

Förord

Med Nusvensk frekvensordbok 4 lägger vi fram resultaten av sista fasen i det forskningsprojekt som lett fram till tre tidigare ordboksdelar (1970, 1971, 1975). Denna fjärde del redovisar ordleden (morfemen) och deras ordbildningsegenskaper och betydelser på fyra olika beskrivningsnivåer. Det undersökta textmaterialet är detsamma som tidigare: en miljon löpande ord ur fem morgontidningar 1965.

I korta drag har projektet haft följande förlopp. Mot bakgrund av erfarenheter från 1961/62 skedde planläggningen 1964. Jag skisserade då främst den första etappen. Huvudlinjerna för fortsättningen drogs upp under arbetet på denna första etapp. De har efterhand naturligtvis preciserats i mycket hög grad.

Arbetet under 1965 stöddes ekonomiskt av Statens humanistiska forskningsråd. Anslag från den nyinrättade Stiftelsen Riksbankens jubileumsfond gjorde det möjligt att 1966 bilda en arbetsgrupp, som fick namnet Forskningsgruppen för modern svenska. Den alldeles övervägande delen av arbetet har utförts inom denna grupp, som hela tiden har varit verksam vid Göteborgs universitet. Gruppen upplöstes formellt 1976 i och med att anslaget från jubileumsfonden upphörde. Det återstående arbetet har utförts på basis av de mera begränsade resurser som har kunnat frigöras inom institutionen för språkvetenskaplig databehandling.

Efterhand som projektarbetet fortskred, vidgades verksamheten. Nya projekt tillkom, och en fastare grund skapades. Forskningsrådet spelade härvidlag en mycket viktig roll. Vid detta inrättades 1970 en forskartjänst i språklig datateknik. Denna följdes 1972 av en personlig professur i språklig databehandling vid rådet. I samband därmed inrättades avdelningen för språklig databehandling (Språkdata) vid Göteborgs universitet. Den fick växande uppgifter och omvandlades 1977 till en självständig enhet, institutionen för språklig databehandling. Professuren överfördes 1979 till Göteborgs universitet och blev en ordinarie professur i språkvetenskaplig databehandling. Institutionens namn ändrades samtidigt till institutionen för språkvetenskaplig databehandling. Ämnet kallas också datalingvistik. Den engelska termen är computational linguistics.

Den del av verksamheten som gäller hantering av stora datorläsbara textmängder – insamling, lagring, bearbetning, tillhandahållande – och uppbyggande av en ordbank fick fast form 1975. Då inrättades Logoteket som ett serviceorgan på nationell basis. Det ingår i institutionens basorganisation.

Indirekt har denna utveckling naturligtvis kommit det aktuella projektet tillgodo. Det gäller såväl teori och metod som teknik. De tekniska förutsättningarna har förbättrats radikalt. Huvuddelen av arbetet på fjärde delen av ordboken har utförts genom interaktiv bearbetning vid textskärmsterminaler kopplade till institutionens specialutrustade dataanläggning.

Utvecklingen har samtidigt inneburit, att mina medarbetares roll har blivit än mer betydelsefull. Projektet har i hög grad främjats av deras praktiska erfarenhet och av deras bidrag till den närmare utformningen av den principiella beskrivningsmodellen.

Det analytiska förfarandet har successivt arbetats fram vid ett stort antal diskussioner inom forskningsgruppen. De efterhand framkommande resultaten har efter granskning likaså diskuterats igenom vid täta överläggningar.

Utöver att delta i den omfattande gemensamma aktiviteten har de enskilda medarbetarna givetvis utfört särskilda uppgifter. Sture Berg har som den ende heltidsanställda inom forskningsgruppen under den aktuella perioden samordnat den dagliga verksamheten och utfört största delen av det praktiska arbetet vid terminalerna. Jonas Löfström har gjort flera genomgångar av materialet, bland annat rörande avledningsmorfologin, och i stor utsträckning samarbetat med Sture Berg i fråga om konsekvenskontroller. Jerker Järborg och Bo Ralph har lämnat mycket betydande bidrag till detaljutformningen av beskrivningsmodellen och bland annat med utgångspunkt i den gemensamt fastlagda principen om kärnbetydelser och semantisk vidgning gjort en differentiering av betydelserna hos formella morfem, som sedan har legat till grund för gruppens slutbearbetning på detta plan. Christian Sjögreen har uppslagsrikt och effektivt svarat för programmeringen. I inledningsskedet och vid genereringen av sättbandet har vi dessutom kunnat utnyttja Rolf Gavares stora erfarenhet. Med det ofta krävande kontrollarbetet har vi haft hjälp av Gerd Bengtsson, Kerttu Simola och Lennart Tegner.

Tyngre bearbetningar av materialet har körts på Göteborgs datacentral för forskning och högre utbildning. Samarbetet med personalen där har varit mycket gott. Medel till körningarna har beviljats av Göteborgs universitet.

Den engelska översättningen av förordet och inledningen har granskats av Sölve Ohlander.

Som tidigare gäller naturligtvis att hela det framtagna materialet finns tillgängligt på magnetband. Här kan också nämnas att textkorporus, en initialalfabetiskt ordnad konkordans till texten och en motsvarande finalalfabetiskt ordnad konkordans är tillgängliga på mikrokort. De kan beställas från institutionen för språkvetenskaplig databehandling, Göteborgs universitet. I denna fjärde del av Nussvensk frekvensordbok har dessutom en fullständig källförteckning intagits.

Vid avslutningen av forskningsprojektet vill jag säga tack till alla dem som genom sina insatser har gjort det möjligt att nå fram. Det gäller först och främst mina medforskare genom åren – dem jag har nämnt här och Martin Gellerstam och Staffan Hellberg. Det gäller också övriga samarbetspartner och inte minst de instanser som har ställt resurser till förfogande.

Göteborg i mars 1980

Sture Allén