



Språkbanken

**A language technology
resource for Norwegian**

Språkbanken tilbyr digitale språkressursar til forskning og språkteknologisk utvikling. Ressursane er fritt tilgjengelege på nett, og samlinga vert stadig utvida.

Språkbanken er ei teneste til den delen av næringslivet som arbeider med utvikling av språkbasert IKT, til forskarar innanfor språkvitskap og språkteknologi, og til offentlege verksemder som utviklar elektroniske løysingar for offentlege tenester. Den skal mellom anna innehalde tekst- og talespråkskorpus, det vil seie store samlingar av tekst og tale i maskinleseleg format. Språkbanken er eit viktig nasjonalt infrastrukturtiltak som skal sikre at det blir utvikla språkteknologiske produkt tufta på norsk språk, og dermed hindre tap av bruksområde for norsk språk på teknologi-avhengige område (St.meld. 2007-2008 *Mål og mening*).

Ressursane til Språkbanken ligg tilgjengelege på nett til fri nedlasting. Døme på tilgjengelege ressursar er leksikalske databasar, mellom anna Norsk ordbank, og taledatabasar. Språkbanken vil etter kvart òg innehalde store tekstkorpus, n-gram for dei skandinaviske språka, dataverktøy til handsaming av elektronisk tekst og tale, med meir. På nettsida finn du dessutan lenkjer til tilgjengelege eksterne språkteknologiressursar.

Språkbanken har sett i gang arbeidet med ulike prosjekt, til dømes eit ordnett for bokmål og nynorsk. Ordnett er viktige ressursar for utviklinga av språkteknologiske program, som automatisk omsetjing, informasjonsattfinning og grammatikk-kontroll. Den første versjonen av ordnettet vil vere tilgjengeleg i desember 2011.

Språkbanken offers a collection of digital language resources for use in research and in the development of language technology. The resources are available free of charge, and the collection is growing continuously.

Språkbanken is a service to language-based ICT developers, to researchers within linguistics and language technology, and to public enterprises developing electronic solutions for public services. It contains corpora of written and spoken language, i.e. large collections of text and speech in machine-readable format. According to the white paper *Mål og mening* (2007-2008), *Språkbanken* represents an important language policy measure for ensuring that language technology products are developed on the basis of the Norwegian language.

All *Språkbanken's* resources are available on our website, free of charge. The available resources are lexical databases, e.g. *Norsk ordbank*, a large word/inflection database for Norwegian, and speech databases. Large text corpora, n-grams for all Scandinavian languages, and different tools will be made available shortly. Additionally, we will provide an overview of resources available elsewhere.

Språkbanken has initiated several projects, among them a project to construct wordnets for Norwegian (both *Bokmål* and *Nynorsk*). Wordnets are important resources for the development of language technology applications – for instance in machine translation, information search and grammar check – as well as in linguistic research, in particular computer linguistics. The first version of the wordnets will be made available in December 2011.

Ta kontakt for meir informasjon!

sprakbanken@nb.no
www.sprakbanken.no

Contact us for more information!

sprakbanken@nb.no
www.nb.no/spraakbanken/english