Machine translation with North Saami as a pivot language

Lene Antonsen and **Ciprian Gerstenberger** and **Maja Kappfjell**
Sandra Nystø Rahka and **Marja-Liisa Olthuis**
Trond Trosterud and **Francis M. Tyers**
Department of Language and Culture
UiT The Arctic University of Norway
firstname.lastname@uit.no

Abstract

Translating from a majority language into several minority languages implies duplicating both translation and terminology work. Our assumption is that a *manual* translation into one of the languages, and machine translation from this one into the other ones, will both cut translation time and be beneficial for work on terminology. We test the functionality of North Saami as a pivot language, with subsequent machine translation into South, Lule and Inari Saami.

1 Introduction

In this paper, we present a workflow with *manual* translation from the majority languages Finnish, Norwegian (and Swedish) into North Saami and subsequent rule-based machine translation (hereafter MT) into the target languages (hereafter TL) South, Lule and Inari Saami. Thus North Saami is source language (SL) for the MT system and pivot language for the overall evaluation[1]. The system is based upon grammatical analysis of sme transfer lexica, lexical-selection rules, and transfer rules for the syntactic differences between the languages. We deemed the rule-based approach a good fit for closely related languages with complex morphology and very few parallel texts.

In the remainder of the paper we delineate the linguistic and theoretical background of the project (Section 2), give an overview of the project (Section 3), describe the evaluation method of the systems (Section 4) and discuss different aspects of the evaluation method (Section 5). Finally, we point out the importance of such systems both for research and for society (Section 6).

[1] We will refer to the working languages by their language code: sma, sme, smj and smn for South, North, Lule and Inari Saami, as well as nob and fin for Norwegian Bokmål and Finnish.

2 Background

The Saami branch of the Uralic language family consists of 6 literary languages, 4 are included in the present article. sme is the largest one, it has 25,700 speakers in Norway, Sweden and Finland. The smaller languages, smn, sma and smj, each count 450–2000 speakers.[2] With the exception of sma, the neighbouring Saami languages are to some extent mutually intelligible.

All Saami languages are endangered minority languages, having a limited position as an official language in some domains in modern society. There is a continuous shortage of texts, and the lack of both writers and translators is a bottleneck to building full literacy. sme is in a better position than the other languages, especially in Norway, where the imbalance in speaker base is largest, i.e. the proportion of sme speakers to sma and smj speakers is the highest.

Our goal is to explore the use of MT between closely-related languages, for easing the translation bottleneck, by a setup with manual translation to one Saami language and then MT to other Saami languages, instead of manual translation from the three majority languages into several Saami languages (given the lack of MT systems into Saami).

2.1 Previous work

The literature on Saami MT includes several works. Relevant here is an article about an early version of an MT system sme → sma on a limited domain, where sme is used as pivot language for nob to sma translation (Antonsen et al., 2016)[3]. In a study on pivot translation for under-resourced languages, (Babych et al., 2007) a Ukrainian–English RBMT system performes better with the

[2] http://www.ethnologue.com and Pasanen (2015)

[3] The other works are (Tyers et al., 2009) on an early sme → smj system, comparing rule-based and statistical MT, (Wiechetek et al., 2010) on lexical selection rules for the same language pair, and (Trosterud and Unhammer, 2013) on an evaluation of a sme → nob system.

Proceedings of the 21st Nordic Conference of Computational Linguistics, pages 123–131,
Gothenburg, Sweden, 23-24 May 2017.

aid of Russian as a pivot language than one without.

Using Spanish as a pivot language between English and Brazilian Portuguese, (Masselot et al., 2010) shows translators only English original and Brazilian Portuguese MT output. This is a similar approach to ours: the evaluators were shown the fin or nob original and the target language MT output made by translating from the manually translated output in sme.

3 The project

The MT systems were implemented with Apertium (Forcada et al., 2011), which is a highly modular set of tools for building rule-based MT systems. For each language pair, the pipeline consists of the following modules[4]:

- morphological analysis of the SL by means of a Finite-State Transducer (hereafter FST)
- disambiguation and syntactic analysis with Constraint Grammar (hereafter CG)
- lexical transfer (word translation of the disambiguated source)
- lexical selection (choice of contextually appropriate lemma)
- one or more levels of FST-based structural transfer (reordering and changes to morphological features)
- generation of TL by means of FST

Figure 1 offers an overview of the modules and shows the output on each processing step.

3.1 Resource challenges

The backbone of the MT system is the lexical mapping, which is implemented as a dictionary between SL and TL. The described MT project deals with pairs of minority languages. As before the project there were no dictionaries between Saami languages, the resources had to be compiled in various ways.

Due to the proximity between sme and smj, it was possible to map the sme lexical entries into smj by means of a transliteration FST. The output was then post-edited by a native speaker of smj. This simple yet effective method ensured that the SL lexicon was congruent with the TL lexicon. However, this shallow lexical mapping is not possible for Saami languages that are by far more different than sme, as it is the case with sma and smn.

[4]For a presentation of the grammatical analysers and generators, see Antonsen et al. (2010) and Antonsen and Trosterud (forthcoming).

The dictionary between sme and sma was built by crossing the sme–nob with the nob–sma dictionary, both compiled at Giellatekno. The coverage of the resulting sme–sma dictionary has been incrementally extended during the system development work

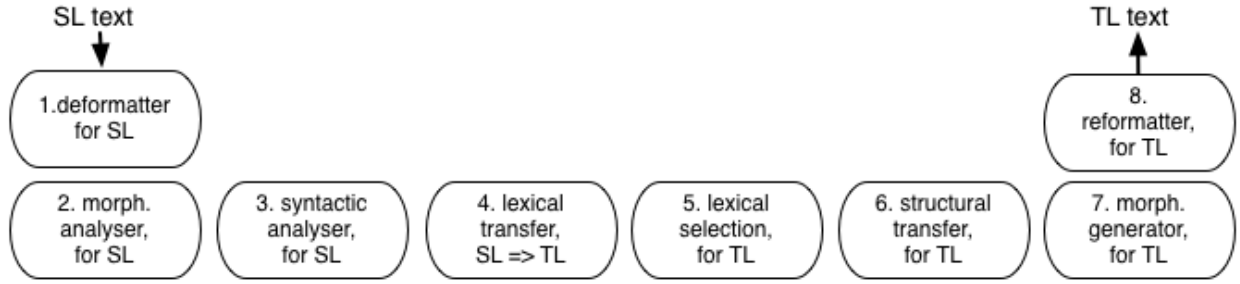
The most difficult case was the compilation of the sme–smn lexical resource. The candidate word pairs were created by mapping sme–fin onto the fin–smn, and since the fin–smn dictionary gave several smn translations for each entry, the resulting sme–smn had about 3.62 translations for each smn. Since cognates were in most cases the best candidates, we calculated the Levenshtein distance (Levenshtein, 1965) between the sme form and a version of the smn candidate that was orthographically adjusted to sme, and sent the highest scoring candidate(s) to be manually corrected. As an illustration, for the sme entry *bahčit* 'to milk' there were two smn candidates: *paččeeđ* 'to milk' and *cuskâdiđ* 'to stop a milk-feeding animal from giving milk'. After adjusting for regular sound changes, *paččeeđ* gave a Levenshtein distance of 3, and *cuskâdiđ* of 6, and thus *paččeeđ* was chosen. In the Saami languages, proper nouns are inflected for case, and heuristic recognition of names is thus not sufficient. Therefore, 80-90% of the bilingual dictionaries were devoted to proper noun pairs.

After a manual check, the sme–smn dictionary was ready for use in MT. All three dictionaries have been incrementally extended and refined during the system development.

3.2 Linguistic challenges

3.2.1 Linguistic differences

Generally, the grammatical differences between the Saami languages are minor. However, with only 7 cases, the pivot language, sme, is the one with the smallest case inventory. Of these, nouns and pronouns in accusative share forms with the genitive, and numerals in accusative share forms with the nominative. smn shares the system of grammatical and local cases with sme, but has two extra cases: partitive and abessive (corresponding to the sme postpositional phrase *N haga* ('without N')). smn also makes a distinction between accusative and genitive, most notably in the plural. sma and smj have a richer case system than sme: their genitive and accusative forms are always distinct from each other. Moreover, unlike the locative case syncretism in sme for in and from spatial relations, these languages encode the two different



SL: Son bargá vuođđoealáhusas.

1–3. ^Son<prn><pers><p3><sg><nom><@SUBJ→>$ ^bargat<vblex><tv><indic><pres><p3><sg><@+FMAINV>$
^vuođđu<n><sem_semcon><cmp_sgnom><cmp>+ealáhus<n><sem_domain><sg><loc><@←ADVL-ine>$

4–5. ^Son<prn><pers><p3><sg><nom><@SUBJ→>/Dïhte<prn><pers><p3><sg><nom><@SUBJ→>$
^bargat <vblex><tv><indic><pres><p3><sg><@+FMAINV>/barkedh<v><iv><inf><@-FMAINV>$
^vuođđu <n><sem_semcon><cmp_sgnom><cmp>/våarome<n><sggencmp><cmp>$
^ealáhus<n><sem_domain><sg><loc><@←ADVL-ine>/jieliemasse<n><sem_domain><sg><loc><@←ADVL-ine>$

6. ^Dïhte<prn><pers><p3><sg><nom>$ ^våarome<n><cmp_sgnom><cmp>+jieliemasse<n><sg><ine>$
^barkedh<vblex><tv><indic><pres><p3><sg>$

7–8. TL: Dïhte våaromejieliemassesne barka.

Figure 1: Translation pipeline and processing example for *Son bargá vuođđoeláhusas* 'He works in-the-primary-sector' from sme to sma

relations by inessive and elative case, respectively. Hence, given the case syncretism of the SL, one of the challenges for the MT system is to make the contextually correct case distinction in the target language.

In NP-internal agreement in sme, the adjective does not agree with its head noun, but gets a separate *attributive* form, invariant in the different cases, but marking membership in the NP. In principle, all the other Saami languages have the same system, but there are some differences. smn, on the one hand, has a richer system of semi-agreement for large sub-parts of its adjectives, whereas sma often replaces adjective loanwords with genitive nouns. smj is here closer to sme.

As in fin, negation is expressed by a negation verb in the Saami languages. smn and sme have the same system as fin: in the present tense, the negation verb combines with a form identical to the imperative, while in past tense, it combines with a form identical to the perfect participle. In contrast, smj and sma have an older system, where the negation verb itself is inflected for tense while the main verb is identical to the imperative irrespective of tense.

Regarding syntax, sme and smn are quite similar, whereas in smj and especially in sma there is a strong tendency towards SOV word order, where sme and smn have SVO. With a verb complex *auxiliary + verb* (AV), the sme and smn may also have SAOV in addition to SAVO, which is dependent upon the information structure of the sentence.

For NP structure and treatment of given and new information, sma also differs most from the rest. As for verbs, despite minor differences between SL and TL, the inventories of non-finite verbs are rather similar, which enabled a one-to-one mapping of verb forms.

3.2.2 Analysis of the pivot language

In order to cope with as many of the above-mentioned challenges, we enriched the input in the SL with pieces of information needed for the appropriate choice in the specific TL.

This has been realised partly by adding extra tags in the CG (the syntactic module), and partly by adding parallel paths to the FST (the morphological module). The sme accusative–genitive syncretism may exemplify this: this ambiguity is disambiguated in the syntactic analysis.

The issue of one-to-many mapping of the sme locative case, which should be translated either as inessive or elative into sma and smj, was solved by adding an extra tag to the adverbials in the syntactic analysis, marking the ambiguity between inessive and elative. This way eased the choice of the contextually appropriate case in the TL output.

Additionally, locative was also marked for *habitive*[5] in the syntactic module. Correct mark-

[5]The possessive construction as in 'I have a book.'

ing of habitive versus other adverbials is only relevant for sma, which uses genitive instead of the sme locative. In smj the habitive case is inessive, which is the default translation for the locative if it does not have a tag for elative in the syntactic analysis.

Adding extra tags was also the solution for the frequent sme particle *ge*, which is used for both negative and positive polarity. This extra marking eased the choice of the appropriate forms in smn and sma because these two TLs feature different clitics for polarity marking.

Other case assignment differences between SL and TLs such as **time** and **path** expressions were solved by enriching the SL analysis with tags indicating semantic category (e.g. *Sem/Time*). Semantic tags were used also for structural transfer of sme adposition phrases into sma.

3.2.3 Transfer rules

In a transfer-based MT system, the transfer module takes care not only of the simple lexical transfer but also of any structural discrepancy between source and target language by, e.g., changing morphological attributes, deleting or adding words, or changing word order. Table 1 shows some examples of structural mapping of grammatical patterns between source and target language.

The rules for transforming the word order, e.g., from VX in sme to XV in sma, have to cover all different syntactic constructions that VX is a part of, such as subject ellipsis, progressive constructions, complex objects, and verb phrases as complement to nouns or adjectives. By means of syntactic tags in the sme analysis, the transfer rules build chunks of syntactic phrases, and then the verb is moved past these chunks. Compared to earlier Apertium systems as described in (Antonsen et al., 2016), this is new, and a significant improvement. Unlike the MT systems for smn and smj that contain a similar amount of rules, the sma MT system has three times as many rules. This is due to the syntactic differences between sma and the other Saami languages.

rule type	sme		
	→smn	→smj	→sma
modifying/ chunking	63	75	171
word order	7	24	37
macro rules	38	12	96
total	108	111	304

Table 1: Transfer rules for each of the language pairs. Macro rules modify morphological attributes, as a part of ordinary rules.

4 Evaluation

We evaluated the output of the MT systems in three steps. First, we estimated the lexical coverage, then we analysed and evaluated the amount on editing on the the MT output text via the pivot language. Finally, the evaluators were asked to compare post-editing to translation of a similar text from the majority language, yet **without** access to any MT output text.

4.1 Word coverage

To measure the system coverage, we used a corpus of 8.9 million words, consisting of texts on the Saami school system in Finland as well as administrative texts from the Saami Parliament of Norway. As Table 2 shows, the difference in coverage between the three language pairs is minimal[6].

	coverage	**dynamic comp.**	**dynamic deriv.**
sme–sma	0.938	0.557	15 types
sme–smj	0.940	0.558	22 types
sme–smn	0.944	0.822	26 types
Average	0.941	0.670	

Table 2: Coverage of text corpus (1.0 = 100%)

In Table 2, dynamic compounding means that the system translates any N + N compound. This makes up more than 8% points in coverage for smn and a little more than 5% for the other languages. Another significant difference is in how many dynamic derivations (= all stems are optionally directed to a set of derivational affixes) are transferred from SL to each TL: 26 dynamic derivation types to smn, 22 types to smj and only 15 types to sma.

As indicated by the amount of similarity in dynamic word formation, sme–smn is the most similar language pair both for compounding and derivations, while the largest differences are found between sme and sma.

[6]Note that for sme–sma, this is an improvement over the 87.4% reported in (Antonsen et al., 2016).

	sma	smj	smn	**Total**
WER	0.57	0.46	0.37	0.42
PER	0.45	0.39	0.32	0.35
PER/WER	0.79	0.85	0.86	0.83

Table 3: WER - all languages

	sma	smj	smn
Lexical selection	0.33	0.42	0.38
Word form correction	0.18	0.17	0.28
Word generation correction	0.01	0.11	0.03
Insert/delete/move word	0.43	0.26	0.26
Punctuation	0.04	0.04	0.03
Total	1.00	1.00	1.00

Table 4: Distribution of correction types

4.2 Word Error Rate

4.2.1 Evaluation setup

For the quantitative evaluation, we selected one text in nob and one in fin that had already been manually translated into sme. Since the coverage was measured in a separate test (see Section 4.1), we added the missing sme words into each of the systems. Using the MT systems, we translated the sme text with a nob original into sma and smj, and the sme text with a fin original into smn.

For each language pair, we had three evaluators, who were all professional translators. Each evaluator received both the nob or fin original and the MT output. The task was then to produce a good target language text, either by correcting the MT version or by translating the original. As two evaluators did not post-edit they are treated separately in Section 4.4.

For each evaluator, we calculated Word Error Rate and Position-independent Word Error Rate (hereafter WER and PER) of the MT version as compared to the post-edited text. WER is defined as the number of words being corrected, inserted, or deleted in the post-edited text. PER differs from WER in ignoring word-order changes. Thus, a WER of 10% means that every tenth word has been changed in one way or another in the post-edited text. Average WER and PER values for all evaluators for the different languages are shown in Table 3.

The best values were found for smn, which was also the language with the smallest WER/PER difference. sma had the highest values (i.e. worse results). sma is also the language with the largest WER/PER difference. Given the word order differences between sma and the other Saami languages, these values were as expected.

In order to get a better picture of the challenges, we looked at five different categories for each language pair. This gave the picture in Table 4.

sma stands out with word order being the largest category, for the two others lexical selection is largest, whereas word generation is problematic for smj. We comment on the different types below.

4.2.2 Lexical selection

sme–sma had more lexical selection changes than the other pairs and there was also less consensus among the evaluators as to what to change to. In no instances did the every evaluator agree what to replace the MT suggestion with. Either they disagreed on whether to replace the MT suggestion, or they differed as to what to replace it with. An example of the former is where one evaluator accepted *evtiedimmienuepieh* for *utviklingsmuligheter* ('development possibilities'), where the other one wanted *evtiedimmiehille* (*nuepie, hille* meaning 'possibility', *nuepie* also 'offer'). An example of the latter type was *bærekraftig* 'sustainable', where the MT *gaarsje* was replaced, either with *nænnoes* 'solid' or with *jïjtjeguedteldh* 'self carrying'. Similar examples were also found for sme–smn and sme–smj.

A closer investigation of lexical choice by the evaluators shows that usually the lexemes found in the MT output were retained, indicating that the bilingual dictionary is solid. In the cases where the correct lexeme was not chosen by the system, evaluators did not agree on which was most appropriate.

4.2.3 Word form correction

The choice of wrong forms in the TL output had several causes. Often, the correcting of word form was due to lexical selection, replacing a verb may, for instance, result in changing case for the adverbial as well.

Another reason was difficulties in the SL input analysis, i.e., mistakenly resolved ambiguities. sme features a series of systematic homonymies such as gen vs. acc, inf vs. prs.pl1 as well as sg.com vs. pl.loc. These homonymies can not be preserved into any of TLs: the one–to–many map-

nob	De siste **fem årene** ...
sme	Maŋimus **viđa jagis** ...
sma-mt	Minngebe **vïjhtene jaepesne**...
sma-e1	Minngemes **vïjhte jaepie** ...
sma-e2	Minngemes **vïjhte jaepesne**...
sma-e3	Daah minngemes **vïjhte jaepieh**...
smj-mt	Maŋemus **vidán jagen**...
smj-e1	Maŋemus **vidán jagen** ...
smj-e2	Maŋemus **vidá jage** ...
smj-e3	Maŋemus **vihtta jage** ...
	The last **five years**...

Table 5: Translation of the phrase *de siste fem årene* 'the last five years'

ping problem. As sme input, the nom–acc ambiguous form *dieđusge* 'of course' in the context of the gen–acc ambiguous form *bohccuid* 'reindeers' as in *Ailu lea maiddái oaidnán luonddus máŋggaid ealliid, bohccuid dieđusge* ('Ailu has also seen in-the-nature many animals, reindeers of course'), triggers the wrong gen form *poccui* in smn, instead of the correct acc form *poccuid*. Similar errors were found for the other language pairs as well.

An example of the amount of variation is the translation of the phrase *de siste fem årene* 'the last five years' into sma and smj. As presented in Table 5, all six evaluators gave different versions of the phrase, and only one of them agreed with the MT output (for smj) . This demonstrates that the languages involved have weak norms.

4.2.4 Word generation correction

Word form generation correction occurs when there is a correct analysis of the input, there is a correct mapping in the bilingual dictionary, but some word forms in the TL are not generated properly or the evaluator prefers another possible normative form. Generation corrections constituted the smallest type of the post-editing corrections. This indicates that each transducer is an accurate representation of the grammar of the language it models. The FST use for the proofing tools of the different Saami languages also supports this observation. smj stands out with the worst results here, this is partly due to different orthographic conventions for smj in Norway and Sweden.

4.2.5 Reordering, addition, and deletion

A common type of word addition is the addition of grammatical words. Thus, for the original *stimulere til etablering av nye næringer innenfor nye bransjer* ('stimulate the establishing of new businesses within new industries') the sme *ođđa surggiin* 'new industry.loc' (from nob *innenfor nye bransjer* ('within new industries') was rendered with inessive *orre suerkine* by the system. One of the evaluators accepted this, and the other one inserted a postposition instead (*orre suerkiej sistie* 'new industry.pl.gen within.po').

There is no norm for how to write year-numerals in sma and smj, and two of the evaluators for smj and one for sma had added the word 'year' for the case marking, e.g. inessive in front of postposition in smj: *jagen 2000 rájes* 'year-ine 2000 from' 'from 2000'.

In all three Saami languages pro-drop is common, but the pronouns tend to be kept in translations from languages without pro-drop. Both for sma and smj two of the evaluators deleted the third person singular pronominal subject in the same sentence in the MT text.

Word order change was an issue sma and smj. smn sentences, however, kept the sme word order. This was accepted by the evaluators, as expected, given the high degree of syntactic similarity between the two languages.

4.3 Qualitative evaluation

In addition to the text discussed in the previous section (Text B), the evaluators got another, equally-sized text (Text A) in the original language (nob/fin), without a machine-translated version. The level of difficulty of Text A was estimated to be similar to that of Text B. In addition to post-editing or translating Text B to the target language, the evaluators were asked to translate Text A. The second part of the evaluation consisted in comparing the two tasks: translation with and without the help of a pivot language. This step was carried out via a questionnaire[7] containing three multiple choice questions (cf. Table 6):

1. Compare the time you spent on the two texts, Text A (translating from scratch) and Text B (using the MT version).
2. How did you use the MT version?
3. Do you think that such an MT program will be useful for you as a translator?

In addition, there were two open questions: The evaluators were asked to comment upon the terms suggested by the MT system that cannot be found

[7]The URL to the original texts sent out will be provided after review.

Time spent	sma	smj	smn	Σ
more time on A than B	0	3	1	4
same amount of time on both	2	0	2	4
more time on B than A	0	0	0	0
How did you use the MT version?	sma	smj	smn	Σ
I used it for post-editing	2	2	3	7
I translated from scratch ...				
... but used it to find terms	0	1	0	1
... but it was of some help	1	0	0	1
It is so bad that I cannot use it	0	0	0	0
Is this MT program useful?	sma	smj	smn	Σ
Yes, ...				
even as it is now	3	3	3	9
only after much improvement	0	0	0	0
only when almost perfect	0	0	0	0
No, I do not think so	0	0	0	0

Table 6: Answers to multiple choice questions

in relevant term collections, and they were invited to comment freely upon their experience with using the MT program.

Both the sma and smj evaluators appreciated the new terms suggested by the MT system, although, in several instances, they would not have used the terms proposed.

Except for one smn evaluator, who had no comments, all others had positive overall comments to the program. It was of 'great help', it did the job of looking up all unknown words, and it was able to consistently give a good translation, where a human translator might get bored and fall back to just copying the nob syntax.

4.4 Translating from scratch

One sma and one smj evaluator did not post-edit, instead, they translated the text from scratch, yet using the MT output as a reference. Both had considerably higher WER results that the evaluators who have post-edited the MT output. It seems that MT output post-editing in itself gives rise to solutions closer to the MT output, thus closer to the pivot language sme.

A case in point is when the nob original writes about *en analyse som Telemarksforskning har gjennomført for Sametinget*, 'an analysis which T. has conducted for the Saami parliament'. Both the MT and the two evaluators post-editing the output write *mej Telemarksforskning tjïrrehtamme Saemiedigkien åvteste*, on a par with the sme *lea čađahan Sámedikki ovddas*. The third evaluator, writing from scratch, finds a drastically different solution. In this translation, Telemarksforsking conducts an analysis which the Saami parliament supports (*maam Saemiedægkan dorjeme*). Again, for the wording choices of the translated text, there is a difference between post-editing an MT output and translating from scratch.

5 Using a pivot language

The first manual step in the translation process, from the original to the pivot language, has clearly had an influence on the result. To investigate the impact of this influence on the current translation process, we compare the WER results in Section 4.2 from a parallel evaluation from sme to smn, yet, this time measured not against the fin original, but against the MT source language itself. Where the two-step translation process gave WER and PER values of 0.37 and 0.32, respectively (cf. Table 3), the corresponding values for a similar translation from North Saami as SL were 0.11 and 0.11, more than three times as good.

In retrospect, we see the following as a weakness in the evaluation. In the first step, from nob and fin to sme, we deliberately chose actual translations, in order to make what we saw as a realistic setup. The next step, from sme to the target Saami language, we conducted as described in 4.2.1. The result was that the two translation steps served different functions: The first step made a sme text for a concrete set of readers in a concrete setting, whereas the last step was part of a decontextualised evaluation process. Rather than aiming at a realistic case only for the first step, we should have ensured the same function across the whole translation chain, either by having translators translate (accurately) from nob/fin to sme, or by correcting the sme translation ourselves.

The syntactic analysis of the pivot language is crucial for the generation or the correct target language sentence and the importance of a correct syntactic analysis increases with larger syntactic differences between pivot and target language. The smj and sma evaluation texts were the same, and nine bad suggestions in the sma MT output text were due to incorrect analysis: five because of incorrect or deficient disambiguation and four because of incorrect syntactic tag. Due to syntactic similarities between sme and smj, the same nine errors in the input analysis caused only three errors in the smj MT output text.

The target languages of this study are continuously suffering from the lack of adequate terminology, especially concerning the modern society and special fields. For example 'archery', fin *jou-*

siammunta, was translated into sme with *dávgebissuin báhčin* ‘shooting with bow gun’, wherefrom it can be taken into smn with *tävgipissoin pääččim*.

Secondly, also some idiomatic expressions could be created using MT. The fin expression *toiminnallisilla rasteilla* ‘at functional posts’ (along the trail) can not be literally translated. The sme translator has chosen for *doaimmálaš bargobádji* ’functional workshop’. The same expression *toimâlâš pargopääji* can also be used for smn:

fin	toiminnallisilla rasteilla tutustutaan muun muassa riistanhoitoon.
sme	Doaimmalaš bargobájiid áigge mánát besset oahpásmuvvat fuođđodikšumii.
mt	Toimâlij [pargopääjái ääigi] párnááh peesih uápásmuđ pivdoelleetipšomân.
e12	Toimâlij pargopääjáin párnááh peesih uápásmuđ pivdoelleetipšomân.
e3	Toimâlij pargopääjái ääigi párnááh peesih uápásmuđ pivdoelleetipšomân.
tr	During the workshops the children get to know how the wild animals are treated.

Offering literal translations, dynamic compounding and derivation from sme, the program successfully suggests adequate terms or other translation solutions. This is possible while a perfect equivalence at word level between the pivot language and the TL exists. This phenomenon was pointed out by several evaluators, especially the sma ones, as a positive experience with MT translations.

6 Conclusion

We have presented a project in which we built three rule-based MT systems to translate from sme to sma, smj and smn, respectively. Each of the systems was tested for coverage and three evaluators post-edited the MT translations and gave feeback on the system quality via a questionnaire.

All the MT systems were judged as useful by the evaluators, especially with respect to terminology. All but two evaluators used the MT output as a basis for post-editing, rather than writing from scratch. Half of the evaluators found post-editing time-saving, the rest found it equally fast as manual transtion.

A central problem was the lack of a stable norm in the target languages, both with respect to terminology, orthography and syntax, which made it hard to present a translation that could gather consensus among the evaluators. The lion’s share of the errors still came from the pivot translation not following the original. With manual translations into the pivot language being closer to the original text, we anticipate the present setup to improve considerably.

Acknowledgments

This work was financed by Norsk forskingsråd (grant No. 234299), the Kone Foundation as well as our university. Thanks to Erika Sarivaara for work with the smn transducer.

References

Lene Antonsen and Trond Trosterud. forthcoming. Ord sett innafra og utafra – en datalingvistisk analyse av nordsamisk. Norsk Lingvistisk Tidsskrift.

Lene Antonsen, Trond Trosterud, and Linda Wiechetek. 2010. Reusing grammatical resources for new languages. In *Proceedings of LREC-2010*, Valetta, Malta. ELRA.

Lene Antonsen, Trond Trosterud, and Francis Tyers. 2016. A North Saami to South Saami machine translation prototype. *Northern Europe Journal of Language Technology*, 4.

Bogdan Babych, Tony Hartley, and Serge Sharoff. 2007. Translating from under-resourced languages: comparing direct transfer against pivot translation. In *Proceedings of the MT Summit XI*, pages 29–35.

Mikel L. Forcada, Mireia Ginestí-Rosell, Jacob Nordfalk, Jim O’Regan, Sergio Ortiz-Rojas, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Gema Ramírez-Sánchez, and Francis M. Tyers. 2011. Apertium: a free/open-source platform for rule-based machine translation. *Machine Translation*, 25(2):127–144.

V. I. Levenshtein. 1965. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady 10, 707–710, trans. from Doklady Akademii Nauk SSSR*, 163:845–848.

Francois Masselot, Petra Ribiczey, and Gema Ramírez-Sánchez. 2010. Using the apertium spanish-brazilian portuguese machine translation system for localization. In *Proceedings of the 14th Annual Conference of the European Association for Machine Translation, EAMT10*.

Annika Pasanen. 2015. *Kuávsui já peeivičuovâ. ’Sarastus ja päivänvalo’ : Inarinsaamen kielen revitalisaatio*. Uralica Helsingiensia, Helsinki.

Trond Trosterud and Kevin Brubeck Unhammer. 2013. Evaluating North Sámi to Norwegian assimilation RBMT. In *Proceedings of the Third International Workshop on Free/Open-Source Rule-Based Machine Translation (FreeRBMT 2012)*, volume 3 of *Technical report*, pages 13–26. Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg.

Francis Tyers, Linda Wiechetek, and Trond Trosterud. 2009. Developing prototypes for machine translation between two Sámi languages. In *Proceedings of the 13th Annual Conference of the European Association for Machine Translation, EAMT09*, pages 120–128.

Linda Wiechetek, Francis Tyers, and Thomas Omma. 2010. Shooting at flies in the dark: Rule-based lexical selection for a minority language pair. *Lecture Notes in Artificial Intelligence*, 6233:418–429.