Improved Norwegian Bokmål Translations for FLORES

Petter Mæhlum
Language Technology Group
University of Oslo, Norway
pettemae@ifi.uio.no

Anders Næss Evensen
Disputas
anders@disputas.no

Yves Scherrer
Language Technology Group
University of Oslo, Norway
yves.scherrer@ifi.uio.no

Abstract

FLORES+ is a collection of parallel datasets obtained by translation from originally English source texts. FLORES+ contains Norwegian translations for the two official written variants of Norwegian: Norwegian Bokmål and Norwegian Nynorsk. However, the earliest Bokmål version contained non-native-like mistakes, and even after a later revision, the dataset contained grammatical and lexical errors. This paper aims at correcting unambiguous mistakes, and thus creating a new version of the Bokmål dataset. At the same time, we provide a translation into Radical Bokmål, a sub-variety of Norwegian which is closer to Nynorsk in some aspects, while still being within the official norms for Bokmål. We discuss existing errors and differences in the various translations and the corrections that we provide.

1 Introduction

This paper describes our submission to the WMT 25 open language data shared task, where participants were asked to contribute to open dataset collections such as FLORES+, the MT Seed dataset or other parallel datasets. We have chosen to focus on the Norwegian Bokmål part of the FLORES+ dataset, as the authors notice non-fluencies in the dataset in 2024, and notified the original authors. These issues were attempted resolved in a process that lead to additional errors, which are the ones that form the basis of this paper. In addition to correcting these translations, we translate the resulting Norwegian Bokmål dataset into a version of a specific variety of written Norwegian called radical Bokmål. Having these two normed varieties can be beneficial for experiments where variation in Norwegian spelling norms is important. We summarize some of the encountered errors in the newest Bokmål translations, and show some results on several machine translations baselines for the new and existing Norwegian versions.

2 The Norwegian Language and its Writing Norms

Norwegian is one of the official languages of Norway, along with Sámi languages and Norwegian Sign Language. It is a North-Germanic language historically descendant of Western Norse, but following large Saxon and East Norse influences, is largely mutually intelligible with its neighbors Swedish and Danish, and more different from Icelandic and Faroese. However, following centuries of having Danish as Norways national language, nationalist movements in the late 19^{th} century lead to the establishment of two written standards: Landsmål (today **Nynorsk**), which was based on dialects "untainted" by Danish, and Rigsmål (Riksmål, today **Bokmål**), which was Norwegianized Danish. Nynorsk historically aimed at preserving Norwegian-specific features, which means that Saxon and Danish influences are less pronounced in Nynorsk than in Bokmål.

2.1 Conservative, Moderate and Radical Bokmål

Within both written norms however, considerable variation exists. This variation is not arbitrary, and generally follows typical patterns, leading to what is known as **norm clusters** (nor. *normklynger*) (Dyvik, 2009). On the Bokmål side, perhaps the most common sub-norm is **Moderate Bokmål** (MBM) (or Conservative Bokmål, CBM)[1], which is what dominates especially formal Norwegian discourse. This variety is known for Danish-like conjugational and declensional patterns: preterite in *-et*, no feminine nominal endings (with a few exceptions). On the other hand, **Radical Bokmål** (RBM) aims at a style closer to how many people in

[1]While the terms *moderate* and *conservative* are used synonymously by some authors, some prefer to reserve *conservative* for the most extreme (Danish-like) Norwegian, and reserve *moderate* for a norm that allows for some radical elements.

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 882–890
November 8–9, 2025.

(Eastern) Norway speak, and adopts features shared with Nynorsk (NN) in some regards: obligatory feminine marking for articles and noun declension, preterite in *-a* (NN also *-a*) where MBM has *-et*, and neuter definite plurals in *-a* (NN also *-a*) where MBM has *-ene*. There are also some sound correspondences, with a difference between diphthongs and monophthongs being especially common. For example, RBM has *mjølk* where MBM has *melk* (NN *mjølk)*. As illustrated in the artificial example below, this affects both morphology and syntax, with Nynorsk added for comparison. Note how while RBM is said to be closer to Nynorsk, this is not to say that all RBM forms exist in Nynorsk, as exemplified by *skog* (MBM, NN) and *skau* (RBM).

(1) MBM: I helgen var jeg på hytten i skogen og danset med min søster.

RBM: I helga var jeg på hytta i skauen og dansa med søstera mi.

NN: I helga var eg på hytta i skogen og dansa med søstera (or systera) mi.

The degree to which a writer follows these patterns differs, leading to many possible variations within this spectrum. In our efforts to provide a radical form, we chose the most radical form as presented in the official Norwegian dictionary *Bokmålsordboka*.[2]

3 The FLORES Dataset

The FLORES dataset is an evaluation dataset for multilingual machine translation, consisting of a *dev* and a *devtest* part with about 1000 sentences each. The dataset is multiparallel and English-centric: the original sentences are in English, and all other language variants were produced by translation. Several versions were made available over time, reflecting efforts to increase language coverage and address quality issues.

FLORES101 was the first version of FLORES, covering 101 languages, including Norwegian (Goyal et al., 2021). While the authors claimed that the sentences were "[...] translated in 101 languages by professional translators through a carefully controlled process", we observed severe quality problems with the Norwegian Bokmål translations. See further discussions in 3.1.

[2]Bokmålsordboka. The Language Council of Norway (Språkrådet) and the University of Bergen https://ordbokene.no.

FLORES200 was a continuation from both FLORES101 and Guzmán et al. (2019), with an increased coverage of 200 languages (NLLB Team, 2022). The Norwegian sentences appear to be unchanged between FLORES101 and FLORES200. The FLORES200 translations were used, among others, in the Belebele (Bandarkar et al., 2024) benchmark. Quality problems in the former therefore directly affect results reported on the latter dataset. We have used the Bokmål sentences from FLORES200 both as an aid in correcting the translations, and as a point of comparison agains the new dataset as a whole. These sentences initially struck the authors as unnatural, with exampled reported such as translating *iron* (the metal) as *strykejern* (eng. 'clothes iron'), and (judicial) *court* as *hoff* (eng. royal court).

FLORES+ The responsibility for the FLORES datasets was eventually moved to the Open Language Dataset Initiative.[3] As a result, the updated versions are referred to by FLORES+ and published on HuggingFace.[4]

In January 2024, the authors of this paper reached out to the original FLORES101 authors to express concern over the quality of the Norwegian Bokmål dataset, based on FLORES200. Following this, the dataset was updated, as indicated by a changelog note from November 11^{th} 2024[5]. This note informs that the Norwegian version has been updated after quality assessment, but with no further information. Going through these changes, we see, however, that not all errors were corrected, and that new ones were introduced. Correcting these errors is the main focus of this paper.

3.1 Problems with the Norwegian FLORES Translations

The Norwegian version published as part of FLORES101 (referred to as **BM1**) contained a range of issues, some of which were amended in the FLORES+ version (referred to as **BM2**).

When comparing the BM1 and BM2, the naturalness of the sentences have improved in some cases, but there are still multiple mistakes left in the dataset. Overall, BM1 used a syntax that was less influenced by the original English. In some

[3]https://oldi.org/

[4]https://huggingface.co/datasets/facebook/flores

[5]https://huggingface.co/datasets/openlanguagedata/flores_plus/blob/main/CHANGELOG.md#20--2024-11-11

cases, BM2 is even worse than BM1. Every single sentence in the Bokmål dataset was changed during this edit, but it seems like the BM2 translations show signs of not having referenced the BM1 translations, as in several cases where both Nynorsk and BM1 have a correct translation, but BM2 still mistranslates, for example the English *engraver*, which is translated correctly as *gravør* in NN and BM1, is erroneously translated as *graver* 'digger' in BM2. While the Nynorsk dataset also had changes in 21 sentences, this turned out to only be differences in trailing spaces.

Certain errors are so pronounced that it is difficult to conceive they were produced by a professional translator, or even by a fluent speaker of Norwegian. This can be illustrated by example 2 where the English word 'bill' has been translated as *lovforslag* 'bill (judicial)', while in the new translation it has been changed to *regning* 'bill, receipt'. The terms are not ambiguous as in English, and the result is comical.

(2) BM1: Det opprinnelige **lovforslaget** ble utarbeidet av tidligere ordfører i São Paulo [...]

BM2: Den opprinnelige **regningen** ble utarbeidet av tidligere borgermester i São Paulo [...]

EN: The original **bill** was drafted by former mayor of São Paulo, Marta Suplicy.

These examples make it clear that despite the corrections being made in 2024, not all mistakes were fixed, and some new ones were introduced. This prompted us to critically assess the BM2 translations and correct them. In the following section we describe our correction process, then follow with a brief overview of error types and key statistics. We refer to our corrected Bokmål translations of FLORES+ as **BM3**, and to the Radical Bokmål version as **BM3R**.

4 Translation Correction

We introduce our methodology and discuss some of the encountered errors. See Appendix A for selected example sentences in all languages involved in this process.

4.1 Methodology

Our aim for this effort is not a full retranslation of the English, but rather to take the BM2 translations as a starting point. We assume that the BM2 translations follow the FLORES translation guidelines[6], which do not allow any AI tools. They should therefore provide an appropriate point of departure, being the most recent translations. We aim mostly at correcting *obvious* grammatical and lexical mistakes in the dataset, while keeping the structure and otherwise correct lexical choices of the original translators intact. However, in more severe cases, the other sentences were used as reference, especially the NN sentences, as they overall hold a much higher quality level, though not free of errors. Some concrete breaches of the translation guidelines were also corrected, notably cases where units of measure were translated and converted. There were also cases of named entities that were not changed, even though an established Norwegian spelling exists, such as Eng. *Pythagoras* vs. Nor. *Pytagoras*.

Measured in error correction on the most recent Bokmål translations, 64.4% of the sentences in the devtest split contained errors that were fixed, with 70.8% for the dev split. We make an attempt at avoiding corrections due to matters of choice, but do correct in the following cases:

1. Grammatical mistakes
2. Clearly mistranslated terms
3. Orthographic or punctuation mistakes
4. Misunderstanding of context, etc.
5. Mismatching radicalness

In some cases, errors in the Nynorsk were discovered, but they were not corrected. We urge others to revisit the Nynorsk translations.

Two annotators worked on the correction task, with about 25 hours of work for each annotator. Both were native Norwegian speakers with some background in professional translation. The datasets were split in two, with each person correcting their half, before quality checking the other person's corrections. It was not the translators' intention to re-translate, simply to correct the existing translations, basing the new translations on these existing ones to provide a more fluent and correct sentence, and thus improving the overall quality of the dataset. Changes in radicalness were only done in cases where it did not match the dataset overall.

In the rare cases where none of the earlier translations are correct, the translators allow themselves

[6] https://oldi.org/translation-guidelines.pdf

to retranslate. This is mostly seen in the case of specific jargon or common misunderstandings, in places where the English syntax is kept in the translations despite being ungrammatical, or if there were no appropriate translations for them to base themselves on. Therefore some cases with English-like syntax but no grammatical or lexical mistakes have been kept.

When the communicative intent of the English sentence is not hindered by small errors, these errors are not carried over in the Norwegian translation.

Although difficult to avoid completely, the translators have attempted to avoid letting stylistic preferences affect the correction. For example in cases where the translations is passable, but the translator would prefer another word, we have avoided correcting them, except in cases where the sentence sounds very unnatural. All cases were discussed between the two annotators. In some borderline cases, where it is difficult to argue in disfavor of the original translation, the sentences are kept, as with the cases discussed above. Semi-authoritative sources such as the Norwegian Wikipedia, the Norwegian encyclopedia Store Norske Leksikon (SNL) or books in the National Library (NB) collection were used to guide term usage.[7]

4.2 Types of errors

In order to give an impression of some of the mistakes found necessary to correct by the annotators, we attempt to summarize some of the more common ones in broad groups.

Term Coinage and Anglicisms The original translator has in several instances made up words with little to no previous usage, in cases where there are clearly preferred terms. Examples include *meteordusj* 'meteor shower' for meteorsverm lit. 'meteor swarm', or in the case of *martianer* in 3, which needs to be rewritten as *fra Mars*.

(3) [...] rundt 34 var **martianer** i opprinnelsen.

[...] about 34 have been verified to be **martian** in origin.

Direct Translation Similar to coinages, but where a coinage might be seen as a creative way to translate a term into Norwegian, where the translator is unaware of an existing or more commonly used term, the direct translations (ie. translating word-by-word) lead to clear errors, as the resulting word has a completely different meaning in Norwegian. Sometimes these translations might pass as understandable anglicisms, while at times they are nonsensical. This is especially typical for fixed expressions. In 4, the English phrase *on its own* is translated as *alene* 'alone', which does not share the same use. In 5, the English *in a big way* is translated to Norwegian *på en stor måte*, which does not make sense.

(4) Madagaskar er den klart største, og et kontinent **alene** når det gjelder dyreliv.

Madagascar is by far the biggest, and a continent **on its own** when it comes to wildlife.

(5) Araberne førte også islam til landene, og det tok **på en stor måte** i Komorene og Mayotte.

The Arabs also brought Islam to the lands, and it took i**n a big way** in the Comoros and Mayotte.

Misunderstanding Context In a few cases, a Norwegian word might be a possible translation of an English word, but the translator misunderstands the context, and translates the wrong sense of the word. This is different from the two above in that the word is actually correct, but not the correct translation in this case and context. In 6, the English *peers* is translated as NB *jevnaldrende* (lit. same-aged), a term used especially when discussing peers in primary school and similar cases. It cannot be used in the sense of professional peers, which is the intended meaning here. In 7, the understood context of the elided 'flair' has caused the translator to use the word *afrikaner*, 'African (person)', instead of the adjective. Most lexical errors fall in this category.

(6) Generelt sett kan to atferd oppstå når ledere begynner å lede sine tidligere **jevnaldrende**.

Generally speaking, two behaviors can emerge as managers begin to lead their former **peers**.

(7) [...] fordi den har mer arabisk stil enn en **afrikaner**.

[...]because it has more of an Arabic flair than of an **African**.

[7] Wikipedia:https://no.wikipedia.org/wiki/Forside, SNL: https://snl.no/, NB: https://www.nb.no/

Agreement Norwegian Bokmål exhibits agreement between some parts of speech where English does not, such as past participles, adjectives and some determiners. The authors found cases of mismatching agreement in the BM2 texts, such as "en naturlig forekommende encellede marine organisme.", where *encellede* 'single cell' and *marine* 'marine' are plural forms, while *en* 'a' and *organisme* 'organism' are singular. The correct form would be *encellet* and *marin*.

Subject-Possessor Mismatch These are cases when the translator fails to use the correct possessor. This is especially clear in Norwegian, as there is a difference in whether the grammatical subject of a sentence is the possessor or not. In 8 the third person possessor *sitt* is ungrammatical due to "nylige eksempler på [...] arbeid" being the subject of the subordinate clause, and in this case, the possessor should have been *hans* (eng. his).

(8) **Han** var også engasjert i gravering av sedler i mange land, nylige eksempler på **sitt** arbeid, inkludert statsministerportretter [...]

He was also engaged in engraving banknotes for many countries, recent examples of **his** work including the Prime Ministerial portraits [...]

Incorrect Noun Gender Several nouns have been used with the wrong grammatical gender, for example, *sexet* (neut.) 'the sex', instead of *sexen* (masc.), and *giftet* (neut.) 'the poison' instead of *giften/gifta* (masc./fem.) and *det største anskaffelsen* (neut.) 'the largest acquisition', instead of *den store anskaffelsen* (masc.). In the latter case, the mismatch is especially clear, as the noun is already declined in the masculine definite form, causing an agreement error.

The gender mismatch also extends to anaphoric pronouns, as in ***Et motbåtskip*** *av Avenger-klasse [...].* ***Den*** *er tildelt [...]*, where the first noun is neuter, while the referring pronoun is common gender. This is different from the agreement point above, in that these cases mistake the gender of the nouns themselves, not just the modifying elements.

Syntactical errors While English is syntactically close to Norwegian, there are certain constructions that when directly translated become ungrammatical or unnatural. An example is seen in 9, where in English an appositioned place name can function as an indicator of origin, while this has to be rewritten in Norwegian, for example as *jazz-spiller fra Utah* 'Jazz player from Utah', or in 10, where the subordinate clause initialized by a single past participle is very marked in Norwegian and must usually be rewritten.

(9) NBAs beslutning fulgte en **Utah Jazz-spiller** [...]

The NBA's decision followed a **Utah Jazz player** testing positive for the COVID-19 virus.

(10) **Født i Hong Kong** studerte Ma ved New York University [...]

Born in Hong Kong, Ma studied at New York University

Repetitions A final type of error is the case where multiple words in English have been translated to the same word in Norwegian. An example is seen in 11. These need to be rewritten to avoid repetition if no good synonyms can be found in Norwegian.

(11) ulovlige handlinger som dommere, **advokater, advokater og advokater** har gjort i løpet av de foregående årene.

[...] illegal actions that judges, **lawyers, solicitors and attorneys** have done during the previous years.

Minor Mistakes In addition to the issues above, there are other minor mistakes. However, one striking aspect about all of these is that they are rarely encountered with native speakers, as these are core components of Norwegian grammar, and not infrequent or rare phenomena.

4.3 Other Correction Issues

When correcting, we have focused mainly on improving the latest Bokmål translations, as these are supposed to be improvements on the earlier translations. They are mostly in a standard moderate variety, allowing for feminine definite forms of very frequent feminine nouns, as is typical in moderate BM, but varying in consistency when it comes to some less frequent words and other potentially radical forms.

In the case of many loanwords in Norwegian, both older spellings and more Norwegianized spellings are allowed in some cases. For many old loanwords, these spellings are naturalized and

one might even think about their original spellings, such as *byrå* (fr. bureau) 'office' and *sjåfør* (fr. chauffeur) 'driver', while for many more recent words, especially from English, the Norwegian alternatives can sometimes be more marked. The translators of the most recent Bokmål dataset seem to have a preference for keeping original spellings, and we have kept them. However, in the radical versions, we have used the more Norwegian versions, leading to differences such as streame vs. strømme 'to stream', container vs. konteiner 'container', etc.

4.4 Radical version

Following the correction of the BM2 version into BM3, we then convert these into radical Bokmål (BM3R).

As described above, Radical Bokmål is a subvariety of Bokmål, where radical options are chosen. These options refer to lemma varieties sanctioned by the dictionary Bokmålsordboka, while trying to stay as close as possible to the normative guidelines put forth by the association for Radical Bokmål[8]. The differences broadly fall into three categories:

Sound Correspondences Many optional forms are based on differences in diverging sound changes that are semi-regularly executed in Bokmål. These are found broadly across parts of speech, as in *melk* (MBM)/*mjølk* (RBM) and *fløte* (MBM) and *fløyte (RBM-*[9]

(12) [...] masse krem**fløte** (ikke **melke**skum) og te blir servert uten **melk**. (MBM)

[...] masse krem**fløyte** (ikke **mjølke**skum) og te blir servert uten **mjølk**. (RBM)

Another large category is morphology. In our case, this especially applies to the endings discussed above, in addition to using the deverbal nominal suffix *-ing* instead of the more MBM-coded *-else*, where these are listed as equal in the Bokmål dictionary.

(13) [...] en viktig del av **opplevelsen**. (MBM)

[...] en viktig del av **opplevinga**. (RBM)

Finally, the only syntactic change is when rewriting possessive constructions with a bare genitive 's' in MBM, which are preferred as prepositional phrases in RBM, as in 14.

(14) **Tigerens brøl** er ikke [...] (MBM)

Brølet til tigeren er ikke [...] (RBM)

[8] https://bokmal.no/

[9] Note that marking a word as MBM/RBM in this case is for convenience, but it is not the case that all radical versions are equally strong indicators of RBM as this is a continuum.

BLEU	BM1	BM2	BM3	BM3R
BM1	–	32.31	35.59	32.54
BM2	32.37	–	73.73	64.89
BM3	35.63	73.65	–	86.89
BM3R	32.56	64.79	86.84	–
chrF	**BM1**	**BM2**	**BM3**	**BM3R**
BM1	–	62.65	63.76	62.22
BM2	61.58	–	84.81	81.51
BM3	62.96	85.21	–	94.77
BM3R	61.32	81.74	94.59	–

Table 1: Similarity metrics for the various translations using BLEU and chrF (rows refer to hypotheses, columns to references).

	MADLAD		NLLB		OPUS-MT	
Ref.	BLEU	chrF	BLEU	chrF	BLEU	chrF
BM1	33.44	62.03	31.31	59.53	34.91	62.96
BM2	62.11	79.72	58.98	76.82	69.50	84.38
BM3	56.12	75.59	52.01	72.16	60.06	77.75
BM3R	49.83	72.88	46.12	69.62	53.01	74.81

Table 2: Translation scores for three Bokmål system outputs, using the four different reference translations.

5 Experiments

5.1 Distances between translations

To quantify the differences between the Bokmål translations, we compute BLEU and chrF scores between pairs, taking one as a reference and the other as the hypothesis. Table 1 presents the results.

We observe that BM1 is most different from all other translations, with differences of around 30 BLEU points and 20 chrF points. BM3 is relatively similar to BM2, which is not surprising due to the majority of corrections being done on these sentences. The highest similarities are observed between the moderate and radical BM3 translations, suggesting that lexical and morphosyntactic contexts that allow variation are relatively rare in the dataset. Both metrics follow roughly the same pattern.

5.2 Machine translation experiments

We measure the impact of the updated translations on machine translation evaluation in two experiments, using English-to-Bokmål and Bokmål-to-English MT systems, respectively.

First, we translate the English FLORES dev set to Norwegian Bokmål and evaluate the output on all four reference translations. This shows how much the evaluation scores of a single translation can vary according to the reference used. We use three MT systems to translate from English to Bokmål: MADLAD-400-3B-MT[10], NLLB-200-distilled-1.3B[11], and OPUS-MT[12]. Table 2 shows the results.

It can be seen that BM1 provides the lowest translation scores, suggesting that the reference translations are too different from the ones produced by off-the-shelf MT systems. On the other hand, the highest scores are obtained when using BM2 as a reference; this could hint to increased translationese in this dataset.

The moderate reference yields higher scores than the radical reference, which suggests that the translations produced by the three MT systems is more similar to the (more widely used and less marked) moderate variant. The radical reference impacts BLEU score slightly more than chrF score, as many moderate/radical differences occur at subword level (e.g., inflectional endings).

The three MT models provide output of similar quality, with OPUS-MT outperforming MADLAD and NLLB. Interestingly, the score differences across models are more pronounced with BM2 and BM3 than with BM1. BM1 seems therefore of limited usefulness to discriminate between different MT systems.

Second, we use the four Bokmål translations as input to Norwegian-to-English translation systems and evaluate the English outputs using the English FLORES reference. We again use MADLAD-400-3B-MT and NLLB-200-distilled-1.3B, as well as an OPUS-MT model covering the opposite translation direction[13]. The results are presented in Table 3.

[10] https://huggingface.co/google/madlad400-3b-mt

[11] https://huggingface.co/facebook/nllb-200-distilled-1.3B

[12] https://huggingface.co/Helsinki-NLP/opus-mt-tc-bible-big-deu_eng_fra_por_spa-gmq

[13] https://huggingface.co/Helsinki-NLP/opus-mt-tc-big-gmq-en

	madlad3b		nllb1.3bdist		opusmt	
Input	BLEU	chrF	BLEU	chrF	BLEU	chrF
BM1	43.69	68.02	40.48	64.73	44.10	68.26
BM2	63.98	80.68	58.97	76.82	66.66	81.94
BM3	56.18	75.24	52.19	71.87	58.38	76.44
BM3R	55.16	74.73	51.08	71.13	57.61	75.85

Table 3: Translation scores for English system outputs produced from various Bokmål inputs.

The results quite closely reflect those of the English-to-Norwegian translation, with the exception of the moderate/radical distinction, which almost has no impact on the models' capacity to translate the text to English.

We find that the initial translation (BM1) severely underestimates the models' true MT capabilities, both when used as a reference and as an input text. The updated version provided by the FLORES team (BM2) yields the highest scores, whereas the translations provided in this paper (BM3) lie in between.

An explanation for some of this effect, is that we still observe translationese tendencies in the translated texts, for example, all three systems provide the following translation in example 15 (madlad3b adds *av*, the others do not). The word (chemical) *element* is *grunnstoff* in Norwegian, and anglicisms like this might inflate the scores for the earlier translations.

(15) Du kan også ha legeringer som inneholder små mengder (av) ikke-metalliske elementer som karbon. (PRED)

You can also have alloys that include small amounts of non-metallic elements like carbon. (EN)

Det finnes også legeringer som inneholder små mengder ikke-metalliske grunnstoffer som karbon. (GOLD-MBM)[14]

6 Conclusion

Even after its initial correction, several obvious and non-native-like mistakes remained in the FLORES+ Bokmål dataset. Our attempt has corrected the most obvious mistakes, making sure that there are at least no grammatical or lexical mistakes in the dataset, without introducing excessive changes to

[14] Adding *av* is acceptable.

the work done by the professional translators. We hope that these corrections make results from these datasets more reliable.

On a more personal note, this is not the first time the authors experience problems with context and understanding coming in the way when translating datasets that are supposed to be the basis of massive-parallel datasets. We urge the creators of such original datasets to perhaps add clarifying remarks where there might be misunderstandings. Following the observation that close to 70% of all sentences in the corrected dataset contained at least one lexical or grammatical error, we recommend earlier users of the dataset to reevaluate results used on this dataset. There is also some reason to doubt the claims that all these translations were indeed done by professional translators, and we hope that future dataset creators will use the native professional communities to gain valuable feedback in these situations.

Limitations

We observe that the Nynorsk overall holds a much better quality, but that this dataset would also benefit from a round of corrections by someone qualified. We urge a native Nynorsk writer with translation experience to do a similar check of the Nynorsk data. We also acknowledge that some cases were difficult to translate due to a lack of domain knowledge, especially alongside ambiguous English original sentences, and that the focus of this effort was to remove clear errors.

Acknowledgments

This work was supported by the HPLT project which has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070350. The contents of this publication are the sole responsibility of its authors and do not necessarily reflect the opinion of the European Union.

References

Lucas Bandarkar, Davis Liang, Benjamin Muller, Mikel Artetxe, Satya Narayan Shukla, Donald Husa, Naman Goyal, Abhinandan Krishnan, Luke Zettlemoyer, and Madian Khabsa. 2024. The belebele benchmark: a parallel reading comprehension dataset in 122 language variants. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 749–775, Bangkok, Thailand. Association for Computational Linguistics.

Helge Dyvik. 2009. Å navigere i skriftspråkets rom. om normklynger i bokmål og nynorsk. *Språknytt*, 37(3):15–21.

Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato, Francisco Guzmán, and Angela Fan. 2021. The flores-101 evaluation benchmark for low-resource and multilingual machine translation.

Francisco Guzmán, Peng-Jen Chen, Myle Ott, Juan Pino, Guillaume Lample, Philipp Koehn, Vishrav Chaudhary, and Marc'Aurelio Ranzato. 2019. Two new evaluation datasets for low-resource machine translation: Nepali-english and sinhala-english.

James Cross Onur Çelebi Maha Elbayad Kenneth Heafield Kevin Heffernan Elahe Kalbassi Janice Lam Daniel Licht Jean Maillard Anna Sun Skyler Wang Guillaume Wenzek Al Youngblood Bapi Akula Loic Barrault Gabriel Mejia Gonzalez Prangthip Hansanti John Hoffman Semarley Jarrett Kaushik Ram Sadagopan Dirk Rowe Shannon Spruit Chau Tran Pierre Andrews Necip Fazil Ayan Shruti Bhosale Sergey Edunov Angela Fan Cynthia Gao Vedanuj Goswami Francisco Guzmán Philipp Koehn Alexandre Mourachko Christophe Ropers Safiyyah Saleem Holger Schwenk Jeff Wang NLLB Team, Marta R. Costa-jussà. 2022. No language left behind: Scaling human-centered machine translation.

A Dataset example

Selected examples from the corrected datasets, along with BM1 (FLORES200), BM2 (FLORES+), corrected version (BM3), radical version (BM3R) and Nynorsk (NN).

EN	BM1	BM2	BM3	BM3_RAD	NN
The truck driver, who is aged 64, was not injured in the crash.	Lastebilsjåføren på 64 år, ble ikke skadet i styrten.	Lastebilsjåføren, som er 64 år gammel, ble ikke skadet i ulykken.	Lastebilsjåføren, som er 64 år gammel, ble ikke skadet i ulykken.	Lastebilsjåføren, som er 64 år gammel, ble ikke skada i ulykka.	Lastebilsjåføren på 64 år vart ikkje skada i kollisjonen.
During his trip, Iwasaki ran into trouble on many occasions.	I løpet av reisen sin kom Iwasaki i trøbbel ved flere begivenheter.	Under reisen hans kom Iwasaki i vanskeligheter ved mange anledninger.	Under reisen sin kom Iwasaki i vanskeligheter ved mange anledninger.	Under reisa sin kom Iwasaki i vanskeligheter ved mange anledninger.	Under turen hans møtte Iwasaki på problem ved fleire høve.
In just two weeks the Americans and Free French forces had liberated southern France and were turning towards Germany.	I løpet av bare to uker hadde amerikanerne og frie franske styrker frigjort den sørlige delen av Frankrike og vendte seg mot Tyskland.	På bare to uker hadde amerikanerne og de franske franske styrkene frigjort Sør-Frankrike og vendte seg mot Tyskland.	På bare to uker hadde amerikanerne og de frie franske styrkene frigjort Sør-Frankrike og vendt seg mot Tyskland.	På bare to uker hadde amerikanerne og de frie franske styrkene frigjort Sør-Frankrike og vendt seg mot Tyskland.	På berre to veker hadde amerikanarane og sjølvstendige franske styrkar frigjort den sørlege delen av Frankrike, og var på veg mot Tyskland.