

OM ÖVERSÄTTNING MED HJÄLP AV DATORPROGRAM

Pro gradu-avhandling i nordisk
filologi vid Jyväskylä universitet
Våren 2002

Anne Kinnunen
Paula Takala

JYVÄSKYLÄN YLIOPISTO

Humanistiska fakulteten	Institutionen för nordiska språk
Författare Kinnunen Anne, Takala Paula	
Titel Om översättning med hjälp av datorprogram	
Ämne nordisk filologi	Typ av avhandling pro gradu
Avhandlingen färdig Vårterminen 2002	Antalet sidor 61
<p>Sammandrag</p> <p>Syftet med denna avhandling är att undersöka betydelsen av översättningsprogram och översättningsminnen som hjälpmedel för översättare. Vår avsikt är att utreda den potentiella nyttan av dessa system i förhållande till de besvär och kostnader som deras användning förutsätter. Detta förhållande granskar vi från två olika håll. För det första undersöker vi två olika programs funktioner och kommenterar hur lätt eller svårt det är att använda dem. För det andra reder vi ut hurdana krav som ställs på texterna som skall översättas med hjälp av detta slags program.</p> <p>I teoridelen presenterar vi maskinöversättningens historia. Vi tar fram de allra första försöken med maskinell översättning och fortsätter med utvecklingen ända till de nutida kommer-siella översättningsprogrammen. Vi presenterar också de viktigaste systemen och inriktning-arna på maskinöversättningens område.</p> <p>I analysdelen utför vi några översättningstester med ett maskinöversättningsprogram och ett översättningsminne och utvärderar deras användbarhet som hjälpmedel för en frilansöver-sättare. Vi använder som testmaterial färdigt översatta manualer av IBM-skrivare.</p> <p>Resultaten visar att TranSmart är mycket användarvänligt medan Trados är utmanande på grund av sin komplexitet. Båda programmen förutsätter att texterna som skall översättas fyller vissa speciella krav för att deras användning skall vara effektiv. TranSmart passar inte som sådant till yrkesmässig översättning fast själva översättningskvaliteten är bra. Ett stort hinder för programmets användning i översättningsarbetet är den besvärliga efterbe-handlingen. Trados är däremot lämpligt för yrkesmässig översättning som sådant om kraven på dess effektiva användning fylls.</p>	
Uppslagsord Maskinöversättning, översättningsminne, Kielikone TranSmart, Trados	
Bibliotek/Förvaringsplats Institutionen för nordiska språk	
Övriga uppgifter	

INNEHÅLL

1 INLEDNING	5
2 MASKINÖVERSÄTTNINGENS HISTORIA	7
2.1 Första generation	7
2.2 ALPAC (Automatic Language Processing Advisory Committee)	9
2.3 Andra generation	11
2.4. Aktuella inriktningar	12
3 OLIKA MASKINÖVERSÄTTNINGSSYSTEM.....	13
3.1 Mekaniseringsgrad av översättning.....	13
3.2 Datorstödd översättning	14
3.3 Automatisk översättning	15
3.3.1 Interaktion med översättaren	15
3.3.2 Direktöversättningssystem	15
3.3.3 Interlinguainriktade system	17
3.3.4 Transferinriktade system	19
3.2.4 Några andra system	23
4 SYFTE, MATERIAL OCH METODER	24
4.1 Problemställningar	24
4.2 Kielikone TranSmart	25
4.2.1 Allmänt	25
4.2.2 Transfer som grundteknologi	26
4.2.3 Taggningar och lexikon.....	27
4.3 Trados.....	28
4.3.1 Versioner	28
4.3.2 Komponenter	29
4.3.3 Fuzzy matching	30
4.4 Frilanspaket	31
4.5 Metoder	32
5 KRAV PÅ ORIGINALTEXT	33
5.1 Originaltext.....	34
5.2 Behovet av olika slags översättningskvalitet	34
5.3 Lämpliga texter	36
6 BESKRIVNING AV ORIGINALTEXTER	37
7 ANVÄNDAR-ORIENTERAD TESTNING.....	37
8 TRANSMART	38
8.1 Användbarhet	38
8.2 Funktionalitet	40
8.2.1 Exakthet.....	41
8.2.2 Lämplighet.....	44
9 TRADOS	45
9.1 Användbarhet	45
9.1.1 Förståelighet	45
9.1.2 Lärbarhet.....	46
9.2 Effektivitet.....	49
9.3 Funktionalitet	50

9.3.1 Exakthet.....	51
9.3.2 Lämplighet.....	52
10 DISKUSSION	53
LITTERATUR	59

1 INLEDNING

Behovet av översättning ökar ständigt i dagens informationssamhälle. Till exempel utgör olika språkservice 40-50 % av de löpande utgifterna i institutionerna i Europeiska unionen. Största delen av dem är utgifter för översättning och tolkning. (Arnold 1995, <<http://xlwww.essex.ac.uk/MTbook/HTML/node5.html>>.) Kravet på hög kvalitet i EU-dokument är en självklarhet men nuförtiden finns det många andra texttyper och situationer där det skulle räcka med en helt grov översättning, till exempel vad gäller www-sidor och e-post. Det finns inte ens tillräckligt många mänskliga översättare som skulle kunna klara av allt som det nutida informationsflödet ställer inför oss. Där kan teknologin hjälpa.

Utvecklingen inom maskinöversättningen har svarat på utmaningar som informationssamhället har ställt. Utbudet på översättningsteknologi är nuförtiden omfattande och den har genomgått en särutveckling. Under de senaste årtiondena har det kommit ut på marknaden hjälpmedel för översättare. Dessa så kallade översättningsminnen sägs öka effektiviteten i översättningsarbetet samtidigt som de inte ger avkall på idealen av en högkvalitativ översättning. I dessa program är det ju människan som översätter och programmet är endast ett hjälpmedel som ”kommer ihåg” hur man tidigare har översatt samma eller liknande meningar. Å andra sidan har också allt flera maskinöversättningssystem blivit kommersiella på grund av en ny slags efterfrågan t.ex. i elektronisk kommunikation och hantering av informationsflödet. Dessa automatiska översättningssystem strävar inte efter perfekt kvalitet utan efter en god råöversättning.

Syftet med vårt arbete är att betrakta kombinationen människa och översättningsminne samt kombinationen maskinöversättning och efterbehandling. Översättningsminnen används redan vidsträckt i synnerhet på det tekniska området medan själva maskinöversättningssystemet troligen är mindre kända som hjälpmedel i det traditionella översättningsarbetet. Vår avsikt är att utreda den potentiella nyttan av dessa system i förhållande till de besvär och kostnader som deras användning förutsätter. Detta förhållande försöker vi granska från två olika håll. För det första undersöker vi två olika programs funktioner och kommenterar hur lätt eller svårt det är att använda dem. För det andra försöker vi utreda hurdana krav ställs på texterna som skall översättas med hjälp av dessa program.

I teoridelen skall vi presentera maskinöversättningens historia, som vi anser vara ganska viktig för att man skall kunna förstå den påfallande utvecklingen i branschen. I kapitel 2 tar vi fram de allra första försöken med maskinell översättning och fortsätter med utvecklingen ända till de nutida mera brukbara översättningsprogrammen. Efter det skall vi i kapitel 3 presentera de viktigaste systemen och inriktningarna på maskinöversättningens område samt försöka beskriva lite deras verksamhet.

I kapitel 4 presenterar vi närmare ett par olika översättningsprogram: *TranSmart*, som är ett maskinöversättningsprogram, och *Trados*, som är ett översättningsminne. I analysdelen försöker vi utvärdera deras funktion och användbarhet på en generell nivå. Vi kommer att pröva dessa hjälpmedel för översättare i praktiken för att kunna få en djupare syn på saken. Här använder vi som testmaterial färdigt översatta manualer av IBM-skrivare.

Före presentering av resultaten av individuella tester riktar vi uppmärksamheten på originaltext i kapitlen 5 och 6. Först betraktar vi denna fråga på basis av teori. Efter det beskriver vi originaltexterna i vårt testmaterial. I kapitel 7 presenterar vi närmare våra testtyper och metoden i presentering av resultat. Resultaten av våra tester presenteras i kapitlen 8 och 9, båda program i sitt eget kapitel. Till slut anger vi i kapitel 10 våra egna subjektiva intryck och jämförelser samt avväger om dessa program är eller kan bli något beaktansvärt hjälpmedel i översättarens arbete.

2 MASKINÖVERSÄTTNINGENS HISTORIA

I och med datorn fick man idén att översättning skulle kunna utföras maskinellt. Man trodde att det skulle vara lätt. Allt man skulle behöva var att överföra ett par ordböcker till en elektronisk form. Maskinöversättning blev ett nytt forskningsområde och fick snabbt mycket understöd, också finansiellt.

Många forskare inom maskinöversättningens område har redan hunnit ägna hela sitt arbetsliv åt en fråga full av utmaningar: hur våra tankar konstrueras och om man kan lära en maskin tänka. Deras insats har gett impuls till nya idéer och idag finns det redan flera mera eller mindre fungerande system på marknaden.

2.1 Första generation

De första fröna till utvecklingen av en dator hade knappt utformats när man började drömma om automatisk översättning. Två forskare, den ena i Frankrike, den andra i Ryssland, patenteerade sina översättningsmaskiner 1933 utan att de visste om varandra. Det är dock ryssen P. Smirnov-Trojanskij som har satt mera påtagliga spår i minnen av sina efterträdare. Fast hans egen maskin var i stort sett bara en ord-för-ord översättare, introducerade han tre faser av maskinell översättning som senare också tekniskt sett blev realistiska. (Hutchins & Somers 1992, 5.) De tre faserna är *analys*, *transfer* och *generering*. Trojanskij påstod också att det finns ett logiskt innehåll som är gemensamt för alla språk. Därför är det möjligt att forma en universell, logisk representation mellan vilka som helst två språk. På det viset introducerade han ytterligare ett begrepp som används idag, *interlingua*. (Buchmann, 1987, 4.) De tre faserna av maskinöversättning förklaras mer i detalj i samband med olika strategier för maskinöversättning.

Trojanskij var före sin tid (Hutchins & Somers 1992, 5). Det tog ännu ungefär tjugo år innan maskinöversättning väckte uppmärksamhet bland andra forskare, och då hade man redan glömt det preliminära arbetet som Trojanskij hade gjort (De Roek 1987, 42). I Amerika utvecklade A.D. Booth och R.M. Richens 1947 ett datorprogram som var ett mekaniserat, tvåspråkigt översättningssystem. Ett år senare förändrade de systemet med att indela uppslagsordet i ordstam och affix. De hade både lingvistiska och datormässiga motiv till detta. Å ena

sidan fäste man inte stor uppmärksamhet vid syntax på den tiden, och den dominerande uppfattningen var att man skulle lyckas med att översätta semantiska enheter efter varandra. Å andra sidan var datorkapaciteten, i synnerhet minneskapacitet och snabbhet, ett ständigt bekymmer för dem som medverkade i maskinöversättningsprojekt. Den nya metoden minskade betydligt antalet uppslagsord i elektroniska ordböcker. (Buchmann 1987, 4-5.)

W. Weaver utvärderade i sitt memorandum år 1949 arbetet som A.D. Booth och R.M. Richens hade gjort. Hans egen synpunkt reflekterade hans åsikt att maskinöversättning liknade avkodning. (Buchmann 1987, 6.) Weaver föreslog att man skulle använda avkodningsmetoder, som daterade sig från krigstiden, samt statistisk analys och informationsteori i forskningen av maskinöversättning. Ytterligare framhävde han forskning av logiska och universella drag i språk (Hutchins & Somers 1992, 6.) I syfte att underlätta förståelighet av översättningar föreslog Weaver att man skulle dra nytta av statistiska undersökningar av semantik för att minska ambiguitet av vissa ord och fraser. Ambiguiteten var ju ett av de största problemen som då sysselsatte forskare. (Buchmann 1987, 6.)

Weavers memorandum gav flera universitet och tekniska institut i USA en impuls till forskningsprojekt. Den första heltidsforskaren av maskinöversättning, Y. Bar-Hillel, var utnämnd 1951. Redan ett år senare organiserade Bar-Hillel en konferens där riktningar av kommande forskningsarbete utformades. (Hutchins & Somers 1992, 6.) Den var den första konferensen av detta slag och den samlade ihop arton lingvister och ingenjörer i datateknik. En av frågorna som togs upp var mekanisering av syntaktisk analys för den hade datormässigt blivit möjlig på grund av utvecklingen på området. E. Reifler hade föreslagit 1950 det första trovärdiga schemat för en mekaniseringsprocess av texter. Han introducerade begreppen *för-* och *efterbehandling*. Med förbehandling menade han att man skulle markera med speciella tecken de punkter som man visste att de skulle vara problematiska beträffande grammatik eller betydelse av enskilda ord. I efterbehandlingsfasen skulle en översättare se till att den maskinöversatta texten blev förståelig. Lingvistiskt sett betydde mekanisering visserligen att datorn bara automatiskt identifierade syntaktiska ordklasser och syntaktiska funktioner av fraser (fraser som fungerade som substantiv, adjektiv, adverb eller verb). (Buchmann 1987, 6-8.)

På den första konferensen presenterade man också idéer om användning av *reglerade språk* (texter som är skrivna för att översättas med maskin, dvs. med att använda förenklad struktur

och förminska möjligheter till ambiguitet) och *subspråk* (med en begränsad terminologi och grammatik på något specialområde, t.ex. väderleksrapporter).

Första generationens maskinöversättning baserade sig i stort sett på ord-för-ord-översättning. Typiskt för dessa system var att lingvistisk planering varken var konsekvent eller logisk, man hade inte separerat lingvistiska data och algoritmer (instruktioner enligt vilka ett program utför olika processer, t.ex. adderar två inmatade siffror), och varje problem löstes som ett fall för sig. Trots det resulterade ord-för-ord-översättning också i positiv utveckling: å ena sidan var dess höga krav på minneskapacitet inte utan inverkan på datateknik, och å andra sidan utvecklades omfattande elektroniska ordböcker. (Buchmann 1987, 15-16; Hutchins & Somers 1992, 48.)

På 50- och 60-talen trodde man trots allt fast på framtidsutsikter av maskinöversättning i många forskningsgrupper. Forskningen var dubbelriktad: somliga grupper koncentrerade sig fortfarande på "direktöversättning" med att använda empiriska, ofta statistikbaserade metoder medan andra grupper inriktade sig redan på mera teoretiska applikationer, så kallade transfer- och interlinguasystem (Hutchins & Somers 1992, 6).

Den gradvisa övergången från empirisk till teoretisk forskning bevisar att några forskare började inse otillräckligheten av ord-för-ord-översättning. De insåg att utvecklingen av maskinöversättningssystem utan syntaktisk analys inte längre var möjlig. Bar-Hillel trodde att helt automatisk maskinöversättning av hög kvalitet var omöjlig redan i princip. Kommersiellt var maskinöversättning nog användbar antingen som grova system för bruk i situationer där översättning av lägre kvalitet skulle räcka eller för att användas som ett hjälpmedel och tidsvinst för en översättare som skulle efterredigera de maskinöversatta texterna. (Buchmann 1987, 15-18.)

2.2 ALPAC (Automatic Language Processing Advisory Committee)

50-talets optimism förändrades till besvikelse när man blev tvungen att inse komplexiteten av lingvistiska problem (Hutchins 1995). I och med mera formella lingvistiska teorier, t.ex. formell beskrivning av syntaktiska strukturer av Noam Chomsky år 1957, hade forskare inom maskinöversättning insett att de saknade den teoretiska basen för forskningen. Försök att fylla

denna brist invercade däremot på faktumet att det inte fanns några operationella system av hög kvalitet i bruk i början av 60-talet. (Buchmann 1987, 16, 19.)

1964 bildades ALPAC (Automatic Language Processing Advisory Committee) för att undersöka maskinöversättningens utsikter. Kommittén organiserades av de statliga finansiärer som ville se om deras insats, hittills ungefär 20 miljoner dollar, hade varit till nytta. ALPAC hade en rent kommersiell målsättning i undersökningen och den räknade till och med noga om manuell eller maskinell översättning blir dyrare. (Buchmann 1987, 20; Hutchins 1995.) I ALPAC-rapporten år 1966 konstaterades det att maskinell översättning var långsammare, dyrare och inexaktare än en manuell översättning (ALPAC 1966, 19-20).

Maskinöversättning var då ännu ganska outvecklad och därför hade den inga riktiga möjligheter att konkurrera mot hårda finansiella beräkningar. Man måste komma ihåg att datorerna på den tiden inte var mera komplicerade än elektroniska kalkylatorer redan på 80-talet, inte heller hade man utvecklade programmeringsspråk i bruk utan man använde ett enkelt symboliskt maskinspråk. (Buchman 1987, 20; Hutchins & Somers 1992, 72.)

ALPAC kom till den slutsatsen att det inte fanns användbar maskinöversättning och kanske inte ens blir det. Kommittén föreslog att man borde göra slut med allt finansiellt stöd till maskinöversättning. ALPAC rekommenderade att man i stället för att forska i maskinöversättning hellre borde koncentrera sig på att utveckla t.ex. generell datalingsvistik och maskinella hjälpmedel för översättare. (ALPAC 1966, 32-34.) De stora illusioner som man hade om maskinöversättning kulmineras i ALPAC-kommitténs rapport:

...when, after eight years of work, the Georgetown University MT project tried to produce useful output in 1962, they **had to resort to postediting**. [vår framhävnig] (ALPAC 1966, 19)

Efter många års forskningsarbete hade man inte lyckats producera färdig text av hög kvalitet utan mänsklig intervention.

Många ansåg ALPAC-rapporten vara såväl knapp och kortsynt som partisk. Den anklagade i onödan att efterbehandlingen gjorde maskinöversättningen så dyr att det inte var lönsamt. Man behöver ju kontrollera också de manuellt gjorda översättningarna om man vill ha hög kvalitet. ALPAC togs så seriöst att den förorsakade en stilla period i maskinöversättningen. (Hutchins 1995.) Inverkan var allvarligast i USA där de följande 25 åren har ansetts vara

nästan helt framgångslösa på maskinöversättningens område. I Europa hade ALPAC också en kraftig inverkan, men där varade dess effekt bara i drygt ett tiotal år (Pugh 1992, 19, 24.) ALPAC hade präglat maskinöversättningen som misslyckad och dessutom fick den ingen finansiering till utvecklingsarbetet (Buchmann 1987, 20).

2.3 Andra generation

Efter ALPAC flyttade maskinöversättningen i stor utsträckning från USA till Europa och Kanada. I Kanada opererade TAUM-projektet, som hade framgång och gav ut till exempel Météo-systemet till översättning av väderleksrapporter. (Hutchins & Somers 1992, 7-8.) I USA var de tio åren efter ALPAC-rapporten relativt passiva i maskinöversättningen. Då undersöktes dock områden som t.ex. datalingvistik och artificiell intelligens, som också senare spelade en viktig roll för maskinöversättning. (Warwick 1987, 25.)

Det fanns två olika inställningar till maskinöversättning på tiden efter ALPAC. Den ena forskade i maskinöversättning på ganska generell nivå medan den andra ville skapa operationella översättningssystem för bruk. De operationella systemen fick stöd och finansiering av privata företag medan universiteten fortsatte den generella undersökningen. Exempel på dåtida operationella system var SYSTRAN och LOGOS. Den generella undersökningen tog hänsyn till teoretisk utveckling i lingvistik och tog med idén om djupstruktur och universalier i språk. På det sättet försökte man utveckla en interlinguabaserad modell för maskinöversättning: man kan nämligen tänka att det finns en mellanform – interlingua – mellan källspråklig och målspråklig översättningstext, t.ex. djupstruktur. (Warwick 1987, 24.) Interlingua-tänkandet förlorade en del av sin framgång när transfer-tänkande tycktes ha bättre framtid. (Se avsnitt 3.2.3; Hutchins 1995).

År 1975 började man intressera sig för maskinöversättningen på nytt på grund av framgång och ökat finansiellt stöd. Synen på maskinöversättningen hade blivit mera realistisk och lyckade resultat, till exempel med TAUM-METEO, inspirerade finansiärerna att investera. De ungefär tio åren efter ALPAC hjälpte att bilda grunden för den andra generationen i maskinöversättningen. (Warwick 1987, 25-26.)

Undersökningen i slutet av 1970-talet riktade sig åt tre olika håll: 1) till utveckling av transferbaserade system, som bildades utgående från de tidigare interlinguabaserade systemen, 2) till utveckling av nya typer av interlingua-system och 3) till undersökning av tekniker och utvecklingslinjer för artificiell intelligens. (Hutchins 1995.)

På 1980-talet tog maskinöversättningen stora steg framåt. Många nya system utvecklades och de blev kommersiellt mer intressanta. SYSTRAN installerades vid stora inrättningar (t.ex. NATO) och blev det mest framgångsrika programmet i maskinöversättningen. Störst var den kommersiella succén av maskinöversättning på 1980-talet i Japan. (Hutchins 1995.)

På 1980-talet infördes maskinöversättningssystem för mikrodatorer. Ofta var de inte så utvecklade men kunde ändå utnyttjas som värdefulla hjälpmedel för översättare. Ett av de första programmen för mikrodatorer var ALPS-systemet som hade många olika funktioner: flerspråklig ordbehandling, automatisk ordbok och terminologisk konsultering samt interaktiv översättning. (Hutchins 1995.)

Många forskare hade ända till början av 1980-talet som mål att utveckla helautomatiska översättningssystem. Tanken ansågs dock vara för ambitiös och man koncentrerade sig sedan på mera realistiska mål. (Pugh 1992, 17-18.)

2.4. Aktuella inriktningar

Till slutet av år 1989 hade de traditionella *regelbaserade* översättningssystemen (både transfer- och interlinguabaserade system) förlorat en del av sin värderade ställning när de korpusbaserade metoderna hade kommit i bruk. I början var man förvånad över att resultat kunde bli så bra med hjälp av statistik och korpus av tidigare översatta texter. Vilket var nytt var att dessa system inte alls använde semantiska regler i textanalysen. (Hutchins 1995.)

Regelbaserade system gick framåt trots att uppmärksamheten kanske mest var hos kunskapsbaserade system (se avsnitt 3.2.4). Både transfer- och interlinguasystem undersöktes och förbättrades. På 1990-talet tog man i beaktande hur man skulle kunna förbättra översättningskvaliteten i maskinöversättningen. Problemet med kvaliteten ansågs vara en av de största svå-

righeterna i maskinell översättning. Ett stort utvecklingsområde som undersöktes på 90-talet var också översättning av talat språk. (Hutchins 1995.)

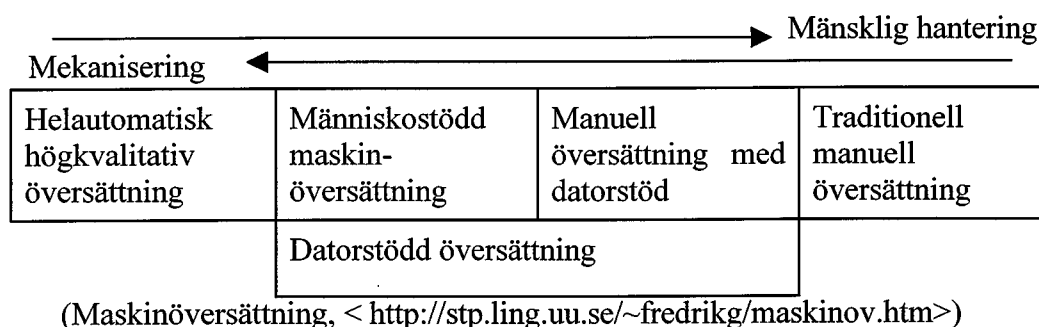
1990-talet kom med ny entusiasm mot maskinöversättningen. Stora företag intresserade sig för maskinella översättningstjänster som kända system (Systran, METAL, Logos) erbjöd. På samma tid kom nya system för mikrodatorer som också fick en massa användare. Ett stort genombrott i maskinöversättningens utveckling var då uppfattningen att professionella översättare hellre behöver hjälpmedel med sitt eget arbete än helautomatiska maskinöversättningsprogram med bristfällig kvalitet: "translator's workbench" eller "translator's workstation" utvecklades. (Hutchins 1995.)

Dessa hjälpmedel för översättning kombinerar t.ex. flerspråklig ordbehandling, medel för att ta emot och överföra elektroniska dokument, textigenkänningsmedel och översättningsminne. Översättningsminnet möjliggör att översättaren kan jämföra tidigare översättningar om bestämda fraser eller satser med sin dåvarande översättning. Nuförtiden används översättningsminnen vidsträckt i översättningen av t.ex. manualer. Dessa omarbetas ofta bara delvis och kan därmed dra nytta av översättningsminnet som har den originala texten i minnet (Hutchins 1999.)

3 OLIKA MASKINÖVERSÄTTNINGSSYSTEM

3.1 Mekaniseringsgrad av översättning

Mera konkreta målsättningar ledde till att man i början av 80-talet inledde utvecklingen av omfattande datahjälpmedel för professionella översättare. I vissa fall gör dessa hjälpmedel översättningsarbetet lättare och snabbare, i synnerhet därför att översättaren kan välja vilka hjälpmedel han eller hon vill använda vid varje tillfälle. Dagens översättningsmetoder kan indelas i fyra kategorier beroende på i vilken mån maskinen och människan koopererar. Det har illustrerats i Figur 1.



Figur 1 *Mänsklig och maskinell översättning*

Vi behandlar inte traditionell översättning. Datorstödd översättning och olika system för automatisk översättning beskrivs i följande avsnitt.

3.2 Datorstödd översättning

Manuell översättning med datorstöd (CAT - Computer Aided Translation) kan i vidare utsträckning betyda all hjälp som en dator kan ge människan: korrekturläsning, grammatik- eller stilrevidering (Hutchins & Somers 1992, 149). Oftast betyder CAT dock mänsklig översättning som använder datorstöd i själva översättningsprocessen. Datorn använder korpusar av tidigare översatta text (dvs. översättningsminnen) och bjuder alternativ till översättaren. Kända program som utnyttjar sådana översättningsminnen är till exempel *Trados*, *Transit* och *Déjà Vu*. (Computer Aided Translation, <<http://www.isp.nl/Us/cattxt.html>>.)

Ett CAT-program processar textsegment som programmet sedan integrerar i sin databas. Databasen växer ständigt med översättningen men ändå kan den hitta översättningsförslag till översättaren mycket snabbt. Nuförtiden kan programmet ofta föreslå ett segment även om det inte perfekt motsvarar det källspråkliga uttrycket. Då kan programmet visa motsvarigheten i procenttal. (Computer Aided Translation, <<http://www.isp.nl/Us/cattxt>>.)

System som kallas "translators' workstations" kan delas in i tre nivåer på basis av deras funktion. Nivå ett utnyttjar ord- och termbehandling samt telekommunikation. På nivå två finns det därtill textanalys, automatisk sökning i ordboken samt synkroniserad tvåspråkig textsökning. Nivå tre närmar sig redan automatisk maskinöversättning. (Melby 1992, 147.)

3.3 Automatisk översättning

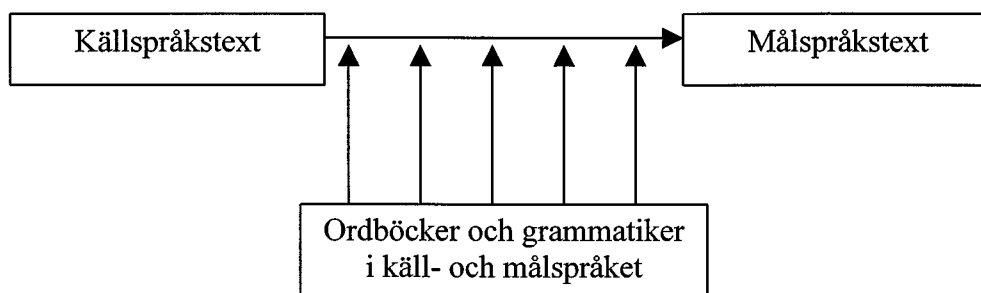
3.3.1 Interaktion med översättaren

Automatisk översättning omfattar maskinell översättning antingen utan eller med hjälp av människan. De flesta maskinöversättningsprogram i dag utnyttjar mänsklig hjälp och bara några få program arbetar helt utan mänsklig påverkan. Automatisk översättning som människan deltar i i någon mån kan också kallas människostödd maskinöversättning. Vi räknar inte denna översättningstyp som datorstödd översättning (som i figur 1) utan som automatisk maskinöversättning. Helautomatisk högkvalitativ översättning (se figur 1) är fortfarande ett ideal, inte en realitet. Maskiner kan ännu inte tänka.

Människan kan vara med i översättningen redan i förbehandlingen av källtexten, interaktivt under själva processen eller i efterbehandlingen av målspråkstexten. Ofta är också förbehandlingen interaktiv när programmet förutser möjliga kommande problem i översättningen av vissa texter. Efterbehandlingen av översatta texten kan också vara interaktiv: programmet frågar efter människans råd på ställen där det fanns några tvivelaktiga uttryck eller oklarheter under översättningen (Hutchins & Somers 1992, 150.) I dag kan en maskin nå en viss nivå i översättningen men människan måste ofta behandla översättningstexten antingen på för- eller efterhand för att språket skulle bli tillräckligt naturligt.

3.3.2 Direktöversättningssystem

Historiskt sett är ett direktöversättningssystem det första systemet i maskinöversättningen. (Tucker 1987, 22). Jfr:



Figur 2 *Direktöversättningssystem*

Direktöversättningssystem är alltid utvecklade för ett speciellt språkpar. Den huvudsakliga principen i dessa system är att ord och syntax analyseras bara i den utstreckning som är nödvändigt för disambiguering och identifikation av lämpliga uttryck och ordföljd i målspråket. (Hutchins 1986, 54.) Direktöversättningssystem baserar sig vanligtvis inte på någon generell lingvistisk teori. Varje mening översätts steg för steg, så att output av varje steg blir input för det nästa steget. Det finns vanligen ungefär tio steg. De väsentligaste stegen är listade nedan.

1. Ord i källtexten söks upp i lexikonet och de analyseras morfologiskt
2. Homografer identifieras
3. Sammansättningar identifieras
4. Nominal- och verbfraser identifieras
5. Idiomatiska uttryck processas
6. Prepositioner processas
7. Subjekt och predikat identifieras
8. Syntaktiska ambiguiteter identifieras
9. Måltexten genereras och processas morfologiskt
10. Ord och fraser i måltexten omorganiserar

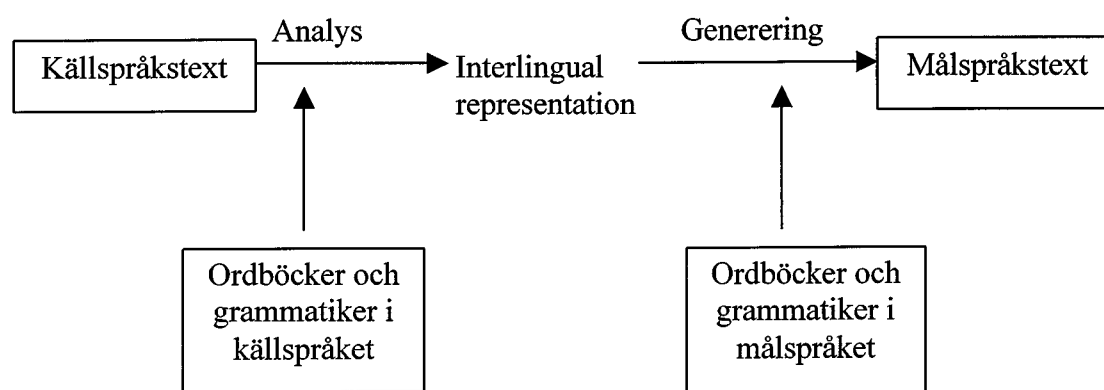
Olika system skiljer sig litet från varandra vad användning och ordning av dessa steg berör. (Tucker 1987, 22-23.)

Direktmetoden är tung. Den stannar inte ifall den inte känner igen ord eller grammatiska konstruktioner. Den passerar helt enkelt dessa punkter och överför dem orörda till måltexten. I värsta fall producerar den text som är obegriplig i målspråket. Systemet kan dessutom inte utvidgas eller modifieras utan ansträngning därför att översättningsprocessen inkluderar komplicerade, interaktiva regler. Det är utvecklat för översättning i en riktning, t.ex. från finska till engelska. (Arnold 1995.)

Hutchins och Somers (1992, 72-73, 175) jämför kvaliteten i traditionella direktöversättningssystem med resultat som skulle kunna väntas om någon som bara har elementära kunskaper om målspråkets grammatik försökte att översätta en text med hjälp av en enkel, tvåspråkig ordbok. Direktmetoden används fortfarande modifierad t.ex. i Systran, som har varit i bruk från 70-talet. Allmänt taget var direktmetoden dock misslyckad, och lingvistiskt mera sofistikerade system utvecklades, nämligen interlingua- och transfermetoder.

3.3.3 Interlinguainriktade system

Interlinguametoden utvecklades före transfermetoden. Tanken var att skapa en "genuint" abstrakt representation som skulle kunna tillämpas mellan vilka som helst språk. Med andra ord skulle källspråktexten analyseras så noggrant att den interlinguala mellanrepresentationen skulle kunna användas direkt för generering av målspråktext. (Hutchins & Somers 1992, 73.) Hutchins (1986, 55) presenterar interlinguasystemet i form av en figur som illustreras nedan något modifierad.



Figur 3 *Interlinguasystem*

Figur 3 illustrerar den ideala situationen i interlinguainriktade system. Den interlinguala representationen inkluderar all information, både lexikal och syntaktisk, i abstrakt form. Målspråktexten kan genereras från mellanrepresentationen utan att "se tillbaka" till källtexten (Hutchins & Somers 1992, 73). I praktiken har man inte hittills kunnat skapa ett genuint interlingvalt översättningssystem, utan åtminstone en tvåspråkig ordbok har fogats till systemet i samband med den interlinguala representationen.

Att forma en abstrakt representation av ordens betydelse är kanske det största problemet i interlinguatanken för tillfället. Olika språk skiljer sig från varandra i förhållande till hur de delar världen in i begrepp, vilket förorsakar svårigheter när man försöker att hitta den abstrakta lexikala representationen. (Arnold 1995, <<http://clwww.essex.ac.uk/MTbook/HTML/node36.html>>.) Problem uppstår när ett språk representerar ett begrepp med ett ord medan ett annat språk har flera ekvivalenter för detta begrepp. Se följande exempel (en del av exemplen har lånats från Hutchins & Somers 1992, 101):

1. Engelska <i>wall</i>	Svenska	<i>vägg</i>
	Tyska	<i>Wand</i> (inne i en byggnad) <i>Mauer</i> (ute)
2. Engelska <i>river</i>	Svenska	<i>älv</i> (i de nordiska länderna) <i>å</i> (mindre) <i>flod</i> (utanför de nordiska länderna)
	Franska	<i>rivière</i> (en allmän term) <i>fleuve</i> (viktigaste floder, som rinner ut till havet)

Man har två möjligheter att behandla problemet. Antingen väljer man arbiträrt ett språk vilkas "representation av världen" ska ligga till grund för interlingua eller tar man i beaktande alla skillnader som finns i språk. Det senare fallet är dock problematiskt därför att programmeraren av ett interlinguasystem inte kan behärska alla språk och skillnader mellan dem. Om han planerade ett system som inkluderade t.ex. engelska, svenska, tyska och franska, borde han ha två separata "former" för engelska *wall*, och tre för *river*. Om han senare ville foga japanska till systemet, borde han ha vetat att man behöver nio former för engelska *wear* beroende på vad man klär på sig (skor, hatt, byxor osv.). (Arnold 1995, <<http://clwww.essex.ac.uk/MTbook/HTML/node36.html>>.)

Det har ingen betydelse vilken form man ger till begreppen, den kan bestå av ord i ett språk (engelska, esperanto), nummer eller koder av ett påhittat språk (Arnold 1995, <<http://clwww.essex.ac.uk/MTbook/HTML/node36.html>>). Det är inte praktiskt genomförbart att inkludera i interlingual representation all information som skulle behövas för generering av vilket som helst språk. I stället är denna representation planerad med avseende på de språk som redan existerar i systemet eller som man har haft för avsikt att foga senare.

Interlinguametoden, såsom transfermetoden, skiljer sig från direktmetoden t.ex. vad behandling av ordföljd gäller. Direktmetoden bevarar källspråkets ordföljd och till slut omorganiserar den med hjälp av vissa regler. De två senare metoderna analyserar däremot syntax till en abstrakt form varifrån målspråkets ordföljd kan genereras med hjälp av målspråkets grammatikregler. (Arnold 1995, <<http://clwww.essex.ac.uk/MTbook/HTML/node35.html>>.) När undersökning av interlinguametoden initierades verkade Chomskys teori om djupstruktur lovande. Senare insåg man att dess orientering mot ytstrukturer av individuella språk förorsakade att den inte var tillräckligt abstrakt för tillämpning i interlinguasystem. (Hutchins & Somers 1992, 119.)

Intresset att utveckla interlinguametoden är förståeligt: det är betydligt lättare att foga nya språk till interlinguasystem än till transfersystem. Om ett interlinguasystem är rätt planerat kan man foga ett nytt språk med att skapa två nya komponenter (en analys- och en genereringskomponent) medan ett transfersystem kräver ytterligare en transferkomponent för båda riktningar. Behov för dessa transferkomponenter ökar exponentiellt. (Arnold 1995, <<http://clwww.essex.ac.uk/MTbook/HTML/node36.html>>.) Tabell 1 illustrerar behov av transferkomponenter i transfersystem i relation till antal språk som finns i systemet:

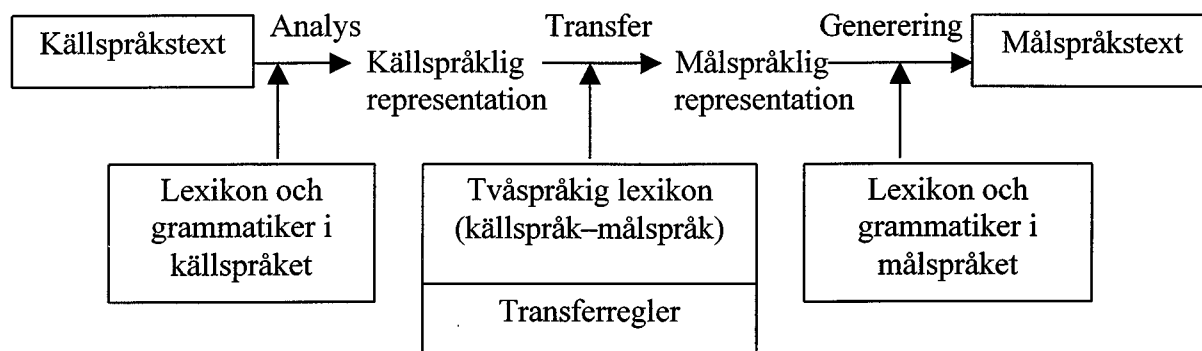
Tabell 1 *Antalet transferkomponenter i transfersystem*

Antal språk	Antal transferkomponenter
3	6
5	20
9	72

Svårigheter med att hitta en abstrakt representation för ord och syntaktiska processer och relationer har förorsakat att interlinguametoden är mera fruktbar som en bas för teoretisk undersökning än som en metod i operationella system (Arnold 1995, <<http://clwww.essex.ac.uk/MTbook/HTML/node36.html>>).

3.3.4 Transferinriktade system

I transferinriktade system utförs översättning i tre faser som maskinöversättningens pionjär Smirnov-Trojanskij introducerade redan på 30-talet: analys, transfer och generering. Hutchins (1986, 55) representerar transfersystemet med en figur som har illustrerats nedan



Figur 4 *Transfersystem*

I *analysfasen* skapas en mellanrepresentation som ändå är språkberoende. Först bearbetas källtexten preliminärt, t.ex. med att identifiera och markera skiljetecken i texten. Sedan analyseras texten morfologiskt, vilket innebär att orden sätts i grundform (t.ex. böcker → bok [+PLUR], sprang → springa [+IMPERFEKT]). I det nästa steget taggas orden, dvs. att de markeras med ett visst antal karakteristika som består av grammatisk kategoritillhörighet (t.ex. substantiv, adjektiv), semantiska egenskaper (t.ex. mänsklig, abstrakt) och syntaktiska egenskaper (t.ex. ackusativ, demonstrativ). Den slutgiltiga semantiska och syntaktiska analysen av satser resulterar i en form av ett analysträd. När satserna i källtexten har transformerats till trädstrukturer, börjar *transferfasen*. I denna fas ersätts källspråkets lexikala enheter med motsvarande enheter i målspråket i den mån som det är möjligt. Transfer kan utföras också på syntaktisk eller semantisk nivå då trädstrukturerna förändras så att de blir förenliga med målspråkets normer. I *genereringsfas* fingranskas satserna i enlighet med målspråkets normer. Det vill säga att t.ex. ordföljden korrigeras och orden böjs i förhållande till huvudorden. (Surfa på ditt modersmål - Maskinöversättning på Internet, <<http://www.svenska.sics.se/oversattning/mt-uppsats.htm>> .)

När källspråkets ord och strukturer inte har några direkta motsvarigheter i målspråket, behöver man regler för behandling av dessa enheter. Problematiska fall av *lexikal transfer* kräver ofta undersökning av kontext. Till exempel engelska *know* har två motsvarigheter i franska (*savoir, connaître*) och i tyska (*kennen, wissen*)

- (1.) I know the right answer.
Je connais la bonne réponse.
Ich kenne die richtige Antwort.

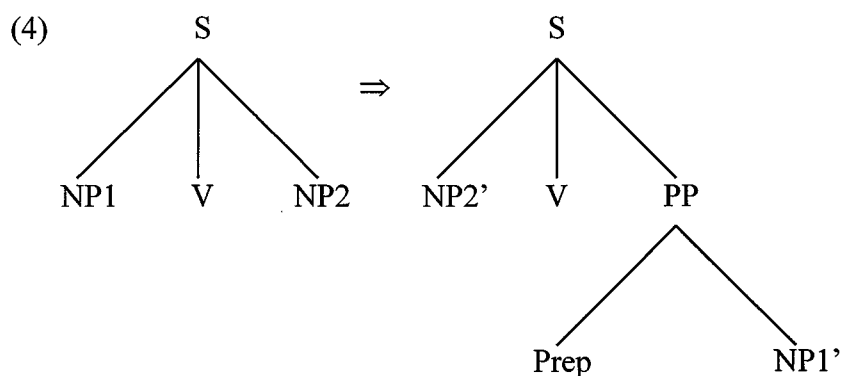
- (2.) I know what the right answer is.
Je sais quelle est la bonne réponse.
Ich weiß, was die richtige Antwort ist.

Valet av motsvarighet för engelska *know* i franska och tyska beror på det direkta objektet. Om det direkta objektet består av en nominalfras, väljer man orden *connaître* och *kennen*, om det direkta objektet däremot består av en underordnad fras, motsvaras det av orden *savoir* och *wissen*. (Hutchins & Somers 1992, 100, 113.) Interlinguametoden skiljer sig från transfermetoden vad gäller behandling av motsvarigheter av denna typ. I transfersystem behandlas dessa frågor under transferfasen med hjälp av speciella regler medan motsvarande frågor i interlinguasystem måste behandlas redan i analysfas utan referens till ett visst målspråk. Med andra ord måste det bestämmas redan i analysfasen av den engelska texten vilket slags "vetande" det är frågan om. (Hutchins 1986, 56.)

Satsernas struktur motsvarar inte heller alltid varandra mellan olika språk. Om man jämför till exempel den engelska satsen (3) med dess motsvarighet i franska och tyska, ser man att strukturerna inte stämmer överens

- (3) Jones likes the film.
Le film plaît à Jones.
Der film gefällt dem Jones.

Hutchins och Somers presenterar en "transferregel" för den franska satsen i form av en trädstruktur. Trädstrukturen illustreras nedan något simplificerad.



Subjektet av det engelska verbet *like* blir indirekt objekt i franska och dativobjekt i tyska medan objektet av det engelska verbet blir subjekt både i franska och tyska. (Hutchins & Somers 1992, 113-114.) Om analysen fördjupas, kan storleken och komplexiteten av komparativ grammatik förminska. Till exempel skulle man med att använda *semantiska relationer* ha kunnat eliminera behovet för regeln som har illustrerats ovan. (Arnold 1995, <<http://clwww>.

essex.ac.uk/MTbook/HTML/node35.html>.) I alla tre satser skulle *Jones* vara ”upplevaren” och orden *the film / le film / der Film* skulle anses som ”tema” i dessa satser.

Den huvudsakliga skillnaden mellan direkt-, transfer- och interlinguametoderna är storleken av de tre huvudkomponenterna (analys, transfer och generering). Hutchins och Somers illustrerar detta med ett pyramiddiagram

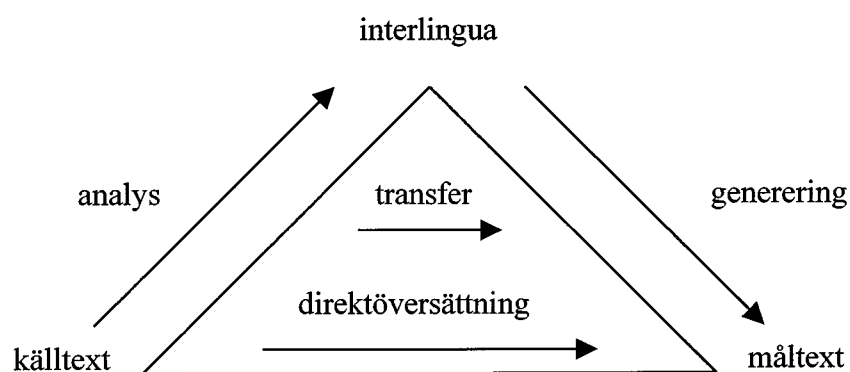


Figure 5 *Pyramiddiagram*

Diagrammet illustrerar också att ju mera källtexten analyseras desto enklare blir transferkomponenten. Direktmetoden står i ena ändan med minimal analys och maximal transfer medan interlinguametoden står i andra ändan med maximal analys och minimal (eller ingen) transfer. (Hutchins & Somers 1992, 106-107.)

I operationella system föredras transfermetoden framför interlinguametoden. Det finns åtminstone två orsaker till detta. För det första är det problematiskt att skapa språkberoende representationer av texter. För det andra blir analys- och genereringskomponenter mycket komplicerade därför att de avlägsnas från käll- och målspråkets karakteristika. I transfersystemen är dessa komponenter enklare för mellanrepresentationen består språkberoende. Dagens maskinöversättningssystem är i allmänhet baserade på en blandning av lingvistiska teorier. Vanligen utnyttjar de transformationsgrammatik eller dependensteori som bas men de kan vara mycket modifierade på grund av påverkan från andra teorier eller av datormässiga krav. (Hutchins & Somers 1992, 76, 81.)

3.2.4 Några andra system

Det har också kommit några andra system i bruk under de senaste tiderna. Dessa system är ofta mindre utnyttjade än de tidigare nämnda bättre kända system. Deras resultat kan trots det vara relativt bra och de är säkert värda att ta upp.

Det finns tre typer av så kallade korpusbaserade översättningssystem. Alla tre använder en stor tidigare ihopsatt korpus av översättningstexter. Den *kunskapsbaserade översättningsmodellen* översätter text med hjälp av exempelöversättningar i korpusen (Somers 1992, 199.) Den försöker representera översättningstextens semantiska innehåll. Texten är möjlig att översättas när man förstår dess innehåll. Den källspråkiga texten förvandlas till en abstrakt interlingua med t.ex. kasusroller (agent, tema) och sedan översätts den till målspråket. (Maskinöversättning, <<http://stp.ling.uu.se/~fredrikg/maskinov.htm>>.)

Exempelbaserad maskinöversättning används vid sidan av kunskapsbaserade översättningsmetoder. Två språk har en parallellkorpus (dvs. en tvåspråkig korpus av översättningspar) och båda språken har analyserats i översättningsenheter och dessa har annoterats på samma sätt. Systemet alltså översätter genom att den imiterar redan existerande översättningar ur en massiv korpus och tillämpar dem i fortsättningen. (Maskinöversättning, <<http://stp.ling.uu.se/~fredrikg/maskinov.htm>>.) Resultatet blir ett översättningsmönster som direktöversättningssystemet då kan komplettera (Arnold 1995, <<http://clwww.essex.ac.uk/Mtbook/HTML/node93.html>>). Fördelen med detta system är att resultatet ganska säkert blir exakt därför att översättningsmodellen i korpusen har gjorts av professionella översättare (Hutchins 1995).

Statistikbaserad översättning baserar sig enbart på statistik (Maskinöversättning, <<http://stp.ling.uu.se/~fredrikg/maskinov.htm>>). Den använder en stor massa text på en korpus och jämför med hjälp av statistiken hur bra (i procenttal) till exempel ord- eller ordgrupp-ekvivalenter på käll- och målspråk motsvarar varandra (Somers 1992, 200-201). Systemet är användbart speciellt vid sidan av andra maskinöversättningssystem (Maskinöversättning, <<http://stp.ling.uu.se/~fredrikg/maskinov.htm>>).

Det finns säkert också andra intressanta översättningssystem. Vi har nämnt bara några av de mest kända metoder som borde iakttas. De här metoderna utnyttjas sällan som sådan utan de kombineras ofta med ett annat översättningssystem.

4 SYFTE, MATERIAL OCH METODER

4.1 Problemställningar

Syftet med vårt arbete är att undersöka betydelsen av översättningsprogram och översättningsminnen som hjälpmedel för översättare. Vad översättningsprogram gäller har vi avgränsat ämnet betydligt beträffande programmens användningsområde. Vi betraktar översättningsprogram inte i förhållande till det område på vilket de redan har börjat befästa sin ställning. Utvecklingsarbetet av maskinöversättning har nämligen fått nya tillämpningsområden som har framkallats av informationssamhället. Programmen används redan till exempel i sökverktyg på Internet, i översättning av e-post, eller som råöversättningar för att underlätta genomgången och bearbetningen av stora informationsmassor. På dessa ställen kunde utnyttandet av mänskliga översättare bli för dyrt och dessutom för långsamt, om inte omöjligt. Alla dessa uppgifter är oftast sådana där översättningsprogrammet, dvs. maskinen, inte konkurrerar med mänskliga översättare.

Vi intresserar oss för den aktuella kvalitetsnivån av dessa program jämfört med den traditionella mänskliga översättningen. Vi har valt två mycket olika hjälpmedel för översättare som forskningsobjekt och försöker jämföra dessa program på en generell nivå. På grund av programmens olikhet kan man inte direkt jämföra dem med varandra. Båda programmen är avsedda för att underlätta översättningsarbetet, men de angriper temat helt olikt, vilket för sin del försvårar jämförelsen. Därför har vi för avsikt att jämföra de två programmen närmast på en allmän nivå och undersöka hur de lämpar sig som hjälpmedel för en enskild översättare.

Vi skall undersöka fördelar av de två programmen och behandla dessa i förhållande till t.ex. kostnader eller besvär som de kräver. Centrala frågor i vår undersökning är till exempel följande: Hur besvärligt är det att använda programmet? Underlättar eller försvårar programmet översättarens arbete med hänsyn till förberedelse, tid som behövs för själva översättningen och efterbehandling? Vilken nytta kan man dra av programmen och i vilken mån? Hurdan är råöversättningens kvalitet?

I början presenterar vi de två maskinöversättningsprogrammen allmänt vart i sitt eget avsnitt. Därefter utför vi några översättningstester med de båda programmen och på samma gång utvärderar vi programmens användbarhet i allmänhet. På grund av olikheten av de två pro-

grammen använder vi olikartade tester för de båda programmen med hänsyn till översättarens behov. Till slut jämför vi på basis av våra egna erfarenheter programmets värde och användbarhet för en enskild översättare och försöker på en allmän nivå kartlägga deras egenskaper i förhållande till varandra.

4.2 Kielikone TranSmart

4.2.1 Allmänt

Kielikone är ett av de två största företagen inom språkteknologin i Finland. Bland Kielikones produkter finns det program för språkvård, elektroniska ordböcker och ett maskinöversättningssystem. Karakteristiskt för utvecklingsprocessen av deras maskinöversättningssystem har varit att de från första början har samarbetat med tre kundföretag: Nokia Telecommunications Oy, Rautaruukki Oy och Trantex Oy. Dessa företag har deltagit i processen med att hela tiden ge feedback. (Kieliteknologia Suomessa 1998, <<http://www.csc.fi/kieliteknologia/raportti/raportti.htm>>.)

Det finns två versioner av översättningssystemet Kielikone Transmart. Den ena (TranSmart Standard) passar för allmänt bruk medan den andra (TranSmart Professional) är inriktad åt professionell användning. Krav på programvara i användningen av TranSmart Professional är

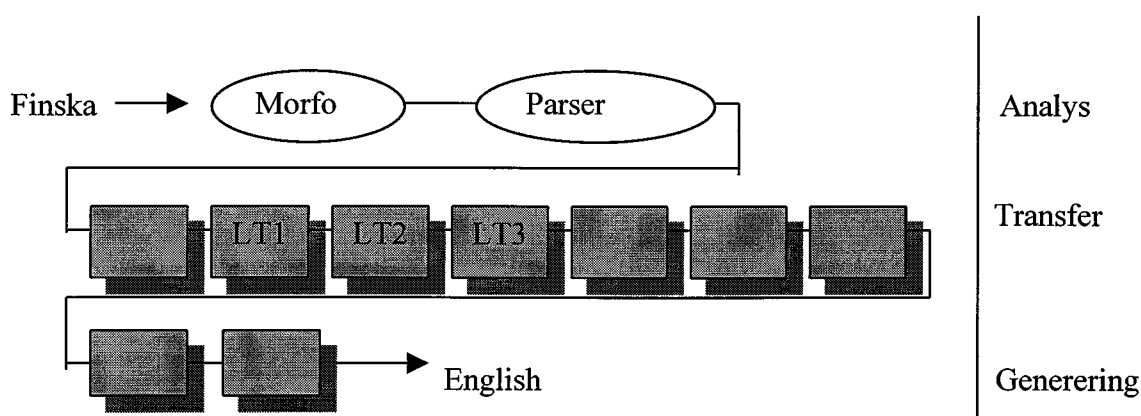
- Windows (version 95, 98, Me, 2000, NT 4.0 eller nyare)
- MS Word (version 6.0, 7.0, 97, 2000 eller nyare)
- Minst 32 megabytes primärminne
- Cirka 100 MB på hårddisken

(Kielikones hemsida, <http://www.kielikone.fi/kielikone/tuotteet/transmart_laitteistoalusta.shtml> Hämtat 23 oktober 2001.)

Vi undersöker TranSmart Standard. Detta innebär att programmet innehåller endast det allmänna lexikonet, därutöver kan vi skapa ett dokumentspecifikt lexikon. Vi har inte tillgång till andra lexikon. Grundteknologin, som är vårt forskningsobjekt, är densamma i båda versioner.

4.2.2 Transfer som grundteknologi

Kielikone använder transfersystemet som grund i sin maskinöversättningsteknologi. Som lingvistisk teori används dependensteorin. Under översättningsprocessen transfereras källspråkliga dependensstrukturer till målspråkliga. Systemet har tre faser: analys, transfer och generering. Analysdelen för sin del består av två faser: för det första analyseras ordformerna morfologiskt, för det andra analyseras satsstrukturerna enligt dependensmodellen. Transferfasen i TranSmart baserar sig på en deterministisk parsning som utförs i lineär tid. Till slut utförs generering. Arnola et al. (1996a) illustrerar TranSmart-systemet med en figur som presenteras nedan något simplificerad



Figur 6 *Transfer-systemet i TranSmart*

Transferfasen i TranSmart är indelad i flera separata deluppgifter. Deluppgifterna behandlas i en kedja av enhetsapplikationer så att varje enhet hanterar en viss deluppgift. I Kielikones maskinöversättningsteknologi utförs också den lexikala transferen i dessa enhetsapplikationer. Av figur 6 framgår att transfer aktiveras sju gånger i TranSmart. Tre enheter hanterar lexikal transfer. Transfersystemet kan när som helst förändras genom att definiera nya deluppgifter och tillägga nya applikationer. TranSmart är utvecklat utgående från en språkoberoende teknologi. Denna teknologi kan vidareutvecklas för ett nytt språkpar med mindre investeringar än det första språkparet från finska till engelska. (Arnola et al, 1996a; 1996b.)

Kielikone har också utvecklat ett översättningsminne på basis av dependensteorin. De traditionella översättningsminnen lagrar de källspråkliga segmenten med deras ekvivalenter i sitt minne. I fortsättningen jämför dessa översättningsminnen det aktuella segmentet med de

segment som redan finns i minnet. Kielikones översättningsminne lagrar däremot de källspråkliga segmenten i sitt minne i form av dependensträd. Detta översättningsminne jämför de tidigare lagrade dependensträden med de träd som det bildar ur den aktuella texten. (Arnola et al. 1996a.) För tillfället är detta översättningsminne inte integrerat i de två kommersiella versionerna av TranSmart.

4.2.3 Taggningar och lexikon

I TranSmart finns det ett system inbyggt som taggar problemfyllda ord med speciella färger eller streck. Det finns fyra typer av översättningsproblem på vilka programmet ber användaren att fästa uppmärksamhet. Om TranSmart inte känner till ord som finska ord, streckar programmet under de ifrågavarande orden med prickar. De övriga orden för vilka programmet inte hittar någon ekvivalent i sina lexikon stryks över. Kielikones översättningssystem försöker också lösa problemet som produktiviteten av finska sammansättningar förorsakar. *Hiha-tehdas* till exempel är ett helt korrekt finskt ord fast det inte finns i någon ordbok. Det är ett exempel på ord som är hypotetiskt, men som bra skulle kunna användas om någon kom på att grunda en sådan fabrik. På de områden som utvecklas snabbt bildas det nya ord med en motsvarande logik i en sådan takt att lexikonarbetet inte hinner med. Om TranSmart inte hittar basformen av ett substantiv i sina lexikon, försöker det analysera det som en sammansättning formad av två enskilda ord. Om programmet hittar ekvivalenter för de enskilda orden i sammansättningen, översätter det båda två separat och markerar dem med en dubbel understrykning. Ytterligare markeras uttryck som är kulturbundna med att både stryka över dem och strecka under dem med prickar. Användaren får välja om han i stället vill att dessa översättningsproblem markeras med vissa färger. (Hyvönen & Arnola 1998.)

Semantiska problem har Kielikone försökt att lösa med att utnyttja lexikon på ett sätt som underlättar disambiguering av ord. Det grundläggande systemet inkluderar bara ett allmänt lexikon. Andra mera specifika lexikon skapas alltid efter kundernas verksamhetsområde. Programmet söker ord i lexikon i följande ordning: 1. Dokumentspecifikt lexikon (om användaren har skapat en) → 2. Användarspecifikt lexikon (Allmän terminologi i företaget, t.ex. namn på produkter) → 3. Domänspecifika lexikon (t.ex. IT-lexikon, finansiellt lexikon, lexikon som ansluter sig till olika processer i produktion) → Ett allmänt lexikon. Med detta system ersätts delvis en omfattande semantisk djupanalys. I en text som handlar om informationstek-

nologi, översätter systemet ordet *tavu* genast som 'byte' och inte som 'stavelse', om IT-domänspecifikt lexikon är i bruk. (Hyvönen & Arnola 1998.)

4.3 Trados

4.3.1 Versioner

Trados är ett språkteknologiföretag som erbjuder tjänster och produkter för översättnings- och lokaliseringsprocesser. Trados Translator's Workbench är en av de fyra mest använda arbetsstationerna för översättare (Hutchins, <<http://ourworld.compuserve.com/homepages/WJHutchins/AMTA-96.htm>>).

Egentligen finns det två system, *Trados 5* och *Trados 5 Freelance* av vilka den ena är avsedd för stora organisationer och den andra för frilansöversättare. Trados 5 fungerar i olika nätverksomgivningar och kan tillämpas på alla språk som finns i Windows-applikationer medan Trados 5 Freelance är ett enskilt system som fungerar med högst fem språk. Dessa språk definieras vid installeringen. De båda systemen inkluderar fyra kärnkomponenter: WorkSpace för skapandet och hanteringen av översättningsprojekt, WinAlign för alignment, MultiTerm för hantering av termer och översättningsminnet Translator's Workbench. Ytterligare finns det komponenter som möjliggör översättning i andra omgivningar, t.ex. i PowerPoint, FrameMaker, Interleaf. Det är också möjligt att använda Trados för att översätta internetfiler. (Trados Translation guide.)

För att nå optimal effektivitet rekommenderas att antingen operativsystemet Windows NT eller Windows 2000 används. Krav på programvara är:

- Windows 95, Windows 98, Windows ME, Windows NT eller Windows 2000
- Minst 64 megabytes primärminne
- Microsoft Office 97 eller Microsoft Office 2000.

(Trados hemsida, <http://www.trados.com/products/fast_track_system_requirements.asp>.)

Vi undersöker en demoversion av Trados 5 Freelance. Därför har vi några restriktioner i dess bruk som vi måste ta i beaktande. Den största begränsningen är att vi kan översätta bara unge-

fär 100 översättningssegment med översättningsminnet *Translator's Workbench*. En annan begränsning av en demoversion är att resultat av *WinAlign* inte kan lagras eller exporteras.

4.3.2 Komponenter

Det grundläggande draget i *Translator's Workbench* är att det lagrar alla översättningar, som senare kan återanvändas. Översättningsprocessen sker i segment som vanligen består av en sats. Programmet lagrar varje segment med dess målspråkliga ekvivalent i översättningsminnet. I fortsättningen jämför programmet alltid det segment som skall översättas med tidigare inlagrade segment. Om programmet har en motsvarighet i minnet, ger det ett översättningsförslag åt översättaren. Översättaren kan acceptera förslaget som sådant, då nästa segment framträder på fältet. Översättaren kan också modifiera förslaget och först senare acceptera det. (Description of the individual TRADOS programs, <http://partners.trados.com/download/Trados_Product_Descriptions.pdf>; Trados hemsida, <http://www.trados.com/products/deep_track_translation_memory.asp>.)

MultiTerm är ett verktyg för hanteringen av terminologi. Användaren kan inkludera termer i flera språk i *MultiTerm* med klassificerande eller beskrivande information enligt behov. *MultiTerm* kan konstrueras med att manuellt tillägga enheter eller med att importera färdiga ordlistor i databasen. Termbankens struktur kan när som helst modifieras efter användarens vilja. Sökfunktionen i *MultiTerms* databas fungerar i båda riktningar, dvs. att användaren kan söka termer på vilket som helst språk som finns i systemet. *MultiTerm* söker automatiskt termer i det aktiva segmentet. Om det finns några termer, visar programmet dessa och översättaren kan överföra dem i dokumentet med ett klick. (Description of the individual TRADOS programs, <http://partners.trados.com/download/Trados_Product_Descriptions.pdf>.)

Med *WinAlign* kan man utnyttja tidigare översatta texter som inte har gjorts med Trados. *WinAlign* jämför originaltexten och dess översättning och matchar individuella segment med deras ekvivalenter. Matchningarna lagras i en textfil och därefter överförs de till översättningsminnet. (Description of the individual TRADOS programs, <http://partners.trados.com/download/Trados_Product_Descriptions.pdf>.)

WorkSpace är ett verktyg med vilket man kan skapa och hantera individuella översättningsprojekt. I *WorkSpace* kan användaren lagra all material från kunden (t.ex. en källtext och olika slags tilläggsmaterial) under ett namn i en projektlista. *WorkSpace* inkluderar bland andra punkten *Translate with TagEditor* 'Översätt med TagEditor'. Om till exempel allt material av ett översättningprojekt av ett HTML-dokument har lagrats på denna punkt, öppnar *WorkSpace* automatiskt den fil som skall översättas i *TagEditor*. Samtidigt öppnar *WorkSpace* *MultiTerm*-databasen och *Translator's Workbench* med ett eventuellt översättningsminne. (Translationzone.com, <<http://www.translationzone.com/Scripts/WebObjects.dll/TZ.woa/wa/ProductsDirectAction/freelanceEdition>>.)

4.3.3 Fuzzy matching

Utöver fullständig överensstämmelse (100 % match) kan översättningsminnet söka efter ungefärliga motsvarigheter för det aktiva källsegmentet. Användaren får sätta ett tröskelvärde i procenttal på basis av vilket programmet går igenom de redan lagrade segmenten i sitt minne. Om tröskelvärdet är t.ex. 70 %, söker översättningsminnet segment som har minst 70 % motsvarighet med det källsegment som skall översättas. Detta system kallas *fuzzy matching*. För ordet *multilingual* skulle sökning med denna strategi kunna ge olika resultat beroende på tröskelvärdet:

100 %	{multilingual}
80 %	{multilingual}
60 %	{multilingual, bilingual}
40 %	{multilingual, bilingual, MultiTerm}

(Laboration 4: Översättning med översättningsstöd, <<http://stp.ling.uu.se/educa/kurssidor/ist99/istl4.html>>.)

För att översättaren skall kunna notera olikheterna mellan källsegmentet och fuzzy match-förslaget, markerar programmet dessa med olika färger. Ord som inte motsvarar varandra markeras med *gul*. Om fuzzy match-förslaget skiljer sig från källsegmentet så att det har färre eller flera ord än källsegmentet, framhävs dessa med *grå* färg. *Blå* används för att indicera att en del av segmentet har placerats olikt från källsegmentet (t.ex. uttrycket till exempel). (Trados *Translator's Workbench* Online Help.)

Det finns en option för automatisk substituering av nummer, akronymer eller andra variabler, som påverkar fuzzy match-förslaget. Om användaren har valt denna option, behandlar Translator's Workbench som 100 % match ett segment i sitt minne som skiljer sig från det aktuella segmentet i källtexten bara i förhållande till dessa variabler. Med andra ord förändrar programmet sitt förslag så att det ersätter automatiskt t.ex. en akronym som finns i dess minne med den nya akronymen i källtexten. (Trados Translator's Workbench Online Help.)

4.4 Frilanspaket

Lokalisering av ett maskinöversättningssystem betyder att man lär systemet det speciella ordförrådet och ovanliga uttryck som förekommer i en text. Lokaliseringen kan ha stor effekt på översättningskvaliteten. (Honkonen 2001.) Vanligtvis lokaliseras översättningsprogram alltid till de företag eller organisationer som köper det. Detta betyder att t.ex. färdiga lexikon skapas enligt företagets behov.

För individuella översättare skulle Kielikone enligt Kaarina Hyvönen (personlig e-post) bjuda en TranSmart Professional-version. Paketet skulle lokaliseras språkligt med hänsyn till översättarnas specialområde och kunder. Typiskt lokaliseras t.ex. olika namn, termer och mycket generella ord enligt individuella behov. Allt annat skulle översättarna göra själva, fast Kielikone kunde undervisa dem i att upprätthålla sina lexikon. Företeelser som översättarna inte själva kan tillägga till sina lexikon, måste de skilt beställa från Kielikone. Kaarina Hyvönen medger att lokalisering höjer priset av programmet ganska betydligt: när grundpaketet (TranSmart Standard utan lokalisering) kostar ungefär 650 euro (3900 mark), blir priset av TranSmart Professional med fullständig lokalisering redan cirka 3300 euro (20 000 mark).

Trados kan såsom TranSmart också lokaliseras för varje företag individuellt. Translator's Workbench-översättningsminnet som inkluderar MultiTerm Professional skulle kosta ungefär 3000 euro (18 000 mark). Trados utbildar organisationer för cirka 1100 euro (6600 mark) per dag. Det arrangerar också kurser som kostar ungefär 220 euro (1300 mark) per person. (Balkan et al., 1997.)

4.5 Metoder

Vi har valt tekniska texter som testmaterial och vi tror att de innehåller mycket terminologi och är upprepande. Som en teoretisk grund för evalueringen har vi stött på en forskningsrapport om evaluering av system för processning av naturliga språk (Natural language processing, NLP), EAGLES Evaluation of natural language processing systems.

Vi använder som testmaterial IBM-manualer som vi har hittat på Internet. Materialet är ursprungligen skrivet på engelska men vi utnyttjar den finskspråkiga versionen som källtext i alla fall. Det var ytterst svårt att hitta lämpligt material för vår undersökning. Vårt material måste vara i elektronisk form, vi behöver samma text både på finska och på engelska och texten borde också vara teknisk. Materialet av IBM fyller alla dessa krav och det lämpar sig idealt för våra behov. Vi har valt två IBM-manualer på finska och motsvarande manualer på engelska. De är två olika manualer för samma IBM-skrivare. Vi antar att dessa två texter skulle lämpa sig mycket bra till översättningsminne. Texterna borde ha likadana drag, dvs. de borde t.ex. inkludera många samma termer. Den andra är en bruksanvisning för skrivarens användare och den andra en bruksanvisning för operatören. Dessa texter är översatta från de engelska bruksanvisningarna *IBM Network Color Printer: User's Guide* och *IBM Network Color Printer: Operator's Guide*. Vi undersöker de finskspråkiga manualerna i fall det skulle finnas skillnader i dem. Vi försöker ta reda om dessa möjliga skillnader i den finska källtexten kommer att påverka översättningen som ett av våra testprogram, TranSmart, producerar.

Först skall vi utföra en testöversättning med översättningsprogrammet TranSmart. Därefter skapar vi ett kortfattat dokumentspecifikt lexikon och då översätter vi källtexten på nytt. Med detta försöker vi kartlägga möjliga skillnader eller förbättringar som framträder mellan dessa två översättningsversioner. Vi vill undersöka om översättningskvaliteten betydligt kan förbättras med hjälp av specifika lexikon. Vi intresserar oss också för påverkan av dokumentets utseende. Vi skall ta reda på hur redigeringar eller brist på dem kommer att påverka översättningsresultatet av TranSmart. För det första låter vi programmet översätta texten som sådan. Därefter tillsätter vi en tabell i texten såsom den ursprungligen ser ut. På det sättet testar vi om programmet klarar av att hantera text i tabellen och om den översätter texten bättre eller sämre när den står i tabellen. Detta visar också tydligt om TranSmart bevarar de ursprungliga redigeringarna, dvs. om tabellen ser likadan ut före och efter översättningsprocessen. Till slut

tar vi ett urval ur de översatta texterna och utvärderar råöversättningens kvalitet i de valda meningarna.

Med Trados-översättarens arbetsstation testar vi först om källtexterna är upprepande. Detta gör vi så att vi börjar med ett översättningsminne som är helt tomt (såsom det också skulle vara vid inköp). Vi tror att programmet under översättningsprocessen förr eller senare kommer att föreslå översättningsförslag för olika källspråkliga segment, naturligtvis beroende på hur pass upprepande vår källtext egentligen är. Ju mera upprepning källtexten har, desto flera ekvivalenter kommer översättningsminnet att föreslå och desto tidigare. Efter detta lagrar vi denna första översättning som översättningsminne i Trados Translator's Workbench för att kunna dra nytta av den i fortsättningen. Då försöker vi översätta den andra IBM-manualen med hjälp av detta översättningsminne. Vi har för avsikt att evaluera om det blir avsevärt nyttigare att översätta texten med ett översättningsminne av samma område eller om nyttan inte blir så speciellt stor. Vi skall också försöka skapa ett litet översättningsminne med WinAlign och använda vidare en del av en IBM-manual till detta. Vi kommer att testa om detta alignment är svårt att förverkliga och vi försöker också testa om man kan dra nytta av det.

Vi kommer också att undersöka båda programmen beträffande deras användbarhet eller nytthet. Samtidigt som vi testar programmens olika funktioner skall vi betrakta programmen ur en nybörjares synvinkel, dvs. hur lätt eller svårt det är att lära sig att använda dessa program. Med en nybörjare menar vi en människa som inte har någon tidigare erfarenhet av maskinöversättningssystem eller översättningsminnen. Ingentera av oss har någonsin använt ett sådant datorprogram och därför passar vi bra som testare.

5 KRAV PÅ ORIGINALTEXT

I maskinell översättning spelar källtexten en viktig roll för översättningskvaliteten. Olika texttyper passar bättre eller sämre till maskinöversättning eller till översättningsminne. Man bör alltid ta hänsyn till vad för en källtext man har för att man skall kunna evaluera dess potentiella lämplighet för de ovannämnda verktygen. Därför har vi till att börja med närmat oss denna fråga på basis av teorin. Vi har velat sätta oss in i teorin om texttyper och texternas karakteristika.

5.1 Originaltext

Det finns olika slags texttyper, t.ex. administrativa texter, lagtexter eller tekniska texter, vilka definierar avsikten med texten. Texttypen återspeglar också vilken slags terminologi texten innehåller. Terminologin kan vara generell, domänspecifik eller företagsspecifik. I somliga organisationer kan en texttyp vara dominerande, medan flera texttyper kan förekomma i andra i lika stor utsträckning. Dessutom varierar texttypen och terminologin individuellt i varje text. (EAGLES 1995, <<http://www.issco.unige.ch/ewg95/node162.html>>.)

Originaltexten karakteriseras vanligen med drag som

- upprepande
- idiomatisk
- rik på terminologi
- fattig på terminologi

Om individuella termer, ord, fraser och till och med hela meningar förekommer vid upprepade tillfällen i en text oftare än i genomsnitt, kan *upprepning* vara ett karakteriserande drag i denna text. Satsstrukturerna och ordföljden är i dessa texter mera konsekventa än vanligt. En teknisk text har typiskt mycket upprepning. En *idiomatisk* text är i många förhållanden en motsats till upprepande text. I en idiomatisk text används många olika ord och termer som synonyma för variationens skull. Variationen kan också synas i satsstrukturer. Litterära texter och reklamtexter är ofta idiomatiska. Texter som är *rika på terminologi* innehåller rikligt med speciell terminologi på ett visst fack- eller verksamhetsområde. I sådana texter kan också ett ord från generellt lexikon få en speciell betydelse. Dessa karakteristika är typiska för vetenskapliga och tekniska texter. *Fattig på terminologi* är däremot inte en direkt motsats till rik på terminologi. Om texten är fattig på terminologi kan den bestå av endast allmänt ordförråd, eller att standardtermer eller domänspecifika termer är ersatta med alternativa uttryck. (EAGLES 1995, <<http://www.issco.unige.ch/ewg95/node163.html>>.)

5.2 Behovet av olika slags översättningskvalitet

Traditionellt har översättningens mål varit att producera en högkvalitativ ekvivalent till källtexten. Översättningar av denna typ är avsedda för publicering eller annan slags distribuering t.ex. inom ett företag eller en organisation. I princip har maskinöversättning strävat efter att

fylla detta behov från början fast det har förutsatt efterbehandling av den maskinellt översatta texten. På grund av utvecklingen av arbetsstationer för översättare under de senaste åren har användningen av maskinell hjälp i högkvalitativ översättning ökat avsevärt. Behovet av översättning i dagens samhälle har dock fått nya dimensioner. Det har uppstått ett nytt behov av översättningar av lägre kvalitetsnivåer. Översättningar av litet lägre kvalitet och i synnerhet av enklare stil utgör en ny typ av översatta texter som används av människor som vill veta det väsentliga innehållet i ett dokument så snabbt som möjligt. När PC-baserade översättnings-system blev billigare ansåg en del av användarna att de hellre kunde dra nytta av översättningar på litet lägre kvalitetsnivå än ingen översättning alls. Maskinöversättning har så småningom börjat hitta sitt ställe i kommunikationssituationer där en traditionell översättning inte kommer i fråga, t.ex. i översättning av www-sidor, e-post eller till och med elektroniska chat-listor. Maskinöversättning är redan oersättlig i system som behandlar stora informationsflöden t.ex. i sökverktyg. (Hutchins 1999.)

Kvaliteten av måltexten kan nuförtiden indelas i fyra kategorier:

- råöversättning
- översättning av normal kvalitet
- översättning av särskilt bra kvalitet
- omformulering (adaptation) av originaltext

Det centrala innehållet av källtexten kommer fram i en *råöversättning*. Den översatta texten måste vara förståelig men den kan dock ha skrivfel eller grammatiska fel. Typiska källtexter för råöversättning är till exempel vetenskapliga sammandrag. Den traditionella mänskliga översättningen hör till *översättning av normal kvalitet*. Måltexten är grammatiskt korrekt och relativ flytande fast stilen ibland kan vara stel. Tekniska manualer är ofta av denna typ. *Översättning av särskilt bra kvalitet* är både flytande och idiomatisk. Den har anpassats helt till målspråkets kulturella kontext. Reklamtexter och litteratur är typiska exempel av denna kvalitet. *Omformulering av originaltext* är inte en trogen översättning utan en ny text som har skrivits på basis av en eller flera källtexter. Den följer inte nödvändigtvis källtextens meningsstruktur men översättaren kan i stället ha omformulerat text eller lämnat bort någonting. (EAGLES 1995, <<http://www.issco.unige.ch/ewg95/node165.html>>.)

5.3 Lämpliga texter

Översättningsminnet passar för texter som på ett eller annat sätt återanvändas senare. Ett exempel på återanvändning är *uppdateringar*. Någon gång kan översättaren få en uppdaterad version av källtexten mitt under översättningsprocessen. Den konventionella uppdateringen och uppdateringen med ett översättningsminne skiljer sig från varandra. Med översättningsminnet behöver översättaren inte gå igenom alla segment utan programmet känner igen de bearbetade eller förändrade segmenten. Ett annat exempel på återanvändning är *revisioner*. Kunderna kan ha förändrat en manual bara lite för att den skall kunna tillämpas till en ny produktmodell och vill en ny översättning av denna manual så snart som möjligt. Översättningsminnet passerar segment som är identiska med den gamla versionen. Förutom uppdateringar och revisioner kan översättaren också utnyttja tidigare översatta dokument som översättningsminne om de avsevärt liknar det aktuella dokumentet. (Webb 1992.)

Fast återanvändning av texter är det viktigaste kriteriet på användbarheten av ett översättningsminne, ökar den relativa nyttan av översättningsminnet om texten innehåller mycket repetition. Repetitionen kan bestå av upprepade ord, fraser eller till och med hela kapitel. (Webb 1992.)

Maskinöversättningsprogram ställer högre krav på källtextens form om man strävar efter bra kvalitet. Översättningsresultatet försämras av lingvistiska problem såsom lexikal ambivalens, komplexa syntaktiska konstruktioner, lexikala skillnader i betydelse, ellipser och ogrammatiska konstruktioner (Maskinöversättning, <<http://stp.ling.uu.se/~fredrikg/maskinov.htm>>). Med andra ord borde texterna som skall översättas maskinellt vara syntaktiskt enkla och grammatiskt korrekta. Meningarna borde vara korta och tydliga. Den ideala situationen är att texten från början är planerad att översättas maskinellt. Då används ofta subspråk eller reglerade språk. Vanligen är texterna nog inte planerade för att översättas med maskin. Maskinöversättningens output kan i alla fall användas som hjälpmedel för översättare i form av en råöversättning.

6 BESKRIVNING AV ORIGINALTEXTER

De IBM-texterna som vi har som testmaterial befinner sig ursprungligen i form av en pdf-fil. När dessa texter överfördes till Microsoft Word, förlorade de största delen av sina redigeringar. Den mest påfallande förändringen var att tabellerna inte bevarade sin form utan bara texten utan ramarna flyttade till Word. Överhuvudtaget har texten i Word inga mellanslag mellan olika kapitel. Texten har dock bevarat fonterna, fetstilen och kursiveringen samt förteckningar. Vi har tagit 15 sidor av de båda två manualerna som vårt forskningsmaterial.

De två IBM-manualerna skilde sig mera beträffande sakinnehållet än vi trodde. Den ena manualen (Operator's Guide) innehöll anvisningar för användning och underhåll medan den andra (User's Guide) bestod av definitions- och installeringsanvisningar. Den förra texten hade flera långa och någorlunda komplicerade satser. Satserna var korta närmast bara i förteckningar där satsstrukturen i allmänhet verkade enklare. Den innehöll relativt många termer, t.ex. på skrivarens enhetsdelar. Den senare texten bestod till stor del av koncisa anvisningar som skulle utföras fas för fas. I denna text fanns det knappast några termer, med undantag av några ADB-termer. De två texterna hade sålunda inte samma termer. Detta innebär att vi inte skulle kunna testa det första lexikonet, som vi hade tänkt att skapa, i översättningsprocessen av den andra texten.

7 ANVÄNDAR-ORIENTERAD TESTNING

Vår synpunkt från början har varit att granska två hjälpmedel för översättare, TranSmart och Trados, med utgångspunkt i en frilansöversättares behov. För att undersöka programmets nyttighet har vi utfört flera tester på de båda programmen. Dessa tester kan inkluderas under termen *systematisk testning*. Den systematiska testningen syftar på all testning som har som mål att undersöka beteendet av programvaror i speciella förhållanden med vissa, väntade resultat (EAGLES 1995, <<http://www.issco.unige.ch/ewg95/node36.html>>).

I våra tester kombineras två typer av systematisk testning: *uppgiftsorienterad testning* och *benchmark-testning*. Med uppgiftsorienterad testning undersöks om programvaran uppfyller de uppgifter som har definierats på förhand. Det huvudsakliga målet är att med hjälp av relevant input utvärdera den allmänna funktionaliteten av programmen. Å andra sidan granskas

också kvaliteten av output. Typiskt för denna testtyp är att testerna inte misslyckas om de avbryts på grund av problem i utförandet av en uppgift. Man kan dokumentera de problem som uppstått och upprepa testuppgiften. Med benchmark-testning undersöks systemens prestationsförmåga. Den kan tillämpas på enstaka funktioner, på systemmodulerna eller på hela programmet. Ett exempel på benchmark-testning är att undersöka användbarheten av ett program, dvs. undersöka hur lätt det är att förstå och lära sig programmet. Resultaten av benchmark-testning anges ofta med tid eller antalet tryckningar på tangenter eller klickningar med mus som behövs för att utföra en uppgift. (EAGLES 1995, <<http://issco.unige.ch/ewg95/node36.html>>, <<http://issco.unige.ch/ewg95/node85.html>>, <<http://issco.unige.ch/ewg95/node157.html>>.)

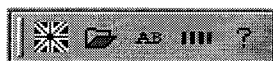
Det finns en standard, ISO 9126 Standard, som definierar kvalitetsdrag som lämpar sig för evaluering av programvaruprodukter. Standarden inkluderar sex sådana drag, som är avsedda att vara heltäckande. Härav följer dock av att de är omfattande som begrepp. (EAGLES 1995, <<http://www.issco.unige.ch/ewg95/node14.html>>.) Vi tillämpar dessa kvalitetsdrag direkt eller något anpassade för våra ändamål i evalueringen av maskinöversättningssystemet TranSmart och översättarens arbetsstation Trados. Vi rubricerar vår presentation av resultat med kvalitetsdrag som passar för våra ändamål. Under varje rubrik anger vi en ISO-standard som en ram för vår presentation. Vår evaluering av resultaten är till stor del klassificerande, till exempel svar på frågan hur bra ett program utför en uppgift. Också resultaten av benchmark-testningar anger vi med klassificerande värden. De klassificerande värdena är subjektiva, våra egna intryck av programmens nyttighet. Några resultat presenterar vi också kvantitativt.

8 TRANSMART

8.1 Användbarhet

A set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users.
(ISO 9126: 1991, 4.3)

En av fördelarna av TranSmart är dess användarvänlighet. Vid installering skapar programmet ett extra verktygsfält i Word, samt en TranSmart-meny. Det finns bara fem knappar i verktygsfältet, se nedan



Med de två huvudsakliga funktionsknapparna kan man låta programmet översätta ett dokument och sedan kan man hämta det översatta dokumentet till granskning. Den tredje knappen är avsedd för TranSmarts egna översättningsmarkeringar. Användaren kan välja mellan markeringar med olika färger eller streck. Med att trycka på den fjärde knappen medan TranSmart översätter kan översättningsprocessen följas. Vid den femte knappen finns det instruktioner. Det finns några ytterligare funktioner i TranSmart-menyn. Man kan t.ex. skapa och editera ett dokumentspecifikt lexikon eller ta bort översättningsmarkeringarna.

Själva översättningsprocessen är enkel och snabb. Man behöver bara i Word öppna det dokument som skall översättas. Sedan väljer man knappen Käännä för att översätta dokumentet. Ett fält öppnas där programmet frågar efter vissa val, t.ex. om markeringar skall visas med färg eller streck. Med klickning på OK börjar översättningsprocessen. När processen är slut, kan man hämta råöversättningen.

Såsom användning av TranSmart i allmänhet, är skapandet av ett dokumentspecifikt lexikon enkelt. Det finns en option för redigering av ett dokumentspecifikt lexikon i TranSmart-menyn. När det dokument som skall översättas är öppet, kan ett sådant lexikon skapas. Ett fönster öppnas. Ett ord tillsätts genom att skriva ordet och dess ekvivalent i följd, och därtill märket # tilläggas mellan dem. Efter det att lexikonet är färdigt, kan dokumentet översättas. Det dokumentspecifika lexikonet kan inte lagras eller exporteras.

Översättningsmarkeringarna påverkar programmets användbarhet i efterbehandlingsfasen. De framhäver vissa punkter i råöversättningen och underlättar så efterbehandlingen. De dubbla understrykningarna var de nyttigaste av TranSmarts översättningsmarkeringar. Med denna markering framhävs fall där TranSmart har försökt att översätta finska sammansättningar genom att översätta de två delarna skilda. TranSmart hade lyckats i drygt tio procent av dessa. Med att skapa ett dokumentspecifikt lexikon kunde vi minska ungefär 55 % av dubbelunderstrykningarna. Programmet klarade inte av de sammansättningar som bestod av tre eller flera

delar, om de inte fanns i det allmänna lexikonet, såsom finska *piirtoheitinkalvo* 'transparang' (transparency).

Om ett ord har lämnats orört, märker översättaren det lätt därför att det är överstruket. Största delen av ord som var överstrukna i vårt material bestod av vissa, ofta förekommande initialförkortningar (IBM, CD). Nummerserierna var också överstrukna, såsom vissa finska ord, som hade lämnats orörda i texten. Sådana orörda ord var t.ex. sammansättningar med tre ord (t.ex. *sarjanumerokilpi* 'serienummerskylt'), samt finska böjningsformer som programmet inte kände igen (t.ex. *lailla* 'med lag'). Med understreckning hade markerats ord som var på engelska redan i den finska originaltexten (t.ex. Network Color Printer, QuarkXpress, Select [knapp]). Antalet ord som var understrykna och överstreckade kunde vi knappt minska med hjälp av det dokumentspecifika lexikonet.

Det fanns många fall där det var svårt att förstå varför programmet hade markerat ord med överstrykning eller understreckning. För att ge ett exempel, hade programmet överstrukt initialförkortningen IBM medan det hade understreckat initialförkortningen CMYK (cyan-magenta-yellow-black). Skillnaden är också oklar i markeringar på program, t.ex. QuarkXPress är understreckat medan QuarkXPress 3.2 är överstruket. Dessa skillnader kan inte klassificeras med hjälp av programmets egna förklaringar för markeringarna (känner inte som ett finskt ord - överstrykning, hittar inte i sitt lexikon - understreckning). I alla fall kunde man minska dessa slags översättningsmarkeringar om man hade alla lexikonmöjligheterna i bruk. Sådana ord kunde inkluderas i ett användarspecifikt lexikon.

8.2 Funktionalitet

A set of attributes that bear on the existence of a set of functions and their specified properties. The functions are those that satisfy stated or implied needs. (ISO 9126: 1991, 4.1)

Under begreppet funktionalitet behandlar vi två kvalitetsdrag, exakthet och lämplighet. Vi behandlar dessa drag i separata avsnitt, i vilka vi presenterar översättningskvalitet och grubblar på TranSmarts lämplighet som hjälpmedel för frilansöversättare.

8.2.1 Exakthet

Attributes of software that bear on the provision of right or agreed results or effects. (ISO 9126: 1991, A.2.1.2)

För att undersöka översättningskvaliteten av TranSmarts råöversättningar tog vi ett urval av meningar från våra IBM-manualer. Vi klassificerade meningarna i fyra kategorier: enkla meningar, komplexa meningar, meningar med citattecken och meningar med inskott. Med enkla meningar syftar vi på meningar som inkluderar endast en eller två satser. Komplexa meningar innehåller däremot flera än två satser i vår klassificering. Förutom enkla och komplexa meningar tog vi fram i vårt urval två fall som vi antog välla svårigheter för ett maskinöversättningsprogram: För det första undersökte vi fraser i citattecken (t.ex. avsnittsrubriker). För det andra intresserade vi oss för meningar i vilka det finns en hel fras eller sats mellan subjekt och predikat, dvs. ett inskott.

Antalet meningar försökte vi definiera i förhållande till deras förekomst i texten. Med andra ord tog vi flest enkla meningar (20 stycken) och litet färre komplexa meningar (10 stycken) samt minst meningar som hör till de två sistnämnda kategorierna (5 stycken av båda). Vi tog samma urval från de båda manualerna, dvs. sammanlagt 80 meningar. Dessa meningar valde vi arbiträrt ur texterna, till exempel var tjugonde enkla mening.

Vi evaluerade dessa meningar med hänsyn till deras förståelighet. Förståeligheten indelade vi i en skala av tre alternativ: helt förståelig, otydlig och obegriplig. Kategorin *helt förståelig* förutsatte inte att meningen var grammatiskt korrekt. Meningen borde förstås rätt på basis av råöversättningen utan att man behöver gå tillbaka till originaltexten. Meningarna som evaluerades som *otydliga* stod emellan helt förståelig och obegriplig. Denna kategori inkluderade också fall där våra avvägningar skilde sig från varandra. *Obegripliga* meningar kunde inte alls förstås på basis av råöversättningen.

Ett överraskande problem i originaltexternas redigeringar påverkade betydligt översättningsresultatet. Texterna såg helt normala ut till en början men de översatta meningarna var konstiga. Sedan märkte vi att problemen berodde på radbyten. Se till exempel

1. Viittaukset IBM:n lisensoituihin [radbyte]
 tuotteisiin, ohjelmiin tai palveluihin eivät tarkoita, että vain niitä voi-
 daan [radbyte]
 käyttää.
Motions IBM-licenced [radbyte]
for the products, programmes or services not only they mean that it is
could have [radbyte]
one uses.

2. Kun lähetät tai siirrät työn [radbyte]
 pidätysjonoon, kirjoitin ottaa sen vastaan välittömästi ja tallentaa sen
 kirjoittimen [radbyte]
 umpilevyyn.
When you send or move the work [radbyte]
the printer receives it immediately and records its printer to the
detention queue [radbyte]
to hard disk.

Det ser ut som om TranSmart behandlade texten efter radbytet som ett nytt segment och till och med som ett nytt kapitel och därför översatte det separat. De konstiga redigeringarna kan ha förekommit under överföringsprocessen från pdf-fil till Word-fil.

Efter att ha hittat en förklaring till de misslyckade översättningsresultaten, korrigerade vi redigeringarna och lät TranSmart översätta den korrigerade texten på nytt. Sedan indelade vi meningarna med vår evalueringsskala på basis av deras förståelighet. Nästan 60 % av alla meningstyper var helt förståeliga, dvs. att man kunde förstå dem rätt utan att behöva se tillbaka på originaltexten. Här räknade vi inte med små inexaktheter till exempel i artikelbruket, som bör dock alltid kontrolleras i originaltexten. I allmänhet var kvaliteten av dessa meningar utmärkt.

3. Muissa käyttöympäristöissä saadut tulokset voivat olla hyvinkin erilaisia.
The results that have been obtained in other usage environments can be very different.

4. Huolehdi riittävästä tuuletuksesta ja jäähdytyksestä.
Take care of a sufficient airing and cooling.

Däremot var ungefär 25 % av de valda meningarna obegripliga. Oftast var ordföljden i dem sådan att man inte alls kunde förstå dem eller man förstod dem felaktigt. Sådana problem förekom i mer än hälften av meningarna med citattecken (7/10), i hälften av meningarna med inskott (5/10) och i de komplexa meningarna (4/20):

5. Luvussa 8, ”Apuohjelmien käyttö”, kuvataan, miten käytetään IBM Network Color Printer Spooler- ja Downloader-apuohjelmia.

Chapter 8 ”” shows the use of Apuohjelmas how IBM-is used the Color Printer Spooler and Downloader utility programs Network.

6. Älä pidä kirjoitinta helposti syttyvien materiaalien, kuten verhojen ja kemikaalien, lähellä.

Do not hold a printer, the materials, such as curtains and chemicals, near which light easily.

7. Kun kohdistusmerkki on kalvon oikeassa yläkulmassa, tasaa pino kalvoja tasaisella alustalla siten, että kulmat ovat suorassa.

When a setting right sign is at the right top corner of the membrane, a pile of membranes will level with an even base so that the corners will be suorassa.

TranSmart kunde inte alls förstå citattecken som man också kan se i exempel 5. Det är överraskande att texten inom citattecken inte hölls ihop av programmet. TranSmart behandlar inte heller denna företeelse konsekvent utan citattecken kan finnas var som helst. Detta gör det svårt för en översättare att med tiden lära programmets sätt att behandla motsvarande fall, vilket skulle underlätta förståeligheten. I exempel 6 har TranSmart förväxlat ordföljden så att referensrelationerna har blivit otydliga. Om man tänker på strukturen i den finska meningen, är det lätt att förstå varför ordet *near* har placerats efter inskottet. Den engelska meningen följer strikt ordföljden i den finska originaltexten. Det är däremot konstigt att den relativa satsen *which light easily*, har ryckts lös från sitt huvudord och placerats separat i slutet av meningen. Detta är oväntat i synnerhet därför att programmet har lyckats så bra att bilda en relativsats av den finska frasen *helposti syttyvien*, som är en bestämning till huvudordet *materiaalien*. Exempel 7 är ett typiskt exempel på ambiguitetsproblem. TranSmart har missförstått ordet *kalvo* (transparency; membrane) därför att man har använt den förkortade formen av ordet *piirtoheitinkalvo* ’transparang’. Också ordet *tasaa* ’jämna ut’ vållade svårigheter. Det kan översättas med imperativ såväl som med futurum eller presens. I detta exempel borde ordet ha översatts med imperativ.

Drygt 15 % av meningarna var på något sätt otydliga. Oklarheten uppstod ofta på grund av enstaka ord, som programmet hade översatt felaktigt:

8. Kun se palaa, kirjoittimeen on kytketty virta.

When it returns, a current has been switched on to the printer.

9. Tässä luvussa kerrotaan, kuinka ohjaustaulu toimii ja kuinka näyttössä olevia valikoita käytetään.

In this number it is told how the console operates and menus how in the display are used.

I exempel 8 är den enda oklara punkten verbet *palaa* 'brinner'. Det finska verbet kan också betyda 'kommer tillbaka' och därför har ordet översatts med det engelska verbet *return*. När man känner till denna typ av ambiguitetsproblem av maskinöversättningsprogram, kan man dock komma till den rätta betydelsen. Detsamma gäller exempel 9 där TranSmart har översatt det mångtydiga finska ordet *luku* med det engelska ordet *number* i stället för *chapter*. Å andra sidan liknar översättning av bestämningen *näytössä olevia* översättning av bestämning i exempel 6. I principen har programmet lyckats att översätta satsdelarna men det har placerat dem i en ordning som varken följer finskans eller engelskans ordföljd.

Skillnaderna i originaltexten syntes i någon mån i vårt urval av meningarna. De översatta meningarna i *User's Guide*, som vi antog vara strukturellt enklare, var förståeliga oftare än de i *Operator's Guide*. I den förstnämnda manualen var 17 av de 20 enkla testmeningarna helt förståeliga medan endast 14 av 20 i den sistnämnda manualen. I *User's Guide* fanns det också färre komplexa meningar (33) än i *Operator's Guide* (50).

8.2.2 Lämplighet

Attribute of software that bears on the presence and appropriateness of a set of functions for specified tasks (ISO 9126: 1991, A.2.1.1)

Ur frilansöversättarens synvinkel råder det brist på användarvänlighet beträffande efterbehandling. Efter det att råöversättning är färdig, är dokumentet ett vanligt Word-dokument. Det finns inga funktioner i TranSmart för att underlätta efterbehandlingen. Den enda möjligheten att bearbeta den översatta texten är att skriva på den. Ifall en mening var obegriplig, skulle man gärna gå tillbaka till originaltexten för att jämföra dess satsekvivalent. Man kan nog öppna både originaltexten och råöversättningen i Word men det är besvärligt att hantera texterna på detta sätt. Särskilt svårt är det att hitta de motsvarande punkterna i texterna.

I standardversionen av TranSmart finns det möjlighet att skapa endast ett extra lexikon. Därför att Kielikone rekommenderar den professionella versionen för frilansöversättare, skulle vi

gärna ha prövat också andra lexikon (t.ex. ett användarspecifikt lexikon och ett domänspecifikt lexikon). Principen är nog densamma i hierarkin mellan alla lexikon, dvs. att programmet söker ord först i det lexikon som står först i hierarkin (se avsnitt 4.2.3). Eftersom vi inte hade andra lexikon i vår version av TranSmart, skapade vi ett dokumentspecifikt lexikon. Vi byggde detta lexikon i syfte av att minska antalet TranSmarts egna översättningsmarkeringar.

Det dokumentspecifika lexikonet förbättrade översättningskvaliteten och antalet översättningsmarkeringar minskades. Standardversionen är dock opraktisk därför att detta lexikon inte kan återanvändas. Om standardversionen inkluderade ett lika enkelt lexikonsystem som kunde lagras också för senare bruk, skulle frilansöversättare kunna använda det som hjälpmedel i sitt arbete. Det skulle vara nyttigt om man kunde skapa flera sådana lexikon enligt olika ämnesområden. Med hänsyn till en frilansöversättares resurser skulle en sådan version vara lämpligare som hjälpmedel i översättningsarbetet än TranSmart Professional med lokalisering.

9 TRADOS

9.1 Användbarhet

A set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users.
(ISO 9126: 1991, 4.3)

Vid Trados behandlar vi begreppet användbarhet under två underkategorier: förståelighet och lärbarhet. Under förståelighet kommenterar vi hur lätt eller svårt det är att förstå logiken i programmets funktioner. Under lärbarhet beskriver vi funktionerna och hur lätt eller svårt det är att lära sig använda dem. Vi behandlar dessa i separata avsnitt.

9.1.1 Förståelighet

Attributes of software that bear on the users' effort for recognizing the logical concept and its applicability. (ISO 9126: 1991, A.2.3.1)

Trados är ett oerhört komplex system. Trados inkluderar flera självständiga program som är kompatibla med varandra. Varje program är ett avancerat system i sig. Det tar relativt lång tid innan man begriper hur dessa delprogram fungerar separat och tillsammans med varandra.

Vi började med att bekanta oss med vissa delprogram för sig. För det första försökte vi öppna själva översättningsminnet Workbench separat från Start-menyn. Ett fönster öppnades där det fanns två tomma fält. Vi kunde inte öppna någonting från menyerna i detta fönster, inte heller kunde vi skriva någonting på fälten. Till slut gav vi upp och försökte i stället att använda MultiTerm som vi också öppnade från Start-menyn. Detta lyckades vi inte heller att få att fungera. För det första kunde vi inte välja finska som källspråk och engelska som målspråk. Engelskan fanns där som antagande men finskan hittade vi inte och kunde inte heller tillägga den till språklistan. Vi hade läst att demoversionen av Trados fungerar med endast fem språk som bör väljas vid installering. Därför antog vi att språken borde ha valts vid installering också för MultiTerm. Vi kom dock inte ihåg att ett sådant alternativ förekom vid installeringen.

Vi kunde inte behärska delprogrammen separat så därför kom vi fram till att försöka hantera dem från Workspace. Workspace är ju ett verktyg som är avsett för hantering av översättningsprojekt. På basis av teorin antog vi att vi kunde öppna en Word-fil från Workspace med de två centrala verktygen Translator's Workbench och MultiTerm. Men även om vi redan hade bekantat oss med Workspace i teorin, insåg vi inte fullständigt dess funktionsprincip. Det såg ut som om Workspace endast var ett anteckningsverktyg där en översättare skulle kunna lagra information om ett visst projekt, t.ex. dess namn, allmän beskrivning och deadline. Vi var ändå säkra på att alternativet Översätt med Word, med vilken vi skulle kunna börja översätta vårt testmaterial, borde finnas fast vi inte hittade det.

9.1.2 Lärbarhet

Attributes of software that bear on the users' effort for learning its application (for example, operation control, input, output). (ISO 9126: 1991, A.2.3.2)

Efter tiotals frustrerande försök lyckades vi äntligen öppna översättningsminnet Translator's Workbench tillsammans med vår källtext, men inte genom Workspace. Detta hände när vi en

gång öppnade Word och märkte att det hade uppstått ett extra verktygsfält med Trados funktionsknappar. Sedan öppnade vi också Translator's Workbench och efter det var fortsättningen lätt på basis av den teoretiska informationen som vi hade om översättning med programmet. Verktygsfältet av Trados ser så här ut:



Man börjar med den första knappen, som öppnar det första käll- och målfältet. Källfältet är blått och målfältet är gult. På det gula fältet översätter man det aktuella segmentet. När ett segment har översatts, godkänns det med den sjätte knappen. Denna knapp är den mest använda knappen under översättningsprocessen. Den lagrar översättningen till minnet, stänger fältet och öppnar nästa segment samt söker efter motsvarighet i minnet. Om man av någon orsak inte vill lagra segmentet i översättningsminnet (när t.ex. ett segment har översatts fritt), kan man välja bara att stänga fältet med den nionde knappen. Då kan man öppna det nästa segmentet med den andra knappen i Trados verktygsfält, som också söker efter översättning i minnet. En nyttig funktion är kopiering av källsegmentet till målfältet med den femte knappen, som kan användas om ett källsegment kan överföras oförändrad. Ytterligare finns det några andra funktioner, t.ex. för att man skall kunna söka termer (pilar med en hakparentes över dem) och överföra oföränderliga element såsom grafik (pilar med en hakparentes under dem).

Efter det att vi hade lyckats att pröva hur programmet fungerade i praktiken, återkom vi till problemet med Workspace. Vi insåg hur viktigt organiseringen skulle vara för en professionell översättare som ständigt får nya översättningsuppdrag och som möjligen samtidigt har flera översättningsprojekt på gång. Därför försökte vi hitta mera detaljerade instruktioner om användningen av Workspace i dess hjälp-meny. Vi hittade bara enskilda instruktioner, däremot hittade vi inte heltäckande information om hur man går framåt fas för fas i skapande och hantering av ett projekt.

Vi lyckades till slut genom försök och misstag. Vi hade alltid stannat på ett ställe som verkade vilseledande. Det vilseledande fönstret heter *Add Source Component* 'tillägg källkomponent'. Där kan man tillägga information om källkomponenten, dvs. namn, beskrivning och deadline. Eftersom man inte kunde bläddra och söka efter filer, som man på tidigare punkter hade kunnat göra, hade vi inte fyllt i dessa fält. Vi tänkte att om man inte kan bläddra och välja en

fil, är fönstret avsett bara för anteckningar. Vi gav upp och försökte att gå vidare från ett annat håll därför att det alltid fanns flera alternativ i varje fönster att fortsätta. Men det var just det fönster där man borde ha fortsatt, med att fylla i den relevanta informationen och trycka på OK.

Hjälpmenyer i MultiTerm och WinAlign visade sig inte nyttiga för en nybörjare. Där fanns inga praktiska instruktioner om dessa programs användning fas för fas. Hjälp-menyn är strukturerad enligt meny-funktionerna och varje alternativ har definierats. Denna slags strukturering är praktisk och funktionell för en användare som redan känner till programmen tillräckligt bra. Användaren kan då stöda sig på instruktioner vid behov. Nybörjarna går däremot lätt vilse om de försöker att lära sig använda programmen endast med hjälp av hjälpmenyerna. Förutom hjälp-menyerna utnyttjade vi instruktionsböcker som befinner sig på samma www-sida där en demoversion av Trados kan laddas ner. Det bästa och effektivaste sättet att lära sig använda Trados enligt våra erfarenheter var dock genomgången av instrueringsdemonstrationer. Vi prövade två demonstrationer, en för MultiTerm och en annan för WinAlign. Dessa går igenom alla faser med diabilder för att man kan lära sig grunderna.

Efter att ha följt instruktionerna i instrueringsdemonstrationen av MultiTerm, var det lätt att skapa ett litet lexikon i detta delprogram. Demonstrationen visade steg för steg hur man tillägger termer i MultiTerm samt hur olika optioner definieras och uppdateras. När vi hade gjort lexikonet färdigt hade vi ingen aning hur det borde användas. Vi öppnade vårt översättningsprojekt genom WorkSpace med Translator's Workbench och MultiTerm. När vi översatte texten vidare som vanligt, föreslog MultiTerm inga termer någonstans såsom vi trodde det borde göra. Vi måste titta på demon om Translator's Workbench som förklarade mycket noggrant vad vi inte hade förstått. Vi måste nämligen titta på Workbench-fönstret där programmet visade det aktiva segmentet. Termer som MultiTerm föreslog, visades med röd strykning över ordet och till höger såg man översättningsförslaget. Med en klick på knappen *Get current term* (den näst sista knappen i Trados verktygsfält) kunde man överföra termen i det aktiva segmentet. Ett problem var att Workbench-fönstret inte automatiskt stod på skärmen med den aktuella översättningen utan man måste välja en viss option för att kunna se fönstret hela tiden. Över huvudtaget verkade MultiTerm ganska besvärligt att lära och använda, och användningen var besvärligt, åtminstone i början.

WinAlign lyckades vi nästan använda på egen hand. Vi missade bara en funktion från ett fönster. Med denna funktion borde man ha valt alternativet *Align file names* med vilken man hade kunnat sätta de käll- och målspråkiga filerna att motsvara varandra. Efter det att vi hade gått igenom demon om WinAlign kunde vi använda detta program. Man hade tre alternativ av vilka man måste välja ett på basis av vilket alignment utfördes: stilar, rubriksnumrering och storleken av font. I vårt testmaterial fanns inga stilar och ingen rubriksnumrering, så vi måste välja storleken av font som bas. Det var troligen inte det bästa alternativet och alignment lyckades inte perfekt. Programmet kopplade de motsvarande segmenten ihop med streck. Det fanns dock skillnader mellan originaltexten och dess översättning på ett par sidor i början. Vi började korrigera alignment manuellt vilket var besvärligt. Man kan dock uppdatera alignment efter varje segment som man har korrigerat.

I allmänhet är Trados ett mycket komplicerat system. Antalet funktioner och alternativ som finns i alla delprogram är enormt. Redan inläringen av basfunktionerna av enskilda program är tidskrävande, och därtill måste man kunna hantera dem från WorkSpace. Vi lärde oss grunder av Translator's Workbench, MultiTerm, WinAlign och WorkSpace under den tid som vi utförde individuella tester. I alla fall är det fortfarande svårt att bilda en helhetsbild av Trados och komma ihåg alla steg efteråt. Man måste ta sig tillräckligt tid för inläring av programmet efter dess anskaffning. Trots en grundlig inläring kan man inte förvänta sig att genast kunna använda det effektivt i översättningsarbetet. Dess relativ nytta ökar så småningom.

9.2 Effektivitet

A set of attributes that bear on the relationship between the level of performance of the software and the amount of resources used, and stated conditions. (ISO 9126: 1991, 4.4)

Effektiviteten av WorkSpace kan förstås genast när man lär sig känna dess funktionsprincip. För det första blir projektkontrollen på ett naturligt sätt organiserad om man använder WorkSpace. Allt material på ett projekt kan lagras under ett namn där man hittar dem utan ansträngning och man behöver inte erinra sig var man har lagrat filerna som hör ihop. För det

andra kan man öppna ett projekt snabbt med alla verktyg som behövs i det. Det lönar sig att ordentligt bekanta sig med WorkSpace redan från början.

På grund av restriktionerna i demoversionen kunde vi inte få någon helhetsbild av samarbetet mellan MultiTerm och Translator's Workbench. Hundra översättningssegment räcker inte för att se nyttan av MultiTerm i förhållande till det arbete som krävs för att skapa lexikonet. Både Translator's Workbench och MultiTerm var ju tomma när vi började. I början kräver MultiTerm extra tid och i själva verket gör översättningsarbetet långsammare. Man måste välja de önskade inställningarna, t.ex. språk och ämnesområde. Om man börjar skapa lexikonet medan man översätter, betyder det att man först måste tillägga nya termer i MultiTerm oftare än det ger förslag. Så småningom ökar nyttan när MultiTerm börjar bli heltäckande och man behöver inte tillägga termer annars än då och då. På basis av våra tester är det svårt att estimera hur lång tid det skulle ta innan MultiTerm når nivån i vilken MultiTerm kan användas effektivt.

Effektiviteten av WinAlign beror på formen av dokument. När originaltexten och översättningen noggrant motsvarar varandra, fungerar WinAlign utmärkt och snabbt. Å andra sidan minskar effektiviteten radikalt om WinAlign misslyckas att känna igen överensstämmelsen mellan käll- och måltexten. Problem uppstår om i den översatta texten har tillagts eller borttagits någonting, eller om några meningar har indelats i mindre enheter eller kombineras. Om ett fel uppstår betyder det nog inte nödvändigtvis att alla följande segment är felaktigt kombinerade utan programmet kan återställa den rätta uppställningen senare. Lyckade resultat kan befrämjas med olika alternativ, t.ex. med att välja uppställning på basis av stilar, rubriksnumrering eller storleken av font. Det lönar sig inte att korrigera texterna mycket i WinAlign utan det är förnuftigare att göra de ändringar som behövs utanför detta verktyg. Det är nämligen besvärligt och tidskrävande att manuellt bearbeta alignment i WinAlign.

9.3 Funktionalitet

A set of attributes that bear on the existence of a set of functions and their specified properties. The functions are those that satisfy stated or implied need. (ISO 9126: 1991, 4.1)

Under begreppet funktionalitet behandlar vi två kvalitetsdrag, exakthet och lämplighet. Vi behandlar dessa drag i separata avsnitt. Under exakthet betraktar vi förekomsten av matchningar och användbarheten av fuzzy match-förslag. Under lämplighet grubblar vi på lämpligheten av Trados som hjälpmedel för frilansöversättare.

9.3.1 Exakthet

Attributes of software that bear on the provision of right or agreed results or effects. (ISO 9126: 1991, A.2.1.2)

Ett av våra forskningsobjekt var att undersöka upprepning i vårt testmaterial. Med andra ord räknade vi antalet matchningar som Translator's Workbench föreslog. I denna test hade vi inget färdigt översättningsminne utan programmet började skapa ett från det första översättningssegmentet som vi lagrade till dess minne. I vår test med en manual för en IBM-skrivare föreslog Translator's Workbench i genomsnitt fyra matchningar per sida. Hälften av matchningarna var 100 % ekvivalenter, hälften var av typen fuzzy match med 70 % som tröskelvärde.

De 100 % matchningarna bestod av enskilda ord (rubriker), fotnoter eller hela meningar. Manualen innehöll installeringsanvisningar och därför förekom det några identiska meningar i texten. Med 70 % som tröskelvärde föreslog Trados olika typer av fuzzy matchningar. För det första förekom fuzzy matchningar i förteckningar, som normalt inte är fullständiga satser:

1. Paperin valinta sivu 2-2
Choosing paper page 2-2
2. Paperin säilytys sivu 2-5 → Fuzzy match: *Choosing paper page 2-5*
Storing paper page 2-5

I detta exempel är verben olika. Translator's Workbench har dock substituerat sidnumren automatiskt.

Ett annat exempel på fuzzy match är rubriker som upprepas i fotnoter. Fotnoten inkluderar också ett sidnummer vilket förorsakar att matchningen inte är 100 % motsvarighet. Installeringsanvisningar består någon gång av liknande faser:

3. Kuvaruutuun tulee Connect-ikkuna.

The Connect window appears.

4. Kuvaruutuun tulee Select-ikkuna. → Fuzzy match: *The Connect*

The Select window appears. *window appears.*

I detta exempel måste man bara byta ut ordet *Connect* mot ordet *Select*.

Restriktionen i antalet översättningssegment i demoversionen av Trados förorsakade dock att våra resultat i undersökningen av upprepning i en text inte i något avseende är allmängiltiga. De ungefär hundra översättningssegment som man kan översätta med en demoversion motsvarade i vår test drygt fyra sidor.

9.3.2 Lämplighet

Attribute of software that bears on the presence and appropriateness of a set of functions for specified tasks (ISO 9126: 1991, A.2.1.1)

Även om demoversionen naturligtvis är avsedd bara för att man kan bekanta sig med programmets allmänna funktionsprincip, ville vi med testning av upprepning i en text få en aning om hurdana texter lämpar sig att översättas med Trados. Det såg ut som om antalet matchningar är relativt litet i User's Guide som vi använde i testningen av upprepning. Men på basis av analys av originaltexter verkade det som om den andra manualen innehöll ännu mindre upprepning. De två manualerna liknade inte heller så mycket varandra beträffande innehållet eller termer fast båda av dem var manualer av samma IBM-skrivare. Om en text inte innehåller mera upprepning än vårt testmaterial lönar det sig inte att använda Trados. Programmet kan till och med göra översättningen långsammare än traditionell översättning utan programmet på grund av besvär i användningen. Man måste öppna och lagra segment samt eventuellt tillägga termer och frasdelar. Trados passar bäst till översättning av uppdateringar och revisioner, dessutom måste man ha den tidigare versionen som översättningsminne. Verksamheten av en frilansöversättare måste vara relativt omfattande för att nyttan, besvär och kostnader möter varandra.

Om verksamheten är tillräckligt omfattande, översättningsminnen kan användas effektivt och texterna innehåller repetition, ökar effektiviteten av översättning säkert. På basis av vår un-

dersökning av användbarheten av Trados och testningar med vårt material kan vi dock inte utvärdera i vilken mån dessa faktorer måste förekomma för att användning av Trados lönar sig.

10 DISKUSSION

Vi försöker i detta avsnitt evaluera om en frilansöversättare kan dra nytta av de program som vi har testat. Vi försöker också jämföra programmens användbarhet. Alla våra testresultat är subjektiva. Vi kommer att värdera båda programmen ur vår egen synvinkel, till exempel vilketdera programmet vi själva antar passa bättre för frilanssysselsättning, om något alls. Programmen kan dock inte jämföras likvärdigt därför att TranSmart inte alls har planerats för frilansöversättare och därför passar dess funktioner inte idealt till frilansöversättning. Trados däremot har en frilansversion av programmet.

Vi har utfört ett antal tester med vart och ett av våra testprogram. Våra tester kan verka lite oenhetliga eller spridda och de ger kanske inte någon helhetsbild av programmens användning och möjligheter som hjälpmedel för en frilansöversättare. Vi testade programmens användning under en relativt kort period och dessutom hade vi bara två dokument att översätta. Det skulle kräva en längre undersökningsperiod för att man närmare skulle kunna sätta sig in i funktionerna och effektiviteten av dessa två program. Man skulle också gärna göra egentligt översättningsarbete under undersökningen för att riktigt få reda på programmens nyttighet i yrkesöversättningen.

Ytterligare hade vi inte i vårt pro gradu-arbete fullständiga professionella versioner av programmen. I TranSmart hade vi endast ett allmänt lexikon och ett dokumentspecifikt lexikon i bruk. Normalt lokaliserar programmet dock nästan alltid vid inköpet och då tillsätts användar- och domänspecifika lexikon. Vi hade i bruk en demoversion av Trados och det fanns också restriktioner i den. För det första kunde vi översätta endast 100 översättningsenheter med programmet, vilket vållade många slags problem. För det andra kunde resultaten av WinAlign inte lagras eller exporteras och därför kunde vi endast experimentera med funktioner i WinAlign.

Jämförelsen av termlexikonen är inte direkt möjligt på grund av deras olikhet. I TranSmarts lexikon har det strävats efter enkelhet och snabbhet medan med Trados MultiTerm har det satsats på nyttan på lång sikt med att försöka göra lexikonet heltäckande och mångsidigt. Båda lösningarna är lyckade och de lämpar sig utmärkt för det ändamålet de är avsedda. Enkelheten i TranSmarts lexikon är naturlig därför att man skapar lexikonet egentligen för en maskin. Å andra sidan är termlexikonet i Trados avsett för en människa, då till exempel möjligheten att tillägga definitioner och andra noter är välgrundad.

Vi skapade ett dokumentspecifikt lexikon i TranSmart. Själva skapandet var mycket enkelt och lätt. Det måste man naturligtvis göra på förhand. En frilansöversättare skulle kanske hellre skapa lexikonet så småningom under översättningsarbetet. Ambiguitetsproblem skulle underlättas om programmet föreslog termerna och lät översättaren godkänna dem. Då borde hela översättningen dock ske i segment.

Trados har ett bra termlexikon som passar också för frilansöversättarens behov. Lexikonet föreslår termer och låter översättaren bestämma om termen överförs till texten eller inte. Användaren kan själv tillsätta ord i termlexikonet också under översättningsarbetet samt tillägga förklaringar, kommentarer eller till och med exempelmeningar. Det är alltså relativt jobbigt att skapa ett bra och inventiöst termlexikon, men med tiden kommer det säkert att löna sig.

Med tanke på TranSmart kom vi till det resultatet att en frilansöversättare kanske inte har så stor nytta av att programmet är mycket snabbt. Naturligtvis är snabbheten mestadels en fördel men en översättare skulle säkert hellre ha mera interaktion under översättningen i stället för att programmet först översätter dokumentet i sin helhet efter vilket det efterbehandlas. Med andra ord skulle översättaren kanske föredra att översättningsarbetet sker i segment och inte genast på en gång som med TranSmart.

I översättning med segment kunde översättaren ta del i arbetet och på det sättet minska behovet av mycket grundlig efterbehandling. Det är besvärligt att efterbehandla dokument som har översatts med TranSmart: om till exempel en översatt sats var obegriplig, skulle översättaren vara tvungen att gå tillbaka till originaltexten för att kunna korrigera satsen. Detta är svårt därför att två dokument inte ryms bra på skärmen samtidigt. Ursprungligen hade Kielikone en version av programmet som var konstruerad i syfte av att i synnerhet efterbehandling skulle gå så behändigt som möjligt. Å andra sidan fungerade det inte i Word och det var mera kom-

plicerat i många avseenden. De stora pilotkundföretagen ville inte skapa och redigera dokument i en omgivning och sedan översätta och efterbehandla dem i en annan. (Hyvönen & Arnola 1998.)

För att ett maskinöversättningssystem som TranSmart skall nå en optimal nivå i råöversättning ställs det vissa krav på de texter som skall översättas. Meningarna borde vara grammatiska och de borde bestå av syntaktiskt enkla konstruktioner. Lexikala eller syntaktiska ambiguiteter borde inte förekomma. Dessa krav borde iakttas om man försöker avväga om det lönar sig att köpa TranSmart.

Vi blev inte övertygade om lämpligheten av Trados för vilken som helst översättare. För att effektiviteten av Trados skall synas, ställs det höga krav på sysselsättningens kvalitet. För det första måste sysselsättningen vara tillräckligt omfattande. Trados (såsom TranSmart, om det anskaffas för yrkesmässig användning) är ett relativt dyrt program, vilket innebär att antalet uppdrag eller volymen av individuella uppdrag måste täcka kostnaderna av programmets anskaffning. För det andra krävs det mycket tid för att man skall kunna lära sig använda Trados på ett övergripande och effektivt sätt. Varje delprogram har flera funktioner och alternativ som kan väljas. Åtminstone för en nybörjare är det ofta inte så klart vad som händer med varje val och vilka optioner man nödvändigtvis måste välja. En ordentlig orienteringsperiod måste räknas med i kostnaderna. Det är sannolikt att den normala arbetseffektiviteten under denna period går ner. För det tredje ställs det också krav på de texter som skall översättas. Den ideala situationen skulle vara om man fick översätta revisioner och uppdateringar av vilka man hade färdiga översättningsminnen. I alla fall borde texterna likna varandra så mycket att översättningsminnet kunde användas effektivt, dvs. att åtminstone förekomsten av fuzzy matchningar var rimlig proportionellt sett.

Trados är lämpligt för yrkesmässig översättning som sådant om kraven på dess effektiva användning fylls. En nackdel med Trados är dock att det är svårt att behandla programmet för det är mycket komplicerat. Till exempel skapar det tiotals nya filer under hantering av projekt och under själva översättningen. Man vet inte vilken uppgift dessa filer fyller. Därför känns det som om man borde vara expert i både lingvistik och datateknik för att ordentligt kunna förstå detta program. TranSmart däremot är användarvänligt. Att lära sig använda TranSmart kräver inte någon långvarig orienteringsperiod utan basfunktionerna är få och kan läras in

ganska snabbt. Till skillnad från Trados passar TranSmart inte som sådant till yrkesmässig översättning. Åtminstone efterbehandling borde underlättas.

SAMMANFATTNING

Vi har i teoridelen gått närmare in i maskinöversättningens utveckling ända från början till nutid. Vi har försökt ta upp de mest betydande prestationerna i maskinöversättningens utveckling därför att vi har velat framhäva betydelsen av historien i utvecklingen av de högkvalitativa maskinöversättningssystem och översättningsminnen som finns i dag. Vi har dessutom presenterat de viktigaste typerna av maskinell översättning: både automatiska översättningssystem och hjälpmedel för mänskliga översättare.

I analysdelen har vi testat två program i praktiken. Vi intresserade oss för maskinöversättningens aktuella tillstånd beträffande dess användning i översättningsarbetet. Vi har velat jämföra ett maskinöversättningssystem med ett översättningsminne. Med en sådan jämförelse har vi försökt utvärdera maskinöversättningssystemens potential som ett hjälpmedel för frilansöversättare vid sidan om översättningsminnen som redan finns i bruk inom detta användningsområde. Vi har evaluerat om programmen av denna typ kan bli ett avsevärt hjälpmedel i frilansöversättningen. Med andra ord har vi försökt ta reda på om maskinöversättningen möjligen kunde konkurrera med översättningsminnen som redan nu används i syftet att underlätta översättningsarbetet.

Vi skulle ha velat undersöka ytterligare många drag i båda programmen. Olika lexikon i TranSmart som man kunde få genom lokalisering av programmet skulle intressera oss. Vi skulle vilja se hur dessa lexikon påverkar översättningsresultatet. Vi skulle också vilja pröva på efterredigering av en text översatt av TranSmart för att ta reda på hur besvärligt och tidskrävande den egentligen skulle vara. Själva programmet var nämligen mycket lätt och snabbt att använda.

Vid Trados fick vi inte en ordentlig helhetsbild av funktionerna och därför skulle det vara intressant att fortsätta testa dem i praktiken. Man borde översätta större helheter med hjälp av programmet för att bättre kunna evaluera dess funktioner och egenskaper. Trados har så många funktioner och alternativ att det säkert skulle ta en lång tid att verkligen lära sig känna programmet. Intressant skulle också vara att vidare undersöka repetitionen i texter med Trados.

När man tänker köpa ett maskinöversättningsprogram, vill man säkert att det skall fungera. Ett maskinöversättningsprogram är ofta en stor investering och därför vill man också veta om det är värt pengarna. Det är också viktigt att programmet skulle vara lätt att använda. Om det fanns tillräckligt mycket undersökningsinformation om användbarheten av översättningsprogram, skulle det vara lättare för en eventuell köpare att avväga sina behov i förhållande till nyttan som varje program kunde erbjuda. Då skulle antalet misslyckade uppköp reduceras men sannolikt skulle också efterfrågan av riktigt funktionella program öka.

När nya översättningsprogram kommer ut på marknaden, skulle det vara bra att ha jämförande undersökningar om dessa program så snabbt som möjligt. Nya program borde jämföras med de program som redan finns på marknaden. Detta slags undersökning skulle ge värdefull information åt ett företag eller en enskild översättare som tänker köpa ett översättningsprogram. Nya program borde jämföras med till exempel marknadsledarna Trados, Transit eller Systran för att kunna se vilka av dem som är värda att köpas.

LITTERATUR

- ALPAC (1966). *Language and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee. Washington, D.C: National Academy of Sciences.
- Arnola, H.; Hyvönen, K.; Juntunen, J.-P.; Linnanvirta, T. och Suoranta, P. (1996a). Kielikone machine translation technology and its perspective on the economics of MT. *Proceedings of MT workshop*. Wien.
- Arnola, H.; Hyvönen, K.; Juntunen, J.-P.; Linnanvirta, T. och Suoranta, P. (1996b). Kielikone Finnish-English MT system in practical use. *Translating and the computer*, 18.
- Arnold, D.; Balkan, L.; Meijer, S.; Humphreys, R.L. och Sadler, L. (1995). *Machine translation: an introductory guide*. [www]. Hämtat från <<http://clwww.essex.ac.uk/MTbook/HTML/book.html>>. Hämtat 28 maj 2001.
- Balkan, L.; Arnold, D. och Sadler, L. (1997). *Tools and techniques for machine translation teaching: A survey*. [www]. Hämtat från <<http://clwww.essex.ac.uk/group/projects/MTforTeaching/index.html>> Hämtat 23 oktober 2001.
- Buchmann, B. (1987). Early history of machine translation. I King, M. (red.), *Machine translation today: The state of the art*. Edinburg: Edinburg University Press, 3-21.
- Computer aided translation*. [www]. Hämtat från <<http://www.isp.nl/Us/cattxt.html>>. Hämtat 31 januari 2001.
- De Roek, A. (1987). Linguistic theory and early machine translation. I King, M. (red.), *Machine translation today: the state of the art*. Edinburg: Edinburg University Press, 38-57.
- Description of the individual TRADOS programs*. [www]. Hämtat från <http://partners.trados.com/download/Trados_Product_Description.pdf>. Hämtat 23 oktober 2001.
- EAGLES Evaluation of natural language processing systems Final report*. (1995). [www]. Hämtat från <<http://www.issco.unige.ch/ewg95/ewg95.htm>>. Hämtat 20 juni 2001.
- Honkonen, J. (2001). *Lokalisoinnin vaikutus käännöksen laatuun*. [www]. Hämtat från <<http://www.mikrobitti.fi/nettijatkot/2001/09/kieli/index.html>>. Hämtat 16 oktober 2001.
- Hutchins, W.J. (1986). *Machine translation: past, present and future*. West Sussex: Ellis Horwood Limited.
- Hutchins, W.J. (1995). *Machine translation: a brief history*. [www]. Hämtat från <<http://ourworld.compuserve.com/homepages/WJHutchins/Conchist.htm>>. Hämtat 10 maj 2001.

- Hutchins, W.J. (1999). *The development and use of machine translation systems and computer-based translation tools*. [www]. Hämtat från <<http://ourworld.compuserve.com/homepages/WJHutchins/Beijing.htm>>. Hämtat 19 juni 2000.
- Hutchins, W.J. & Somers, H.L. (1992). *An introduction to machine translation*. London. Academic Press Limited.
- Hyvönen, K. <kaarina@kielikone.fi> ”Re: TranSmart-kysymyksiä!” (23 oktober 2001) Personlig e-post.
- Hyvönen, K. & Arnola, H. (1998). Kielikone takes user-friendliness seriously. *Translating and the computer*, 20.
- Kielikones hemsida. [www]. Hämtat från <<http://www.kielikone.fi>>. Hämtat 23 oktober 2001.
- Kieliteknologia Suomessa*. (1998). (Red.) Miettinen, Manne. [www]. Hämtat från <<http://www.csc.fi/kieliteknologia/raportti/raportti.html>>. Hämtat 16 oktober 2001.
- Laboration 4: Översättning med översättningsstöd*, Hämtat från <<http://stp.ling.uu.se/educa/kurssidor/ist99/ist14.html>>. Hämtat 22 oktober 2001.
- Maskinöversättning*. [www] Hämtat från <<http://stp.uu.se/~fredrikg/maskinov.htm>>. Hämtat 20 juni 2001.
- Melby, A. (1992). The translator workstation. I Newton, J. (red.), *Computers in translation A practical appraisal*. London. Routledge, 147-165.
- Pugh, J. (1992). The story so far: an evaluation of machine translation in the world today. I Newton, J. (red.), *Computers in translation A practical appraisal*. London. Routledge, 14-32.
- Somers, H.L. (1992). Current research in machine translation. I Newton, J. (red.), *Computers in translation A practical appraisal*. London. Routledge, 189-207.
- Surfa på ditt modersmål - Maskinöversättning på internet*. [www]. Hämtat från <<http://www.svenska.sics.se/oversattning/mt-uppsats.htm>>. Hämtat 5 mars 2001.
- Trados hemsida. [www]. Hämtat från <<http://www.trados.com>>. Hämtat 23 oktober 2001.
- Trados translation guide. [www]. Hämtat från <<http://support.trados.com/download.asp>>. Hämtat 23 oktober 2001.
- Trados Translator's Workbench Online Help. Help-meny i programmet Trados Translator's Workbench.
- Translationzone.com. [www]. Hämtat från <<http://www.translationzone.com/Scripts/WebObjects.dll/TZ.woa/wa/ProductsDirectAction/freelanceEdition>>. Hämtat 23 oktober 2001.

- Tucker, A.B. (1987). Current strategies in machine translation research and development. I Nirenburg, S. (red.), *Machine translation: theoretical and methodological issues*. Cambridge: Cambridge University Press, 22-41.
- Warwick, S. (1987). An overview of post-ALPAC developments. I King, M. (red.), *Machine translation today: The state of the art*. Edinburg: Edinburg University Press, 22-37.