



TRANSKRIPSJONSVEILEDNING FOR NOTA-OSLO

Av Kristin Hagen *

Versjon 30.10. 2008

* Denne transkripsjonsveiledningen er stort sett ført i pennen av Kristin Hagen, men er blitt til i et samarbeid mellom alle deltakere på NoTa-prosjektet ved Universitetet i Oslo: vit.ass. Anne Marit Bødal, professor Janne Bondi Johannesen, vit.ass. Fredrik Jørgensen, språkingeniørene Lars Nygaard og Joel Priestley, professor Hanne Gram Simonsen, samt transkribørene Hilde C. F. Haug, Ingunn Indrebø Ims, Signe Laake, Inger Margrethe Seim og Åshild Søfteland. Transkripsjonsveiledningen bygger også på: Du Bois et. al. 1993, Kristoffersen og Molde 2000 og xxx.

0. Om transkripsjonsveiledningen.....	2
1. Bakgrunn	3
1.1 Formålet med et talemålskorpus	3
1.2. Ortografisk transkripsjon.....	3
1.3 Diskusjon: hvordan bør en ortografisk transkripsjon være?.....	4
2. Praktisk veiledning for transkripsjon.....	5
2.1 Gode råd for transkripsjon	5
2.2 Filorganisering	5
2.3 Transkriberingsprogrammet <i>Transcriber</i>	6
2.3.1 Begynne på en ny transkripsjon.....	6
2.3.2 Fortsette med en påbegynt transkripsjonsfil	7
2.3.3 Lag <i>report</i>	7
2.3.4 Definer talere	7
2.3.5 Tidskoding og segmenter	7
2.3.6 Avspilling av lydfile	8
2.3.7 Stavekontroll.....	8
3. Transkripsjonsregler	9
3.1 Om disse reglene	9
3.2 Når reglene ikke strekker til.....	9
3.3 Utvalg og sensitive personopplysninger.....	9
3.4 Markering av ekstrainformasjon	10
3.5 Språkform.....	11
3.5.1 Hovedregler	12
3.5.1.2 Variasjon.....	13
3.5.1.3 Spesialbehandling av pronomen	14
3.5.1.4 Sammentrekking.....	14
3.5.2 To unntak fra hovedregelen	14
3.5.2.1 Pronomenene <i>a</i> og <i>n</i>	14
3.5.2.2 Kjønn beholdes på substantiv	15
3.5.3 Ord som ikke står i Bokmålsordboka.....	15

3.5.3.1 Forkortelser	17
3.5.3.2 Sammensetninger	17
3.5.3.3 Tall	17
3.5.3.4 Navn	18
3.5.3.5 Dialektord.....	19
3.5.3.6 Ord fra andre språk.....	20
3.5.3.7 Nye ord, banning ...	21
3.5.3.8 Ikke-språklige lyder og trykk.....	21
3.5.4 Sitater og ord som er stavet eller lest.....	22
3.5.4.1 Sitater.....	22
3.5.4.2 Ord som er stavet eller lest.....	22
3.5.5 Avbrudd, pauser og uklare passasjer.....	22
3.5.5.1 Avbrutte ord.....	22
3.5.5.2 Uavsluttede ytringer	23
3.5.5.3 Pauser	23
3.5.5.4 Passasjer som er vanskelige å forstå	24
3.5.6 Tegnsetting og stor bokstav	25
3.6 Tur, segment og meningsenhet.....	25
3.6.1 Tur.....	25
3.6.1.1 Replikk og ikke-replikk.....	26
3.6.2.2 Tur og turskifte	27
3.6.2 Meningsenhet og segment.....	27
3.6.3 Innskudd av ikke-replikker, overlapping og delte meningsenheter.....	28
3.6.3.1 Innskudd av ikke-replikker	28
3.6.3.2 Ikke-replikk eller del av replikk?	29
3.6.3.3 Overlapping	30
3.6.3.4 Delte meningsenheter	30
3.6.3.5 Skjulte turer	31
4. Korrektur.....	32
5. Oversikt over hurtigkommandoer, eventer og ikke-språklige lyder.....	33
5.1 Hurtigkommandoer	33
5.2 Skjema over eventer.....	34
5.3 Liste over interjeksjoner	36
5.4 Huskeliste over spesielle ord	37
5.5 Kløyvd infinitiv.....	38
6. Referanser.....	39

0. Om transkripsjonsveiledningen

I avsnitt 1 finner du en generell introduksjon om transkripsjon. Avsnitt 2 er mer detaljert og praktisk, og gir gode råd om transkripsjon samt en kort beskrivelse av hvordan du bruker transkripsjonsprogrammet *Transcriber*. Avsnitt 3 tar for seg de konkrete transkripsjonsreglene. Dette avsnittet bør du ha lett tilgjengelig mens du transkriberer. Avsnitt 4 handler om korrektur. I avsnitt 5 finner du blant annet en oversikt over hurtigkommandoer og eventer (hendelser), og avsnitt 6 inneholder referanser. I avsnitt 7 finner du et lengre transkripsjonseksempel.

For enkelthets skyld har vi i denne veiledningen valgt å bruke de engelske termene fra *Transcriber*. Forhåpentligvis vil dette gjøre det lettere å følge de praktiske instruksjonene.

1. Bakgrunn

1.1 Formålet med et talemålskorpus

For at man skal kunne forske på norsk språk, er det viktig å ha et godt datagrunnlag. Det er upraktisk og tidkrevende å lage spørreskjemaer og andre typer undersøkelser hver gang man er interessert i å finne ut noe bestemt om et særtrekk ved språket. Derfor er det en utpreget fordel å ha store datamengder lagret elektronisk, slik at de lett kan søkes i - såkalte tekstkorpus. For slike korpus har vi i dag gode søkesystemer og gode systemer for grammatisk merking (tagging). Vi har også skriftspråkskorpus av en viss størrelse, men for talemål har vi knapt noe. Skal man studere lydlig forhold, er selvsagt et talemålskorpus det eneste som nytter, men det er grunn til å tro at talespråket ofte skiller seg fra skriftspråket også på andre plan. Vi ønsker derfor å lage et talespråkskorpus for ikke bare lydlig (fonetisk-fonologisk) analyse, men også for morfologisk, syntaktisk og leksikalsk analyse. Dessuten vil vi at det skal være mulig å studere språklig samhandling (dialog) gjennom korpuset. I tillegg til lyd vil vi ta opp video av informantene slik at det blir mulig å se på visuelle aspekter av samtalen.

1.2. Ortografisk transkripsjon

For å kunne søke i et talespråkskorpus må materialet transkriberes. Det er mange måter å transkribere et talespråk på, men ingen av dem er perfekte. En uttalenær, fonetisk transkripsjon kan være nyttig for noen fonologiske studier, men det vil alltid være en mulighet for at akkurat de fenomenene som en bestemt fonolog er interessert i, likevel ikke er markert. Det vil også gjøre korpuset vanskelig å søke i og det er vanskelig å gjennomføre konsekvent, med forskjellige transkribører med ulik faglig og teoretisk bakgrunn. En rent ortografisk transkripsjon vil løse mange av disse problemene. Den vil være lett å utføre, den vil gjøre det lett å søke i korpuset, den vil være lett å tagge automatisk, og lett å lese. Vi velger derfor å bruke en slik transkripsjon i Nota-Oslo.

Ortografisk transkripsjon skal altså gjengi informantens tale så godt det lar seg gjøre innenfor det skriftnormalen tillater. Den skal med andre ord være et best mulig kompromiss mellom krav om nøyaktig gjengivelse av det talte, og skriftmålets normer.

1.3 Diskusjon: hvordan bør en ortografisk transkripsjon være?

Hva ligger så i at vi skal transkribere teksten ortografisk? Dersom vi ser hva andre prosjekter har gjort, ser vi at det er mange mulige løsninger. Nedenfor har vi forsøkt å stille opp fordeler og ulemper ved tre forskjellige typer transkripsjon:

Full norm	Full norm med varianter	Full variasjon med visse begrensninger
+ effektivt + kostnadseffektivt + lett å være konsekvent + lett å arbeide videre med + enkelt å tagge automatisk - av og til må man gjette hvilke normert form som er ment (<i>Jeg så a -> Jeg så hun/henne - Hva het'n -> Hva het han/den</i>) - transkribørene kan ha forskjellig norm - hvilken norm skal man velge? - den endelige transkripsjonen reflekterer ikke det informantene sier - umuliggjør søk etter faktiske former	- ikke fullt så effektivt - mer arbeid - vanskeligere å være konsekvent - krever mer arbeid å lage en norm for variantene og å plukke ut hvilke varianter man skal transkribere + muliggjør søk etter faktiske former + reflekterer i større grad det informantene faktisk sier	+ reflekterer det folk sier - vanskelig å tagge og søke i, kan lett bli rotete

Som oversikten ovenfor viser, har metodene *full norm* og *full norm med varianter* begge både fordeler og ulemper. *Full variasjon med visse begrensninger* er urealistisk å gjennomføre. De prosjektene vi kjenner til, har alle valgt ulike former for *full norm med varianter*, og dette er også vårt mål. Metoden er imidlertid såpass tidkrevende at vi i hovedsak satser på å transkribere materialet etter *Full norm*-metoden. Unntaket er pronomen der vi også skriver inn uttalevariasjoner som avviker fra standard østnorsk uttale (se 3.5.1.3). Siden planlegger vi å gå igjennom materialet med tanke på å skrive inn flere varianter. Til dette håper vi å kunne bruke en automatisk eller halvautomatisk metode. 10 % av materialet vil bli transkribert fonetisk slik det er beskrevet i prosjektsøknaden.

2. Praktisk veiledning for transkripsjon

2.1 Gode råd for transkripsjon

Det er viktig å finne fram til gode rutiner og vaner for transkribering. Målet er at arbeidet skal gå både raskt og feilfritt. Det er ikke sikkert at det som fungerer for en transkribør, fungerer for alle. Nedenfor er likevel noen gode råd du bør merke deg:

- Ta hyppige, korte pauser der du fokuserer øynene på noe annet og strekker på beina
- Arbeid aldri mer enn fire timer av gangen med transkripsjon
- Lær deg flest mulig hurtigkommandoer. Det sparer tid og ikke minst hendene dine
- Hør på ca 30 sekunder av lydfila først for å få en oversikt over samtalen. Ta deretter for deg 10 og 10 sekunder av gangen og transkriber
- Fokuser på det du hører, det er dessverre fort gjort å skrive det en forventer i stedet for det som faktisk blir sagt

2.2 Filorganisering

NoTa-Oslo har sitt eget fellesområde *nota* på serveren *Foni*. Her ligger blant annet alle filene vi skal bruke i transkripsjonen: lydfiler, transkripsjonsfiler og loggfiler samt felles dokumenter som informantbrosjyre og samtykkeerklæring.

Lydfiler: ligger i *Lyd*-mappa. Navnet på lydfilene består av to eller tre ledd forbundet med en lav strek med *wav* som etternavn. Det første leddet er enten *samtale*, *intervju* eller *telefon*. Det neste leddet er informantnummer. Er det *samtale* eller *telefon*, består leddet av to informantnummer med bindestrek mellom. Er det *intervju*, består leddet av informantnummer pluss bindestrek og initialene til intervjueren. Når en transkribør har plukket ut ei fil for transkribering, omdøper han eller hun fila slik at fila får transkribørens initialer først fulgt av en lav strek:

```
samtale_001-002.wav  
KH_samtale_001-002.wav  
  
intervju_001-KH.wav  
KH_intervju_001-KH.wav
```

I det siste eksempelet er *KH* både transkribør med initial først i filnavnet og intervjuer med initial bak nummeret til informanten som er med på intervjuet.

Transkripsjonsfiler: ligger i *Transkripsjon*-mappa og blir laget etter hvert som transkribørene velger en lydfil for transkripsjon. Filene skal hete det samme som lydfilene, men får automatisk *trs* som etternavn av transkripsjonsprogrammet. I *Transkripsjon*-mappa er det to mapper: *Til korrektur* og *Korrekturlest*. Når du er ferdig med å transkribere ei fil, flytter du fila over til *Til korrektur*-mappa. Dette må du gjøre etter at du har gått ut av programmet. Når fila er ferdig korrekturlest (se avsnitt 4), legger du fila inn i *Korrekturlest*-mappa.

Loggfiler: ligger i *Logg*-mappa. Hver transkribør skal ha ei loggfil der arbeidet for hver arbeidsøkt er loggført. Transkribørene bør også ha sin egen problemfil der de noterer alle spørsmål tilknyttet problemmerknader, se 3.2 nedenfor. Loggfilene og problemfilene har transkribørens initialer som fornavn og *logg* og *prob* som etternavn.

KH.logg

KH.prob

Loggfila kan se slik ut:

Dato	timer	Arbeid
11.10.04	6	tatt opp samtale_001-002.wav og samtale_003-004.wav
12.10.04	6	transkribert ½ time av KH_samtale_001-002.wav
...		

2.3 Transkriberingsprogrammet *Transcriber*

Transcriber er et fransk transkriberingsprogram. Vil du lese mer om det, kan du gå inn på nettsiden: <http://trans.sourceforge.net/en/presentation.php>

I dette avsnittet beskriver vi kort hvordan programmet brukes. Nettsiden har ellers en manual som er grei å bruke.

2.3.1 Begynne på en ny transkripsjon

Gå inn i *Lyd*-mappa via skrivebordet. Velg den første fila som er ledig, det vil si den fila som ikke har noe initial først i filnavnet. Omdøp den ved å legge til dine initialer først i filnavnet:

Opprinnelig navn på fila: **samtale_001-002.wav**

Nytt navn på fila: **KH_samtale_001-002.wav**

Deretter går du inn i *Transcriber*. Med <Ctrl-n> eller **New trans** under **File** åpnes ei ny fil for transkripsjon. Du blir først bedt om å velge ei lydfil. Gå til *Lyd*-mappa på *nota*-området og velg den fila du akkurat har gitt nytt navn til. Lagre fila med en gang i mappa *Transkripsjon* ved å bruke <Ctrl-s>. Transkripsjonsfila vil automatisk få det samme navnet som lydfila, men med **trs** som etternavn. Når fila er lagret, velg **Edit episode attributes** under **File**. Her skriver du inn navnet ditt under **Transcribers name**.

2.3.2 Fortsette med en påbegynt transkripsjonsfil

Med <Ctrl-o> eller **Open trans** under **File** åpner du ei påbegynt transkripsjonsfil. Fila ligger i *Transkripsjon*-mappa.

2.3.3 Lag report

Velg **Edit section attributes** under **Segmentation** eller klikk på den røde firkanten der det står *report*. Velg *New Topic* og skriv inn det som passer best av følgende karakteristikker:

Intervju
Samtale – familie
Samtale – venner
Samtale – bekjente
Samtale – ukjente

Trykk deretter *OK*. Er du usikker på hva slags forhold samtalepartnerne har til hverandre, kan du slå opp i *Informantmatrisen* som ligger på fellesområdet under *Dokumenter* og *NoTa-dokument*.

2.3.4 Definer talere

Under **Segmentation** skal du velge **Edit turn attributes** og **Create speaker**. Skriv inn identitetsnummeret til informant. De andre mulige valgene i skjemaet lar du være umarkert.

Hver gang en av informantene tar ordet, trykk <Ctrl-t> og velg informant. Vi har valgt å bruke begrepet *tur* for en slik sekvens der en informant har ordet, se 3.6.1.

2.3.5 Tidskoding og segmenter

Ved å trykke **Enter**, settes det automatisk inn en tidskode i dokumentet slik at lydfila blir forbundet med transkripsjonen. Dette markeres med en grønn runding. Teksten mellom to tidskoder kaller vi *segment*, se 3.6. Nedenfor ser vi en tur som er delt opp i fem segment:

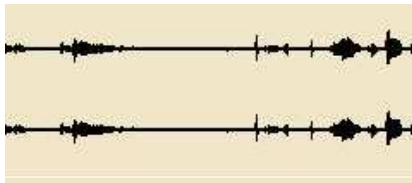
Monica

- ① ja fordi at jeg går rundt og prøver være venner med alle
- ② og det er ikke noe jeg har bevisst prøvd å gjøre
- ③ jeg har prøvd å være hyggelig mot folk
- ④ for det er ingen her inne jeg direkte misliker så hvorfor skulle jeg gå rundt da å være ...
- ⑤ [snusfing] ja, selvfølgelig er det noen jeg liker mer enn andre men det er # jeg kjenner ikke alle godt nok til å ...

2.3.6 Avspilling av lydfile

Nederst i Transcriber-vinduet ser du et spektrogram. Med **Tab** starter og stopper du avspillingen og kan se hvordan en rød strek flytter seg bortover i spektrogrammet. Spektrogrammet kan være til hjelp når du skal sette inn en tidskode.

Dersom du trykker **Shift Tab**, får du avspilt det segmentet markøren står på. Ved **Tab** starter avspillingen på begynnelsen av det segmentet markøren står på, og fortsetter til den blir stoppet. Du kan også markere ett område i spektrogrammet og høre på akkurat dette området ved å trykke **Tab**. Nedenfor er et lite utsnitt av et spektrogram:



2.3.7 Stavekontroll

Når du er ferdig med å transkribere ei fil, skal fila kontrolleres med stavekontrollen i *Transcriber*. Velg **Spell checking** under **Edit**-menyen. Stavekontrollen vil stoppe når den finner et ord som ikke finnes i den ordlista stavekontrollen bruker. Ordlista i stavekontrollen bygger på *Bokmålsordboka*. Når stavekontrollen stopper på et ord, må du likevel i hvert enkelt tilfelle vurdere om stavekontrollen har stoppet på en feil eller på et ord vi har definert som godkjent i denne veiledningen. Noen eksempler:

- er ordet en godkjent sammensetning (se 3.5.3.2) som ikke står i *Bokmålsordboka*, velger du **Ignore** i stavekontrollen. Er dette en sammensetning som er hyppig brukt, kan du velge **Add to dictionary** slik at stavekontroll ikke reagerer på dette ordet neste gang.
- er ordet avbrutt og derfor bare skrevet halvveis ut med bindestrek (se 3.5.5.1), velger du også **Ignore**
- er ordet et dialektord eller et annet ord som ikke står i ordboka, skal det ha korrekt eventmerking (se 3.5.3.5 og 3.5.3.6). Stavekontrollen vil stoppe på ord som ikke har slik merking, og du må gi ordet korrekt event.
- er ordet en feilskrivning, skriver du rett ord i stavekontrollvinduet eller velger korrekt ord fra den lista som stavekontrollen presenterer. Velg så **Replace**.

3. Transkripsjonsregler

3.1 Om disse reglene

Det er viktig at de som skal bruke korpuset, kan stole på at teksten tilfredsstillende bestemte spesifikasjoner, og at disse spesifikasjonene er de samme fra tekst til tekst. Dette stiller store krav til oss som transkribører på et prosjekt der det er flere som skal transkribere. Nedenfor er det satt opp en del regler for transkripsjon. Les dem nøye, og følg dem!

3.2 Når reglene ikke strekker til

Selv om reglene nedenfor er forholdsvis detaljerte, vil vi helt sikkert komme bort i tilfeller der reglene ikke strekker til, enten fordi det ikke finnes noen regel for dette, eller fordi tilfellet er et grensetilfelle der vi er usikre på hva vi skal gjøre. Er du usikker, gjør følgende:

- Transkriber så godt du kan, men sett inn en problem-merknad med **<Alt-p>**. I eksempelet nedenfor er transkribøren usikker på om pausemarkøren # er brukt på riktig måte:

```
han har bestemt seg # ja du at hva? {problem}
```

- Loggfør problemet i loggfile og ta det opp på neste transkripsjonsmøte

Loggfør heller ett problem for mye enn ett for lite. Det er viktig at vi blir enige og løser vanskelige tilfeller på samme måte!

3.3 Utvalg og sensitive personopplysninger

Følgende skal ikke med i transkripsjonen:

- Eventuelt småprat om opptaket/opptakssituasjonen før samtalen eller intervjuet har kommet i gang. Skriv en *innledende bemerkninger*-kommentar **<Alt-i>**:

```
{no speaker}  
{innledende bemerkninger}
```

- Avbrudd i samtalen som for eksempel skyldes at mikrofonen faller av, at noen kommer inn i opptaksrommet og avbryter samtalen, at noen får hosteanfall osv. Skriv *avbrudd*-kommentar **<Alt-a>** og **no speaker** :

```
{no speaker}  
{avbrudd}
```

- Navn på venner, familiemedlemmer, kolleger osv. Vi har egne regler for hvordan vi skriver navn, se 3.5.3.4.
- Sekvenser som inneholder sensitive personopplysninger. Dette er ifølge NSD (Norsk samfunnsvitenskapelig datatjeneste) følgende: *opplysninger om rasemessig eller etnisk bakgrunn, politisk, filosofisk eller religiøs oppfatning, at en person har vært mistenkt, siktet, tiltalt eller dømt for en straffbar handling, helseforhold, seksuelle forhold, medlemskap i fagforeninger.*

Dersom opptaket inneholder slike sekvenser, skal det merkes med en kommentar (<Alt-s>) som forteller at sensitive opplysninger er tatt ut av transkripsjonen:



At navn og sensitive personopplysninger er tatt ut av transkripsjonen, betyr ikke at de er fjernet fra lydfila. I ettertid kommer vi til å søke fram alle steder i lydfila der transkripsjonen er merket med *sekvens med sensitive personopplysninger* og erstatte originallyden med pip eller annen lyd. Dette kommer vi ikke til å gjøre med alle navn. Det viktigste er at navnene ikke er søkbare. Dersom du likevel vurderer at det er uheldig at et navn blir koplet mot en spesiell hendelse, setter du <Alt-s> bak navnet slik at vi kan finne det igjen og fjerne det fra lydfila.

Informantene skal ha beskjed om at de ikke skal snakke om forhold som karakteriseres som sensitive personopplysninger eller om tredjeperson. Vi får derfor håpe at det ikke forekommer så ofte i lydfilene.

3.4 Markering av ekstrainformasjon

All slags ekstrainformasjon som skal angis i transkripsjonene, må gis som *event* (hendelse). **Insert event** finnes under **Edit**-menyen. Vi skal bruke fire typer eventer:

- **noise** (til lyder som hosting, kremting, sukking og klikkelyd) - Hurtigkommando <Ctrl-d>
- **pronounce** (til å markere en meningsenhet (me) som går over flere segment (se 3.6.3) eller til å markere når noe f.eks. er uklart uttalt eller stavet i stedet for å bli uttalt på vanlig måte) - Hurtigkommando <Ctrl-r>
- **language** (til å merke ord som ikke står i *Bokmålsordboka*, f.eks. engelske ord) – Hurtigkommando <Ctrl-l>
- **lexical** (til å spesifisere uttalen på noen ord, først og fremst pronomener, se 3.5.1.3) – Hurtigkommando <Ctrl-e>
- **comment** (til kommentarer som *problem* (se 3.2), *sensitive personopplysninger* (se 3.3) eller *uforståelig* (se 3.5.5.4))

For mange eventer, blant annet for alle kommentarene, har vi laget egne hurtigkommandoer. Har du behov for å sette inn en type kommentar eller ekstrainformasjon som ikke er listet opp i denne veiledningen, behandler du det som en *problem*-merknad (se 3.2). En oversikt over alle eventer og hurtigkommandoer finnes i 5. Hver enkelt event blir nærmere beskrevet med eksempler i punktene under 3.5.

Det skal alltid være mellomrom før og etter eventene. Dersom eventen beskriver ett spesielt ord, skal eventen stå etter ordet. I dette eksemplet beskriver eventen *yes* som engelsk:

```
yes +[lang=English] slik er det
```

Det er ikke meningen at teksten skal leses alene (altså uten lyd og bilde), så det er ikke nødvendig å ta med "teateraktige" eventer som *snur seg*, *løfter koppen*. Er en informant forkjølet, er det heller ikke nødvendig å ta med hvert host eller snufs. Utfordringen blir å ta med ekstrainformasjon som er betydningsfull for meningen med ytringen. Ett host eller kremt er for eksempel ofte en type nøling (*hva var det jeg skulle si...*) som det er viktig å få med i transkripsjonen. Eventer som hosting og latter skal settes der de forekommer i sammenhengen:

```
[latter] det var morsomt
```

Noen eventer kan gå over flere ord – eller over flere tidskoder. Sett inn en event ved det første ordet ved å velge eventtype og den eventen du vil ha (for eksempel *leende*). Marker i tillegg **Start of event** i eventboksen. Etter det siste ordet går du igjen inn i eventboksen og velger samme event, men marker **End of event** før du trykker **OK**.

```
[leende-] det var morsomt [-leende]
```

Dersom du markerer det området som skal "rammes inn" av event-taggen og velger hurtigkommandoen for eventen, vil den automatisk få start- og slutt-tag. Legg merke til at taggene vil legge seg helt inntil ordene den rammer inn slik at du må legge inn mellomrom.

3.5 Språkform

En ortografisk transkripsjon skal gjengi informantens tale så godt det lar seg gjøre innenfor det skriftnormalen tillater. Den skal med andre ord være et best mulig kompromiss mellom krav om nøyaktig gjengivelse av det talte, og skriftmålets normer.

Dersom et system som det som presenteres nedenfor, overhodet skal ha noen hensikt, må reglene følges, selv om resultatet i enkelte tilfeller skulle stride mot språkfølelsen din eller mot det du ellers har lært om hvordan en tekst skal utformes. Husk at resultatet ikke først og fremst skal bli en mest mulig lesbar tekst, men et teknisk søkeredskap i forhold til de digitaliserte innspillingene de er transkripsjoner av.

Det er lett å instinktivt overse gjentakelser og grammatisk inkonsekvens i tale, og det krever litt trening å "undertrykke" den impulsen og høre hva informantene *faktisk* sier.

Vi skal transkribere med bokmål, og følge *Bokmålsordboka* (Wangensteen 2004) med hensyn til hovedformer og klammeformer. Denne regelen har to unntak som vi vil beskrive nedenfor. Nedenfor vil vi også beskrive hva vi skal gjøre med ord som ikke står i Bokmålsordboka.

3.5.1 Hovedregler

Vi følger tre hovedregler for transkripsjon:

Hovedregel 1: Skriv bare former som finnes i Bokmålsordboka

Informanten sier:

jei jikk på vegen
jei mener det ass
ska vi sjå

Vi transkriberer:

jeg gikk på vegen
jeg mener det altså
skal vi se

Hovedregel 2: De syntaktiske reglene for skriftlig bokmål tas ikke hensyn til i den forstand at ordene skal skrives ned i den rekkefølgen de blir uttalt

hva du sa jeg skulle skylle ferdig før du pussa det?

Hovedregel 3: Bruk godkjente former fra Bokmålsordboka selv om de er brukt på ”feil” sted i forhold til det den normative grammatikken sier.

Informanten sier:

henne jikk
vi snakka på det
jei ga det til de
jei bruker ei maskin
da får døm si det sjøl
de er sånn sjokoladepulver
låkk ijen vindu

Vi transkriberer:

henne gikk
vi snakka på det
jeg ga det til de
jeg bruker ei maskin
da får dem si det sjøl
det er sånn sjokoladepulver
lukk igjen vindu (ikke vinduet)

3.5.1.1 Hovedformer og klammeformer

Vi bruker alle godkjente former fra Bokmålsordboka, også klammeformer. Det er altså ikke lov til å skrive *sne*, men det er lov å skrive *mjølken*, *åssen*, *ho* og *trur*. Velg alltid den formen som ligger tettest opp til det originalt uttalte ordet, bruk for eksempel a-endelser der det sies. Det er vanskelig å ha en fullstendig oversikt over hva som er tillatt, f.eks. når det gjelder klammeformer. Ikke bruk tid på å lete etter godkjente former i Bokmålsordboka, men bruk klammeformer du kjenner dersom de ligger nærmest det informanten sier. Det kan lønne seg å ha nettsiden til Bokmålsordboka klar på skjermen slik at du lett kan slå opp det du lurer på angående ortografi og bøyning. Nettsiden finner du her:

(<http://www.dokpro.uio.no/ordboksoek.html>)

Noen hovedformer og klammeformer er lite brukt i skriftlig bokmål. Noen av disse kommer vi imidlertid til å ha hyppig bruk for. Mange i Oslo-området bruker for eksempel både *vart* og

blei i preteritum. Legg merke til at begge er hovedformer i Bokmålsordboka. Kløyvd infinitiv er også vanlig å bruke i deler av Oslo-området. Kløyvd infinitiv er klammeform i Bokmålsordboka, og dermed skal vi transkribere for eksempel *gjøra* og *væra* dersom informanten sier det. I 5.5 er det en liste over ord som kan ha a-infinitiv i kløyvd infinitiv. I 5.4 er det en liste over andre ord du skal være spesielt oppmerksom på og bruke.

Av og til kan det være vanskelig å avgjøre hvilken godkjent form fra Bokmålsordboka vi skal bruke, fordi det informanten sier, ikke likner på noen godkjent form. Dersom ordet kan karakteriseres som muntlig eller dialektnært, er det mest naturlig å velge den formen i Bokmålsordboka som er den antatt mest muntlige, for eksempel *rauk* for *røyk* (presens av *ryke*) eller *flytta* for preteritum *flytte*

Informanten sier:	Vi transkriberer:
sne	snø
røyk	rauk
flytte (pret)	flytta
mjølken	mjølken
åssen	åssen
trur	trur
vart	vart
blei	blei
væra	væra
leva	leva
hu	hun <i>+_[lex=hu]</i>

Les mer om spesialbehandling av pronomen i 3.5.1.3.

3.5.1.2 Variasjon

Å bruke Bokmålsordboka som norm kan i noen tilfeller virke litt brutalt. Det er imidlertid en regel det er greit å forholde seg til når en blir vant til det. Legg merke til at vi skal bruke bokmålsordboka som norm både når det gjelder **fonologisk stammevariasjon** og når det gjelder **uttalevariasjoner i bøyingsuffiks**:

Informanten sier:	Vi transkriberer:
itte	ikke
søvi	sovet
hestær	hester
prate (presens)	prater

3.5.1.3 Spesialbehandling av pronomen

Personlige pronomenes behandles spesielt. Vi transkriberer som vanlig med Bokmålsordboka som norm, men spesifiserer uttalen etter ordet med eventen **lexical** <Ctrl-e>:

```
• jeg tok boka deres +[lex=dems]
```

Det er bare uttale som avviker fra uttalen i standard østnorsk som markeres på denne måten. For eksempel markerer vi ikke at *jeg* uttales *jei* eller at *meg* uttales *mei*. Derimot spesifiserer vi uttalen *hu* for *hun*:

```
hun +[lex=hu] jenta
```

3.5.1.4 Sammentrekking

Sammentrekking markeres ikke.

Informanten sier:	Vi transkriberer:
jei kan'ke gå	jeg kan ikke gå
de æ'kke slik	det er ikke slik
det kom no gutter	det kom noe gutter

Unntaket er *n*, se 3.5.2 nedenfor.

Legg merke til at *no gutter* transkriberes som *noe gutter* fordi *noe* også er en flertallskvantor.

3.5.2 To unntak fra hovedregelen

Det er to unntak fra hovedregelen om at vi bare bruker godkjente former fra Bokmålsordboka:

- Pronomenene *a* og *n*
- Kjønn beholdes på substantiv

3.5.2.1 Pronomenene *a* og *n*

Hunnkjønnspronomenet *a* er svært vanlig i Oslo-området og er ikke et godkjent pronomen i følge Bokmålsordboka. *a* kan både tilsvare *hun* og *henne*:

```
der er a  
jeg så a i går
```

For å se om *a* tilsvare *hun* eller *henne*, måtte en ha sett på konteksten, og ellers kjent til om informanten vanligvis brukte subjektform eller objektform av pronomenen i samme posisjon.

Mange informanter ville for eksempel bruke objektsform i dette tilfellet: *der er dem*, og da ville det ha vært naturlig å anta at *a* i *der er a* tilsvarte *henne*: *der er henne*.

For at vi skal slippe å gjette hva *a* betyr i en kontekst, transkriberer vi i stedet *a* som *n*. Tilsvarende med *n*. *n* kan være en forenkling både for *han* og for *den*. For å slippe å gjette ut i fra konteksten, skriver vi *n*:

jeg så n

n blir alltid skrevet som *n*, også i tilfeller der det er klart hva *n* er en forenkling for:

jeg så n Lars

3.5.2.2 Kjønn beholdes på substantiv

Det andre unntaket gjelder kjønn. *maskin* er et hannkjønnsord i følge Bokmålsordboka, men som vist i hovedregel 3 under 3.5.1 overfor, skriver vi *ei maskin* dersom informanten sier det, fordi vi har valgt å transkribere det informanten sier når det er en godkjent form - og *ei* er en godkjent form i Bokmålsordboka. Tilsvarende skriver vi *maskina* selv om dette ikke er normen. Når vi har valgt å gjøre det, er det både fordi det er logisk å beholde kjønn i bestemt form når det beholdes i ubestemt form, men det er også fordi kjønn er avgjørende for kongruens i tilliggende ord. Kjønn er noe som varierer sterkt fra dialekt til dialekt. Når vi ikke retter opp kjønn, slipper vi også å slå opp substantiv vi er usikre på.

Informanten sier:	Vi transkriberer:
ei maskin	ei maskin
maskina	maskina

3.5.3 Ord som ikke står i Bokmålsordboka

Mange ord står ikke i Bokmålsordboka. Dette kan være dialektord, nyere ord, banning eller lydmalende ord som *eh* og lignende. Disse formene må vi selvfølgelig også transkribere. Nedenfor følger en oversikt over hva vi gjør med slike ord.

Av og til kan det være vanskelig å skille mellom hva som er en lydlig variant av et ord som står i Bokmålsordboka og hva som er et eget leksem som skal behandles spesielt, for eksempel et dialektord (se 3.5.3.5 nedenfor):

Hovedregelen er slik: et ord er et eget leksem dersom det ikke er en lydlig eller morfologisk variant av et leksem som står i Bokmålsordboka.

Er du i tvil om et ord, så diskuter det med en av de andre. Slå aller først opp i ordboka. Både ord som for eksempel *bedritent* og *faen* står faktisk der.

Typiske varianter av leksem som står i ordboka, kan være dialektvarianten *itte* for *ikke* og bøyingsvarianten *bært* for *båret*. Av og til må man ut i fra sammenhengen kunne tolke om et ord er en variant av et leksem eller et eget leksem: Om informanten sier at hun *pikk sett ham* må vi kunne anta at *pikk* er *fikk* og ikke kjønnsorganet. Det samme er tilfelle dersom informanten sier at *han kom bærene*. Her er det neppe bølger det er snakk om, men en variant av *bærende*. Motsatt vil vi anta at *mor* i følgende setning er et eget leksem og ikke leksemet som står oppført i Bokmålsordboka: *mor de var bra asså*. Informanten bruker ordet tydelig som interjeksjon og ikke henvendt til moren sin.

En regel som kan brukes for å avgjøre om et ord skal betraktes som et eget leksem eller som en variant av et annet leksem, er at en ortografisering eventuelt ville bli en ren gjetning uten grunnlag. Ordet *flingre* kan for eksempel bety både *liten bit* og *flis*, og det ville være meningsløst å forsøke å gjette om *flingre* skulle normaliseres til *flis* eller *liten bit* i transkripsjonen. Det samme ville gjelde for *arjkynt* som kan bety både *sint*, *egen* og *sta*. Her skriver vi altså *flingre* og *arkjynt*.

Når ordet ikke kan være en variant av et leksem i Bokmålsordboka, skal det merkes som egen event, **language**, se punktene nedenfor. Unntaket er *a* og *n* (se 3.5.2), som ikke merkes spesielt.

I boksen nedenfor er det både ord som er varianter av ord i Bokmålsordboka og ord som er egne leksem (merket med *lang=X*, se 3.5.3.5 nedenfor):

Informantene sier:	Vi transkriberer:
det er illere det er bedritent vi er itte slik jeg skulle bært sekken jei pikk sett ham han kom bærene med de	det er illere det er bedritent vi er ikke slik jeg skulle båret sekken jeg fikk sett ham han kom bærende med det
de æ arkjynt	det er arkjynt <i>+ [lang=X]</i>
sjå ei flingre!	se ei flingre <i>+ [lang=X]</i>
dette kuper	dette kuper <i>+ [lang=X]</i>
mor de var bra asså	mor <i>+ [lang=X]</i> det var bra altså
jei prøvde og lese jei prøvde å lese de skulle vært artig å prøvd	jeg prøvde og lese jeg prøvde å lese det skulle vært artig og prøvd

De fleste informanter vil bruke *å* både for *å* og *og*. Sier de *og*, skriver vi *og* i tråd med hovedregel 3 (Bruk godkjente former fra Bokmålsordboka selv om de er brukt på "feil" sted). Sier de *å*, transkriberer vi ved å bruke *og/å*-reglene fra skolegrammatikken.

3.5.3.1 Forkortelser

Forkortelser skrives som forkortelser. Forkortelser som *dvd*, *cd* og *pc* skrives med små bokstaver.

```
de sa det på NRK
det ble vedtatt i FN
det var to dvd-er
```

3.5.3.2 Sammensetninger

Mange sammensetninger står ikke i Bokmålsordboka, men sammensetninger skal likevel normaliseres på vanlig måte:

```
a-endelse
l-lyden
bokstavkombinasjonen
trafikksituasjon
barne- og ungdomsarbeider
```

3.5.3.3 Tall

Tall: skrives med bokstaver. Slik får vi med om f.eks. *1600* uttales *ett tusen seks hundre* eller *seksten hundre*. Tall under hundre skrives som ett ord, alle tall over hundre skrives i flere ord. Legg merke til at tallordet *1* skrives uten aksent.

```
tjuefire
trettisju
sekstifire tusen
ni millioner
tre tusen to hundre og tjuefem
en
```

Unntaket er årstall som skrives som ett ord:

```
nittenhundreogtrettiåtte
```

Ordenstall: skrives også i ett ord med bokstaver:

```
tjuefjerde
```

Ordenstall over hundre skrives i flere ord.

Brøker: skrives også med bokstaver:

to tredjedels

Mange informanter bruker den gamle tellemåten: *sju og seksti* eller *fem og tjuende*. Slike tall skriver vi i flere ord. Årstall med gammel tellemåte skrives derimot i ett ord slik som andre årstall: *nittensjuogtjue*. Legg merke til at de frekvente tallordene *syv*, *tyve* og *tredve* ikke er verken hovedformer eller klammeformer i Bokmålsordboka. Sier informanten *syv og tredve*, normerer vi det til *sju og tretti*.

Informanten sier

**sju og seksti
syv og tredve
fem og tjuende
nitten sju og tjue**

Vi skriver

**sju og seksti
sju og tretti
fem og tjuende
nittensjuogtjue**

3.5.3.4 Navn

Vi skriver vanligvis ikke inn navn på personer. I stedet skriver vi koder for hvert enkelt navn, *F* for jentenavn (fornavn eller fornavn pluss etternavn), *M* for guttenavn (fornavn eller fornavn pluss etternavn), og *E* for etternavn (uten fornavn). Navnene får fortløpende nummer etter som de dukker opp i teksten:

Informanten sier:

jei så at Kari Hansen ga Nina dokumentene til Nilsen

Vi skriver:

jeg så at F1 ga F2 dokumentene til E1

Navn på offentlige personer behøver som regel ikke anonymiseres.

Fotballag, stedsnavn, navn på kjæledyr og organisasjonsnavn m.v. skrives normalt inn dersom du ikke vurderer at det er sensitive personopplysninger involvert, se 3.3. Bør et slikt navn anonymiseres, skriv *N1*, *N2* osv. Gatnavn kan normalt stå, men dersom informanten oppgir hele adressen med gatnavn og nummer, erstattes adressen med *N*.

Navn på sanger, filmer og bøker skrives normalt med stor bokstav. Navn på verk skrives i hermetegn når det består av flere ord: "*Blow with the wistle*". Navn på leker som *sisten*, *gjemsel* og *egget*, skrives vi med liten bokstav. Dersom slike navn består av mer enn ett ord, setter vi også hermetegn rundt for å synliggjøre at ordene hører sammen: "*boksen går*", "*slå på ring*", "*hawk og due*".

Dersom et navn skal skrives som navn og ikke som kode, skal det skrives slik "eieren" vanligvis skriver det: *TV2*, *CatoSenteret*, *Ringenes Herre*, *TVNorge*. Dersom det er vanskelig å oppfatte navnet, skriver vi bare *N1*, *N2* osv. NB! Ikke bruk mye tid på navn!

Dersom en av samtalepartnerne bruker navnet på den andre samtalepartneren, bruker vi informantnummeret i stedet for de vanlige *F* og *M*.

Informanten sier:

teve Norge filma på Tjøtta # ikke sant Hans
vi lekte sisten og hawk og due

Vi skriver:

TVNorge filma på N1 # ikke sant 081
vi lekte sisten og "hawk og due"

NB! At navn er anonymisert i transkripsjonen, betyr ikke at de er fjernet fra lydfila. Dette er normalt ikke noe stort problem så lenge navnene ikke er søkbare. Dersom du likevel vurderer at det er uheldig at et navn blir koplet mot en spesiell hendelse, setter du **<Alt-s>** (for *sensitive personopplysninger*) bak navnet slik at vi kan finne det igjen og fjerne det fra lydfila.

3.5.3.5 Dialektord

Selv om informantene snakker dialekt, skal vi så godt vi kan forsøke å finne hva som er standard skrivemåte for ordet i Bokmålsordboka. Sier informanten *mjælka* eller *vøri*, skriver vi *mjølka* og *vært*. Men i enkelte tilfeller er det umulig å finne ordet i ordboka:

den fisken ser gøllei ut (= ekkel, dissete)

I slike tilfeller skriver vi ordet slik det høres ut, og merker det med eventen **language** og *X*, hurtigkommando **<Alt-x>**:

den fisken ser gøllei ut *+ [lang=X]*

Ord fra nynorsk som ikke finnes i Bokmålsordboka, beskrives også med **language=x**.

3.5.3.6 Ord fra andre språk

Dersom et utenlandsk ord eller uttrykk er brukt i en ellers norsk dialog som et alternativ til norsk tale, skal ordet eller uttrykket avmerkes. Engelsk har sin egen hurtigkommando <Alt-e>, for eksempel ved ytringer som denne:

```
yes +[lang=English] slik er det
```

Dersom det er snakk om andre språk, velg **Ctrl-I** og riktig språk fra lista som nås ved å sette musa over **Other** i event-boksen. Trykk til slutt **OK**.

Dersom det er en hel ytring som ikke er norsk, velger du **Ctrl-I** foran det første ordet. Velg riktig språk fra lista, men marker i tillegg **Start of event** i eventboksen før du trykker **OK**. Gå deretter til det siste ordet i ytringen og gjør det samme, men med **End of event**:

```
[lang=English-] we are the world [-lang=English]
```

Ett unntak fra denne regelen er **verk** som merkes i anførselstegn:

```
jeg elsker "Blow with the whistle"
```

Utenlandske ord skrives med godkjent rettskrivning fra det landet ordet kommer fra sammen med **language** og riktig språk.

Unntakene er:

- vi skriver engelske sammensatte ord med bindestrek:

```
jeg planlegger en bachelor-degree +[lang=English] i utlandet
```

- dersom utenlandske ord får norsk bøyning, skriver vi ordet på vanlig måte etterfulgt av bindestrek og norsk bøyingsendelse. Ordet får language-merking som vanlig: : *space-a, battle-er, family-en*

```
jeg safe-a +[lang=English] for family-en +[lang=English]
```

Ord som kan karakteriseres som *nynorsk*, regnes her som dialekt. Det vil si at ordene normaliseres til bokmål slik som andre dialektord gjør. Dersom det ikke finnes noe tilsvarende ord for ordet på bokmål, merker vi det som *language=x* (se 3.5.3.5):

Informanten sier

Vi transkriberer:

han vil ikkje sjå

Han vil ikke se

vedtake var samrøystes

```
vedtaket var samrøystes +[lang=X]
```

3.5.3.7 Nye ord, banning ...

I tillegg til dialektord og utenlandske ord er det mange andre ord som ikke står i Bokmålsordboka. Dette kan være for eksempel nyord (*blogg*) eller slang (*loke*) og andre ord som er vanlig i muntlig språk (*derre*). Når slike ord er brukt, behøver vi ikke bruke tid på å karakterisere dem som typer. Velg eventen **language** og **x** med hurtigkommando <**Alt-x**>.

```
den derre +[lang=X] bloggen +[lang=X]
```

Legg merke til at *derre* ikke står i Bokmålsordboka og skal merkes **language=x**, men at *herre* som i *den herre her* står i ordboka og ikke behøver slik merking.

Noen ganger kan det være vanskelig å avgjøre om et ord skal ha **language=x** eller om det er en ny sammensetning, f.eks. ved avledninger som *regnete* eller *bilaktig*. Dersom du er i tvil, så merk ordet som **language=x**.

Det kan være vanskelig å avgjøre hvordan **language=x**-ordene skal skrives. Hovedprinsippet er at de skal skrives mest mulig uttalenært. Er du i tvil, kan du søke etter den vanligste skrivemåten på Atekst. Du kan også bruke ordbøkene *Kebabnorsk ordbok* (Østby 2005) eller Kunnskapsforlagets *Slangordboka 2005*.

```
de er så kine +[lang=X] på å dra
```

Merk at ord fra engelsk eller andre språk skal skrives på originalspråket. Slike ord får **language=eng**, **language=tysk** osv.

3.5.3.8 Ikke-språklige lyder og trykk

Ikke-språklige lyder som latter, hosting og lignende som er del av en ytring, skal være med i transkripsjonen når de skaper mening i samtalen. Dette kan for eksempel være et host som markerer uenighet. Latter skaper nesten alltid mening, men et ukontrollert host skal ikke være med.

Lyder skal spesifiseres som eventen **noise**. Skriv **Ctrl-d** og velg korrekt event. Det samme gjelder ikke-språklige lyder som går sammen med en ytring (f.eks. *leende*, *gjespende*, osv.) sammen med **Start of event** og **End of event** som beskrevet i 3.4:

```
[gjespende-] jeg er så trøtt [-gjespende]
```

Latter er også en event, men er såpass hyppig at den har fått sin egen hurtigkommando: <**Alt-o**>:

```
[latter] det var morsomt
```

Latter sammen med en ytring, transkriberes som tilfellet *gjespende* ovenfor.

Ta også med bekreftende *mm*, spørrende *hm?* og den nølende *e*-lyden selv om de ikke er oppslag i Bokmålsordboka. Vi har laget en liste over interjeksjoner i avsnitt 5.3. Velg den

skrivemåten som passer lyden best herfra. Finner du ingen skrivemåte som passer, så skriv en problem-merknad <Alt-p>. Trykk markeres ikke.

3.5.4 Sitater og ord som er stavet eller lest

3.5.4.1 Sitater

Sitater settes i anførselstegn. Dette gjelder også hypotetiske sitater:

```
da kjørte jeg den "jeg? hæ?" da ljuger jeg
```

Dersom en informant snakker om uttalen av et ord, regnes dette også som sitat dersom ordet ikke er stavet (se 3.5.4.2).

```
jeg sier ikke "sne" jeg jeg sier "snø"
```

Legg merke til at vi ikke bruker *language X* eller andre eventer som markerer språk inne i sitater.

3.5.4.2 Ord som er stavet eller lest

Staving av ord: I ord som er stavet, skrives bokstavene en for en, og rammes inn av eventen **pronounce**:

```
kutt skrives [pron=stavet-] c u t [-pron=stavet] på engelsk
```

Skriv <Ctrl-r> og velg **stavet** og **Start of event** foran med <Ctrl-r>, **stavet** og **End of event** bak.

Internett skrives tilsvarende med *w w w* og *n o* rammet inn av **pronounce stavet** dersom det er slik det sies. Skriv *punktum* eller *dot* alt etter hva informanten sier. Sier informanten *dot*, skal det markeres med <Alt-e> for engelsk, se 3.5.3.6.

Høytlesing: Dersom noe leses høyt, skal dette markeres med **pronounce <Ctrl-r>** og **høytlesing**.

3.5.5 Avbrudd, pauser og uklare passasjer

3.5.5.1 Avbrutte ord

Når et ord er avbrutt, skal de delene som er uttalt, likevel gjengis i ortografi. Da markerer vi det ved hjelp av en bindestrek (-), som skrives uten mellomrom rett etter den siste bokstaven, for eksempel *han viss-* eller *han s-*. Det avbrutte ordet skal ikke skrives mer ortofont enn det ville blitt skrevet dersom det ikke hadde blitt avbrutt. Vi skal derfor ikke skrive *vo-*, *vo-*, *hvordan*, men *hvo-*, *hvo-*, *hvordan*.

Dersom en person stammer inne i ordet (*inne-ba-bane*) skriver vi det på vanlig normert måte *innebane*. En annen variant av dette er når første delen av en sammensetning gjentas flere ganger, kanskje ikke som avbrudd, men som forsterkning. Dette skriver vi slik: *kjempe-kjempe-kjempedeilig*. Hvis personen derimot begynner på et ord og retter seg selv inne i ordet, skriver vi det slik: *nittenseks- -syttiåtte* eller *yng- -dom*.

Ikke bruk bindestrek for avbrutt ord mer enn absolutt nødvendig. Er du i tvil om en person sier *d-* eller *det*, så skriv hele ordet med *uklart*-tagg bak. Stotring som *m- b- n-* bør heller betegnes med *uforståelig* eller *stønning*. Det er ikke så viktig å få med seg hver enkelt bokstav så lenge det ikke er mulig å forstå hva informanten forsøker å uttrykke.

Informanten sier	Vi transkriberer:
vo- vo- vordan	hvo- hvo- hvordan
inne-ba-bane	innebane
kjempe kjempe deilig	kjempe-kjempe-deilig
nitten seks søttiåtte	nittenseks- -syttiåtte
yng dom	yng- -dom
femte sjetteklasse	femte- sjetteklasse
femtesjetteklasse	femtesjetteklasse (sies som ett ord)

3.5.5.2 Uavsluttede ytringer

Dersom en ytring ikke blir skikkelig avsluttet, og en annen taler overtar, for eksempel ved avbrytelse o.a., skal den avbrutte ytringen avsluttes med tre punktum. Dette gjelder også når en taler forsøker å overta turen, men blir avbrutt og mislykkes.

høres ut som sånn her ...

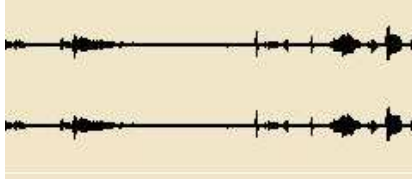
Hvis en taler avbryter seg selv og fortsetter med en annen ytring, skal vi ikke bruke tre punktum. Skriv den påbegynte ytringen og fortsett fortløpende med den neste:

du har du har du har ikke gjort det

Dersom det forekommer pauser i ytringen, se avsnitt 3.5.5.3

3.5.5.3 Pauser

Korte pauser markeres med #, mellomlange pauser med ## og svært lange pauser med ###. En mellomlang pause skal være såpass lang at det i praksis ville ha vært tid for en knappenål å falle i gulvet (ca 1 cm eller lengre i spektrogrammet, se nedenfor). En kort pause er kortere. Det kan være vanskelig å avgjøre når en skal markere korte pauser. En huskeregel kan være at det skal være et markert opphold uten lyd. Som regel vil en pause synes som en rett strek på spektrogrammet. Nedenfor er først en kort og så en mellomlang pause:



og # jeg tror ## det er slik at

En ekstra lang pause ### vil sannsynligvis forekomme forholdsvis sjelden i våre intervjuer og samtaler. Vi kan definere en ekstra lang pause som en pause som er så lang at samtalepartnerne ville synes det begynte å bli pinlig stille.

Pauser som er fylt med lyd, markeres ikke med pausetegn #. Ulike typer nøling som *e* osv markeres med ord fra lista i 5.3. Hosting, snufsing og klikkelyder markeres med eventen **noise** (se 3.5.3.8) når de skaper mening i samtalen. Ellers markeres de ikke. At informanten drar ut vokalene, som i *såååå det eeer slik* markeres ikke, heller ikke andre former for pauser og nøling inne i ord som ved *inne-ba-bane* (se 3.5.5.1).

Pauser forekommer som regel inne i segment.

du har drept # du har # du har drept ham

Mellom segmenter og turer (se 3.6) vil det som regel også være pauser, men disse pausene markerer vi ikke.

3.5.5.4 Passasjer som er vanskelige å forstå

Uforståelig: Skulle det være lydkombinasjoner som du ikke klarer å forstå, markerer du området med en *uforståelig*-kommentar, <Alt-u>.

men takk for at du {uforståelig}

Uklart: Skulle det være passasjer som du tror du forstår, men som du ikke er sikker på, setter du en *uklart*-event foran og bak (Velg **Ctrl-r** sammen med **uklart** og **Start of event** og **End of event** eller marker de uklare ordene og velg hurtigkommando <Alt-k>:

han sa at [pron=uklart-] han visste [-pron=uklart] at de skulle komme

Ett uklart ord markeres med <Alt-k> slik:

1: e dere så jo litt på det det skjemaet i stad da # nei e lista der
2: * mm +[pron=uklart]

3.5.6 Tegnsetting og stor bokstav

Stor forbokstav brukes ikke i begynnelsen av en setning. Navn skrives derimot med stor bokstav. Semikolon, kolon, komma, utropstegn og punktum brukes ikke. Spørsmålstejn brukes som vanlig ved spørsmål, også dersom spørsmålet blir avbrutt. Ved spørsmål man ikke forventer svar på (*det er slik ikke sant det bare blir sånn*), kan spørsmålstejn sløyfes.

```
ja men det er så vanskelig for noen ganger så husker du fra
bilder og så tror du at du husker det men du husker det egentlig
ikke da
```

```
kommer du i morgen?
har du hørt det?
hvem er det?
og han kjører fast egentlig bare på lastebilen eller har n ...?
det er slik ikke sant det bare blir sånn
```

Grensen mellom hva som er navn og hva som er vanlig substantiv er flytende. Navn på snop kan for eksempel være problematisk! Dersom ordet står i ordboka i betydningen snop (*seigmann*), skriver vi det med liten bokstav. Hvis ikke, skriver vi produktnavnet med stor bokstav (*Hobby*).

Når navn er bøydd, skriver vi det fortsatt med stor bokstav og uten bindestrek.

```
gi meg en seigmann
nei men du kan få en Hobby dersom du gir meg Nokiaen min
```

3.6 Tur, segment og meningsenhet

Begrepet *tur* brukes ofte for en sekvens der en person har ordet. Dette begrepet vil vi også bruke i denne veiledningen sammen med begrepene *segment* og *meningsenhet*. Segmenter og meningsenheter deler en tur opp i mindre enheter. De neste avsnittene skal forsøke å klargjøre hvordan vi skal bruke disse tre begrepene, og vi har brukt Linell og Gustavsson (1987) som referanse.

3.6.1 Tur

Linell og Gustavsson definerer *tur* som en periode da en person har ordet. Å ha ordet innebærer at man disponerer talerommet og har rett og eventuell plikt til å ytre seg.

Det er ikke alltid innlysende hvordan man skal dele en samtale inn i turer. Kanskje kan det hjelpe å tenke seg samtalen som en situasjon der samtalepartnerne har en stafettpinne de sender fram og tilbake mellom seg. Samtalepartneren som har pinnen, har også turen. Det skal litt til før en samtalepartner gir fra seg pinnen – den andre samtalepartneren kan kanskje

komme med små utrop eller kommentarer til det den første sier, men stafettpinnen vil fortsatt være hos førstemann. Vi kan også forestille oss at en samtalepartner avslutter sin tur og legger stafettpinnen på bordet uten at den andre tar den opp. Så følger kanskje en lang pause før den første tar pinnen opp igjen og begynner å snakke om noe annet. En slik sekvens vil tilsvare to turer selv om det er den samme samtalepartneren som snakker.

En tur (*turn* i *Transcriber*) settes inn i dokumentet med hurtigkommandoen <Ctrl-t>.

3.6.1.1 Replikk og ikke-replikk

Ofte vil stafettpinnen gå uproblematisk fram og tilbake mellom samtalepartnerne, men mange ganger kan det være vanskelig å avgjøre hvem som har turen. Her strekker ikke stafettpinnemetajoren til, og vi innfører begrepene *replikk* og *ikke-replikk* (fra Linell og Gustavsson, 1987) for å klargjøre turbegrepet:

En *replikk* er en ytring som blir sagt av den som har turen. En *ikke-replikk* blir sagt av den som ikke har turen. Nedenfor vil vi gå nærmere inn på hva som kan sies å være en *ikke-replikk*. Dersom vi har et klart bilde av hva en *ikke-replikk* er, kan vi definere *replikk* som det motsatte. Den samtalepartneren som har turen, er dermed den personen som ytrer replikker. Når den andre samtalepartneren ytrer noe som kan betegnes som en replikk, skifter turen over på vedkommende.

En *ikke-replikk* er ikke et konversasjonelt bidrag, og den påvirker verken turhavers tale i innholdsmessig retning eller turhavers tale med hensyn til aspekt på emne. En *ikke-replikk* kan være av to typer:

- *Feedback*: Taler B signaliserer at hun er enig, forstår og at A skal fortsette. Typisk feedback er ytringer som *ja*, *mm*, *oi*, latter osv.
- *Mislykket avbruddsforsøk*: B forsøker å avbryte A, men mislykkes slik at A beholder turen. (Dersom B lykkes og ”vinner turen”, er det en replikk.)

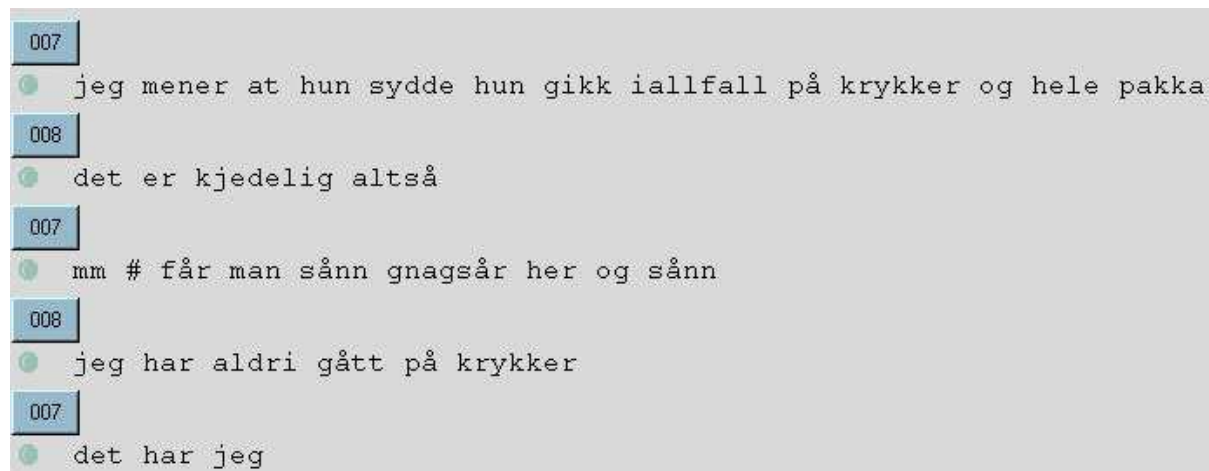
Andre kjennetegn på *ikke-replikker*:

- ikke svar på spørsmål (svar på spørsmål er *replikk*)
- kort latenstid (pause) eller overlapp mellom A og Bs ytringer
- svak lydstyrke eller lite omfang
- ofte omformuleringer eller korte spørsmål

Altså: Dersom noe ikke er feedback eller mislykket avbruddsforsøk, er det en replikk.

3.6.2.2 Tur og turskifte

For å oppsummere: Når vi transkriberer, skal vi dele samtalen opp i *turer* (<Ctrl-t>). En *tur* inneholder en eller flere *replikker* fra samme taler:



The screenshot shows a transcription interface with a light gray background. It displays a dialogue between two speakers, identified by blue boxes with white text: 007 and 008. Each speaker's turn is preceded by a small green circle icon. The text is as follows:

007
jeg mener at hun sydde hun gikk iallfall på krykker og hele pakka

008
det er kjedelig altså

007
mm # får man sånn gnagsår her og sånn

008
jeg har aldri gått på krykker

007
det har jeg

Turskifte oppstår når taler A ved pause eller intonasjon signaliserer at det er B sin tur, og B tar ordet. Det er også turskifte dersom B klarer å overta turen ved å snakke såpass høyt eller overbevisende at A til slutt slutter å snakke. Det er ikke tur-skifte om taler B kommer med ulike former for feedback, se 3.6.3.1.

At samtalepartnerne snakker i munnen på hverandre, er svært vanlig i en naturlig samtale. Ofte er det slik at B kommer med overlappende feedback samtidig med at A snakker i sin tur. Overlappende tale kan også være mislykkede eller vellykkede forsøk fra B sin side på å ta turen. Se 3.6.3.2 for å lese mer om overlappende tale.

3.6.2 Meningsenhet og segment

Vi bruker begrepet *meningsenhet* for den minste enheten en samtale skal deles opp i. I følge Linell og Gustavsson (1987) er en meningsenhet en formelt og/eller semantisk avgrenset enhet som utgjør en ytring eller en del av en ytring. Vi kan bruke følgende kriterier for å dele ytringer inn i meningsenheter:

- Intonasjonskriterier (fallende, stigende og fortsettende tonefall)
- Avslutning av grammatisk enhet (helsetning, delsetning)

I *Transcriber* tilsvarer en meningsenhet nesten alltid et *segment*. Et segment starter med en grønn runding. Et nytt segment startes ved å taste *Enter*.

Monica

- 1: ja fordi at jeg går rundt og prøver være venner med alle
- 2: og det er ikke noe jeg har bevisst prøvd å gjøre
- 3: jeg har prøvd å være hyggelig mot folk
- 4: for det er ingen her inne jeg direkte misliker så hvorfor skulle jeg gå rundt da å være
- 5: [smuvsing] ja, selvfølgelig er det noen jeg liker mer enn andre men det er # jeg kjenner ikke alle

Foran hvert segment, altså ved hver grønne runding, settes det inn en tidskode i *Transcriber* som binder lyden sammen med det som er transkribert. I sluttproduktet til NoTa-Oslo vil segmentet være den minste enheten brukeren kan få som svar dersom han eller hun søker etter spesielle konstruksjoner i NoTa-korpuset. Er du i tvil om hvor du skal sette en segmentgrense etter å ha brukt kriteriene ovenfor, kan det kanskje hjelpe å forestille seg en søkesituasjon. Er segmentet svært kort, vil brukeren få lite informasjon om konteksten til det det søkes etter. Er segmentet svært langt, vil det være tungt å jobbe med. I eksemplet nedenfor er *ja* satt sammen med det påfølgende segmentet fordi et *ja* alene ville blitt et svært kort segment.

Pausemarkøren # viser at det er en pause mellom *ja* og resten av segmentet:

030+029

- 1: ja # eller jeg var veldig keen +[lang=English] på å ta n før
- 2:
- 3: 1:
- 4: 2: mm

Som hovedregel deler vi aldri opp meningsenheter i flere segmenter. Ett unntak er når taler 1 og taler 2 snakker i munnen på hverandre uten at meningsenhetene deres er synkronisert. I slike tilfeller lar vi taler 1s meningsenheter styre samtaleens oppdeling i segmenter, og deler opp taler 2s meningsenheter. For å ta vare på taler 2s meningsenheter, setter vi inn taggen **me** foran og bak meningsenheten. Mer om dette i 3.6.3.3.

007+008

- 1: ja for de regner med at det går sin gang
- 2: [pron=me-] ja det det går
- 3: 1: jeg hadde skikkelig traumer for at jeg ikke kunne klokken og sånn
- 4: 2: sikkert [-pron=me]

3.6.3 Innskudd av ikke-replikker, overlapping og delte meningsenheter

Som framstillingen ovenfor kommer inn på, foregår ikke segment- og turskifte alltid så smertefritt. Nedenfor følger en oversikt over hvordan vi løser ulike kompliserte tilfeller.

3.6.3.1 Innskudd av ikke-replikker

Det er vanlig at taler 2 kommer med en eller flere ikke-replikker som innskudd i løpet av taler 1s tur. Dette er som regel feedback av typen *mm*, *oj*, og liknende. Slike innskudd behandler vi slik:

Sett inn en tidskode ved hjelp av **Enter**. Trykk <Ctrl-t> og velg første taler i skjemaet. Kryss så av for **Overlapping speech** og velg deretter andre taler. (Noen ganger vil programmet selv gjette seg til hvem som er første og andre taler.) Etter å ha trykket **OK**, ser du nå at begge

talerne kommer opp i navnefeltet. Under er talerne nummerert med første og andre taler. Taler 1 har turen, og taler 2 er enten taus eller kommer med ikke-replikker mens taler 1 er taus.

008 + 007

1: husker vi m dere fikk sånn mat på døren
2:
1:
2: ja
1: og så var jeg med og levere det en gang så var jeg så misunnelig
2:
1:
2: mm

Noen ganger vil taler 2s ikke-replikker overlappe med taler 1s replikker, se 3.6.3.2.

Legg merke til at en tur dermed kan markeres på to måter: Enten som en tur der turhaveren har ordet alene slik som her:

Monica

ja fordi at jeg går rundt og prøver være venner med alle
og det er ikke noe jeg har bevisst prøvd å gjøre

Eller en tur der taler 2 har ikke-replikker som innskudd eller overlapping under hele eller deler av 1s tur slik som det forrige eksemplet ovenfor.

I noen av samtalene er det eksempler på at intervjueren også kommer med ikke-replikker. Disse transkriberer vi ikke dersom vi er nødt til å dele opp turen for å få plass til dem. (Som bildene ovenfor viser, opererer *Transcriber* bare med to talere under en tur, derfor er det enklest å kutte den tredje talerens ikke-replikker.)

3.6.3.2 Ikke-replikk eller del av replikk?

Ved turskifter kan det være vanskelig å avgjøre om det som blir sagt i overgangen mellom turene, tilhører den første eller den andre turen:

1: skada av å sitte på fly ikke sant|
2:
029 + 030
1: ja # men det er sånn du får i Norge òg ved å fly små fly
2:

I dette eksemplet kan *ja*-et i den andre turen kanskje også sies å være en ikke-replikk i den første turen. Vi har imidlertid valgt å se det slik at taler 2 aldri disponerer det siste segmentet i taler 1 sin tur alene. Derfor blir det siste taler 2 sier alene, alltid en del av replikken i neste tur.

3.6.3.3 Overlapping

Vi markerer overlapping på samme måte som innskudd. En overlappingstur lages ved å trykke <Ctrl-t> og velge første taler i skjemaet. Kryss så av for **Overlapping speech** og velg deretter andre taler.

I en slik overlappingstur er taler 1 den som har turen og ytrer replikker, taler 2 kommer med ikke-replikker, se klassifiseringen av *replikker* og *ikke-replikker* i 3.6.1.1 ovenfor.

```
1: jeg husker jeg hadde mistet e jeg mistet så mye når jeg var liten
2: vet det
```

Det er ikke så farlig akkurat hvor taler 2 kommer inn med ikke-replikker i taler 1s tur, men vi markerer at en ikke-replikk kommer litt ut i eller på slutten av taler 1s segment, ved å sette inn en stjerne. Forekommer det flere ikke-replikker med pause mellom fra taler 2 i løpet av taler 1s segment, kan ikke-replikkene skilles med stjerne:

```
1: fordi før så var det sånn man var ikke helt sikker på hvem som egentlig var vennene sine og
2: mm * m
1: hvert fall sånn på ungdomsskolen så tar det i hvert fall et par år før du # merker hvem du liksom
nære da
2: * m
```

Det kan også forekomme en stjerne i taler 1s segment. Dette er i så fall i begynnelsen av segmentet dersom taler 2s ikke-replikk starter før taler 1s replikk. Andre pauser i taler 1s segment markeres som vanlig med #.

All overlapping skal med i samtalene – selv om det er tidkrevende å gjennomføre. I intervjuene har vi imidlertid bestemt oss for å ikke transkribere intervjuerens oppmuntrende *hm*-er og andre små ikke-replikker i løpet av informantens tur. Vi tar selvsagt med reelle replikker fra intervjueren.

3.6.3.4 Delte meningsenheter

Når taler 1 og taler 2 snakker i munnen på hverandre uten at meningsenhetene deres er synkronisert, lar vi taler 1s meningsenheter styre samtalens oppdeling i segmenter, det vil si at taler 1 alltid får ”snakke ferdig”. Taler 2s meningsenheter kan dermed bli delt opp over flere segmenter. For å ta vare på taler 2s meningsenheter, rammer vi dem inn med eventen **pronounce** og **me** (meningsenhet). Det er to hurtigkommandoer for dette: <Alt-m> for start på meningsenhet og <Alt-n> for slutt på meningsenhet.

```
007+008
1: ja for de regner med at det går sin gang
2: [pron=me-] ja det det går
1: jeg hadde skikkelig traumer for at jeg ikke kunne klokken og sånn
2: sikkert [-pron=me]
```

Taler 2s meningsenheter kan også være delt over to turer:

008+007

- 1: sånn i julekalender og sånn fikk sånn brukte ting og sånn
- 2: [pron=me-] jeg husker jeg fikk en sjampo

007+008

- 1: for e farget hår jeg hadde ikke farget hår en gang [latter][-pron=me]
- 2: * [latter]

Dette skjer i tilfeller der taler 2 snakker samtidig med taler 1, og der taler 2 i løpet av meningsenheten overtar turen.

Som vi har sett tidligere (3.6.3.3), brukes stjerne som pausetegn i taler 2s ikke-replikker. Når taler 2s meningsenheter er delt opp over to turer slik at taler 2 overtar turen, kan det derimot forekomme vanlig pausetegn i taler 2s meningsenhet. Dette er fordi det som er inne i meningsenheten, er en replikk:

004+003

- 1: ja nei men hvis du er fornøyd så kan vi godt gjøre det ja da
- 2: * ja * mm * [pron=me-] ja ikke sant # det

003+004

- 1: synes jeg også [-pron=me]
- 2:

3.6.3.5 Skjulte turer

Som vi har sett ovenfor, vil en *tur* bestå av taler 1s replikker og taler 2s ikke-replikker. Når talerne snakker i munnen på hverandre, kan det likevel skje at taler 2 sier noe som taler 1 reagerer eller svarer på. Taler 2 har da i følge definisjonen i 3.6.1.1 kommet med en replikk som egentlig burde hatt sin egen tur. Problemet er imidlertid at replikken blir ytret samtidig med at taler 1 også har en replikk, og at taler 1 svarer taler 2 umiddelbart. Vi har valgt å markere dette ved å lage en ny tur – selv om taler 1 og taler 2 er den samme i begge turene:

004+003

- 1: nei e det er ikke så veldig søtt
- 2: er det ikke det?

004+003

- 1: nei det er sånn man har i brød og særlig utlandet Italia Frankrike og sånn så liker de gjerne å brødet med karvefrø # og det e jeg syns det er helt forferdelig
- 2: trodde det var ... * ja * ja

To identiske turbetegnelser etter hverandre vil dermed bety at taler 2 har hatt en replikk på slutten av den første turen slik som overfor, der *er det ikke det?* blir ytret av taler 2 samtidig med *nei e det er ikke så veldig søtt* av taler 1. Taler 1 svarer umiddelbart på taler 2s spørsmål, og dette er altså skilt ut som egen tur.

4. Korrektur

Filer bør korrekturleses av en annen transkribør enn den som opprinnelig transkriberte fila. Når du skal velge ei fil for korrektur, går du inn i *Transcriber* og velger <Ctrl-o> eller **Open trans.** I *Til Korrektur-mappa* under *Transkripsjon* åpner du den første fila som er ledig for korrekturlesing. Fila skal deretter lagres med nytt navn ved hjelp av **Save as** under **File**. Omdøp fila ved å legge til dine initialer først i filnavnet:

Opprinnelig navn på fila: *KH_samtale_001-002.wav*

Nytt navn på fila: *JBK_KH_samtale_001-002.wav*

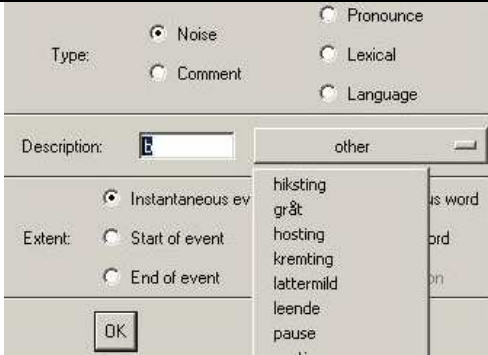
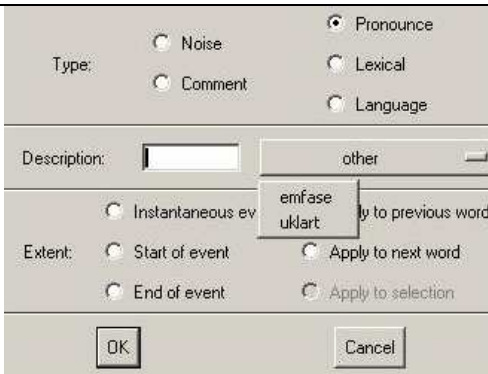
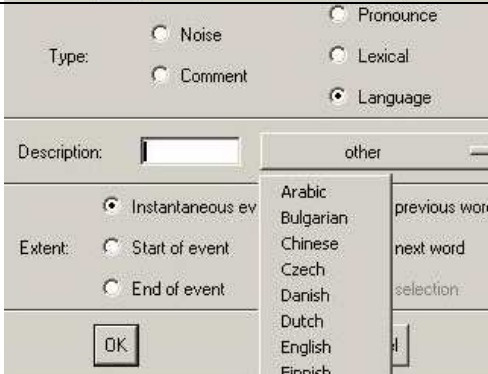
Når du leser korrektur, spiller du av korte sekvenser av gangen. Følg godt med, og spill uklare passasjer om igjen. Finner du feil i transkripsjonen, retter du dem opp. Er du i tvil om noe, setter du inn en problemmerknad (se 3.2). Når du er ferdig med å lese korrektur, legger du fila inn i *Ferdig korrekturlest-mappa*. Dette må du gjøre på skrivebordet når du har gått ut av *Transcriber*. Dersom ei fil inneholder mange utydelige passasjer eller dersom du har endret svært mye i ei fil, legges fila i mappa *Til_andrekorrektur* slik at enda en kan lese korrektur på den.

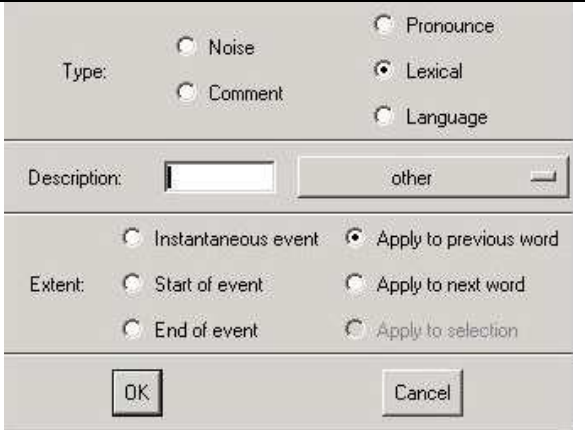
5. Oversikt over hurtigkommandoer, eventer og ikke-språklige lyder

5.1 Hurtigkommandoer

avbrudd	<Alt-a>	{avbrudd}
copy	<Ctrl-c>	
create turn	<Ctrl-t>	
cut	<Ctrl-x>	
delete breakpoint	<Ctrl-Backspace>	
engelsk	<Alt-e>	+ <i>[lang=English]</i>
find/replace	<Ctrl-f>	
innledende bemerkninger	<Alt-i>	{innledende bemerkninger}
insert breakpoint	<Return>	
language	<Ctrl-l>	
latter	<Alt-o>	<i>[latter]</i>
lexical	<Ctrl-e>	
meningsenhet, start	<Alt-m>	<i>[pron=me-]</i>
meningsenhet, slutt	<Alt-n>	<i>[-pron=me]</i>
new trans	<Ctrl-n>	
noise	<Ctrl-d>	
nyord/slang/dialekt ...	<Alt-x>	+ <i>[lang=X]</i>
open trans	<Ctrl-o>	
paste	<Ctrl-v>	
play around cursor	<Alt-Space>	
play/pause	<Tab>	
play segment	<Shift-Tab>	
problem	<Alt-p>	{problem}
pronounce	<Ctrl-r>	
quit	<Ctrl-q>	
save	<Ctrl-s>	
sensitive pers	<Alt-s>	<i>{sekvens med sensitive personopplysninger}</i>
uforståelig	<Alt-u>	{uforståelig}
uklart	<Alt-k>	+ <i>[pron=uklart]</i>
undo	<Ctrl-z>	

5.2 Skjema over eventer

<p>noise</p>	<p><Ctrl-d></p>	<p>Fremre klykkelyd bakre klykkelyd sugelyd labial frikativ labial vibrant sibilant latter gjespende gjesping gråt hosting hviskende kremting leende lydmalende ord nysing pause pusting snufsing stønning syngende sukking trekker pusten</p>	
<p>pronounce</p>	<p><Ctrl-r></p>	<p>høytlesing me stavet uklart</p>	
<p>language</p>	<p><Ctrl-l></p>	<p>arabisk ... engelsk ... svensk ... X</p>	

lexical	<Ctrl-e>	Skriv inn uttalen av pronomener, for eksempel <i>dom</i> , <i>hu</i> , <i>dems</i>	
comment			Brukes bare med hurtigkommandoene <Alt-p>, <Alt-s>, <Alt-u>, <Alt-i> og <Alt-a> for henholdsvis <i>problem</i> , <i>sekvens med sensitive personopplysninger</i> , <i>uforståelig</i> , <i>innledende bemerkninger</i> og <i>avbrudd</i>

5.3 Liste over interjeksjoner

I talemål er det som regel fullt opp av lydmalende ord som *eh, aha, em, ææææ, oiiii* osv. Noen av disse ordene er interjeksjoner som vi finner i Bokmålsordboka (*BMO* i oppstillingen nedenfor), andre er ikke standardiserte. Nedenfor har vi laget vår egen standardiserte liste over interjeksjoner vi skal bruke. Legg merke til at vi ikke markerer lengde på vokalene. To like bokstaver betyr to stavelser. Alle interjeksjonene er satt opp med betydning. Velg den interjeksjonen som ligger nærmest det informantene sier, både når det gjelder uttale og betydning.

Noen ord er bare lydmalende uten spesiell betydning. Dette er ord som rett og slett beskriver lyd, som for eksempel *uuuuuuuu* for ulingen til en sykebil eller *ææææææææ* som skrikingen til en baby. Slike ord skriver vi ikke inn, men karakteriserer dem med eventen **noise** <Ctrl-d> og **lydmalende ord**.

<i>aha</i>	(overrasket) BMO
<i>e</i>	(nøling - uansett lengde på een)
<i>eh</i>	(avstandsindikerende)
<i>ehe</i>	("Jeg forstår" - to stavelser)
<i>em</i>	(nøling)
<i>gud a meg</i>	(overrasket, utrop)
<i>heh</i>	(imponert)
<i>hm</i>	(spørrende, undrende) BMO i betydningen kremting
<i>huff a meg</i>	(beklagelse) BMO
<i>hæ</i>	(spørrende) BMO
<i>jaha</i>	(forsterkende "ja") BMO
<i>m</i>	(nøling, ta til etterretning, nam)
<i>m-m</i>	(benektende)
<i>mhm</i>	("Jeg forstår" - to stavelser)
<i>mm</i>	(bekreftende - to stavelser)
<i>næ</i>	(overrasket, undrende)
<i>nja</i>	(tvilende) BMO
<i>næhei</i>	(forsterkende "nei")
<i>ops</i>	(oi da)
<i>u</i>	(imponert)
<i>uff a meg</i>	(beklagelse) BMO
<i>ææ</i>	(konstaterende - to stavelser)
<i>å-å</i>	("oj")
<i>å ja</i>	(overrasket)

5.4 Huskeliste over spesielle ord

Informanten sier:	Vi transkriberer
atte	at
atte	at e (ved at + nøling)
blei	blei
deluks	de-luxe
denna	denne
detta	dette
én (tallordet)	en
fårse	vorse (fra vorspiel)
ga	ga
guri (interjeksjon)	guri
	hun <i>+[lex=hu]</i>
hu	
herre	herre (den herre bilen)
hvertfall	i hvert fall
jammen (tonem 1)	ja men (= innsigelse)
jammen (tonem 2)	jammen (som i <i>det var jammen bra</i>)
keg	keeg
littegrann	lite grann
no	noe
òg	òg (adverb = også)
ops	obs (= observer/oppmerksom)
ops	ops (= oi da! interjeksjon)
sia	sia
	shit <i>+[lang=English]</i>
shit	
sleng	sleng
sto	sto
tbane	T-bane
tredjeklasse	tredjeklasse
tredje klasse	tredje klasse
vart	vart
viss	hvis

5.5 Kløyvd infinitiv

Verbene nedenfor kan ha a-endelse ved kløyvd infinitiv. * betyr at verbet har to eller flere betydninger, ikke alle har a-ending. Slå opp i Bokmålsordboka!

A	grasa	masa *	svaga
Aga	grava	mata	svala
Aka	gula	P	svara
ala	H	para	T
ana	haka	posa	tala
apa	hata	prata	tana
asa	hela	R	tapa
B	herja	raga *	tela
baka	hesja	raka *	teva
bala	heva	rana	tora *
bana	homma	rapa	trega
basa	hopa	rasa	tørra *
belja	hølja	reka	U
betala	J	remja	umaka
beva	jaga	reva	V
bevara	K	S	vada
bora	kaka	saga	vaka
braka	kara	saka *	vara
bæra	kasa	sela	vasa
D	kava	sitta	vega
daga	kjava	skaka	verja
dala	kjea	skapa	vita
delja	klaka	skara	vima
detta	knaka	skata	vraka
drepa	knasa	skava	væra
dryssa	knisa	skipa	
drønna	komma	skjena *	
E	krasa	skrapa	
eta	krava	skreva	
eva	kreka	skvala	
F	krota	sleva	
fara	kunna	smaka	
fata	kveda	smala *	
flana	L	snaka	
flara	lada	snara	
flata	laga	sova	
flytta	lata *	spara	
freda	lava	speka	
frega	lemma	spora	
G	lepja	spraka	
gala	lesa	spøta	
gana	leva	staga	
gapa	lova *	stava *	
gjøra	M	stega	
glana	maka	st(j)ela	
gleda	mala *	streka	
gnaga	mana *	streva	

6. Referanser

C. Barras, E. Geoffrois, Z. Wu, and M. Liberman . 2000. Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication special issue on Speech Annotation and Corpus Tools*, Vol 33, No 1-2, January.

Du Bois, John, Stepan Schuetze-Coburn, Susanna Cumming, and Danae Paolino. 1993. Discourse Transcription. *Santa Barbara Working papers in Linguistics*, Vol. 4.

Kristoffersen, Gjert og Ann-Kristin Molde. 2000. Norsk Talemålskorpus: Transkripsjonsveiledning. <http://gandalf.hit.uib.no/vol/ntk/Transkripsjonsveiledning.htm>

Linell, Per og Gustavsson, Lennart. 1987. Initiativ och respons. Om dialogens dynamikk, dominans og koherens. SIC 15, Universitetet i Linköping.

MacWhinney, Brian. 2003. CHAT Transcription System Manual <http://chilides.psy.cmu.edu/manuals/CHAT.pdf>

Slangordboka 2005. Kunnskapsforlaget.

Svennevig, Jan, Margareth Sandvik & Wenche Vagle. 1995. *Tilnærminger til tekst. Modeller for språklig tekstanalyse*. Oslo: LNU/Cappelen Akademisk Forlag.

Wangenstein, Boye (red) 2004. *Bokmålsordboka: definisjons- og rettskrivningsordbok*. Oslo: Kunnskapsforlaget

Østby, Andreas Eilert. 2005. *Kebabnorsk ordbok*. Gyldendal.

Bokmålsordboka og Nynorskordboka. <http://www.dokpro.uio.no/ordboksoek.html>

Fra TEI Guidelines: Transcription of Speech <http://www.tei-c.org/Vault/GL/P3/TS.htm>

Transkripsjonsveiledning på *Transcribers* hjemmeside: Manuel du transcripteur <http://www.etca.fr/CTA/gip/Projets/Transcriber/fr/user.html>