å verb indikativ presens 3.p.ent.

- + ordavledning –uvvot
 - passivform av verbet
- + ordavledning –goahtit (inkoativ)
 - begynne å X

It makes it possible to click in text

<http://sanit.oahpa.no/read/>


Sámi riikajoavku galggai čiekčat **čájáhusčiekčamiid** Salašvákki
 spábbačiekčanjoavkku (Tromsø) ja sin presideanttain, čiekčat
 Salašvággi šattai čiekčat čájáhusčiekčamiid avkk
 Sámi riikajoavku ii bohta čájáhus (s.) — utstilling
 Sámi studeanttaid searvi Riikajoavku (SSL) ja sin presideanttain, čiekčan (s.) — fotballkamp Spáb
 čájáhusčiekčama Salašvákk čiekčat (v.) — spenne, sparke /ku či
 á Sal

Behind the dictionaries there are dynamic lookup processes

FST-analysis and translation via a web service



... that can generate word paradigms

Neahttagisáin [Home](#) [Click-in-text](#) [About](#) 

sme→nob

leat (verb)

- o [\(verb\) er](#)

Mu oabbá lea sámegiigáirras.
Søstera mi er sametingsrepresentant.
- o [\(verb\) har](#)

Mus lea beana.
Jeg har en hund.
- o [\(verb\) \(hjelpeverb\)](#)


Mun lean oastán biilla.
Jeg har kjøpt bil.

It go oainne ahte lean lohkame?
Ser du ikke at jeg leser?

	PRESENTS (odnha)	PRETERITUM (kda)	
1.p.ent.	(mun) lean	ledjen	<i>leat</i> verb indikativ presens 2.p.ent.
2.p.ent.	(don) leat	ledjet	<i>leat</i> verb indikativ presens 1.p.ft.
3.p.ent.	(son) lea	lei	<i>leat</i> verb indikativ presens 3.p.ft.
		leai	<i>leat</i> verb indikativ presens nektingsform
		leimme	<i>leat</i> verb infinitiv
1.p.tot.	(moai) letne	leimme	
	(moai) ledne		
2.p.tot.	(doai) leahppi	leicde	
3.p.tot.	(soai) leaba	leigga	
		leigga	
1.p.ft.	(mii) leat	leimmet	
2.p.ft.	(dii) lehpet	leiddet	
3.p.ft.	(sii) leat	ledje	
	presens nektingsform	(odne in) leat	
	preteritum nektingsform	(ikte in) lean	
	perf.partisipp	(lean) leamaš	
		(lean) leamašan	

[Full paradigm →](#)

... that can generate word paradigms

Neahttadigisánit [Home](#) [Click-in-text](#) [About](#) 

sme→nob ▾

nieida (subst.)

- (subst.) [jente](#)
- (subst.) [datter](#)

nieida is a possible form of ...

	ENT.	FLT.	<i>nieida</i> Texts →
nominativ	nieida	nieiddat	nieida subst. ent. nominativ
akkusativ	nieidda	nieiddaid	
genitiv	nieidda	nieiddaid	
illativ	niidii	nieiddaide	
lokativ	nieiddas	nieiddain	
komitativ	nieiddain	nieiddaiguin	
essiv		nieidan	

[Full paradigm →](#)

[More information about this word](#)

You can search in text corpus

Neahttagisáinít Home Click-in-text About

Dictionaries

- North Saami → Finnish
- North Saami → Norwegian**

WRITTEN VARIANT

- Standard (*áčđŋšž*)
- Tolerant (incl. *acdñstz*)

- Norwegian → North Saami
- Finnish → North Saami

Other North Saami dictionaries

Other dictionaries

North Saami (Standard) → Norwegian (⇌ Swap)


á č đ ŋ š t ž

á Search term Search **Search texts**

Search for a word, or an inflected form. Click the word for a list of inflected forms.

You can also search for the word in a text corpus, or for all inflected forms of the word by clicking the 'texts' link next to it in the search results.

You can search in text corpus

Neahttadigisánit [Home](#) [Click-in-text](#) [About](#) 

Dictionaries

North Saami → Finnish

North Saami → Norwegian

WRITTEN VARIANT

Standard (áčđŕšžž)

Tolerant (incl. acđnstž)

Norwegian → North Saami

Finnish → North Saami

Other North Saami dictionaries

Other dictionaries

North Saami (Standard) → Norwegian (= Swap)

á č đ ŕ š ž

á

nieida is a possible form of ...

nieida [Texts -->](#)



nieida subst. ent. nominativ


nieida (subst.)

- o (subst.) jente
- o (subst.) datter

You can search in text corpus

North Saami texts | Lule Saami texts | South Saami texts | Aanaar Saami texts | Skolt Saami texts | More »

Cite Korp | Davvisámi | Norsk | Suomi | English  



10 corpora selected (all) — 28.41M of 28.41M tokens

nieida

Simple Extended Advanced Compare

nieida Search

also as initial part final part and case-insensitive

KWIC: hits per page: 25 | sort within corpora: not sorted | Statistics: compile based on: word | Show word picture show_map

KWIC Statistics Word picture map

Results: 2,621

« 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... » goto_page of 105 Show context

Adaptation from: (does not support extended context)

Ivan Petrof Lavriaa ruovttuodalus ledje eamit, **nieida** ja guokte kvena mat orro su luhtte.
 Veahás margelaš oidne balddáskan nieidaustibat ahte seakka juvddálaš **nieida** dolvojuvvui uksaráiigi olggos Dalsbergstien 3s.
 oahppat ja máhttit atnit doahpagiid eadni, áhčči, **nieida**, bárdni, viellja, oabbá ja áhkku ja áddjá sámegiilli.
 Njeallječieguide 19-24 čállit bearašániid nugomat **nieida**, bárdni, eadni, áhčči, ja áhkku ja áddjá.
 Dát **nieida** lea doadján giedal
 Okta **nieida** lea Ohcejogas eret.
 Lei beallesaddi **nieida**.
 Muhtun geasi **nieida** jávkka, ja olbmot ohce su, muhto eai gávdnan su eai gostege, eaije diehtán gosa lei
 Nisu dajal: * Dát lea du **nieida**, gi lei láhpon.
 Duorastaga dáiddadáhpáhusa láiddesteaddjin lei lassin Heahhtá **nieida**, studeanta Rosa Mären Magga.
 Os lea dattege ahte nuoramus bárdni válddiid badjelasas doalu, muhto dat sáhtii maid leat **nieida** jos son lei nuoramus.
 Iit suorra beroštumi ja duostilvuodain ovddiidit ležaset sámi duogáža – nugo nuorra sámi **nieida** Eli Anne Nystad dagai marjnjá unohas dáhpáhusa Troandimis mannan čavčča.
 5- jahkásas nieidaš lei Bááhveaisápmelaččaid jodiheaddji **nieida**, ja lei gáhcččan rabas dollii ja buollán fasttit.
 Juohke vídát sámi nuorra **nieida** vásiha seksuála vearredaguid ovdal. go lea deavdán 18 jagi.
 Okta **nieida** ja okta bárdni oažžuba goabbáge stipeandda, jus goappašat sohkaabeallit leaba ohci
 Okta **nieida** ja okta bárdni oažžuba goabbáge stipeandda, jus goappašat sohkaabeallit leaba ohci
 Sealggeadnin logai Stockholmmas ja čájeihii juo dalle leat roahkkadis **nieida**, guhte ii ballan oaivillidist dadjalit.
 Okta **nieida** ja okta bárdni oažžuba goabbáge stipeandda, jus goappašat sohkaabeallit leaba ohci

Corpus

Administrative texts

Text attributes

title: Finnmärkkue eatnamiid ja čázadagaid
 geavahepmi historjjálaš
 author's name: Pedersen
 domain: administration

Word attributes

part-of-speech: noun
 baseform: nieida
 grammatical analysis: N.Sg.Nom
 dependency relation: ←SUBJ

Show Dependency Tree

You can search in text corpus

10 corpora selected (all) — 21.11M of 21.11M tokens

Search history

Simple Extended Advanced Compare

part-of-speech is
adverb

grammatical analysis contains
Ind

grammatical analysis contains
Nom

Search within sentence

KWIC: hits per page: 50 sort within corpora: not sorted

KWIC Statistics

Results: 24,792

Previous 1 2 3 4 5 6 7 8 9 10 11 - 495 496 Next Show context

ADMINISTRATIVE TEXTS (does not support extended context)

Álggos Čájejuvvo regioidnákárta ja obbalaš diedut regiivna birra.
Dás lea sáhka smávrit eanan-osiin, mat leat lagamusas orohaga // čearu guohtonguovllu, ja leat dákká nuppe bealde riikaráji.
Danne dárbašit bohccot dagalduvvat fuoddariidda hui árrat, dan báie go mikroorganismat leat olleslogus čovvjijs ja nákejit smoidgeahtit odda fuoddi.
Dálvet guohtu eallu nanas álggá. lea lođiil. iise manat áiktaid duššetá.
Danne geahppu boazu
Dálvet lea jeageguohtun
Nu hedjona ealádatkvalitehta
Vuosttážettiin lea váttis
Nuppáđassii lea váttis
Dás lea sáhka
Dánnehan leage mávssolaš
Nu juohká várreráidu

Dependency Tree

Álggos Čájejuvvo regioidnákárta ja obbalaš diedut regiivna birra .

Corpus

Administrative texts

Text attributes

title: __UNDEF__
year: __UNDEF__
undefined: __UNDEF__
domain: administration
undefined: __UNDEF__

Word attributes

part-of-speech: verb
grammatical analysis: VIV:Ind.Prs.Sg3
dependency relation: Modal verb

Show Dependency Tree

Conclusion

Digital humanities by and for whom?

Conclusion

Digital humanities by and for whom?

Digital humanities for all working within the humanities

Conclusion

Digital humanities by and for whom?

Digital humanities for all working within the humanities

How linguistics and LT can help?

Conclusion

Digital humanities by and for whom?

Digital humanities for all working within the humanities

How linguistics and LT can help?

→ By "cleaning the signal",
and thereby clearing the picture

Thanks

`http://giellatekno.uit.no/`