

NB Tale - Teknisk dokumentasjon

Dette dokumentet beskriver tekniske aspekter ved **NB tale**. Her finner du detaljer om innhold, organisering og filformater. Kompletterende informasjon finnes i følgende filer som ligger i samme mappe som dokumentet du nå leser:

- [01_NB_Tale_Teknisk.pdf](#)
Dette dokumentet
- [02_NB_Tale_Innleseroversikt.pdf](#)
En oversikt over innlesere og grupperingen av disse i NB Tale.
- [03_NB_Tale_Opptakprosedyrer.pdf](#)
Beskrivelse av prosedyrene som er fulgt i arbeidet med opptak av lydmaterialer til ressursen.
- [04_NB_Tale_Annotering.pdf](#)
Beskrivelse av prosedyrene som har blitt brukt i den manuelle utarbeidelsen av annoteringen.
- [05_NB_Tale_Informantdata.xlsx](#)
Informasjon knyttet til hver enkelt taler.

Innhold

- [Versjonshistorikk](#)
- [NB Tale](#)
 - [Modul 1 og 2: Manuskriptlest tale](#)
 - [Modul 3: Spontan tale](#)
- [Lydopptak](#)
- [Merkefiler](#)
 - [XML](#)
 - [Eksempel: example_1.xml](#)
 - [Eksempel: example_3.xml](#)
 - [Elementer på overordnet nivå](#)
 - [Elementer på ordnivå](#)
 - [Element på segmentalt nivå](#)
 - [Element uten varighet](#)
 - [Trans](#)
 - [Eksempel - example_1.trans](#)
 - [Eksempel - example_3.trans](#)
 - [Konvertering fra XML for modul 3](#)
 - [TextGrid](#)

Versjonshistorikk

2015-12-22

- **Nytt:** Modul 3 ferdigstilt
- **Rettet:** Mikrofonnavn i navngivning av lydopptak. I forrige versjon var mikrofonnavnene forvekslet.
- **Rettet:** Et utvalg lydopptak har uheldigvis blitt tatt opp med feil punktprøvningsfrekvens (44,1 khz). Disse har blitt oppkonvertert til 48khz. Logg over hvilke filer dette gjelder kan finnes sammen med taleopptakene.

2014-11-17

- **Nytt:** Modul 2 ferdigstilt

- **Nytt:** Modul 1 ferdigstilt

NB Tale

NB Tale er en norsk akustisk-fonetisk taledatabase utarbeidet av Lingit AS i samarbeid med NTNU. Taledatabasen inngår i Språkbanken som administreres av Nasjonalbiblioteket og består av 3 moduler:

- Modul 1 - Manuskriptlest tale, førstespråk
- Modul 2 - Manuskriptlest tale, andrespråk
- Modul 3 - Spontan tale

Alle tre modulene består av en samling annoterte lydopptak. Teknisk sett er modul 1 og 2 designet, lest inn og annotert på samme måte. Begge er annotert (tidsmerket) på et fonotypisk detaljnivå. Den spontane talen i modul 3 er ortografisk transkribert og setningssegmentert. I tillegg er et sett ekstralingvistiske trekk, som for eksempel pusting og nøling, tidsmerket.

Modul 1 og 2: Manuskriptlest tale

Modul 1 og 2 er uavhengige, men består begge av manuskriptlest tale med tilhørende tidsmerket fonotypisk annotering. Det er i hver modul benyttet et balansert utvalg av innlesere med tanke på dialektale trekk, alder og kjønn. I tillegg er datasettet delt i to uavhengige deler: 1) treningssett og 2) testsett. Det skal ikke være overlapp mellom disse settene, hverken av talere eller innleste setninger, med unntak av 3 kalibreringssetninger.

Hver taler har lest 20 setninger. Disse er delt inn i tre grupper: A, B og C.

- 3 setninger kommer fra sett A (kalibreringssett) og leses av alle talere.
- 12 setninger kommer fra sett B og leses av tre ulike talere
- 5 setninger kommer fra sett C og er unike for hver taler

Totalt består modul 1 og 2 av 7600 setninger, fordelt på 3423 unike tekster og 380⁽¹⁾ talere:

- Modul 1 består av 4800 innleste setninger fordelt på 2163⁽²⁾ unike tekster og 240 talere.
- Modul 2 består av 2800 innleste setninger fordelt på 1263 unike tekster og 140⁽¹⁾ talere.

⁽¹⁾ Reelt består Modul 2 av 137 av 140 planlagte talere, da tre av de planlagte talerne ikke lot seg oppdrive innenfor prosjektets rammer. Det er dermed totalt 377 talere i modul 1 og 2.

⁽²⁾ 4 setninger i Modul 1 har blitt tilegnet ulike identiteter mens de i realiteten er like i innhold. Det er dermed 2159 unike setninger.

Manuskriptdesign

Setningene som leses i modul 1 og 2 er valgt ut med tanke på å inkludere så mange ulike akustiske hendelser som mulig. Dette er en ikke-triviell oppgave fordi de statistiske egenskapene i språket er slik at det finnes en nesten ubegrenset "hale" med meget sjeldne hendelser. Dette betyr at man, selv med flere timer opptak, vil observere at noen hendelser bare forekommer én gang - eller aldri.

Man kan imidlertid velge setninger på en måte som har til hensikt å løfte halen; altså øke antallet sjeldne hendelser i forhold til de høyfrekvente. I dette prosjektet har en såkalt "grådig søkealgorime" blitt brukt for å plukke et passende antall setninger fra et meget stort tekstkorpus. Uttalebeskrivelser for hver eneste setning beregnes automatisk ved hjelp av en tekst-til-tale-motor og algoritmen velger ut setninger på grunnlag av blant annet estimerte verdier for:

- difonidentitet (gjeldende og neste fonem)
- leksikalsk trykk og tonelag for vokaler
- pauser

Modul 3: Spontan-tale

Modul 3 inneholder spontan tale, lest av de samme talerne som ble benyttet i modul 1 og 2. Hver taler har lest inn i snitt to minutter inklusive pauser, og opptakene er gjort i samme sesjon som de manuskriptleste modulene. Lydfilene har altså samme akustiske miljø som den manuskriptleste talen for samme innleser.

- Modul 3 består av 365¹ lydopptak med transkripsjon

⁽¹⁾ av 380 planlagte innlesninger.

Anonymisering

Noen små segmenter av taleopptakene er anonymisert av personvern hensyn. Disse segmentene er gjort lydløse og tidsmerket eksplisitt.

Lydopptak

Lydopptakene er gjort med 48kHz punktprøvningsfrekvens og 16 biters oppløsning med to mikrofoner:

- Hodemikrofon: Sennheiser HS 2-5-1
- Stormembranmikrofon: Shure KSM 44

Opptakene har tre ulike akustiske miljø knyttet til opptakssted:

- Lydstudio ved NTNU (342 talere)
- Universitetet i Tromsø, UiT (25 talere)
- Universitetet i Agder, UiA (10 talere)

Organisering og navngivning av filer

Filene er organisert i kataloger som identifiserer mikrofon, modul og gruppe/region,

`Speech/{MIC}/part_{PART}/group_{GROUP}/`
der

- **MIC** angir mikrofonmodell, enten `sennheiser` eller `shure`
- **GROUP** angir innlesergruppe, et tosifret tall i et lukket intervall [1,24]
- **PART** angir modulen, et tall fra settet {1,2,3}

For modul 1, (gruppe [1-12]), inneholder hver gruppekatalog til sammen 400 filer fra 20 talere.

For modul 2, (gruppe [13-24]), inneholder gruppe 13 og 24 til sammen 400 filer fra 20 talere, mens hver av de resterende har 200 filer fra 10 talere bortsett fra gruppe 15 som har 140 filer lest av 7 talere.

Navnet på hver lydfil identifiserer modul, gruppe, kjønn, aldersgruppe, subid, talertype, setningsgruppe og setningsidentitet, som illustrert ved følgende regulære uttrykk:

`"p(\d)_g(\d{2})_([mfn])([0-2])_(\d)_([tx])-([abc]\d{4}|free).wav"`
der

- parentes 1 angir "delprosjekt" (med hensyn til innlesing)
 - "1" for modul 1 (og modul 3 med innlesere fra modul 1)
 - "2" for modul 2 (og modul 3 med innlesere fra modul 2)
- parentes 2 angir "gruppe"
- parentes 3 angir "kjønn"
 - "m" for hankjønn
 - "f" for hunkjønn
 - "n" for ikke-spesifisert
- parentes 4 angir "aldersgruppe"
 - "0" for uspesifisert
 - "1" for 16-39 år
 - "2" for 40 år eller eldre
- parentes 5 er en "subid" for taler med gitt gruppe, kjønn og aldersgruppe. For modul 1 er dette et tall i et lukket intervall [1-5]
- parentes 6 angir "talertype"
 - "t" for talere i treningssettet
 - "x" for talere i testsettet
- parentes 7 avhenger av om filen er fra de manuskriptleste modulene (1 og 2) eller spontantalemodulen, 3
 - "a" + "setningsid" for en setning i A-settet.
 - "b" + "setningsid" for en setning i B-settet.
 - "c" + "setningsid" for en setning i C-settet.
 - "free" dersom lydfilen representere spontantale fra modul 3.

Eksempler: filnavngivning

Filen "Speech/sennheiser/part_1/group_11/p1_g11_f2_3_t-b0774.wav" angir:

- mikrofon: sennheiser
- delprosjekt: 1
- gruppe/region: 11
- kjønn: f
- aldersgruppe: 2 (40 år eller eldre)
- subid: 3
- talertype: t (fra treningssettet)
- setningsgruppe: b (leses av tre talere)
- setningsid: 0774

Filen "Speech/sennheiser/part_3/group_11/p1_g11_f2_3_t-free.wav" angir spontantalen (i modul 3) for samme innleser som eksempelet over.

Merkefiler

Annoteringsfiler finnes i katalogen `Annotation` og er tilgjengelig i flere formater:

Modul	XML	Trans	TextGrid
Modul 1	part_1.xml	part_1.trans	TextGrid/part_1/**/*.TextGrid
Modul 2	part_2.xml	part_2.trans	TextGrid/part_2/**/*.TextGrid
Modul 3	part_3.xml	part_3.trans	TextGrid/part_3/**/*.TextGrid

I det følgende beskrives først XML-strukturen, deretter sammenhengen mellom denne og de andre formatene.

XML

XML-filene er i samsvar med definisjonen `annotationdoc.dtd`. Følgende to eksempeldokument viser alle elementer i bruk i henholdsvis modul 1 (samt 2) og 3:

- `Annotation/Example/example_1.xml` (modul 1 og 2)
- `Annotation/Example/example_3.xml` (modul 3)

I det påfølgende vises først (delvis) innholdet i eksempelfilene. Deretter beskrives XML-elementene i nærmere detalj.

Eksempel: example_1.xml

Strukturen for modul 1 og 2 er svært enkel. Det brukes totalt 4 element: `annotationdoc`, `annotation`, `word` og `seg`.

```
<?xml version='1.0' encoding='UTF-8'?>
<annotationdoc>
<annotation id="part_1/group_01/p1_g01_f1_3_t-c1903"
  speaker="p1_g01_f1_3_t"
  text="Sterke meiningar har møyrna."
  end="4.87492063492"
  >
  <!-- ... -->
  <word text="har">
    <seg end="3.0494" label="h" start="2.9535" />
    <seg end="3.1809" label="&quot;A;" start="3.0494" />
    <seg end="3.2268" label="4" start="3.1809" />
  </word>
  <!-- ... -->
</annotation>
<!-- ... -->
</annotationdoc>
```

Eksempel: example_3.xml

Modul 3 er i utgangspunktet ortografisk transkribert og segmentert på setningsnivå. Utover dette har bare ekstralingvistiske trekk, slik som innpust, utpust, pause og nøling blitt tidsmerket. Disse kan forekomme både mellom og inne i ord. På grunn av dette har XML-strukturen for modul 3 blitt noe utvidet (sammenlignet med modul 1 og 2).

Eksempelet under viser alle elementene i bruk.

```
<annotationdoc>
  <annotation id="part_3/AGroup/TheSpeaker-free"
    speaker="TheSpeaker"
    end="5.57"
  >
    <!-- ... -->
    <sil start="0" end="0.8" />
    <inhale start="0.8" end="1.5" />
    <word text="Og" />
    <slang text="e" normalized="jeg" />
    <word text="rop" deadend="1" />
    <vowel start="2.0" end="2.3" />
    <word text="sa" start="2.3" end="2.6"/>
    <fp start="2.6" end="2.8" />
    <nasal start="2.8" end="3.1" />
    <titlestart start="3.1" end="3.1"/>
    <complex_word text="Stopp">
      <word text="S" />
      <sil start="3.4" end="3.5" />
      <word text="topp" />
    </complex_word>
    <word text="der" />
    <titleend/>
    <word text="til" />
    <anonymized start="3.8" end="4.2" />
    <comma text="," />
    <exhale start="4.2" end="4.4" />
    <inhale start="4.4" end="4.8" />
    <word text="nummer" />
    <multiword text="fem og tredve">
      <word text="fem" />
      <word text="og" />
      <vowel start="5.4" end="5.7" />
      <slang text="tredve" normalized="tretti" />
    </multiword>
    <incomprehensible start="6.0" end="6.5" />
    <sentence_boundary text="." start="6.5" end="6.5" />
    <exhale start="6.5" end="6.8" />
    <sil start="6.8" end="7.0" />
  </annotation>
  <!-- ... -->
</annotationdoc>
```

Elementer på overordnet nivå

XML	Oblig.	Beskrivelse	I bruk
<annotationdoc>			
Rot-element. Inneholder ett eller flere <code>annotation</code> -elementer.			
<annotation			
id=ID	ja	Unik identifikator for opptak	part_1 part_2 part_3
speaker=SPEAKER	ja	Unik identifikator for innleser	part_1 part_2 part_3
text=TEXT	ja	tekst assosiert med innlesingen	part_1 part_2
end=END	ja	sluttid i sekunder	part_3
>			
Inneholder ett eller flere av følgende sub-element: <code>word</code> , <code>slang</code> , <code>complex_word</code> , <code>multiword</code> , <code>sil</code> , <code>fp</code> , <code>inhale</code> , <code>exhale</code> , <code>nasal</code> , <code>vowel</code> , <code>incomprehensible</code> , <code>seg</code> , <code>titlestart</code> , <code>titleend</code> , <code>comma</code> , <code>sentence_boundary</code>			

Elementer på ordnivå

Det er definert 4 element som representerer ord (eller fraser):

- `word` er basisrepresentasjonen av et ord
- `slang` beskriver et uttalenært ord med en tilhørende normalisert form
- `complex_word` beskriver et ord som inneholder nøling og som derfor har blitt brutt opp i flere deler
- `multiword` beskriver et flerordsuttrykk, som for eksempel tall. Dette elementet har minst to sub-elementer som beskriver ordene i uttrykket

I utgangspunktet er ikke ordene i modul 3 tidsmerket, men de kan bli det indirekte dersom de har ekstralingvistiske naboelement med eksplisitt tidssegmentering.

Språk er, ved behov, angitt med attributten `lang` med verdier i henhold til ISO 639-1.

Nøling er en viktig del av modul 3 og det kan forekomme på flere måter. Dersom et ord (eller deler av et ord)

blir ytret og senere rettet opp av innleseren på en måte som gjør ordet irrelevant for ytringen, så markeres dette med en attributt `deadend`.

I hovedsak forekommer nøling (ekstralingvistiske element) mellom ord, men det finnes også tilfeller der innleser bryter opp ord med nøling slik at ordet består av en sekvens av elementer. Gruppe-elementet `complex_word` har blitt definert for å samle slike bruddstykker til et komplett ord.

En tilsvarende situasjon har man også i for eksempel tallfraser og uttrykk som utskrevet kan bli en sekvens av ord. Slike tilfeller håndteres av gruppe-elementet `multiword`. Her er ikke nøling nødvendigvis til stede, men kan forekomme.

Under defineres de fire ordelementene:

XML	Oblig.	Beskrivelse	I bruk
<pre><word text=TEXT start=START end=END lang=LANG deadend=1 ></pre>	ja	tekst assosiert med et ord starttid i sekunder sluttid i sekunder språkkode, ISO 639-1 Satt hvis ordet ikke er en del av den fullførte setningen.	part_1 part_2 part_3 part_3 part_3 part_3 part_3
Elementet <code>word</code> er basisrepresentasjonen av et ord.			
<pre><slang text=TEXT normalized=NORMALIZED start=START end=END lang=LANG deadend=1 ></pre>	ja ja	uttalenær tekst assosiert med ordet normalisert tekst assosiert med ordet starttid i sekunder sluttid i sekunder språkkode, ISO 639-1 Satt hvis ordet ikke er en del av den fullførte setningen.	part_3 part_3 part_3 part_3 part_3 part_3
Elementet <code>slang</code> er en utvidet ordbeskrivelse som inneholder en obligatorisk normalisert form.			
<pre><complex_word text=TEXT deadend=1 ></pre>	ja	tekst assosiert med uttrykket Satt hvis ordet ikke er en del av den fullførte setningen.	part_3 part_3
Et <code>complex_word</code> representerer et ord som er lest med nøling. Første og siste sub-element skal være enten <code>word</code> eller <code>slang</code> og det skal også finnes minst ett ekstralingvistisk sub-element: <code>exhale</code> , <code>inhale</code> , <code>fp</code> , <code>nasal</code> , <code>vowel</code> , <code>sil</code>			
<pre><multiword text=TEXT deadend=1 ></pre>	ja	tekst assosiert med uttrykket Satt hvis ordet/frasen ikke er en del av den fullførte setningen.	part_3 part_3
Et <code>multiword</code> representerer et flerordsuttrykk, for eksempel et tall. Første og siste sub-element skal være enten <code>word</code> eller <code>slang</code> og det kan også finnes ekstralingvistiske sub-element: <code>exhale</code> , <code>inhale</code> , <code>fp</code> , <code>nasal</code> , <code>vowel</code> , <code>sil</code>			

Element på segmentalt nivå

I modul 3 er det brukt 8 element som har eksplisitt (obligatorisk) segmentering (med varighet): `anonymized`, `exhale`, `inhale`, `fp`, `nasal`, `vowel`, `sil` og `incomprehensible`. Modul 1 har bare benyttet ett element på segmentalt nivå, `seg`, men kan representere samme informasjon ved spesielle `label`-attributtverdier.

XML	Oblig.	Beskrivelse	I bruk
<pre><anonymized start=START end=END /></pre>	ja ja	starttid i sekunder sluttid i sekunder	part_3 part_3

En anonymisert region			
<nasal			
start=START	ja	starttid i sekunder	part_3
end=END	ja	sluttid i sekunder	part_3
/>			
Nasal nølelyd			
<vowel			
start=START	ja	starttid i sekunder	part_3
end=END	ja	sluttid i sekunder	part_3
/>			
Vokalisk nølelyd			
<fp			
start=START	ja	starttid i sekunder	part_3
end=END	ja	sluttid i sekunder	part_3
/>			
Fylt pause			
<inhale			
start=START	ja	starttid i sekunder	part_3
end=END	ja	sluttid i sekunder	part_3
/>			
Innpust			
<exhale			
start=START	ja	starttid i sekunder	part_3
end=END	ja	sluttid i sekunder	part_3
/>			
Utpust			
<sil			
start=START	ja	starttid i sekunder	part_3
end=END	ja	sluttid i sekunder	part_3
/>			
Stillhet			
<incomprehensible			
start=START	ja	starttid i sekunder	part_3
end=END	ja	sluttid i sekunder	part_3
/>			
Uforståelig segment			
<seg			
label=LABEL	ja	X-Sampa-notasjon (eller spesialsymbol)	part_1 part_2
start=START	ja	starttid i sekunder	part_1 part_2
end=END	ja	sluttid i sekunder	part_1 part_2
>			
Et tidsmerket segment med tilhørende X-Sampa-notasjon. Dette brukes bare i modul 1 og 2.			

Element uten varighet

Det finnes noen få strukturelle element som ikke har noen direkte tilknytning til varighet. I utgangspunktet har `sentence_boundary` obligatorisk tidsmarkering, men det finnes noen få unntak i forbindelse med sitater. Legg merke til at `start`- og `end`-attributter må, dersom angitt, ha samme verdi.

XML	Oblig.	Beskrivelse	I bruk
<comma			
text=TEXT	ja	tekst assosiert med elementet (',' eller '--')	part_3
/>			
Et tekstlig komma, satt med en "muntlig" tilnærming.			
<sentence_boundary			
text=TEXT	ja	tekst assosiert med elementet ('!', '...', ':', '?')	part_3

```
start=START          starttid i sekunder
end=END              sluttid i sekunder
/>
```

part_3
part_3

Setningsgrense, satt med en "muntlig" tilnærming.

```
<titlestart/>
```

Starten på et sitat eller tittel. Skal samsvare med en tilhørende <titleend/>.

```
<titleend/>
```

Slutten på et sitat eller tittel. Skal samsvare med en tilhørende <titlestart/>.

Eksempel

```
<titlestart />
<word text="OK" />
<titleend />
<word text="sier" />
<word text="jeg" />
<sentence_boundary text="." start="2.4" end="2.4" />
```

Trans (en variant av HTK-MLF)

En Transcription-fil (.trans) er en variant av Master Label File (MLF) der tidspunkt er angitt i millisekund (i stedet for 100-nanosekund). Filen bruker UTF-8 tegnkoding og definerer et sett med transkripsjoner:

Hver transkripsjon begynner med en unik identifikator i anførselstegn. Deretter følger en liste med tab-separerte informasjonsfelter:

1. starttid i millisekund (eller en _ hvis ikke tilgjengelig)
2. sluttid i millisekund (eller en _ hvis ikke tilgjengelig)
3. Merkelapp på første nivå
4. Merkelapp på andre nivå. Ikke obligatorisk. Man kan ha så mange nivåer man ønsker.

Transkripsjonen avsluttes av en linje med et enkeltstående punktum

I dette prosjektet er trans-filene generert ut fra innholdet i tilsvarende XML-filer. Det er lagt ved et transkripsjonsskript som viser hvordan dette kan gjøres:

```
Annotation/Example/example_xml_to-trans.xslt
```

Man kan f.eks. bruke verktøyet `xsltproc` for å gjøre konverteringen.

Eksempel - example_1.trans

```
#!TRANSCRIPTION!#
"part_1/group_01/p1_g01_f1_3_t-c1903"
0      1840  <sil>  <start>
1840   1945  s      Sterke
1945   2000  t
2000   2095  ""{
2095   2179  4
2179   2305  k
2305   2348  @
2348   2435  m      meiningar
2435   2577  ""{i
2577   2631  n
2631   2681  i
2681   2750  N
2750   2889  A
2889   2954  4
2954   3049  h      har
3049   3181  "A:
3181   3227  4
3227   3324  m      mørrna
3324   3478  ""2y
3478   3579  4
3579   3646  n
3646   3781  A
3781   4875  <sil>  <end>
.
```

Modul 1 er fullsegmentert og tidsinformasjonen på hver linje i transkripsjonen er hentet fra `start-` og `end-` attributtene i et `seg`-element. Merkelappene på første nivå er hentet fra attributten `label`. For første segment i hvert ord finner man også merkelapper på nivå 2. Disse er ordstrenger hentet fra `text`-attributten i det tilhørende `word`-elementet.

Eksempel - example_3.trans

```

#!TRANSCRIPTION!#
"part_3/AGroup/TheSpeaker-free"
0      800 <sil>                <sil>
800    1500 <inhale>            <inhale>
-      -      Og                Og
-      -      e;normalized=jeg    e;normalized=jeg
-      -      rop;deadend=1       rop;deadend=1
-      -      <vowel>            <vowel>
2000   2300 sa                  sa
2300   2600 <fp>                <fp>
2600   2800 <nasal>             <nasal>
3100   3100 <titlestart>        <titlestart>
-      -      S                  Stopp;tag=complex_word
3400   3500 <sil>                <sil>
-      -      topp               der
-      -      der                <titleend>
-      -      <titleend>         til
-      -      til                <anonymized>
3800   4200 <anonymized>        <comma>
-      -      <comma>           <exhale>
4200   4400 <exhale>           <inhale>
4400   4800 <inhale>           nummer
-      -      nummer            fem_og_tredve;tag=multiword
-      -      fem                og
-      -      og                 <vowel>
5400   5700 <vowel>            tredve;normalized=tretti
-      -      tredve;normalized=tretti
6000   6500 <incomprehensible> <incomprehensible>
6500   6500 <sentence_boundary> <sentence_boundary>
6500   6800 <exhale>           <exhale>
6800   7000 <sil>              <sil>
.

```

Konvertering fra XML for modul 3

Eksempelen over viser hvordan de ulike elementene i `example_3.xml` er kodet. For modul 3 er det første merkenivået tilnærmet ordnivå, mens merkelappene på nivå 2 er avhengig av om "ordet" på nivå 1 er selvstendig eller inngår i en større sammenheng. Under vises eksempler på oversettelse fra xml- til trans-representasjon:

De fleste ord står selvstendig og har identiske merkelapper på nivå 1 og 2. Et frittstående `word`-element, som vist under har en merkelapp lik verdien av `text`-attributten.

XML		Trans
<code><word text="Og" /></code>		<code>_ _ Og Og</code>

Dersom et element har flere attributter brukes et semikolon for å skille disse. Elementet `slang`, som har både en `text`- og en `normalized`-attributt, får følgende konvertering:

XML		Trans
<code><slang text="e" normalized="jeg" /></code>		<code>_ _ e;normalized=jeg e;normalized=jeg</code>

Lignende, dersom `deadend`-attributten er aktiv:

XML		Trans
<code><word text="rop" deadend="1" /></code>		<code>_ _ rop;deadend=1 rop;deadend=1</code>

Ekstralingvistiske element får en merkelapp som er lik elementnavnet - innesluttet i vinkelparenteser:

XML		Trans
<code><nasal start="2.8" end="3.1" /></code>		<code>2800 3100 <nasal> <nasal></code>

Spesialelementene som *ikke* har noen fysisk varighet får også merkelapper lik elementnavnet i vinkelparenteser

XML		Trans
<code><titlestart/></code>		<code>_ _ <titleend> <titleend></code>

Som vist over har de fleste element identisk merkelapp på nivå 1 og 2. Dette gjelder ikke gruppe-elementene `complex_word` og `multiword` som bare har verdi på nivå 2 for første del-element (første ord i gruppen). Denne merkelappen er en tekstlig representasjon av hele gruppe-elementet. De resterende del-elementene i gruppen har ingen verdi for nivå 2.

Merkelappen på nivå 2 er satt sammen av to deler. Den første delen er lik verdien av `text`-attributten. En asterisk plasseres i forkant av teksten dersom `deadend`-attributten er satt. (Dette vises ikke i eksempelet.). Dette etterfølges av et semikolon og navnet på gruppe-elementet, som vist under. Legg merke til at "ordene" på nivå 1 ikke er selvstendige, de er bruddstykker delt av en segmentert stillhet:

XML	Trans
<code><complex_word text="Stopp"></code>	
<code><word text="S" /></code>	— — S Stopp;tag=complex_word
<code><sil start="3.4" end="3.5" /></code>	— — <sil>
<code><word text="topp" /></code>	— — topp
<code></complex_word></code>	

Mens et `complex_word` representerer et sammenhengende ord består et `multiword` av flere ord, en frase, for eksempel en utskrevet tallfrase. Elementene kodes i trans-filen på samme måte og, som vist under, mellomrom er på nivå 2 konvertert til understrek (`_`):

XML	Trans
<code><multiword text="fem</code>	
<code>og tredve"></code>	
<code><word text="fem" /></code>	— — fem
<code><word text="og" /></code>	— — og
<code><vowel start="5.4"</code>	— — fem_og_tredve;tag=multiword
<code>end="5.7" /></code>	— — og
<code><slang</code>	5400 5700 <vowel>
<code>text="tredve"</code>	— — tredve;normalized=tretti
<code>normalized="tretti"</code>	
<code>/></code>	
<code></multiword></code>	

TextGrid

TextGrid er et filformat som brukes av Praat. I dette prosjektet er filene generert ut fra innholdet i XML-dokumentene. Hver fil tilsvarer innholdet i ett `annotation`-element og er navngitt på samme måte som de assosierte lydfilene, i henhold til `id`-attributten;