

# **Transcriber**

## DGCSS' udskrivningsmanual

Oktober 2007

### Resume:

Dette dokument beskriver, hvordan udskrivning på DGCSS skal foregå i Transcriber, med henblik på at udskriften senere konverteres til DGCSS' Praat-format. Denne manual beskriver både brugen af Transcriber-programmet og selve udskrivningskonventionerne, så det er ikke nødvendigt at have kendskab til tidligere DGCSS-udskrivningsmanualer for at kunne bruge denne manual.

## Forord

Dette er Sprogforandringscentrets udskrivnings- og korrekturmanual til udskrivningsprogrammet Transcriber. Princippet i manualen er, at udskrivningen skal være simpel, hurtig og konsistent, og det prioriteres højt, at udskrivningen og korrekturlæsningen foretages efter så klare og ens principper, at de færdige udskrifter ikke på nogen væsentlige punkter afviger fra hinanden. Derfor tages der stilling til så mange tvivlsspørgsmål som muligt i selve manualen.

I manualen bliver der gjort rede for både den tekniske del af udskrivningsproceduren i Transcriber, de specifikke konventioner, DGCSS' data udskrives efter, og proceduren for korrekturlæsning af udskrevne filer.

Sprogforandringscentret har oprindeligt brugt programmet Clan til at udskrive i, men skiftede i 2006 over til Transcriber, primært fordi det sikrer en god sammenhæng mellem lyd og tekst, samtidig med at det er et af de mere brugervenlige programmer i genren. Programmet er open source og kan frit hentes fra <http://trans.sourceforge.net>, der også huser online-manualen. Online-manualen beskriver i detaljer, hvordan programmet fungerer, mens nærværende udskrivningsmanual beskriver, hvordan programmet benyttes på DGCSS.

Efter læsning af manualen fra start til slut kan indholdsfortegnelsen bruges som huskeliste under udskrivningsarbejdet.

Manualen er sidst redigeret af:

Torben Juel Jensen  
Janus Møller  
Line Visby Nielsen  
Nina Kanstrup Hansen  
Louise Gad

## Indholdsfortegnelse

UDSKRIVNING .....	4
1. Transcriber .....	4
1.a. Opstart af Transcriber .....	4
Første gang .....	5
Transskriptions-konfiguration .....	5
1.b. Noget om lyd .....	6
1.b.1 Boundaries .....	6
1.b.2 Flyt boundaries .....	7
1.b.3 Slet boundaries .....	7
1.b.4 Navigation .....	7
1.c. Taleture .....	7
1.d. Overlappende tale .....	8
1.e. Når en fil er færdig – hvordan den gemmes .....	10
1.e.1 Kør automatisk stavekontrol .....	10
2. Regler for udskrivning af tale .....	10
2.a. Hovedregel: Der udskrives i almindelig ortografi .....	11
2.a.1 Ordstammer skrives jf. RO .....	11
2.a.2 Bøjningsendelser skrives, som de bliver sagt, men skal stå i RO .....	11
2.a.3 Hvis man ikke kan høre forskel, vælges den korrekte form .....	12
2.a.4 Sammentrækninger skrives ud .....	12
2.b. Afvigelser og tilføjelser i forhold til retskrivningen .....	12
2.b.1 Talord skrives ud .....	12
2.b.2 Store bogstaver bruges kun i begrænsede tilfælde .....	12
2.b.3 Forkortelser skrives, som de siges .....	13
2.b.4 Ord, der ikke findes i RO .....	14
2.b.5 Særlige konventioner for ikke-ord, der skal udskrives som ord .....	14
2.c. Tegnsætning .....	15
2.c.1 Der sættes ikke interpunktionstegn .....	15
2.c.2 Accenter bruges kun i ikke-danske ord .....	15
2.c.3 Underscore _ .....	15
2.c.4 Bindestreg - .....	17
2.c.5 Apostrof .....	17
3. Kommentarer og forhold i taleproduktionen .....	17
3.a. At indsætte en event .....	17
3.b. Lydlige begivenheder .....	18
3.c. Kommentarer .....	18
3.d. Baggrundsstøj .....	19
3.e. Pause # .....	19
3.f. Afbrudt og genoptaget ord markeres med bindestreg - .....	20
3.g. Tvivlsom passage {...} .....	20
3.h. Uforståelig passage xxx .....	20
3.i. Forlængelse : .....	20
3.j. Emfase ! .....	21
4. Ting, der kræver særlig opmærksomhed .....	21
4.a. Er alle ord og morfemer repræsenteret i udskrifterne? .....	21
4.b. Er der for mange ord i udskriften? .....	21
4.c. Afspejles talestrømmens ordrækkefølge direkte i udskriften? .....	21
4.d. Skal der stå 'at' eller 'og'? .....	21

## UDSKRIVNING

I de følgende afsnit beskrives reglerne for udskrivning af lydoptagelser af samtaler i DGCSS' datamateriale, dvs. interviews og gruppesamtaler. At udskrive samtaler er at overføre sproglyde til skrift efter et fastlagt regelsæt, og formålet med udskrivningsprocessen er at gøre indholdet i samtalerne tilgængeligt for bl.a. fonetisk og grammatisk analyse.

I afsnit 1 gives en indføring i, hvordan man bruger Transcriber til udskrivning, i afsnit 2 gennemgås de generelle regler for udskrivning af ordene i samtalen, og afsnit 3 indeholder konventionerne for notering af forhold i taleproduktionen, bl.a. pauser, overlap og afbrydelser, samt udskriverens kommentarer til udskriften. I afsnit 4 listes en række ting, man skal være særligt opmærksom på, når man udskriver talesprog.

### 1. Transcriber

Transcriber-programmet er specielt udviklet til at udskrive radioudsendelser og har en lang række funktioner, der gør det muligt at inddele lydfilet i segmenter (for eksempel udenrigs, sport, vejret og musik) og at tilføje avancerede diskursanalytiske kommentarer. Men vi bruger det kun til den rene udskrivning – og når udskriften er færdig bliver den konverteret til et format, der kan læses af programmet Praat, som benyttes til efterfølgende analyse og opmærkning. Denne konvertering foregår helt automatisk, men den betyder blandt andet, at der ikke kan rettes fejl i den oprindelige udskrift efter konverteringen.

Derfor er det vigtigt at udskrifterne overholder de konventioner, der beskrives i denne manual. I Transcriber findes der en lang række indbyggede kodnings-standarder til at markere forskellige para-lingvistiske fænomener såsom pauser, lyde og ændring af prosodi, men på DGCSS benytter vi en lidt anden kodning end den, der beskrives i Transcribers online-manual og indbyggede hjælp. Det er dels for at lette overgangen for udskrivere, der er vant til vores hidtidige udskrivningskonventioner, men også for at gøre det så enkelt og hurtigt som muligt at udskrive den type samtaler, vores datamateriale består af.

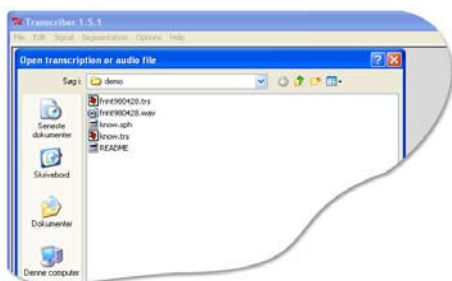
Fordi Transcriber er beregnet til radioudsendelser, hvor der som regel kun er en person, der taler ad gangen, og folk sjældent afbryder hinanden, er det ikke særligt velegnet til gruppesamtaler. Og programmets indbyggede håndtering af overlap er heller ikke velegnet til typiske sociolingvistiske interviews, hvor en stor del af overlappene består af korte returytringer som "ja", "nå" og "mm". Derfor benytter DGCSS en anden notering af overlap – dette beskrives mere indgående i afsnit 1.d.

I det følgende beskrives, hvordan programmet Transcriber benyttes, hvordan man navigerer i en lydoptagelse, og hvordan tale udskrives

#### 1.a. Opstart af Transcriber

Transcriber er et brugervenligt program, der husker, hvem du er, og hvad du lavede sidst, så når du først er i gang, kan du som regel bare fortsætte dit arbejde, hvor du slap.

Når du starter Transcriber, bliver du straks bedt om at åbne en transskription, dvs. en udskriftfil, eller en lydfil. Transskriptionernes filnavne ender med .trs, mens lydfiles navne ender med .wav. Hvis du endnu ikke har skrevet noget, skal du finde en lydfil – dvs. den lydfilet du skal i gang med at udskrive.



Alle lydfilet findes på vores fælles server – populært kaldet L-drevet. Gå ind i mappen "lydoptagelser" og find undermappen for det projekt, der indeholder din lydfilet. Når du har fundet lydfilet, viser Transcriber i et lille vindue, at den indlæser lydfilet. Det kan godt tage 2-3 minutter, og det er generelt en dårlig ide at begynde at skrive noget, før vinduet er væk. Næste gang du indlæser den samme lydfilet, skal du ikke vente nær så længe.

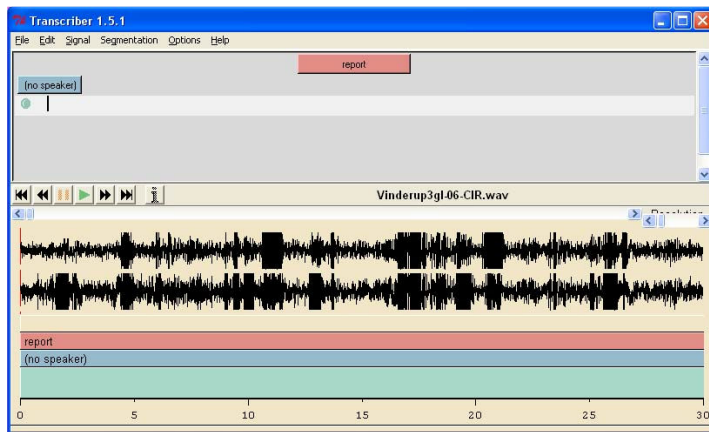


Når lydfilen er læst ind, kan du se, hvordan programmet er opbygget – det er ret ligetil. Øverst vises transskriptionen. En rød boks viser hvilken sektion, du er ved at transskribere, og en blå boks viser hvilken talers tur, der er i gang.

*Bemærk: Programmet har en særhed, der gør, at det ofte bliver vist lidt for langt nede på skærmen. Dobbeltklik på vinduets blå titelbjælke for at få det til at placere sig normalt.*

Under transskriptionen er der nogle rækker til at afspille lyd, og umiddelbart derunder kan du se selve lyden.

Nederst er der tre farvede spor: et rødt, der viser opdelingen i sektioner, et blått, der viser opdelingen i talerture, og et grønt, der viser, hvad de enkelte talere siger. De røde sektioner bliver ikke brugt til noget i DGCSS-øjemed.



## Første gang

Når du åbner programmet første gang, skal du tilpasse det ved at lave nogle få indstillinger, men når du én gang har gennemgået proceduren beskrevet i det følgende, skal du ikke gøre dette mere.

Gå op i menuen "Options", og vælg "General..."

I det vindue, der kommer frem, skal du i "Default scribe's name"-feltet skrive dit fulde navn.

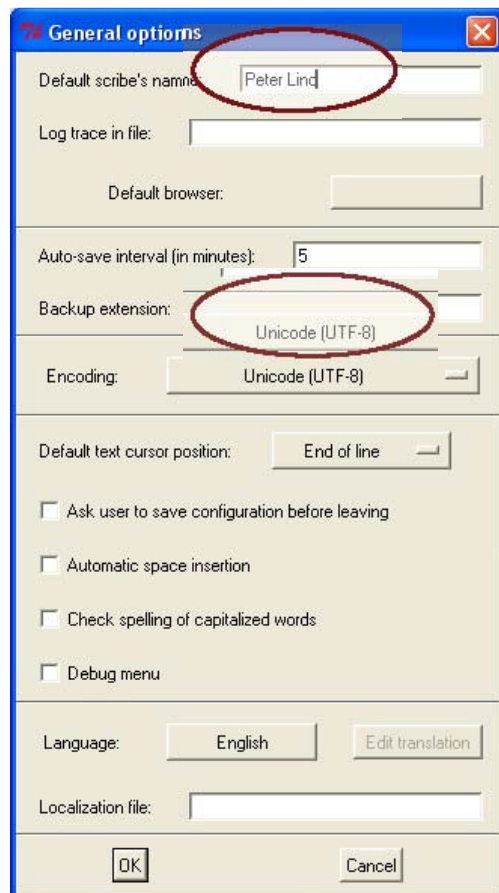
Og i "Encoding"-feltet skal du vælge "UTF-8".

Du behøver ikke udfylde andre felter – men du kan ændre "Auto-save interval" til et lavere tal, hvis du vil have hyppigere backups. Skriv fx 1 for at gemme hele tiden.

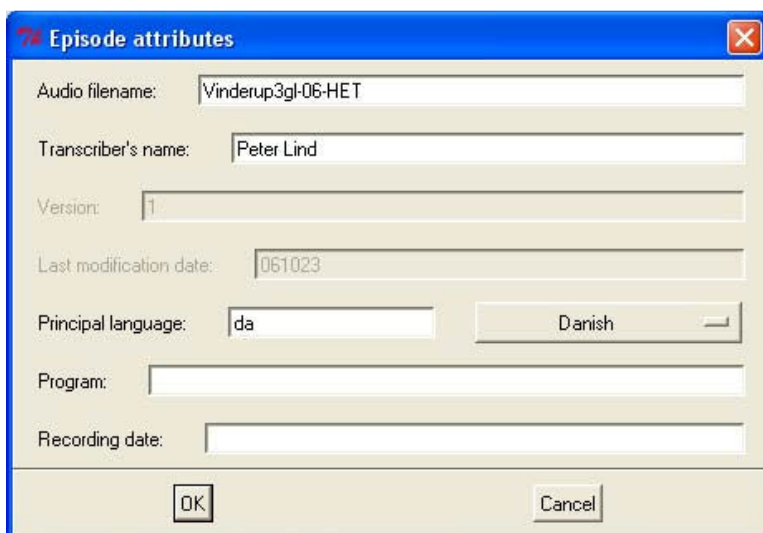
*Bemærk, at den nederste knap, hvor der står "Language", ikke vælger sprog for transskriptionen, men derimod for programmet. Og det er endnu ikke tilgængeligt på dansk.*

Tryk på "OK".

Gå derefter op i menuen "Options" igen, og vælg "Save configuration".



## Transskriptions-konfiguration



Når du har konfigureret programmet, skal du konfigurere selve transskriptionen.

Gå op i menuen "File", og vælg "Edit episode attributes..."

En transskription i Transcriber hedder en "episode".

Forhåbentlig er filnavnet for lyden korrekt, ligesom dit navn også burde være taget fra det, du lige har indtastet. Hvis ikke, så ændr det.

Tryk på knappen ud for "Principal language", og vælg "Danish" – forudsat at transskriptionen primært er på dansk selvfølgelig. Det valgte sprog bestemmer, hvilken stavekontrol der skal benyttes.

De to sidste felter, "Program" og "Recording date", er beregnet på radioprogrammer og er derfor ikke relevante for udskrifterne.

Tryk "OK" – og du er klar til at udskrive.

### 1.b. Noget om lyd

Lydoptagelsen ligger som en lydfil, separat fra udskriften, men alligevel koblet til den. Nederst i Transcribers vindue kan man "se" lyden i form af et billede over svingningerne. Ideelt set burde det være til at se præcis, hvornår der er tale, og hvornår der ikke er – idet alt det sorte er lyd, og alt det ikke-sorte er stilhed. Jo højere de sorte streger går op og ned, jo kraftigere er lyden det sted i optagelsen.



Der vises altid kun et lille udsnit af lydoptagelsen, og med scrollbaren, der ligger lige oven over sporet, kan man vælge, hvilket udsnit det skal være. Med den lille scrollbar ude til højre kan man zoome ind og ud på lyden og dermed vise et større eller mindre udsnit.

Desværre har Transcriber en tendens til at gøre de sorte streger alt for små eller alt for store, så de er ikke altid lige anvendelige. Men det er tit nødvendigt at kunne zoome ind eller ud for at flytte boundaries præcist.

#### 1.b.1 Boundaries

Samtidig med at man kan se lyden i form af de sorte svingninger, kobler udskriften sig også til lyden. Hver gang man trykker **Return**, bliver der skiftet linje i teksten, og hver linje forbindes til et interval i lyden. Den grønne boble i starten af hver linje svarer til et skel imellem to intervaller nede i det grønne talespor nederst i vinduet. I eksemplet herunder bliver der altså sagt "mm hh men har de så meget tid" i intervallet fra 5:40 til lige knap 5:43. Intervaltiden er angivet i minutter og sekunder.

A screenshot of the transcription interface. The top part shows a list of transcription lines with green circles indicating boundaries. The text includes: "ikke men # min mor hun arbejder nogle gange hjemme foran computeren", "# ofte", "mm hh men har de så meget tid", "eller har de meget fritid din far og mor", "m- altså # min far han # ja: men han har altså han ha- kommer hjem klokken # sådan seks stykker om aftenen eller sådan noget". Below the text is a control bar with play/pause buttons and a volume icon. Underneath is the audio waveform. At the bottom is a timeline with colored segments corresponding to the transcription lines. The segments are labeled with the text and their start/end times: "ikke men # min mor hun arbejder nogle gange hjemme foran computeren" (5:36-5:38), "# ofte" (5:38-5:40), "mm hh men har de så meget tid" (5:40-5:42), and "eller har de meget fritid din far og mor" (5:42-5:44). The segments are color-coded: blue for the first, green for the second, and red for the last two.

Man bruger **TAB**-tasten til at starte og stoppe afspilningen af lyden, og en tynd rød streg (lydcursoren) bevæger sig mod højre i billedet af lyden. Når man så trykker **Return**, bliver der indsat et 'boundary' på det sted. Derfor kan man ikke trykke **Return**, når den røde lydcursor står lige oven i et skel – der skal nemlig være et vist mellemrum mellem to boundaries.

Brug **shift+TAB** for at afspille lyden i det aktuelle interval fra den første boundary igen.

Det er vigtigt at indsætte boundaries ret nøjagtigt. Der skal stå det samme på en linje, som der bliver sagt i intervallet. En sætning kan godt deles op i flere boundaries.

Der er ingen faste regler for, hvor mange boundaries der skal være, eller hvor tit de skal forekomme, men jo oftere de forekommer, og jo kortere de er, jo bedre. Sæt som minimum et boundary, hver eneste gang der er en pause i talen, og hver gang der er en eller anden form for ophold i talen, en tøyen, en vejtrækning, eller noget andet, der klart kan defineres som ikke værende en del af et ord. Der kan næsten ikke sættes for mange boundaries – ideelt set burde der være et for hvert eneste ord, men det vil sjældent være praktisk muligt.

### 1.b.2 Flyt boundaries

Når man har sat et boundary, så kan det let flyttes til et andet sted, dvs. tidspunkt, i lydfilen. Placer musen over intervalgrænsen i det grønne talespor, og tryk på den midterste museknap (altså rullehjulet), og hold den nede, mens boundaryet flyttes til højre eller venstre. Musen behøver ikke pege præcist på intervalgrænsen, den skal bare være i nærheden.

### 1.b.3 Slet boundaries

For at slette et boundary helt – altså slå to tekstlinjer (og dermed også intervaller) sammen – skal man placere sig i starten af linjen og trykke **ctrl+backspace** (**backspace** kaldes også **slet tilbage** på visse tastaturer). Man kan også gå op i menuen "segmentation" hvor boundaries kaldes "breakpoint", og vælge "delete breakpoint".

Hvis man ikke placerer sig i starten af linjen, når man trykker på **ctrl+backspace**, sletter man også det tegn, der stod til venstre for tekstcursoren.

Sletning af et boundary kan ikke gøres om, så selv om det ikke er nogen drastisk ændring, skal man lige sikre sig, inden man gør det, fx ved at sørge for, at lyd cursoren er placeret præcist ved boundaryet, så man hurtigt kan indsætte et nyt.

### 1.b.4 Navigation

Man kan navigere frem og tilbage i lyden med scrollbaren oven over lydbilledet, og efterhånden som udskriften bliver mere og mere færdig, kan man frit skifte mellem at navigere i tekst eller lyd.

Når tekstcursoren bliver flyttet op og ned imellem de forskellige linjer (med **pil-op** og **-ned** tasterne), bliver hele den aktuelle linje vist med en lidt lysere baggrund, ligesom intervallet i det grønne talespor også bliver fremhævet. Lydcursoren bliver automatisk placeret i starten af intervallet – ved boundaryet.

Ved at holde **ctrl** inde sammen med **pil-op** eller **-ned**, kan man skifte til forrige eller næste talers tur, altså hoppe frem og tilbage i det blå talerspor.

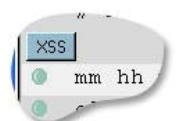
Man kan også markere et interval direkte i lydsporet ved at trække musen over det område man vil høre. Det bliver så markeret med en mørkrosa baggrund, og efterfølgende tryk på **TAB**-tasten vil kun afspille dette interval.

Der er endnu flere navigationsmuligheder til rådighed i "Edit", "Signal" og "Segmentation"-menuerne, og derudover virker de fleste piletast-kombinationer kendt fra andre programmer.

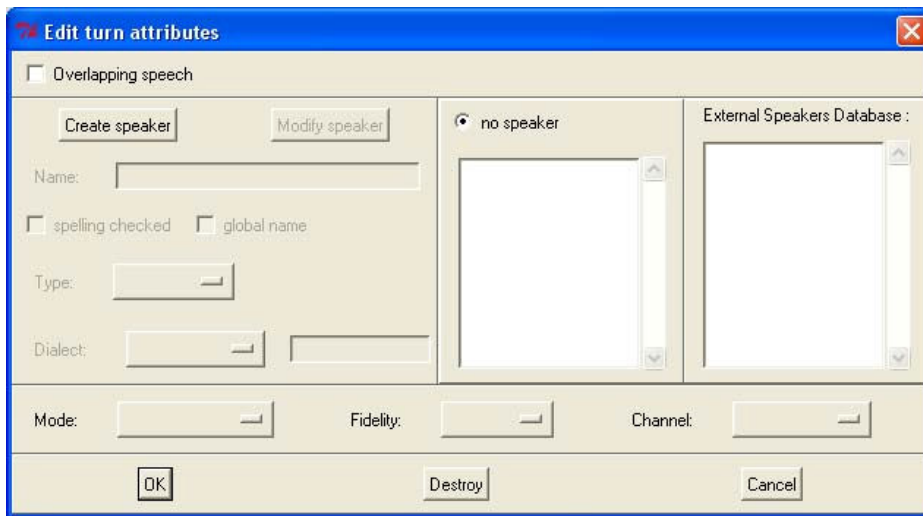
### 1.c Taleture

Ud over at være opdelt i tidsintervaller med passende boundaries, skal talen også fordeles på taleture. En taletur er ikke nødvendigvis en særlig fast defineret størrelse, det er blot et tidsinterval, hvor en enkelt taler taler uden større indblanding fra andre talere.

Taleture er vist i det blå talerspor nederst i Transcribers vindue og bliver vist med en tilsvarende blå boks i starten af hver talers tur. En taletur har altid navn efter den pågældende talers ID-kode, som består af tre store bogstaver. Som regel har interviewerens en ID-kode der begynder med X. For det meste passer ID-koden med talerens initialer, men det er ikke altid til at regne med, så sørg for at checke ID-koderne med informant-databasen på <http://dgcssintranet>.



Når der påbegyndes en ny transskription i Transcriber, bliver den første taletur altid tildelt (no speaker) – tryk på den blå boks, eller tast **ctrl+alt+T** for at ændre det. Så får du et vindue hvor du kan redigere "turn attributes", dvs. specificere talerne.



Tryk "Create speaker", og skriv talerens ID-kode i "Name:"-feltet. Du skal ikke udfylde andre felter. Det kan være en fordel at oprette begge eller alle talere på en gang, så du har både den interviewede og intervieweren. En taler kan altid ændres – hvis du for eksempel har kaldt intervieweren XSS, men senere opdager, at det var XMM, så åbn dette vindue som beskrevet, vælg XSS, klik "Modify speaker", og ændr navnet.

Når der under udskrivningen skiftes taletur, så lav et boundary og tryk **ctrl+T** for at genkalde ovenstående vindue. Når der kun er to talere, vil programmet som regel sørge for at vælge den modsatte, så du kan bare trykke **Return**, når du har trykket **ctrl+T**. Hvis der blev valgt den forkerte taler, så ret det ved enten at klikke på den blå boks eller ved at trykke **ctrl+alt+T**.

*Bemærk: Nogle gange vises ovenstående vindue uden "OK"-, "Destroy"- og "Cancel"-knappe – det er uvist hvorfor, men du skal så blot tage fat i bunden af vinduet og trække det lidt større. "Destroy"-knappen vises ikke altid – tryk på den blå boks for at være sikker på at få en "destroy"-knap.*

Hvis der er et større ophold i lydoptagelsen, hvor der ikke er nogen, der taler – for eksempel på grund af en lang pause, eller støj udefra – så lav en taletur og sæt den til (no speaker).

Hvis der optræder en ny ukendt taler i en samtale, så opret vedkommende som UKE – standardnotation for en ukendt taler – men forsøg efterfølgende at finde ud af, hvem vedkommende er, og ændr derefter navnet. Hvis der er flere forskellige ukendte talere, kan de hedde UK1, UK2, og så fremdeles.

### 1.d. Overlappende tale

Intervaller i lyden skal helst være opdelt, så der kun er en person, der taler i hvert interval. Men fra tid til anden overlapper talerne hinanden. Især kan små returytringer som "ja", "nå" og "mm" overlappes et enkelt ord helt eller delvist.

Som nævnt skal teksten på en linje svare nøje til det, der bliver sagt i linjens interval i lydoptagelsen. Og når der er flere, der taler samtidig, så skal der være tekst for hver taler på samme linje. Transcriber har en indbygget funktion til at markere, at en taletur tilhører to forskellige talere – men da den både er besværlig at bruge, og det sjældent er en hel tur, der er overlappet – så benytter vi en lidt anden notation på DGCSS.

I det følgende eksempel siger SKO: "nej nej han spiller også på den og", og XHH overlapper hans "på den og" med "ha ja det". Derfor skrives "på den og" og "ha ja det" på samme linje i SKO's taletur, men "ha ja det" sættes ind mellem < og > tegn for at angive, at det ikke er ham, der siger det.



```
XHH
● det vil sige sådan at øh din far han kan ikke holde fingrene fra den
SKO
● nej nej han spiller også
● på den og <XHH: ha ja det>
XHH
● ha kunne jeg
```

< og > betyder altid, at det er en anden taler end den, der har turen, der siger det pågældende.

Når der kun er to talere, er det altid indlysende, hvem der overlapper, men hvis der er flere talere, så skal man skrive den overlappende talers ID-kode inde i < >. Fx er det i ovenstående eksempel noteret, at det er XHH, der overlapper SKO's taletur. Det betyder, at man stort set altid skal notere i overlappet, hvem der overlapper, da der i nye filer oftest er stemmeprøver m.m. og dermed mere end to talere i filen.

Der må ikke stå anden tekst på linjen end det, der overlapper hinanden. Følgende kan altså **ikke** lade sig gøre:

```
XHH
● det vil sige sådan at øh din far han kan ikke holde fingrene fra den
SKO
● nej nej han spiller også på den og <XHH: mm>
XHH
● ha kunne jeg
```

Eksemplet viser, at XHH overlapper SKO's "nej nej han spiller også på den og" med et tilsyneladende meget langt "mm".

Den overlappende tekst i skarpe parenteser (< >-tegnene) skal altid stå sidst på linjen – det gør det nemmere for analysegrupperne at læse teksterne. Det vil sige, at der ikke må være tekst både før og efter et overlap, medmindre det er andre taleres < > tegn, med indsatte ID-koder. Fx:

```
SKO
● som jeg så
● #
● sidder der og kigger og
● studerer <XHH: mm> <XSS: javel> <XMN: interessant>
● lidt i
XHH
● mm
```

Det betyder, at man **ikke** kan skrive følgende, da der ikke kan stå tekst på begge sider:

```
XHH
● skolen er det noget du kan lide
JSH
● ja det kan jeg
● det meste af tiden
● eller <XHH: ja> #
● sådan noget
```

Man kan ikke skille et ord ad, der udtales uden selvfrydelse. Derfor kan følgende eksempel **ikke** lade sig gøre:

```
XHH
● har du nogle søskende
JSH
● ja jeg har to
● øh Peter
● han er syvogtyve
● og vild med cirkushe-
● -ste <XHH: aha>
```

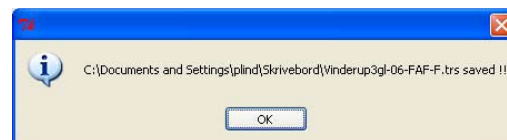
Men er der tale om en selvfærdighed (se afsnit 3.f.), kan følgende godt lade sig gøre:

```
XHH
● ses i tit
JSH
● nej vi ses efter-
● #
● -hånden <XHH: mm>
● sjældent faktisk
```

### 1.e. Når en fil er færdig – hvordan en fil gemmes

Da du indstillede konfigurationen, satte du et interval for, hvor tit transskriptionen skulle gemmes. Første gang dette tidsinterval nås, bliver du spurgt, hvor transskriptionsfilen skal gemmes – gem da altid på dit skrivebord.

Undervejs under udskrivningen gemmer Transcriber automatisk din transskription, og du kan også trykke **ctrl+S** for at gemme en ekstra gang. Så popper der en lille besked op om, at du har gemt – bare tryk **OK** eller **Return** til den.



Når du er færdig med dagens arbejde, skal du aflevere filen på dit skrivebord til versionsstyringsystemet. Det beskrives i en anden vejledning.

#### 1.e.1 Kør automatisk stavekontrol

Gå op i menupunktet "Edit", og vælg "Spell checking". Herefter skal du vælge enten at ignorere ("Ignore") eller erstatte ("Replace") de fejl, stavekontrollen finder. Hvis der popper en fejlmeddelelse op om, at den ikke kan finde ordbogen, så gå ind i menupunktet "File", vælg "Edit episode attributes...", og skriv "Danish" i feltet "Principal language".

Når du er nået slutningen af transskriptionen, luk da stavekontrollen, og gem filen.

## 2. Regler for udskrivning af tale

Formålet med at udskrive samtaler er som tidligere nævnt at gøre det muligt at analysere indholdet i samtalerne på forskellige niveauer. Når samtalerne er udskrevet, skal man kunne finde alle forekomster af bestemte ord, ordformer, konstruktioner mv. på tværs af materialet ved at søge efter dem, hvorefter man kan se, hvilken sammenhæng de bliver brugt i, og høre, hvordan de bliver udtalt. Det er derfor yderst vigtigt, at *alle* ord og ordformer, der siges, udskrives, og at de udskrives i overensstemmelse med DGCSS' konvention. For at sikre konsistens i udskrivningen er hovedprincippet i udskrivningskonventionen, at Retskrivningsordbogens (RO) regler følges. På denne måde kommer teksterne til at se mere normale ud, end de egentlig er; men det opvejes af, at lydfilen altid er tilgængelig via kobling med udskriften.

I afsnit 2.a. beskrives de overordnede konventioner for udskrivning af ord, hvor hovedreglen er, at RO følges. I afsnit 2.b. beskrives de enkelte afvigelser fra og tilføjelser til RO's regler, og i afsnit 2.c. beskrives de tegn udover bogstaver, der må bruges i forbindelse med ord.

## 2.a. Hovedregel: Der udskrives i almindelig ortografi

Tale gengives i ortografi, hvilket overordnet vil sige, at alle ord, der bliver sagt, som hovedregel udskrives, som de står opført i RO. Alle bogstaver i alfabetet må bruges, herunder bogstaver med umlaut, fx ü og ö, hvis disse indgår i den korrekte form af ordet. Ordbogen findes i elektronisk form på alle centrets computere.

### 2.a.1 Ordstammer skrives jf. RO

Ordstammer udskrives altid efter RO, uanset hvordan ordet udtales. Der skrives:

ikke (uanset om det udtales fx [ik], [ing], [ikke], [et] eller [æ'])  
til (uanset om det udtales fx [til] [t] eller [te])  
skulle, have, kunne (uanset om det udtales [sku], [ha] og [ku] eller på anden måde)  
kan du have det (og ikke fx [ka du ha'ed])  
kysse (uanset om præteritumformen udtales [kysse], [kysset] eller [kysset])  
meget (uanset om det udtales fx [møj], [møje], [my] eller [mad])  
jeg (uanset om det udtales fx [jɑ] eller [a])  
synes (uanset om præsensformen udtales fx [sys], [synes] eller [syns])  
huske (uanset om infinitiven udtales fx [hus], [husk] eller [håw])  
social (og ikke fx [sjosjal])  
håndklæde (og ikke fx [hångklæd])  
det værger jeg mig ved (hvis ordet udtales som "hverver" og ikke som "vægrer")

### 2.a.2 Bøjningsendelser og artikler skrives, som de bliver sagt, men formen skal stå i RO

Reglen for udskrivning af bøjningsendelser og artikler er, at man skriver dem, som de bliver sagt, hvis formen findes i RO, men uden hensyntagen til, om bøjningsendelsen eller artiklen ifølge RO eller konservativt standarddansk passer med det enkelte ords bøjningsmuligheder og kongruens i sætningen.

Det betyder, at ordene skal skrives i den form, der svarer til udtalen, også selv om der er tale om en anden bøjningsform end den, der ifølge RO – eller i (konservativt) rigsmål – ville forekomme i de pågældende syntaktiske omgivelser. Fx:

alle de andre i byen de var **gammel** (skriftsprogsnormen: "gamle")  
det er bare noget man skal **kan** (skriftsprogsnormen: "kunne")  
det er noget med at **skal** være derovre når der lukkes op og at **skal** hjem når den holder (skriftsprogsnormen: "skulle")  
jeg skal ud at købe **en** køleskab (skriftsprogsnormen: "et")

Princippet gælder også i forbindelse med bøjningsendelser, som ifølge RO slet ikke forekommer i forbindelse med det pågældende ord, men som er korrekte for andre ord fra samme ordklasse. Fx:

hatten **hængte** på knagen (RO: "hang")  
de **liggede** ved poolen (RO: "lå")  
for **havde** vi ikke **kommen** i forgårs så ved jeg snart ikke (RO: "var ... kommet")  
han **svorede** han ville gøre det (RO: "svor")  
hun havde et **lillat** tørklæde på (RO: "lilla")  
det er et **privilegie** (RO: "privilegium")  
jeg **faldede** ned (RO: "faldt")  
han **bedte** mig om at gå (RO: "bad")

Endelser, som ikke er korrekte for andre ord fra samme ordklasse, udskrives ikke. Der tilføjes i stedet en kommentar. Fx:

og så skete det alligevel {udtales: alligevellet} (om brugen af kommentarer se afsnit 3.c)

Den eneste undtagelse fra reglen om, at artikler og bøjningsendelser skrives efter RO, er den foranstillede bestemte artikel 'æ'; den noteres som "æ". Fx:

jeg gik en tur med **æ** hund

En vestjysk ytring, der udtales (i retning af) "ka du håw te du kyst mig en awt- en awten a mødt dig dernije i æ kjær", skal derfor skrives:

kan du huske til du kyssede mig en aft- en aften jeg mødte dig dernede i æ kær

Ved udskrivning af dialektale træk skal man følge ovenstående regler (og reglerne i denne manual i det hele taget), og der henvises i øvrigt til en separat vejledning i udskrivning af dialekt med tilhørende lydeksempler (L:\lydoptagelser\lydeksempler\vinderup eksempler).

### **2.a.3 Hvis man ikke kan høre forskel, vælges den korrekte form**

I tilfælde, hvor det ikke ud fra udtalen er muligt at afgøre hvilken form, der er tale om, skrives den form, som svarer til skriftsprøgnormen. Der skrives:

"jeg syntes om ham dengang" (men "jeg synes han er sød")

"jeg tror vi kører til Italien" (men "vi skal vist køre til Italien")

[å] skrives "og", når det er en konjunktion, og "at", når det er et infinitivmærke (se også afsnit 4.d.).

### **2.a.4 Sammentrækninger skrives ud**

Sammentrækninger af ord i talesproget skrives ud, selv om ordene ikke udtales separat. Der skrives:

der er kaffe på kanden (og ikke "der kaffe på kanden")

jeg er tørstig (og ikke "jeg tørstig")

Det gælder også sammentrækninger af udenlandske ord. Der skrives:

I will be loving you allways {engelsk sangtitel} (og ikke "I'll be loving you allways" - om brugen af kommentarer se afsnit 3.c.)

Gigi Di Agostino spiller i aften (og ikke "Gigi D'Agostino spiller i aften")

Forlængelser, der skyldes sammentrækning, markeres ikke. Der skrives:

det er min tur (og ikke "de:t er min tur")

## **2.b. Afvigelser og tilføjelser i forhold til retskrivningen**

Som beskrevet ovenfor gengives tale som hovedregel i almindelig ortografi. I forhold til denne hovedregel er der dog en række afvigelser eller tilføjelser til retskrivningsnormen, som det er nødvendigt at overholde for at lette det senere analysearbejde. Disse afvigelser og tilføjelser beskrives her.

### **2.b.1 Talord skrives ud**

Der må ikke skrives tal. Når en tale fortæller om noget, der skete d. 18/5 1993, og nævner datoen og årstallet i sin fortælling, skal det skrives, som det siges. Den nævnte dato kan bl.a. have følgende former:

den attende maj nitten hundrede og treoghalvfems

den attende maj nitten hundrede treoghalvfems

attende maj nitten treoghalvfems

attende i femte nitten treoghalvfems

De to første eksempler er medtaget, da det i RO er valgfrit, om man skriver og'et. Vi skriver, hvad der siges, dvs. siger taleren "hundrede og to" skrives og'et, ellers ikke.

I overensstemmelse med RO skrives talord under 100 i et ord, mens talord over 100 deles op. Dette gælder naturligvis også ordenstal, fx "for syv hundrede og syttende gang".

### **2.b.2 Store bogstaver bruges kun i begrænsede tilfælde**

I modsætning til i retskrivningen udskrives alle ord som hovedregel med små bogstaver. I udskrifterne må der kun bruges store bogstaver i begrænsede tilfælde, nemlig i:

En ny linje må kun begynde med stort bogstav, hvis det første, der siges, falder inden for reglerne nedenfor, dvs. er en af de tre nævnte typer.

## Enkelte pronominer

I, De, Dem, Deres.

## Proprier

Navne på bl.a. personer, steder, institutioner og virksomheder, film- og bogtitler, hvor stort begyndelsesbogstav er obligatorisk i retskrivningen. Der skrives:

Anders  
Vinderup  
Københavns Amt  
Vejdirektoratet  
Miljøministeriet  
Næstved Tidende  
Det Kongelige Teater  
Borte med blæsten (i titler skrives kun begyndelsesbogstavet i første ord med stort)

Uofficielle kaldenavne udskrives dog med småt, da de i praksis er svære at identificere. Der skrives:

forsvaret  
søerne  
bornholmerbåden

## Bogstavnavne

Bogstavnavne optræder især i forbindelse med navne, der staves, og i akronymer (dvs. forbogstavsforkortelser), Der skrives:

H C Andersen (uden punktummer, med mellemrum)  
mit navn staves C H R I S T O P H E R  
bolle\_Å (om brugen af underscore se afsnit 2.c.3)  
U\_land  
K\_salat  
D\_S\_B  
U\_S\_A

Reglen om, at bogstavnavne skrives med stort, gælder også i de akronymer, som ifølge RO skal skrives med små bogstaver. Der skrives:

L\_P  
T\_V  
I\_D  
C\_D\_afspiller  
W\_C'et (selv om W udtales [ve] skrives W) – (om brugen af apostrof se afsnit 2.c.5)

### 2.b.3 Forkortelser skrives, som de siges

Med hensyn til forkortelser skal der i udskriften ordret stå det, taleren siger. Der må derfor ikke bruges forkortelser, medmindre ordene udtales forkortet. Der skrives:

og så videre og så videre (og ikke "osv. osv.")  
for eksempel (og ikke "f.eks." eller "fx")  
forhenværende (og ikke "fhv.")  
vedrørende (og ikke "vedr.")

Afkortede former, der udtales som egentlige ord i samtalen, fx "hr.", "cand.mag." og "maks." skal skrives, som de siges. Selv om forkortelsespunktum er obligatorisk ifølge RO, skrives dette ikke. Der skrives:

den gode hr Jakobsen  
en cand mag stud (men "cand\_mag\_studerende" – om brugen af underscore se afsnit 2.c.3)  
de fyrer den maks af

Bemærk at forkortelser med bogstavudtale og obligatorisk forkortelsespunktum ligeledes skrives uden forkortelsespunktum. Der skrives:

P\_T  
S\_U

Reglen om, at forkortelser skrives, som de udtales, gælder også akronymer, altså forbogstavsforkortelser. Når akronymer udtales som ord, i stedet for at bogstaverne nævnes ved navn, fx når der tales om "Kua", "Nato" og "Ruc", behandles de som almindelige ord. Der sættes ikke apostrof mellem forkortelser, der udtales som ord, og bøjnings- og afledningsendelser (se afsnit 2.c.5). Der skrives:

Kua (og "på Kuas gange")  
Nato (og "Natos ledelse")  
Ruc (og "hun er Rucer")

Med hensyn til akronymer, der udtales bogstav for bogstav, ligesom i "DSB", "TDC" og "USA", se 2.c.3 om brugen af underscore.

#### 2.b.4 Ord, der ikke findes i RO

Hvis der i optagelsen siges ord, som ikke findes i RO, skrives de i henhold til Den Danske Ordbog (DDO), som findes på centret samt på Afdeling for Dialektforsknings Bibliotek.

Ord fra fremmedsprog. Det vil typisk sige engelske låneord, men også lån fra andre sprog. De skal skrives i henhold til det pågældende sprogs retskrivning, medmindre de er opført i RO. På centret står en udgave af Chamber's Thesaurus, og det er den ordbog, man skal slå op i, hvis man er i tvivl om stavemåden af et engelsk ord. I forbindelse med ord på andre sprog anvendes tilsvarende ordbøger, som kan findes på Afdeling for Dialektforsknings Bibliotek. Man kan eventuelt notere i en kommentar, hvor ordet kommer fra. Men som det fremgår af 3.c., skal der kun skrives en kommentar, hvis den forekommer relevant for samtalen.

Dialekt. Hvis man støder på dialektale ord, som ikke er anført i RO eller DDO, skal man se i wikien under artiklen "Dialektord". Hvis ordet ikke er anført der, skal man henvende sig til udskrivnings-koordinatoren. Bemærk, at dette gælder dialektord, som repræsenterer et andet opslagsord end et standarddansk ord, fx "tøs" i betydningen "synes", men ikke ord, som blot er udtalevarianter af ord, der findes i RO, jf. afsnit 2.a. Dialektord skrives med en kommentar om ordets oprindelse. Der skrives:

jeg så en **jylkat** {bornholmsk for pindsvin} (om brugen af kommentarer se afsnit 3.c.)

Nonsensord. Ord, der ikke er anført i RO eller DDO og som ikke er låneord eller dialekt, udskrives som ord ud fra RO's principper for retstavning. Der skrives:

hun siger stadig **basgetti**  
han har lige købt en splinterny **bimmer**

Proprier. Dvs. bl.a. navne på steder, personer, firmaer og institutioner. De skal skrives korrekt, så de efterfølgende kan identificeres. Navne, som udskriveren ikke kender stavemåden på, og navne, som har flere mulige stavemåder, skal derfor tjekkes i informantbasen (<http://dgcssintranet/informantbase>) eller i andre opslagsværker, så det så muligt er den korrekte form, der skrives. Det gælder fx stednavne som Appenæs (som skrives med 'p' og ikke 'b'), personnavne som Kristoffersen (som kan skrives med både 'K' og 'Ch') og Hüttemeier (som, hvis der er tale om tv-kogekonen, staves med 'ü' og 'i').

Til identifikation af stednavne kan bruges [www.google.com](http://www.google.com), [www.krak.dk](http://www.krak.dk) eller funktionen "Find et sted" på Kort- og Matrikelstyrelsens hjemmeside [www.kms.dk](http://www.kms.dk). Brug også wikiartiklerne "Stednavne i ..." (<http://dgcssintranet/wiki>). Til personnavne kan der desuden bruges De Gule Sider ([www.dgs.dk](http://www.dgs.dk)), hvis der er tale om mere særprægede navne.

#### 2.b.5 Særlige konventioner for ikke-ord, der skal udskrives som ord.

I dette afsnit gennemgås konventionerne for nogle af de mest udbredte fænomener i talesproget, der ikke normalt markeres i ortografien.

Der er en række meningsbærende lyde, der ikke anerkendes som ord i RO, men som skal gengives på en standardiseret måde. De standardiserede gengivelser ses nedenfor, og de skrives i udskriften ligesom almindelige ord. Dette er de eneste ikke-ord, der må – og skal – skrives i udskriften.

ha	Alle former for latter
hh	Ind- eller udånding, som har en tydelig interaktionel funktion i samtalen, fx som en bekræftelse på et spørgsmål eller et forsøg på at komme ind i samtalen. Hørlig vejrtrækning, som ikke har en klar interaktionel funktion i samtalen, skal <i>ikke</i> udskrives. 'Ja' sagt på indånding udskrives derfor 'ja', ikke 'hh ja'.
mm	Returytring. Skrives altid med to – og kun to – m'er uanset varighed. Længdeprikker bruges ikke. Der skal ikke tages stilling til, om lyden er bekræftende, afvisende eller neutral. Dvs. der skrives aldrig: "mm." og "m:m". To på hinanden følgende mm'er i betydningen 'nej' udskrives 'mm', ikke 'mm mm'.
sh	Tyssen
uh	Udbrud, uanset om det kan beskrives som positivt eller negativt, og uanset vokalens lyd. Men udbrud, der findes i DDO, skal udskrives korrekt. Det gælder for eksempel 'årh', 'ej', 'ah' og 'puh', der er de mest almindelige. Der findes desuden en liste over interjektioner (udråbsord) i Wikien.
øh	Alle former for fyldte pauser, dvs. pauser med fonering, uanset kvalitet og varighed. Længdeprikker bruges ikke. 'Øhm' udskrives 'øh'. Det samme gælder 'm:' 'n:' eller 's:' og lignende, der ikke er en forlængelse af sidste lyd i det foregående ord eller første lyd i det kommende ord.
nja, njo	"Tøvende bekræftelse" udtalt [nja] og [njo] (evt. med vokalforlængelse), der ikke kan betragtes som et afbrudt 'nej' efterfulgt af et 'ja' eller 'jo'. Varianterne [mja] og [mjo] udskrives henholdsvis 'nja' og 'njo'.

Lydfiler med eksempler på de enkelte lyde findes i L:\Lydoptagelse\Lydeksempler.

Hosten, rømmen og andre lyde, som ikke findes i listen ovenfor, men som muligvis er betydningsbærende, skal derfor ikke udskrives. Til gengæld skal de noteres som "lydlig begivenhed", jf. afsnit 3.b.

## 2.c. Tegnsætning

Af hensyn til maskinlæsbarheden må der i udskrifterne ikke stå andre tegn end de bogstaver, der angiver, hvad der siges, og de få tegn, der beskrives nedenfor.

### 2.c.1 Der sættes ikke interpunktionstegn

Tegnene punktum, komma, kolon, semikolon, spørgsmålstegn og udråbstegn bruges ikke som interpunktionstegn. Kolon, udråbstegn og andre tegn må dog bruges, som beskrevet nedenfor.

### 2.c.2 Accenter bruges kun i ikke-danske ord

Accenttegn skal kun anvendes i forbindelse med ikke-danske ord, som ikke findes i RO. Ikke-obligatoriske accenttegn, som kan sættes i ord som fx "turne", "ide" og "cafe", noteres ikke.

### 2.c.3 Underscore \_

Underscore bruges til at sammenkoble alle orddele, der tilsammen har status af ét ord. Ordene kendetegnes bl.a. ved, at de kun har ét hovedtryk:

**I forkortelser med bogstavudtale.** Der skrives:

D\_S\_B

C\_D

P\_T (uden forkortelsespunktum – jf. afsnit 2.b.3)

S\_tog (om brugen af underscore i sammensætninger med bogstavudtalte forkortelser se nedenfor)

W\_C'et (selv om W udtales [ve] skrives W) - (om brugen af apostrof se afsnit 2.c.5)

Y\_M\_C\_A {engelsk forkortelse} (om brugen af kommentarer se afsnit 3.c.)

**I stedet for fælles orddele.** Der skrives:

ottende\_og\_niende\_klasserne (men "ottende og niende klasse")  
social\_og\_sundhedsassistent

**I sammensætninger med bogstavudtalte forkortelser.** Der skrives:

E\_mail\_adresse (om brugen af underscore i sammensætninger hvor det ene led indeholder en underscore se nedenfor)  
S\_tog (og "S\_togs\_station")

**I sammensætninger og afledninger med tungere affikser, når det ene led består af udtalte bogstavnavne.** Der skrives:

bolle\_Å  
A\_klassen

**I gruppesammensætninger.** Der skrives:

væg\_til\_væg\_tæppe  
science\_fiction\_roman (men "science fiction")  
to\_hundrede\_kilometer\_grænsen (men "to hundrede kilometer")

**I gruppeafledninger med tungere suffikser, når første led er skrevet i mere end ét ord.** Der skrives:

Peter\_Pan\_agtig (men "Peter Paner", da flertalsendelsen er let, jf. RO § 63.8)  
skrid\_og\_lad\_mig\_være\_i\_fred\_agtig (men "skrid og lad mig være i fred", "sukkeragtig" og "han er sådan # agtig")

**I stedet for bindestreg i oprindelig fremmede sammensætninger.** Der skrives:

drive\_in  
walkie\_talkie

**I betydningen "fra ... til ..." eller "mellem ... og ...".** Der skrives:

tre\_halv\_fire (men "tre\_halv\_fire\_tiden")  
en\_fem\_seks\_hundrede\_stykker  
en\_returbillet\_København\_Århus

**I sammensætninger med sidestillede led.** Der skrives:

køkken\_alrum  
Lolland\_Falster  
græsk\_katolsk  
olie\_eddike

**I stedet for skråstreg i sammensætninger med alternative led.** Der skrives:

og\_eller  
enten\_eller

**I stedet for obligatorisk bindestreg i proprier.** Der skrives:

Svend\_Erik\_Hermansen (for Svend-Erik Hermansen)  
Winnie\_Larsen\_Jensen (for Winnie Larsen-Jensen)



**I sammensætninger, når det ene led indeholder en underscore.** Der skrives:

C\_twenty\_one\_koncert {engelsk navn på band} (om brugen af kommentarer se afsnit 3.c)  
U\_S\_A\_turen  
drive\_in\_biograf  
plastik\_walkie\_talkie  
M\_P\_tre\_afspiller  
olie\_eddike\_dressing

**Underscore bruges ikke til at sammenkoble orddele i vejnavne og proprier trods tryktab.** Der skrives:

H C Andersens Boulevard  
Store Kongensgade  
Søværnets Operative Kommando  
Anders Fogh Rasmussen

#### **2.c.4 Bindestrøg -**

**Bruges kun til at markere selvafbrydelser.** Se afsnit 3.f.

**Bindestrøg bruges ikke** i sammensætninger med usædvanlige eller svært læselige ord- og bogstavsammenstød, der ifølge RO kan skrives med bindestreg. Der skrives:

semesterstartsfest  
stresssymptomer  
firserpermanentdulle

Det samme gælder sammensætninger med danske og engelske ord. Der skrives:

trancemusik {engelsk orddele} (om brugen af kommentarer se afsnit 3.c.)  
hotmailadresse {engelsk orddele}  
pokerchips {engelsk orddele}

Bindestrøg bruges heller ikke i sportsresultater. Der skrives:

Brøndby F\_C\_K to nul

#### **2.c.5 Apostrof skrives kun, hvor den er obligatorisk**

Apostrof skrives kun, hvor den er obligatorisk i retskrivningen, dvs. hvor den sammenkæder bøjningsendelser med citatord, bogstavnævne eller akronymer, samt hvor den markerer genitiv af ord, der ender på s, z eller x. Der skrives:

der er en del men'er ved sagen  
hendes A'er er usædvanligt flade  
C\_D'en (om brugen af underscore se afsnit 2.c.3)  
U\_S\_A's præsident  
T\_V'et (bestemt form: svarer til "tv'et", i modsætning til T\_V\_et, der svarer til kanal nummer 1)  
dette hus' indretning  
Randers' erhvervsliv  
Marx' skrifter  
den moderne jazz' udvikling

### **3. Kommentarer og forhold i taleproduktionen**

Oftest sker der ting i løbet af en samtale, som det er væsentligt at få med i udskriften, for at læseren kan forstå, hvad der tales om. Derudover er der en række forhold i samtaledeltagernes taleproduktion, der skal noteres med særlige symboler i udskriften. Hvordan disse ting gøres, beskrives nedenfor.

#### **3.a. At indsætte en event**

En event bruges til enten at beskrive talerens lyde, der ikke har en konventionel betegnelse, eller til at indsætte kommentarer om talernes udtale m.m. For at gøre dette, skal man trykke **ctrl+D**.

Dette åbner et vindue, hvor man kan vælge enten "noise" (for eksempler se afsnit 3.a.) eller "comment" (for eksempler se afsnit 3.b). De resterende funktioner: "pronounce", "lexical", "language" og "named entities" bruges ikke.

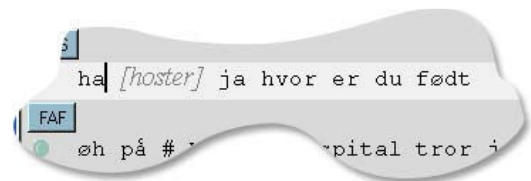
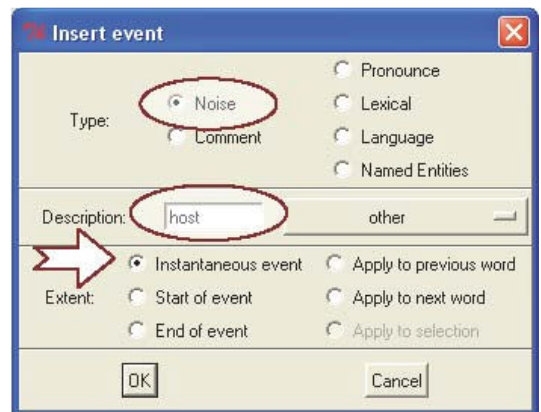
Lyden eller kommentaren skrives i "Description"-feltet, og de skal gerne være formuleret i et så kort og letforståeligt sprog som muligt.

Sørg for, at der er valgt "Instantaneous event" i "Extent"-området nederst i vinduet.

Noise-events bliver vist inde i teksten, i firkantede parenteser, i

kursiv, og i en grå farve, men de er ikke en del af teksten, og du kan ikke redigere dem ligesom den omliggende tekst. Men du kan klikke på eventen og få samme vindue frem som vist ovenfor, hvor du kan redigere alle detaljer.

Det samme gælder comment-events. Forskellen er, at de vises i krøllede parenteser.



### 3.b. Lydlige begivenheder

Hvis en taler siger noget, der ikke kan skrives som et ortografisk ord, eller som mm, hh, øh, ha eller et af de andre konventionelle udtryk beskrevet i 2.b.5, så skal det skrives som en "event" og som "noise".

En noise-event er fx et host, et nys eller en rømmen, men kan også være mere avancerede udtryk. Det vil sige, at man kun bruger "noise" til at beskrive talernes lyde og ikke andre lyde i omgivelserne, såsom kopper, der sættes på bordet, eller telefoner, der ringer. De beskrives som "comment".

For at gøre udskriverne så ensartede som muligt, har de mest brugte udtryk en fast betegnelse:

Hoster	Dækker al hoste
Rømmer sig	Dækker al rømmen, dvs. man kan ikke skrive "hrm" eller "rømmen"
Nyser	Dækker al nysen
Smasker	Dækker alle læbelyde, mundlyde, smasken og 'tsk'

### 3.c. Kommentarer

Forskellige former for kommentarer, som er relevante i forbindelse med samtalen, fx særligt bemærkelsesværdig udtalevariation, brug af fremmedsprog, problemer med lyd kvaliteten af optagelsen og den slags, som ikke hører hjemme i gengivelsen af, hvad taleren siger, kan noteres.

Kommentarerne er nyttige, når samtalerne skal analyseres.

Undlad at bruge tid på at kommentere udtalevariationer, fx at "tutor" udtales "tjutor", eller at "fanden" udtales "fan", men angiv, når taleren fx skifter til vestjysk i en sekvens, hvor der ellers tales københavnsk, eller at et ord udtales på fx engelsk. Der skrives:

jeg så den her film forleden Le fabuleux destin de Amélie Poulain **{fransk filmtitel}**

har du nogensinde set en jylkat **{bornholmsk for pindsvin}**

så spillede vi ude på fodboldbanerne **{udtales med 's': fodboldsbanerne}**

Kommentarer bruges også til at notere skift i udsigelsesform. Skift i udsigelsesform kan beskrives som en ændring i talerens stemme, enten ved at vedkommende imiterer en anden talers stemme eller parodierer en bestemt måde at tale på. Angiv som minimum det første og det sidste ord i den passage, hvor taleren har skiftet udsigelsesform, eller sæt en kommentar både efter første og sidste ord. Skift i udsigelsesform er væsentligt i sig selv, men også vigtigt at angive af hensyn til den fonetiske analyse.

### 3.d. Baggrundsstøj

Hvis en event ikke kan tilskrives en bestemt taler, så opret så vidt muligt en ny taletur til den, og vælg (no speaker). Hvis det ikke kan lade sig gøre – hvis lyden for eksempel overlapper med noget af talen, så skriv *global:* i starten af eventens description-felt, og beskriv ellers lyden som normalt.

Hvis lyden overlapper flere taleture, skrives der '*global:*', både når lyden starter, og når lyden slutter, men ikke i de mellemliggende taleture.

Ved noise-events (se 3.b.) skrives:

```
XAB
● hvis du lige vil kigge på de her sætninger [global start: sang]
● #
● øh du skal læse dem op som du siger dem [global slut: sang]
```

Ved comment-events (se 3.c.) skrives:

```
XAB
● jeg synes maden er dejlig {global start: fløjtelyd}
CDU
● okay
XAB
● altså når den steges på den helt rigtige måde
CDU
● ja men det kan du da have ret i {global slut: fløjtelyd}
```

### 3.e. Pause #

Pause noteres med "havelågen" – med mellemrum både før og efter symbolet. Pauser skal om muligt have deres eget boundary. Hvis det ikke er muligt, skal en pause altid placeres sidst i et boundary. Helt korte pauser behøver ikke nogen boundary-markering, men skrives blot inde mellem talerens øvrige ord. Længere pauser, der ikke tilhører en specifik taler – fx hvis den ene taler kortvarigt forlader interviewet – skal angives i en separat taletur, hvor taleren skal være sat til (no speaker). Pause efter endt taletur tilhører enten den sidste taler eller (no speaker), aldrig den næste taler.

Hvis fx en telefon ringer i en længere pause, og det er betydningsbærende for samtalen, skal lyden af telefonens ringen have sit eget boundary, hvor telefonens ringen beskrives i en comment-event:

```
XAB
● ja det kan jeg godt se
[no speaker]
● #
● {telefonen ringer}
● #
XAB
● skal du ikke tage den
```

Det vil sige, at der ikke kan skrives en pause og telefonens ringen i det samme boundary. Hvis lyden ikke er betydningsbærende, noterer man det som en lang pause uden kommentarer.

### 3.f. Afbrudt og genoptaget ord markeres med bindestreg -

Bindestreg bruges, når en taler afbryder sig selv, fx når vedkommende stammer eller skifter ordvalg, eller når en taler bliver afbrudt af en anden. Bindestregen må kun bruges, når afbrydelsen forekommer *inden for et ord*. Der må ikke være mellemrum mellem det sidste bogstav i det uafsluttede ord og bindestregen, men der skal være mellemrum efter bindestregen. Der skrives:

Karen Margrethe hed Jakobsen og **vil-** var fra øh Svebølle foruden han har job jo som automekaniker ja som **autome-**

I visse tilfælde vil et selvfølgelig afbrudt ord blive færdiggjort efter en kort pause, uden at første del gentages. I sådanne tilfælde indføres en bindestreg i begyndelsen af den genoptagne orddel. Der skrives:

dengang de stadig var **arbejds- # -dygtige** hvis man kan sige det sådan

Her er det væsentligt at skelne mellem fuldendelse og egentligt skift af ord. Vær især opmærksom på dette, hvis det, der siges efter selvfølgelig afbrydelsen, udgør et selvstændigt ord. Vurderes denne anden del til at repræsentere et skift i ordvalg snarere end en fuldendelse af et påbegyndt, men selvfølgelig afbrudt ord, skal der ikke skrives bindestreg foran ordet.

Alle selvfølgelig afbrydelser udskrives som de ord, udskriveren tolker, personen ville have sagt. Der skrives:

hun hedder vist Sti- (tolkning: afbrudt proprium)  
jeg ved ikke hv- (tolkning: afbrudt spørgeord)

### 3.g. Tvivlsom passage {...}

Der sættes krøllede parenteser omkring en passage, hvor der er tvivl om, hvad der bliver sagt. Tvivlsomme passager skal altid afsluttes inden for den taletur, hvori de er påbegyndt. Der skrives:

øh de første tre måneder af mit liv der boede jeg længere nede ad **{Katholmvej}**

Husk, at det er almindelige krøllede parenteser, der skal sættes, og ikke dem, der vises i forbindelse med kommentarer. Hvis du kun er usikker på nogle af ordene i en passage, så nøjes med at sætte parentes om de ord, du er usikker på.

### 3.h. Uforståelig passage xxx

Tre, og kun tre, x'er bruges til at markere steder i udskriften, hvor en taler siger noget, man ikke kan forstå.

Man skal ikke forsøge at angive, hvor mange ord eller stavelser, man tror passagen indeholder. Hvis det uforståelige tydeligvis indeholder flere meningsbærende orddele, må man gerne notere den meningsbærende del, man kan høre, i kombination med xxx. Der skrives:

så sagde jeg hun skulle gå til **xxx\_vejen**

I så fald skal der bruges understrebe til at sammenkoble orddelene. Alternativt skal det noteres som en tvivlsom passage jf. 3.g ovenfor.

### 3.i. Forlængelse :

Kolon bruges til at markere forlængede vokaler eller konsonanter. Tegnet sættes umiddelbart efter det bogstav, der svarer til den lyd, der forlænges. Der bruges kun et kolon, uanset længden. Der skrives:

ja:  
men:  
s:ynes  
kunne:  
vaske:r (og ikke "vasker:")

Der sættes aldrig kolon efter bogstaver, der ikke udtales. Former som "det:", "fald:er" og "lea:se" er umulige. Der sættes heller ikke kolon efter "mm" og "øh", uanset varighed.

### 3.j. Emfase !

Udråbstegn bruges til at markere emphatisk tryk, altså ord med særlig kraftig fremhævelse. Udråbstegnet placeres efter ordet, aldrig midt i ordet. Der skrives:

hun er helt! vildt dygtig (og ikke "hun er he!lt vildt dygtig")  
og alle! stod bare og gloede på mig (og ikke "og alle stod bare og gloede på mig")

Der skal ikke være mellemrum mellem ordet og udråbstegnet, men der skal være mellemrum efter udråbstegnet.

Brug emfasemarkering med omtanke og kun til at fremhæve helt klare tilfælde af prosodisk emfase. Brug ikke udråbstegn til at markere udråb som i retskrivningen.

### 4. Ting, der kræver særlig opmærksomhed

Ud over almindelig retstavning er der en række ting, man skal være særligt opmærksom på, når man udskriver talesprog.

#### 4.a. Er alle ord og morfemer repræsenteret i udskrifterne?

Kontroller omhyggeligt, at alle de ord, du som lytter mener er til stede i talestrømmen, faktisk også skrives. Vær her særligt opmærksom på sammentrækninger (assimilationer); specielt sammentrækningen af hhv. 'det' og 'er' og 'der' og 'er', jf. 2.a.4. I sådanne tilfælde skal det udeladte 'er' indsættes.

der er ikke mere kaffe (og ikke "der ikke mere kaffe")  
det er i orden (og ikke "det i orden")

Det er dog også vigtigt at sikre sig, at udskriften ikke er blevet "hyperkorrekt". Ifølge DGCSS' udskrivningsregler går den faktiske udtale forud for RO's regler for bøjning, jf. afsnit 2.a.2.

#### 4.b. Er der for mange ord i udskriften?

Kun ved udskrivning af sammentrækninger må udskriftet indeholde flere ord end talen. Der udskrives, hvad der siges – ikke det, der burde siges eller forventes sagt. Således udskrives "at" kun, hvis det siges.

#### 4.c. Afspejles talestrømmens ordrækkefølge direkte i udskriften?

Kontroller omhyggeligt, at du ikke er kommet til at overføre den i skriftsproget korrekte ordstilling til udskriften, fx i forbindelse med ledsætninger, der er blevet sagt med hovedsætningsordstilling:

man skal også tænke på at der er ikke varme med i det  
selvfølgelig er det ikke sådan jeg har ikke tid til noget samvær med andre

Dette er særlig vigtigt at være opmærksom på, da hovedsætningsordstilling er en variabel i centrets undersøgelser af syntaks.

#### 4.d. Skal der stå 'at' eller 'og'?

Ordene 'og' og 'at' udtales ofte ens, nemlig som vokalen i "kom!" (≈ å). I tilfælde hvor det ikke er muligt at høre forskel på to former, er det vigtigt, at man er konservativ med hensyn til, hvordan det udskrives.

Hovedreglen er her, at når lyden "å" står umiddelbart foran infinitiv, skrives "at", ellers skrives "og". Man skal altså skrive "at" i sætninger som disse:

vi forsøgte at nå det  
det er næsten ikke til at holde ud  
husk at ringe hjem

Hovedreglen gælder også i de tilfælde, hvor "å" i forbindelse med retningsadverbialer betyder det samme som "for at":

jeg går ud at lege  
vi vil tage ud at spise efter forestillingen

I alle andre tilfælde, dvs. ved sideordnede infinitiver og foran ikke-infinitiver, noteres "å" som "og":

vi skal spartle og male  
hun kommer og besøger os  
jeg vil og skal nå det

Bemærk: Reglen om, at "å" skrives som "at" foran en infinitiv gælder i forhold til skriftsprogsnormen. Da gengivelsen af den faktiske udtale går forud for skriftsprogsnormen i DGCSS's udskrifter, skal et verbum, der ifølge skriftsprogsnormen skulle være i infinitiv-form, i visse tilfælde skrives som en præsensform. Fx "det er et spørgsmål om at kan" (i forhold til skriftsprogsnormen: "det er et spørgsmål om at kunne"). I sådanne tilfælde skal "å" skrives "at", selv om det ikke står foran en ortografisk infinitivform.

Ligeledes skal sætningen "det er noget med å skal være derovre når der lukkes op å skal hjem når den holder" (som ifølge skriftsprogsnormen skrives: "det er noget med at skulle være derovre når der lukkes og (at) skulle hjem når den holder"), udskrives enten:

det er noget med at skal være derovre når der lukkes op og skal hjem når den holder

eller

det er noget med at skal være derovre når der lukkes op og at skal hjem når den holder

Dette er afhængigt af, om der i sidste del af sætningen er en eller to "å"-lyde før "skal".