

# Att undersöka språkmöten med datorn

Lars Borin

## Vad är språkmöten och hur märker man av dem?

Det som jag i rubriken syftar på med termen *språkmöten* studeras i språkvetenskapen under några olika rubriker; man talar bland annat om *kontaktlingvistik*, *översättningsfenomen* och *inlärarespråk*. I samtliga dessa fall handlar det om att man kan visa att ett språk på något sätt påverkar ett annat. De olika rubrikerna beror på att det rör sig om olika sociala situationer (mer om detta strax).

Språkmöten sker alltid inne i människor, därför att det är inne i människor som språket finns och det är också där språkförändringar börjar. Man kan dock bara utforska språkmöten genom att undersöka deras verkningar i konkreta språkhandlingar, i tal eller skrift. Här gör man ungefär som astronomerna när de skiljer ut planeter, kometer, asteroider och rymdskepp från fixstjärnorna (som kallas så just för att de verkar orörliga): Man jämför en ögonblicksbild av en del av himlavalvet med en annan tagen vid en annan tidpunkt.<sup>1</sup> På ett liknande sätt kommer man åt en sorts effekt av språkmöten. Man utgår från ett känt språkssystem (enligt en nedtecknad beskrivning eller den egna språkkunskapen), och iakttar detaljer som verkar förändrade. Naturligt nog är det lättast att peka på ett annat språk som källa när någon detalj har *kommit till* i språket. Sådana detaljer kan man finna på alla möjliga språkliga områden:

- i ordförrådet, med lånord, nya användningar av existerande ord, etc.
- i fraseologin, med nya fasta uttryck, nya bundna konstruktioner, etc.
- i syntaxen, med nya konstruktionstyper
- i texters uppbyggnad, med nya retoriska strukturer, nya genrer, etc.

När vi istället har att göra med *förlust* av ord, uttryck, konstruktioner etc., så blir det svårare att direkt visa att ett annat språk har orsakat förlusten; här blir det mer fråga om ett resonemang som stöds av indicier.

I vilket fall som helst så är sådana här språkliga detaljförändringar förhållandevis lätta att upptäcka. Språkmöten manifesterar sig dock också i en mer subtil form, nämligen som en skillnad i distribution. Här handlar det inte om några klara skillnader av den typ som ett nytt ord i språket innebär, utan om tendenser. Språket är ju inte bara något som tillhör individen, utan även ett socialt fenomen. Precis som andra sociala fenomen kan man således undersöka det med statistiska metoder, för att på det sättet se sådant som inte syns annars. För att ta ett exempel ur en annan social domän: Diskriminering

i samhället syns ofta klarare när man tittar på ett stort antal fall. Då kanske man upptäcker att en viss grupp (kvinnor, romer, invandrare, äldre) är systematiskt under- eller överrepresenterade i något eller några sammanhang, i förhållande till dess andel av befolkningen. Man skulle t.ex. förvänta sig att c:a 50% av professorerna vid ett svenskt universitet vore kvinnor, eftersom c:a 50% av den svenska befolkningen utgörs av kvinnor, att c:a 1% av poliserna i Sverige skulle vara romer, etc., men så är det inte. Nu måste jag understryka att observationen att en grupps representation i ett visst sammanhang avviker från vad man skulle förvänta sig, givet gruppens andel av en större gemenskap, är ett faktum som kräver *tolkning* – i form av en plausibel förklaring – för att vara användbart. I de exempel som jag just har givit och andra liknande fall som man stöter på i medierna behöver förklaringen inte vara diskriminering. Vad jag vill illustrera med exemplen är bara att tendenser är något annat än enskilda fall, och kräver en annan undersökningsmetod. Det är inte anmärkningsvärt att man i det enskilda fallet – t.ex. vid en anställning – väljer en svensk man bland, säg, 40 kandidater, precis som det inte är anmärkningsvärt att en tärning visar en sexa när man kastar den en gång. Men om man undersöker 1000 fall och det visar sig att en svensk man alltid har valts eller att tärningen alltid har fallit med sexan uppåt, då anar man ugglor i mossen.

På samma sätt kan man använda sig av statistik över ett stort antal språkhandlingar för att se om man på det viset kan upptäcka tendenser som inte går att urskilja i de enskilda språkhandlingarna. Här kan man också undersöka språket i två kontexter – vid två tidpunkter, i två sociala sammanhang, i två texttyper, etc. – för att se om det finns statistiskt påvisbara systematiska skillnader mellan de två kontexterna.

Här ska jag alltså berätta om hur datorn kan användas för att hjälpa språkvetaren att undersöka språkmöten. Jag ska närmare bestämt visa hur mina kollegor och jag har tagit datorn till hjälp för att undersöka konkreta exempel på tre olika sorters språkmöten. De typer av språkmöten jag ska ta upp brukar inom språkvetenskapen diskuteras under de redan nämnda rubrikerna

- (1) *kontaktlingvistik* eller *språkkontaktforskning*
- (2) *translationese* eller *översättningsfenomen*
- (3) *inlärarspråk* eller *interimspråk* eller *interferens*

Dessa termer ska jag förklara lite längre fram, i de sammanhang där jag behandlar dem. Först måste jag dock förklara några andra termer som vi kommer att behöva för att kunna tala om datorns roll i det här sammanhanget.

## Några datorlingvistiska termer

Mitt eget forskningsområde kallas *datorlingvistik*, och jag brukar ofta förklara dess ändamål som ”att få datorer att göra vettiga saker med språk”. Med ”vettiga saker” menar jag ungefär sådant som i någon mening förutsätter förståelse av, eller åtminstone kunskap om, språk. Inom datorlingvistiken studerar man bl.a. automatisk översättning (som man kan se i en del webbsökmotorer), språk- och grammatikkontroll (som i en del ordbehandlingsprogram), innehållsbaserad dokumentsökning och dokumentklassificering, och datorstödd språkinlärning.

Precis som andra områden håller sig datorlingvistiken med en uppsättning analytiska metoder och verktyg, och termer för dessa. Några sådana termer som vi kommer att behöva är:

**korpus:** en (stor) mängd text insamlad enligt vissa principer (t.ex. representativitet). Korpusen ska vara stor för att man ska kunna undersöka den med statistiska metoder, och beroende på vilka språkliga fenomen man vill undersöka rör det sig om några tusen till hundratals miljoner löpord. En vanlig korpusstorlek idag är någon till några miljoner ord.

**parallellkorpus:** en korpus med texter och deras översättningar. Parallellkorpusen innehåller alltså minst två språk, men kan innehålla mer än en översättning av samma källspråkstext, till mer än ett målspråk.

**jämförbar korpus:** en korpus med originaltexter på två eller flera språk eller språkformer, där texterna överensstämmer med avseende på sådana faktorer som stil, ämnesområde och genre.

**korpuslingvistik:** ett delområde av datorlingvistiken där man utvecklar datamaskinella verktyg för att undersöka språket i korpusar.

**konkordans:** ett av de just nämnda datamaskinella korpusverktygen, en sorterad lista över (valda) textord i de kontexter (rad eller mening) de förekommer i korpusens texter; se figur 1.

**(ordklass)taggning:** (automatiskt) förse orden i en text med uppgifter om ordklass och böjningsform. En *(ordklass)taggad korpus* är således en korpus där varje textord har försetts med sådan information. Den mest kända svenska taggade korpusen, SUC, eller *Stockholm Umeå Corpus* (Ejerhed och Källgren 1997), omfattar en miljon löpord fördelade på ett antal texttyper och genrer som tillsammans antas ge en god bild av modernt svenskt skriftspråk.

**länkning:** (automatisk) ’hopkoppling’ av översättningsmotsvarigheter i en parallellkorpus. Vanliga språkliga enheter som länkas är meningar (*meningslänkning*; se figur 2) och ord (*ordlänkning*; se figur 3).

## Exempel 1: kontaktlingvistik

Inom kontaktlingvistik eller språkkontaktforskningen undersöker man framför allt i ett historiskt perspektiv – det är alltså ett forskningsfält som gränsar till den s.k. historiska lingvistikens – hur språk påverkar varandra på alla plan. Typiskt handlar det om två (eller flera) språkgemenskaper som lever nära inpå varandra, med en betydande grad av tvåspråkighet. Man brukar anse att resultatet av språkpåverkan framför allt betingas av sociala omständigheter, eller maktförhållanden, rent ut sagt. Det finns en språkvariant som har hög prestige och en eller flera med låg prestige – där språkformens prestige egentligen beror på dess talares sociala status – och normalt utövar högprestigevarianten inflytande på lågprestigevarianten, men inte tvärtom, av den enkla orsaken att det mest brukar vara lågprestigevariantens talare som är tvåspråkiga; som exempel kan vi ta samisktalande områden i Sverige där samerna i allmänhet också talar svenska, medan svenskarna oftast inte behärskar samiska.

Under en längre tid har jag samlat en korpus med skriven finsk romani, alltså de finska romernas, eller zigenarnas, språk (Vuorela och Borin 1998; Borin 2000). De finska romerna är en grupp om ungefär 10.000 personer, varav ungefär två tredjedelar lever i Finland och en tredjedel i Sverige. Deras språk, finsk romani, är ett s.k. nyindiskt eller indoariskt språk, besläktat med det gamla kulturspråket sanskrit och sådana moderna indiska språk som hindi och bengali. Romaniforskningen har alltid varit av stor betydelse för de lingvister som utforskat språkkontakter, eftersom de olika romanivarianterna så uppenbart har påverkats av olika språk genom tiderna, samtidigt som de alla har bevarat en kärna av indisk grammatik och indiskt ordförråd. De har alltså hjälpt forskarna att teoretisera om vad i språket som lättare påverkas i vissa sorters språkkontaktsituationer och vad som är mer motståndskraftigt.

För finsk romani finns sedan 1971 en officiell ortografi, och det har producerats en hel del text på språket (även före 1971), i form av tidningsartiklar (främst i månadspublikationen *Romano Boodos*, som ges ut av den religiösa organisationen *Romano Missio*), evangelieöversättningar, lärobokstexter och en del annat textmaterial. Finsk romani är en lågprestigevariant (även om den inte är en alldeles typisk sådan; se Vuorela och Borin 1998) och man förväntar sig alltså att finsk romani ska uppvisa påverkan från högprestigespråket finska, något som jag alltså empiriskt undersöker med hjälp av korpusmaterial.

Korpusen innehåller både originaltext från tidningar och läroböcker, och översatt text, främst evangelier. För de sistnämnda finns naturligtvis också parallella versioner på andra språk, där den finska versionen troligen ligger till grund för översättningarna. Även om jag betecknar korpusen som en korpus med skriven finsk romani, så innehåller den också en liten mängd nedtecknat intervjumaterial, alltså snarast talspråk. Denna korpus använder jag

således för att undersöka finskt inflytande på finsk romani, där jag för närvarande tittar närmare på två språkdrag, nämligen bruket av bestämd artikel och utvecklandet av ett genitivkasus.

Europeisk romani och tidigare även finsk romani har en framförställd bestämd artikel *o* 'maskulinum singularis' / *i* 'femininum singularis' / *o* 'pluralis' (formerna varierar en aning beroende på vilken romanivarietet man talar om). Artikeln fungerar alltså som engelska *the* eller tyska *der/die/das/die* eller franska *le/la/les*. Denna 'gamla' artikel har helt försvunnit ur modern finsk romani, troligen under finskt inflytande, eftersom det finska standardspråket inte har några artiklar.<sup>2</sup> En ny bestämd artikel verkar dock vara under utveckling, där *dola*, en stelnad form av ett demonstrativt pronomen används på ungefär samma sätt som det finska pronomenet *se* 'den, den där', som i finskt talspråk under lång tid har utvecklat sig i riktning mot ett slags bestämd artikel.<sup>3</sup> Här erbjuder korpuslingvistik ett utmärkt arbetsredskap i form av de tidigare nämnda konkordanserna. Man kan undersöka hur *dola* används genom att generera en meningskonkordans för det ordet ur korpusen; se figur 1.

```
\id AR:1940/00[PA1=S]0048
\tx Doi hin dui xeel presiba buuribosta, dola hispata.
    [Det är två tusen (i) kostnad för boendet, för det huset.]
\id AR:1940/00[PA2=S]0028
\tx Doori vela iek siivi ta thouvela dola kentos.
    [Där kommer en kvinna och tvättar det barnet.]
```

**Figur 1:** Del av meningskonkordans för ordet *dola* ur den finska romanikorpusen. Beläggstället markeras med "\id" och meningen med "\tx". Den svenska översättningen finns inte i korpusen, utan är tillagd här.

Parallellkorpusmaterialet har jag använt framför allt i länkad form. En automatisk meningslänkning av versionerna på finsk romani och finska av Johannevangeliets början visas i figur 2. Den meningslänkningsmetod som använts är utvecklad av Erik Tjong Kim Sang, tidigare vid Institutionen för lingvistik, Uppsala universitet, nu vid universitetet i Antwerpen (Tjong Kim Sang 1999).

Number of matches: 1482

# id2 xid2 xid3

(1) Vaagos sas Lau, ta Lau sas Deevelesko neere, ta Lau sas Deevel.

[I början var Ordet, och Ordet var hos Gud, och Ordet var Gud.]

(2) Alussa oli Sana.

(2) Sana oli Jumalan luona, ja Sana oli Jumala.

# id3 xid4

(1) Jou sas vaagos Deevelesko neere.

[Det var först hos Gud.]

(2) Jo alussa Sana oli Jumalan luona.

**Figur 2:** Början av meningslänkning av Johannesevangeliet på finsk romani och finska. Den svenska översättningen finns inte i korpusen, utan är tillagd här.

Det är olämpligt att endast använda översatt text för att visa på språkkontaktfenomen, eftersom översatt text ofta uppvisar påverkan från källspråket även i fall där man inte alls brukar tala om språkkontakt (se nästa avsnitt). Därför använder jag parallelltextmaterialet närmast som 'hypotesgenerator'. Jag letar i det (med hjälp av parallella konkordanser, alltså konkordanser gjorda på menings- och ordlänkat material) efter ord och konstruktioner där man har anledning att misstänka att finsk romani kan ha påverkats av finska, men studerar sedan också hur dessa ord och konstruktioner uppträder i det enspråkiga korpusmaterialet. Parallellbeläggen är sedan värdefulla när man ska redovisa hur påverkan kan ha gått till, eftersom de tillhandahåller konkreta exempel på konstruktionsparallellism, som i figur 3.

[ur Johannes 1:1:]

**romani** Deevel=es -ko neere

**finska** Jumala -n luona

'Gud' -GEN 'hos'

[d.v.s. 'hos Gud']

**Figur 3:** Exempel ur parallellkorpus på användning av postposition (*neere* 'hos') som styr (ny) genitiv i finsk romani, som en exakt parallell till finska (där konstruktionen är mycket gammal). Den svenska översättningen finns inte i korpusen, utan är tillagd här.

Tidigare hade finsk romani inget genitivkasus – i likhet med de flesta andra europeiska romanivarieteter – utan genitivens plats intogs av ett så kallat *possessivt adjektiv*, ungefär som om vi på svenska inte skulle kunna säga *Europas historia*, utan alltid bara *den Europeiska historien*.<sup>4</sup> Det possessiva adjektivet hade lika många former som andra adjektiv, nämligen former för kongruens med huvudord i maskulinum (singularis *-ko* / pluralis *-go*) och fe-

mininum (-*ki/-gi*) singularis samt i pluralis (-*ke/-ge*). I nutida finsk romani finner man bara en form, den som tidigare markerade att huvudordet i ägar-konstruktionen var maskulinum singularis (-*ko/-go*). Misstanken att vi i det moderna språket har att göra med ett regelrätt genitivkasus styrks av att det nu förekommer postpositioner – mot tidigare enbart prepositioner; finskan däremot har fler postpositioner än prepositioner – som styr den nya genitiv-formen, något som inte kunde inträffa i det äldre språksystemet; se figur 3.

## Exempel 2: 'translationese'

Det är välkänt och okontroversiellt bland översättningsforskare att översatta texter inte är likadana som originaltexter skrivna direkt på språket ifråga.<sup>5</sup> Martin Gellerstam vid Göteborgs universitet kallar i en uppsats denna företeelse för "translationese" (Gellerstam 1985). Translationese (eller kanske 'översättningska') är *inte* detsamma som språkfel, utan språket i en översättning är nog oftast korrekt vad gäller enskilda språkdrag. Snarare handlar det om avvikande distributionella mönster, om under- eller överanvändning av ord, konstruktioner o.dyl. På senare år har språkvetare börjat undersöka translationese med hjälp av korpusmaterial. Oftast har det rört sig om undersökningar av lexikaliska förhållanden, alltså användningen av vissa ord, t.ex. hur *anlända* överanvänds i svenska översättningar från engelska (enligt mönster från engelska *arrive*).

I samarbete med min kollega Klas Prütz vid Institutionen för lingvistik, Uppsala universitet, har jag försökt utvidga korpusbaserade undersökningar av translationese även till det syntaktiska planet. Vi har nämligen arbetat med ordklassstaggade korpusmaterial, och undersökt frekvenser av ordklasser och ordklassföljder (2–5 ord i följd) i

- (1) svensk originaltext (ur *Invandrartidningen*); c:a 100.000 ord
- (2) engelsk översättning av denna (*News and Views*); något fler ord
- (3) brittisk engelsk originaltext (tidningstext); c:a 100.000 ord
- (4) amerikansk engelsk originaltext (tidningstext); c:a 100.000 ord

Här har vi alltså både en parallellkorpus (1–2) och två jämförbara korpusar (1–3 och 2–3). Korpusarna taggades med jämförbara ordklassuppsättningar (vilket naturligtvis underlättades av att svenska och engelska är så pass närbesläktade språk), alla ordklasskvenser av längden 1–5 extraherades ur de taggade korpusarna och räknades, och ett antal specialprogram användes sedan för att göra följande bearbetningar av materialet:

- (1) Den översatta texten jämfördes med den brittiska engelska originaltexten (eftersom översättningen var att karakterisera som brittisk snarare än amerikansk engelska), och alla ordklasskvenser som förekom med

signifikant högre eller lägre frekvens (enligt Mann-Whitney, ett statistiskt standardtest) i den översatta texten sparades.

(2) Den svenska originaltexten jämfördes med den brittiska engelska originaltexten, och alla ordklasssekvenser som förekom med signifikant högre eller lägre frekvens i den svenska texten sparades.

(3) Den amerikanska engelska originaltexten jämfördes med den brittiska engelska originaltexten, och alla ordklasssekvenser som förekom med signifikant högre eller lägre frekvens i den amerikanska engelska originaltexten sparades.

(4) De tre listorna jämfördes med varandra på följande sätt: (4a) alla sekvenser som fanns i den tredje listan togs bort från de andra två listorna (på det viset hoppades vi undvika att eventuella amerikanismer skulle klassas som översättningsfenomen; (4b) alla ordklasssekvenser som var gemensamma för återstoden av första och andra listan letades fram och fördes till resultatlistan.

Dessa bearbetningar gav som resultat några tiotal kvarvarande ordklasssekvenser, representerande både överanvändning och underanvändning i förhållande till den brittiska engelska originaltexten. Här ska jag bara ge några exempel på översättningsfenomen som vi tror oss ha påvisat med denna metod. Engelska översatt från svenska (och svensk originaltext) uppvisade i jämförelse med engelsk originaltext fler adverb totalt sett, och fler prepositioner i första meningspositionen. Det sistnämnda hör rimligen ihop med att man på svenska gärna placerar något annat än subjektet på första plats i meningen, t.ex. ett rums- eller tidsuttryck, som ju ofta inleds med en preposition. På engelska finns den möjligheten också, men den utnyttjas i betydligt mindre utsträckning; se figur 4.

**För** att bli svensk folkmusiker tog Ale en ovanlig omväg.

**I** Malmö, där han bodde, blev han bekant med en invandrad grekisk musiker som satte en bouzouki i händerna på Ale. [*Invandrantidningen*]

**In** becoming a Swedish folk musician, Ale Möller took a strange path. He lived **in** Malmö and got to know an immigrant Greek musician who placed a bouzouki in his hands. [*News and Views*]

**Figur 4:** Prepositioner (*för*, *i*) i första meningspositionen i svensk originaltext och i (automatiskt meningslänkad) översättning till engelska.



### Exempel 3: inlärarspråk

På samma sätt som man utforskar översättningsfenomen med hjälp av korpusar, har man på senare tid också börjat samla in så kallade *inlärarkorpusar*. Dessa innehåller språkproduktion från andraspråksinlärare, och syftet med dem är att man med deras hjälp ska kunna genomföra empiriska undersökningar av inlärarspråk, eller *interimspråk*, som i andraspråksinlärningsforskarnas terminologi är en sammanfattande benämning på språkstadierna mellan absolut nybörjarnivå och full andraspråksbehärskning. En intressant och delvis kontroversiell fråga när det gäller interimspråk är vilken roll *interferens* (negativ påverkan från inlärarens modersmål) spelar i dess utveckling.

De inlärarkorpusundersökningar som hittills har presenterats (t.ex. flera uppsatser i Granger 1998) har – precis som korpusbaserade undersökningar av översättningsfenomen – mest rört sig på det lexikala planet, och såvitt jag vet uteslutande behandlat målspråket. Interferens består ju, precis som translationese, i en relation mellan två språk, inlärarens modersmål och det språk hon håller på att lära sig. Därför verkar det rimligt att man skulle kunna undersöka interferens på samma sätt som man undersöker translationese.

Därför har min kollega Klas Prütz och jag tagit oss före att tillämpa samma metod som vi använde för vår översättningsundersökning (se föregående avsnitt) även på ett material bestående av

- (1) en korpus med skriftlig produktion av avancerade svenska inlärare av engelska (studenter på Engelska institutionen vid Uppsala universitet), Uppsala Student English-, eller USE-korpusen (Axelsson 2000).
- (2) en korpus med originaltexter på standardsvenska (SUC, eller *Stockholm Umeå Corpus*; Ejerhed och Källgren 1997)
- (3) en korpus med engelska originaltexter (BNC, eller *British National Corpus*; Burnard 1999)

Bearbetningen av undersökningsresultaten pågår fortfarande, men preliminärt kan vi säga följande om inlärarspråk jämför med översatt språk:

- Inte helt oväntat fanns det fler skillnader mellan originaltext på målspråket och inlärartext än mellan originaltext och översatt text.
- Det fanns många fler fall av underanvändning än av överanvändning, d.v.s. inlärarspråket uppvisar mindre syntaktisk variation än originaltext på målspråket, vilket nog också är vad man förväntar sig.
- De intressantaste resultaten finns förstås bland de konkreta skillnaderna, av vilka vi här ger ett litet axplock. I inlärarspråket iaktogs: (1) överanvändning av konjunktioner, adverb och finita verb, (2) underanvändning av particip och gerundier, (3) underanvändning av substantiv+substantiv-sekvenser i meningens början.

## Slutord

Sammanfattningsvis hoppas jag att jag har kunnat visa att datorn är ett värdefullt redskap för undersökningar av språkmöten. Speciellt gäller det storskaliga empiriska undersökningar av korpusmaterial, där det i praktiken är helt omöjligt att genomföra samma undersökningar manuellt. Korpusverktygen utvecklas också ständigt, till en stor del naturligtvis därför att datorerna i rask takt blir snabbare och kraftfullare, men också därför att det pågår ett intensivt metodutvecklingsarbete inom denna del av datorlingvistik. Bland annat blir det allt vanligare med så kallade parsade korpusar – eller ’trädbanker’ (*treebanks* på engelska), alltså korpusar där man inte bara har taggat de enskilda orden med deras ordklass och eventuellt böjningsform, utan också med information om hur de hör ihop och vilken roll de spelar i större syntaktiska enheter, eller *fraser*, som substantivfraser, prepositionsfraser, satser o.s.v. Det vore högst naturligt och önskvärt att utvidga de jämförelser som jag har redovisat i de föregående avsnitten till att omfatta också sådana större syntaktiska enheter. Då kunde man t.ex. fråga direkt ”Vilken sorts fras kommer först i meningen?”, snarare än som nu indirekt, via frågan ”Vilken ordklass har det första, andra, o.s.v. ordet i meningen?”. Detta är också vad vi planerar för framtiden i vår forskning om översättnings- och inlärarespråk.

I min framställning har jag koncentrerat mig på metodfrågor och på själva vetenskapen, alltså på hur undersökningarna har genomförts och vad de har gett för resultat i form av ny kunskap. Däremot har jag inte alls berört frågan om vad denna kunskap kan användas till. Därför ska jag avslutningvis helt kort nämna några tillämpningsområden för sådan kunskap:

- Den kunskap man får genom automatisk länkning av parallellkorpusar kan användas för att förbättra arbetsredskapen för översättare, både manuella redskap som lexika och kontrastiva grammatiker, och automatiska redskap som maskinöversättningssystem.
- Arbetet med inlärarespråk kan leda till bättre och mer individ- eller gruppanpassad språkundervisning, liksom till bättre hjälpmedel för språkinlärare.
- En mycket viktig och ofta förbisedd tillämpning för den typ av resultat och de metoder jag har redogjort för här är utarbetande av resurser (ordböcker, grammatiker, läromedel, terminologi) för små, ofta hotade språk (Allwood och Borin 2001).

## Noter

- 1 Tekniken man använder är densamma som i tecknad film: Man växlar snabbt mellan de två bilderna på samma visningsyta. Objekt som har rört sig på himla-valvet mellan de två exponeringarna kommer då att skenbart hoppa fram och tillbaka.
- 2 Den 'gamla' artikeln fanns fortfarande i den språkform som beskrevs av Thesleff (1901), men är idag helt försvunnen. Det är nog inte bara artikellösheten i det finska standardspråket som har haft inflytande här, utan även det faktum att finska inte har någon genusdistinktion.
- 3 Även om det fortfarande råder delade meningar i finsk språkvetenskap om hur man ska klassificera konstruktionerna med *se* (se Juvonen 2000).
- 4 Den stora bokstaven i "Europeiska" står där med avsikt; här handlar det om ett tänkt alternativ till den svenska genitivkonstruktionen, och då kunde egennamn mycket väl tänkas behålla sin ortografiska markering.
- 5 När jag läste språk på universitetet, var det i textläsningskurserna av den anledningen alltid förbjudet att läsa översättningar. Märk dock att man *inte* brukar säga att språket i översatta texter är sämre än det i originaltexter, utan bara anorlunda!

## Referenser

- Allwood, Jens och Borin, Lars. 2001. "Datorer och språkteknologi som hjälpmedel i bevarandet av romani – Computers and language technology as an aid in the preservation of Romani". Plenarföreläsning vid symposiet *Romani som undervisningsspråk • Romani as a language of education*, Göteborgs universitet, 19–20/1 2001.
- Axelsson, Margareta Westergren. 2000. "USE – The Uppsala Student English Corpus: an instrument for needs analysis". *ICAME journal* 24. 155–157.
- Borin, Lars. 2000. "A corpus of written Finnish Romani texts". I: Ó Cróinín, Donncha (red.), *LREC 2000. Second international conference on language resources and evaluation. Workshop proceedings. Developing language resources for minority languages: reusability and strategic priorities*, 75–82. Athens: ELRA.
- Borin, Lars och Prütz, Klas. 2001. "Through a glass darkly: part of speech distribution in original and translated text". I: Daelemans, Walter; Sima'an, Khalil; Veenstra, Jorn och Zavrel, Jakub (red.), *Computational linguistics in the Netherlands 2000*, 30–44. Amsterdam: Rodopi.
- Borin, Lars och Prütz, Klas. under utarb. "By their fruits ye shall know them: interference in a learner language corpus".
- Burnard, Lou (red.). 1999. *Users reference guide for the BNC Sampler*. Published for the British National Corpus Consortium by the Humanities Computing Unit at Oxford University Computing Services.
- Ejerhed, Eva och Källgren, Gunnel. 1997. "Stockholm Umeå Corpus version 1.0, SUC 1.0". Department of Linguistics, Umeå University.

- Gellerstam, Martin. 1985. "Translationese in Swedish novels translated from English". I: Wollin, Lars och Lindquist, Hans (red.), *Translation studies in Scandinavia. Proceedings from the Scandinavian symposium on translation theory (SSOTT) II*, 88–95. Lund University.
- Granger, Sylviane (red.). 1998. *Learner English on computer*. London: Longman.
- Juvonen, Päivi. 2000. *Grammaticalizing the definite article*. Department of Linguistics, Stockholm University.
- Thesleff, Arthur. 1901. *Wörterbuch des Dialekts der finnländischen Zigeuner*. Helsinki.
- Tjong Kim Sang, Erik. 1999. "Aligning the Scania corpus". I: Sågvall Hein, Anna (red.), *Working papers in computational linguistics & language engineering 18. Reports from the ETAP project: Converting, aligning and tagging for ETAP*, 1–7. Department of Linguistics, Uppsala University.
- Vuorela, Katri och Borin, Lars. 1998: "Finnish Romani". I: Ó Corráin, Ailbhe och Mac Mathúna, Séamus (red.), *Minority languages in Scandinavia, Britain and Ireland*, 51–76. Acta Universitatis Upsaliensis, Studia Celtica Upsaliensia.