

## **Omsetjingsminne frå Nynorsk pressekontor (2011 – 2022)**

Dette korpuset inneheld omsetjingar frå bokmål til nynorsk av nyhendetekst frå Norsk telegrambyrå (NTB). Tekstene er omsette av Nynorsk pressekontor (NPK), som nyttar den såkalla Nynorskroboten til automatisk omsetjing av tekster frå bokmål til nynorsk, og korrigerer feila roboten gjer manuelt før publisering. Les meir om dette hjå Nynorsk pressekontor (<https://www.npk.no/>).

Materialet skriv seg frå perioden februar 2011 til desember 2022. Av opphavsrettslege grunnar er omsetjingseiningane randomiserte. Ei omsetjingseining svarar stort sett til eit setningspar.

Materialet finst på to ulike filformat: TMX og TSV. Begge filene inneheld dei same setningspara. Totalt inneheld korpuset 1.069.440 setningspar.

- TMX-fila er på TMX 1.4-format (ein variant av XML). I filene er kvar enkelt omsetjingseining (TU – Translation Unit) merkt med språk og tidsperiode.
- TSV-fila er TAB-separert og inneheld tre kolonner: periode – bokmålstekst – nynorsktekst.

## **Translation memories from Nynorsk News Press Agency (2011–2022)**

This translation memory contains translations of news text from Norwegian Bokmål to Norwegian Nynorsk. The texts are produced by the Norwegian News Agency (<https://www.ntb.no/about-ntb>) and translated by Nynorsk News Press Agency (<https://www.npk.no/>), who translate the texts automatically (using the so called Nynorsk Robot), then correcting the translations manually before publication.

The material in this corpus was produced in the period from February 2011 to December 2022. For copyright reasons, the translation units (in most cases corresponding to sentence pairs) have been randomized.

The corpus is provided in two formats: TMX and TSV. Both files contain the same sentence pairs. In total, the material consists of 1,069,440 translation units/sentence pairs.

- The TMX file follows the TMX 1.4 XML standard. Each sentence pair corresponds to one translation unit (TU) which is annotated with time period and language code.
- The TSV file is tab-separated and contains three columns: time period – source language (Norwegian Bokmål) – target language (Norwegian Nynorsk).