Sámi Language Technology at the University of Tromsø

Lene Antonsen, Ciprian Gerstenberger, Trond Trosterud, Linda Wiechetek

http://giellatekno.uit.no/

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

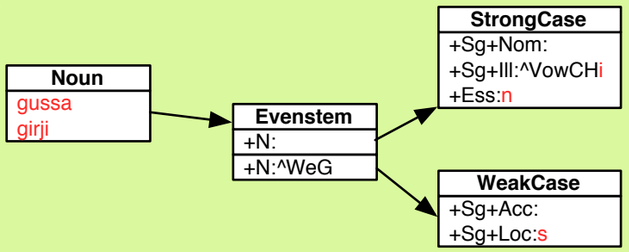
Two-level morphology transducers are used for morphonology, and finite-state transducers for morphology and lexicon. The North Sámi transducer comprises 102.000 lexical entries and 367.000 states.

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.

Lexicon and concatenative morphology

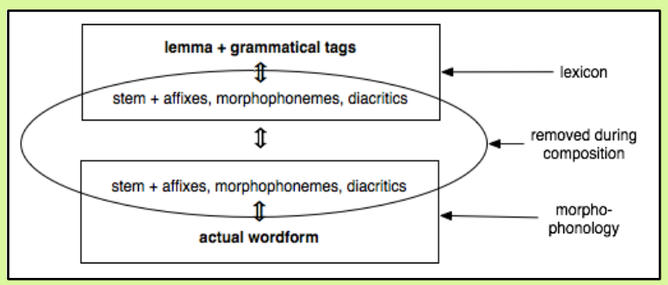
Lexicon and concatenative morphology are organised as tuples <upper level:lower level>. The upper level represents lemma and grammar features, while the lower level represents stem, affixes and triggers for morphonological processes.



Morphonological rules

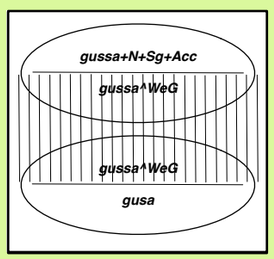
Consonant gradation and vowel change are handled by a set of context-sensitive morphonological rules, triggered either by phonological context alone, or by morphologically induced triggers, here WeG for weak grade and VowCH for vowel change:

```
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     gussa+N+Sg+Gen

girjji
girjji   girji+N+Sg+Acc
girjji   girji+N+Sg+Gen

girjái
girjái   girji+N+Sg+Ill
girjái   girjái+A+Sg+Ill
girjái   girjái+A+Sg+Nom
```

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.'

```
"<Mii>"
    "mii" Pron Interr Sg Nom
    "mii" Pron Rel Sg Nom
    "mun" Pron Pers Pl1 Nom
"<eat>"
    "ii" V IV Neg Ind Pl1
"<leat>"
    "leat" V IV Ind Prs Pl3
    "leat" V IV Ind Prs Sg2
    "leat" V IV Inf
    "leat" V IV Ind Prs ConNeg
    "leat" V IV Ind Prs Pl1
"<dan>"
    "dat" Pron Pers Sg3 Acc
    "dat" Pron Dem Sg Acc
    "dat" Pron Pers Sg3 Gen
    "dat" Pron Dem Sg Gen
"<muitalan>"
    "muitalit" V* TV Der2 Der/eapmi N Sg Gen
    "muitalit" V TV Actio Acc
    "muitalit" V TV Actio Gen
    "muitalit" V TV Ind Prs Sg1
    "muitalit" V TV PrfPrc
    "muitalit" V TV Actio Nom
"<.>"
    "." CLB
```

V Pl1 follows, hence *Pron Pers Pl1*.

Preceeding *V Neg*, hence *ConNeg*.

Not *Dem* (no *A* or *N* following). Not *Gen* (not following *Pr* or *Num*, nor preceedes *N, A, Po*). Hence *Pron Acc*.

PrfPrc, since it is complement to a Neg verb.

```
"<Mii>"
    "mun" Pron Pers Pl1 Nom @SUBJ> #1->2
"<eat>"
    "ii" <aux> V IV Neg Ind Pl1 @FAUX #2->0
"<leat>"
    "leat" <aux> V IV Ind Prs ConNeg @IAUX #3->2
"<dan>"
    "dat" Pron Dem Sg Acc @OBJ> #4->5
"<muitalan>"
    "muitalit" <mv> V TV PrfPrc @IMV #5->3
"<.>"
    "." CLB #6->6
```

Syntactic functions @SYNTAG>

Rules:

Nom gets **@SUBJ>** if there is no finite verb to the left, but 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**.