

Status for arbeidet med norske språkteknologitermer

Koenraad De Smedt
University of Bergen and Unifob AKSIS

Nordterm 2009, København
12. juni 2009

Denne teksten er en oppsummering med hyperlenker av foredraget på workshopen om terminologi for språkteknologi i Norden, organisert av Torbjørg Breivik, i anledning Nordterm 2009. Målet er å gi en kort oversikt over status for arbeidet med norske språkteknologitermer.

Kanskje den eldste listen som inneholder norske språkteknologitermer er en nispråklig ordliste som ble laget ca. 1999 i regi av to europeiske nettverk under Socratesprogrammet, *Advanced computing in the humanities* (under ledelse av Koenraad De Smedt, Universitetet i Bergen) og *The landscape of future education in speech communication sciences* (under ledelse av Gerrit Bloothoof, Universiteit Utrecht). Denne listen inneholder 41 termer på ni språk: dansk, engelsk, fransk, islandsk, nederlandsk, norsk, spansk, svensk og tysk. Målet var å støtte mobilitet innen høyere utdanning i Europa innenfor språk- og taleteknologi, slik at Erasmus-studenter kunne bli kjent med terminologi som er i bruk i undervisning, selv om de fleste lærebøkene er på engelsk. Utvalget av såvel termer som språk var mer eller mindre tilfeldig og listen ble til som et resultat av dugnadsarbeid utført av nettverkene medlemmer. Denne listen er blitt kopiert på andre steder på internett.

Ca. 2002 ble en norsk og en engelsk termliste publisert av Senter for informasjon, læring og kunnskapstjenester (SILK), et initiativ av Voss kommune i samarbeid med språkteknologimiljøet på Voss. Disse to listene inneholder 90 termer med forklaringer; de er separate og inneholder ingen lenker mellom norske og engelske termer. SILK er nå avviklet og listene er ikke lenger holdt vedlike.

Deretter ble de sistnevnte to listene gjennomgått og kvalitetssikret av Gisle Andersen og Torbjørg Breivik, med innspill fra flere forskere. De engelske termene ble satt i relasjon til de norske slik at det ble laget en tospråklig liste med 82 termer på norsk og engelsk og beskrivelser på norsk. Dette arbeidet skjedde i kontekst av Norsk Dokumentasjonssenter for Språkteknologi, ett av de fem dokumentasjons-

sentre som ble opprettet i Norden i kontekst av det nordiske forskningsprogrammet innen språkteknologi (2000–2004, utvidet til 2005).

Norsk Dokumentasjonssenter for Språkteknologi har også produsert en oversettelse av terminologien som ble brukt i LT World, et tysk initiativ. Tilsvarende oversettelser ble forresten laget for alle de fem store nordiske språkene i kontekst av et initiativ i regi av NorDokNet om en tverrspråklig søkemotor for språkteknologisk informasjon på de fem nordiske dokumentasjonssentrene.

Norsk Dokumentasjonssenter for språkteknologi har samlet tre av ovennevnte lister, samt en lenke til den tverrspråklige søkemotoren, på sin nettside om terminologi. Disse tre listene er også lenket fra Termportalen, en ny norsk nettside om terminologi.

I konklusjon er det per idag flere lister med norsk språkteknologisk terminologi tilgjengelig på nettet, men utvalget av termer er lite og terminologisk informasjon er begrenset til korte beskrivende tekster. Ingen av de eksisterende listene inneholder annen terminologisk informasjon, f.eks. om subdomene, bruksområde, register, og semantiske relasjoner (synonymer, hyponymer, meronymer osv.). Spesielt hvis man vil dekke både språk- og taleteknologi kan det være nyttig å ha slik ekstrainformasjon, blant annet fordi noen termer (f.eks. 'grammar') kan ha ulike spesifikke betydninger i kontekster knyttet til ulike subdomener.

De eksisterende listene har blitt produsert i kontekst av flerspråklige initiativer, slik at det de norske termene er knyttet til oversettelser på engelsk og andre språk. Listen som er basert på LT World er oversatt til de fem store nordiske språkene og kan derfor et godt utgangspunkt for videre arbeid.

Videre arbeid med språkteknologisk terminologi i en nordisk kontekst trenger etter min mening forankring i et stabilt organ. Nye initiativer bør stimulere til bidrag fra hele fagmiljøet f.eks. gjennom et Wiki-basert opplegg for oppbygging av en termbase. En slik termbase bør være flerspråklig, utvidbar og kan med fordel bruke eksterne lenker, spesielt til artikler i Wikipedia. For at en nordisk termbase skal ha en merverdi i forhold til Wikipedia, bør ekstra informasjon om f.eks. subdomene, bruksområde og semantiske relasjoner inkluderes.