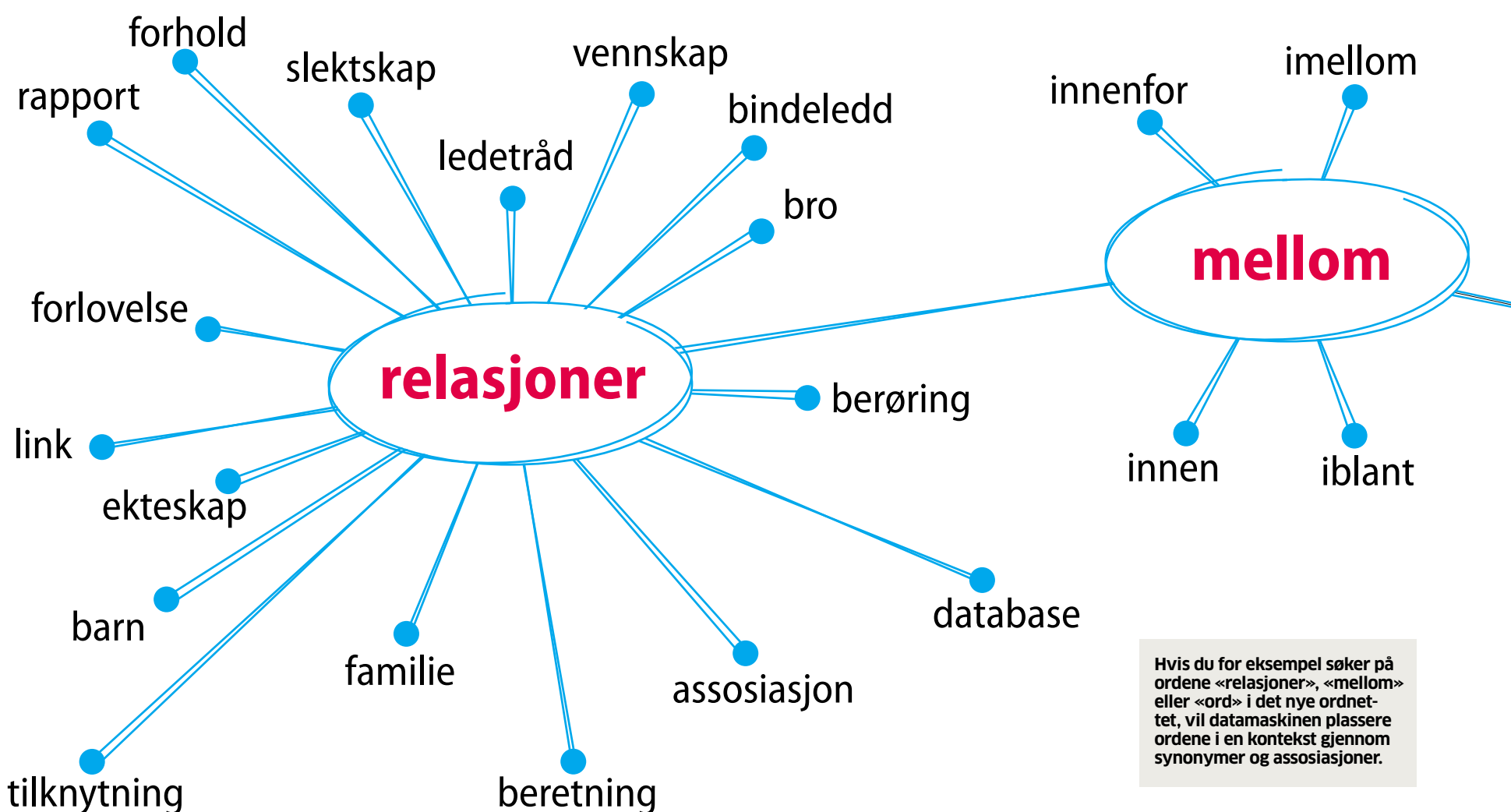


INNSIKT SPRÅKTEKNOLOGI

innsikt@aftenposten.no


GRAFIKK: SVEIN EIDE

En ny, norsk språkbank skal sikre at PC-en, mobiltelefonen og kjøleskapet forstår hva vi sier og gir oss intelligente svar.

LIV BERIT TESSEM

Norsk språk mangler en avansert programvare som gjør hverdagen lettere, blant andre for funksjonshemmede. Også norsk næringsliv vil dra nytte av bedre språkteknologi.

Den nye, norske banken veier ikke sin vekt i gull, men i ord. Den er en del av Nasjonalbiblioteket og skal utvikle grunnlaget for såkalte digitale taleverktøy. Banken skal være basen for våre fremtidige samtaler med for eksempel mobiltelefon og lesebrett. For at våre digitale verktøy på sikt skal kunne gi oss intelligente svar, lages et ordnett, en ordsamling for bokmål og nynorsk, der ordene er forbundet i nettverk av betydninger. Slike nettverk gjør norskspråklige søk på nett mer intelligente. Googler du «hund», får du ikke bare treff i søkemaskinen på dokumenter som inneholder ordet, men også det som er skrevet om schæfere, settere og terriere, om husdyr, om hvalper, tisper og om rovdyr. Det er som en ordbok, men istedenfor å være strukturert alfabetisk, er den strukturert etter ordenes betydning.

Offentlig ansvar

I over 20 år har norsk næringsliv og forskere i samarbeid med Språkrådet arbeidet med å bygge opp en egen norsk språkteknologi. Sammenlignet med store språk som engelsk, tysk og spansk er vi med knapt fem millioner norsktalende et lite lukrativt marked. Stortingsmeldingen Mål og mening argumenterte i 2008 for at utvikling av norsk språkteknologi er et offentlig ansvar. Nasjonalbiblioteket, med sitt tusen års perspektiv, fikk i oppgave å etablere Språkbanken. Etter et års forberedende arbeid, er nå en mengde digitale språkressurser på vei. I dag omfatter samlingen digital tale, tekst og leksikalske ressurser, i løpet av

Fakta

Norsk språkbank

- Er en del av Nasjonalbiblioteket.
- Ble opprettet i 2010 som et språkpolitisk tiltak etter stortingsmeldingen Mål og mening (2008)
- Inneholder en digital samling språklige råvarer og verktøy, en infrastruktur som er nødvendig for å lage språkteknologiske produkter og tjenester.
- Banken skal inneholde tekst, tale, leksikalske data, terminologi og verktøy for språkteknologisk bruk i både nynorsk og bokmål.
- Den utvides fortløpende og skal være fritt tilgjengelig på nett for næringsliv, forskere og offentlige virksomheter som arbeider med utvikling av digitale språktjenester.

det kommende året vil Språkbanken åpne slusene for å ta i mot en uendelig ordflom: I slutten av oktober utlyste Språkbanken et prosjekt for å utvikle en grunnleggende akustisk-fonetisk taledatabase på norsk.

– Det er et stort prosjekt fordi en slik taledatabase må ta hensyn til dialektvariasjonene i norsk, sier Johanne Ostad, senior rådgiver i Språkbanken. – Opplest tale må være grunnlaget for taledatabasen, og den må forstå ulike dialekter. I tillegg må den også representere spredning i alder og kjønn og trolig alminnelige talefeil i norsk.

Lite kjent

Norge er et av verdens mest digitaliserte samfunn, både arbeidsliv og fritid er basert på at elektroniske verktøy og nettverk fungerer optimalt. Det ligger store kommersielle muligheter i en bedre språkteknologi, men kostnadene med å bygge opp den nødvendige mengden av kvalitetssikrede

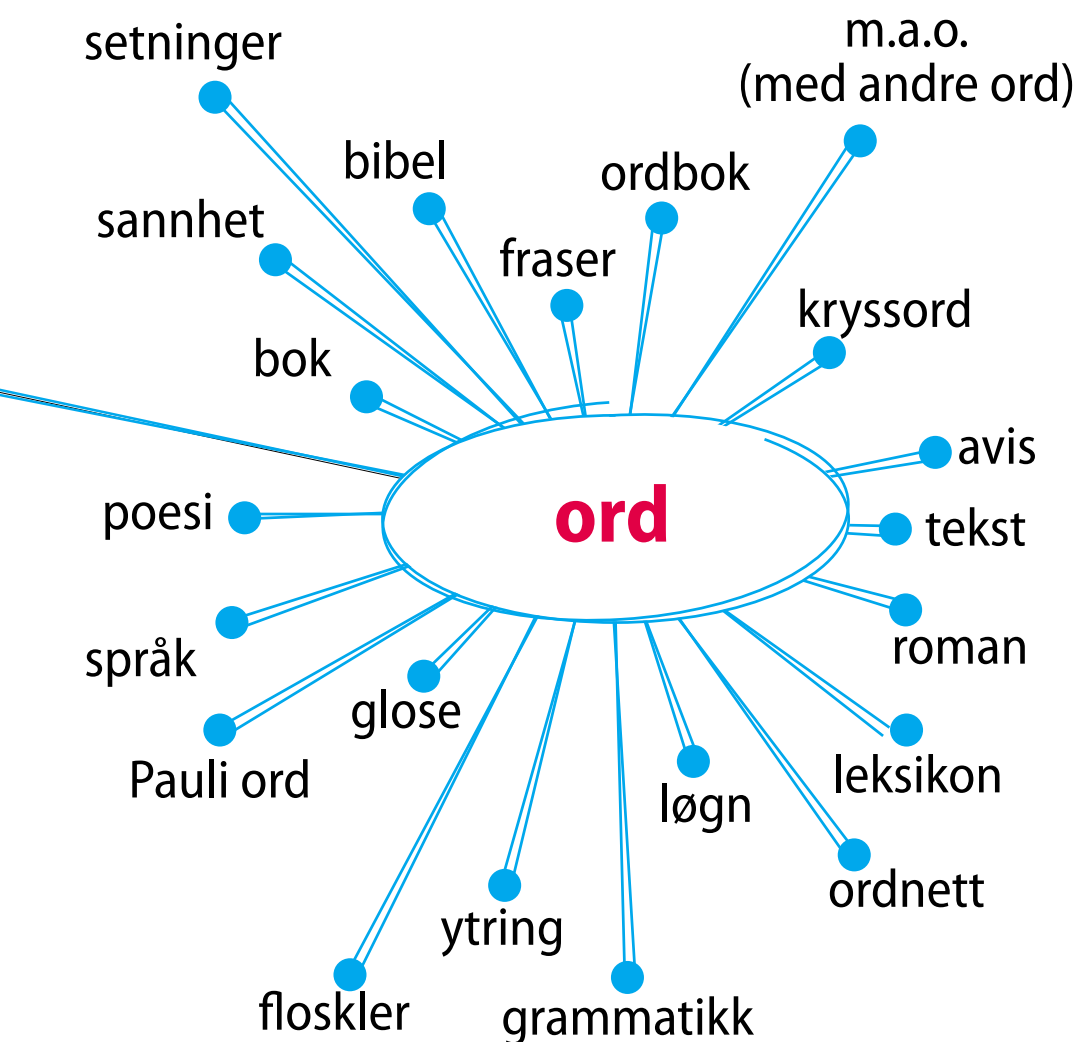
Digital flom av ord fyller Språkbanken

» Inntil nå har det for eksempel bare vært en drøm å få en datamaskin til å forstå vanlig talespråk.

Arnfinn Muruvik Vonen, direktør i Språkrådet

Lyst på mer innsikt?

Da er et abonnement på Innsikt, Aftenpostens månedsmagasin, midt i blinken. aftenposteninnsikt.no



Fakta

Ordnett

- Et nytt ordnett vil inneholde 50 000 begrep som kan inneholde flere ord og vil gjøre det lettere å gjøre intelligente søk på nettet.
- Nasjonalbiblioteket har inngått avtale med Kaldera Språkteknologi A/S for å utvikle et ordnett for bokmål og nynorsk.
- Ordnett er nyttig for utvikling av språkteknologiske program slik som stave- og grammatikkontroll, automatisk oversetting og datalingvistikk.
- Ordnett er basert på skandinavisk samarbeid og strukturen er basert på det danske DanNet
- Det største ordnettet er WordNet som er utviklet ved Princeton University i USA.

ord i banken har vært for høy til at entusiastiske lingvister og teknologer har maktet oppgaven alene. En stor forskningsrapport fra det europeiske nettverket META-NET, som arbeider for et flerspråklig informasjonssamfunn i Europa, har undersøkt den språkteknologiske statusen av 30 språk. For Norges del konkluderer forskergruppen META-NORD ved Universitetet i Bergen slik:

- Norge har et underforbruk av produkter og tjenester som kan gjøre hverdagen lettere for mange. Det er liten kjennskap til hva som finnes, kvaliteten er usikker, og det er ofte restriksjoner på bruken.
- Ressursgrunnlaget, mengden av innsamlede ord, er for liten, og to målformer gjør det ekstra kostbart
- Flere gründerbedrifter er rede til å utvikle salgbare verktøy når nye digitale språkressurser er klare til bruk.
- Det er et godt grunnlag for å satse på språkteknologi, men næringsliv, fors-

kningssektoren og myndigheter må gå sammen for å unngå at norsk taper terreng som bruksspråk på flere områder.

– Vi ser at det nå er i ferd med å skje en stor utvikling i taleteknologi knyttet til mobiltelefon og lesebrett, mange av dem har allerede i praksis droppet tastatur, og snart vil det bli helt naturlig å bruke stemmen når vi skal få teknologien til å virke, sier professor Koenraad De Smedt ved prosjektet META-NORD, Universitetet i Bergen.

Demokratisk problem

Dataprogrammereren Hege Haveraaen er en av mange som vil oppleve at mulighetene for å delta i samfunnslivet vil endre seg radikalt til det bedre med norsk språkteknologi. Hun har vært trygdet i fire år på grunn av blant annet kronisk senebetennelse i skulderen og armer. Hun har testet ut det engelske talespråksprogrammet Dragon Naturally Speaking.

– Jeg forstår engelsk godt, men min uttale er ikke god nok til at et engelskspråklig taleprogram forstår meg, sier Haveraaen.

– Dessuten er det meste av det jeg ønsker å bruke PC-en til basert på norsk, og jeg opplever at vi som av ulike årsaker ikke kan kommunisere skriftlig på internett blir stengt ute av samfunnet. Nesten all informasjon fra stat og kommune skjer digitalt, uten et godt norsk talegjennkjenningsprogram er denne informasjonen utilgjengelig for oss. Det er et alvorlig demokratisk problem for hele samfunnet. Mye kontakt med familie og venner foregår nå på sosial medier, mange blir stengt ute fra dem uten eget norsk taleprogram.

Arntfinn Muruvik Vonen, direktør i Språkrådet, er glad for at Språkbanken er etablert: – Vi er i startfasen av en språkteknologisk utvikling som vil gi oss helt nye muligheter i fremtiden. Inntil nå har det for eksempel bare vært en drøm å få en datamaskin til å forstå vanlig talespråk.

ZOOLOGI



FOTO: INTERNATIONAL MAMMOTH COMMITTEE

■ Slik ser den verdensberømte babymammuten Lyuba ut inni. Mammuten levde i arktiske områder i Russland for 42 000 år siden og lå utrolig godt bevart da hun ble oppdaget i mai 2007. Nå er hun CT-skannet i regi av International Mammoth Committee. Lyuba ble antagelig tatt av og begravet levende under et jordskred.

ASTRONOMI

Utelot kosmisk oppdagelse

I nesten et århundre har mannen som fikk Hubble-teleskopet oppkalt etter seg, astronom Edwin Hubble, fått æren for oppdagelsen av at universet ekspanderer. Men det var faktisk den belgiske kosmologen Georges Lemaitre som i 1927, to år før Hubble, først påpekte akkurat dette. Problemet var at da han selv oversatte sitt materiale til engelsk for Royal Astronomical Society utelot han selv avsnittet om oppdagelsen, skriver Nature.

ARKEOLOGI

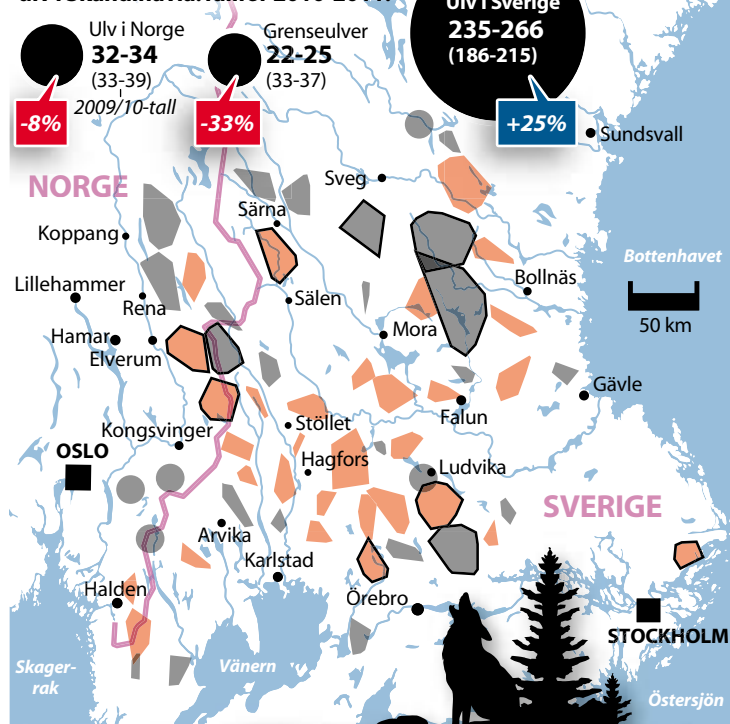
Arkeologisk drøm

Ved hjelp av luftfotoer og satellittbilder har arkeologer fra University of Leicester i England kartlagt spor etter mer enn 100 gårder, landsbyer og borglignende bygninger i Libya, skriver forskning.no. Byggverkene stammer fra mellom år 0 og 500 år før Kristus og er reist av den førislamske Garamantes-sivilisasjonen.

– Disse bosetningene er ikke blitt kartlagt under Gaddafi-regimet, sier professor David Mattingly til ScienceDaily.

Færre ulv i Norge - flere i Sverige

Totalt finnes det mellom 289 og 325 ulv i Skandinavia. Tall for 2010-2011.



VALPER: Det ble registrert tre valpekull i Norge (31 i Skandinavia). **DØDE:** 57 ulv ble drept eller funnet døde i Skandinavia i 2010-2011. 12 ulver ble drept i Norge. En av toget, to av bil, fire ble skutt ved lisensjakt, fem ved skadefelling og en i nødverge.

KILDE: Rovdata-Årsrapport Ulv 2010-2011

nyhetsgrafikk.no