

## Annotering: tegninventar og retningslinjer

### Modul 1 og 2

Koden som er brukt er X-SAMPA, og inventaret er inspirert av det som ble brukt i TABU.0 (Amdal og Ljøen 1995).

#### Fonotypisk inventar

Symbol	Eksempelord	Transkripsjon	Symbol	Eksempelord	Transkripsjon
<b>Vokaler</b>			<b>Plosiver</b>		
A	latt	l" At	p	pil	p"i:l
A:	lat	l" A:t	b	bil	b"i:l
e	lett	l" et	t	til	t"il
e:	let	l" e:t	d	dill	d"il
i	litt	l" it	k	krill	k4"il
i:	lit	l" i:t	g	grill	g4"il
u	bukk	b"uk	t`	erte	""{t`@
u:	bok	b"u:k	d`	ferdig	f""{d`i
}	lund	l"}n	? <sup>1</sup>	eige aksjeselskap	""{ig@ ? "AkS@selsk%A:p
}::	lun	l"}:n	c	ikke	"ic
y	lynne	l""yn@	J\	kladd	kr""AJ\
y:	lyne	l""y:n@			
{	verre	v""{4@	<b>Frikativer og approksimanter</b>		
{:	være	v""{:4@	f	fin	f"i:n
2	søtt	s"2t	v	vin	v"i:n
2:	søt	s"2:t	w	wine (eng)	w"Ain
O	vott	v"Ot	s	sal	s"A:l
O:	våt	v"O:t	S	sjal	S"A:l
@ <sup>2</sup>	være	v""{:4@	C	kjøp	C"2:p
I	thin (eng)	T"ln	j	gi	j"i:
U	foot (eng)	f"Ut	h	hi	h"i:
3:	world (eng)	W"3:ld	R <sup>3</sup>	rød	R"2:
V	but (eng)	B"Vt	r\	rød (f.eks. Lofoten/Veste rålen), red (eng)	r\ "2:, r\ "ed
<b>Diftonger</b>			z	freeze (eng)	fr\ "i:z
{i	sein	s" {in	Z	vision (eng)	v" iZn _=
el	sein (trøndersk / nordnorsk); late (eng)	s" eln, l" elt	c__C <sup>4</sup>	kjole (sunnmørsrk)	c__C""u:l@

<sup>1</sup> Glottalt lukke [ʔ] forekommer typisk mellom vokaler.

<sup>2</sup> Trykklett e (schwa)

<sup>3</sup> Vanlig bakre r / skarre-r

<sup>4</sup> Affrikat, typisk i sunnmørske dialekter.

<b>Forts. Diftonger</b>			<b>Forts. Frikativer og approksimanter</b>		
A}	sau	s" A}	t_ S <sup>5</sup>	church (eng)	t_ S"2:t_ S
aU	skåla (sogn); mouth (eng)	sk" aUIA, m" aUT	d_ Z <sup>6</sup>	George (eng)	"d_ ZO:d_ Z
2}	sau (trøndersk)	s"2}	T	thin (eng)	"TIn
Ai	kai	k" Ai	D	this (eng)	"DIs
2y	øy	"2y	x	ach (tysk)	"Ax
Oy	koie	k"" Oy@			
ji	hui	h"ji			
@U	show (eng)	S"@U			
ui	hoi	h "ui	<b>Nasaler, lateraler, rullelyd og flapp</b>		
			m	mål	"mO:r`
			n	nål	"nO:r`
<b>Diakritisk tegn (ikke eget segment)</b>			J	mann (trønder / nordnorsk)	"mAJ
_ =	ballen	b" Aln_ =, b" A:l_ =	N	sang	"sAN
<b>Prosodi (ikke egne segmenter)</b>			l	lat	"lA:t
"	bønder	b"2n@4	L	alle (trønder / nordnorsk)	""AL@, "AL
""	bønner	b""2n@4	4 <sup>7</sup>	rød	"42:
:	håpet	h"O:p@	r <sup>8</sup>	blå	"br`O:
%	bønnestengel	b""2n@st% eN@l	n`	hjørne	""j2:n`@
			l`	herlig	""h{:l`i

## Forklaringer

Alle tegnene under har det til felles at de ikke kan stå som egne segmenter, men må knyttes til et lydskrifttegn

[ " ]: hovedtrykk, tonelag 1. Tegnet plasseres til venstre for vokalen i den aktuelle stavelsen.

[ "" ]: hovedtrykk, tonelag 2. Tegnet plasseres til venstre for vokalen i den aktuelle stavelsen.

[ % ]: bitrykk (dvs. øvrige trykksterke stavelser i et ord, som kommer i tillegg til hovedtrykkstavelsen). Tegnet plasseres til venstre for vokalen i den aktuelle stavelsen.

[ : ]: lengde. Kun trykksterke vokaler som er lange skal ha lengdetegn. Konsonanter får ikke lengdetegn. Eksempel: "hoppet" transkriberes [h"Op@], mens "håpet" transkriberes [h"O:p@]

[ \_ = ]: Stavelsesbærende konsonant. Brukes når en vokal har forsvunnet, men stavelsen fortsatt er til stede.

<sup>5</sup> Engelsk affrikat.

<sup>6</sup> Engelsk affrikat.

<sup>7</sup> Vanlig tungespiss-r

<sup>8</sup> Tjukk l

### Ikke-lingvistisk lyd

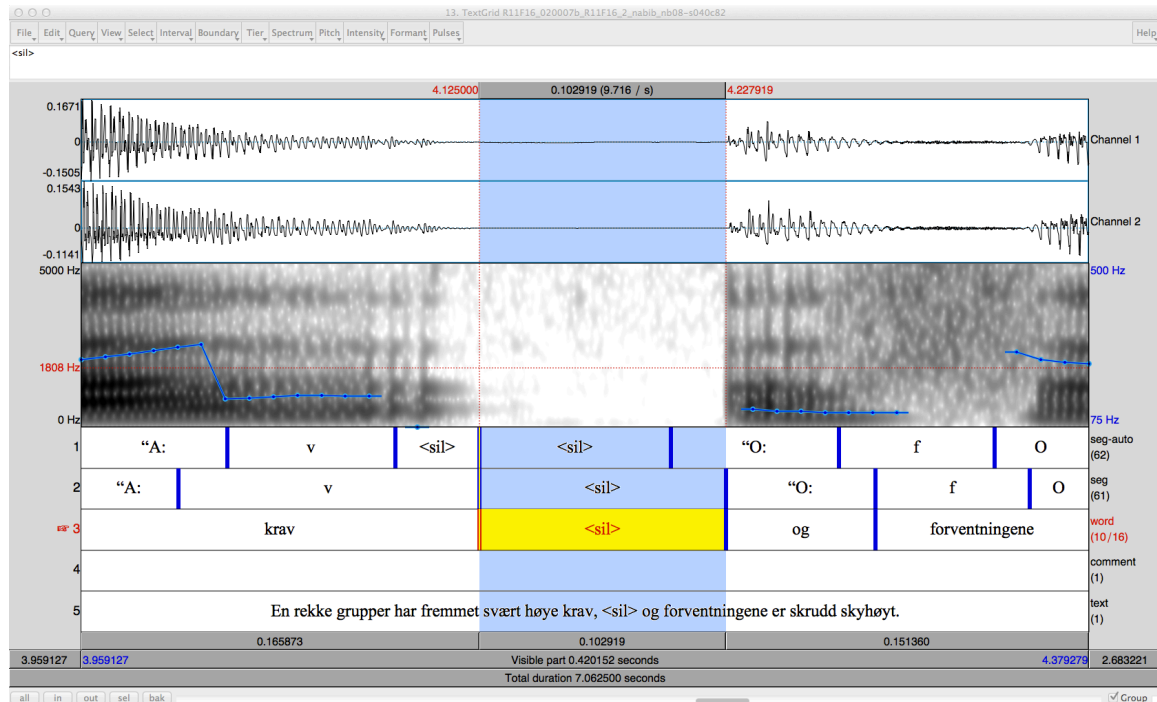
Talestøy skal merkes på samme måte som de lingvistiske lydene. Alle tegn for ikke-lingvistiske lyder skal stå i vinkelparenteser. Disse kodene skal inn på følgende tre nivåer:

- (2) ”seg”: som egne segmenter
- (3) ”word”: som egne segmenter
- (5) ”text”: koden settes inn i teksten på riktig sted, men uten at det defineres segmentgrenser

Fenomen	Transkripsjon	Kommentar
Stille pause	<sil>	Minimum 40 ms varighet Nivå 5: <sil> ved kommagrense plasseres til høyre for komma
Innpust	<inhale>	Nivå 5: <inhale> ved kommagrense plasseres til høyre for komma
Utpust	<exhale>	Nivå 5: <exhale> ved kommagrense plasseres til venstre for komma
Vokalisk nøling	<vowel>	
Nasal nøling	<nasal>	
Uspesifisert lyd fra innleser (latter, smatting, hosting osv)	<fp>	
<i>Før setningsstart</i>	<START>	<i>Merkes kun på nivå ”word”</i>
<i>Etter setningsslutt</i>	<END>	<i>Merkes kun på nivå ”word”</i>

### Ord-nivå (word)

Segmenter på ordnivå som ikke er vanlige ord, kan enten fjernes i sin helhet eller erstattes med ikke-lingvistiske koder som <sil>, <inhale> osv. (se over). Eksempler på segmentkoder vi ønsker at erstattes eller fjernes, er komma ( , ), tankestrek ( - ) osv. (Disse segmentene har alle koden <sil> på segmentnivå (nivå 1 og 2)). Eksempel på tilfelle der komma (,) på nivå 3 er erstattet med <sil>. <sil> er også satt inn på nivå 5.



Ordsegmenter som består av flere ord skal deles opp i enkeltord. Eksempler er preposisjonsfraser som "inn i", som altså skal deles opp i to separate segmenter.

### Prosodisk annotering

Alle ord i taledatabasen er i utgangspunktet annotert med hovedtrykk og tilhørende tonelag. Mange ord kan uttales med ulik trykkplassering og/eller ulikt tonelag avhengig av f.eks. dialektale sosiolektale eller idiolektale preferanser. Den manuelle annoteringen har etter beste evne beskrevet faktisk realisert trykkplassering og tonelag i de enkelte tilfeller.

Videre finnes det et lite antall enstavingsord i materialet som har blitt betydelig redusert (vokalbortfall eller kun schwa).

## Modul 3

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. Det ekstralingvistiske tegninventaret er beskrevet i <01\_NB\_Tale\_Teknisk.pdf>, side 5-7, med eksempel på transkripsjonsfil på side 8-9 i samme fil.

### Merking når uttale avviker mye fra ortografi

En del uttaler av ord avviker mye fra det bokmålske uttrykket til ordene. Vi gir definerte typer av slike avvik en uttalenær gjengivelse, deretter ordet på bokmål i parentes.

Typer avvik som gjengis uttalenært er:

- 1) *Sterkt dialektale ord som i stor grad men på ulike måter skiller seg fonetisk fra bokmålsord med tilsvarende semantisk innhold....*
  - a. "høyveirsbikåri" (øyenstikker)
  - b. "hærle" (ganske)
  - c. "likar" (bedre)
- 2) *Ord uttalt med ulikt antall stavelser og ulik rotvokal:*
  - a. "held" (holder)
  - b. "kjem" (kommer)
  - c. "slepp" (slipper)
  - d. "søv" (sover)
  - e. "fær" (drar)
  - f. "gjenge" (går)
  - g. "vore"/"vorre"/"vori"/"vorri"(vært)
- 3) *Ord uttalt med ulik rotvokal og minst ett annet uttaleavvik fra ortografi:*
  - a. "dreg" (drar)
  - b. "sjå" (se)
  - c. "dokter" (dere)
  - d. "drivi" (drevet)
  - e. "nokke"/"nokka" (noe)
- 4) *Ord uttalt med ulikt antall stavelser, men lik rotvokal:*
  - a. "itj" (ikke)
  - b. "mytj" (mye)
- 5) *Ord uttalt med ulikt første fonem og ulik rotvokal:*
  - a. "kallas" (hvilken)
  - b. "kjeme" (kommer)
  - c. "kossen" (hvordan el. hvilken)
- 6) *Ord som ikke finnes i bokmålsordboken, men som enten er oppslagsord i nynorskordboken, eller har falt ut av bokmålsordboken pga rettskrivingsreformer o.l.*

- a. "oppteken" (opptatt)
- b. "førebygg" (forberedt)
- c. "verte" (bli)
- d. "tyve"(tjue)
- e. to\_og\_"tyve" (tjue)

### Eksempler på uttaleavvik som ikke merkes spesielt

Det er ikke praktisk mulig å lage en ortografisk gjengivelse som er både 100% lydnær og konsistent. En del typer uttaleavvik er følgelig ikke spesifikt merket, men kun gjengitt i bokmålsform. Eksempler på slike tilfeller:

- vanlig apokope (ordfinale stavelsebortfall)
  - ligger uttalt "ligg"

### Referanser

Amdal, I. og Ljøen, H. (1995). TABU.0 - en norsk telefonspråkdatabase. FoU R 40/95. Telenor FoU

X-SAMPA: <http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm>