Morphology-based Sámi Proofing Tools

✓ÁBČ

The Divvun project is a project under the Norwegian Sámi Parliament to make proofing tools for Sámi, which means spell checkers and automatic hyphenators for North, Lule and South Sámi. The Divvun tools work on Linux, MacOS X and Windows in the most common office applications.

The Divvun project is cooperating closely with the Sámi language technology group Giellatekno at the University of Tromsø. We use the same source files, like two-level rule files for morpho-phonology and lexicon files, and the same infrastructure for development and testing.

Compounding

Sámi has dynamic noun + noun compounding. We utilize both two-level rules and combinatorial restriction rules to get a correct compounding grammar. Within this compounding grammar, compounds are dynamically accepted and suggested by the spellers.

Here are some issues relating to compounding:

The case of the first part: *loddemoahti* (nom., 'flock of birds') vs. *lottibeassi* (gen., 'birds nest'). The case can be one or more of NomSg, GenSg, or GenPl.

Vowel reduction or not: *lottečivga* ('young bird') vs. *lottibeassi* ('birds nest'). Vowel reduction is either prohibited, obligatory, or facultative.

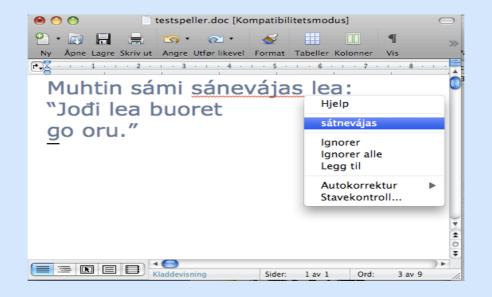| Part of generated paradigm for *loddi* ('bird')	
loddi+N+Sg+Nom	loddi
loddi+N+Sg+Gen	lotti
loddi+N+Sg+Acc	lotti
loddi+N+Sg+Ill	loddái
loddi+N+Sg+Loc	lottis
loddi+N+Sg+Com	lottiin
loddi+N+Ess	loddin

The case of the first part of a compound

The compounding case is mostly lexically specified, although the default is not:

```
loddi ('bird') ;   +SgNomCmp +SgGenCmp
sátni ('word') ;   (unspecified, default: +SgNomCmp)
```

This will give us:

```
loddemoahti ('flock of birds')  nom. first part OK
lottečivga  ('young bird')      gen. first part OK
sátnevájas  ('proverb/saying')  nom. first part OK
*sánevájas  ('proverb/saying')  gen. first part NOT OK
```


Stem vowel reduction

A. Obligatory reduction:
```
loddi AIGI ;
loddemoahti vs *loddimoahti (nom. 'flock of birds')
```

B. No reduction:
```
johtu ALBMILONG ; ('circulation')
johtučálus vs *johtočálus (nom. 'circular')
```

The two words are inflected using two different lexica: **AIGI** and **ALBMILONG**. The inflection lexicon **AIGI** has triggers that via two-level rules regulate the vowel reduction.