

Rolf Gavare  
SPRAKDATA

## LEXIKOGRAFISK ALFABETISERING

### Inledning

Föredraget behandlar de principfrågor som aktualiseras vid krävande datamaskinell alfabetisering av lexikaliska och encyklopediska material och ger förslag för den praktiska implementeringen. Även om särskild hänsyn tas till svenska språkets konventioner, så har grundprinciperna relevans även för andra språk med alfabetisk skrift.

Vad menar vi då med alfabetisering? Frågan kan i förstone tyckas överflödig. När man kommer in på detaljfrågor visar det sig emellertid ofta att man inte har någon etablerad praxis eller inte är medveten om vilka kriterier man använder för det alfabetiska ordnandet.

Med ett alfabet brukar vi mena ett språks uppsättning av bokstäver med en viss hävdvunnen intern ordningsföljd. Dessa konventioner är språkberoende. Alfabetisering är alltså, lite vagt formulerat, ordnandet av ord eller ordförbindelser i alfabetisk följd. Problemet är nu bland annat att ord inte enbart består av alfabetiska skrivtecken utan även av siffror, interpunktionstecken, diverse symboler, främmande skrivtecken, etc., utan självklar plats i alfabetet. Dessutom tillmäts inte alla skrivtecken lika stor betydelse — somliga tecken påverkar inplaceringen i den alfabetiska ordningsföljden mindre än andra.

Alfabetiskt ordnade register behöver ofta hanteras av datorer, vilket ställer särskilda krav i fråga om konsekvens, entydighet och formaliserbarhet på principerna för det alfabetiska ordnandet. De standardprogramvaror som finns för alfabetisk filering uppvisar emellertid stora brister; ofta blir det till och med problem att jämställa versaler och gemena skrivtecken eller v och w eller att få den svenska bokstavsordningen ...zääö. De diakritiska tecknen och junkturerna klaras inte alls av på något tillfredsställande sätt. För en del användningsområden kan man acceptera vissa ofullkomligheter i alfabetiseringen, men för tillämpningar med större krav på lexikologisk stringens (såsom olika slag av konkordanser, ordlistor, ordböcker, encyklopedier, personnamnsregister och bibliografiska kataloger) krävs formaliserade alfabetiseringsprinciper som motsvarar vetenskapliga krav.

Alfabetiseringsreglerna skall uppfylla den grundläggande förutsättningen att kunna ge en algoritmisk beskrivning som

kan implementeras i datorprogram. Tyvärr saknas ännu en tillräckligt specifik standard för lexikografisk alfabetisering. Denna översikt kan förhoppningsvis bidra med synpunkter inför ett sådant standardiseringsarbete.

### Allmänna överväganden

Ett fundamentalt krav på principer för alfabetisering är att slumpmässig ordning mellan (grafiskt) olika enheter inte skall tolereras. Ordningföljden skall alltid vara förutsägbar och entydig. Detta innebär att man skall få exakt samma resultat om samma material alfabetiseras på nytt. En delvis slumpmässig interfilering vållar i många sammanhang komplikationer.

Då de traditionella alfabetiseringsprinciperna är otillräckliga och man behöver generalisera och utvidga dem, bör man sträva efter att finna lösningar som har lingvistisk och psykologisk relevans.

Rent allmänt kan sägas att alfabetiseringsregler först skall ge besked i två förberedande steg: vilka element som alfabetiseringen skall ta hänsyn till och vilka teckenföljder som skall jämföras då ordningen mellan enheterna skall fastställas. Man skall sträva efter att åstadkomma en uttömmande och särskiljande uppsättning termer för sorteringen.

I det första steget gäller det således att etablera de s.k. rangord som skall vara sorteringsgrundande. Olika tillämpningsområden har mycket olika konventioner för vad som skall utgöra rangord. Som ett exempel kan nämnas den bibliografiska fileringen, där bland annat vissa prepositioner och konjunktioner samt bestämda och obestämda artiklar (oavsett språk) inte får användas som första rangord; vissa artiklar kan dock vara homografa med exempelvis pronomen eller räkneord, vilka kan bli rangord. Ur den stora mängden bibliografiska regler kan också nämnas att tecknet & skall alfabetiseras som och, og, and, und, et, osv. alltefter titelns språk och att siffertal ordnas i stigande numerisk ordning. (Det förekommer också att man inordnar siffertalen enligt en alfabetisk form på det aktuella språket.)

Ett annat viktigt tillämpningsområde rör namnregister över personer, företag, etc. Här gäller ofta på likartat sätt att vissa led i personnamn negligeras (af, van, van der, von, etc.) och man använder inte attribut av typen Allmänna, Rikets, Statens, Svenska, Sveriges, osv. som första rangord.

I den lexikografiska alfabetiseringen är normalt alla ord som ingår i en ordboksartikels uppslagsform (och ev. ordklassbeteckning) rangord. (Underordnade uppslagsord eller -fraser samt hänvisningar kan dock vara primärt inplacerade med huvudordet som rangord i stället för förstaordet.) Eftersom vi här koncentrerar oss på de lexikografiska tillämpningarna, så behöver vi inte närmare gå in på hur rangorden etableras.

I det andra förberedande steget vid alfabetiseringen måste vi alltså bestämma i detalj vilka tecken som jämförelsen skall grunda sig på. Annorlunda uttryckt gäller det att etablera en ordningsform ur de rangord man tidigare valt ut. Detta led i alfabetiseringsproceduren har direkt koppling till det implementeringstekniska, vilket gäller genereringen av de sorteringsnycklar som datorn behöver för att styra ordnandet.

Ordningens form etableras i lexikografiska sammanhang i allmänhet enligt teckenprincipen (stavningsprincipen). Denna innebär att enheternas inplacering i den alfabetiska ordningen helt grundar sig på stavningen, inte uttal, betydelse, vanligaste stavningsform eller något annat.

Andra principer för etablerandet av rangordens ordningsform, som t.ex. efter ordens fonetiska form eller normalform (kanonisk form), kräver i allmänhet att varje element manuellt förses med en lämplig ordningsform. Här ligger också en stor risk för osäkra eller inkonsekventa bedömningar. Personnamnsregister är ofta alfabetiserade enligt normalformsprincipen (t.ex. telefonkatalogen) eller fonetisk form (exempelvis variantregister över släktnamn). Stavningsprincipen kommer här att ligga till grund för de fortsatta resonemangen.

[Beroende på materialets art, omfattning och presentationsform (internt arbetsmaterial, ordböcker för publicering, databaser, etc.) kan det finnas anledning att göra andra bedömningar än dem som fastslås i de följande avsnitten. Avsikten har emellertid i första hand varit att komma fram till ett så konkret förslag som möjligt vad gäller behandlingen av stora lexikaliska material.]

### Ordvis eller teckenvis jämförelse?

Grundprincipen för sorteringsgången är att jämförelsen sker ord-för-ord eller tecken-för-tecken och vid likhet upprepar man samma procedur för nästa element osv. När två enheter skiljer sig från varandra endast genom att det ena har extra tecken i slutet anser vi den kortaste komma först — "ingenting före någonting".

Skall man då alfabetisera teckenvis eller ordvis? Eller, annorlunda uttryckt: I vilken mån skall man ta hänsyn till ordgränserna då man alfabetiserar? Här går det inte att ge något självklart och allmängiltigt svar — det beror på materialets art och, inte minst, på det aktuella språket. Rent allmänt kan sägas att om språkbruket vacklar mellan särskrivning och sammanskrivning (med eller utan divis) så är tecken-för-tecken-metoden (där alltså ordmellanrummen är av underordnad betydelse) oftast att föredra därför att man då får samlat de förekommande varianterna på ett ställe; då man inte är säker på om ett uttryck är sär- eller hopskrivet slipper man alltså att söka på flera ställen. I engelskspråkiga register tillämpas därför i de flesta fall teckenvis alfabetisering.

För svenskans vidkommande torde i allmänhet ord-för-ord-principen ge bäst resultat. Flerordiga uttryck behandlas då inte som hopskrivna utan förs in direkt efter det första ordet som simplex och ordnas sinsemellan i andra hand på följande ord.

<u>ord-för-ord</u>	<u>tecken-för-tecken</u>
god	god
god frejd	godartad
god man	god frejd
god tro	godhet
godartad	godkänna
godhet	god man
godkänna	godmodig
godmodig	gods
gods	godtaga
godtaga	god tro

Ett viktigt problem i samband med sorteringsgången gäller frågan om på vilket stadium man skall ta hänsyn till accentuerade tecken, skiljetecken, osv. Vid alfabetiseringen sker, som vi strax skall få se, ordnandet i flera steg där man successivt tar hänsyn till bland annat variantformer av bokstäver, diakritiska tecken, versaler och jukturer. Vid den teckenvisa alfabetiseringen kan dessa modifieringar emellertid endast beaktas då hela ordningsformen i övrigt är identisk med en annan och inte på ordnivån, vilket är mer naturligt från lingvistisk synpunkt. Även detta faktum talar alltså för den ordvisa gången.

När alfabetiseringen sker ordvis är det således viktigt var ordseparatorer förekommer i de element som skall ordnas. Ett fall där bruket i svenskan kan vackla är abbreviationer och akronymer, vilka betraktas som flerordsuttryck om de skrivs med ordmellanrum: osv. respektive o. s. v. placeras därför in på olika platser, likaså SAS och S A S. (Problemen känner de flesta igen från telefonkatalogen.)

I vissa fall kan man vilja ge även andra tecken än ordmellanrum denna ordseparatorande status, exempelvis logogram av typen & och - (s.k. till-minus; ev. även divis). I alfabetiska register över ämnens kemiska formler kan man göra en "ord-för-ord"-sortering där versalitet fungerar som separator (C CO CO<sub>2</sub> CaCO<sub>3</sub>, ej C CaCO<sub>3</sub> CO CO<sub>2</sub>).

Frågan om ordvis eller teckenvis alfabetisering måste alltså avgöras från fall till fall med hänsyn till språk och materialets art. För de fortsatta resonemangen (som i första hand gäller svenskan) utgår vi i allmänhet från att alfabetiseringen sker ordvis.

Vad som tidigare sagts beträffande ordningsgången i en alfabetisk sortering — att man jämför tecknen genom orden från ordbörjan till ordslut — gäller den normala, initialalfabetiska sorteringen. Ord som börjar på likartat sätt kommer då att stå intill varandra i den sorterade listan.

I en finalalfabetisk sortering jämförs orden i stället från ordslutet och åt vänster, så att alla ord med likartat slut (samma sista sammansättningsled, avledningssuffix, böjningsändelse, etc.) grupperas samman. För speciella ändamål (såsom korsordsordlistor och morfemordnade ordlistor) kan man naturligtvis införa speciella regler för i vilken ordning jämförelserna skall ske. För att kunna lösa problemen med behandlingen av diakritiska tecken, versaler, junkturer, osv. permuterar man helt enkelt tecknen i ordningsformen på önskat sätt. Sorteringsnyckeln genereras sedan på normalt vis och några andra åtgärder behövs inte.

### Alfabetisk sortering av bokstäver

Vad som ligger till grund för en alfabetisk sortering är naturligtvis bokstävernas konventionella ordningsföljd i alfabetet. Såväl uppsättningen grundbokstäver som deras inbördes ordning varierar som bekant från språk till språk. Svenskan har aå\* b c d eé f g h i j k l m n o p q r s t u v w x yü z å ä ö, danskan och norskan har a b c d eé f g h i j k l m n o p q r s t u v w x y z æ ø å, isländskan aá b dö eé f g h ii j k l m n oö p r s t uú v x yý z þ æ ö, engelskan a b c d e f g h i j k l m n o p q r s t u v w x y z, tyskan aä b c d e f g h i j k l m n oö p q r s t u v w x y z, franskan aâ b ç d eéâê f g h iï j k l m n oö p q r s t uù v w x y z, spanskan aá b c ch d eé f g h ii j k l ll m n ñ oó p q r s t uú v x y z, osv.

Om vi återgår till svenskan, kan vi notera att det förekommer bokstäver som egentligen bär diakritiska tecken — i, j, å, ä, ö — men som konventionellt betraktas som egna, självständiga bokstäver. När vi i fortsättningen talar om bokstäver med diakritiska tecken avses inte denna grupp av etablerade grundbokstäver i det traditionella alfabetet.

Vi kan vidare se att där förekommer ligaturer och digrafer av typen æ (da., no., isl.), ch och ll (sp.) och ij (holl.), vilka i respektive språk, av fonematiska skäl, betraktas som enkla bokstäver i alfabetiskt hänseende, men som av andra språk oftast betraktas som likvärdiga med de särskrivna bokstäverna: æ = æ etc. (Tyskt ß = sz behandlas emellertid som ss och franskans e behandlas alltid som oe.)

I vilken mån skall då variantformer av bokstäver (i svenskan exempelvis w eller ü) och diakritiska tecken påverka den alfabetiska inplaceringen av ett ord? Skulle det vara acceptabelt att sätta w efter v och é efter e? I så fall skulle vi exempelvis få ordningen kilovara kilovis kilovolt kilovoltampere kilowatt kilowattimme respektive ide ideal ... ideologi idé idéassociation ... idévärld idiom. Detta är inte tillfredsställande. Vid den alfabetiska sorteringen måste hänsyn primärt tas endast till grundbokstaven, utan hänsyn till om den i själva verket förekommer i en variantform eller är accentuerad. Först om de jämförda orden i övrigt är lika, skall man beakta bokstavsvarianter och diakritiska tecken.

\*) Bokstäver som här står grupperade har samma primära sorteringsvärde.

Ett likartat problem gäller distinktionen mellan gemena och versala tecken. Skulle gemen alltid gå före versal (dvs. a A b B c C ...) så skulle t.ex. näve komma före Na och sten långt före Sten. På motsvarande sätt som nyss nämnts skall man därför alfabetisera primärt på en form där grundbokstaven inte ger någon distinktion mellan gemen och versal och först om orden i övrigt är lika bör man låta gemen, som språkligt sett får anses vara det omarkerade fallet, gå före versal.

Saammanfattningsvis kan alltså sägas beträffande alfabetisering av ord med enbart alfabetiska tecken att sorteringen i första hand skall grunda sig på en normaliserad form, där variantformer jämställs med grundbokstaven, diakritiska tecken negligeras och gemener och versaler betraktas som likvärdiga. Vid ordlikhet tas sedan hänsyn till variantformer, diakritiska tecken och versalitet, i nämnd ordning. Att det är lämpligt att testa i just denna ordning beror på att dessa distinktioner i de flesta fall specificerar en följd med avtagande semantisk differentieringsförmåga; den semantiska skillnaden mellan ordpar som sving / swing, tvist / twist, vatten / watten eller Myller / MÖller är oftast större än den mellan exempelvis a / à, filen / filén, jubileum / jubiléum, supen / supén eller Rosen / Rosén, som i sin tur är större än skillnaden gemen—versal: göteborgskonstnär / Göteborgskonstnär, tv / TV eller Saab / SAAB (i synnerhet om även "öakta" versaler finns med i materialet, t.ex. från ord som stått först i en mening eller först i en propriell syntaga). Det räcker således inte med en linjär ordningsföljd för alfabetets grundbokstäver, utan även de underordnade nivåerna av variantformer, diakritiska tecken, osv., måste beaktas och i sin tur få en intern rangordning.

Med tanke på att man i möjligaste mån vill gruppera ihop semantiskt närstående ord, är det viktigt att även jämförelsen med avseende på varianter, accenter och versaler sker i ordning från vänster till höger på ett sådant sätt att ord med störst likhet från ordbörjan räknat får stå tillsammans.

#### Behandlingen av abbreviationer och logogram

En speciell status bland orden har de olika typerna av förkortningar. De alfabetiska ordförkortningarna, abbreviationerna, är mycket ofta flertydiga (typen el., f., fr., min.). Vissa abbreviationer har också fått samma status som normala, självständiga ord och böjs som sådana: kolla, bil, stins, Saab, laser. Det är därför rimligast att abbreviationerna inordnas på samma sätt som övriga ord, dvs. enligt stavningsprincipen, oavsett vad förkortningarna härstammar från.

För logogrammen finns liknande skäl mot en inplacering grundad på en tänkt alfabetisk transkription. Men eftersom dessa tecken (av typen §, &, \*, %, osv) inte har någon hävdvunnen plats i alfabetet får de ordnas i en särskild följd, lämpligen före alfabetet (så att de lätt uppmärksammas; ord som står

efter alfabetets slut glöms lätt bort). I undantagsfall kan logogrammen även ingå som led i hybrida graford: 96X-ig, \$-kursen, 273°K, 24x36.

Vissa logogram kan, som tidigare nämnts, anses vara ordseparerande, t.ex. &, - ("till"; om man vill betrakta detta tecken som logogram) och matematiska tecken.

Enstaka tecken från främmande alfabet kan också betraktas som logogram. Så exempelvis bruket av grekiska tecken i naturvetenskap:  $\gamma$ -strålning,  $\beta$ -karoten,  $\Omega$ ,  $\mu\text{m}$ . I speciella fall kan man överväga möjligheten att i stället sortera på en translittererad form. Vi berör detta i avsnittet om suppleringsinformation för sortering.

### Junkturgrafemens behandling

Till kategorin jukturer räknar vi i detta sammanhang alla tecken som inte är bokstäver, diakritiska tecken, logogram eller siffror. Specialtecken och symboler utan uttalsform räknar vi således hit. Förutom de vanliga skiljetecknen (med agglutinerande, abbrevierande eller syntaktisk funktion) kan nämnas tryckaccenter, segmentgränser, parenteser och citations-tecken av olika slag.

Då man skall bedöma hur jukturerna skall behandlas i alfabetiseringen måste man bland annat ta hänsyn till att jukturerna normalt inte utläses, att bruket av t.ex. divis i sammansatta ord ofta är vacklande (all-rum / allrum, icke-religiös / ickerreligiös, u-land / uland, Väst-tyskland / Västtyskland, etc.), likaså användningen av punkt i ordförkortningar (U.S.A. / USA, d.v.s. / dvs., kr. / kr) och kolon, apostrof och vissa andra skiljetecken vid förkortning, böjning och avledning (typen sta'n, så'n, Lars', '82, TV:n, SSU:are, H:son), för att nu ta några exempel.

Jukturerna bör därför påverka den alfabetiska sorteringsordningen endast då man på grundval av den rent alfabetiska delen av orden inte kan avgöra ordningen och man följer då en intern rangordning mellan junkturgrafemen. Också i detta fall bör man sträva efter att gruppera samman enheter som börjar på likartat sätt.

Vid teckensvis alfabetisering bör ordmellanrum och annan utslutning betraktas som jukturer.

Det kan ibland vara önskvärt att även ge vissa jukturer ordseparerande status (exempelvis ' och - i franska material). Detta beaktas i så fall lämpligen redan då ordningsformen skapas.

### Problemen med siffertalen

Hur stort inflytande skall siffertalen ha på den alfabetiska ordningen? Vid alfabetiseringen bör målsättningen vara att,

inom ramen för ett formaliserbart förfarande, försöka gruppera samma semantiskt närstående ord. Man har därför större skäl att hålla samma grupperna

00-talet	2-åring	2-sidig
10-talet	21-åring	4-sidig
1000-talet	50-åring	8-sidig
1800-talet	75-åring	96-sidig
1930-talet		
80-talet		

än en grupp med större semantiska skillnader, som

50-lapp  
50-metersbassäng  
50-minutersprogram  
50-talslåt  
50-åring  
50-årsdag.

När man ser detta ur språkvetenskaplig synvinkel bör slutsatsen bli att siffrorna inte skall vara primärt sorteringsgrundande, som de alfabetiska tecknen, utan endast påverka ordningen då den alfabetiska delen av orden är identisk, på samma sätt som junkturerna. Att på något sätt väga om junkturerna eller siffrorna påverkar den semantiska uttolkningen mest, är nog att gå för långt; den vinst som eventuellt skulle kunna uppnås genom införandet av ytterligare en nivå i sorteringen uppvägs säkert inte av att användaren då kan få svårt att orientera sig i förteckningen. Det förefaller därför rimligt att utföra sorteringen på två huvudnivåer — först på den alfabetiska delen av ordet (inklusive logogram) och därefter på övriga tecken, dvs. junkturer och siffror. Ord av typen 491, 0,5, 10:50, 1981/82, etc., som saknar alfabetisk del kommer då följaktligen att inordnas först.

I en lexikologisk ordförteckning finns det inga skäl att frångå teckenprincipen och ordna de numeriska leden i talordning. Skulle talordning införas måste rimligen även decimalbråk, allmänna bråk, romerska tal m.m. ordnas efter samma princip. Problemet är att en mycket stor andel av enheterna kan vara homonyma eller polysema: 1,250 kan vara "ett komma ..." eller "ett tusen ...", 2:10 kan vara priset 2 kronor och 10 öre eller 1/5, 2/3 kan vara 2 mars eller en musikalisk taktangivelse eller ett matematiskt förhållande, 19.30 kan vara ett klockslag eller ett pris osv. ofta har också den numeriska delen av ett ord en proprielikhande funktion, som i årtal eller varubeteckningar (1981, 80-talet, 343, 4-2-4-systemet, 4711, 5A, 8:an).

Att generellt inordna numeralerna enligt en fullständig alfabetisk form är självfallet i allmänhet ogörligt, bland annat på grund av de nämnda homonymiproblemen och problem med böjda och avledda numeraler (skall 1 inordnas som en, ett, första eller förste; skall 1800 tolkas som ettusenåttahundra, ettusen åtta hundra, artonhundra eller möjligen adertonhundra?).



Den enda rimliga lösningen, för våra syften, är alltså att strikt tillämpa teckenprincipen på siffrorna och junkturerna då den alfabetiska delen av orden är lika.

För speciella tillämpningar finns alltid möjligheten att sortera efter en manuellt supplerad ordningsform. I t.ex. bibliografiska sammanhang, där man har tillgång till ordens kontext i titeln, kan detta vara genomförbart. Man bör emellertid vara mycket restriktiv med inordning på en form som ej är explicit och fullständigt entydig, förutom att man även av andra skäl vill minimera den manuella insatsen.

### Supplering och negligering av information för sortering

I vissa sammanhang kan det finnas skäl att åsätta en enhet en ordningsform som kanske inte kan genereras automatiskt, med formaliserbara kriterier. Som tidigare berörts kan detta gälla exempelvis inordnandet av romerska tal (i t.ex. regentlängder vill man kanske ha ordningen I, II, III, IV, V ... och inte I, II, III, IV, IX, V), behandlingen av ord med icke-latinska skrivtecken, samsortering av olika stavningsvarianter under en normalform eller bibliografisk filering där bland annat vissa ord ej får vara rangord. Det är därför ibland av vikt att man har möjlighet att markera att vissa tecken skall negligeras i rangorden, och omvänt, att man skall kunna modifiera rangord eller suppleras med ytterligare rangord för att åstadkomma en ordningsform som ger den avsedda ordningen. Dock bör man alltid se upp med manuella ingrepp som innebär en risk för inkonsekvent behandling av enheterna.

### Sammanfattning

För att kort sammanfatta denna översikt över de viktigaste problemen vid automatisk alfabetisering av lexikografiskt material kan sägas följande. I ett första steg utväljes de rangord (uppslagsform, ev. ordklass etc.) som skall alfabetiseras. I nästa steg etableras en ordningsform på grundval av rangorden. Ordningensformen har exempelvis spjälkat æ till ae, eventuellt translittererat þ till th, ð har blivit ss, i spanska material har ch och ll fått unika koder och ordseparatorer kan ha inskjutits kring vissa logogram och jukturer. Är sorteringen en annan än initialalfabetisk permuteras sedan denna ordningsform. Därefter genereras den sorteringsnyckel som skall användas vid den maskinella jämförelsen. Sorteringsnyckeln skall vara konstruerad på sådant sätt att hänsyn i en första omgång endast tas till alfabetiska tecken (och logogram) och först då de alfabetiska tecknen överensstämmer beaktas övriga tecken (jukturer och siffror).

För att återgå till den alfabetiska delen, tillgår jämförelsen där på fyra underordnade nivåer: först jämförs grundbokstäverna i orden, därefter (dvs. vid likhet mellan orden) tas hänsyn till eventuella varianttecken, diakritiska tecken och

versaler, i nämnd ordning. Inom varje nivå skall således finnas en intern rangordning mellan tecknen (mellan grundbokstäverna, variantformerna, de diakritiska tecknen, gemener/versaler respektive övriga, icke-alfabetiska tecken). Jämförelsen bör ske ordvis och på ett sådant sätt att ord som börjar på likartat sätt ordnas intill varandra.

Avslutningsvis bör framhållas att det är viktigt att man, åtminstone i mer omfattande alfabetiska register, tydligt beskriver de alfabetiseringsprinciper man tillämpat. Åtminstone tills dess att en standard på området har etablerats.

[En något utförligare presentation med bibliografiska referenser och en modell för implementering (med flödesplan och kodtabeller) finns i Rapport nr 14 från Språkdata. Den kan beställas från Språkdata, Göteborgs universitet, Norra Allégatan 6, S-413 01 Göteborg.]