

Transkription och direktglossning av dialektinspelningar i SveDiaSyn

I SveDiaSyn finns det två tillvägagångssätt då inspelningar ska överföras till skriftspråk: transkription och direktglossning. En transkription innebär att man gör en ljudnära överföring till skriftspråk och direktglossning innebär att man inte tar hänsyn till uttal, utan gör en direkt överföring till svensk skriftstandard (med vissa viktiga undantag).

transkription: Å sen samtidit såm dän låckar till säj fågel så # hämtar dän ju fågel då ##

direktglossning: Och sedan samtidigt som den lockar till sig fågel så hämtar den ju fågel då

Vi kommer att transkribera dialekter som ligger förhållandevis långt från rikssvenska, medan andra dialekter kommer att direktglossas. Transkriberade dialekter glossas i ett senare skede.

De särdrag som utmärker dialekterna kommer att specificeras i en kort kommentar för varje dialekt – därför är det önskvärt att varje transkribent noterar dialektala egenheter som kan nämnas i denna kommentar. Se vidare nedan.

En del filer är rätt långa – bearbeta då bara de första 30 minuterna av ljudfilen. Många ljudfiler börjar med upplästa ordlistor; börja i så fall inte glossa/transkribera förrän det spontana talet kommer igång. Uppge var din text börjar i CHAT-filens början (alltså efter 40 min. 32 sek, till exempel). Om en ljudfil visar sig vara oväntat svårbearbetad så ska det noteras – i sådana fall stämmer ju inte våra arbetstidsberäkningar. Detsamma gäller om en fil är oväntat enkel att jobba med.

Transkriberade filer ska märkas med <tr> i filnamnet. Ett exempel: sprogeOM3tr.cha. Glossade filer ska inte märkas på detta sätt (se vidare nedan).

Innan du börjar transkribera kan det vara bra att lyssna igenom ljudfilen, så att du vet ungefär vad du har att vänta och så att du under transkriptionen inte plötsligt upptäcker att du sitter och lyssnar till en spännande berättelse och glömmer att transkribera...

TRANSKRIPTIONSANVISNINGAR

Transkriptionen kan beskrivas som en ljudenlig stavning som ligger nära en fonematisk transkription. Det finns dock vissa uppenbara begränsningar eftersom vissa ljuddistinktioner inte markeras, som exempelvis främre och bakre /r/. Vi har bedömt att läsbarheten och tillgängligheten av transkriptionen är det primära, på bekostnad av intressanta ljud-distinktioner som inte kan markeras. Principer för transkriptionen:

- försök komma så nära uttalet som möjligt i transkriptionen. Det bör alltså finnas en så direkt koppling som möjligt mellan bokstav och ljud.
- de allofoner som inte framgår av transkriptionen kan finnas med som kommentarer och tips på vad man särskilt kan lyssna efter, t.ex. variant av sje-ljud, tjocka /l/ eller förekomsten av både främre och bakre /r/ som i t.ex. västgötamål.
- dialektala ord: när du stöter på ett dialektalt ord finns det två strategier för att hantera det. Om du vet vad ordet betyder, och om det har en direkt etymologisk motsvarighet i rikssvenska så ska du använda det rikssvenska ordet och ange det dialektala inom hakparentes:

fähus [föjs]

Om du däremot inte vet vad ordet betyder, och/eller om en direkt motsvarighet saknas så ska du skriva det dialektala ordet och markera det som dialektalt:

föjs@d

Du kan också, med denna senare metod, infoga ett frågetecken eller en möjlig översättning inom hakparentes:

föjs@d [fähus?]

KONSONANTER

R-ljud stavas alltid <r> oavsett kvalitet.

L-ljud stavas alltid <l> oavsett kvalitet.

J-ljud stavas alltid <j>. Ex. <juren, jorde, ejentlien>.

G-ljud stavas alltid <g>; undvik därför oftast finalt <g> i t. ex. <rikti, vikti, kånsti>.

Sje-ljudet (som i *station, kanske, choklad, sjö*) stavas alltid <sj>. Ex. <stasjon, kansje, sjoklad>.

Supradentaler (som i slutljudet i *fors, bord, kort, sorl*) markeras inte särskilt, utan skrivs enligt vanlig skriftstandard: <rt>, <rs>, <rl>, <rd>. Ex. <bårt, fårs>.

Tje-ljudet (som i *kök, kjol, tjatträ*) stavas alltid <tj>. Ex. <tjök, tjol, tjatträ>.

Konsonantreduktioner är vanliga i talspråk. Två exempel på vanliga konsonantreduktioner:

Ex. 1. <ja # de ble ju inte så mycke nä man inte hade bil å barna va små >

Ex. 2. <å få fö si å gå ut å så>

Några bokstäver och bokstavskombinationer har komplexa relationer till fonemen. För transkriptionerna gäller följande:

Grafemet <c> stavas alltid <s>. Ex. <presis, syklåp> (undantag: nedanstående <ch>, samt de fall då <c> ska återges med <k>: *cupfinal*, första förekomsten i *cancer* etc.).

Digrafen <ch>, som i *och*, stavas alltid <ck>.

Grafemet <x> stavas alltid <ks>. Ex. <eksakt, seks>.

VOKALER

Tecknet <e> ska reserveras för dialekter med verkligt slutet [e], använd annars <ä>; var konsekvent (så långt det är möjligt).

Ex. 1. <reven freste> för *räven fräste*.

Ex. 2. <dä ä inte mä mä dä> för *det är inte mer med det*.

Tecknet <o> betyder alltid sluten vokal (som i *ost, mod, bok*); öppnare vokal skrivs <å> (<båll, båta>)

Diftonger: de mest iöronfallande diftongerna (tex skånska /u/, /o/ och /a/, eventuellt även /i/, /y/ och /e/) markeras med två vokaltecken.

Ex. från Fårö: <dåm har trei åolika värsjåoner åm samma bit>

DIREKTGLOSSNINGSANVISNINGAR

Vid direktglossning gör man en nedteckning av inspelningarna som inte är ljudnära. Alltså stavas ord så som de stavas i rikssvensk standard. Det finns två mycket viktiga punkter att tänka på:

- ordföljden i inspelningen ska återges exakt.
- antalet ord i inspelningen ska återges exakt. Man ska alltså göra en ord-för-ord-nedteckning av dialektinspelningen, utan att lägga till eller ta bort ord.

Båda dessa principer kan leda till att glossningen inte motsvarar standardsvenska – detta är en nödvändighet för att eventuella syntaktiska särdrag i de respektive dialekterna ska vara synliga i glossningen. I norrländska dialekter förekommer till exempel vissa sammansättningar (se ex. 2 nedan) som man ska återge på samma sätt i glossningen. Notera också negationens placering i bisatsen i ex. 2.

Ex. 1. transkription: <Nä ä i fel litä ed>
glossning: <Nej det är väl lite det>

Ex. 2. transkription: <Men e jer läiga bra ad **änt** e kåmm **våkkerveire** firi missomar.>
glossning: <Men det är lika bra att **inte** det kommer **vackervåddret** före midsommar>

I vissa fall kan det vara svårt att finna motsvarigheter till dialektala konstruktioner – notera i så fall dessa svårigheter och gå vidare i glossningen.

Markera pauser också vid glossning – det gör ev. senare transkription lättare.

KODER

Vid transkription är det brukligt att ange pauser, ohörbart tal och så vidare med kodtecken. Det finns förutbestämda kodtecken i Chat/Clan (nedladdningsbara manualer finns på <http://www.cnts.ua.ac.be/childes/>), men de vanligaste och viktigaste är förtecknade nedan.

Pauser i meningar markeras med <#>, <##> eller <###> (kort paus, lång paus och extralång paus).

Ex. <å sen på vinten få man ju ta in då # fö då gå då ju i ide ## åck ### eh ##
ejentlien säjs de att man ska lägga då i tjylskåpet öve vinten>

Ohörbart tal markeras med <xx> (ett ord) eller med <xxx> (flera ord).

Oavslutade meningar eller omstarter markeras med <[-]>

Ex.1. <så ja säje [-] å likdant mä kläder å allting>

Ex.2. < min fåssa han jobba ju me [-] de e # ejentlien så e de>

Påståenden, frågor och utrop/imperativer markeras med punkt, frågetecken och utropstecken. Varje replik i Chat-format ska avslutas med ett skiljetecken.

Tvekljud skrivs som <eh, hm, em> osv. Försök begränsa variationen.

Ex. 1.<för de # eh ## de finns ju inte undelag>

Ex. 2.< å # eh # sen ä de väl # ja, såna dä småjåbb man kan ta, vikariat å så där. ja ## eh #
ska kansje jåbba såm # em # fritisledare>

Upprepningar markeras i transkriptionen.

Ex. <så de, men de va ju en, en # arbetsam tid # på de, de viset, att de va mycke handarbete>

Stakningar markeras i transkriptionen.

Ex. < inte en männisja såm vå vågade bygga där sen>

Skratt markeras inte.

Dialektala ord: vid dialektala ord så som *glyttarna*, ange att det är en dialektal form med kodtecknet <@d>. Skriv alltså <glyttarna@d>. I kommentaren kan sådana ord förklaras. Alternativt kan man ge en förklaring inom hakparentes (om man är säker på vad ordet betyder) – se ovan. I en del dialekter är emellertid andelen dialektala ord mycket hög – då ska inte denna metod användas.

Intervjuarens tal bör också transkriberas.

Ignorera så många störningar som möjligt t.ex. bakgrundsljud i pauser, intervjuarens feedback-signaler, onödiga frågor eller kommentarer från intervjuaren.

Kommentarer och förklaringar kan sättas inom hakparenteser. Använd bara kommentarer och förklaringar om texten annars är obegriplig.

Extra kodtecken

Följande kodtecken kan också användas när det är nödvändigt:

upprepning: <ord/fras/mening>[/]

bästa gissning: <ord/fras/mening>[?]

talet självdör: <ord/fras/mening>+...

avbrott: <ord/fras/mening>+./

Använd dessa tecken restriktivt.

CHAT-FILEN

Hur sätter man bullets?

Arbeta i Sonic Mode, markera en bit ljud och klicka sedan på den lilla symbolen *s* som finns längst till vänster på ljudgrafiken – sätt först markören på den plats i texten där du vill sätta en bullet, alltså vid yttrandets slut.

Namn på filer och deltagare?

Använd namnen som finns i ljudfilen; OW1, OM2 och så vidare. Deltagarnas riktiga namn får *inte* skrivas ut! Skriv NN när den som ska intervjuas presenterar sig. Intervjuaren/-na kan kallas INT (och eventuellt INT2).

Namnge CHAT-filerna efter ljudfilerna. Exempel:

ljudfil: ste_om_1sp.wav namn på CHAT-fil: stenbergOM1.cha

Ortnamnet ska alltså skrivas i sin helhet i CHAT-filens namn.

Headers?

Använd följande mall för headers:

@Begin
@Languages: sv
@Participants: OM1 äldre_man1 Adult, INT intervjuare Investigator
@ID: sv||||male||||
@Coder: Henrik Rosenkvist
@Location: Stenberga
@Comment: ljudfilen finns i SweDias material, i Stenbergamappen
(ste_om_1sp.wav)
@Warning: transkription, INT ej transkriberad utan glossad

Man ska alltså ange vem som gjort transkriptionen/glossningen, ljudfilens namn och orten där inspelningen gjorts. Efter @Warning ska man ange om filen innehåller en transkription eller en glossning.

Samtidigt tal?

Det är inte nödvändigt att markera samtidigt tal. Skriv replikerna var för sig, även om de överlappar i tiden. Bulletmarkeringarna ska dock överlappa (så att man kan koppla hela yttandet till respektive del av ljudfilen).

Om delar av det överlappande talet inte kan uppfattas så använd markeringen för ohörbart tal (<xxx>).

Yttrande/makrosyntagm?

En del repliker i materialet är väldigt långa. Bryt då upp dem, så att ett och samma yttrande delas upp i makrosyntagmer. Exempel:

- *INT: när man bestämmer sig för att renovera ett sådant hus liksom kan du beskriva ordningen vad är det man gör för någonting till att börja med?
- *OM1: jaaa för å få byggnadsvårdspengar från länsstyrelsen så måste man ju ta kåntakt mä ## kulturenheten åck Per-Olof Renkvist såm ä länsantikvarie då.
- *OM1: å å diskutera fram va såm ska jöras å få en plan på det hela ## åm dä ä nånting värt att bevara över huve taget.
- *OM1: men just här i byn ä ju länsstyrelsens inställning att detta ä ju ett storåbjekt såm ska satsas på till hunnra präsent.
- *OM1: så dä kårmer ju å bli hus efter hus här nu ### såm tas itu mä.
- *OM1: så dää ## ett tre såm du säger där åppe ## dä dää ## ska ja väl ha hand om nästa år ## så ja anställer en snickare då så vi ska jöra i årning dä tillsammans.

OM1 talar här ganska länge, men tack vare att hans långa yttrande bryts upp blir det lättare att göra en syntaktisk analys; det blir också enklare att hitta i materialet eftersom vi kan sätta fler bulletar.

<e >eller <ä>?

Det kan vara svårt att välja mellan <e> och <ä> när man transkriberar. Ordet *med* kan ju transkriberas som <me, mä, med, mäd>. Detta är emellertid inte en viktig sak för just detta projekt – välj den variant som du tycker ligger närmast i början av transkriptionen och använd den sedan i hela transkriptionen.

En svensk CHAT-manual finns på:

<http://www.sol.lu.se/humlab/education/uploads/clantrans.pdf>

RAPPORTERING/INLÄMNING AV FILER

Skicka färdiga Chat-filer till:

Henrik.Rosenkvist@nordlund.lu.se

Sammanställ dina arbetade timmar när du tycker att det passar, och skicka därefter sammanställningen, som bör innehålla en specifikation av vad du har gjort, till Lars-Olof (Lars-Olof.Delsing@nordlund.lu.se), så kommer pengarna så småningom. Räkna med att det tar lite tid.