

### Norwegian-English Parallel Corpus from Public Web Sites

These two corpora contain parallelized text from the public web sites nav.no, nyinorge.no and skatteetaten.no. These web sites provide information in both Norwegian Bokmål and Nynorsk, and parts of this is translated into English. The corpora were made by Paul Meurer and Andrew Salway at the University of Bergen for the Language Bank.

The following report describes how the corpora were composed, and provides technical information:

- [https://www.nb.no/sbfil/dok/20180402\\_report.pdf](https://www.nb.no/sbfil/dok/20180402_report.pdf)

### Norwegian Bokmål—English corpus

The Norwegian Bokmål—English corpus contains material from skatteetaten.no, nav.no and nyinorge.no in three separate files. Total word/token and sentence count:

	nob	eng
token	359.401	448.717
sentence	26.771	26.693

### Norwegian Nynorsk—English corpus

The Norwegian Nynorsk—English corpus contains material from nav.no and skatteetaten.no in two separate files. Total word/token and sentence count:

	nno	eng
token	289.722	353.837
sentence	21.056	20.998

The corpora are also available in the Clarino Bergen Centre's corpus management and analysis system Corpuscle (<https://clarino.uib.no/korpuskel/>).

### Norsk--engelsk parallellkorpus frå offentlege nettstader

Desse to korpusa inneheld parallellisert tekst frå dei offentlege nettstadane skatteetaten.no, www.nav.no og www.nynorge.no. Desse nettstadane publiserer informasjon på både bokmål og nynorsk, og delar av dette er omsett til engelsk. Korpusa vart laga for Språkbanken av Paul Meurer og Andrew Salway ved Universitetet i Bergen.

Denne rapporten beskriv korleis korpusa vart laga og inneheld teknisk informasjon:

- [https://www.nb.no/sbfil/dok/20180402\\_report.pdf](https://www.nb.no/sbfil/dok/20180402_report.pdf)

### Bokmål—engelsk-korpus

Bokmål—engelsk-korpuset inneheld material frå nav.no, nynorge.no og skatteetaten.no i tre separate filer. Antal ord (token) og setningar:

	nob	eng
ord	359.401	448.717
setningar	26.771	26.693

### Nynorsk—English-korpus

Nynorsk—engelsk-korpuset inneheld material frå nav.no og skatteetaten.no i to separate filer. Antal ord (token) og setningar:

	nno	eng
ord	289.722	353.837
setningar	21.056	20.998

Korpusa er òg tilgjengelege i korpushandsamingssystemet Korpuskel ved Clarino Bergen Centre (<https://clarino.uib.no/korpuskel/>).

### **Norsk--engelsk parallellkorpus fra offentlige nettsider**

Disse to korpusene inneholder parallellisert tekst fra de offentlige nettsidene skatteetaten.no, www.nav.no og www.nynorge.no. Disse nettsidene publiserer informasjon på både bokmål og nynorsk, og deler av dette er oversatt til engelsk. Korpusene ble laget for Språkbanken av Paul Meurer og Andrew Salway ved Universitetet i Bergen.

Denne rapporten beskriver hvordan korpusene ble laget og inneholder teknisk informasjon:

- [https://www.nb.no/sbfil/dok/20180402\\_report.pdf](https://www.nb.no/sbfil/dok/20180402_report.pdf)

### **Bokmål—engelsk-korpus**

Bokmål—engelsk-korpuset inneholder materiale fra nav.no, nynorge.no og skatteetaten.no i tre separate filer. Antall ord (token) og setninger:

	<b>nob</b>	<b>eng</b>
<b>ord</b>	359.401	448.717
<b>setninger</b>	26.771	26.693

### **Nynorsk—English-korpus**

Nynorsk—engelsk-korpuset inneholder materiale fra nav.no og skatteetaten.no i to separate filer. Antall ord (token) og setninger:

	<b>nno</b>	<b>eng</b>
<b>ord</b>	289.722	353.837
<b>setninger</b>	21.056	20.998

Korpusene er også tilgjengelige i korpushåndteringssystemet Korpuskel ved Clarino Bergen Centre (<https://clarino.uib.no/korpuskel/>).