A South Sámi Grammar Checker for Stopping Language Change

Linda Wiechetek
UiT Norgga árktalaš universitehta
`linda.wiechetek@uit.no`

Maja Lisa Kappfjell
UiT Norgga árktalaš universitehta
`maja.l.kappfjell@uit.no`

Abstract

We have released and evaluated the first South Sámi grammar checker *Gram-Divvun*. It corrects two frequent error types that are caused by and causing language change and a loss of the language's morphological richness. These general error types comprise a number of errors regarding the adjective paradigm (confusion of attributive and predicative forms) and the negation paradigm. In addition, our work includes a classification of common error types regarding the adjective and negation paradigms and lead to extensive grammatical error mark-up of our gold corpus. We achieve precisions above 71% for both adjective and negation error correction.

1 Introduction

Language change is a natural process caused by various factors in all languages. Indigenous languages are in a special situation, as they typically need to compete with a majority language, which is used by the bilingual language user more often and in more domains. South Sámi is in a critical situation that requires concrete measures so that morphological richness is taught to the next generation and does not get lost. While we do not think we can stop language change altogether, we do think that we can provide necessary grammatical support for South Sámi writers when other help is not available. A language community that wants to preserve certain language structures, will only be able to so if someone can give feedback to language learners, both L1 and L2.

The school system does not provide sufficient language support for the South Sámi language. Students have only a few hours a week to learn the language. The teachers of South Sámi have to select what they teach, which are typically the topics that are satisfactorily described in the grammar books, such as the verbal and nominal paradigms. Other topics, such as the adjective and negation paradigms, on the other hand, are described very superficially and lack information about the variation in the spoken language. A grammar checker that corrects grammatical errors of the latter types can deliver feedback and thereby improve grammatical knowledge in these areas.

In this article, we focus on two very frequent grammatical error types of morphological forms that the language community wishes to preserve, which has been expressed in professional meetings of teachers and translators. Some of these tendencies have been decided on by the Sámi normative institution.[1] Those include adjective inflection and inflection of verbal periphrastic negation. An investigation in 2018 (Kappfjell and Trosterud, forthcoming) showed tendencies of adjective classes being reduced from four to two classes. Blokland and Inaba (2015) cover negation in South Sámi and shows that at least four non-traditional paradigms of past tense copula negation are used in contemporary text. There are strong tendencies in the language community itself to preserve the traditional paradigm as it is presented in the written grammar. (Bergsland, 1994)

The first South Sámi gammar checker, *Gram-Divvun*, has been released 31st May 2023 and is freely available for MS Word and GoogleDocs.[2] We encourage the use of our proofing tools in schools and other educational institutions, publishing houses and the Sámi government.

---

[1] `https://sametinget.no/sprak/sami-giellagaldu/?sprak=14`

[2] `https://divvun.no/en/korrektur/gramcheck.html`

2 Background

2.1 Language situation

According to Blokland and Hasselblatt (2003, p.110), there are about 2,000 ethnic South Sámi, of which approximately 300-500 are South Sámi speakers. There are two major varieties in South Sámi: northern (or Asele) South Sámi and South (or Jamtland) South Sámi (Sammallahti, 1998, p.24), but the differences between the two are minor, and limited mostly to phonetics and morphology. South Sámi has a written standard, which is adhered to especially (children's) published fiction. South Sámi is an official language in altogether four communities in Norway: Aarborte (Norwegian: Hattfjelldal), Snåase, (Norwegian: Snåsa), Raarvihke (Norwegian: Røyrvik) and Plaassja (Norwegian: Røros) in Norway. In Sweden there are 10 South Sámi communities; Bïerje (Swedish: Berg), Kraapohke (Swedish: Dorotea), Herjedaelie (Swedish: Swedish: Härjedalen), Krokome (Swedish: Krokom), Luspie (Swedish: Storuman), Straejmie (Swedish: Strömsund), Upmeje (Swedish: Umeå), Vualtjere (Vilhelmina), Älvdaelie (Swedish: Älvdalen) and Ååre (Swedish: Åre). There are some minor differences between the orthographies used in Sweden and Norway, e.g. the letter *ä* is used in Sweden where the letter *æ* is used in Norway.

There is a lack of standardization and clarification regarding grammatical variants that are due to language change and simplification. Adjectives and negation paradigms, which we will deal with in this article, are exemplary cases of these changes.

2.2 Technical background

The technological implementation of the grammar checker is based on rule-based natural language processing: finite-state automata (*FST*) for morphological analysis (Beesley and Karttunen, 2003; Lindén et al., 2013; Pirinen and Lindén, 2014) and constraint grammar (Karlsson, 1990; Didriksen, 2010) for syntactic and semantic as well as other sentence-level processing. In our work, we use the free open source implementation VISLCG-3 (Bick and Didriksen, 2015). The South Sámi tools are publicly available[3]. It is part of a (multilingual infrastructure (Moshagen et al., 2013) *https://github.com/giellaltGiellaLT*), which

---

[3]`https://github.com/giellalt/lang-sma/`

includes 130 languages.

GramDivvun first analyzes the morphological structure of a text together with part-of-speech tagging, and displays all homonymy of a given form. In addition lexical semantic tags are added to (especially) nouns. A number Constraint Grammar modules are then used for ambiguous tokenisation of compounds, ordinals and abbreviations, morpho-syntactic disambiguation of word form homonymy, valency additions and lastly error detection and correction. Error detection and correction is accomplished by means of a set of hand-written rules that first identify an erroneous form in a given morpho-syntactic context, labels it, and then exchanges an incorrect tag combination with another one, which then is used to generate the correct form. The full modular structure is described in Wiechetek (2019). As Figure 1 shows, *GramDivvun* is realized in the right-hand column and gives feedback and suggestions for South Sámi errors as in this case the negation form *Ean*.

Our work started out with collecting error sentences according to error type. Those sentences were used to develop rules for *GramDivvun* and typically contain between 10 and 70 examples that are relevant to a certain error type. These regression tests are used for developing and quality ensuring our tool, cf. Wiechetek et al. (2021). Regression testing shows that error correction for both negation and adjective forms look promising with precisons of up to 80% when starting our work.

2.3 Motivation

A recent survey shows that language technology is used to a far greater extent by minority languages and indigenous languages than by state-bearing majority languages such as Norwegian. (Trosterud, 2019) The size of the language community also plays a role: South Sámi use language technology aids to a far greater extent than North Sámi. Language technology tools are therefore central to the revitalization of South Sámi, and our goal is to be able to provide good tools to the South Sámi language community. One of the authors is a member of the South Sámi language community that serves as a reference for linguistic questions regarding grammar and the lexicon. Competent speakers with clear language intuitions are essential for a language community. South Sámi school
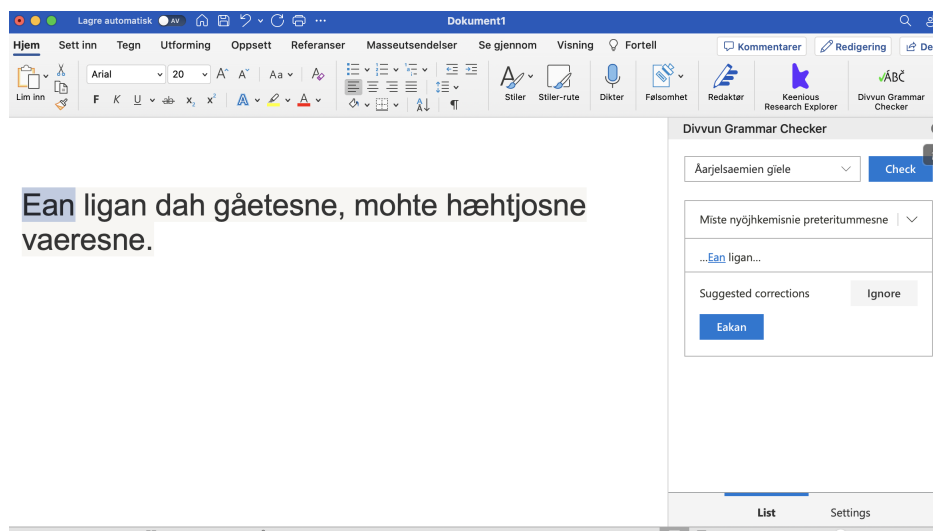
Figure 1: *GramDivvun* integrated in MS Word

children of the 80s who were taught by Anna Jacobsen,(Jacobsen, 2013) (teacher in Hattfjelldal) and Ella Holm Bull (teacher in Snåsa) had a strong grammarian with clear expectations of how correct language should be as guidance.(Kappfjell, 2014) When language experts from the past generation pass away, the bearers of this knowledge disappear. In a reality where South Sámi is not used as frequently in daily life as it used to be, we need other tools to ensure that feedback for correct and incorrect language is available. Otherwise, there is a lot of insecurity about it and instead of using the language, people keep quiet and do not dare to write.

3 The South Sámi grammar checker

3.1 Adjective errors

South Sámi grammars that write about the adjective system often state that the adjective paradigm is unclear. In the dictionaries and in the text corpora, there is a lot of variation.

According to earlier grammarians, two-syllable adjectives usually have two forms in the positive, one of them ending in a vowel and the other of them ending in *-s*.

These two forms can be attributive or predicative forms. Alternatively, there can be only one form for both attributive and predicative. According to earlier grammarians, the comparative forms are built on the predicative form. However, in today's South Sámi there are also comparatives built on attributive forms. Table 1 shows all four attribute-predicative combinations are those

according to these grammars (Lagercrantz, 1926; Bergsland, 1946; Hasselbrink, 1981-1985; Magga and Mattsson Magga, 2012).

Attributive	Predicative
vowel (*buerie*)	vowel (*buerie*)
vowel (*skïemtje*)	-s (*skïemtjes*)
-s (*vihkeles*)	vowel (*vihkele*)
-s (*båeries*)	-s (*båeries*)

Table 1: Adjective paradigms in positive

In addition to that, some of the adjective forms can also be adverbs. The predicative form *vihkele* 'important' for example is homonymous to the adverbial form. Other adjectives have more part-of-speech homonymies. *buerie* 'good' is for example both attributive and predicative form of an adjective, but can at the same time also be a noun. The form *båetije* 'coming' is both an adjective, deverbal noun and a present participle of a verb.

Kappfjell and Trosterud (forthcoming) show that text collections of modern South Sámi exhibit others tendencies of adjectives inflection than its mentioned on the earlier grammars. They come to the conclusion that modern South Sámi shows the same system as before, but the attribute is more frequent than a predicative: 60% vs. 30%. The other tendency is that instead of four adjective classes, there are only two of them where attributive and predicative are homonymous, either ending in a vowel or in -s. The investigation shows, that predicative and attribute forms are the same in 98.4% of the cases. Only 8.7% of the adjec-

tive types display variation. This system appears to be very stable and consistent. However, there is a desire in the language community to revert the system and go back to and teach morphological richness to new generations, as the author of this paper can confirm.

We have to keep in mind that South Sámi language orthography was approved in 1978, and there has been a careful revitalization at the Sámi schools in Snåsa and Hattfjelldal at the Norwegian side of South Sámi area. There are approximately 500 speakers, but only 1/10 actually write the language as well. South Sámi training has been deficient in that it has been cut short to only a few hours, and the teachers have thus not been given the space they have needed to be able to provide complete training in the most important grammatical systems.

(1)   Saemien kultuvre lea **gånkaladtje** jïh
        Sámi   culture is royal.PRED and
        *tjaebpies.
        beautiful.ATTR
        'The Sámi culture is royal and beautiful.'

For a rule-based grammar checker this means that we need to distinguish between adjectives that have one form for both attributive and predicative forms and those that differ in their forms. We resolve this by adding an early rule to the syntactic analyzer module preceding the grammar checking rules. The rule below adds a secondary tag <AttrPred> to each adjective with both an attributive (Attr) reading and a predicative reading in the same cohort. Since this rule precedes all disambiguation rules, both readings are still available, and the tag ensures that this information is kept throughout the analysis.

```
SUBSTITUTE (A) (A <AttrPred>)
TARGET A
IF (0 Attr LINK 0 (A Nom));
```

The error detection rules are ADD-rules. They add an error tag, here *&msyn-adj-attr-pred* to the erroneous form in a syntactic context. There are different syntactic contexts that require different types of rules. The one below pays attention to a nominative subject to its left and a possible copula between the adjective and the copula. Since copulas can be dropped in South Sámi, the subject can be an important marker. In addition it excludes a noun to its right.

```
ADD (&msyn-adj-attr-pred)
TARGET (A Attr) IF
(*-1 Nom
BARRIER (*) - REALCOPULAS - Ela)
(NEGATE 0 ATTR-PRED-A
OR A + Sg + Nom OR A-ATTR-ONLY)
(NOT 1 N) ;
```

The second context below is a visible copula that can be either by itself or together with a negation verb. If the subject is dropped, the copula is the decisive marker for predicative forms. Again we do not want a noun to the right of the adjective. This rule explicitly asks for an end of sentence after the adjective form.

```
ADD (&msyn-adj-attr-pred)
TARGET (A Attr) IF
(NEGATE 0 ATTR-PRED-A OR
A + Sg + Nom OR A-ATTR-ONLY)
(1 EOS)
(*-1 (Neg Ind) OR
REALCOPULAS BARRIER NOT-ADV-PCLE);
```

The third case is a coordination context where the predicative adjective is coordinated with another predicative adjective, which shows that the form should be predicative rather than attributive.

```
ADD (&msyn-adj-attr-pred)
TARGET (A Attr) IF

(-1 CC LINK *-1 Nom
BARRIER (*) - REALCOPULAS)
(NEGATE 0 ATTR-PRED-A
OR A + Sg + Nom OR A-ATTR-ONLY)
(NOT 1 N);
```

3.2   Negation errors

Standard negation in South Sámi utilizes a negative auxiliary and a connegative form of the lexical verb. The basic paradigm usually presented in grammars, cf. Bergsland (1946, pp.169–170), Hasselbrink (1981-1985, p.145), Magga and Mattsson Magga (2012, p.38), is one where the negative auxiliary has two moods (indicative and imperative) and two simple tenses (present and past tense) The connegative form ends in -h and is homonymous with the second person singular of the imperative. Depending on inflection type, it may also be identical to the second person singular or the third person plural of the present indicative. (Blokland and Inaba, 2015)

But according to Blokland and Inaba (2015), in different (Southern vs. Northern) dialects there are diverging inflectional patterns for negation, some of which are not according to the norm. Four of those are constructions regarding the past tense of the negation auxiliary *ij* together with the connegative verb form *leah* 'be'.

Table 2 shows the typical errors which are in South Sámi texts. In Table 2, Blokland describes the variants of Table 2 as variants which are in use in the Northern South Sámi area. We have one rules to correct the errors in Table 2, one for the negation verb *ij* making it agree with the person and number of the subject instead of treating it as an uninflected adverb (just as in Norwegian, where negation is an adverb like in English).

Error	Correct	Translation
∗**ij** lim	im lim	'I was not'
ih lih	ih lih	'you were not'
ij lij	ij lij	's/he was not'
∗**ij** limen	ean limen	'we two were not'
∗**ij** liden	idien liden	'you two were not'
∗**ij** ligan	eakan ligan	'they two were not'
∗**ij** limh	ibie limh	'we were not'
∗**ij** lidh	idie lidh	'you were not'
∗**ij** lin	eah lin	'they were not'

Table 2: Paradigm for erroneous negation constructions (type 1 - past tense)

The error type in Table 3 is corrected for the connegative past tense form *lih* - only past tense connegative forms of the copula (not any other main verb) - should agree with the negation verb. However, as a common error, the copula connegative form *lih*, which is second person singular and ends in *-h* is used for the whole paradigm. The rule *msyn-NegPrt-lih-congruence* corrects this error type.

Error	Correct	Translation
im *lih	im lim	'I was not'
ih lih	ih lih	'you were not'
ij *lih	ij lij	's/he was not'

Table 3: Paradigm for erroneous negation constructions (type 2)

The negation in Table 4, is an older form of negation documented in Bergsland (1994), which is not included in the current norm. Since it is not

very frequent in spoken and written South Sámi, we have not developed any rules for it yet. The connegative 3rd person form *leam* is used instead of *lij*. This form is now only analyzed as the first person singular present tense. It would be interesting to investigate if the past tense use is related to the North Sámi *lean* past tense connegative of *leat* 'to be'.

Typical error 3	Correct	Translation
im lim	im lim	'I was not'
ih lih	ih lih	'you were not'
ij ***leam**	ij **lij**	's/he was not'

Table 4: Paradigm for erroneous negation constructions (type 3)

Bergsland (1994) describes the variants in Table 5 as Southern variants. Even though this negation system is not so frequently used in the South Sámi text collection, it is frequent in oral speech, as reported by one of the authors, who is herself a member of the South Sámi language community. We therefore expect this error type to become more frequent in writing in the future with increasing South Sámi publications. The negation verb in this error type follows the paradigm for main verbs (as opposed to the paradigm for copulas). That means it uses the form *idtjim/eedtjem* (which is used as a negation verb with main verbs) instead of *im* (which is used for copulas). The connegative form of the past tense copula on the other hand is not inflected (as it is done with main verbs) while it should agree in person and number with the negation verb. This error type is dealt with by two rules in *GramDivvun*, one for the negation verb and the other one for the connegative form.

(2)   Mohte *eakan   *edtjigan
         but   NEG.PRES.3DU be.PRED.3DU
         juakadidh.
         separate
         'They should not seperate.'

In habitive constructions expressing possession, there is a form of *lea* 'to be' agreeing with the possessed item in number and person, and the possessor in genitive case. Typical errors regard the agreement between the copula and the possessed item as in ex. (3), where 3.Sg *ij* 'is' should be 3.Pl because of plural *måvhkah* 'trousers'. This agreement error is corrected by a separate rule since it typically appears in habitive construction.

Error	Correct	Translation
*eedtjem lih	im lim	'I was not'
*eedtjh lih	ih lih	'you were not'
*eedtji lih	ij lij	's/he was not'
*eedtjemen/	ean limen	'we two were not'
*eedtjien lih		
*eedtjeden lih	idien	'you two were not'
	liden	
*eedtjeben lih	eakan	'they two were not'
	ligan	
*eedtjuvh lih	ibie limh	'we were not'
*eedtjede lih	idie lidh	'you were not'
*eedtjen/	eah lin	'they were not'
*eedtjies lih		

Table 5: Paradigm for erroneous negation constructions (type 4)

(3) ***ij**   leah   dov
     NEG.PRES.3SG be.CONNEG you.GEN
     naan   rööpses   måvhkah
     some/red trouser.NOM.PL
     'You do not have any red trousers.'

One error type regards the negation verb itself. In past tense it should be in congruence with the subsequent past tense connegative form. In example (4), the form *ean* (1.Du) should actually be *eakan* (3.Du) as in ex. (5) as the connegative form *ligan* is a third person dual.

(4) ***Ean**   ligan   dah gåetesne,
     NEG.1DU be.PAST.3DU this home.INE.SG
     mohte hæhtjosne   vaeresne.
     this   cabin.INE.SG mountain.INE.SG
     'They were not at home, but in the cabin in the mountain'

(5) **Eakan** ligan dah gåetesne, mohte
     hæhtjosne vaeresne.
     NEG.3DU be.PAST.3DU this home.INE.SG
     this cabin.INE.SG mountain.INE.SG
     'They were not at home, but in the cabin in the mountain'

GramDivvun detects the error by means of an *ADD*-rule that adds a label to a past tense negation form (*Prt ConNeg*) if the negation verb to the left of it agrees in number and person with it. Each *ADD*-rule pairs with one or several *COPY*-rules, which pick up on the error tag, copy the morphological tag and lemma combination that makes out a form, and exchange the unwanted tags with the desired ones. The *COPY*-rule below exchanges

second or third person singular with first person singular. The second *COPY*-rule exchanges first or third person singular with second person singular.

```
ADD (&msyn-ConNegPrt-congruence)
TARGET (Prt ConNeg) + $$SG-PERS IF
(-1 ("ij" Prs Neg) - $$SG-PERS) ;


COPY (Sg1 &SUGGEST) EXCEPT
(Sg2 &msyn-ConNegPrt-congruence)
OR (Sg3 &msyn-ConNegPrt-congruence)
TARGET (&msyn-ConNegPrt-congruence)
IF (-1 Sg1);

COPY (Sg2 &SUGGEST) EXCEPT
(Sg1 &msyn-ConNegPrt-congruence)
OR (Sg3 &msyn-ConNegPrt-congruence)
TARGET (&msyn-ConNegPrt-congruence)
IF (-1 Sg2);
```

A second typical error is the use of the third person singular form of the negation verb *ij* as a default, as in example (6). Here the first person dual form of the connegative form *limen* shows the actual person and number of the verb phrase, and the negation verb should agree with it, i.e. *ij* should be changed to *ean* (1.Du).

(6) ***Ij**   limen   månnoeh
     NEG.3SG be.CONNEG.PAST.1DU there
     desnie.

     'We were not there.'

```
ADD (&msyn-Neg-VFinitt-ConNeg)
TARGET (Ind Prs) + $$ALL-PERS
OR (Ind Prt) + $$ALL-PERS
(-1 ("ij" Prs Neg) + $$ALL-PERS)
(NEGATE 0 ConNeg) ;
```

A third type changes a finite verb form to a connegative verb form, cf. ex. (7). Here, *edtjigan* should be changed to *edtjh*, and subsequently the tense of the negation verb *eakan* should be changed to past tense as marked by the connegative, i.e. *idtjigan*.

(7) Mohte **eakan**   edtjigan
     But   NEG.PRS.3DU will.PAST.3DU
     juakadidh.
     separate
     'But they would not separate.'

4 Evaluation

The evaluation is based on a part of *SIKOR*, the South Sámi corpus,(sik) containing administra-

tive, law, religious, non-fiction, fiction, and science texts. The evaluation corpus is marked up for the following error types - spelling errors, morphosyntactic errors, syntactic errors, formatting errors, real word errors, etc. It consists of a publicly available corpus, *FREECORPUS* (34,512 words) and a part that is restricted by copyright *BOUNDCORPUS* (166,483 words). For evaluation purposes we use the marked-up parts of them, hence *FREECORPUS* and *BOUNDCORPUS*.

The results of the evaluation are shown in Table 6. The quality is measured using basic precision and recall, such that recall $R = \frac{t_p}{t_p+f_n}$ and precision $P = \frac{t_p}{t_p+f_p}$, where t_p is a count of true positives, f_p false positives, t_n true negatives and f_n false negatives.

Adjective rules include both way confusions between attributive and predicative singular/plural form (attr>pred, attr>pred.pl) and pred>attr), a confusion between attributive forms and adverb derivations (attr>adv). Negation rules include (tense and person) errors of the negation verb and errors of the connegative form. The latter can be finite forms or infinitives. Errors can also be application of the main verb paradigm (connegative ending in -h) for the copula paradigm. While the main verb connegative form does not inflect for person and tense, the copula paradigm inflects for person and tense.

Precision for adjective and negation errors are both above 71%. Recall is above 79%. We expect precision to raise to close to 90% after fine-tuning the rules and fixing the last issues of corpus mark-up. The corpus shows that both error types are frequent (188 and 68 errors respectively) and their correction is relevant for the language community. All rules have been released May 31st 2023 and are freely available for the South Sámi language community. It needs to be marked-up for grammatical errors of the type we are investigating. Previous versions did not include certain types of mark-up for the following reasons: 1) The norm had not beeb clear at that point of time. 2) Manual mark-up is cumbersome, and not all error instances are easy to detect.

When further investigating the reasons for the shortcomings of our tool we found the following: In ex. (8) attributive *guelhties* is erroneously corrected to predicative *guelhtie*. The reason for that is that rules are missing a condition for possible coordination. This can easily be specified and cor-

	Precision	**Recall**	**# Err**
Adjective errors	71.81%	85.99%	188
Negation errors	75.00%	79.69%	68

Table 6: Evaluation of the South Sámi grammar checker on *FREECORPUS*

rected.

(8) Bovtside   leah   **guelhties** jih
reindeer.ILL be.3SG cool.ATTR and
gaaloes   giesie hijven.
rainy.cool.ATTR summer good.PRED
'For the reindeer, a cool and rainy summer is good.'

In ex. (9) there is another false positive. Even though the adjective *aelhkie* 'simple' precedes a noun, it is not attributive. Instead, it is part of an infinitive construction of the type 'it is easy to + infinitive'. Therefore the adjective should have the predicate form. However, being an SOV language, in South Sámi, the infinitive can be preceded by an object, here *ditnie-laejkiem* 'tin wire', which leads to the adjective being adjacent to the noun, a typical attributive context.

(9) Ij   leah   **aelhkie**
be.3SG be.CONNEG easy.PRED
ditnielaejkiem giesedh.
tin.wire.ACC   pull.INF
'it is not easy to pull a tin wire.'

The previous example is a recurrent false positive type, just as in ex. (10), where predicative *vihkele* 'important' is erroneously corrected to attributive *vihkeles* since it is followed by a noun. However, this is an infinitive construction with an object before the infinitive just as in the previous example, and the predicative form of the adjective is correct.

(10) lea   **vihkele**   saemiengïelem
be.3SG important.PRED Sámi language.ACC
åtnose   bertedh bievnese-   jïh
use.ILL adjust   information and
gaskesadteme teknologijisnie
communication technology.ACC
'it is important to adjust the Sámi language for use in information and communication technology'

Another false positive caused by homonymy, in this case between adjective and verb, is the markup of the present participle form *båetije* 'coming' as in ex. (11).

(11) *båetije   saemien *siebredahken
      come.PRES.PTC Sámi.GEN society.GEN
      dïejveldimmine
      discussion.INE
      'in future debates about the Sámi society'

All three syntactic contexts can easily be included in error correction rules as exceptions.

As there are many more negation error types, negation rule shortcomings are the following. One issue negation rules have not been paying attention to is the homonymy between finite and a connegative forms like *lij* 's/he was' in ex. (12), resulting in a false positive. *GramDivvun* tries to correct the form based on the assumption that it is a finite form. However, a negative condition excluding possible connegatives, should take care of this problem.

(12) Saemien   *siebredahken tseegkemisnie
      Sámi.GEN society.GEN   building.INE,
      ij   **lij**   gaajhkide
      NEG.PAST.3SG be.PAST.CONNEG all.ILL
      saemientjïertide seamma nuepie
      Sámi.groups.ILL same   possibility
      'In building the Sámi society, there were not the same opportunities for all Sámi groups'

5   Conclusion

In this article we present the first South Sámi grammar checker for adjective and negation error correction, which are both very frequent error types in contemporary writing and speech. Our evaluation on an error marked-up corpus confirms these tendencies. The loss of language arenas in a bilingual society and insufficient grammar teaching in schools have contributed to interference errors and a loss of morphological distinctions. One of the consequences is the use of a simplified adjective paradigm. The negation paradigm, on the other hand, displays a lot of variation in use, both regarding the negation verb and the connegative form. There are errors where the copula copies the main verb paradigm, others where the negation verb is used as an adverb, or agreement is neglected.

The grammar checker tool plays an important role in language revitalization as wished by the language community, implementing normative decisions by means of much needed grammatical feedback. *GramDivvun* shows precisions above 71% for adjective and negation form correction. *GramDivvun* for Microsoft Word and Google Docs has been released in May 2023 and is freely available for download. Future plans include improvement of existing error type correction and correction of other frequent error types.

References

SIKOR uit the arctic university of norway and the norwegian saami parliament's saami text collection, version 06.11.2018. `http://gtweb.uit.no/korp`. Accessed: 2018-11-06.

Kenneth R Beesley and Lauri Karttunen. 2003. Finitestate morphology: Xerox tools and techniques. *CSLI, Stanford*.

Knut Bergsland. 1946. *Røros-lappisk grammatikk: et forsøk på strukturell språkbeskrivelse / av Knut Bergsland*. H. Aschehoug Co. (W. Nygaard); Cambridge, Mass.: Harvard University Press, Oslo.

Knut Bergsland. 1994. *Sydsamisk Grammatikk*. Davvi Girji o.s., Karasjok. 1. utgave 1982.

Eckhard Bick and Tino Didriksen. 2015. CG-3 – beyond classical Constraint Grammar. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NoDaLiDa 2015)*, pages 31–39. Linköping University Electronic Press, Linköpings universitet.

Rogier Blokland and Cornelius Hasselblatt. 2003. *The endangered Uralic languages*. John Benjamins Publishing Company.

Rogier Blokland and Nobufumi Inaba. 2015. https://doi.org/10.1075/tsl.108.14blo *Negation in South Saami*, pages 377–398. Uppsala University University of Turku.

Tino   Didriksen.   2010. http://visl.sdu.dk/cg3/vislcg3.pdf (Accessed 2017-11-29) *Constraint Grammar Manual: 3rd version of the CG formalism variant*. GrammarSoft ApS, Denmark.

Gustav Hasselbrink. 1981-1985. *Oårj'elsaamien baaguog'ärjaa, Skrifter utgivna genom Dialekt- och folkminnesarkivet i Uppsala Ser. C, Lapskt språk och lapsk kultur nr 4, Lundequistska bokhandeln*. Uppsala.

Anna Jacobsen. 2013. *Mijjen Gïele*. Sijti Jarnge, Hattfjelldal.

Lena   Kappfjell.   2014.   *Tjaalehvuekieh*. ČálliidLágádus, Kárášjohka-Karasjok.

Maja Kappfjell and Trond Trosterud. forthcoming. Åarjelsaemien gööktelïhtse adjektijvi gramatihke. page .

Fred Karlsson. 1990. Constraint grammar as a framework for parsing unrestricted text. In *Proceedings of the 13th International Conference of Computational Linguistics*, volume 3, pages 168–173, Helsinki.

Eliel Lagercrantz. 1926. *Sprachlehre des Westlappischen nach der Mundart von Arjeplog. Mémoires de la Société Finno-ougrienne LV.* Helsinki.

Krister Lindén, Erik Axelson, Senka Drobac, Sam Hardwick, Juha Kuokkala, Jyrki Niemi, Tommi A Pirinen, and Miikka Silfverberg. 2013. Hfst—a system for creating nlp tools. In *International workshop on systems and frameworks for computational morphology*, pages 53–71. Springer.

Ole Henrik Magga and Lajla Mattsson Magga. 2012. *Sørsamisk grammatikk*. Davvi girji, Kárásjohkka.

Sjur N. Moshagen, Tommi A. Pirinen, and Trond Trosterud. 2013. Building an open-source development infrastructure for language technology projects. In *NODALIDA*.

Tommi A. Pirinen and Krister Lindén. 2014. State-of-the-art in weighted finite-state spell-checking. In *Proceedings of the 15th International Conference on Computational Linguistics and Intelligent Text Processing - Volume 8404*, CICLing 2014, pages 519–532, Berlin, Heidelberg. Springer-Verlag.

Pekka Sammallahti. 1998. *The Saami Languages: An Introduction*. Davvi Girji, Karasjohka.

Trond Trosterud. 2019. Kva bruker vi minoritetsspråksordbøker til? ein studie av brukarloggane for tolv tospråklege ordbøker. *LexicoNordica*, pages 177–198.

Linda Wiechetek, Sjur Nørstebø Moshagen, Børre Gaup, and Thomas Omma. 2019. Many shades of grammar checking – launching a constraint grammar tool for North Sámi. In *Proceedings of the NoDaLiDa 2019 Workshop on Constraint Grammar - Methods, Tools and Applications*, NEALT Proceedings Series 33:8, pages 35–44.

Linda Wiechetek, Flammie A Pirinen, Børre Gaup, and Thomas Omma. 2021. https://aclanthology.org/2021.iwclul-1.6 No more fumbling in the dark - quality assurance of high-level NLP tools in a multi-lingual infrastructure. In *Proceedings of the Seventh International Workshop on Computational Linguistics of Uralic Languages*, pages 47–56, Syktyvkar, Russia (Online). Association for Computational Linguistics.