

Giellatekno – Centre for Sámi Language Technology

Giellatekno is a research centre specialised in formal linguistics for Sámi and to a certain extent also the other morphologically complex circumpolar languages.

Giellatekno focuses on well-defined, grammar-based analysis methods that are robust enough to form the core of linguistic tools for online and offline applications.

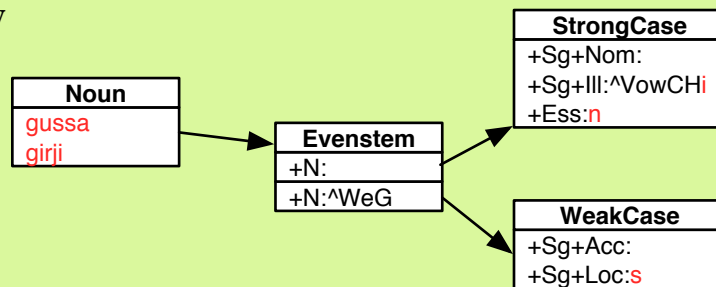
Lexicon, morphology and morphonology

Two-level morphology transducers are used for morphonology, and finite-state transducers for

morphology and lexicon. The North Sámi transducer comprises 102.000 lexical entries and 367.000 states.

Lexicon and concatenative morphology

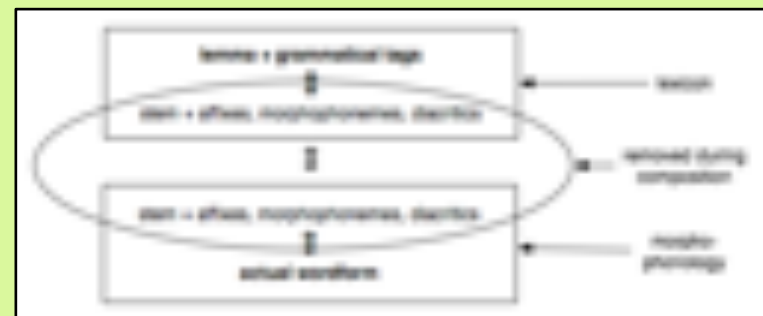
Lexicon and concatenative morphology are organised as tuples <upper level:lower level>. The upper level represents lemma and grammar features, while the lower level represents stem, affixes and triggers for morphological processes.



Morphological rules

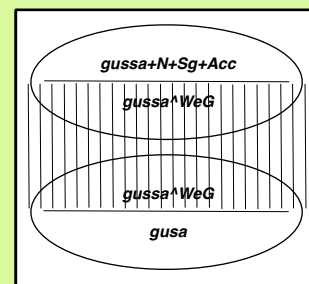
Consonant gradation and vowel change are handled by a set of context-sensitive morphological rules, triggered either by phonological context alone, or by morphologically induced triggers, here WeG for weak grade and VowCh for vowel change:

```
ss -> s   || _ Vow* WeG ;
rj -> rjj  || _ Vow* WeG ;
i  -> á    || _ VowCh  ;
```



Compilation

Compilation removes intermediate representations between morphological and morphonological transducers. The result is a morphological transducer for both word form analysis and generation.



gusa	gusa	gussa+N+Sg+Acc
gusa	gusa	gussa+N+Sg+Gen
girjji	girjji	girji+N+Sg+Acc
girjji	girjji	girji+N+Sg+Gen
girjái	girjái	girji+N+Sg+Ill
girjái	girjái	girjái+A+Sg+Ill
girjái	girjái	girjái+A+Sg+Nom

Constraint Grammar

Morphological disambiguation as well as assigning syntactic and dependency tags is

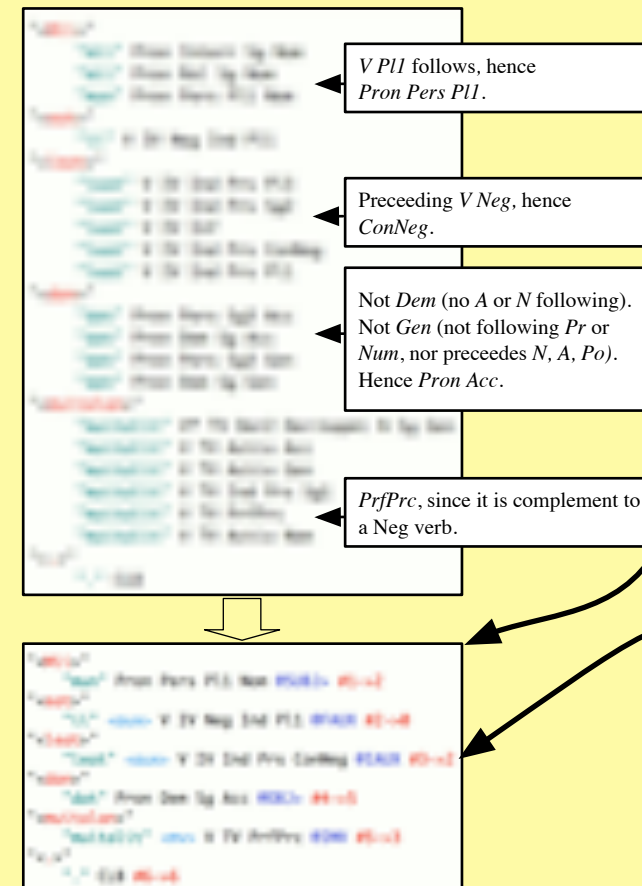
achieved with Constraint Grammar vislcg3, <http://visl.sdu.dk/> The North Sámi CG grammar consist of appr. 3500 rules.

Disambiguation

The correct morphological analysis for each word form is chosen according to context.

Example:

Mii eat leat dan muitalan. 'We haven't told it.'



Syntactic functions @SYNTAG>

Rules:

Nom gets @SUBJ> if there is no finite verb to the left, and a finite verb to the right, and no barrier intervening.

The Neg verb gets @FAUX.

ConNeg verb gets @IAUX if it is a possible auxiliary and has a Neg verb to the left and a participle to the right.

Acc gets @OBJ> if there is no transitive verb to the left.

PrfPrc gets @IMV if there is a copula or orrut to the left, without other participles or Actio Essive intervening, and it is not an auxiliary with a following Inf verb.

Dependency tree #1->2

The root node points to 0, each other node points to its mother node.

Rules:

The mother of @SUBJ> is the first finite verb to the left #1->2.

A finite verb with no mother is the root #2->0.

The mother of a nonfinite @IAUX is the first finite @FAUX to the left #3->2.

The mother of @OBJ> is the first transitive main verb to the right #4->5.

The mother of an infinite verb is the first @FAUX or @IAUX to the left #5->3.