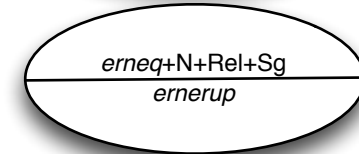


Greenlandic

The Greenlandic transducers consists of serial finite-state rewrite rules for morphonology, and finite-state transducers for morphology and lexicon.



```
define cliticAassim
[ i -> a | | a _ %CLIT ] ;
asavai+LI -> asavaali
```



```
define fricativeassim
[ q -> r | | Vow _
[ %> Segment | %CLIT ] ] ;
```

Practical applications

The transducers form the backbone of several practical applications.



Kukkunიაat, the Greenlandic spellchecker

Giellatekno and Oqaasileriffik have made a spellchecker for Greenlandic, based upon a morphological transducer, integrated in MS Word by Lingsoft.



Converting Kleinschmidt to modern spelling

okausísalo ilait pâsisimalâlerêrsimagavkit okautsîka
paitsôrpatdlârkunagit sapíngisavnik tasamanimiut pâsisi-
naussait atorsínaussáka sule amigaraluakissut ilâtigut
okauseralugit ima okalugpunga

oqaasiisalu ilaat paasisimalaalereersimagakkit oqaatsikka
paatsoorpallaaqqunagit sapinngisannik tasamanimiut
paasisinaasaat atorsinnaasakka suli amigaraluqisut ilaatigut
oqauseralugit ima oqalupunga

Syntactic analysis

Disambiguation, syntactic functions and dependency relations are done using the CG3 formalism and the visleg3 constraint grammar compiler, <http://visl.sdu.dk>.

FST gives all possible morphological analyses.

Disambiguator picks the relevant analyses and adds grammatical functions.

Dependency gives the tree structure.

Kangeq Nuup kitaaniippoq.
(Kangeq is west of Nuuk.)

```
"<kangeq"
"Kangeq" N Abs Sg
"<nuup"
"Nuuk" N Prop Rel Sg
"nuuk" N Rel Sg
"<kitaaniippoq"
"kiti" N* Lok Sg 3SgPoss IP V Ind 3Sg
"kiti" N* Lok Pl 3SgPoss IP V Ind 3Sg
"<.>"
"." CLB
"<kangeq"
"Kangeq" N Abs Sg @SUBJ>
"<nuup"
"Nuuk" N Prop Rel Sg @POSS
"<kitaaniippoq"
"kiti" N* Lok Sg 3SgPoss IP V Ind 3Sg @PRED
"<.>"
"." CLB
"<kangeq"
"Kangeq" N Abs Sg @SUBJ> #1->3
"<nuup"
"Nuuk" N Prop Rel Sg @POSS #2->3
"<kitaaniippoq"
"kiti" N* Lok Sg 3SgPoss IP V Ind 3Sg @PRED #3->0
"<.>"
"." CLB #4->4
```

Flag diacritics in the morphological analysis of Iñupiaq

Problem:

- Verbs are intransitive or transitive.
- Most derivational affixes are common to both stem types.
- Intransitive and transitive verbs have different inflectional affixes.
- The transducer should remember stem type, in order to avoid duplicating the derivational component.

Solution:

- Stems are first flagged according to transitivity and stem type, IV and TV.
- They get derivation and mood markers according to stem type.
- They then get personal endings and clitics, where endings are marked with transitivity flags.
- Finally they pass through a filter that stops flag mismatches.

```
LEXICON IV
@P.IV.ON@ junctionIV ;

LEXICON TV
@P.TV.ON@ junctionTV ;

LEXICON junctionIV
v-der ;
verb_intr ;

LEXICON junctionTV
v-der ;
verb_tr ;

LEXICON v-der ! hundreds...
+NIAQ:%>niaq v-der-cont ;

LEXICON v-der-cont
@R.IV.ON@ verb_intr ;
@R.TV.ON@ verb_tr ;

LEXICON verb_intr
+V+Ind+Prs:%>tu indpers ;
+V+Ind+Prt:%>tua indpers ;

LEXICON indpersonaffix
+1Sg:%>na clit ;

LEXICON verb_tr
+V+Ind+Prs:%>ge indpers_tv_12 ;

LEXICON indpers_tv_12
+1Sg+2Sg0:%>kpiñ clit ;

LEXICON clit
+GUUQ:%>guuq f ;
+LU:%>lu f ;

LEXICON f
< "@. [IT]V.ON@ " > # ;
! Deletes iv/tv combinations
```

The current state of the transducers

The Greenlandic transducer contains 85.000 lexical entries and 160.000 states. Tested on a corpus of 270.000 words of running texts, it recognizes 81% of the wordforms.

The Iñupiaq transducer is still in an experimental state, it contains 1990 lexical entries and 2911 states.

Converting Greenlandic spelling to IPA

Syllabic weight (1, 2, 3) follows Kleinschmidts law. High tone is marked with *.

Kaffimik aallerpoq "s/he fetched coffee" =>
¹kaf.fi.mik a*:¹t.¹tɛp.¹pɔq

Sunaaffa anereersut "it turned out they had left" =>
su.³na:f.fa a.nɛ.³ɛɛ:s.¹sut

Iñupiaq and Greenlandic analysers online

Both the Iñupiaq and Greenlandic analysers are available online, both for analysis and source code download.

iglu: iglu+N+Abs+Sg

iglu N+Abs+Sg iglu
iglu N+Abs+Sg+1Sg igluga
iglu N+Abs+Sg+2Sg iglun
iglu N+Abs+Sg+3Sg iglua
iglu N+Abs+Pl iglut
iglu N+Abs+Pl+1Sg iglutka
iglu N+Abs+Pl+2Sg iglutiñ
iglu N+Abs+Pl+3Sg iglunji
iglu N+Abs+Du igluk
iglu N+Abs+Du+1Sg iglukka
iglu N+Abs+Du+2Sg iglukiñ
iglu N+Rel+Sg iglum
iglu N+Rel+Pl+3Sg iglunjisa
iglu N+Trm+Sg iglumun
iglu N+Trm+Pl iglunun
iglu N+Loc+Sg iglumi
iglu N+Loc+Pl igluni
iglu N+Loc+Du iglunni iglugni

Iñupiaq analyzer

Copy text into the window below, and the program will give you all possible morphological analyses of the wordforms in question. If you choose Disambiguate, it will make an attempt at choosing the analysis which suits the given context (warning: the development of the disambiguation program hasn't really started yet).

You may also convert Iñupiaq text from the 8-bit Iñupiaq Dictionary encoding to Unicode (if your Iñupiaq text contains some of the letters e, f, x, b, you need to do this). Choose "Convert".

Type the word forms:

agnigisuuq

Give all word forms

Convert

Send form Reset form Character coding: utf-8 latin 1

agnigisuuq agnigiq+V+Ind+Prs+3Sg

Copyright © Sámi language technology project
<http://giellatekno.uit.no/>

Future possibilities

Machine translation from Greenlandic to Inuktitut and Iñupiaq, intelligent dictionaries, synthetic speech...