

Sture Allén
 Språkdata
 Göteborgs universitet

VAD ÄR DATALINGVISTIK?

Det som inte kunde bli av för min del i Köpenhamn kanske kan realiseras i Trondheim i stället: att inleda en diskussion om ämnesområdet. Nu kan jag också anknyta till min genomgång av frågan vid sommarens forskarkurs i Reykjavik. Här följer några förberedande synpunkter.

Efterhand som ämnet har vuxit fram under de senaste decennierna, har man haft lite olika tankar om dess innehåll. En del har velat ge det en mycket vid latitud, alltför vid enligt min mening. Det skulle då innefatta alla användningar av datamaskiner, vid vilka programmet utnyttjar någon lingvistisk princip. Andra har velat begränsa det till utveckling av lingvistiska algoritmer, vilket jag ser som alltför inskränkt. Det finns naturligtvis fler ståndpunkter.

För egen del har jag tidigt betonat att det rör sig om utveckling och tillämpning av algoritmiska metoder i syfte att vinna insikt i naturligt språk. Det är alltså fråga om ett språkvetenskapligt ämne av i princip allmän natur.

När det gäller att karakterisera området närmare, kan man ta sin utgångspunkt i begreppet språkbrukare. I det språkliga flödet gäller det för språkbrukaren att analysera och förstå språkyttringar som riktar sig till honom och att själv producera språkyttringar i förekommande fall.

Vilka resurser behöver språkbrukaren för att kunna utföra detta? Först och främst måste han ha lingvistisk förmåga. Det innebär att han känner till språkets lexikaliska och grammatiska uttrycksmedel och dess textuella grundprinciper.

Han måste också ha vad man kan kalla kognitiv förmåga. Det innebär bland annat att han kan organisera inkommande upplysningar och dra slutsatser ur dem. Det innebär exempelvis också att han kan planera de språkyttringar som skall produceras. Intressanta resultat i de här avseendena har kommit fram på området artificiell intelligens under senare år.

Vidare måste språkbrukaren ha en viss stilistisk förmåga. Det krävs att han har en uppfattning om vad som är vanligt och ovanligt i olika sammanhang, eftersom man exempelvis inte uttrycker sig likadant i ett privatbrev som i ett högtidstal.

Han måste också ofta kunna dra nytta av redan existerande texter eller textfragment för att göra sig gällande i språksituationen.

Till detta kommer att han måste ha kunskaper om världen, låt oss kalla det encyklopedisk förmåga. Två viktiga aspekter gäller huvuddragen i världens byggnad (riktningar, material osv.) och inträffade händelser i urval.

Man kan göra tankeexperimentet att placera ett programsystem på språkbrukarens plats i språksituationen. Det kommer då att kräva resurser av liknande slag som språkbrukaren utnyttjar för att kunna analysera och generera språkyttringar. Det är som jag ser det inom denna ram som datalingvistikens rör sig.

Språkvetenskapen förfogar idag inte över de redskap som krävs på något av de fyra områden jag nämnde, inte heller över de algoritmer som är en förutsättning för programsystemet. Modellen är i själva verket ett incitament till forskning på vida fält inom den angivna ramen. Det är högst ovisst om man någonsin når ända fram. De problem som skapas är emellertid av mycket stort intresse. Från språkvetenskaplig synpunkt knyter intresset sig i första hand till den lingvistiska och den stilistiska förmågan. Under överskådlig tid får man tänka sig system som strävar åt modellens håll som interaktiva.

En intressant aspekt är att datamaskinen kan användas för att bygga upp de resurser som programsystemet behöver. Detta kan ske på olika sätt. Man kan experimentera med exempelvis lingvistiska regelsystem för att utveckla dem. Man kan göra studier av enskilda verk eller stora textmängder för att få ett grepp om de faktiska språkförhållandena.

I grunden är det fråga om att språket betraktas som en process. Detta får konsekvenser på både det metodologiska och det teoretiska planet. En språklig teori framstår i detta ljus som en teori om analys och syntes av naturligt språk med hjälp av en uppsättning språkbrukarresurser. I själva verket befinner vi oss kanske i ett paradigmskifte, en övergång från en strukturorienterad till en processororienterad inriktning.