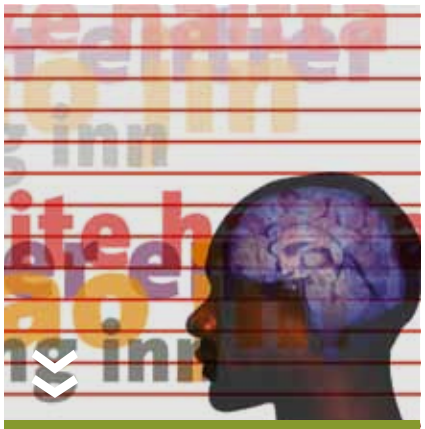
Status: Under etablering

Felles infrastruktur for norske og europeiske språkdatabaser

Med CLARINO realiseres en felles infrastruktur for norske språkdatabaser som kan kobles opp mot europeiske baser i ESFRI-prosjektet CLARIN. Dette åpner for store forskningsmuligheter innenfor språkvitenskap og er nyttig for flere fag innenfor humaniora og samfunnsvitenskap.



Prosjekt:
CLARINO – Common Language Resources and Technology Infrastructure Norway

Mer informasjon:
www.clarin.eu

Infrastrukturen er et viktig verktøy for forskning på språk som studieobjekt og som bærer av innhold. Den vil ha betydning for forskning, utvikling og undervisning innenfor en rekke felt som ikke bare omfatter språkteknologi og språkvitenskap, men også andre fagfelt. CLARINO er et viktig ledd i arbeidet med å bygge opp historiske og nåtidige elektroniske språkressurser.

Norske forskere vil få tilgang til et mangfold av norske og europeiske språkdatabaser. Data som tidligere var vanskelige å få tak i, blir dermed tilgjengelige for mange forskningsmiljøer. Norge har allerede sterke fagmiljøer innenfor språkvitenskap, og denne infrastrukturen vil kunne bidra til internasjonalt ledende forskning.

Når norske språkdatabaser kobles opp mot europeiske baser, åpner det seg store muligheter for interessante komparative studier. Man kan sammenligne språkbruk både historisk og over landegrensene. Store tekstbaser vil bli søkbare. Når man søker på norske ord, vil man kunne få svar på betydning i andre språk til ulike tider. Samtidig vil de norske dataene bli tilgjengelige for det internasjonale forskningssamfunnet.

CLARINO vil ha betydning for forskningsmiljøer og bedrifter som lager språkteknologiske systemer for å organisere kunnskap og utvikle terminologi. Slik kan man utvikle programvare beregnet på det norske markedet og sørge for at norske hjelpemidler og løsninger for universell utforming blir tilgjengelige. Miljøer som utvikler flerspråklige teknologier, som f.eks. maskinoversettelse, vil også kunne dra nytte av infrastrukturen.

Gjennom et felles kodespråk vil infrastrukturen kunne legge til rette for såkalt semantisk web. Det betyr at store datamengder fra svært mange kilder kan knyttes sammen tematisk og bli allment søkbare på nettet.

ESFRI-node
Fullt utbytte av CLARINO oppnås dersom det kobles opp mot ESFRI-prosjektet CLARIN.

Nasjonal koordinator
Universitetet i Bergen.

Partnere
Alle sentrale språkvitenskapelige miljøer i Norge er konsortiepartnere i prosjektet, og bidrar med årsverk og utstyr. Nasjonalbiblioteket og UNINETT AS vil også inngå i samarbeidet.

Finansiering
Spørsmålet om norsk deltakelse i det internasjonale prosjektet er til politisk behandling. Forskningsrådet gir 25 millioner kroner til utvikling av CLARINO, mens konsortiet bidrar med 16,3 millioner kroner i egenfinansiering. 1,6 millioner kroner kommer fra andre kilder. Dette dekker etableringen og medlemskostnader i den europeiske infrastrukturen CLARIN-ERIC. Kostnadene dekkes deretter av konsortiet.

Tidsplan
CLARINO skal etableres i løpet av fem år og stå ferdig i 2017.