

17 Språkteknologiska initiativ

Den allmänna datoriseringen gör att språkteknologiska produkter och tjänster blir allt mer centrala i samhället. Utvecklingen inom talteknologin är snabb, till stor del beroende på att tillämpningar används inom telekommunikationsbranschen. Framväxten av Internet och stora dokumentdatabaser har samtidigt gett ökad betydelse åt texthantering. Frågor kring maskinöversättning och automatisk textförståelse har därmed blivit allt viktigare.

Kommittén skall följa och stimulera initiativ på språkteknologiområdet samt föreslå åtgärder som främjar översättningsverktyg och andra språktekniska hjälpmedel. I detta kapitel behandlar vi våra förslag för att främja den språkteknologiska utvecklingen.

Följande punkter om språkteknologiska initiativ ingår i vårt förslag till handlingsprogram för svenska språket:

- Arbetet med att bygga upp en nationell språkbank i form av gemensamma, svenska text- och taldatabaser samt lexikala resurser m.m. skall fortsätta.
- Forskning och utveckling inom språkteknologiområdet skall förstärkas.
- Språkvården bör som en del i arbetet för språkteknologin fortsätta med sitt arbete med språklig standardisering.
- Ett samordnande, språkteknologiskt sekretariat skall inrättas.
- Maskinöversättning för svenska skall utvecklas.

17.1 Insatser för att främja den språkteknologiska utvecklingen

17.1.1 Vad menas med språkteknologiska produkter och tjänster?

Med språkteknologi (eller språkteknik) avses all tillämpning och användning av språkkunskap vid utveckling av datorsystem som kan känna igen, förstå, tolka och generera mänskligt språk i dess olika former. Även datorbaserade hjälpmedel för att uttrycka sig bättre i tal eller skrift samt vissa verktyg som hjälper människor att tala med eller skriva till varandra via en dator är en del av språkteknologin.

Språkteknologi utgörs av både teknik (som programvara) och språkresurser (korpus) som programvaran utnyttjar, t.ex. maskinläsbara texter och ordböcker. (Se vidare bilaga 4.)

Utbudet av språkteknologiska produkter och tjänster har ökat betydligt under senare år. För det talade språket kan det handla om taligenkänning (t.ex. att genom tal ge ett kommando till telefonen) eller talproduktion/talsyntes (t.ex. konstgjort tal som läser upp ett elektroniskt brev). För det skrivna språket kan det handla om textförståelse (t.ex. skrivstöd såsom stavnings- och grammatikkontroll, automatiskt sammandrag samt kontroll av stil) eller textgenerering (t.ex. att färdiga väderleksrapporter skapas utifrån väderdata). Språkteknologi kan också handla om t.ex. maskinöversättning

och dokumenthantering.

Det språkteknologiska utvecklingsarbetet sker till stor del inom näringslivet, vid t.ex. dataföretag. Ett stort antal företag arbetar inom det språkteknologiska området i Sverige och i de övriga nordiska länderna. Det finns i dag ingen statistik över hur stor denna bransch är.

17.1.2 Myndigheter och institutioner med språkteknologisk verksamhet

Språkteknologiska frågor handläggs i dag av en rad olika aktörer inom de närings-, forsknings- och kulturpolitiska områdena. Språkteknologisk forskning och utveckling pågår vid flera universitet och högskolor. Grundutbildning i datalingsvistik och språkteknologi etc. ges vid universiteten i Göteborg, Linköping, Lund, Stockholm, Umeå och Uppsala samt högskolan i Skövde.

Göteborgs universitet har utsetts till värdhögskola för en s.k. forskarskola inom det språkteknologiska området. Verksamheten startade den 1 juli 2001 och har som mål att tillföra det svenska samhället doktorer med unik kompetens, att höja utbildningsnivån generellt i språkteknologi i Sverige samt att skapa en bred tvärvetenskaplig plattform för forskarutbildning i språkteknologi. Vid Göteborgs universitet finns även Språkbanken (se nedan). Institutionen för datavetenskap vid Linköpings universitet är en av de största i norra Europa. Man forskar inom ett brett spektrum av områden, däribland flera med anknytning till språkteknologi. Som exempel kan nämnas artificiell intelligens, informationssystem och datalingsvistik (se bl.a. Ahrenberg 1998).

Stöd till forskarinitierad grundforskning inom samtliga vetenskapsområden ges numera av Vetenskapsrådet. Målet för verksamheten är att tillförsäkra Sverige en position som ledande forskningsnation. Vetenskapsrådet skall främja tvär- och mångvetenskap, verka för förnyelse och jämställdhet samt stimulera till samverkan och kommunikation. Vetenskapsrådet har en total budget på ca 2 miljarder kronor per år för forskning och forskningsinformation. Övriga aktörer som kan nämnas är framför allt Statskontoret, Verket för innovationssystem (Vinnova) och IT-kommissionen.

Statskontorets uppgift är att utreda och utvärdera olika frågor på regeringens uppdrag samt att förnya statsförvaltningen med hjälp av IT. Språkteknologifrågor ligger inte inom det prioriterade området för Statskontoret, däremot finns anknytningar till några aktuella och tidigare projektaktiviteter, bl.a. 24-timmarsmyndigheten och olika EU-program. IT-kommissionen är en statlig kommitté som är regeringens rådgivare i IT-frågor. Den skall arbeta för att sprida information om de problem och möjligheter som utveckling och användning av informationsteknik innebär. På IT-kommissionens

sekretariat arbetar sex projektledare, en informationsansvarig samt en kanslichef. IT-kommissionens uppdrag gäller till och med den 30 juni 2003 då en slutrapport skall lämnas till regeringen. Vinnova är från och med år 2001 den statliga finansiären av behovsmotiverad forskning och utveckling inom teknik, arbetsliv, kommunikation och transporter. Organisationen har byggts upp av motsvarande verksamheter som tidigare bedrevs av bl.a. Närings- och teknikutvecklingsverket (Nutek).

Nutek inrättade år 1996 ett kompetenscentrum för talteknologi vid Kungliga tekniska högskolan, Centrum för talteknologi (CTT). CTT är en plattform för samarbete mellan företag, forskning och organisationer med intresse inom området. Samarbetet fokuseras på talbaserade dialogsystem, automatisk taligenkänning och multimodal talsyntes. I den etapp av projektet som startade den 1 juli 2001 deltar 22 företag och andra intressenter.

Nutek drev i samverkan med Humanistisk-samhällsvetenskapliga forskningsrådet (HSFR) två språkteknologiska ramprogram. Det första ramprogrammet (1990–1996) omfattade både talat och skrivet språk och var inriktat på såväl grundforskning som tillämpad forskning. Programmet resulterade i förbättrad språkteknologisk infrastruktur och till ökade kontakter mellan olika forskargrupper i Sverige och mellan universitet och industri. Det andra ramprogrammet (1997–1999) hade som mål att utveckla och förbättra metoder och tekniker för bearbetning av naturligt språk, att främja utveckling, utvärdering och spridning av verktyg och dataresurser för svenska språket, att bidra till att det kommer fram praktiskt användbara resultat i form av data, metoder, verktyg och system samt att knyta samman kompetens i tvärvetenskapliga nätverk och främja samarbete mellan olika aktörer inom forskning, utbildning, näringsliv och samhället i övrigt. Programmet satte igång ett antal stora projekt som omfattade flera forskargrupper och kompetenser. Som en uppföljning av dessa två ramprogram driver Vinnova under perioden 2001–2004 ett språkteknologiskt program för forskning, utveckling och demonstration. Programmet syftar till att öka kunskapen om hur man utvecklar IT-system baserade på förståelse av språkliga processer. Detta kan leda till bl.a. bättre verktyg för översättning, informationssökning, dokumenthantering och kommunikation. Programmet omfattade 2 miljoner kronor år 2001 och anslaget för år 2002 är 5 miljoner kronor. Språkteknologiprogrammet är därmed förhållandevis litet i förhållande till Vinnovas övriga program för att stödja forskning och utveckling.

I samarbete med IT-kommissionen och Vinnova arrangerade kommittén ett seminarium den 24 augusti 2001 om svensk språkteknologi. Seminariet, som samlade ett stort antal deltagare, hade som syfte att presentera området språkteknologi, dess möjligheter och begränsningar, att diskutera vad som bör göras för att stimulera den språkteknologiska utvecklingen i Sverige samt att diskutera

framtida organisation för statens hantering av språkteknologifrågorna.
En sammanfattning av seminariet finns i bilaga 4.

17.1.3 Språkteknologi inom nordiska och europeiska program

Nordiska program

Nordiska ministerrådet har beviljat anslag för ett femårigt nordiskt språkteknologiskt forskningsprogram. Det startade år 2000 och syftar till att ”profilere det nordiske sproglige fællesskab og sikre/forøge anvendelsen af de nordiske sprog inden for sprogteknologien, men samtidig støtte de nordiske samfund som en part af det større, omgivende lingvistiske samfund” samt att ”sikre god nordisk sprogteknologi til brugerne” (Sprogteknologi). Budgeten för programmet är 5 miljoner danska kronor per år och medlen används till nätverksbyggande, forskarkurser och pilotprojekt.

Ett nybildat nätverk om språkteknologi och informationssökning i Norden (SPINN) har bildats för att stärka forskningen kring språkliga resurser för informationsinhämtning. Nätverket har fått bidrag från det nordiska språkteknologiprogrammet. Nätverket skall verka för att de forskningsinstitut i Norden som arbetar med att etablera språkliga resurser för bl.a. informationsinhämtning, ordböcker och sökmotorer, kan utbyta erfarenheter och inleda ett samarbete.

De nordiska samarbetsministrarna tillsatte sommaren 1999 en oberoende arbetsgrupp, en s.k. vismanspanel. Vismanspanelen fick i uppdrag att analysera och ge förslag på det nordiska samarbetets framtid. I sin rapport Öppet för världens vindar – Norden 2000 konstaterar panelen bl.a. att engelskan i allt högre grad blir det språk som används inom IT. I Nordiska ministerrådets uppföljning av rapporten (Ny nordisk dagsorden, 2001) diskuteras behovet av en samnordisk utveckling av språkteknologiska verktyg såsom översättningsprogram och teknologi för att understödja användningen av nordiska språk. Där diskuteras även stöd till kontakter mellan universitet och högskolor kring forskning med syftet att utveckla specialkunskap i Norden på det IT-baserade språkteknologiska området.

EU-program

Språkteknologi och mångspråkighet har varit ett område för olika typer av insatser och program också inom EU. Stöd till språkteknologisk forskning och utveckling lämnas inom de olika forskningsramprogrammen. I det nu pågående femte ramprogrammet har insatser för språkteknologin fått stort utrymme inom temat Informationssamhället. Särskild tonvikt läggs på flerspråkig kommunikation, naturlig interaktivitet och tvärspråklig informationshantering. EU har dessutom haft särskilda program för att främja flerspråkigheten. Programmet MLAP (Multilingual Action Plan) löpte åren 1994–96. Programmet MLIS (Multilingual Information

Society) genomfördes under åren 1996–99. Målen för MLIS var att höja medvetenheten om och stimulera tillhandahållandet av flerspråkiga tjänster genom att använda språkteknologier, resurser och standarder, skapa goda förhållanden för utvecklingen av språkindustrier samt minska kostnaderna för informationsöverföring mellan språk, särskilt för små och medelstora företag och att bidra till den språkliga mångfalden.

Insatser för svenskspråkigt innehåll på nätet

Inom den svenska IT-politiken har mycket hittills varit inriktat på teknik, till mindre del på vad nätet innehåller. En av de stora uppgifterna blir att fylla bredbanden med ett viktigt och attraktivt innehåll på svenska språket. Svenska medier och kulturinstitutioner såsom bibliotek, museer och arkiv har stora mängder material som i varierande grad är tillgängligt via Internet. Från statens sida har satsningar på detta gjorts vid flera tillfällen, bl.a. genom satsningen på Kulturnät Sverige, Sesamprojektet och Kulturarvs-IT.

I programmet för det digitala innehållet, det s.k. eContent-programmet, ingår EU:s insatser för att stimulera europeiskt innehåll på nätet. I detta program betonas just kopplingen mellan språk, kultur och innehåll. Det är ett flerårigt gemenskapsprogram för stimulans och utnyttjande av europeiskt digitalt innehåll på det globala nätet samt för främjande av den språkliga mångfalden i informationssamhället. EU-programmet har både kultur- och näringspolitiska förtecken. Det syftar bl.a. till att främja den språkliga och kulturella mångfalden i informationssamhället samt till att öka exportmöjligheterna för innehållsföretag genom kulturell och språklig anpassning. Programmet omfattar åren 2001–2004.

17.1.4 Språkresurser som hittills byggts upp i Sverige

Språkteknologin är i hög grad avhängig av tillgången på stora mängder språkdata. Språkresurserna består av både inspelat tal, texter och lexikala resurser som ordböcker. Insamling av språkdata är i många fall dyrt och den svenska marknaden är liten. I Sverige finns ändå relativt mycket material och olika försök har gjorts att samla detta för att underlätta för brukarna.

Språkbanken är en del av Institutionen för svenska språket vid Göteborgs universitet. Vid denna bygger man upp och ger tillgång till språkliga referensmaterial i form av texter och lexikaliska data. Riksdagen fattade 1975 beslut om att inrätta en språkbank med nationell inriktning ("Logoteket"). För ändamålet inrättades två tjänster och särskilda medel avsattes (prop. 1975:1, bil. 10). Öronmärkningen av medel till Språkbanken har numera upphört. Avdelningen finansieras i dag av Humanistiska fakulteten och är knuten till Institutionen för svenska språket.

Språkbanken har erfarenhet av att tillhandahålla material på

Internet. Texter av olika slag och från olika tider finns tillgängliga via Internet (<http://spraakbanken.gu.se/>) i en form där varje ord presenteras i sitt sammanhang (s.k. konkordans). Materialet omfattar texter på ca 75 miljoner ord, från nutida texter från dagspress och skönlitteratur till äldre material som Strindbergs produktion och medeltida rimkrönikor. Texterna är åtkomliga gratis över nätet i konkordansform och antalet sökningar ligger på ca 5 000–6 000 per dag.

Arbete pågår för att göra språkbankens texter lättare åtkomliga för en mer systematisk sökning efter språkliga kategorier. Denna typ av sökning förutsätter ordklassmärkning av texterna. En sådan korpus finns redan tillgänglig via Språkbanken (det s.k. Parolematerialet på ca 25 miljoner ord).

Språkbankens referensmaterial har spelat en viktig roll i lexikografiska och lexikologiska projekt, exempelvis Nationalencyklopedins ordbok (1995) och Svenska Akademiens ordlista.

Dessutom har Språkbanken medverkat i EU-projekt med inriktning på både uppbyggnad av systematiskt insamlade texter (korpuser), som i Parole-projektet och i dess lexikaliska fortsättning Simple-projektet, där ett bestämt ordmaterial har klassificerats med avseende på betydelse. I båda fallen har arbetet skett i samarbete med ett tiotal europeiska länder. Bland andra europeiska projekt kan nämnas uppbyggnaden av en terminologisk databas med drogtermer, Aventinus.

Korpusuppbyggnad omfattar också länkning (mening för mening) av översättningar mellan svenska och främmande språk. Sådana länkade, parallella texter är viktiga underlag för forskning och utbildning inom översättningsområdet. Samarbete pågår inom Humanistiska fakulteten i Göteborg för att bygga upp sådana parallella texter till flera europeiska språk.

I detta sammanhang kan också Parallellkorpusprojektet vid universiteten i Linköping, Uppsala och Göteborg (PLUG) nämnas.

Det syftade till utveckling, evaluering och tillämpning av program för att utvinna översättningsdata ur parallella texter med svenska som källspråk eller målspråk. Med översättningsdata avses här översättningar av ord, fraser, andra meningssegment och hela meningar. De primära tillämpningarna är maskinöversättning och datorstödd översättning, översättningsdatabaser, flerspråkiga webblexikon och översättarutbildning.

Lexikaliska databaser ingår i Språkbanken och några av dessa är fritt tillgängliga över nätet. Först och främst gäller det Svenska Akademiens ordbok, som medger ordsökningar baserade på bl.a. kategorier som tid och långivande språk. Andra fritt tillgängliga källor är Söderwalls Ordbok öfver svenska medeltidsspråket, Samhällstermer på invandrar språk (polska, serbokroatiska, turkiska) och den tidigare nämnda Aventinus-databasen över drogtermer.

Bland övriga korpusar och textarkiv kan nämnas t.ex. Affärsvärldens tidningsdatabas, Runebergprojektet vid Linköpings universitet och Mediearkivet. Inom talområdet finns ett stort korpusmaterial vid Institutionen för lingvistik vid Göteborgs universitet (Göteborg spoken language corpora). Andra samlingar finns vid CTT vid Kungliga Tekniska Högskolan i Stockholm.

17.1.5 Språklig standardisering och terminologi behöver utvecklas mer

IT och inte minst Internet har inneburit stora förändringar av hur vi hanterar informationsutbyte och kommunikation. Fokus kommer alltmer att ligga på informationsinnehåll och på att göra information tillgänglig för nya tjänster och funktioner. Det handlar om tillgänglighet, publicering i många olika medier samtidigt och maskinläsbarhet. Teknik räcker dock inte för att bygga framtidens informationsstrukturer. Om olika typer av verksamheter och affärsprocesser skall kunna dra nytta av tekniken krävs också väldefinierade terminologier, vilket bl.a. Terminologicentrum TNC har pekat på (Bidragsansökan för år 2002).

Även IT-kommissionen och Statskontoret har uppmärksammat behovet av att utveckla en enhetlig terminologi. IT-kommissionen pekar i Hur blir en ny infostruktur motorn i e-Sverige? (SOU 2000:123) på hur viktigt det är att utveckla tillgången till information i samhället. En förutsättning för att hitta rätt information är att det finns en märkning av informationen. Ett väsentligt arbete är att beskriva hur märkningen skall göras och hur översättning mellan olika scheman av begrepp och termer för olika ändamål kan ske. För att man skall kunna hämta uppgifter automatiskt och för att informationen skall kunna flyta fritt mellan olika tillämpningar bör informationen vara maskinläsbar. Det behövs enligt IT-kommissionen ett standardiseringsarbete som gör digital information maskinläsbar och maskinbegriplig. Vidare är enhetliga begrepp och terminologier en väsentlig förutsättning för t.ex. handel. Det är viktigt med ett klart och entydigt språk för att undvika missförstånd mellan köpare och säljare.

TNC menar att man har relevant kompetens och erfarenhet för att som en aktiv part medverka i de olika utvecklingsinsatser som IT-kommissionen föreslår (Bidragsansökan för år 2002). IT-kommissionen har i en skrivelse till regeringen (Om behovet av insatser för utvecklingen av samhällets informationsstruktur, 2001) föreslagit att Statskontoret bör ges i uppdrag att bereda formerna för hur ett strategiskt utvecklingsarbete kan ske samt konkret samordna insatserna för standardisering, terminologi och sökfunktioner/metadata.

17.1.6 Det norska exemplet

Den norska regeringen behandlade språkteknologiska frågor i propositionen Målbruk i offentlig teneste (Stortingsmelding nr 13, 1997–98). I propositionen föreslås bl.a. att Norsk språkråd byggs ut till att också omfatta ett sekretariat för språkteknik: ”Dette er tenkt som eit verkemiddel for å sikre at norsk språk i framtida kan tene som eit fullverdig uttrykksmiddel også i arbeidet med avansert teknologi. Dette er viktig for å førebyggje såkalla domenetap for norsk språk.” Språkrådet har kunnat anställa två personer med uppgift att fungera som en knutpunkt för språkvården, forskningen och näringslivet.

Norsk Språkråd lade år 2001 fram ett förslag till handlingsprogram för det norska språket och IT (Handlingsprogram for norsk språk og IKT, 2001). I handlingsprogrammet lämnas förslag på olika mål, bl.a. att norska språket skall användas inom språkteknologin och dess produkter och tjänster i Norge. Alla med norska som modersmål skall ha lika tillgång till norskspråkiga produkter och tjänster inom IT-området. Norsk språkteknologi skall utvecklas på grundval av forskning och kunskap om talat och skrivet språk samt på nationell teknikkompetens. Det norska språket i programvaror och på norska webbsidor skall bli bättre. Bland de viktigaste åtgärdsförslagen är att upprätta en norsk språkbank för att samla, tillrättalägga och tillgängliggöra gemensamma norskspråkiga resurser för text och tal. Nationella forsknings- och utbildningsprogram inom ämnesområdena ”språkteknologi” och ”norska språket och IT” skall etableras och allianser mellan näringsliv och forskning för att utveckla språkteknologiska produkter skall främjas. Ett sektorsövergripande program/organ skall upprättas för att samordna nationella åtgärder inom språk, kultur och IT. Reglerna för offentlig upphandling skall ändras för språket i programvaror. Ett nordiskt center för språk och språkteknologi bör inrättas. Norsk IT-terminologi bör samlas, utvecklas och tillgängliggöras. Utvecklingen av elektroniska utbildningsprogram bör prioriteras och stödjas.

17.1.7 Bedömningar och förslag

Våra förslag:

- Arbetet med att bygga upp en nationell språkbank i form av gemensamma, svenska text- och taldata-baser samt lexikala resurser m.m. skall fortsätta.
- Forskning och utveckling inom språkteknologiområdet skall förstärkas.
- Ett samordnande, språkteknologiskt sekretariat skall inrättas.

Vår bedömning:

- Språkvården bör som en del i arbetet för språkteknologin fortsätta

med sitt arbete med språklig standardisering.

Utvecklingen inom språkteknologin innebär både möjligheter och risker för svenska språket. Vi har i kapitel 5 pekat på att det finns risker för att den språkteknologiska utvecklingen kan leda till ytterligare domämförluster för det svenska språket. I detta kapitel vill vi däremot betona de möjligheter som vi menar att den språkteknologiska utvecklingen innebär. Sammanfattningsvis gör vi bedömningen att språkteknologins möjligheter är stora, men att det är viktigt att från statens sida stödja utvecklingen av svenska språkresurser genom en samlad insats för att undvika framtida domämförluster inom området.

Från språk- och kulturpolitisk utgångspunkt kan utvecklingen av språkteknologiska produkter och tjänster ses som ett medel för att behålla och stärka den språkliga och kulturella mångfalden i Europa. Genom en fortsatt utveckling av språkteknologiska produkter och tjänster kan man stärka de små språkens ställning och ge alla likvärdig tillgång till information, oavsett språk. Maskinöversättning kan underlätta för oss svensktalande att på sikt kunna tala och skriva på vårt eget språk och ändå kunna kommunicera med personer med annat språk och få tillgång till information i andra länder. Om maskinöversättning utvecklas även för det svenska språket kan det bli möjligt för personer som inte fullt ut behärskar främmande språk att både få och ge information på det globala nätet. Med maskinöversättningar kan svenska språket få en större roll i informationssamhället. Från förvaltningspolitisk synpunkt är utvecklingen inom språkteknologin viktig, inte minst när det gäller översättningsarbete inom EU och utvecklingen av 24-timmarsmyndigheten.

Från näringspolitisk synpunkt är språkteknologi en framtidsbransch med industriell potential. I dagens läge står dock flera av de språkteknologiska företagen inför svåra lönsamhetsproblem. Att det finns en inhemsk språkteknologisk industri är av stor vikt för utveckling och anpassning till svenska förhållanden. Det kan antas att internationella aktörer har ett mindre intresse av att utveckla och anpassa produkter och tjänster för svenska förhållanden. Språkteknologin kan även användas för att t.ex. genom maskinöversättningar minska konkurrensnackdelar för svenska företag gentemot engelskspråkiga länder. I kapitel 14 behandlar vi språkteknologins stora betydelse för utvecklingen av nya handikapphjälpmedel.

Staten bör ta ansvar för språkteknologiskt utvecklingsarbete

I det språkteknologiska utvecklingsarbetet behövs ett nära samarbete mellan näringsliv, forskning och myndigheter. Det är därvid nödvändigt att definiera de olika parternas roller. Enligt vår bedömning är följande ett statligt åtagande:

att se till att språkresurser på svenska – både i skrift och tal – samlas in, lagras, bearbetas och tillhandahålls i en nationell språkbank, bl.a. för att underlätta framtagandet av svenskspråkiga

produkter och tjänster, men också för att kunna utnyttjas av forskningen, myndigheter och den språkintresserade allmänheten

- att ge stöd till svenskspråkig språkteknologisk forskning och utveckling, både den grundforskning som äger rum vid universiteten och högskolorna och som får stöd av bl.a. Vetenskapsrådet och den tillämpade forskning som bl.a. Vinnova stöder
- att samordna och på olika sätt stödja utvecklingsarbetet, bl.a. genom att ta initiativ till samarbete mellan näringsliv, forskning och myndigheter. Särskilt bör utvecklingen av avancerade maskinöversättningssystem för svenska språket främjas. Produktutveckling m.m. bör dock primärt lösas på marknadens villkor. Det bör vara företag inom det språkteknologiska området som med hjälp av de språkresurser och den forskning som staten finansierar försörjer den svenska marknaden med svenskspråkiga produkter.

Arbetet med att bygga upp en nationell språkbank skall fortsätta

En av de viktigaste åtgärderna för att stödja svenskspråkig språkteknologi är att stödja uppbyggnaden av svenska språkresurser.

Enligt vår mening är språket ett nationellt intresse, och likaså är det ett nationellt intresse att svenska språkresurser samlas in, lagras, bearbetas och tillhandahålls. Det är därför ett statligt åtagande att säkerställa att så sker. Språkbanken vid Göteborgs universitet har sedan 1975 i uppgift att samla in, lagra, bearbeta och tillhandahålla elektroniskt läsbara texter. Vi föreslår att arbetet med att bygga upp en nationell språkbank i form av gemensamma, svenska text- och taldataresurser samt lexikala resurser m.m. skall fortsätta. Syftet bör vara att göra språkresurserna allmänt tillgängliga för så många som möjligt.

En nationell språkbank bör ses som en viktig del i den infrastruktur som byggs upp för informationstekniken. Detta är enligt vår mening en mer effektiv åtgärd än att satsa på t.ex. statliga bidrag till utveckling av enskilda produkter. Utvecklingsarbetet bör inriktas på att göra denna nationella språkbank så komplett som möjligt (tal- och textdatabaser för nutida svenska, parallell- eller översättningskorporusar samt ordböcker, termlistor, uttalsordböcker och skrivregler m.m.).

Den nationella språkbanken skall alltså stödja utvecklingen av en svenskspråkig språkteknologi. För att främja en sådan utveckling är det viktigt att språkresurser hålls allmänt tillgängliga för att kunna återanvändas och vidareutvecklas av såväl forskare som kommersiella projekt. Det är viktigt att denna åtgärd är konkurrensneutral och kommer alla språkteknologiska företag och forskare till godo. Tillgången till den nationella språkbanken måste därför göras så bred som möjligt. Databaserna bör vara i första hand gratis och i andra hand billiga att använda för så många som möjligt. Vi anser att detta är en viktig förutsättning för att språkteknologin skall kunna utvecklas även på svenska.

Den nationella språkbanken skall emellertid inte bara vända sig till språkteknologiska företag och forskare, utan också vara en webbplats för svenska språkresurser på det globala nätet som riktar sig till den breda allmänheten. Det är språkpolitiskt viktigt att den nationella språkbanken får ett folkbildande uppdrag och kan tillgängliggöra material av intresse för alla språkbrukare (se vidare kapitel 19). En annan viktig avnämning är den offentliga förvaltningen.

Arbetet med att vidareutveckla en nationell språkbank underlättas av att det i Sverige redan finns relativt mycket material. Huvudsakligen finns detta vid Språkbanken vid Göteborgs universitet. I detta sammanhang är det naturligtvis också en stor fördel att svenska språket är ett så väldokumenterat språk. Det finns dock behov av ytterligare korpusar, t.ex. stora talspråkskorpusar med regionala varieteter, brytning etc. som underlag för språkteknologin (röststyrd teknik etc.).

Den nationella språkbanken bör ges funktionen av en portal för svenska språkresurser, vars uppgift blir att tillhandahålla de olika korpusarna m.m. Detta innebär att insamlingen, lagringen och bearbetningen kan ske hos olika aktörer. Ansvaret för att bygga upp och sedan driva den nationella språkbanken kan förläggas till olika aktörer. Vi har i våra överväganden utgått från följande. En första utgångspunkt är att Språkbanken vid Göteborgs universitet redan fungerar som en nationell språkbank och att där finns ett mycket omfattande material som måste utgöra en av grundstenarna till en vidareutvecklad nationell språkbank. Språkdata vid Institutionen för svenska språket är en central aktör och den kompetens som finns där är en resurs som måste tillvaratas på bästa sätt.

Förutom vid Språkbanken finns olika korpusar vid andra forskningsinstitutioner. Det är viktigt att en bred repertoar av korpusar tillgängliggörs genom den nationella språkbanken. En andra utgångspunkt är att den vidareutvecklade nationella språkbanken måste få bred tillgänglighet. Till kommittén har det framförts, bl.a. av representanter för språkteknologisk verksamhet, att det i dag finns tillgänglighetsproblem vad gäller främst lexikala databaser. Det är ett nationellt intresse att svenska språkresurser på ett enkelt sätt kommer forskare, företag och andra i hela landet till godo.

En tredje utgångspunkt är att det är en fördel om den som skall ansvara för den nationella språkbanken har bred kompetens när det gäller både språk- och teknikfrågor samt har goda kontakter med både forskarvärlden, myndigheter och det språkteknologiska näringslivet. Eftersom språkresurserna bör komma alla till godo är det vår bedömning att den nationella språkbanken bör hanteras av en statlig myndighet eller statligt finansierad aktör, utan vinstintressen. Den nationella språkbanken innebär bl.a. att länkar till redan befintliga svenska språkresurser samlas och koordineras i en portal för

svenska språkresurser. Det ingår i den nationella språkbankens uppgifter att säkerställa att alla får tillgång till dessa. I kapitlet om språkvårdens framtida organisation återkommer vi till frågan om vem som skall ges i uppdrag att ansvara för den fortsatta uppbyggnaden och driften av den nationella språkbanken.

Ett konkret nordiskt samarbete i språkbanksfrågor bör initieras

Under utredningsarbetet har det framkommit att det finns behov av nordiskt samarbete kring språkteknologiskt utvecklingsarbete. Vi anser att den fortsatta uppbyggnaden och samordningen av tal- och textdatabaser bör kunna ske inom ramen för nordiskt samarbete, både vad gäller resurser och metoder/teknik. Denna åsikt har framförts även från norsk sida (Handlingsplan for norsk språk og IKT, 2001). Norsk Språkråd menar att de nordiska länderna bör kunna ”ha gjensidig nytte av deler av de ressursene og verktøyene som utvikles i en slik sammenheng”. Det bör vara en uppgift för den som skall bygga upp och administrera den nationella språkbanken att initiera ett konkret nordiskt samarbete i dessa frågor.

Rättighetsfrågor bör beaktas i resursberäkningar

Vi vill peka på att korpusar kan vara mycket dyra att ta fram. Detta gäller särskilt talspråskorpusar. En del av kostnaderna för en korpus är förknippade med rättigheter. Detta kan gälla t.ex. förlag som har rättigheter till ordlistor eller forskningsinstitutioner som tagit fram vissa material. Rätten till de gemensamma språkresurserna är en viktig fråga som måste lösas för att Sverige skall få en väl fungerande språkteknologisk infrastruktur. Enligt vår bedömning måste texter som läggs in i den nationella språkbanken kunna utnyttjas fritt. Den som ges uppgiften att bygga upp och administrera den nationella språkbanken måste därför också hantera rättighetsfrågorna. Det är viktigt att beakta detta när de ekonomiska resurserna för detta förslag beräknas (se kapitel 21).

Språkvården bör fortsätta med sitt arbete med språklig standardisering

En annan viktig förutsättning för utvecklingen inom IT och språkteknologin är ett fortsatt arbete med språklig standardisering. Detta kan sägas innefatta två aspekter: dels arbetet med terminologier, dels en mer allmän normbestämning. I den fortsatta IT-utvecklingen kommer betydelsen av klara och entydiga termer att öka, både i det utvecklingsarbete som pågår inom IT-branschen och i gränssnittet mellan användare och maskin/nät.

För att användningen av Internet skall bli enklare är det viktigt att språkvården blir aktiv när det gäller att standardisera bruket av svenskspråkiga ord och uttryck. Språkvården bör enligt vår bedömning fortsatt uppmärksamma dessa frågor.

Termdatabaser bör tillgängliggöras genom språkbanken

Vi har i bl.a. kapitel 6 redovisat våra bedömningar och förslag om statens ansvar för terminologiarbetet och behovet av olika termdatabaser,

inte minst flerspråkiga sådana. Det har i utredningsarbetet framkommit att det finns behov av standardiserade och kvalitetssäkrade terminologidatabaser, inte minst för att Internet skall fungera även på svenska. Terminologicentrum TNC har framfört idén om att inrätta en rikstermbank för att tillgängliggöra termer från t.ex. myndigheter för allmänhet, översättare, tjänstemän m.fl.

Vi delar den bedömningen och menar att språkvårdens terminologiarbete bör samordnas och inriktas på att ta fram en rikstermbank. En sådan rikstermbank bör tillgängliggöras genom den nationella språkbanken. Även de flerspråkiga termbaser som vi föreslår i kapitel 6 bör tillgängliggöras genom språkbanken.

Forskning och utveckling inom språkteknologiområdet skall förstärkas
Vi anser att språkteknologi är ett område som har stor potential att vidareutvecklas. Behovet av forskning och utveckling inom området är fortsatt stort, medan de statliga resurser som avsatts för området är mycket begränsade.

Från Vetenskapsrådets sida ges pengar till grundforskning. Vi föreslår att Vetenskapsrådet ges i uppdrag att utlysa medel för riktad forskning inom språkteknologiområdet. Detta skulle exempelvis kunna pågå under fem år.

Inom området informations- och kommunikationsteknik har Vinnova ett program för språkteknologi. De program som Vinnova/Nutek har ansvarat för under senare år har haft stor betydelse för den språkteknologiska utvecklingen i Sverige. Programmen omfattar dock förhållandevis små belopp och vi anser att mer statliga resurser borde satsas på detta område. Vi föreslår därför att resurserna för forskning och utveckling inom språkteknologiområdet ökar.

Vi föreslår vidare att ansvaret för framtida program för att stödja forskning inom området och därtill hörande medel behålls vid Vinnova, men att det språkteknologiska sekretariat som föreslås nedan ges ett avgörande inflytande över hur medelen skall fördelas. Det är naturligt att stödet till forskning och utveckling inom området hanteras administrativt inom Vinnova som har kompetens och erfarenhet inom området, men att hänsyn tas till statens övriga insatser för språkteknologin när beslut fattas. Detta innebär att ett språkteknologiskt sekretariat tillsammans med Vinnova skall fatta beslut om hur medlen skall fördelas. Vi återkommer till denna fråga i samband med språkvårdens framtida organisation.

För att forskning skall bli bra måste det finnas en god grundutbildning. Det är därför viktigt att det finns resurser för språkteknologisk utbildning. Språkteknologi har en tvärvetenskaplig karaktär och kan därför ha svårt att hävda sig i konkurrensen med områden som ses som mer centrala inom olika fakulteter. Det är

viktigt att se till att språkteknologiområdet inte faller mellan stolarna. De nordiska länderna ligger generellt sett långt framme på ITområdet. Det har dock ägt rum en snabb utveckling inom språkteknologin för de större europeiska språken som har fått tillgång till avancerade informationsbehandlingsverktyg, medan de nordiska språken saknar sådana verktyg och därmed riskerar att halka efter på informationsökningens område. Orsaken är att språken i Norden är relativt sett små och marknaden därmed för liten. För att även svenska skall ingå i flerspråkiga produkter och tjänster är det viktigt att svensk forskning och utveckling deltar i det internationella samarbetet, inte minst inom Norden och EU. Det bör finnas vinster – såväl ekonomiska som verksamhetsmässiga – att göra genom att vidareutveckla i första hand det nordiska samarbetet.

Vi föreslår därför att nordiskt samarbete liksom samarbete mellan forskning och näringsliv skall främjas i de framtida stödprogrammen. Forsknings- och utvecklingsprojekt bör kunna genomföras i samverkan med industrin. Vidare bör svenskt deltagande i programmet för det digitala innehållet och andra EU-program uppmuntras. Det bör vara en uppgift för det språkteknologiska sekretariat som föreslås nedan att verka för detta.

Ett samordnande, språkteknologiskt sekretariat skall inrättas

Det behövs ett nära samarbete mellan forskning, högre utbildning, näringsliv och myndigheter för att föra samman den teknik samt de allmänna språkkunskaper och språkresurser i form av både svensk skrift och svenskt tal som behövs för att utveckla språkteknologin. Språkteknologi berör olika branscher med bakgrund i olika forskningsdiscipliner. Det statliga ansvaret för området är splittrat mellan olika departement, myndigheter och institutioner. Från branschen efterfrågas ett samordnande organ.

Nutek har i en rapport (Maskinöversättning, 1999) till regeringen föreslagit att det bör utarbetas en långsiktig strategi för främjande av svensk språkteknologi. En myndighet/organisation bör utpekas som koordinator för strategiarbetet och ha som uppgift att följa utvecklingen inom området, bygga upp och förvalta resurser för svenska språket samt sprida information och stödja erfarenhetsnätverk. Svenska språknämnden har i sin principskiss om svensk språkvård och Språknämndens framtida organisation pekat på behovet av ett språkteknologiskt språkråd med representanter för branschen, forskning och myndigheter.

Vi delar Nuteks och Språknämndens bedömning och föreslår att staten blir mer aktiv inom det språkteknologiska området genom att inrätta ett samordnande, språkteknologiskt sekretariat. Syftet skall vara att stärka statens insatser och samordna dessa med andra insatser inom det språkteknologiska området.

Ett samordnande språkteknologiskt sekretariat bör genomföra åtgärder, följa utvecklingen samt sprida information. Andra uppgifter

för ett sådant sekretariat är att t.ex. medverka i beslutsfattandet om statens direkta stöd till forskning och utveckling inom området (som i dag handläggs av Vinnova). Ett språkteknologiskt sekretariat bör även ha överblicken över tillgången på svenska språkresurser och följa t.ex. utvecklingen av språkteknologiska hjälpmedel för funktionshindrade.

Sekretariatet bör även ha en konsumentupplysande funktion och kunna fungera som ett rådgivande organ gentemot framför allt offentliga institutioner. De språkteknologiska produkter och tjänster som finns t.ex. för översättning till och från svenska har ännu som regel mycket specifika användningsområden. Användare måste därför veta om det finns produkter för just de aktuella uppgifterna, annars ger språkverktygen inte den förväntade effektiviseringen. Samtidigt måste språkverktygen hitta just sina specifika potentiella användare och köpare. Det är ett av flera skäl till att det från flera håll har framförts behov av testning av språkteknologiska produkter, t.ex. i form av både språklig kontroll och certifiering av produkter. I bl.a. en marknadsundersökning från MATS-projektet¹ (Marknadsundersökning, 2001) pekas på behovet av att t.ex. data- och affärstidskrifter publicerar testresultat. Detta skulle enligt vår bedömning påskynda framväxten av en mer effektiv marknad för bra och användbara produkter och tjänster. Vi menar att det språkteknologiska sekretariatet måste ha en god överblick över produktutbud, praktiska erfarenheter och testresultat för svenska språkverktyg. I det konsumentupplysande arbetet bör ingå att följa och informera om utbud och marknadsläge, ha överblick över genomförda tester och andra studier, följa och samla erfarenhet av användningen av olika verktyg samt uppmärksamma icke tillgodosedd efterfrågan på svenska produkter och stimulera till utveckling av nya produkter eller anpassning av befintliga verktyg till svenska.

En språkteknologisk strategi bör tas fram

Nutek har föreslagit att det bör utarbetas en långsiktig strategi för främjande av svensk språkteknologi. Vi delar den bedömningen. För att vi skall kunna utnyttja språkteknologins potential – både för språkpolitiken, forskningspolitiken, näringspolitiken, förvaltningspolitiken och handikappolitiken – bör det språkteknologiska sekretariatet utarbeta en språkteknologisk strategi för att stärka kompetens och utveckling inom det språkteknologiska området.

¹ Projektet Methodology and Application of a Translation System (MATS) drivs i samarbete mellan Institutionen för lingvistik vid Uppsala universitet, Scania CV AB och Explicon. Det finansieras genom Vinnova.

En inventering av den språkteknologiska branschen bör göras

Vidare bör en närmare inventering av den språkteknologiska branschen göras, bl.a. av vilka produkter och tjänster som finns tillgängliga på svenska språket, vad som är bra och vad som saknas. Språkteknologiska produkter och tjänster är enligt vår mening en viktig del av infrastrukturen.

Nordiskt samarbete bör utvecklas inom språkteknologin

Vi vill även här peka på att olika former av nordiskt samarbete inom språkteknologiområdet borde kunna medföra flera fördelar. Att samarbeta inom området borde vara kostnadseffektivt och utvecklande för både näringsliv och forskning. Strukturen i språken är lika, vilket gör det förhållandevis enklare att anpassa mellan olika nordiska språk än från engelska till något av de nordiska språken. Dessutom är flerspråkiga korpusar till stor nytta vid utvecklandet av flerspråkiga produkter och tjänster. Vi menar att det bör vara en av uppgifterna för det språkteknologiska sekretariatet att i samarbete med de övriga nordiska länderna realisera ett utökat nordiskt samarbete.

17.2 Datorbaserad översättning

17.2.1 Vad menas med maskinöversättning och översättningsminnen?

Genom internationaliseringen av samhället har krav på översättningar mellan olika språk ökat. Inom översättningsområdet kan datorer användas i olika utsträckning, allt ifrån direkt maskinöversättning till att fungera som ett stöd för översättarens arbete.

Maskinöversättning innebär att datorn med hjälp av lexikon, grammatiska regler m.m. genomför en tolkning av strukturen hos meningarna i den text som skall översättas och därefter skapar en text på målspråket. De enkla maskinöversättningssystemen översätter ord för ord, medan mer avancerade transfersystem i flera steg översätter meningar som fristående enheter. I dag används sådan maskinöversättning dels som läs- och kommunikationsstöd för att tillgängliggöra innehållet i texter skrivna på ett främmande språk, t.ex. direktöversättning av webbsidor, dels som delsystem i en mer omfattande översättningsprocess, där översättningen utgör råmaterial för vidare bearbetningar.

I en del fall (speciellt om ämnesområdet är väl avgränsat) kan utfallet bli tämligen tillfredsställande, medan i andra fall resultatet blir sämre. Ett väl genomarbetat system kan garantera kvalitet på vissa aspekter där manuell översättning kan missa (t.ex. sifferuppgifter). Vid vårt seminarium om språkteknologi gjorde forskare bedömningen att upp till 70 % av en facktext kan maskinöversättas godtagbart (se vidare bilaga 4).

Förutom direkt maskinöversättning finns olika typer av översättningsverktyg som omfattar t.ex. elektroniska lexikon, termdatabaser och översättningsminnen. Översättningsstationer är integrerade system för professionellt bruk. De omfattar vanligen flerspråkig ordbehandling och översättningsstöd av olika slag samt ibland också ett maskinöversättningssystem. Till översättningsstationen kan även referensmaterial av olika slag finnas kopplade, t.ex. textsamlingar på käll- och målspråk som behandlar det ämnesområde som översättningen gäller jämte sökfunktioner. Tidigare

gjorda översättningar kan finnas tillgängliga i översättningsminnet eller vara åtkomliga för sökningar på annat vis.

17.2.2 Nutek föreslår en satsning på svensk språkteknologi

Nutek har på regeringens uppdrag utrett tekniska och ekonomiska förutsättningar för att ta fram en svensk basmodul för maskinöversättningar av standardiserade förvaltnings- och affärsdokument (Maskinöversättning, 1999). Av rapporten framgår bl.a. att den största efterfrågan på översättningar i Sverige gäller från svenska till engelska. Genom EU-medlemskapet har dock även behovet av översättning till svenska ökat. Därtill kommer den växande sektorn av s.k. lokalisering (översättning och anpassning av programvaror och manualer) som görs åt företag som exporterar till Sverige. I rapporten görs bl.a. en genomgång av existerande prototyper till maskinöversättningssystem för eller med svenska.

Nutek menar i rapporten att den tekniska utvecklingen inom området maskinöversättning har varit långsam. Även de bästa av dagens system har begränsad översättningsförmåga. De är dock i många fall tillräckliga som läs- och kommunikationsstöd och som delsystem för råöversättning i samband med textproduktion, antingen som fristående system eller som komponenter i stödsystem för översättare. Under senare år har alternativa tekniker börjat utprövas. Dessa metoder verkar enligt Nutek kunna övervinna vissa av transfermetodens brister, men finns ännu inte som kommersiella produkter.

Den viktigaste faktorn för maskinöversättningssystemets översättningsförmåga är systemets lexikon och regler. Ju större och bättre dessa är, desto bättre blir översättningarna. För att förbättra systemen måste man enligt Nutek utnyttja den översättningskunskap som ligger latent i existerande översättningar. Nutek konstaterar i sin rapport att behovet av svenska maskinöversättningssystem i det lägre segmentet – dvs. mer osofistikerade produkter – kan tillgodoses av marknaden, men att utvecklingen av avancerade maskinöversättningssystem för svenska är eftersatt.

Man menar vidare att den viktigaste förutsättningen för utvecklingen av avancerade maskinöversättningssystem är tillgången på översättningsdatabaser. Slutligen menar Nutek att det finns ett behov av ett konsumentupplysningsorgan på språkteknologins område. Nutek föreslår att det utarbetas en långsiktig strategi för att främja svensk språkteknologi. Denna skulle innehålla följande åtgärder:

- att översättningsdatabaser byggs upp
- att befintliga system utvärderas
- att ett integrerat översättningssystem byggs upp
- att en myndighet/organisation utpekas som koordinator för strategiarbetet och de ovan föreslagna åtgärderna. Denna skulle också ha till uppgift att följa utvecklingen inom området, bygga

upp och förvalta resurser för svenska språket samt sprida information och stödja erfarenhetsnätverk.

Det tidigare nämnda MATS-projektet bygger vidare på Nutekrapportens förslag om behovet av mer kunskap om tillgängliga forskningsprototyper. Inom MATS-projektet har man gjort en marknadsundersökning av den kommersiella marknadspotentialen för ett maskinöversättningssystem med svenska som käll- eller målspråk. I rapporten Marknadsundersökning (2001) pekas bl.a. på att maskinöversättningar kan leda till sänkta kostnader, snabbare produktion samt konsekventare och mer exakta översättningar. Det konstateras att det inte finns någon lösning för högkvalitativ maskinöversättning till eller från svenska på marknaden i dag. Av undersökningen framgår att hälften av de svarande företagen skulle föredra att låta skribenterna skriva på svenska, om det är deras modersmål, för att det höjer texternas kvalitet. Förutsättningen var att man skulle ha tillgång till en fungerande maskinöversättning. Merparten av de övriga angav att man konsekvent arbetar med sitt koncernspråk. Undersökningen pekar sammanfattningsvis bl.a. på att maskinöversättningssystemen bör testas och att resultaten av dessa bör spridas. Det pekas vidare på att myndigheter bör välja att införa systemet.

17.2.3 Maskinöversättning inom EU för vissa språk

Inom EU:s olika organ läggs stora resurser på översättning av olika dokument och tolkning mellan de elva officiella språken. Den totala produktionen av översättningar till svenska inom hela EU-verksamheten ligger på drygt 200 000 sidor per år, oftast från engelska men även i betydande grad från franska. (Se kapitel 4.)

Europeiska kommissionen har under de senaste 25 åren arbetat med att utveckla och anpassa ett flerspråkigt hjälpmedel för maskinöversättning för internt bruk, EC Systran (se bl.a. EC Systran, 2001). Maskinöversättning har utvecklats för åtta av EU:s officiella språk (dock inte för alla s.k. språkpar). För svenska, danska och finska finns maskinöversättning över huvud taget inte tillgänglig inom EU:s ram. För närvarande omfattar EC Systran 18 språkpar, samtliga med engelska eller franska. Valet av språkpar har gjorts utifrån olika kriterier:

- efterfrågan: arbetsspråken engelska och franska dominerar
- språklig släktskap: språkpar som franska–spanska är billigare och lättare att utveckla
- medlemsstaternas vilja att samfinansiera projekt är också av betydelse, därav utvecklingen av språkparen engelska–grekiska och grekiska–franska.

Maskinöversättningar används av handläggare på kommissionens sakenheter för tre syften:

- Snabböversättning av korta, repetitiva och ofta brådskande texter med standardiserad struktur och terminologi. Godtagbar översättningskvalitet uppnås efter det att texten korrigerats av

en person med målspråket som modersmål. Texterna används för internt bruk.

- Snabb genomläsning av texter på ett språk som beställaren inte behärskar.
- Framtagande av utkast på något annat språk än beställarens modersmål eller huvudspråk.

Översättare inom kommissionen använder maskinöversättningen som grund för en mer genomarbetad översättning av texten. Översättningens kvalitet varierar mellan olika språkpar. Resultaten för översättningar mellan franska och spanska bedöms som mycket goda. Spanska översättare använder EC Systran dagligen. Maskinöversättning har också god kvalitet där översättare kan missa (siffror etc.). Kvaliteten beror dock på vilken typ av källtext som skall översättas och kvaliteten på källtexten (stavning, format, syntax). Enligt kommissionen kan översättare som regelbundet använder sig av maskinöversättning förkorta översättningstiden med upp till en tredjedel om de rätta förutsättningarna finns (språkpar, texttyp, ämnesområde, stil).

Översättarna använder ett avancerat gränssnitt med namnet EURAMIS (European Advanced Multilingual Information System) som gör det möjligt att återanvända tidigare översättningsarbete genom översättningsminnen. Med utgångspunkt i tidigare översättningar kan datorn gå igenom texter som skall översättas och föreslå en översättning för de meningar där tillräckligt stor matchning uppnås. Denna kan sedan bearbetas av översättaren. Inte minst i EU-sammanhang där ofta ganska strikt bundna texter översätts ger ett sådant förfarande stora tidsbesparingar genom att mycket inte behöver nyöversättas varje gång. Resultatet är starkt beroende av dels storleken på de översättningsminnen som byggts upp, dels kvaliteten på de översättningar som lagras där. För att använda maskinöversättning och översättningsminne kombinerat kan översättaren arbeta med översättningsstationen Translator's Workbench (TWB). TWB understöder bl.a. svenska, men en begränsning vid användningen för översättarna är att det inte finns någon maskinöversättning för svenskan.

Den gemensamma finansieringen av nya språkpar har handlagts inom ramen för MLIS-programmet. I början av år 2001 undertecknades samarbetsavtal med den nederländska och den portugisiska regeringen för utveckling av språkparen nederländska–engelska/franska och portugisiska–engelska/franska. Inom ramen för programmet Human Language Technologies (HLT) har projekt startats för översättning till och från tjeckiska, ungerska och polska.

Vad gäller det nederländska utvecklingsprojektet kan följande nämnas (Machine Translation for Dutch, 2001). Nederländska finns med sedan länge i EC Systran, men har tidigare inte utvecklats och inte heller använts så mycket. Den tekniska utvecklingen

har dock gjort att de flesta översättarna numera använder översättningsminnen. Maskinöversättning kan integreras i den miljön och användas för de delar av en text där översättningsminnen saknas. Genom att efterredigering ändå behövs när man arbetar med översättningsminnen är det inget stort merarbete för användarna att efterredigera också maskinöversatt text. Vid det nederländska valet av system gjordes en offentliga upphandling som företaget Systran vann. Det nederländska projektet kostar 2,4 miljoner euro och pågår i två år (motsvarande 50 personår). Det finansieras till 50 % av företaget Systran och till 50 % av ett konsortium bestående av kommissionen, Nederländska språkförbundet samt regeringarna i Nederländerna och Flandern. Kommissionens medfinansiering tas från det nu avslutade MLIS-programmet.

De nordiska medlemsstaternas språk är ännu inte med i något färdigutvecklat maskinöversättningssystem inom EU. Fler förslag till samarbete har lagts fram när det gäller danska, men ännu har de inte gett något resultat. Tilläggas kan att det i Danmark har utvecklats ett maskinöversättningssystem för patent (PaTrans). MLIS-programmet finansierar utvecklingen av språkparen finska–engelska och engelska–finska i samarbete med ett finländskt företag. Enligt uppgift har man kommit långt främst med riktningen finska-engelska. Vid kommitténs studiebesök i kommissionen framkom att det finns önskemål från bl.a. svenska översättare om språkparen engelska–svenska och franska–svenska.

För den som översätter är naturligtvis också sådant som termdatabaser, elektroniska lexikon och flerspråkiga ordbehandlingsprogram viktiga hjälpmedel. Inte minst viktigt är här termarbete i vid mening, och terminologisk utveckling är även i andra avseenden av stor betydelse för unionens kommunikation. En omfattande terminologisk verksamhet bedrivs också inom EU. Flera stora termdatabaser har byggts upp, bl.a. Eurodicautom vid kommissionen och TIS vid rådet.

17.2.4 Bedömning och förslag

Vårt förslag: Maskinöversättning för svenska skall utvecklas. Från språkpolitisk utgångspunkt kan maskinöversättningar ses som ett medel för att stärka svenska språket och undvika ytterligare domänförluster eller kanske till och med återerövra domäner, t.ex. när det gäller dagens författande av texter på engelska inom svenska företag och myndigheter. Genom att göra tillgången till maskinöversättning mer allmän t.ex. på Internet, kan var och en använda svenska språket i större utsträckning – både genom att skriva texter på svenska som sedan översätts och genom att låta datorn översätta enklare utländska texter till svenska.

För att detta skall fungera är det viktigt att gemensamma, svenska språkresurser och då särskilt s.k. parallellkorpusar byggs

upp. Även flerspråkiga termbaser är nödvändiga språkresurser. Den viktigaste faktorn för maskinöversättningssystemets översättningsförmåga är systemets lexikon och regler. Ju större och bättre dessa är, desto bättre blir översättningarna. Vi har därför i föregående avsnitt föreslagit att en nationell språkbank, innehållande bl.a. fritt tillgängliga översättningskorpusar, skall vidareutvecklas.

Behovet av maskinöversättningar bör vara stort inom det alltmer internationaliserade näringslivet. Inom flera branscher är tempot högt och man varken hinner eller har råd att översätta texter, material och program manuellt. Automatiserad översättning kan snabba upp och förbilliga översättningar betydligt. Möjligheten att använda maskinöversättningar kan innebära att även små och medelstora svenska företag kan närma sig den internationella marknaden. Väl fungerande maskinöversättningar innebär dessutom stora tidsvinster jämfört med manuella översättningar. Bra översättningar av t.ex. manualer är en del av produktkvaliteten. Genom språklig och kulturell anpassning av svenska produkter till utländska konsumenters önskemål kan nya marknader öppnas för svenska företag. Maskinöversättning är också viktig inom förvaltningen. Tillgång till maskinöversättning för svenska bör t.ex. kunna stärka svenskans bruk inom EU.

Det är stor skillnad mellan att översätta mer allmänna texter och att översätta facktexter inom givna ämnesområden. Dagens maskinöversättningar fungerar bäst för standardiserade facktexter. Det behövs mycket grundforskning för att vidareutveckla maskinöversättningar.

Maskinöversättning för svenska skall utvecklas

Det är viktigt att Sverige som ledande IT-nation är med och utvecklar detta område. Ju tidigare fungerande system för maskinöversättningar tas fram, desto större är chansen att göra domänvinster. Vi menar att mer avancerade maskinöversättningssystem för svenska skall tas fram inom ramen för det nordiska samarbetet och/eller i samarbete med EU-kommissionen. Maskinöversättning bör – utifrån såväl språk-, närings-, förvaltnings- som forskningspolitiska utgångspunkter – göras tillgänglig för både offentlig förvaltning och för näringslivet samt för allmänheten. Det finns ett översättningsbehov av såväl förvaltningsdokument som affärsdokument, bruksanvisningar m.m., framför allt i syfte att små och medelstora företag skall kunna bredda sin marknad utanför Sverige. För enklare maskinöversättningssystem bör marknaden själv kunna ta fram de produkter och tjänster som efterfrågas. Men utvecklingen av avancerade maskinöversättningssystem för svenska är eftersatt. Därför behövs insatser för att främja fortsatt forskning och utveckling inom området. Insatserna för att bygga gemensamma språkresurser och tillgängliggöra dessa är en viktig förutsättning för ett fortsatt utvecklingsarbete.

Vi föreslår att Vinnova i samarbete med språkvården ges i uppgift att leda arbetet med att ta fram maskinöversättning för

svenska. Vinnova är den myndighet som på kort sikt har bäst kompetens för detta arbete. På lite längre sikt bör det föreslagna språkteknologiska sekretariatet kunna ta över denna uppgift.

Arbetet bör genomföras i samarbete med andra nordiska länder och EU-kommissionen. Vår kartläggning har visat att det finns intresse inom kommissionen för att utveckla maskinöversättning för svenska, men samtidigt är finansieringsmöjligheterna mindre klara än tidigare. Tillgången på flerspråkiga, fritt tillgängliga resurser som grund för modellen är god inom EU-kontexten. Även inom den nordiska sfären finns på olika håll intresse för ett utvecklingsprojekt. Samtidigt bör mycket av det utvecklingsarbete som gjorts både i Sverige, andra nordiska länder och inom EU kunna användas för att utveckla svensk maskinöversättning. Utvecklingsarbetet bör samfinansieras av näringslivet, svenska staten och förslagsvis andra nordiska länder och EU-kommissionen.