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

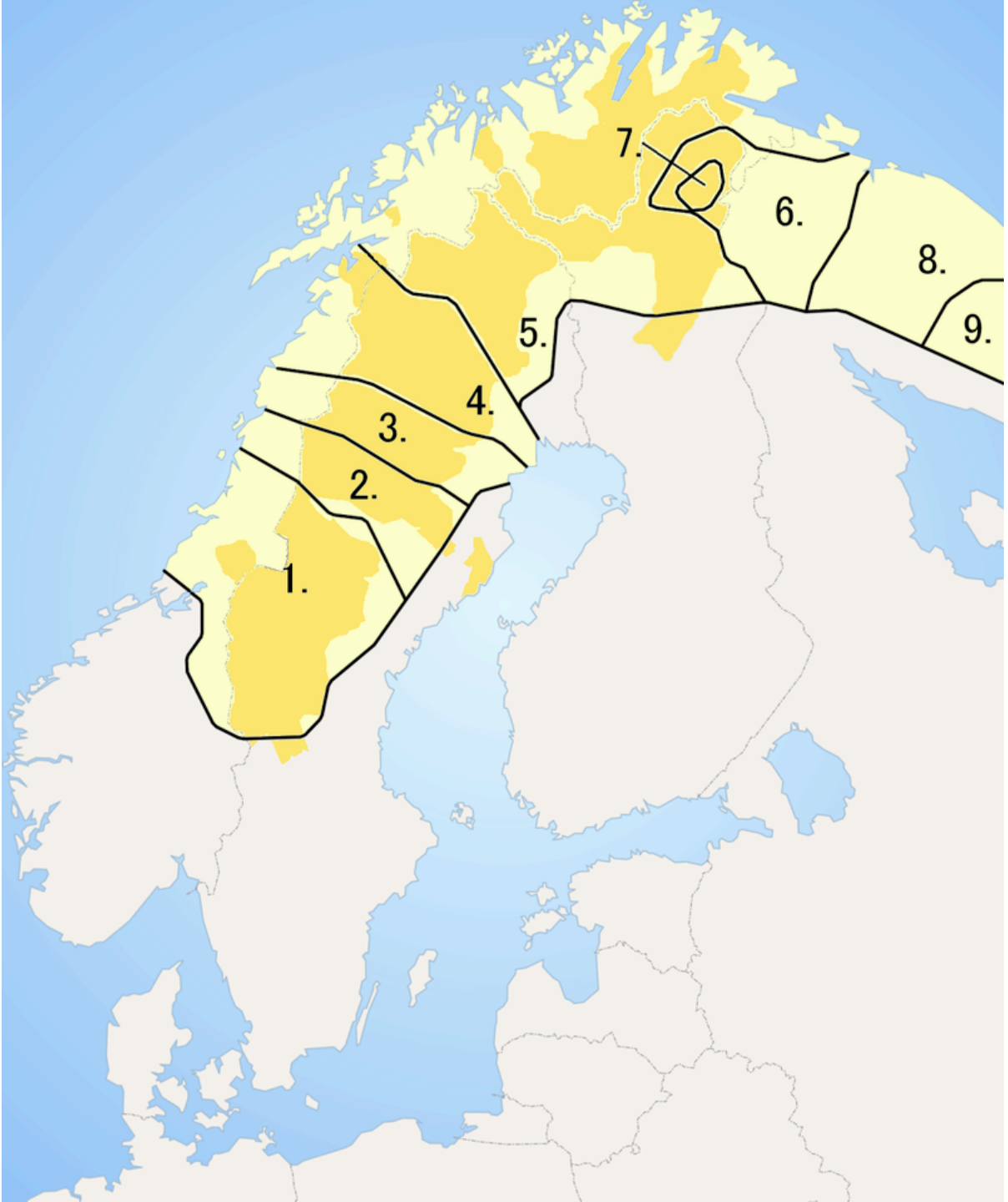
University of Alberta

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

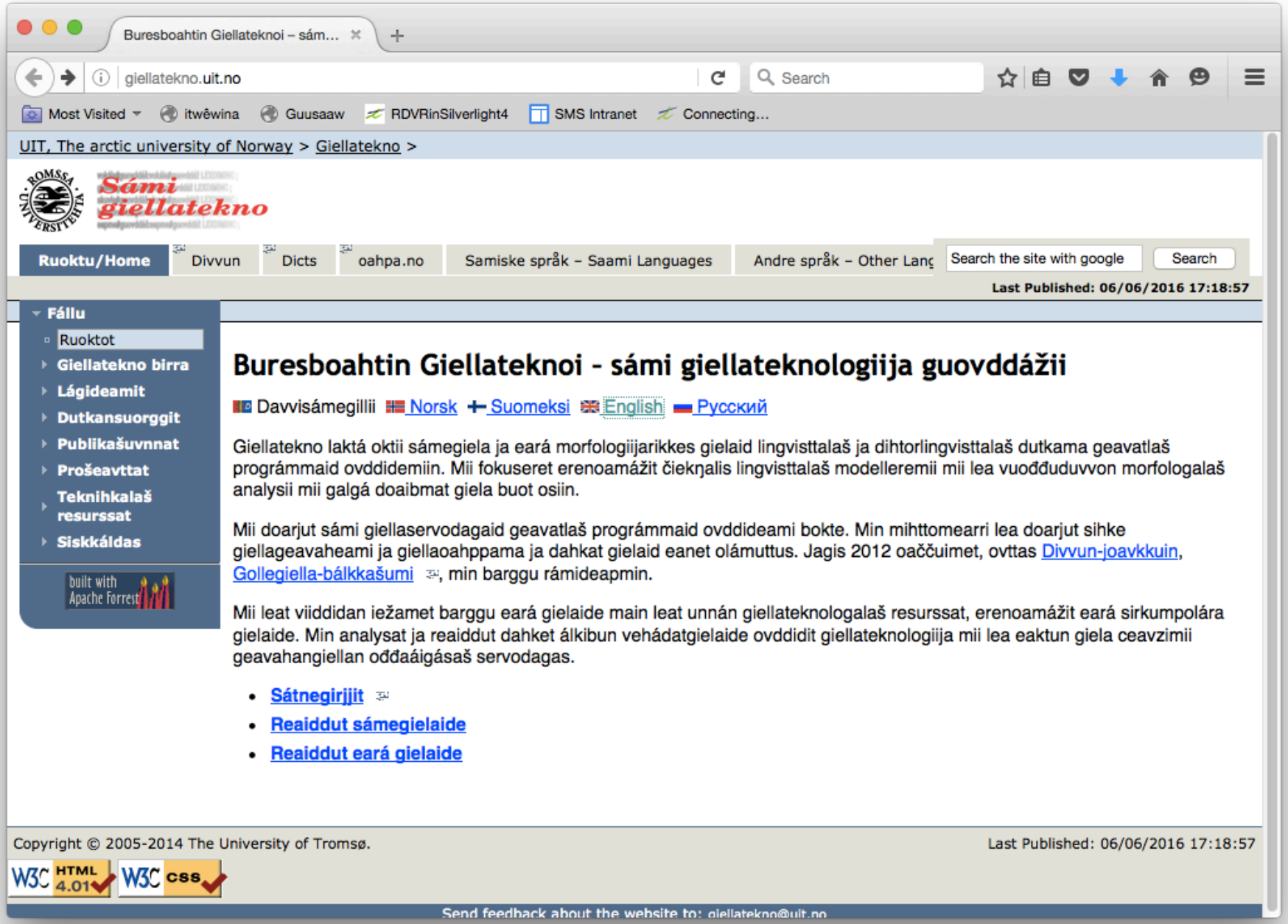
Giellatekno and Divvun have worked over a decade on developing language technological tools for the indigenous Saami languages spoken in Northern Scandinavia

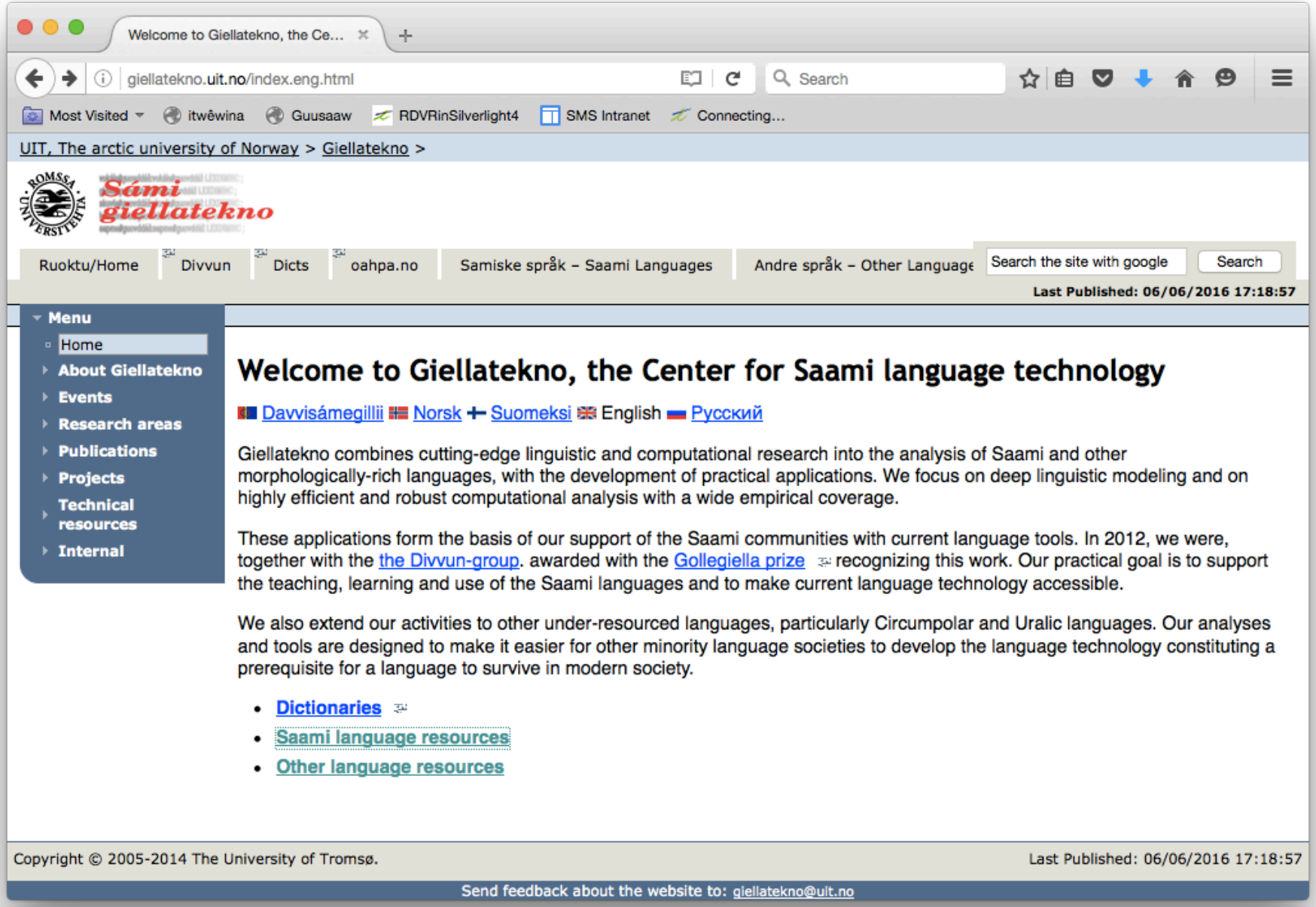


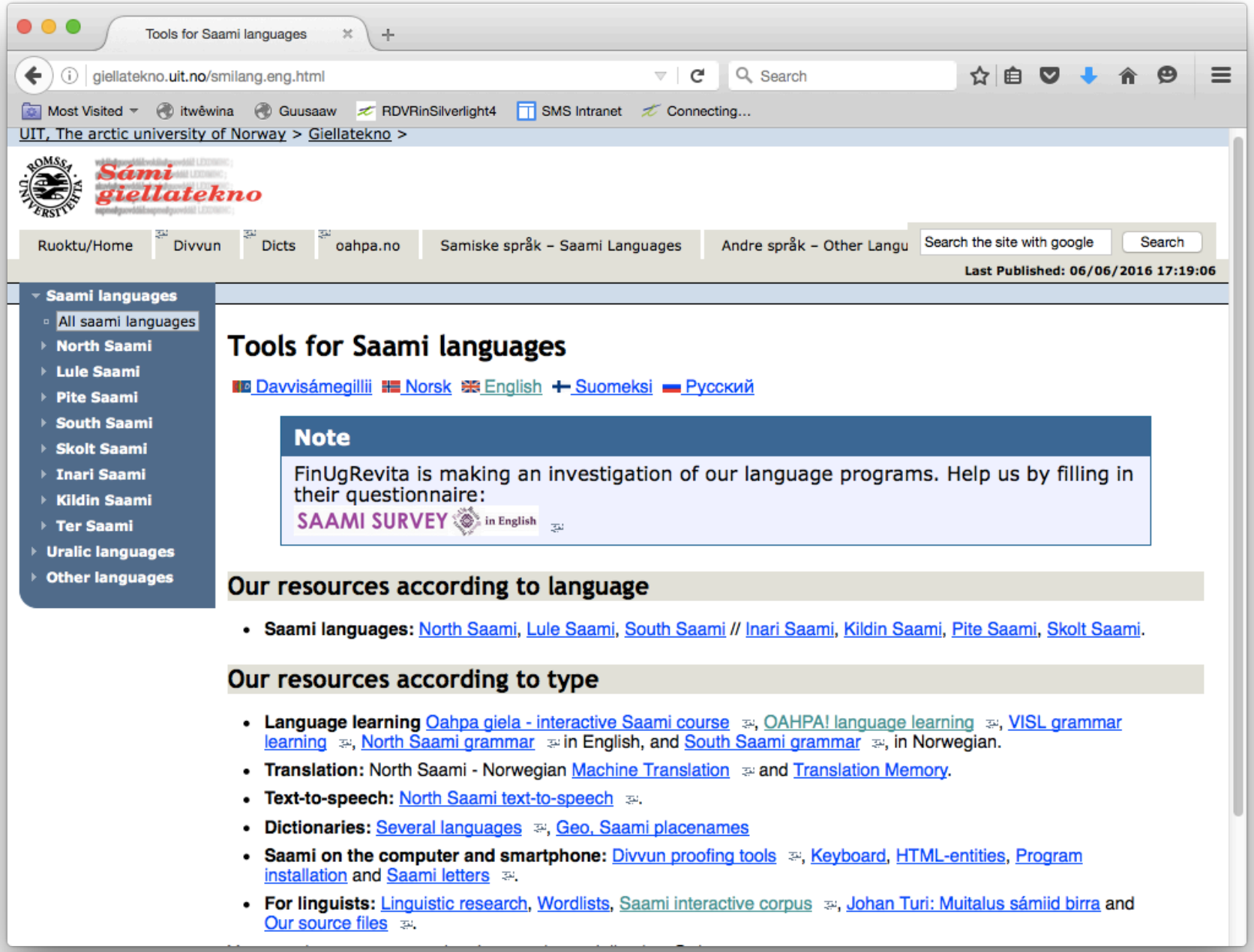
Historically verified distribution of the Saami languages:

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 Giellateknoi – sám…
giellatekno.uit.no
Search
Most Visited
itwêwina
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
Teknihkalaš resurssat
Siskkáldas
built with Apache Forrest
Buresboahtin Giellateknoi - sámi giellateknologiija guovddážii
Davvisámegillii Norsk Suomeksi English Русский
Giellatekno laktá oktii sámegiela ja eará morfologiijarikkes gielaid lingvisttalaš ja dihtorlingvisttalaš dutkama geavatlaš prográmmaid ovddidemiin. Mii fokuseret erenoamážit čieknalis lingvisttalaš modelleremii mii lea vuođđuduvvon 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 olámuttus. Jagis 2012 oačcuimet, ovttas Divvun-joavkkuin, Gollegiella-bálkkašumi, min barggu rámideapmin.
Mii leat viiddidan iežamet barggu eará gielaide main leat unnán giellateknologalaš resurssat, erenoamážit eará sirkumpolára gielaide. Min analysat ja reaiddut dahket álkibun vehádatgielaide ovddidit giellateknologiija mii lea eaktun giela ceavzimii geavahangiellan ođđaáigásaš servodagas.
Sátnegirjjit
Reaiddut sámegielaide
Reaiddut eará gielaide
Copyright © 2005-2014 The University of Tromsø.
Last Published: 06/06/2016 17:18:57
W3C HTML 4.01
W3C CSS
Send feedback about the website to: giellatekno@uit.no


Welcome to Giellatekno, the Ce...
giellatekno.uit.no/index.eng.html
Search
Most Visited
itwêwina
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 Language
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
Search
Most Visited
itwêwina
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, OAHPA! 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. Schweitzer, University of Alaska Fairbanks

Arctic peoples subdivided according to language families

Indo-European family
- Germanic branch

Uralic family
- Finno-Ugric branch
- Samoyedic branch

Altaic 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

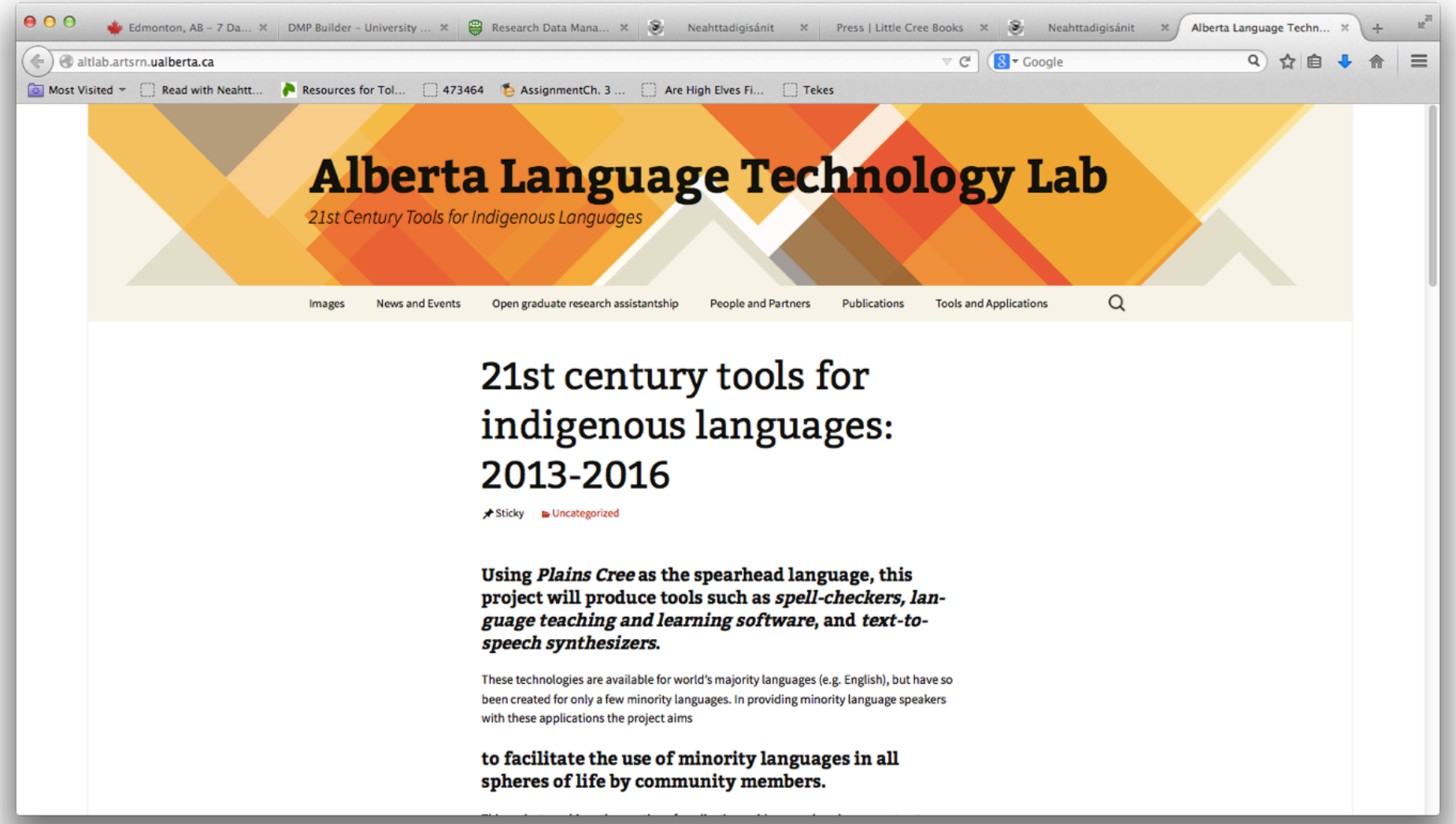
Notes:

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).

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



WWW: altlab.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

gámíɬ

“they’re swimming” (PROGRESSIVE)

Tsuut’ina; Cook (1984: 278–281)

Templatic morphology

gámíł

"they're swimming" (PROGRESSIVE)

Incorporated PP	Adverbial	Iterative (*ná-*)	Incorporated stem	Distributive (*dà-*)	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

Tsuut'ina; Cook (1984: 278–281)

Templatic morphology

gámíł

“they’re swimming” (PROGRESSIVE)

1	ABSTRACT PARTS	**gi-**	**yi-**	**míł**
		they	he,she,it:PROG	swim:PROG

Tsuut’ina; Cook (1984: 278–281)

Templatic morphology

gámíł

“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 → **gá**míł (*i*-Deletion)

Tsuut’ina; Cook (1984: 278–281)

Templatic morphology

gámíł

“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 (*ɣi*-Augmentation)
2. gi- á- míł → **gá**míł (*i*-Deletion)

3 ACTUAL WORDS

gámíł
“they’re swimming”

Tsuut’ina; Cook (1984: 278–281)

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-**	íł	"I'm swimming"
2S	**yí-**	íł	"you're swimming"
3S	**yá-**	íł	"he/she/it is swimming"
1P	**yaà-**	íł	"we're swimming"
2P	**yas-**	íł	"you guys are swimming"
3P	**gá-**	íł	"they're swimming"
4	**ts'á-**	íł	"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'ámíł			ts'idádlùw	4		xats'ánáł
With *s-*				"Be almost swallowing O" (PROG):			"Be taming, training O" (PROG):		
	1S		yiswùsh	1S		diyisnìł	1S		áyisdíł
	2S		yíswùsh	2S		diyísnìł	2S		áyísdíł
	3S	O	yáswùsh *(O yiswùsh?)*	3S	O	dásnìł	3S	O	áyisdíł
	1P		yaàwùsh	1P		diyaànìł	1P		áyaàdíł
	2P		yaswùsh	2P		diyasnìł	2P		áyasdíł
	3P	O	gáswùsh	3P	O	gidásnìł	3P	O	agásdíł
	4		ts'áswùsh	4		ts'idásnìł	4		áts'ásdíł
With *i-*	"Be running along" (PROG):			"Be starting to run along" (PROG):			"Be running along home" (PROG):		
	1S		yistł'áł	1S		didiyistł'áł	1S		náyistł'áł
	2S		yíítł'áł	2S		didiyíítł'áł	2S		náyíítł'áł
	3S		yiitł'áł	3S		didáátł'áł	3S		náyiitł'áł
	1P		yaàtł'áł	1P		didiyaàtł'áł	1P		náyaàtł'áł
	2P		yastł'áł	2P		didiyastł'áł	2P		náyastł'áł
	3P		gáátł'áł	3P		gididáátł'áł	3P		nágáátł'áł
	4		ts'áátł'áł	4		ts'ididáátł'áł	4		náts'áátł'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]:nànáa=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.tl'ò	INTR-yii-POT; ! (POT; yii-POT)

Dene/Tsuut'ina FST structure – lexical tier

Outer Lexical prefix	Outer Inflectional prefix	Middle Lexical Prefix	Middle Inflectional Prefix	Inner Lexical Prefix	Inner Inflectional Prefix	Stem
0		0		0		tsiy
0						tsày
0						tsíł
ná→						←chish
ts'á→						←zíd
ts'á→						←zid
ts'á→						←zíl
tsáná→						←zhiizh
nà→		gu→		di→		←tłod
nà→						←tłàt
nà→						←tłíł
nàná→						←tłiizh
nìná→						←tłiizh

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	SUBJECT:	tsiy
0	Distributive				1Sg	tsày
0					2Sg	tsíł
ná→					3Sg	←chish
					1Pl	
ts'á→					2Pl	←zíd
ts'á→			DIRECT			←zid
ts'á→			OBJECT			←zíl
tsáná→					ASPECT	←zhiizh
					Imperfective	
nà→		gu→		di→	Perfective	←tłod
nà→					Progressive	←tłàt
nà→					Repetitive	←tłíł
nàná→						←tłiizh
nìná→					CONJ-CLASS CLASSIFIERS	←tłiizh

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→	0 Distr:dà	0	0 3Pl:gi 4Sg:t'si	0	0-Ipfv-no 0-Ipfv-inner 0-Ipfv-outer 0s-Ipfv-no 0s-Ipfv-inner 0s-Ipfv-outer 0i-Ipfv-no 0i-Ipfv-inner 0i-Ipfv-outer …	tsiy tsày tsíł ←chish ←zíd ←zid ←zíl ←zhiizh
ts'á→ ts'á→ ts'á→ tsáná→						
nà→ nà→ nà→ nàná→ nìná→		gu→		di→	yi-y-Pfv-no yi-y-Pfv-inner yi-y-Pfv-outer …	←tłod ←tłàt ←tłíł ←tłiizh ←tłiizh

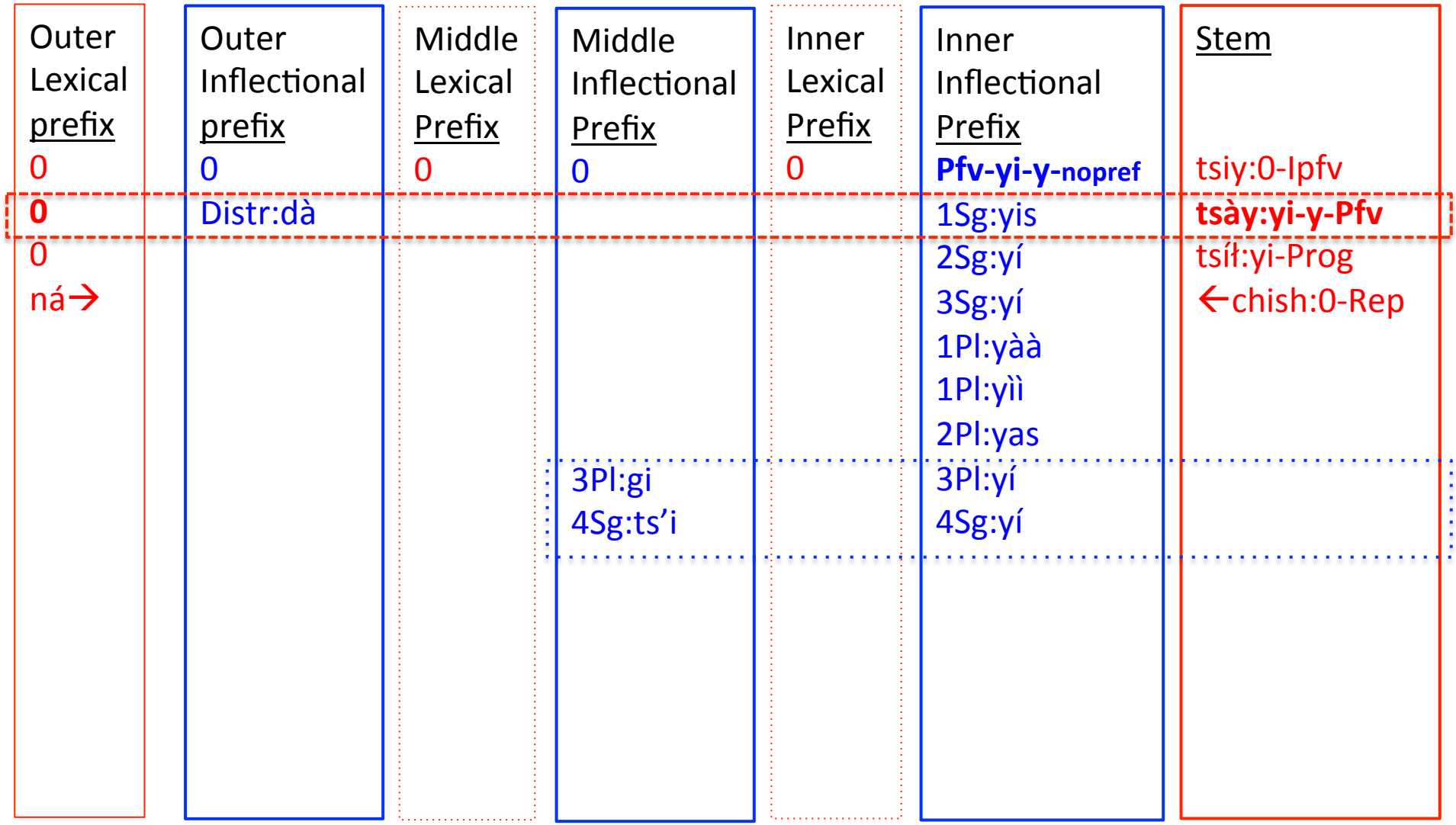
Dene/Tsuut'ina FST structure – inflectional ←→ lexical tier

Outer Lexical prefix	Outer Inflectional prefix	Middle Lexical Prefix	Middle Inflectional Prefix	Inner Lexical Prefix	Inner Inflectional Prefix	Stem
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→					0s-Ipfv-no	←chish:**0-Ipfv**
					0s-Ipfv-inner	
ts'á→					0s-Ipfv-outer	←zíd:**ni-Ipfv**
ts'á→					0i-Ipfv-no	←zid:**ni-Pfv**
ts'á→					0i-Ipfv-inner	←zíl:**yii-Prog**
tsáná→					0i-Ipfv-outer	←zhiizh:**ni-Ipfv**
					...	
nà→		gu→		di→	yi-y-Pfv-no	←tłod:**0i-Ipfv**
nà→					yi-y-Pfv-inner	←tłàt:**sii-Pfv**
nà→					yi-y-Pfv-outer	←tłíł:**yii-Prog**
nàná→					...	←tłiizh:**0i-Rep**
nìná→						←tłiizh:**0i-REp**

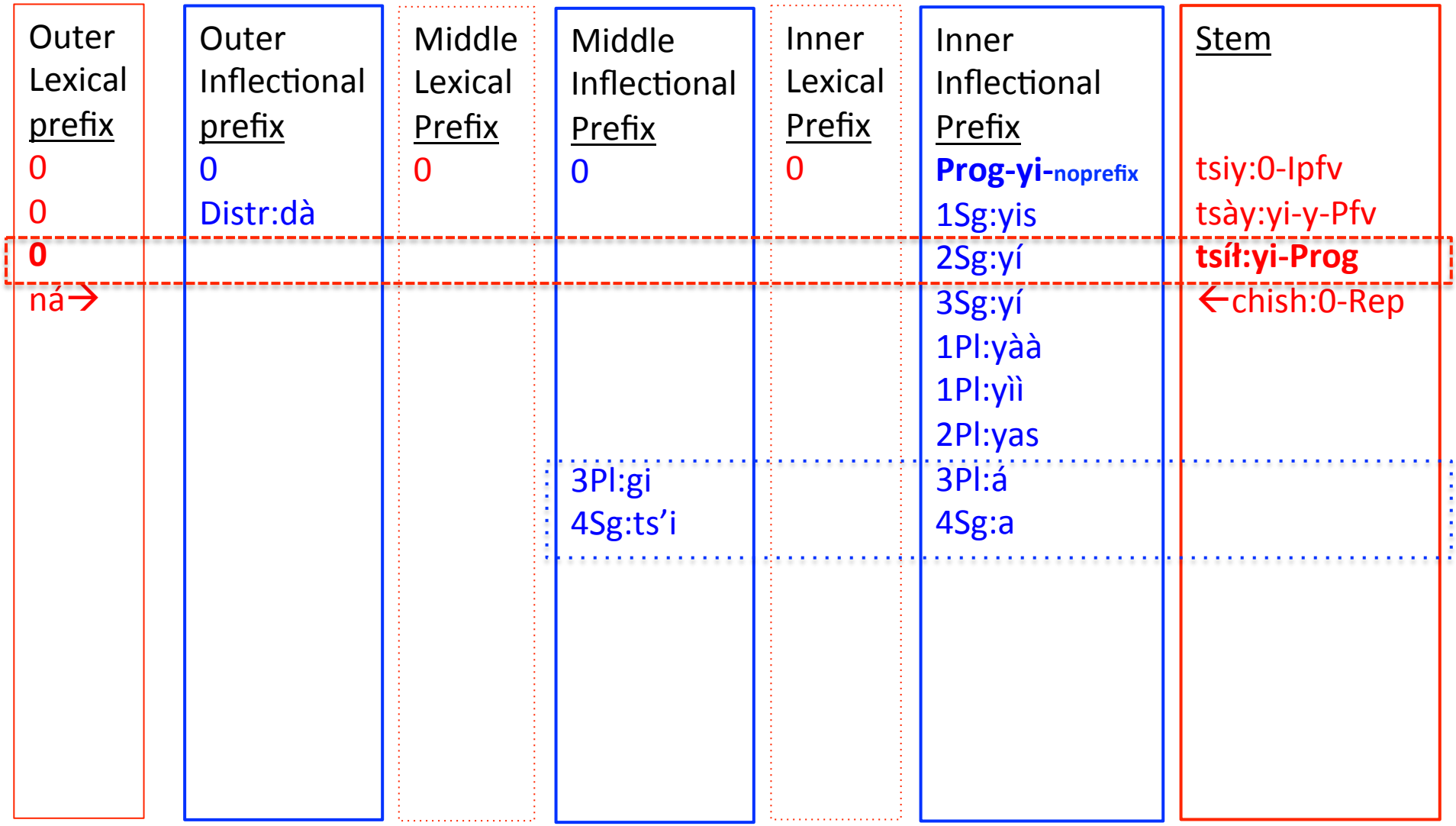
Dene/Tsuut'ina FST structure – inflectional ←→ lexical tier: *tsiy* 'cry'

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

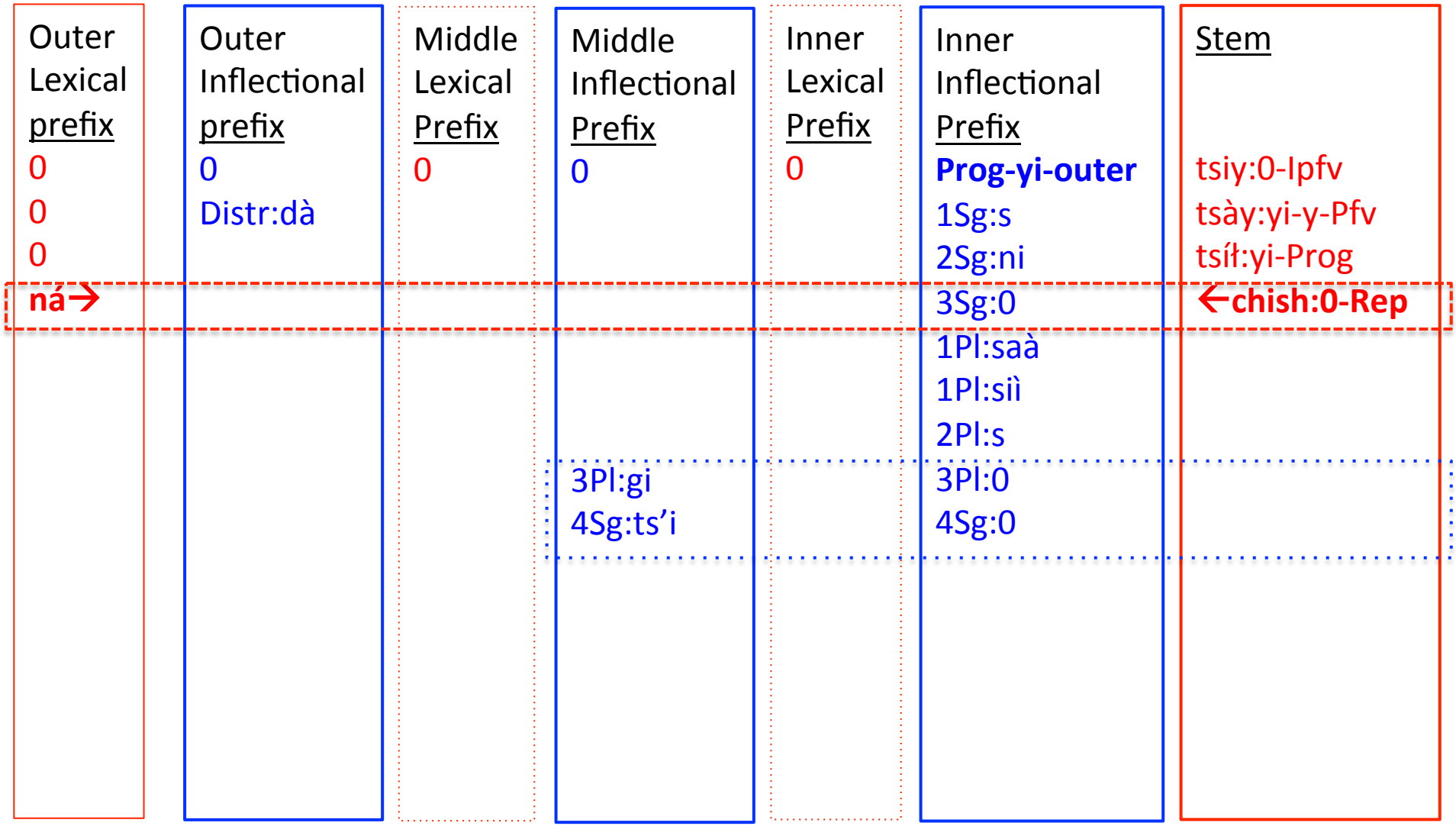
Dene/Tsuut'ina FST structure – inflectional ←→ lexical tier: *tsiy* 'cry'



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



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



inflectional ←→ 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	← tłòt:sii-Pfv
nà→		gu→		di→	2Sg:í	← tłíł:yii-Prog
nàná→		gu→		di→	3Sg:i	←tłiizh:0i-Rep
nìná→		gu→		di→	1Pl:aà	←tłiizh:0i-Rep
					1Pl:ìì	
					2Pl:as	
			3Pl:gi		3Pl:0	
			4Sg:ts'i		4Sg:0	

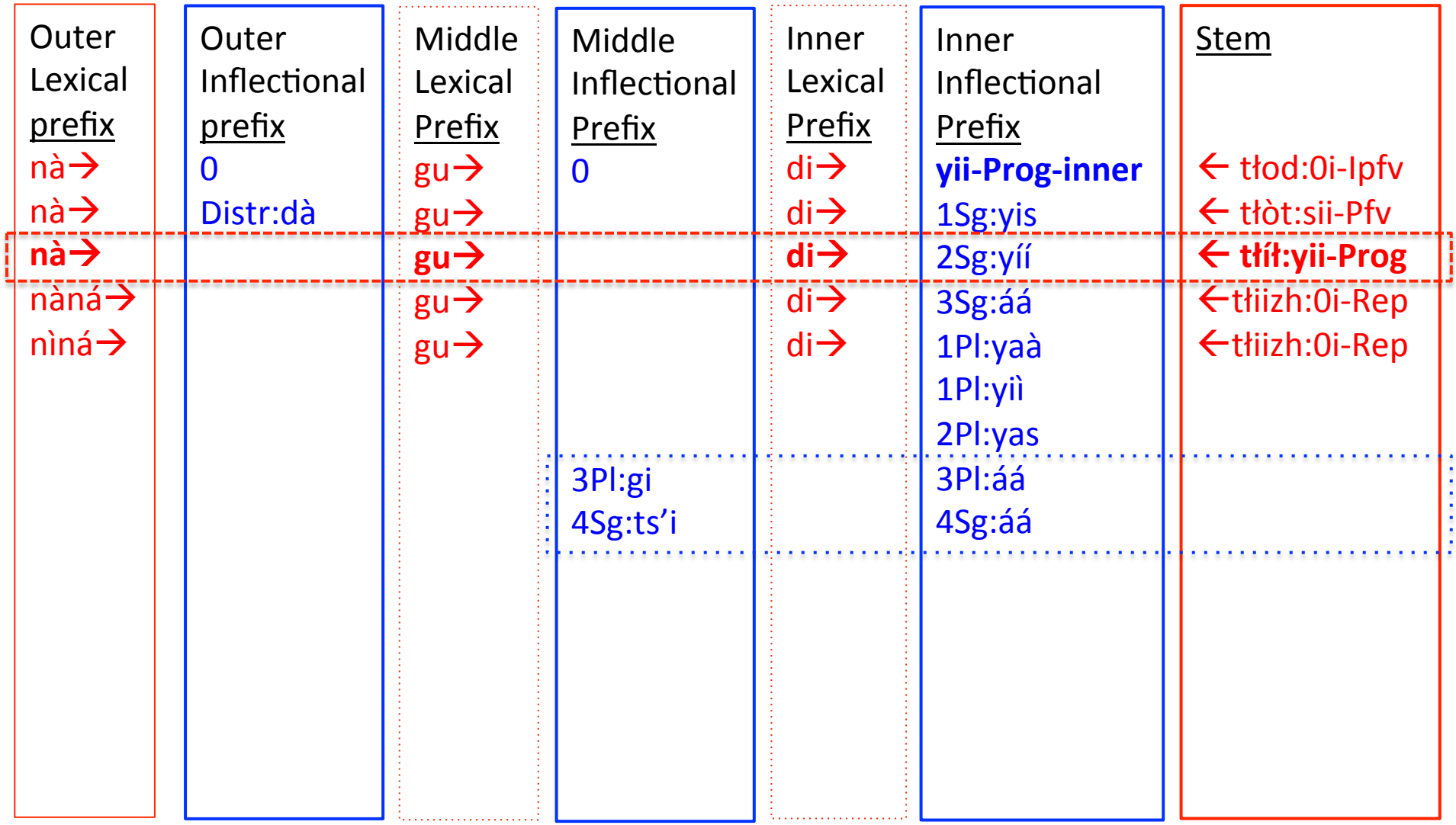
inflectional ←→ lexical tier:

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→	**sii-Pfv-inner**	← tłod:0i-Ipfv
nà→	Distr:dà	**gu→**		**di→**	1Sg:ìsis	**← tłòt:sii-Pfv**
nà→		gu→		di→	2Sg:ìsíí	← tłíł:yii-Prog
nàná→		gu→		di→	3Sg:ìs	←tłiizh:0i-Rep
nìná→		gu→		di→	1Pl:ìsaà	←tłiizh:0i-Rep
					1Pl:ìsiì	
					2Pl:ìsas	
			3Pl:gi		3Pl:ìs	
			4Sg:ts'i		4Sg:ìs	

inflectional ←→ lexical tier:

nàgudiitłod 'jump-down'



inflectional ←→ 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	← tłòt:sii-Pfv
nà→		gu→		di→	2Sg:í	← tłíł:yii-Prog
nàná→		**gu→**		**di→**	3Sg:i	**←tłiizh:0i-Rep**
nìná→		gu→		di→	1Pl:aà	←tłiizh:0i-Rep
					1Pl:ìì	
					2Pl:as	
			3Pl:gi		3Pl:0	
			4Sg:ts'i		4Sg:0	

inflectional ←→ lexical tier:

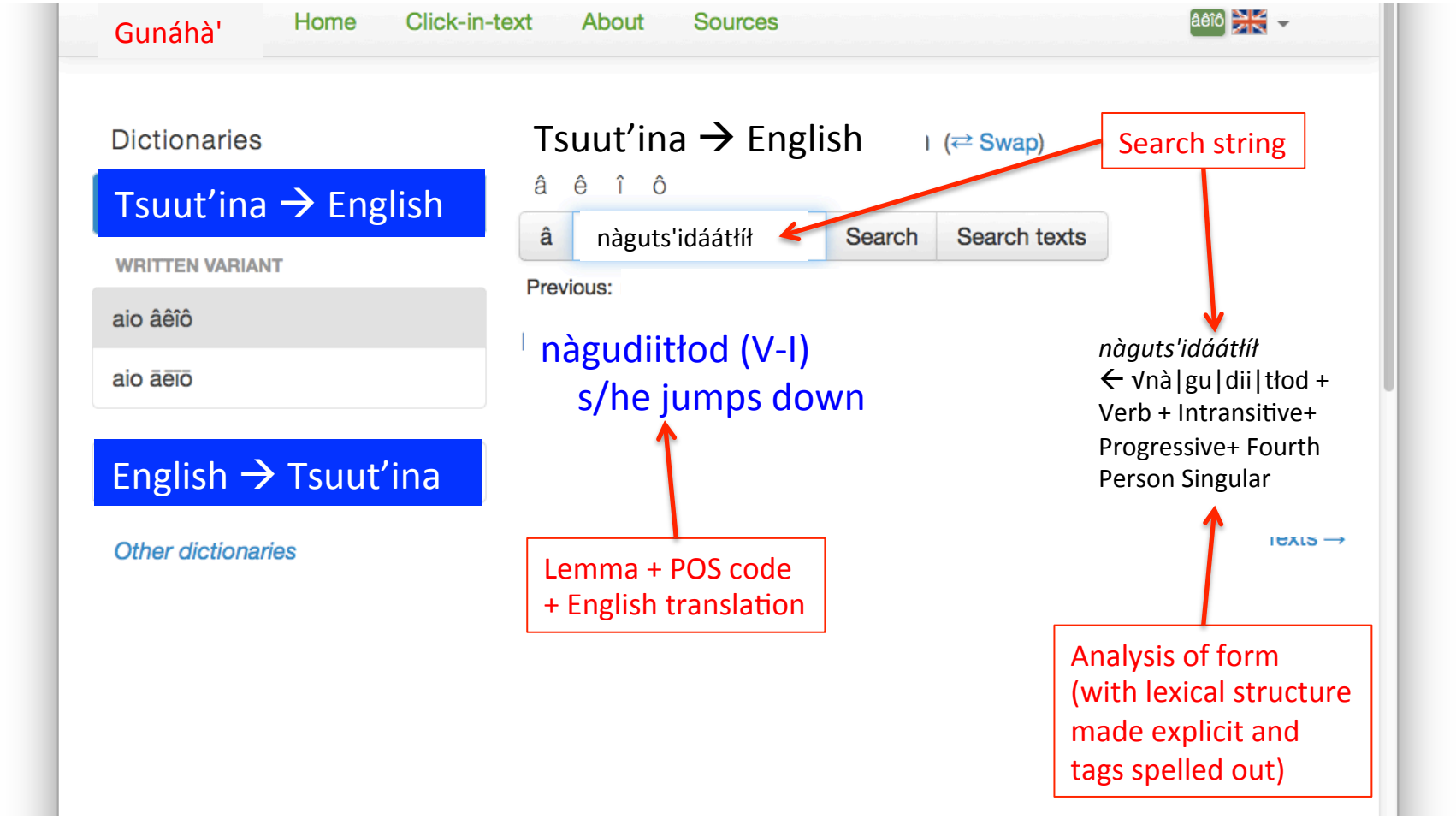
nàgudiitłod 'jump-down'

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	← tłòt:sii-Pfv
nà→		gu→		di→	2Sg:í	← tłíł:yii-Prog
nàná→		gu→		di→	3Sg:i	←tłiizh:0i-Rep
nìná→		**gu→**		**di→**	1Pl:aà	**←tłiizh:0i-Rep**
					1Pl:ìì	
					2Pl:as	
			3Pl:gi		3Pl:0	
			4Sg:ts'i		4Sg:0	

Computational model → genuinely useful language applications

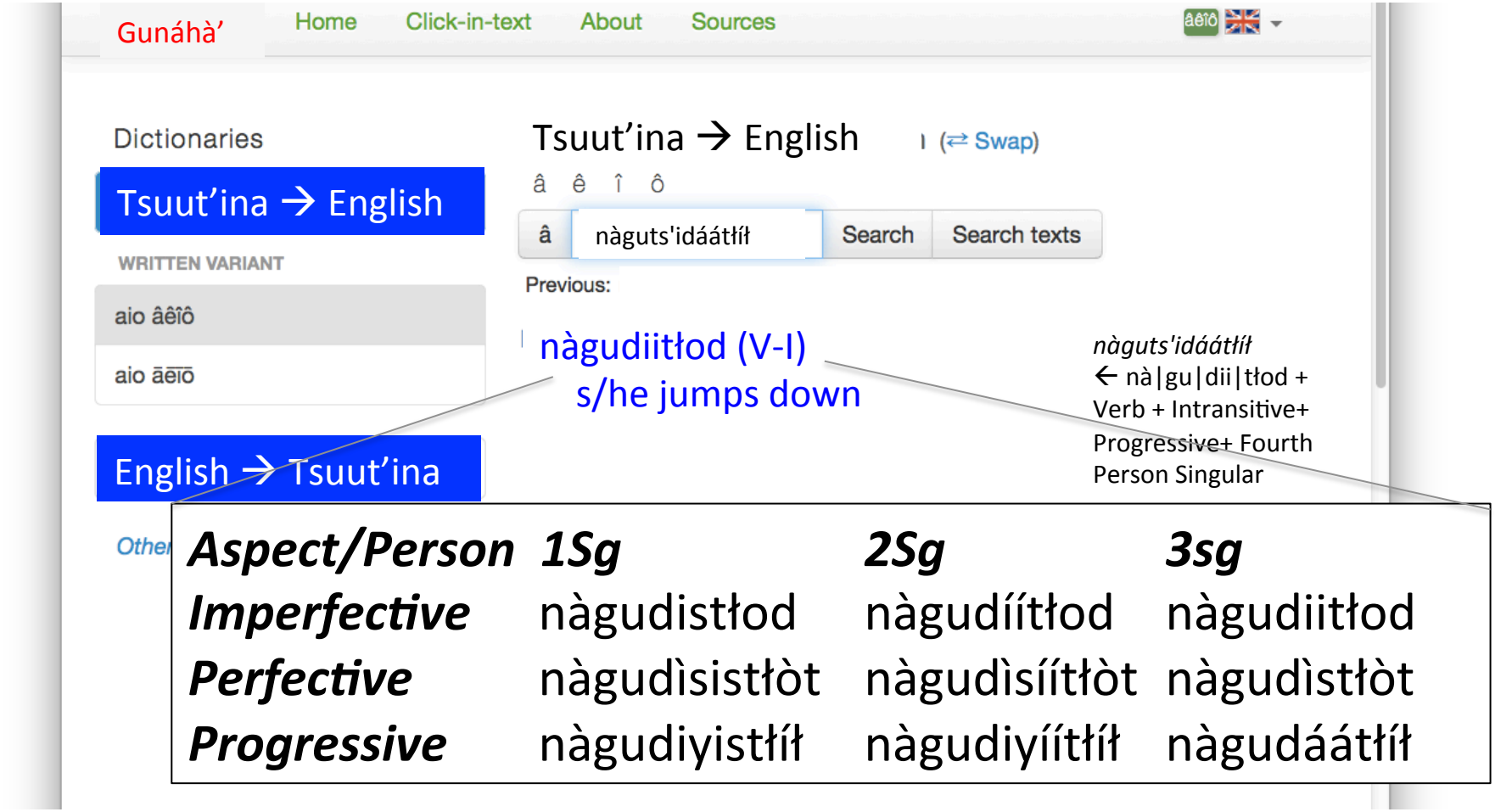
- intelligent web-based dictionary: I-DICT
- intelligent computer-aided language learning: ICALL
- spell-checking

Tsuut’ina – I-DICT
Search with inflected form



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

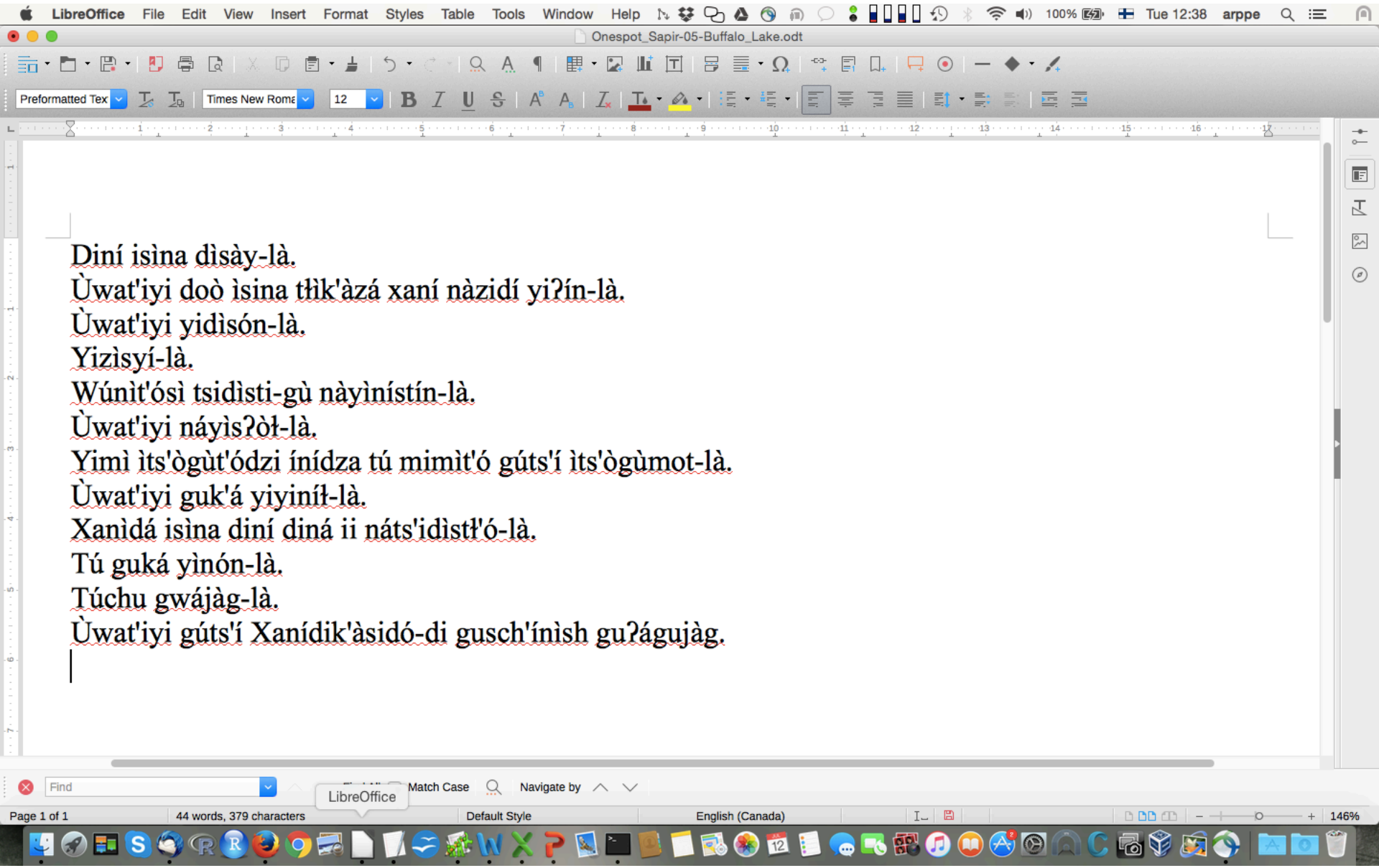
Tsuut'ina – I-DICT – paradigm generation



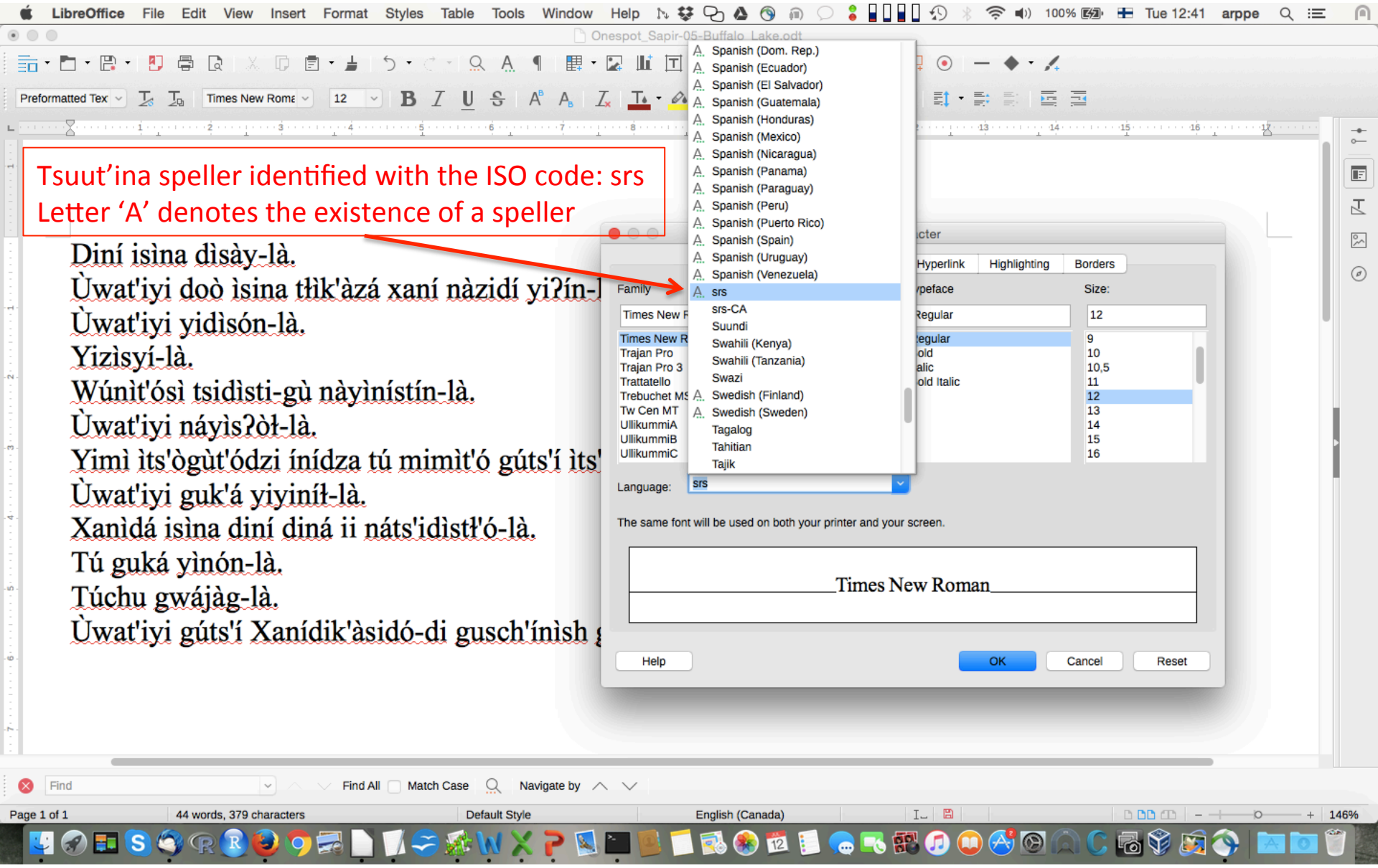
Aspect/Person	***1Sg***	***2Sg***	***3sg***
Imperfective	nàgudistłod	nàgudíítłod	nàgudiitłod
Perfective	nàgudìsistłòt	nàgudìsíítłòt	nàgudìstłòt
Progressive	nàgudiyistłíł	nàgudiyíítłíł	nàgudáátłíł

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

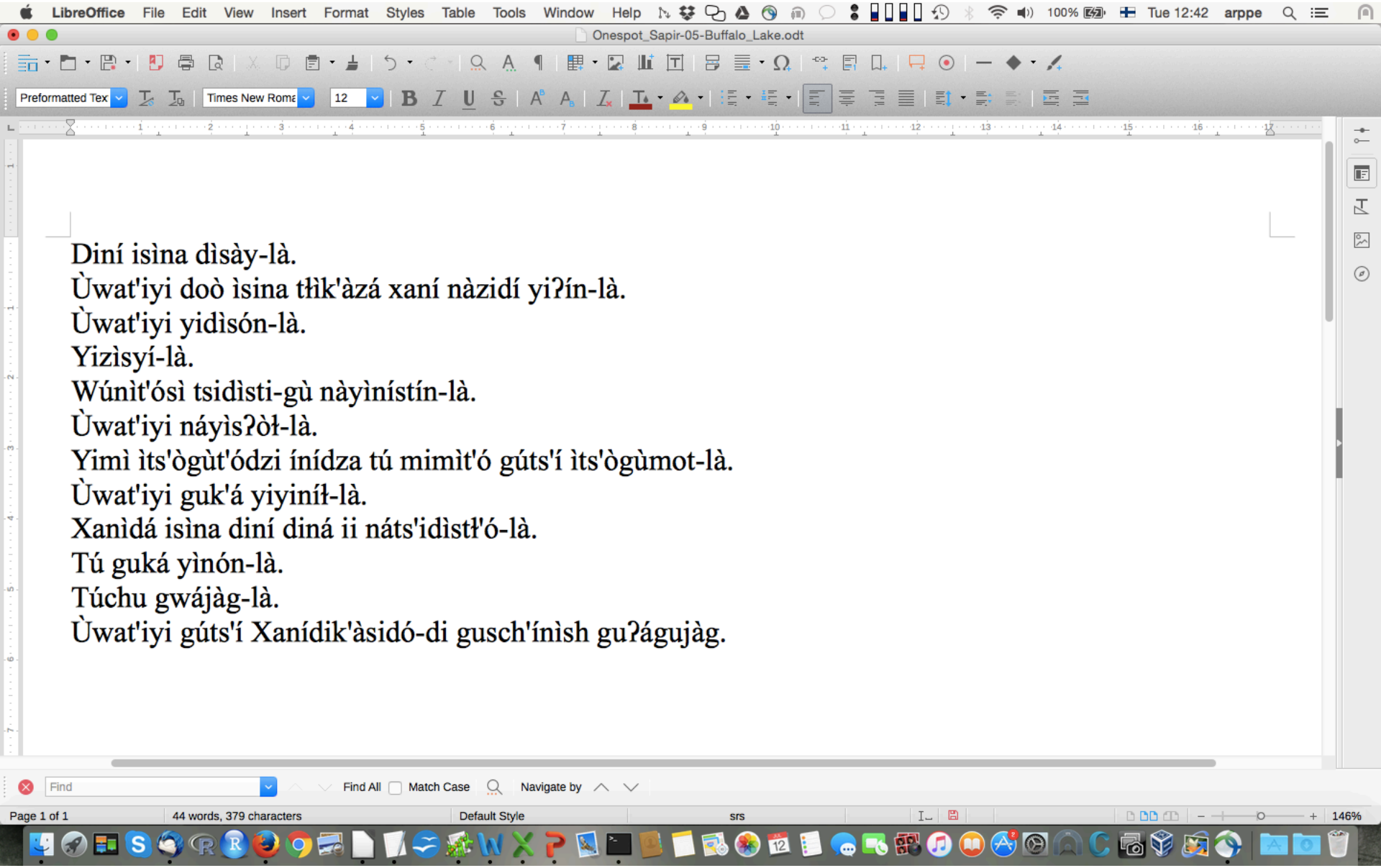
Useful applications – spell-checking



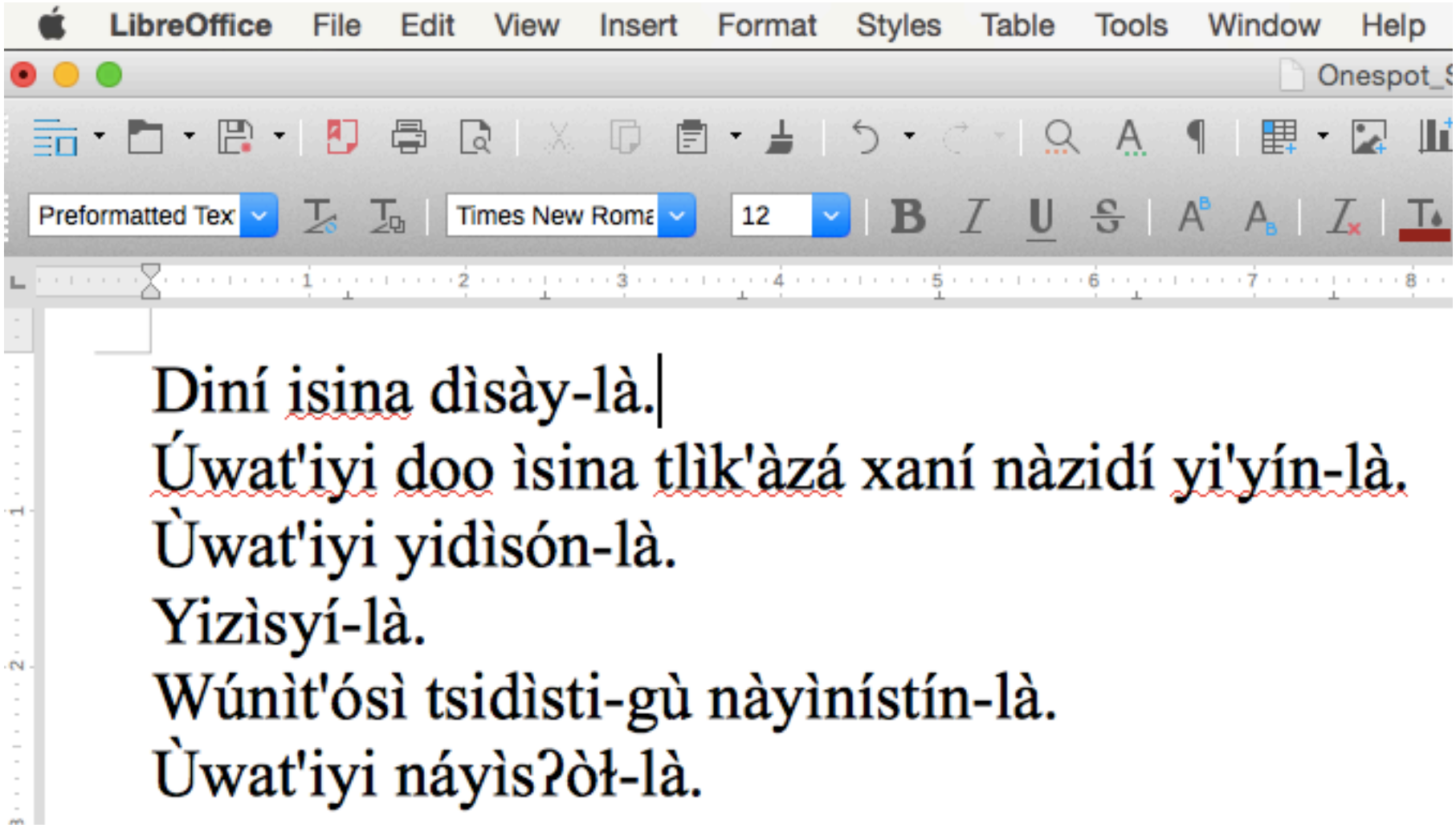
Tsuut’ina – spell-checking



Tsuut’ina – spell-checking

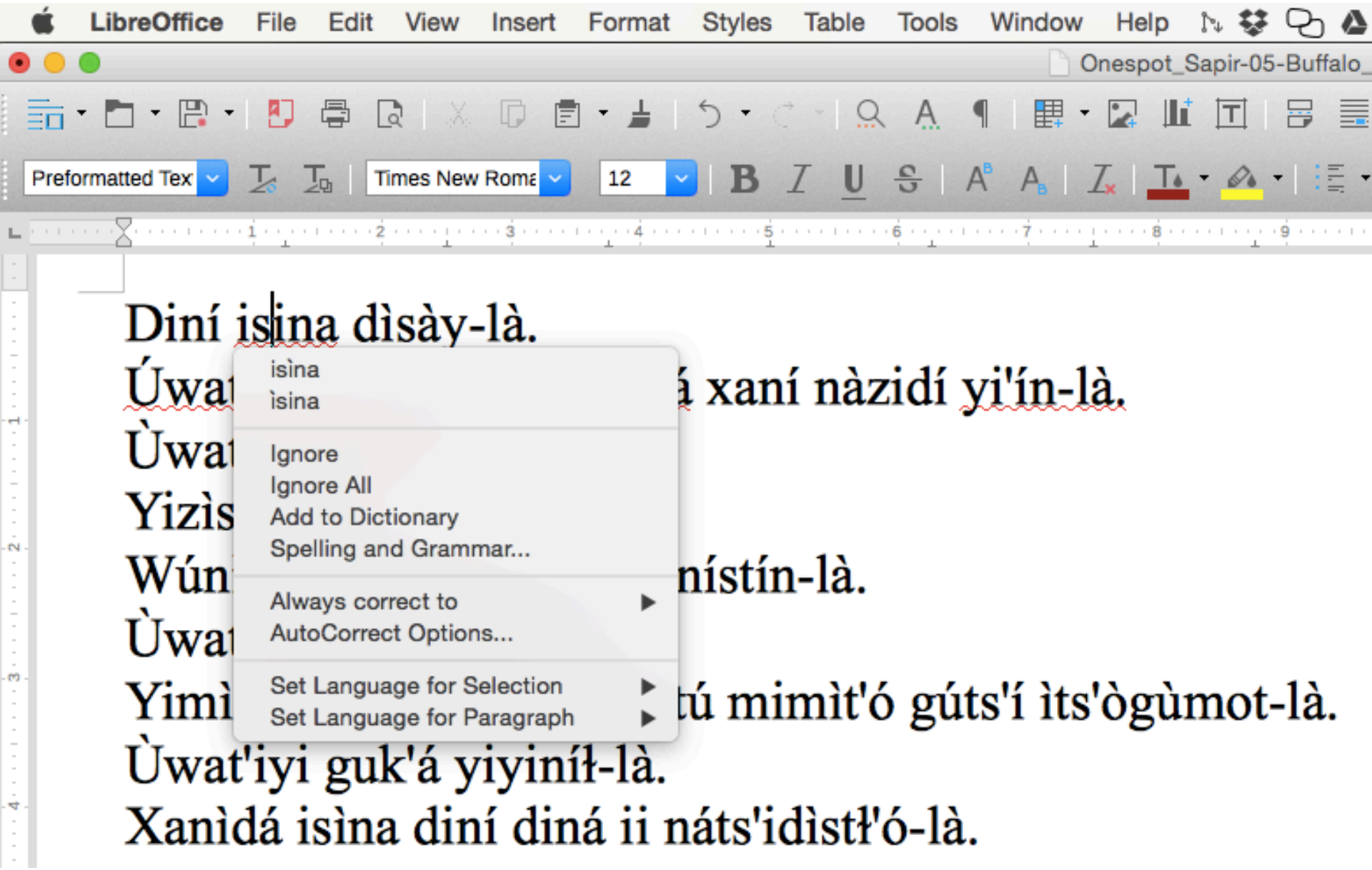


Tsuut’ina – spell-checking – typo correction

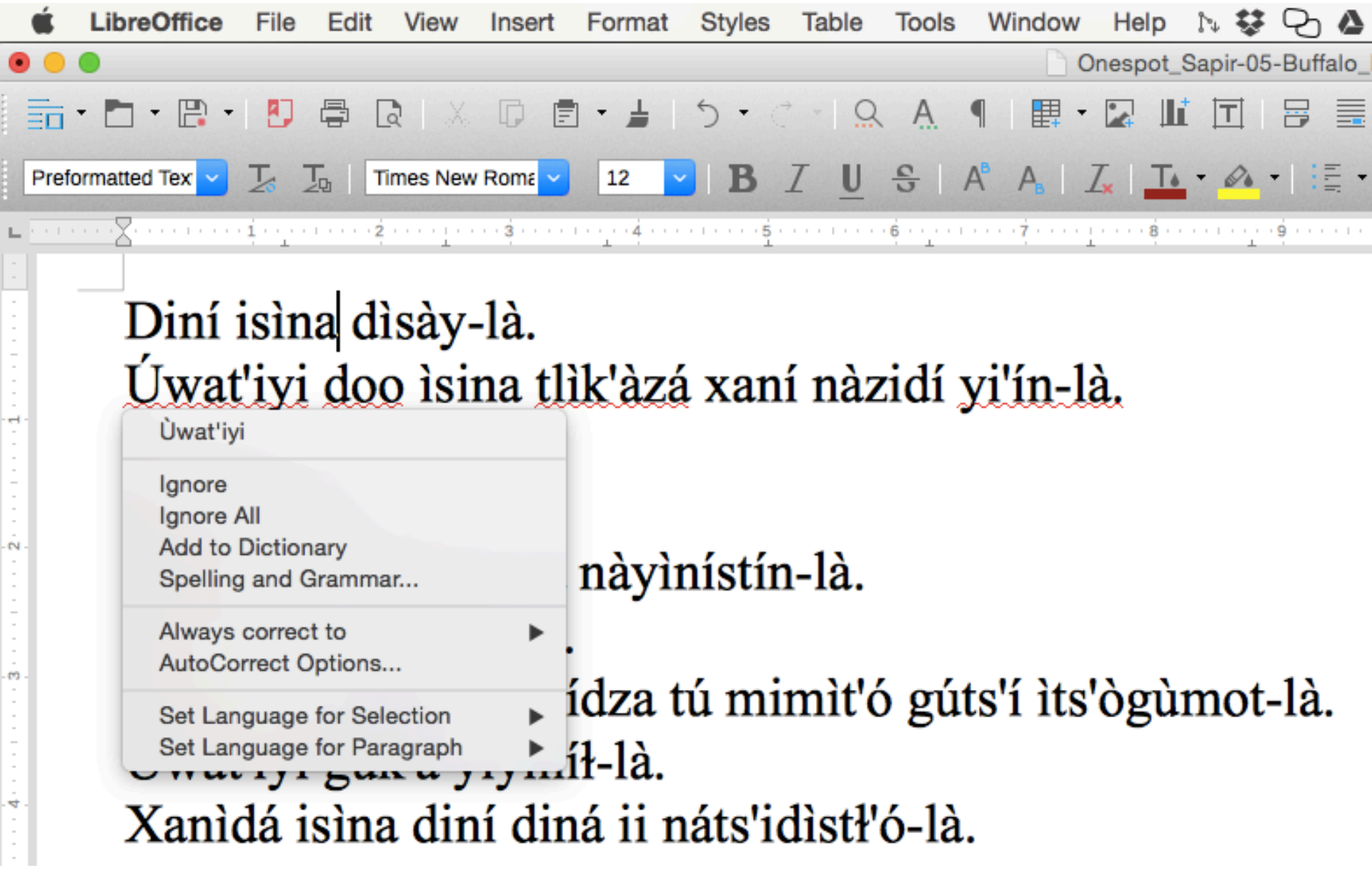


Typos: (1) missing tone marking: *isina*, *doo*; (2) wrong tone marking (Ú<Ù): *Úwat’iyi*;
(3) plain letter instead of diacritic one (l<ł) : *tlìk'àzá*; (3) apostrophe for glottal stop (’<ʔ): *yi'yín-là*

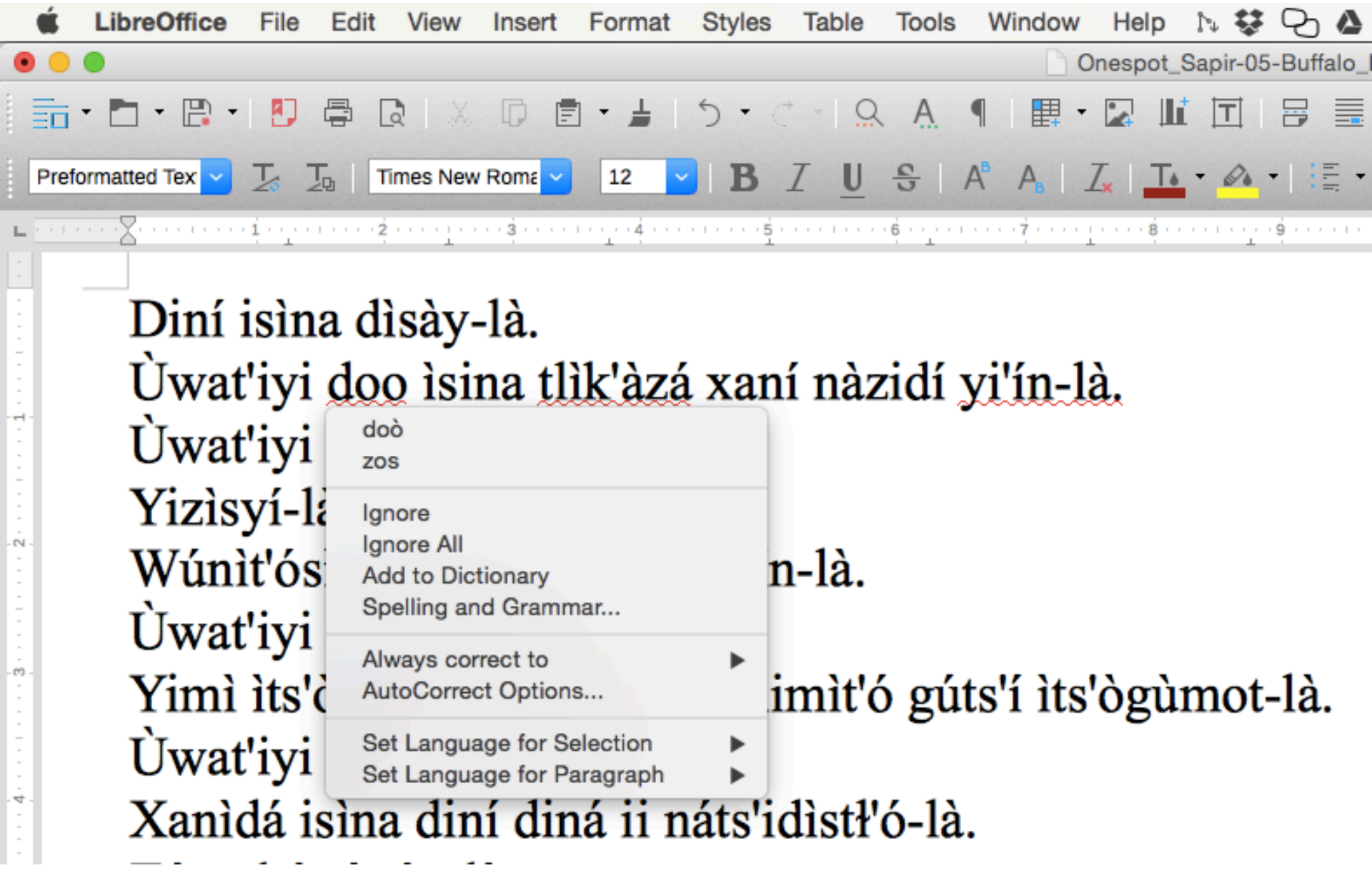
Tsuut’ina – spell-checking



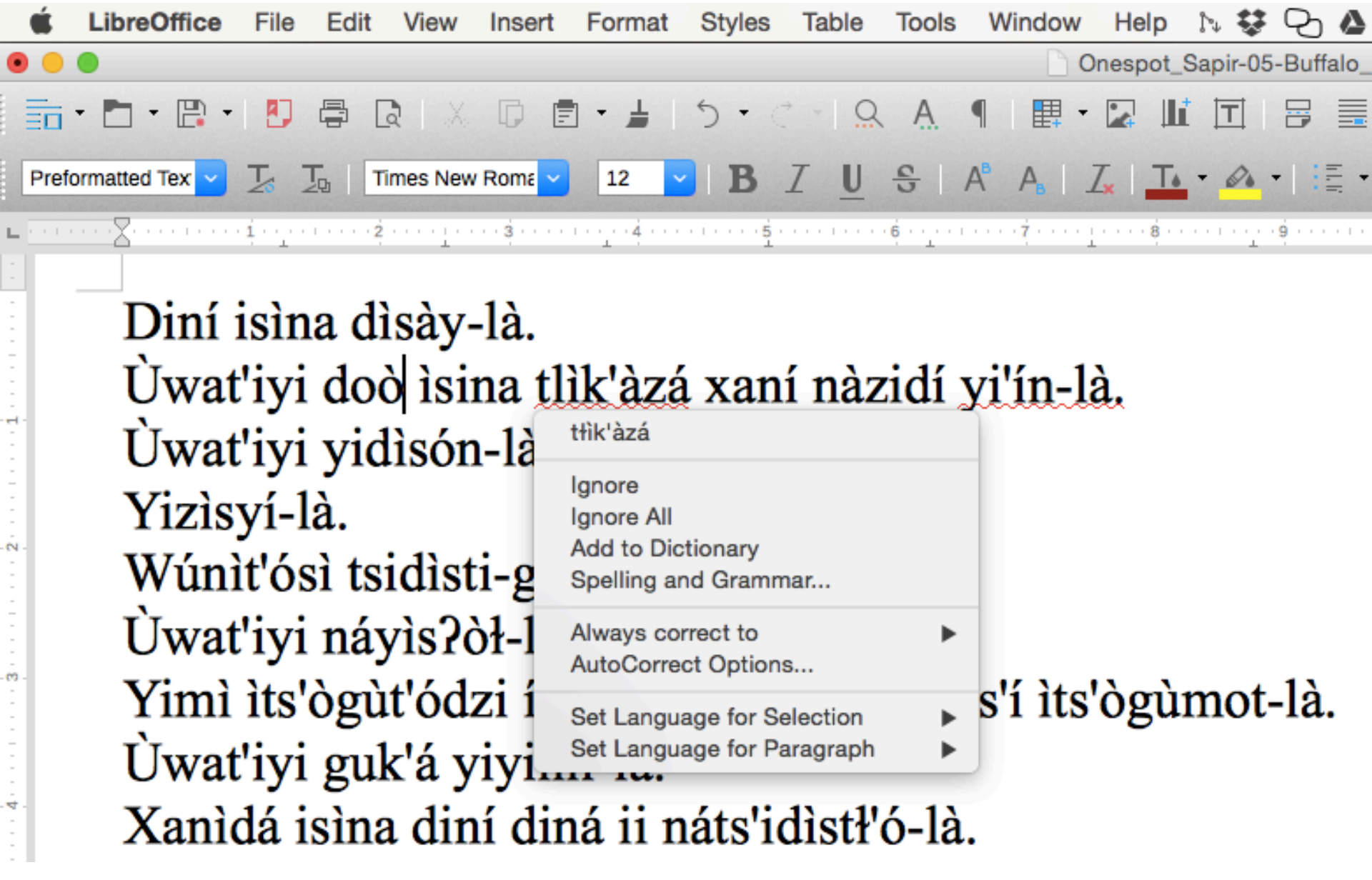
Tsuut’ina – spell-checking



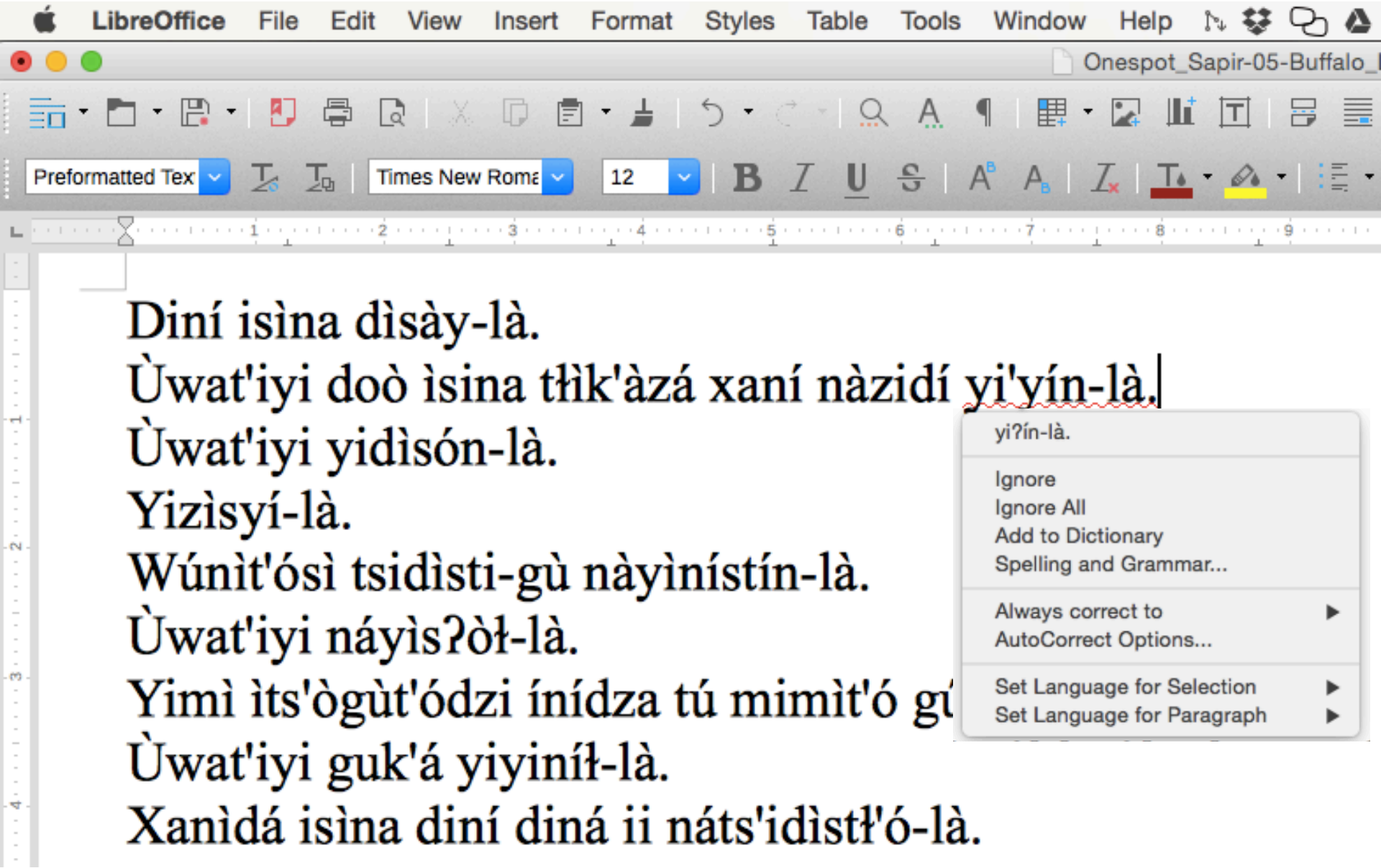
Tsuut’ina – spell-checking



Tsuut’ina – spell-checking



Tsuut’ina – spell-checking



→ 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.