

Computational modelling of verbs in Dene languages

The case of Tsuut'ina

Christopher Cox, Måns Huldén, Miikka Silfverberg, Jordan Lachler, Sally Rice, Sjur N. Moshagen, Trond Trosterud & Antti Arppe

Yukon Education

University of Colorado, Boulder

University of Helsinki

University of Alberta

Giellatekno & Divvun – UIT Arctic University of Norway

Circumpolar collaboration:

We are building on work done by the Giellatekno and Divvun research and development teams at the University of Tromsø

Giellatekno and Divvun have worked over a decade on developing language technological tools for the indigenous Saami languages spoken in Northern Scandinavia



Arctic peoples subdivided according to language families

Indo-European family
Germanic branch

Uralic family

Finnic-Ugric branch
Samoyedic branch

Altaiac family

Turkic branch
Tungusic branch

Chukotko-Kamchatkan fam.

Isolated languages
(Ketic and Yukagir)

Eskimo-Aleut family

Inuit group (of Eskimo br.)
Yupik group (of Eskimo br.)
Aleut branch

Na-Dene family

Athabaskan branch
Eyak branch
Tlingit branch

— Arctic circle

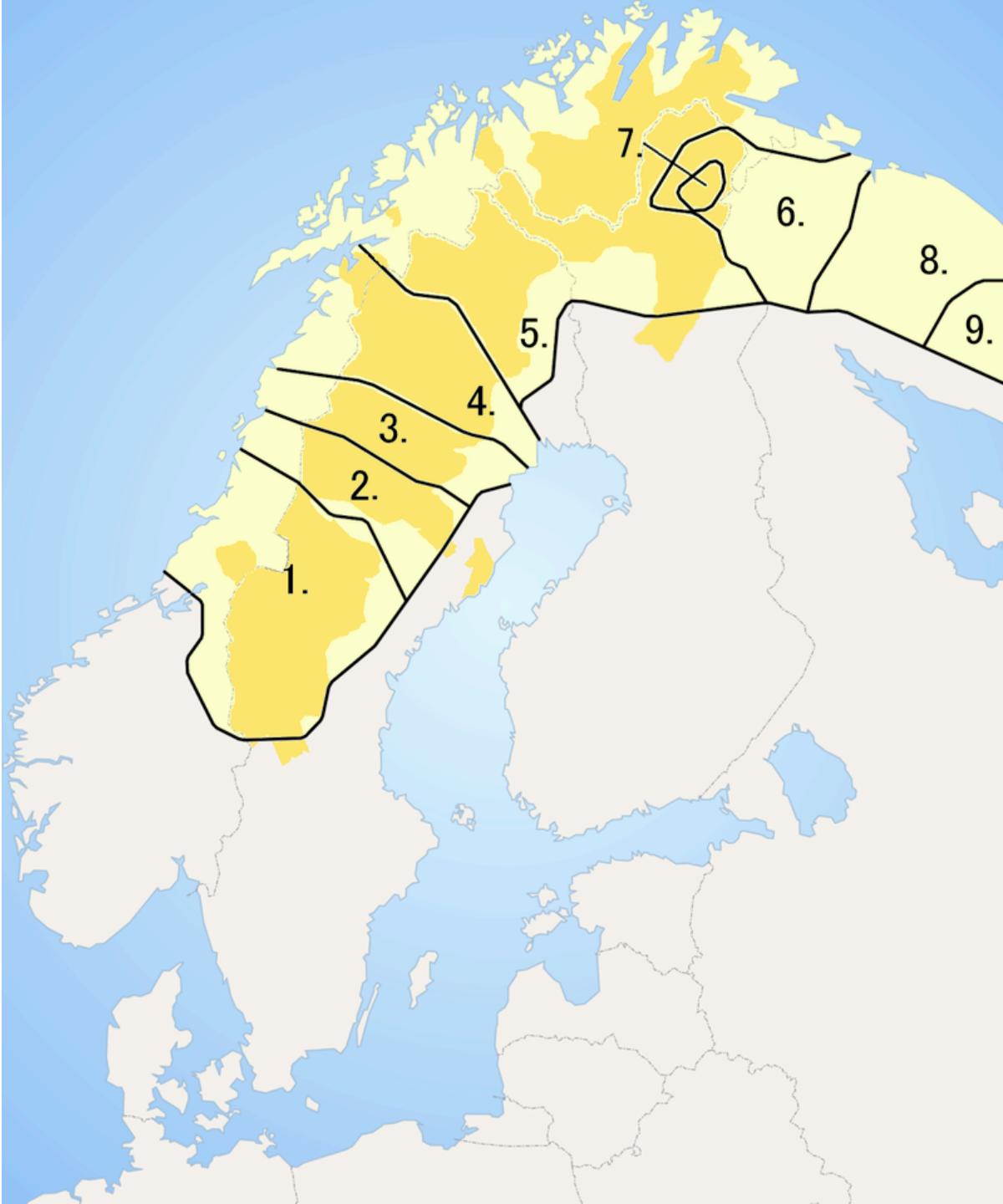
— Arctic boundary according to AMAP

Notes:

Areas show colours according to the original languages of the respective indigenous peoples, even if they do not speak their languages today.

Overlapping populations are not shown. The map does not claim to show exact boundaries between the individual language groups.

Typical colonial populations, which are not traditional Arctic populations, are not shown (Danes in Greenland, Russians in the Russian Federation, non-native Americans in North America).



Historically verified distribution
of the Saami languages:

1. Southern Sami (~600)
2. Ume Sami (~20)
3. Pite Sami (~20)
4. Lule Sami (~1000-2000)
5. Northern Sami (~20k)
6. Skolt Sami (~420)
7. Inari Sami (~300)
8. Kildin Sami [~500)
9. Ter Sami (~2-10)

Darkened area represents
municipalities that recognize
Sami as an official language.

Buresboahtin Giellatekno – sám... +

giellatekno.uit.no | C Search | ☆ | ☰

Most Visited itwēwina Guusaaw RDVRinSilverlight4 SMS Intranet Connecting...

UIT, The arctic university of Norway > Giellatekno >

 **Sámi
giellatekno**

Ruoktu/Home Divvun Dicts oahpa.no Samiske språk – Saami Languages Andre språk – Other Lang Search the site with google Search

Last Published: 06/06/2016 17:18:57

▼ Fállu

- Ruoktot
- Giellatekno birra
- Lágideamit
- Dutkansuorggit
- Publikašuvnnat
- Prošeavttat
- Teknikhalaš resurssat
- Siskkáldas

built with Apache Forrest

Buresboahtin Giellatekno - sámi giellateknologija guovddážii

 Davvisámegillii  Norsk  + Suomeksi  English  Русский

Giellatekno laktá oktii sámejela ja eará morfolođijarikkes gielaid lingvistalaš ja diitorlingvistalaš dutkama geavatlaš prográmmaid ovddidemiin. Mii fokuseret erenoamážit čiekjalis lingvistalaš modelleremii mii lea vuodđuduuvvon morfologalaš analysii mii galgá doaibmat giela buot osiin.

Mii doarjut sámi giellaservodagaid geavatlaš prográmmaid ovddideami bokte. Min mihttomearri lea doarjut sihke giellageavaheami ja giellaoahppama ja dahkat gielaid eanet olámmuttus. Jagis 2012 oačuimet, ovttas [Divvun-joavkkui](#), [Gollejiella-bálkkašumi](#), min barggu rámideapmin.

Mii leat viiddidan iežamet barggu eará gielaise main leat unnán giellateknologalaš resurssat, erenoamážit eará sirkumpolára gielaise. Min analysat ja reaiddut dahket álkibun vehádatgielaide ovddidit giellateknologija mii lea eaktun giela ceavzimii geavahangiellan odđaigášaš servodagas.

- [Sátnegirjít](#)
- [Reaiddut sámegielaise](#)
- [Reaiddut eará gielaise](#)

Copyright © 2005-2014 The University of Tromsø. Last Published: 06/06/2016 17:18:57

Send feedback about the website to: giellatekno@uit.no

Welcome to Giellatekno, the Ce... +

giellatekno.uit.no/index.eng.html

Search

UIT, The arctic university of Norway > Giellatekno >

U. ROMSSA UNIVERSITETTA Sámi giellatekno

Ruoktu/Home Divvun Dicts oahpa.no Samiske språk - Saami Languages Andre språk - Other Languages Search the site with google Search

Last Published: 06/06/2016 17:18:57

Menu

- Home
- About Giellatekno
- Events
- Research areas
- Publications
- Projects
- Technical resources
- Internal

Welcome to Giellatekno, the Center for Saami language technology

 [Davvisámegillii](#)  [Norsk](#)  [Suomeksi](#)  [English](#)  [Русский](#)

Giellatekno combines cutting-edge linguistic and computational research into the analysis of Saami and other morphologically-rich languages, with the development of practical applications. We focus on deep linguistic modeling and on highly efficient and robust computational analysis with a wide empirical coverage.

These applications form the basis of our support of the Saami communities with current language tools. In 2012, we were, together with the [the Divvun-group](#), awarded with the [Gollegiella prize](#) recognizing this work. Our practical goal is to support the teaching, learning and use of the Saami languages and to make current language technology accessible.

We also extend our activities to other under-resourced languages, particularly Circumpolar and Uralic languages. Our analyses and tools are designed to make it easier for other minority language societies to develop the language technology constituting a prerequisite for a language to survive in modern society.

- [Dictionaries](#)
- [Saami language resources](#)
- [Other language resources](#)

Copyright © 2005-2014 The University of Tromsø.

Last Published: 06/06/2016 17:18:57

Send feedback about the website to: giellatekno@uit.no

Tools for Saami languages

giellatekno.uit.no/smilang.eng.html

Most Visited | itwēwina | Guusaaw | RDVRinSilverlight4 | SMS Intranet | Connecting...

UIT, The arctic university of Norway > Giellatekno >

 **Sámi
giellatekno**

Ruoktu/Home | Divvun | Dicts | oahpa.no | Samiske språk – Saami Languages | Andre språk – Other Langu | Search the site with google | Search | Last Published: 06/06/2016 17:19:06

Saami languages

- All saami languages
- North Saami
- Lule Saami
- Pite Saami
- South Saami
- Skolt Saami
- Inari Saami
- Kildin Saami
- Ter Saami
- Uralic languages
- Other languages

Tools for Saami languages

 [Davvisámegillii](#)  [Norsk](#)  [English](#)  [+ Suomeksi](#)  [Русский](#)

Note

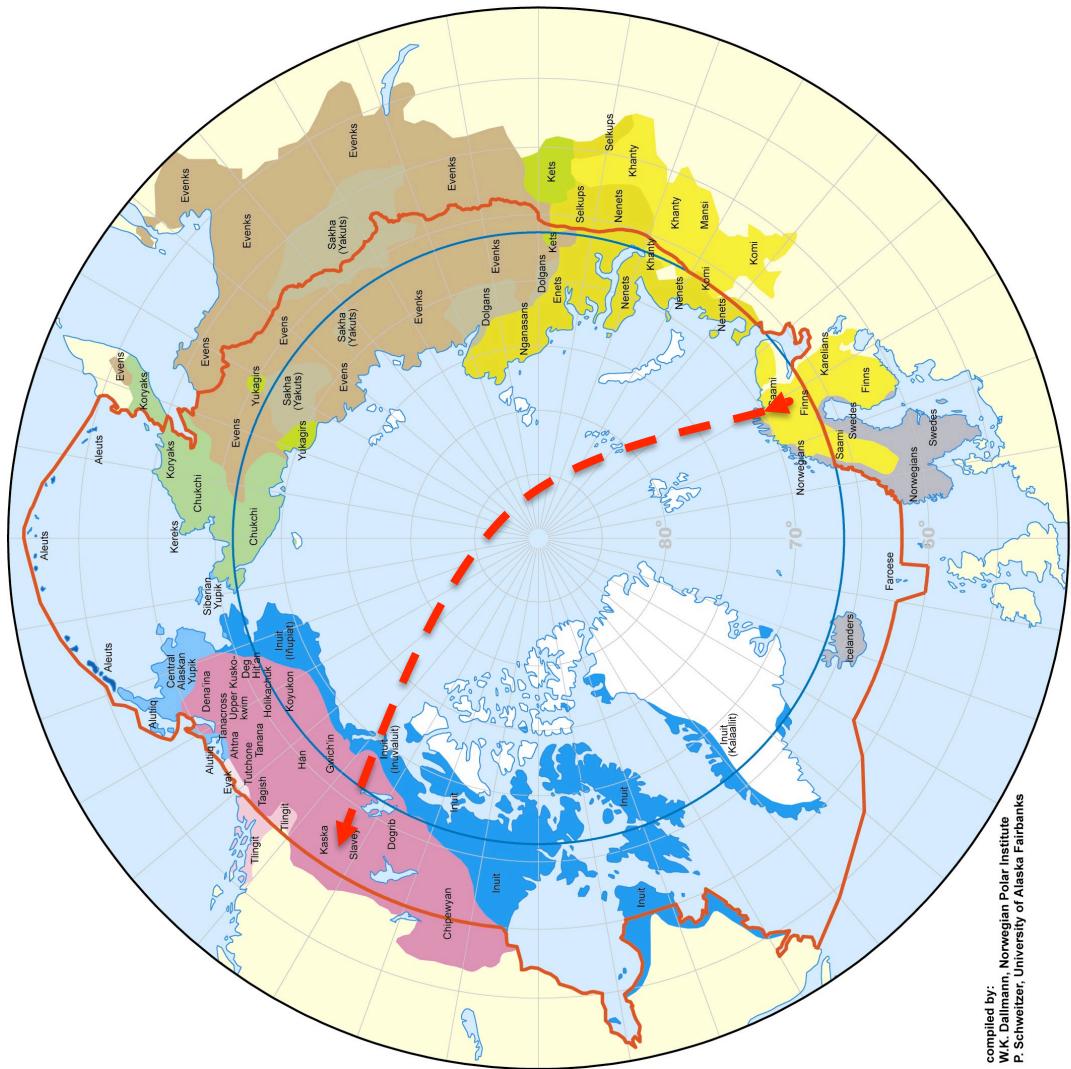
FinUgRevita is making an investigation of our language programs. Help us by filling in their questionnaire:
[SAAMI SURVEY](#)  in English 

Our resources according to language

- Saami languages: [North Saami](#), [Lule Saami](#), [South Saami](#) // [Inari Saami](#), [Kildin Saami](#), [Pite Saami](#), [Skolt Saami](#).

Our resources according to type

- Language learning [Oahpa giela - interactive Saami course](#) , [OAHPAI language learning](#) , [VISL grammar learning](#) , [North Saami grammar](#)  in English, and [South Saami grammar](#)  in Norwegian.
- Translation: North Saami - Norwegian [Machine Translation](#)  and [Translation Memory](#).
- Text-to-speech: [North Saami text-to-speech](#) .
- Dictionaries: [Several languages](#) , [Geo](#), [Saami placenames](#)
- Saami on the computer and smartphone: [Divvun proofing tools](#) , [Keyboard](#), [HTML-entities](#), [Program installation](#) and [Saami letters](#) .
- For linguists: [Linguistic research](#), [Wordlists](#), [Saami interactive corpus](#) , [Johan Turi: Muitalus sámiid birra](#) and [Our source files](#) .



compiled by:
W.K. Dallmann, Norwegian Polar Institute
P. Schweizer, University of Alaska Fairbanks

Arctic peoples subdivided according to language families

Indo-European family	Isolated languages (Ketic and Yukagir)
Germanic branch	Eskimo-Aleut family
	Inuit group (of Eskimo br.)
	Yupik group (of Eskimo br.)
	Aleut branch
Uralic family	Na-Dene family
Finnno-Ugric branch	Althabaskan branch
Samoyedic branch	Eyak branch
	Tlingit branch
Altaiic family	
Turkic branch	
Tungusic branch	
Chukotko-Kamchatkan fam.	

- Notes:**
- Areas show colours according to the original languages of the respective indigenous peoples, even if they do not speak their languages today.
 - Overlapping populations are not shown. The map does not claim to show exact boundaries between the individual language groups.
 - Typical colonial populations, which are not traditional Arctic populations, are not shown (Danes in Greenland, Russians in the Russian Federation, non-native Americans in North America).

Work with North-American Indigenous languages undertaken at ALTLab (in U of Alberta)

The screenshot shows a web browser window with the URL altnlab.artsrn.ualberta.ca in the address bar. The page features a large, colorful graphic background with overlapping triangles in shades of yellow, orange, and red. The main title "Alberta Language Technology Lab" is displayed in a large, bold, black font. Below it, the subtitle "21st Century Tools for Indigenous Languages" is shown in a smaller, italicized black font. A navigation menu at the top includes links for "Images", "News and Events", "Open graduate research assistantship", "People and Partners", "Publications", "Tools and Applications", and a search icon. The central content area contains a large heading "21st century tools for indigenous languages: 2013-2016". Below this, there are two paragraphs of text. The first paragraph discusses the project's goal of using Plains Cree as a spearhead language to develop tools like spell-checkers, language teaching software, and text-to-speech synthesizers. The second paragraph explains the project's aim to facilitate the use of minority languages in all spheres of life by community members, noting that such technologies are available for most world languages but have been created for only a few minority languages.

Alberta Language Technology Lab
21st Century Tools for Indigenous Languages

Images News and Events Open graduate research assistantship People and Partners Publications Tools and Applications

21st century tools for indigenous languages: 2013-2016

Sticky Uncategorized

Using *Plains Cree* as the spearhead language, this project will produce tools such as *spell-checkers*, *language teaching and learning software*, and *text-to-speech synthesizers*.

These technologies are available for world's majority languages (e.g. English), but have so been created for only a few minority languages. In providing minority language speakers with these applications the project aims

to facilitate the use of minority languages in all spheres of life by community members.

WWW: altnlab.artsrn.ualberta.ca

Basic Language Resource Kits for Endangered Languages – EL-BLARK

- Indigenizing the technological resource questions (Arppe et al. 2016):
 - 1) What types of relevant linguistic data resources are likely to be available as a result of field/community linguists' work?
 - 2) What human language technology applications may be of most practical value in supporting the continued use and revitalization of these languages within their communities?
 - 3) Together, (1) and (2) again determine the possible and necessary technological module components

Reasonably expectable data resources

- 1) descriptions of morphology and syntax
 - basic sketches → comprehensive detailed grammars with explicit descriptions of inflectional paradigms and syntactic constructions
- 2) bilingual lexical descriptions
 - translations to a majority language
 - basic word lists → full-scale comprehensive lexical databases (information on paradigm class and semantic restrictions)
- 3) narrative text collections
 - in either printed or electronic format
 - with or without accompanying spoken recordings
- 4) recordings of spoken language
 - carefully pronounced individual words and sentences → multi-participant native speaker discourse, and narratives of various types
 - at best, may be annotated

Practical, useful applications: I-DICT

- Intelligent, web-accessible dictionaries (I-DICT)
 - lexical databases + computational morphological analysers and generators
- computer-aided language learning (CALL) applications
 - exercises with morphological alternations in context (intelligent CALL, or ICALL)
- writers' tools
 - spell-checkers + grammar-checkers
- speech synthesis

ALTLab – languages

- Dene
 - Tsuut’ina
- Algonquian
 - Plains Cree, Northern East Cree, Odawa (Ojibwe)
- Iroquoian
 - Cherokee
- Others/Isolates
 - Northern Haida

Computational tools for Dene languages

- Applying existing infrastructure to Dene languages requires adaptation—especially for **modelling Dene verbs**:
 - Overall structure, with outer (*disjunct*), inner (*conjunct*), and stem ‘zones’ of verb (cf. Kari 1989) not computationally difficult to model
 - More challenging: extensive morphological fusion in subject-aspect inflection immediately preceding the stem (cf. K. Rice 2005: 404–407)

Templatic morphology

gámił

“they’re swimming” (PROGRESSIVE)

Templatic morphology

gámił

“they’re swimming” (PROGRESSIVE)

Incorporated PP	Adverbial	Iterative (<i>ná-</i>)	Incorporated stem	Distributive (<i>dà-</i>)	Object	3P Subject	Thematic	Aspect ₂	Aspect ₁ (Mode)	1/2P Subject	Classifier	STEM
12	11	10	9	8	7	6	5	4	3	2	1	0
					gi-				yi-			míł

Templatic morphology

gámił

“they’re swimming” (PROGRESSIVE)

1

ABSTRACT
PARTS

gi- yi- míł
they he,she,it:PROG swim:PROG

Templatic morphology

gámił

“they’re swimming” (PROGRESSIVE)

1

ABSTRACT
PARTS

gi- yi- míł
they he,she,it:PROG swim:PROG

2

ABSTRACT
RULES

1. gi- yi- míł → gi- á- míł (*yi*-Augmentation)
2. gi- á- míł → gámił (*i*-Deletion)

Templatic morphology

gámił

“they’re swimming” (PROGRESSIVE)

1 ABSTRACT
PARTS

gi- yi- míł
they he,she,it:PROG swim:PROG

2 ABSTRACT
RULES

1. gi- yi- míł → gi- á- míł (*y*i-Augmentation)
2. gi- á- míł → gámił (*i*-Deletion)

3 ACTUAL
WORDS

gámił
“they’re swimming”



The conjugation-primary aspect-subject portion of the verb combines in ways that are not always predictable.

K. Rice (2001)

The conjugation-primary aspect-subject portion of the verb combines in ways that are not always predictable. [...] The non-systematic combinations of conjugation- primary aspect-subject suggest that in at least some cases this stretch of the verb should be treated as a single unit, or portmanteau morph, with complex meaning.

K. Rice (2001)

Paradigmatic approaches

- From a computational perspective, simpler to treat inner inflection as pre-composed ‘chunks’ (portmanteaux):
 - Significantly reduces morphophonemics—only junctures between zones of the verb generally need to be modelled
 - *Precedents:* Young & Morgan (1987), Faltz (1998), McDonough (2000) for Navajo; Leer (1999, *inter alia*) for comparative Na-Dene; etc.

Paradigmatic Approach

gámíł

“they’re swimming” (PROGRESSIVE)

yi-Progressive (no other prefixes)		
1s	yismíł	“I’m swimming”
2s	yímíł	“you’re swimming”
3s	yámíł	“he/she/it is swimming”
1P	yaàmíł	“we’re swimming”
2P	yasmíł	“you guys are swimming”
3P	gámíł	“they’re swimming”
4	ts’ámíł	“someone’s swimming”

Paradigmatic Approach

gámíł

“they’re swimming” (PROGRESSIVE)

yí Progressive (no other prefixes)			
1s	yis-	gámíł	“I’m swimming”
2s	yí-	gámíł	“you’re swimming”
3s	yá-	gámíł	“he/she/it is swimming”
1P	yaà-	gámíł	“we’re swimming”
2P	yas-	gámíł	“you guys are swimming”
3P	gá-	gámíł	“they’re swimming”
4	ts’á-	gámíł	“someone’s swimming”

Table 7: yi- Progressive verbs

	No prefix			Inner prefix			Outer prefix		
Basic	"Be swimming" (PROG):			"Be laughing" (PROG):			"Be becoming like that" (PROG):		
	1s	yismíł		S	diyisdlùw		1s	xayisnáł	
	2s	yímíł		S	diyídlùw		2s	xayínáł	
	3s	yimíł		S	dádlùw		3s	xayináł	
	1P	yaàmíł		P	diyaàdlùw		1P	xayaànáł	
	2P	yasmíł		P	diyasdlùw		2P	xayasnáł	
	3P	gámíł		P	gidádlùw		3P	xagánáł	
	4	ts'ámił		P	ts'idádlùw		4	xats'ánáł	
With s-	"Be almost swallowing O" (PROG):			"Be almost swallowing O" (PROG):			"Be taming, training O" (PROG):		
	1s	yiswùsh		1s	diyisníł		1s	áyisdíł	
	2s	yíswùsh		2s	diyísnił		2s	áyísdíł	
	3s	O	yáswùsh (<i>O yiswùsh?</i>)	3s	O	dásnił	3s	O	áyisdíł
	1P	yaàwùsh		1P	diyaànił		1P	áyaàdíł	
	2P	yaswùsh		2P	diyasnił		2P	áyasdíł	
	3P	O	gásywùsh	3P	O	gidásnił	3P	O	agásdíł
	4	ts'áswùsh		4	ts'idásnił		4	áts'ásdíł	
With i-	"Be running along" (PROG):			"Be starting to run along" (PROG):			"Be running along home" (PROG):		
	1s	yist'áł		1s	didiyist'áł		1s	náyist'áł	
	2s	yíít'áł		2s	didiyíít'áł		2s	náyíít'áł	
	3s	yiit'áł		3s	didáát'áł		3s	náyiit'áł	
	1P	yaàt'áł		1P	didiyaàt'áł		1P	náyaàt'áł	
	2P	yast'áł		2P	didiyast'áł		2P	náyast'áł	
	3P	gáát'áł		3P	gididáát'áł		3P	nágáát'áł	
	4	ts'áátl'áł		4	ts'ididáát'áł		4	náts'áátl'áł	

Computational model

- Finite-State Transducers: FST (e.g. Beesley & Karttunen 2003, Huldén 2009, Lindén et al. 2011)
 - Advantages – well-known computational beasts
 - fast and efficient
 - calculus exists for powerful manipulations
 - allow rule-based definition of paradigms for various verb types (however one wants to define these)
 - portability to different operating systems and platforms
 - easy integratability with other applications (e.g. as spell-checking modules in word-processors)
 - Disadvantages
 - only as good as the available linguistic descriptions
 - too powerful?
 - do all paradigm forms make (equally) sense for all individual verbs within a paradigm type?
 - this is problem with any rule-based system

Structuring the Dene verb

- lexical meaning can be expressed either in single stems, or through fixed, (semi-)arbitrary collocations of prefixes and stems
- however, for each “meaning” the location(s) of these lexical prefixes are known – three locations/boundaries:
 - “inner”, “middle”, “outer”
- Grammatical information on subject, object, etc., indicated by morphemes interspersed within the lexical elements, though at known locations
 - “inner”, “middle”, “outer”

→ two tiers

- lexical & inflectional

Inflectional tier

- inner prefixes
 - portmanteau subject person-number-aspect forms
- “middle” prefixes
 - direct object markers
 - third person plural/fourth person subject prefixes (“outer” subjects)
- outer prefixes
 - distributive plural marker dà-

Lexical tier

- |_|_|_tsiy ‘cry’
 - No lexical prefixes
 - |_|_di_tł'áh ‘run’
 - (only) one lexical prefix *-di-* in the inner position
 - |_gu_|_náh ‘talk’
 - (only) one lexical prefix *-gu-* in the middle position
 - ts'á_|_|_zíd ‘wake up’
 - (only) one lexical prefix *ts'á-* in the outer position
 - tsí_|_di_tł'á ‘run-away’
 - two lexical prefixes in the outer and inner positions
 - nà_gu_di_tłod ‘jump down’
 - three lexical prefixes in all three positions
-
- Vertical lines denote “empty” lexical prefix “slots”/“boundaries”
 - Underscores denote inflectional prefix chunk “slots”

Inflectional tier – “inner” prefixes

- In Dene languages, the inflection that appears in the pre-stem syllable is often analyzed as being morphologically complex, typically representing a combination of 4-6 distinct morphemes.
- In this model, this pre-stem inflection is not fully decomposed into its component morphological parts, but rather treated as forming several distinct sets of portmanteau inflectional markers → “inner” prefixes
- Each portmanteau inflectional marker in these paradigms aggregates several pieces of information:
 - Subject person and number (1-3 persons singular, 1-2 persons plural)
 - Aspect and mode;
 - Conjugation class (i.e., *si-*, *ni-*, *yi-*, or none, representing one of the historical Dene aspect/conjugation class markers)
 - Voice/valence (i.e., the historical voice/valence markers or "classifiers" that fuse phonologically with the preceding prefixes)

Stem alternations → aspect

- itsiy → tsiy INTR-0-IPFV
- itsiy → tsày INTR-yi-y-PFV
- itsiy → tsíł INTR-yi-PROG
- itsiy → ná_|_|_chish INTR-0-IPFV
- stem alternations/aspect associated with a particular set of morpheme chunks in the various slots – for *itsiy* 'cry' the following ones:
 - *0-Imperfective*
 - *yi-y-Perfective*
 - *yi-Progressive*

LEXICON Verbstems

FST-internal notation for lexical prefixes, inflectional prefix positions, and inflectional morpheme chunk “classes”

itsiy[cry]:tsiy	INTR-0-IPFV; ! (IPFV; 0-IPFV)
itsiy[cry]:tsày	INTR-yi-y-PFV; ! (PFV; yi-PFV-yi)
itsiy[cry]:tsíł	INTR-yi-PROG; ! (PROG; yi-PROG)
itsiy[cry]:ná=chish	INTR-0-IPFV; ! (REP; 0-IPFV)
ts'ázid[wake-up]:ts'á=zíd	INTR-ni-IPFV; ! (IPFV; ni-IPFV)
ts'ázid[wake-up]:ts'á=zid	INTR-ni-PFV; ! (PFV; ni-PFV)
ts'ázid[wake-up]:ts'á=ził	INTR-yi-PROG; ! (PROG; yi-PROG) CHECK
ts'ázid[wake-up]:ts'áná=zhiizh	INTR-ni-IPFV; ! (REP; ni-IPFV)
nàgudiitłod[jump-down]:nà=gu_di.tłod	INTR-0i-IPFV; ! (IPFV; 0i-IPFV)
nàgudiitłod[jump-down]:nà=gu_di.tłòt	INTR-sii-PFV; ! (PFV; sii-PFV)
nàgudiitłod[jump-down]:nà=gu_di.tłíł	INTR-yii-PROG; ! (PROG; yii-PROG)
nàgudiitłod[jump-down]:nàná=gu_di.tłiizh	INTR-0i-IPFV; ! (REP; 0i-IPFV)
nàgudiitłod[jump-down]:nìná=gu_di.tłiizh	INTR-0i-IPFV; ! (REP; 0i-IPFV)
tsídiitł'á[run-away]:tsí=di.tł'á	INTR-0i-IPFV; ! (IPFV; 0i-IPFV)
tsídiitł'á[run-away]:tsí=di.tł'ò	INTR-sii-PFV; ! (PFV; sii-PFV)
tsídiitł'á[run-away]:tsí=di.tł'áł	INTR-yii-PROG; ! (PROG; yii-PROG)
tsídiitł'á[run-away]:nátsí=di.tł'ásh	INTR-0i-IPFV; ! (REP; 0i-IPFV)
tsídiitł'á[run-away]:tsí=di.tł'ò	INTR-yii-POT; ! (POT; yii-POT)

Dene/Tsuut'ina FST structure – lexical tier

Outer Lexical prefix	Outer Inflectional prefix	Middle Lexical <u>Prefix</u>	Middle Inflectional Prefix	Inner Lexical Prefix	Inner Inflectional Prefix	<u>Stem</u>
0 0 0 ná→		0		0		tsiy tsày tsít ←chish
ts'á→ ts'á→ ts'á→ tsáná→						←zíd ←zid ←zíl ←zhiizh
nà→ nà→ nà→ nàná→ nìná→		gu→		di→		←tl̥od ←tl̥at ←tl̥ít ←tl̥iizh ←tl̥iizh

Dene/Tsuut'ina FST structure – inflectional tier

Outer Lexical prefix	Outer Inflectional prefix	Middle Lexical Prefix	Middle Inflectional Prefix	Inner Lexical Prefix	Inner Inflectional Prefix	Stem
0 0 0 ná→	Distributive	0	SUBJECT: 3PI 4Sg	0	SUBJECT: 1Sg 2Sg 3Sg 1PI 2PI	tsiy tsày tsít ←chish ←zíd ←zid ←zíl ←zhiizh
ts'á→ ts'á→ ts'á→ tsáná→			DIRECT OBJECT		ASPECT Imperfective Perfective Progressive Repetitive	
nà→ nà→ nà→ nàná→ nìná→		gu→		di→	CONJ-CLASS CLASSIFIERS	←tłod ←tłàt ←tłít ←tłiizh ←tłiizh

Dene/Tsuut'ina FST structure – inflectional tier

Outer Lexical <u>prefix</u>	Outer Inflectional <u>prefix</u>	Middle Lexical <u>Prefix</u>	Middle Inflectional <u>Prefix</u>	Inner Lexical <u>Prefix</u>	Inner Inflectional <u>Prefix</u>	<u>Stem</u>
0 0 0 ná→	0 Distr:dà	0	0 3Pl:gi 4Sg:t'si	0	0-Ipfv-no 0-Ipfv-inner 0-Ipfv-outer 0s-Ipfv-no	tsiy tsày tsít ←chish
ts'á→ ts'á→ ts'á→ tsáná→					0s-Ipfv-inner 0s-Ipfv-outer 0i-Ipfv-no 0i-Ipfv-inner 0i-Ipfv-outer	←zíd ←zid ←zíl ←zhiizh
nà→ nà→ nà→ nàná→ nìná→		gu→		di→	... yi-y-Pfv-no yi-y-Pfv-inner yi-y-Pfv-outer ...	←tłod ←tłàt ←tłít ←tłiizh ←tłiizh

Dene/Tsuut'ina FST structure – inflectional ← → lexical tier

Outer Lexical prefix	Outer Inflectional prefix	Middle Lexical Prefix	Middle Inflectional Prefix	Inner Lexical Prefix	Inner Inflectional Prefix	<u>Stem</u>
0	0	0	0	0	0-Ipfv-no	tsiy:0-Ipfv
0	Distr:dà		3Pl:gi		0-Ipfv-inner	tsày:yi-y-Pfv
0			4Sg:ts'i		0-Ipfv-outer	tsíł:yi-Prog
ná→					Os-Ipfv-no	chish:0-Ipfv
ts'á→					Os-Ipfv-inner	zíd:ni-Ipfv
ts'á→					Os-Ipfv-outer	zid:ni-Pfv
ts'á→					Oi-Ipfv-no	zíl:yii-Prog
tsáná→					Oi-Ipfv-inner	zhiizh:ni-Ipfv
					Oi-Ipfv-outer	
nà→		gu→		di→	...	
nà→					yi-y-Pfv-no	tłod:0i-Ipfv
nà→					yi-y-Pfv-inner	tłàt:sii-Pfv
nà→					yi-y-Pfv-outer	tłíł:yii-Prog
nàná→					...	tliizh:0i-Rep
nìná→						tliizh:0i-REp

Dene/Tsuut'ina FST structure – inflectional \longleftrightarrow lexical tier: *tsiy* 'cry'

Outer Lexical prefix	Outer Inflectional prefix	Middle Lexical Prefix	Middle Inflectional Prefix	Inner Lexical Prefix	Inner Inflectional Prefix	<u>Stem</u>
0	0	0	0	0	Ipfv-0-nopref	<i>tsiy:0-Ipfv</i>
0 0 ná→	Distr:dà				1Sg:is 2Sg:ni 3Sg:i 1Pl:isaà 1Pl:isiì 2Pl:as 3Pl:0 4Sg:0	tsày:yi-y-Pfv tsíł:yi-Prog ←chish:0-Rep
			3Pl:gi 4Sg:ts'i			

Dene/Tsuut'ina FST structure – inflectional ←→ lexical tier: *tsiy* 'cry'

Outer Lexical <u>prefix</u>	Outer Inflectional <u>prefix</u>	Middle Lexical <u>Prefix</u>	Middle Inflectional <u>Prefix</u>	Inner Lexical <u>Prefix</u>	Inner Inflectional <u>Prefix</u>	<u>Stem</u>
0	0	0	0	0	Pfv-yi-y-nopref	<i>tsiy:0-lpfv</i>
0	Distr:dà			1Sg:yis		<i>tsày:yi-y-Pfv</i>
ná→				2Sg:yí		<i>tsíł:yi-Prog</i>
			3Pl:gi	3Sg:yí		<i>←chish:0-Rep</i>
			4Sg:ts'i	1Pl:yàà		
				1Pl:yìì		
				2Pl:yas		
				3Pl:yí		
				4Sg:yí		

Dene/Tsuut’ina FST structure – inflectional ←→ lexical tier: *tsiy* ‘cry’

Outer Lexical prefix	Outer Inflectional prefix	Middle Lexical Prefix	Middle Inflectional Prefix	Inner Lexical Prefix	Inner Inflectional Prefix	Stem
0	0	0	0	0	Prog-yi-noprefix	tsiy:0-IPfv
0	Distr:dà			1Sg:yis	tsày:yi-y-Pfv	
0				2Sg:yí		tsíł:yi-Prog
ná→			3Pl:gi 4Sg:ts'i	3Sg:yí 1Pl:yàà 1Pl:yìì 2Pl:yas 3Pl:á 4Sg:a	←chish:0-Rep	

Dene/Tsuut'ina FST structure – inflectional \longleftrightarrow lexical tier: *tsiy* 'cry'

Outer Lexical <u>prefix</u>	Outer Inflectional <u>prefix</u>	Middle Lexical <u>Prefix</u>	Middle Inflectional <u>Prefix</u>	Inner Lexical <u>Prefix</u>	Inner Inflectional <u>Prefix</u>	<u>Stem</u>
0	0	0	0	0	Prog-yi-outer	tsiy:0-Ipfv
0	Distr:dà				1Sg:s	tsày:yi-y-Pfv
0					2Sg:ni	tsíł:yi-Prog
ná →				3Sg:0	1Pl:saà	←chish:0-Rep
			3Pl:gi		1Pl:sì	
			4Sg:ts'i		2Pl:s	
					3Pl:0	
					4Sg:0	

inflectional \longleftrightarrow lexical tier: *nàgudiitłod* ‘jump-down’

Outer Lexical prefix	Outer Inflectional prefix	Middle Lexical Prefix	Middle Inflectional Prefix	Inner Lexical Prefix	Inner Inflectional Prefix	<u>Stem</u>
nà →	0	gu →	0	di →	0i-Ipfv-inner	← tłod:0i-Ipfv
ná→	Distr:dà	gu→		di→	1Sg:is	← tłot:sii-Pfv
nà→		gu→		di→	2Sg:í	← tlıł:yii-Prog
nàná→		gu→		di→	3Sg:i	← tlıizh:0i-Rep
nìná→		gu→		di→	1Pl:aà	← tlıizh:0i-Rep
					1Pl:iì	
					2Pl:as	
					3Pl:0	
					4Sg:0	

inflectional \longleftrightarrow lexical tier: *nàgudiitłod* ‘jump-down’

Outer Lexical prefix <i>nà</i> \rightarrow	Outer Inflectional prefix 0	Middle Lexical Prefix <i>gu</i> \rightarrow	Middle Inflectional Prefix 0	Inner Lexical Prefix <i>di</i> \rightarrow	Inner Inflectional Prefix <i>sii-Pfv-inner</i>	Stem
<i>nà</i> \rightarrow	Distr:dà	<i>gu</i> \rightarrow		<i>di</i> \rightarrow	1Sg:ìsis	\leftarrow tłod:0i-Ipfv
<i>nà</i> \rightarrow		<i>gu</i> \rightarrow		<i>di</i> \rightarrow	2Sg:ìsíí	\leftarrow tłòt:sii-Pfv
<i>nàná</i> \rightarrow		<i>gu</i> \rightarrow		<i>di</i> \rightarrow	3Sg:ìs	\leftarrow tñíl:yii-Prog
<i>nìná</i> \rightarrow		<i>gu</i> \rightarrow		<i>di</i> \rightarrow	1Pl:ìsaà	\leftarrow tñiizh:0i-Rep
					1Pl:ìsìí	\leftarrow tñiizh:0i-Rep
					2Pl:ìsas	
					3Pl:ìs	
					4Sg:ìs	

inflectional \longleftrightarrow lexical tier: *nàgudiitłod* ‘jump-down’

Outer Lexical <u>prefix</u>	Outer Inflectional <u>prefix</u>	Middle Lexical <u>Prefix</u>	Middle Inflectional <u>Prefix</u>	Inner Lexical <u>Prefix</u>	Inner Inflectional <u>Prefix</u>	<u>Stem</u>
nà →	0	gu →	0	di →	yii-Prog-inner	← tłod:0i-Ipfv
nà →	Distr:dà	gu →	di →	1Sg:yis	← tłot:sii-Pfv	
nà →		gu →	di →	2Sg:yíí	← tlıł:yii-Prog	
nàná →		gu →	di →	3Sg:áá	← tliizh:0i-Rep	
nìná →		gu →	di →	1Pl:yaà	← tliizh:0i-Rep	
			3Pl:gi	1Pl:yìì		
			4Sg:ts'i	2Pl:yas		
				3Pl:áá		
				4Sg:áá		

inflectional \longleftrightarrow lexical tier:

nàgudiitłod ‘jump-down’

Outer Lexical <u>prefix</u>	Outer Inflectional <u>prefix</u>	Middle Lexical <u>Prefix</u>	Middle Inflectional <u>Prefix</u>	Inner Lexical <u>Prefix</u>	Inner Inflectional <u>Prefix</u>	<u>Stem</u>
nà→	0	gu→	0	di→	0i-Ipfv-inner	← tłod:0i-Ipfv
nà→	Distr:dà	gu→	di→	1Sg:is	← tłot:sii-Pfv	
nà→		gu→	di→	2Sg:í	← tlıł:yii-Prog	
nàná→		gu→	di→	3Sg:i	← tlıizh:0i-Rep	
nìná→		gu→	di→	1Pl:aà	← tlıizh:0i-Rep	
			3Pl:gi	1Pl:iì		
			4Sg:ts'i	2Pl:as		
				3Pl:0		
				4Sg:0		

inflectional \longleftrightarrow lexical tier: *nàgudiitłod* ‘jump-down’

Outer Lexical prefix	Outer Inflectional prefix	Middle Lexical Prefix	Middle Inflectional Prefix	Inner Lexical Prefix	Inner Inflectional Prefix	Stem
nà→	0	gu→	0	di→	0i-Ipfv-inner	← tłod:0i-Ipfv
nà→	Distr:dà	gu→		di→	1Sg:is	← tłot:sii-Pfv
nà→		gu→		di→	2Sg:í	← tlıł:yii-Prog
nàná→		gu→		di→	3Sg:i	← tlıizh:0i-Rep
nìná→		gu→		di→	1Pl:aà	← tlıizh:0i-Rep
			3Pl:gi 4Sg:ts'i		1Pl:iì 2Pl:as 3Pl:0 4Sg:0	

Computational model → genuinely useful language applications

- intelligent web-based dictionary: I-DICT
- intelligent computer-aided language learning: ICALL
- spell-checking

Tsuut'ina – I-DICT

Search with inflected form

The screenshot shows the I-DICT application interface. At the top, there is a navigation bar with links for "Home", "Click-in-text", "About", and "Sources". A language selection dropdown shows "âēñō" and "English". Below the navigation bar, there are two main search boxes: "Tsuut'ina → English" and "English → Tsuut'ina". The "Tsuut'ina → English" box is highlighted with a blue background. In the search string input field, the Tsuut'ina word "nàguts'idáátlít" is entered. To the right of the input field are two buttons: "Search" and "Search texts". A red box labeled "Search string" points to the input field. A red arrow points from the "Search string" box to the input field. Below the search box, there is a "WRITTEN VARIANT" section containing "aio âēñō" and "aio āēñō".

Tsuut'ina → English

Search string: nàguts'idáátlít

Search texts

Previous:

nàgudiitłod (V-I)
s/he jumps down

Lemma + POS code + English translation

Analysis of form (with lexical structure made explicit and tags spelled out)

nàguts'idáátlít
← vñà|gu|dii|tłod +
Verb + Intransitive+
Progressive+ Fourth
Person Singular

N.B. The above visualization is a mock-up of an actual I-DICT application yet under development

Tsuut'ina – I-DICT – paradigm generation

The image shows a mock-up of an I-DICT application interface. At the top, there is a navigation bar with links for "Gunáhà", "Home", "Click-in-text", "About", "Sources", and language selection ("âēñō" and "EN"). Below the navigation bar, there are two main sections: "Dictionaries" and "Tsuut'ina → English".

In the "Tsuut'ina → English" section, the search bar contains the Tsuut'ina word "nàguts'idáátłít". Below the search bar, the result "nàgudiitłod (V-I) s/he jumps down" is displayed, with a detailed morphological analysis: "nàguts'idáátłít" (← nà|gu|dii|tłod + Verb + Intransitive+ Progressive+ Fourth Person Singular).

On the left side, under "WRITTEN VARIANT", there are two options: "aio âēñō" and "aio āēñō".

At the bottom, there is a table titled "Aspect/Person" showing paradigm generation:

	1Sg	2Sg	3sg
Imperfective	nàgudistłod	nàgudíítłod	nàgudiitłod
Perfective	nàgudìsistłòt	nàgudìsíítłòt	nàgudìstłòt
Progressive	nàgudiyistłít	nàgudiyíítłít	nàgudáátłít

N.B. The above visualization is a mock-up of an actual I-DICT application yet under development

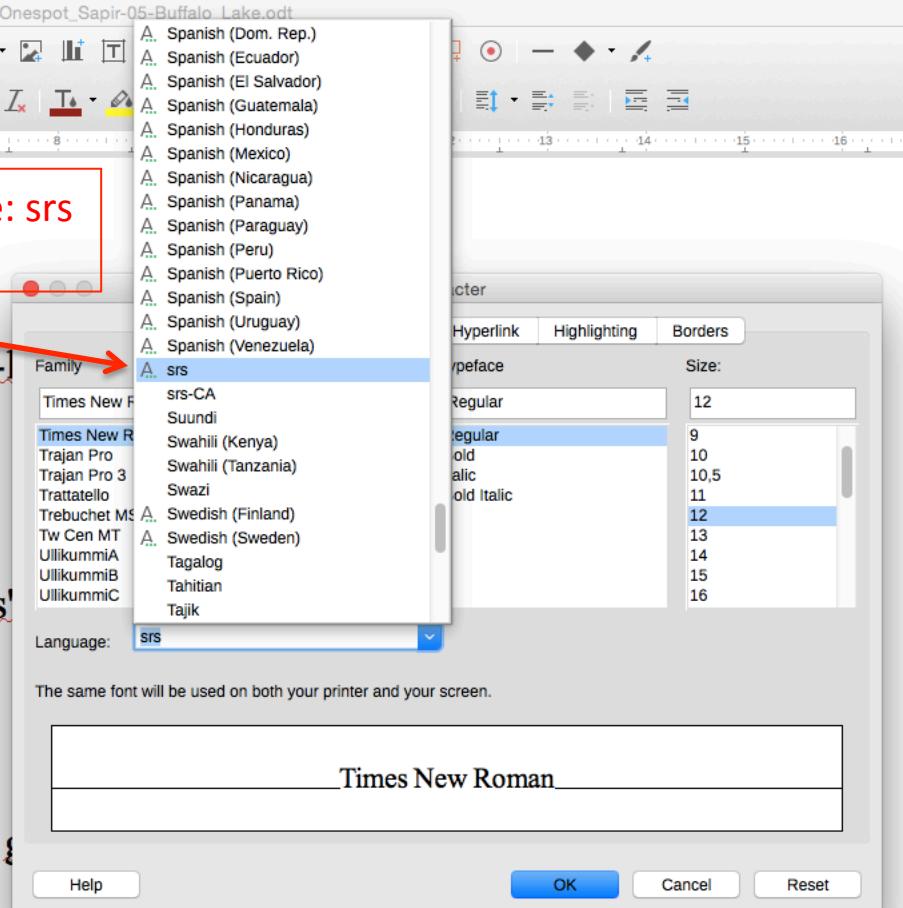
Useful applications – spell-checking

Diní isìna dìsày-là.
Ùwat'iyi doò isina thìk'àzà xaní nàzidí yi?ín-là.
Ùwat'iyi yidìsón-là.
Yizìsyí-là.
Wúnìt'ósì tsidìsti-gù nàyìnìstín-là.
Ùwat'iyi náyìs?òl-là.
Yimì ìts'ògùt'ódzi ínídza tú mimìt'ó gúts'í ìts'ògùmot-là.
Ùwat'iyi guk'á yiyinít-là.
Xanídá isìna diní diná ii náts'idìstl'ó-là.
Tú guká yìnón-là.
Túchu gwájàg-là.
Ùwat'iyi gúts'í Xanídik'àsidó-di gusch'ínìsh gu?águjàg.

Tsuut'ina – spell-checking

Tsuut'ina speller identified with the ISO code: srs
Letter 'A' denotes the existence of a speller

Diní isìna dìsày-là.
Ùwat'iyi doò isina thik'àzà xaní nàzidí yi?ín-là.
Ùwat'iyi yidìsón-là.
Yizisyí-là.
Wúnit'ósì tsidìsti-gù nàyìnìstín-là.
Ùwat'iyi náyìs?òl-là.
Yimì its'ògùt'ódzi ínídza tú mimít'ó gúts'í its'í.
Ùwat'iyi guk'á yiyinítlà.
Xanìdá isìna diní diná ii náts'idìstl'ó-là.
Tú guká yìnón-là.
Túchu gwájàg-là.
Ùwat'iyi gúts'í Xanídik'àsidó-di gusch'ínish g-



Font Family: Times New Roman
Size: 12
Language: SRS

The same font will be used on both your printer and your screen.

Help OK Cancel Reset

Find All Match Case Navigate by Default Style English (Canada)

Page 1 of 1 44 words, 379 characters

Mac OS X Dock icons

Tsuut'ina – spell-checking

Diní isìna dìsày-là.
Ùwat'iyi doò isìna thìk'azá xaní nàzidí yi?ín-là.
Ùwat'iyi yidìsón-là.
Yizìsyí-là.
Wúnìt'ósì tsidisti-gù nàyìnistín-là.
Ùwat'iyi náyìs?òl-là.
Yimì ìts'ögùt'ódzi ínídza tú mimìt'ó gúts'í ìts'ögùmot-là.
Ùwat'iyi guk'á yiyinít-là.
Xanìdá isìna diní diná ii náts'idìstl'ó-là.
Tú guká yìnón-là.
Túchu gwájàg-là.
Ùwat'iyi gúts'í Xanídik'àsidó-di gusch'ínìsh gu?águjàg.

Tsuut'ina – spell-checking – typo correction

Diní isina dìsày-là.
Úwat'iyi doo isina tlik'azá xaní nàzidí yi'yín-là.
Ùwat'iyi yidìsón-là.
Yizìsyí-là.
Wúnìt'ósi tsidìsti-gù nàyìnìstín-là.
Ùwat'iyi náyìs?òł-là.

Typos: (1) missing tone marking: *isina*, *doo*; (2) wrong tone marking (Ú<Ù): *Úwat'iyi*; (3) plain letter instead of diacritic one (l<ł) : *tlik'azá*; (3) apostrophe for glottal stop ('<?) : *yi'yín-là*

Tsuut'ina – spell-checking

Diní isìna dìsày-là.

Úwati xaní nàzidí yi'ín-là.

Ùwat' nistín-là.

Yizìs tú mimìt'ó gúts'í ìts'ògùmot-là.

Wúnì Yimiì t'ó gúts'í ìts'ògùmot-là.

Ùwat'iyi guk'á yiyinił-là.

Xanidá isìna diní diná ii náts'idistł'ó-là.

A context menu is open over the word 'isìna' in the first sentence. The menu options are:

- isìna
- isìna
- Ignore
- Ignore All
- Add to Dictionary
- Spelling and Grammar...
- Always correct to
- AutoCorrect Options...
- Set Language for Selection
- Set Language for Paragraph

Tsuut'ina – spell-checking

Diní isìna dìsày-là.
Úwat'iyi doo ìsina tlìk'àzà xaní nàzidí yi'ín-là.

Úwat'iyi

- Ignore
- Ignore All
- Add to Dictionary
- Spelling and Grammar...

nàyìnístín-là.

ídza tú mimìt'ó gúts'í ìts'ògùmot-là.

ítl-là.

Xanidá isìna diní diná ii náts'idistl'ó-là.

A screenshot of the LibreOffice Writer application interface. The title bar shows "LibreOffice" and the file name "Onespot_Sapir-05-Buffalo_". The toolbar includes various document management and styling tools. The main content area displays Tsuut'ina text. A context menu is open over the word "Úwat'iyi", listing options for spelling and grammar. The menu items are: Úwat'iyi, Ignore, Ignore All, Add to Dictionary, Spelling and Grammar..., Always correct to AutoCorrect Options..., Set Language for Selection, and Set Language for Paragraph. The text in the document is in Times New Roman font, size 12. The menu has a light gray background with black text, and the selected item "Úwat'iyi" is highlighted.

Tsuut'ina – spell-checking

Diní isìna dìsày-là.

Ùwat'iyi doo ìsina tlìk'àzá xaní nàzidí yi'ín-là.

Ùwat'iyi

Yizìsyí-là

Wúnit'ós

Ùwat'iyi

Yimì ìts'ògùmot-là.

Ùwat'iyi

Xanìdá isìna diní diná ii náts'ídìstl'ó-là.

The screenshot shows a LibreOffice Writer window with a document containing Tsuut'ina text. A right-click context menu is open over the word "tlik'azá", which is underlined with a red dotted line indicating it's a misspelling. The menu options include: doò, zos, Ignore, Ignore All, Add to Dictionary, Spelling and Grammar..., Always correct to, AutoCorrect Options..., Set Language for Selection, and Set Language for Paragraph. The menu is partially cut off on the right side.

Tsuut'ina – spell-checking

Diní isìna dìsày-là.

Ùwat'iyi doò| isina tlìk'azá xaní nàzidí yi'in-là.

Ùwat'iyi yidìsón-là

Yizìsyí-là.

Wúnìt'osì tsidìsti-g

Ùwat'iyi náyìs?òl-1

Yimì its'ògùt'ódzi i

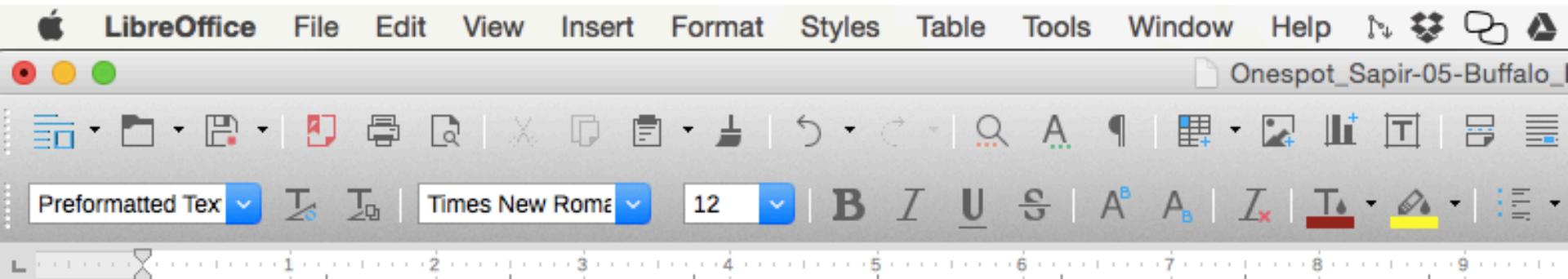
Ùwat'iyi guk'á yiyi....

Xanidá isìna diní diná ii náts'idistl'ó-là.

A context menu is open over the word 'tlìk'azá'. The menu items are:

- tfìk'azá
- Ignore
- Ignore All
- Add to Dictionary
- Spelling and Grammar...
- Always correct to AutoCorrect Options...
- Set Language for Selection
- Set Language for Paragraph

Tsuut'ina – spell-checking



Diní isìna dìsày-là.

Ùwat'iyi doò ìsina tlik'azá xaní nàzidí yi'yín-là.

Ùwat'iyi yidìsón-là.

Yizìsyí-là.

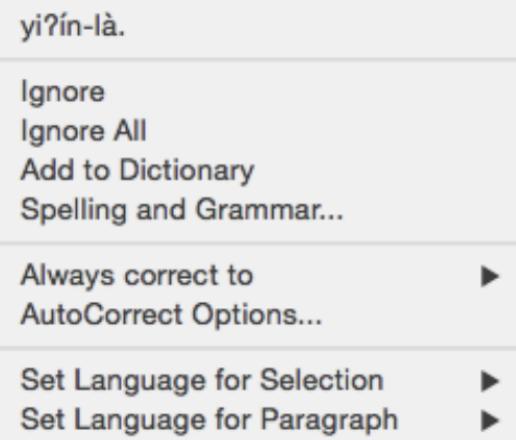
Wúnìt'ósi tsidìsti-gù nàyìnístín-là.

Ùwat'iyi náyìs?òł-là.

Yimì ìts'ògùt'ódzi ínídza tú mimìt'ó gi'

Ùwat'iyi guk'á yiyiníł-là.

Xanidá isìna diní diná ii náts'idistl'ó-là.



→ Other – Dene – languages?

- All the software is open source
- start-up work for other (Dene) languages?
- computational modeling framework
 - several days of collaboration by field linguists and computational linguists
- full computational model for some language
 - several months of work by a linguist
- end-user-friendly applications per language
 - several months of work by a linguist and a several weeks of work by a programmer
- BUT: all this builds upon decades of linguistic documentation by/with Elders and other native speakers
- AND: the tools benefit from continued improvement and fine-tuning

siyísgaas – kiitos

References

- Arppe, Antti, Jordan Lachler, Trond Trosterud, Lene Antonsen & Sjur N. Moshagen. 2016. Basic Language Resource Kits for Endangered Languages: A Case Study of Plains Cree. Workshop on Collaboration and Computing for Under-Resourced Languages (CCURL), 23 May 2016, Language Resources and Evaluation Conference (LREC), Portoroz, Slovenia.
- Cook, Eung-Do. 1984. *A Sarcee grammar*. Vancouver, BC: University of British Columbia Press.
- Faltz, Leonard. 1998. *The Navajo verb. A grammar for students and scholars*. Albuquerque, NM: University of New Mexico Press.
- Huldén, Mans. 2009. Foma: a finite-state compiler and library. Association for Computational Linguistics, 29–32.
- Kari, James. 1975. “The disjunct boundary in the Navajo and Tanaina verb prefix complexes.” *International Journal of American Linguistics* 41(4), 330–345.
- Kari, James. 1989. “Affix positions and zones in the Athapaskan verb complex: Ahtna and Navajo.” *International Journal of American Linguistics* 55(4), 424–454.
- Leer, Jeff. 1999. “Comparative Athabaskan class notes.” Unpublished manuscript CA965L1999a. Fairbanks, AK: Alaska Native Language Archive.
- Lindén Krister, Erik Axelson, Sam Hardwick, Miikka Silfverberg, and Tommi Pirinen. 2011 HFST - Framework for Compiling and Applying Morphologies. Proceedings of Second International Workshop on Systems and Frameworks for Computational Morphology (SFCM), 67-85.
- McDonough, Joyce. 2000. “On the bipartite model of the Athabaskan verb.” In Fernald, Theodore B. and Paul R. Platero (eds.), *The Athabaskan languages: Perspectives on a Native American language family*, pp. 139–166. Oxford University Press.
- Rice, Keren. 2001. “Slave (Northern Athabaskan).” In Andrew Spencer and Arnold M. Zwicky (eds.), *The handbook of morphology*, 648-689. Malden, MA: Blackwell.
- Rice, Keren. 2005. “A typology of good grammars.” *Studies in Language* 30(2), 385–415.
- Young, Robert W. and William Morgan. 1987. *The Navajo language*. Albuquerque, NM: University of New Mexico Press.