

Annotationer i LANCHART-korpusset

Philip Diderichsen, Torben Juel Jensen, Marie Maegaard, Janus Spindler Møller, Natalie Carmen Hau Sørensen, Peter Agerlin Trolle, Aviaja Klarskov Skotte-Nielsen.

Sprogforandringscentret, primo 2023

Senest opdateret 21.3.2023

Dette er en oversigt over annotationerne i LANCHART-korpusset til brug for alle der har brug for ret detaljeret information om hvad annotationerne betyder.

Der er to hovedafsnit i oversigten: Tiers og tags samt Metadata. Tiers og tags-afsnittet beskriver de annotationslag (tiers) der er i korpusset. Metadata-afsnittet beskriver de samtale- og informantmetadata fra informantbasen der også indgår som søgbare kategorier i korpusset.

Se afsnit 1 nedenfor for en kort beskrivelse af vigtige grundlæggende begreber og forhold.

Indholdsfortegnelse

1 Vigtige grundlæggende begreber og forhold.....	2
2 Tiers og tags.....	3
2.1 Overordnede tiers	3
2.2 IIV-tiers	20
2.3 Fonetiktiers.....	26
2.4 Grammatiktiers.....	34
2.5 Dialekttiers	61
2.6 Køgeprojektet.....	66
2.7 Sprogholdningstiers.....	67
2.8 Hanne Sæderup	68
2.9 Vollsmose ... (LaPUR).....	70
2.10 AMDA (Amerikadansk: AmDa, ArgDa, CanDa)	75
2.11 Ukendte tiers.....	79
2.12 Tiers der er skjult.....	80
2.13 Tiers der er slettet.....	82
3 Metadata	84
3.1 Særlige metadata – segmentering og varighed.....	84
3.2 Traditionelle metadata.....	84

1 Vigtige grundlæggende begreber og forhold

Tiers, tokens og tags

Et **tier** er et annotationslag hvor forskerne har opmærket udskrifterne med fonetiske, grammatiske og andre kategorier. Annotationerne er typisk lavet i det fonetiske annotations- og analyseprogram Praat, hvor et annotationslag hedder et tier. Et ord kaldes i denne sammenhæng et **token**. Tokens kan i forskellige tiers annoteres med forskellige kategorier kaldet **tags**. Når ét token annoteres med ét tag, kaldes det **ordkodning**. Når et tag spænder over flere tokens, kaldes det **sekvenskodning**.

Tiernavne i parentes

Hvert tier er beskrevet under en brugervenlig betegnelse der også vises i korpuset. Hvis den er forskellig fra Praat-tiernavnet, står Praat-navnet i parentes. Eksempel: "Simpel ordklasse (RedPoS)". Det er også navnet i parentes der vil gå igen i kolonneoverskrifterne i Korps dataeksport.

BIO-opmærkning

Sekvenskodede Praat-tiers er omkodet til en variant af det såkaldte BIO-format ('Begin, In, Out', ofte også kaldet IOB-format), der gør at man kan finde første, sidste og evt. mellemliggende tokens i en given annoteret sekvens. I praksis opnås dette vha. de følgende præfikser og suffikser:

- Præfiks: Løbenummer (1_, 2_, 3_ osv.)
- Suffiks: "_I" for 'in', "_E" for 'end'.

Når eksempelvis en sekvens af tre tokens oprindeligt er annoteret med kategorien *dialektcitat*, vil de tre tokens være annoteret således: 1_dialektcitat_I, 2_dialektcitat_I, 3_dialektcitat_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

Tomme værdier

Hvis et token mangler en annotation i et givet tier (det forekommer ret hyppigt), vil det have den søgbare værdi "__UNDEF__" (for 'undefined'). Hvis man fx vil specificere at et givet ord IKKE er annoteret med noget i tieret "AN-markering", skal man lave en søgning på tokens hvor "AN-markering" er "__UNDEF__".

Underscore erstatter mellemrum

Mellemrum i annotationer giver fejl i korpusdatabasen. Mellemrum i de originale Praat-annotationer er derfor erstattet af underscore. Fx vil generisk pronomen-koden "G AS" være erstattet af "G_AS".

Statusniveau

Vi arbejder med tre forskellige statusniveauer for tiers:

- **Åben** betyder at et givet tier må lægges helt åbent ud uden passwordbeskyttelse.
- **Forskerniveau** betyder at et givet tier kan gøres tilgængeligt på lanchartkorp.ku.dk, som er passwordbeskyttet under individuelle brugernavne og passwords.
- **Privat** betyder at et givet tier er individuelt beskyttet af et 'projektpassword' som kun gives til medarbejdere inden for et givet projekt.

2 Tiers og tags

Her dokumenteres alle annotationslag (tiers) i LANCHART-korpusset. LANCHART-korpusset er delt op i et antal projektkorpusser med navne som LANCHART_AMAGER, LANCHART_BYSOC osv. Nogle tiers findes i alle projektkorpusser, andre kun i nogle af projektkorpusserne. Til den første gruppe hører de "overordnede tiers" der er beskrevet nedenfor.

2.1 Overordnede tiers

- Kontaktperson: Torben Juel Jensen, tjuelj@hum.ku.dk
- Status: Åben. Alle tiers er åbne for alle.

2.1.1 Ordform (ortografi)

- Ordkodet

Udskrift i almindelig ortografi af det, som informanter, interviewere og forbipasserende siger i interviewet/gruppesamtalen/selvoptagelsen etc. Der findes et tier for hver taler i optagelsen (informantkoden står i parentes efter 'ortografi').

For yderligere oplysninger henvises til artiklen Udskrivningsmanual og selve udskrivningsmanualen, der altid ligger i sin nyeste version på DGCSS' hjemmeside under Publikationer\Manualer og rapporter\Manualer.

2.1.2 Kommentarer

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Dette tier indeholder ikke-talerspecifikke kommentarer. OBS: Indeholder til dels registreringer af fejl.

Eksempel på sekvenskodning: Når eksempelvis en token er annoteret med "Spring i optagelsen", vil dette token få følgende tag: 1_Spring_i_optagelsen_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.1.3 Comments

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Talerspecifikke kommentarer - tieret er knyttet til en specifik taler. I det oprindelige tiernavn i Praatfilerne (Textgrids) står talerkoden i parentes efter "Comments".

Eksempel på sekvenskodning: Når eksempelvis en token er annoteret med "siges leende", vil dette token få følgende tag: 1_siges_leende_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.1.4 Usikker transskription (uncertain transcription)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Tieret er knyttet til en specifik taler - talerkoden står i parentes efter 'uncertain transcription'.

Tieret genereres automatisk og laver boundaries ved de ord i ortografien, der er udskrevet som tvivlsomme. I Praat-filen markeres de med '?'.
I BIO-opmærkning ser en sekvens af usikker transskription sådan ud: "1_?_I 2_?_I 3_?_E".

I nogle tilfælde er det dog en række enkeltord efter hinanden der er markeret som tvivlsomme. I sådanne tilfælde ser det sådan ud: "1_?_E 1_?_E 1_?_E".

2.1.5 IPA-lydskrift (real_IPA)

- Ordkodet

Annotationslaget IPA-lydskrift indeholder en translitterering til IPA-tegn (underliggende repræsenteret som numeriske HTML-koder) af Praat-IPA-koderne i tieret IPA. Annotationerne forekommer ikke som et tier i de underliggende TextGrids, kun i Korp og partiturvisevisningen som en hjælp til at tyde de ellers noget obskure lydskriftkoder.

2.1.6 IPA

- Ordkodet

Dette tier indeholder en automatisk genereret lydskrivning af ordene i tieret *Ortografi* med koder der af Praat kan fortolkes som IPA-lydskrifttegn. Der er et IPA-tier for hver taler. Lydskriften er genereret på baggrund af en standardudtale af et givet ord svarende til hvad man ville kunne finde i en ordbog. Det betyder at lydskriften ikke nødvendigvis stemmer overens med lydsiden.

Praat-IPA-koderne kan ses i kolonnen "Kode i Praat-dokumentation" i tabellen nedenfor. De er i nogle tilfælde dog blevet modificeret i korpus (jf. "Praat-kode i korpus"-kolonnen). De tilfælde, hvor Praat-koden er modificeret, er markeret med fed i skemaerne nedenfor.¹

¹ For yderligere information om IPA- og TtT-notationen i Korpus henvises der til manualen for Sprogforandringscentrets program Phonix, som er en ordklasse- og lydskrifttagger. Phonix-manualen er ikke offentligt tilgængelig, men kan udleveres ved efterspørgsel. Vær i så fald opmærksom på, at der er fejl i nogle af Phonix-manualens tabeller ift. notation af Praat-koder jf. oversigten nedenfor.

Fejl i Phonix-manualen	Fejl	Det korrekte
Tabel 1	[ø] skrives som \yc, men denne praatkode repræsenterer lyden [ɣ].	Praatkoden for [ø] er \o/ (i Korpus er den \o\, se skema nedenfor).
	[œ] skrives som \o/, men denne praatkode repræsenterer lyden [ø].	Praatkoden for [œ] er \oe.
	[œ] skrives som \oe, men denne praatkode repræsenterer lyden [œ].	Praatkoden for [œ] er \Oe.

En samlet oversigt over Praats IPA-koder kan findes her: https://www.fon.hum.uva.nl/praat/manual/Phonetic_symbols.html

Vokaler

IPA	Kode i Praat-dokumentation	Praat-kode i korpus	TtT	Eksempel
i	i	i	i	vi, hvile, hvil
e	e	e	e	det, dele, del
ɛ	\ef	\ef	E	gæs, næse
æ	\ae	\ae	z	være, vane
a	a	a	a	vand, mand
ɑ	\as	\as	A	bark, bak
ɣ	ɣ	ɣ	ɣ	lys, tys, nyse
ø	\o/	\o\ (kan ikke vises i Praat)	q	øl, pølse
œ	\oe	\oe	Q	søn, høne, løn
œ	\Oe	\Oe	x	grøn, trøje
u	u	u	u	guld, gule, fuld
o	o	o	o	tone, rod
ɔ	\ct	\ct	c	kul, kål, kåbe
ʌ	\vt	\vt	C	tom, som
ɒ	\ab	O	X	vores, bære
ə	\sw	\sw	0	-e ("schwa")

Konsonanter

IPA	Praat-kode i dokumentation	Praat-kode i korpus	TtT	Eksempel
b	b	b	b	by, op
d	d	d	d	dø, at
g	\gs	g	g	gå, tak

	[ə] skrives som \ic, men denne praatkode repræsenterer lyden [ɪ].	Praatkoden for [ə] er \sw.
Tabel 2	[ç] skrives som \sh, men denne praatkode repræsenterer lyden [ʃ].	Praatkoden for [ç] er \cc.
Tabel 3	Der står /cc.	Der burde stå \cc.
	Der står /rc.	Der burde stå \rc.

v	v	v	v	ve, ulv
f	f	f	f	fe, hof
s	s	s	s	sø, es
ç	\cc	\cc	S	sjat, tusch
h	h	h	h	hest, hø
m	m	m	m	mand, mus
n	n	n	n	næse, nisse
l	l	l	l	lå, æsel
ʀ	\ri	\rc (bliver til r i Praat)	r	rat, rå
p	p	p	p	på, pølse
t	t	t	t	tå, taske
k	k	k	k	ko, kat
j	j	j	J	jul, haj
ʀ*	\rc	\ri (bliver til ʀ i Praat)	R	mor, ser
w	w	w	w	hav, av
ð	\dh	\dh	D	hid, tid
ŋ	\ng	\ng	N	gang, hænge

* Finalt burde det være \at\nv for et ikke-syllabisk a-schwa [ɐ].

Diakritika

IPA	Praat_dokument	Praat_fundet	TtT	Forklaring
'	\'1	\'1	2	tryk
:	\:f	\:f	:	forlængelse
:ʔ	\?g	\?g		stød

2.1.7 TtT

- Ordkodet

TtT er en måde at skrive IPA med mere gængse tegn. Der er altså en en-til-en korrespondance mellem IPA-tiers og TtT-tiers. I tabellerne ovenfor vises hvilke TtT-tegn, der svarer til hvilke IPA-koder.

For yderligere information om TtT-notationen mv. henvises der til Phonix-manualen (se fodnote på s. 4).

Eksempler

familiemenneske

IPA: {fam\ '1il\ ?gj\ swm\ efn\ swsg\ sw}

TtT: [fam2il!J0mEn0sg0]

døgnberedskab

IPA: {d\ '1\ vtjnbe\ rc\ ae\ dh\ ?gsg\ ae\ :f\ ?gb}

TtT: [d2CJnberzD!sgz: !b]

øveweekend

IPA: {\1\ o\ \:fw\ swWi\ :fg\ efn\ ?gd\ vt}

TtT: [2q:w0Wi:gEn!dC]

etagebyggeri

IPA: {et\1\ ae\ :f\ ?g\ cc\ swbyg\ vt\ rci\ :f\ ?g}

TtT: [et2z: !S0bygCri: !]

2.1.8 Ordklasse (PoS)

- Ordkodet

I dette tier tagges hvert token ud fra ordklasse (Part of Speech, PoS). Til det bruges de såkaldte **PAROLE-tags**. De består af op til 11 tegn, der fungerer som positioner for en række faste kategorier, hvoraf position 3-11 skifter efter den overordnede ordklasse. En kort forklaring af PAROLE-tags er at finde nedenfor, mens en detaljeret beskrivelse kan tilgås her:

https://korpus.dsl.dk/documentation/PAROLE-dokumentation/parole.doc_da.pdf

OBS: Ordklasse-tieret skal bruges med forsigtighed; ordklassekategorierne i PAROLE-tagsættet er meget detaljerede, men tieret er blevet opmærket automatisk, og man er derfor nødt til at tage forbehold for fejlanoteringer.

De forskellige ordklasser og deres tilhørende kategorier

Ordklasse	Underkategori	3	4	5	6	7	8	9	10	11
Substantiv (N)	C, P	Genus: C, N, ., -	Numerus: S, P, ., -	Kasus: G, U	=	=	Bestemthed: D, I, ., -			
Verb (V)	A, E	Modus: D, M, F, G, P	Tempus: R, A, -	=	Numerus: S, P, ., -	Genus: C, ., -	Bestemthed: D, I, ., -	Transkategorisering: R, A, ., -	Diatese: A, P, -	Kasus: G, U, ., -
Adjektiv (A)	N, C, O	Komparation: P, C, S, A, -	Genus: C, N, ., -	Numerus: S, P, ., -	Kasus: G, U, -	=	Bestemthed: D, I, ., -	Transkategorisering: R, U, -		
Adverbium (R)	G	Komparation: U, P, C, S, A								
Præposition (S)	P									
Konjunktion (C)	C, S									
Pronomen (P)	D, I, T, P, O, C	Person: 1, 2, 3, -	Genus: C, N, ., -	Numerus: S, P, .	Kasus: G, U, N	Ejernerum: S, P, ., -	Refleksivitet: N, Y, ., -	Stilleje: U, O, F, P, -		
Interjektion (I)	=									
Unik (U)	=									

Andet (X)	A, F, P, R, S, X									
-----------	------------------	--	--	--	--	--	--	--	--	--

Første position i et PAROLE-tag angiver den overordnede ordklasse.

Anden position angiver en underkategori af den overordnede ordklasse. De mulige værdier vil derfor variere fra ordklasse til ordklasse. Har man f.eks. med et adjektiv/numeralie at gøre, vil der på anden position kunne benyttes en af følgende værdier: N (Normal), C (Cardinal) eller O (Ordinal). Er det i stedet en konjunktion der er tale om, vil der i stedet kunne benyttes følgende værdier på anden position: C (Coordinating) eller S (Subordination).

Værdierne for position 3-11 varierer som tidligere nævnt fra ordklasse til ordklasse (se ovenstående tabel eller Figur 4 på s. 8 i den ovenfor linkede vejledning). Derudover vil en værdi i nogle tilfælde ikke være relevant eller tydeligt defineret for et bestemt ord eller for en hel ordklasse. I så tilfælde benyttes følgende pladsholdere:

=	Benyttes når en værdi ikke er relevant for en hel ordklasse
-	Benyttes når en værdi ikke er relevant for det givne ord
.	Benyttes når en værdi ikke er entydig for det givne ord

Et eksempel på brugen af "." er i substantiver der kan optræde i både fælles- og intetkøn. Ord som *virus* og *parameter* er eksempler på sådanne ord, og de ville derfor beskrives med et punktum på den plads, der beskriver grammatisk køn (medmindre der eksplicit står fx *et parameter*).

Nedenfor ses en tabel over alle de forskellige PAROLE-tags, som kan bruges i PoS-tieret. Eksemplerne er som udgangspunkt taget fra korpus. I tilfælde hvor der ikke var nogen gode eksempler i korpus, er nye eksempler blevet konstrueret. Disse er skrevet i kantede parenteser.

De forskellige PoS-tags i PAROLE-tagsættet

PoS (PAROLE)	Description	RedPoS	Example
AC---G=--	adjective cardinal genitive. Alt: AC---G	NUM_GEN	[Det er de <u>tos</u> projekt]
AC---U=--	adjective cardinal unmarked-case. Alt: AC---U	NUM	<u>to</u> personer
ANA---=-R	adjective normal absolute-superlative adverbial-use. Findes kun som ANA---	ADJ	en af de klasser der er <u>allermindst</u> sammenhold i
ANA..U=DU	adjective normal absolute-superlative unmarked-case definite unmarked-use. Alt: ANA..U	ADJ	den <u>allerstørste</u> virksomhed
ANC---=-R	adjective normal comparative adverbial-use. Alt: ANC--- eller ANC	ADJ	som nu ikke eksisterer <u>mere</u>
ANC.PU=.U	adjective normal comparative plural unmarked-case unmarked-use. Alt ANC.PU	ADJ	der var gået <u>flere</u> år
ANC.SU=IU	adjective normal comparative singular unmarked-case indefinite unmarked-use. Alt ANC.SU	ADJ	der var ikke <u>mere</u> plads derude

ANC..G=.U	adjective normal comparative genitive unmarked-use	ADJ_GEN	[Nu er det den/de <u>ældres</u> tur]
ANC..U=.U	adjective normal comparative unmarked-case unmarked-use. Alt: ANC..U	ADJ	deres regning bliver ikke <u>større</u> end min
ANP---=-R	adjective normal positive adverbial-use. Alt: ANP--	ADJ	det kunne <u>godt</u> være
ANPCSU=IU	adjective normal positive common singular unmarked-case indefinite unmarked-use. Alt: ANPCSU	ADJ	en <u>halv</u> dag
ANPCSU=.U	adjective normal positive common singular unmarked-case unmarked-use. Alt: ANPCSU	ADJ	være religiøs på sin <u>egen</u> måde
ANPNSU=IU	adjective normal positive neuter singular unmarked-case indefinite unmarked-use. Alt: ANPNSU	ADJ	det er sgu rimelig <u>sjovt</u>
ANPNSU=.U	adjective normal positive neuter singular unmarked-case unmarked-use. Alt: ANPNSU	ADJ	det er jeres <u>eget</u> problem
ANP.PG=.U	adjective normal positive plural genitive unmarked-use. Alt: ANP.PG	ADJ_GEN	for <u>manges</u> vedkommende
ANP.PU=.U	adjective normal positive plural unmarked-case unmarked-use. Alt: ANP.PU	ADJ	de/nogle <u>små</u> børn
ANP.SG=DU	adjective normal positive singular genitive definite unmarked-use. Alt: ANP.SG	ADJ_GEN	den <u>enkeltes</u> ideologi
ANP.SU=DU	adjective normal positive singular unmarked-case definite unmarked-use. Alt: ANP.SU	ADJ	det var for <u>hele</u> skolen
ANP.SU=IU	adjective normal positive singular unmarked-case indefinite unmarked-use. Alt: ANP.SU	ADJ	det er en <u>amerikansk</u> rapper
ANP.SU=.U	adjective normal positive singular unmarked-case unmarked-use. Alt: ANP.SU	ADJ	da jeg var <u>lille</u>
ANP..G=.U	adjective normal positive genitive unmarked-use. Alt: ANP..G	ADJ_GEN	[Nu er det den/de <u>stores</u> tur]
ANP..U=.U	adjective normal positive unmarked-case unmarked-use. Alt: ANP..U	ADJ	det var i <u>sidste</u> uge
ANS---=-R	adjective normal superlative adverbial-use. Alt: ANS---	ADJ	det jeg var <u>mest</u> interesseret i
ANS.PU=DU	adjective normal superlative plural unmarked-case definite unmarked-use. Alt: ANS.PU	ADJ	sådan er de <u>fleste</u> faktisk
ANS.PU=.U	adjective normal superlative plural unmarked-case unmarked-use. Alt: ANS.PU	ADJ	dem der får <u>flest</u> point
ANS.SU=DU	adjective normal superlative singular unmarked-case definite unmarked-use. Alt: ANS.SU	ADJ	for det <u>meste</u> sidder vi og snakker
ANS.SU=IU	adjective normal superlative singular unmarked-case indefinite unmarked-use	ADJ	-

ANS..U=DU	adjective normal superlative unmarked-case definite unmarked-use. Alt: ANS..U	ADJ	<i>de/den <u>bedste</u> koncerter/koncert</i>
ANS..U=.U	adjective normal superlative unmarked-case unmarked-use. Alt: ANS..U	ADJ	<i>det var <u>lettest</u> synes jeg</i>
AO---G=--	adjective ordinal genitive. Alt: AO---G	NUM_ORD_GEN	<i>Christian den <u>fjerdes</u> oprindelige byggestil</i>
AO---U=--	adjective ordinal unmarked-case. Alt: AO---U	NUM_ORD	<i><u>syvende</u> klasse</i>
CC	conjunction coordinative	SKONJ	<i>karakterer <u>og</u> årskarakterer</i>
CS	conjunction subordinative	UKONJ	<i>sjoovere <u>end</u> syvende klasse</i>
I=	interjection. Alt: I	INTERJ	<i>ja</i>
NCCPG==D	noun common common-gender plural genitive definite. Alt: NCCPG	N_GEN	<i>det er meget <u>forældrenes</u> valg</i>
NCCPG==I	noun common common-gender plural genitive indefinite	N_GEN	<i><u>forældres</u> skænderier</i>
NCCPU==D	noun common common-gender plural unmarked-case definite. Alt: NCCPU	N	<i>i <u>ferierne</u> plejer du at være sammen med dem</i>
NCCPU==I	noun common common-gender plural unmarked-case indefinite	N	<i>vi skal optræde tre <u>gange</u></i>
NCCPU==.	noun common common-gender plural unmarked-case	N	<i>de/nogle <u>pårørende</u></i>
NCCSG==D	noun common common-gender singular genitive definite. Alt: NCCSG	N_GEN	<i><u>byens</u> historie</i>
NCCSG==I	noun common common-gender singular genitive indefinite	N_GEN	<i>en <u>kammerats</u> mor</i>
NCCSU==D	noun common common-gender singular unmarked-case definite. Alt: NCCSU	N	<i>så skal vi i <u>biografen</u></i>
NCCSU==I	noun common common-gender singular unmarked-case indefinite	N	<i>de stod på <u>række</u></i>
NCNPG==D	noun common neuter-gender plural genitive definite. Alt: NCNPG	N_GEN	<i>i <u>årenes</u> løb</i>
NCNPG==I	noun common neuter-gender plural genitive indefinite	N_GEN	<i>to <u>minutters</u> udvisning</i>
NCNPU==D	noun common neuter-gender plural unmarked-case definite. Alt: NCNPU	N	<i><u>ordene</u> har fået en anden værdi</i>
NCNPU==I	noun common neuter-gender plural unmarked-case indefinite	N	<i>i <u>femten</u> år</i>
NCNSG==D	noun common neuter-gender singular genitive definite. Alt: NCNSG	N_GEN	<i>alle <u>landets</u> skoler</i>

NCNSG==I	noun common neuter-gender singular genitive indefinite	N_GEN	<i>dit livs <u>big</u> time vilde experience</i>
NCNSU==D	noun common neuter-gender singular unmarked-case definite. Alt: NCNSU	N	<i>jeg tager <u>toget</u> til Glumsø</i>
NCNSU==I	noun common neuter-gender singular unmarked-case indefinite	N	<i>der er et <u>problem</u></i>
NC.PU==D	noun common plural unmarked-case definite	N	<i>de her <u>virus</u></i>
NC.PU==I	noun common plural unmarked-case indefinite. Alt: NC.PU	N	<i>noget <u>virus</u></i>
NC.SU==I	noun common singular unmarked-case indefinite. Alt: NC.SU	N	<i>det er til <u>gavn</u> og glæde</i>
NC..G==.	noun common genitive	N_GEN	-
NC..U==I	noun common unmarked-case indefinite. Alt: NC..U	N	<i>Vi har snakket lidt i <u>øst</u> og i <u>vest</u></i>
NC..U==.	noun common unmarked-case	N	<i>lad dem da få <u>lov</u></i>
NP--G==-	noun proper genitive. Alt: NP--G	EGEN_GEN	<i><u>Danmarks</u> bedste jord</i>
NP--U==-	noun proper unmarked-case. Alt: NP--U	EGEN	<i>jeg er født i <u>Århus</u></i>
PC--PG---	pronoun coordinative plural genitive	PRON_REC_GEN	<i>vi skrev <u>hinandens</u> navne op</i>
PC--PU---	pronoun coordinative plural unmarked-case	PRON_REC	<i>vi kender <u>hinanden</u></i>
PD-CSG--U	pronoun demonstrative common singular genitive unmarked-style	PRON_DEMO_GEN	<i>[beskriv <u>dennes</u> form]</i>
PD-CSU--O	pronoun demonstrative common singular unmarked-case obsolete	PRON_DEMO	<i>[hvad skete der egentlig siden <u>hin</u> aften, da telefonen kimede]</i>
PD-CSU--U	pronoun demonstrative common singular unmarked-case unmarked-style	PRON_DEMO	<i>lejligheder som <u>denne</u> her</i>
PD-NSU--U	pronoun demonstrative neuter singular unmarked-case unmarked-style	PRON_DEMO	<i>jeg skulle have ordnet <u>det</u> kød derude</i>
PD-.PG--U	pronoun demonstrative plural genitive unmarked-style	PRON_DEMO_GEN	<i>[beskriv <u>disses</u> former] (søgt)</i>
PD-.PU--U	pronoun demonstrative plural unmarked-case unmarked-style	PRON_DEMO	<i>under <u>de/disse</u> forhold</i>
PD-..U--U	pronoun demonstrative unmarked-case unmarked-style	PRON_DEMO	<i>[det må du/de <u>selv</u> om]</i>
PI-CSG--U	pronoun indefinite common singular genitive unmarked-style	PRON_UBST_GEN	<i>når det er <u>ens</u> egne børn</i>
PI-CSU--U	pronoun indefinite common singular unmarked-case unmarked-style	PRON_UBST	<i>det var <u>en</u> jeg fik</i>

PI-C.N--U	pronoun indefinite common nominative unmarked-style	PRON_UBST	<i>alle de her idræts-skader <u>man</u> hører om</i>
PI-NSU--U	pronoun indefinite neuter singular unmarked-case unmarked-style	PRON_UBST	<i>du må ikke foran-dre <u>noget</u></i>
PI-.PG--U	pronoun indefinite plural genitive unmarked-style	PRON_UBST_GEN	<i>nogen <u>andres</u> for-ældre</i>
PI-.PU--O	pronoun indefinite plural unmarked-case obsolete	PRON_UBST	<i>det er <u>somme</u> tider hårdt</i>
PI-.PU--U	pronoun indefinite plural unmarked-case unmarked-style	PRON_UBST	<i><u>nogle</u> dejlige na-boer</i>
PI-..G--U	pronoun indefinite genitive unmarked-style	PRON_UBST_GEN	<i>for <u>nogens</u> vedkom-mende</i>
PO1CSUPNF	pronoun possessive 1st-person common singular unmarked-case plural nonreflexive formal	PRON_POSS	<i>for hun er en af <u>vor</u> egne</i>
PO1CSUSNU	pronoun possessive 1st-person common singular unmarked-case singular nonreflexive unmarked-style	PRON_POSS	<i>deres regning er mindre end <u>min</u></i>
PO1NSUPNF	pronoun possessive 1st-person neuter singular unmarked-case plural nonreflexive formal	PRON_POSS	<i>billeder fra <u>vort</u> bryllup</i>
PO1NSUSNU	pronoun possessive 1st-person neuter singular unmarked-case singular nonreflexive unmarked-style	PRON_POSS	<i>det er <u>mit</u> arbejde</i>
PO1.PUPNF	pronoun possessive 1st-person plural unmarked-case plural nonreflexive formal	PRON_POSS	<i>i <u>vore</u> dage</i>
PO1.PUSNU	pronoun possessive 1st-person plural unmarked-case singular nonreflexive unmarked-style	PRON_POSS	<i>det står i <u>mine</u> pa-pirer</i>
PO1..UPNU	pronoun possessive 1st-person unmarked-case plural nonreflexive unmarked-style	PRON_POSS	<i>de skulle ud i <u>vores</u> køkken</i>
PO2CSUSNU	pronoun possessive 2nd-person common singular unmarked-case singular nonreflexive unmarked-style	PRON_POSS	<i>han er <u>din</u> mand</i>
PO2NSUSNU	pronoun possessive 2nd-person neuter singular unmarked-case singular nonreflexive unmarked-style	PRON_POSS	<i>du passer <u>dit</u> ar-bejde</i>
PO2.PUSNU	pronoun possessive 2nd-person plural unmarked-case singular nonreflexive unmarked-style	PRON_POSS	<i>har du fået lavet <u>dine</u> lektier</i>
PO2..UPNU	pronoun possessive 2nd-person unmarked-case plural nonreflexive unmarked-style	PRON_POSS	<i>en forskel i <u>jeres</u> sprog</i>
PO2..U.NP	pronoun possessive 2nd-person unmarked-case nonreflexive polite	PRON_POSS	<i>De lader <u>Deres</u> mand om det</i>
PO3CSUSYU	pronoun possessive 3rd-person common singular unmarked-case singular reflexive unmarked-style	PRON_POSS	<i>på <u>sin</u> vis</i>

PO3NSUSYU	pronoun possessive 3rd-person neuter singular unmarked-case singular reflexive unmarked-style	PRON_POSS	<i>hver passer <u>sit</u></i>
PO3.PUSYU	pronoun possessive 3rd-person plural unmarked-case singular reflexive unmarked-style	PRON_POSS	<i>sørge for <u>sine</u> børn</i>
PO3..UPNU	pronoun possessive 3rd-person unmarked-case plural nonreflexive unmarked-style	PRON_POSS	<i><u>deres</u> regning bliver ikke større</i>
PO3..USNU	pronoun possessive 3rd-person unmarked-case singular nonreflexive unmarked-style	PRON_POSS	<i><u>hans/hendes</u> søn kom hjem fra Oslo</i>
PP1CPN-NU	pronoun personal 1st-person common plural nominative nonreflexive unmarked-style	PRON_PERS	<i>sådan noget havde <u>vi</u> da</i>
PP1CPU-.U	pronoun personal 1st-person common plural unmarked-case unmarked-style	PRON_PERS	<i>så kigger han på <u>os</u></i>
PP1CSN-NU	pronoun personal 1st-person common singular nominative nonreflexive unmarked-style	PRON_PERS	<i><u>jeg</u> har fået lov</i>
PP1CSU-.U	pronoun personal 1st-person common singular unmarked-case unmarked-style	PRON_PERS	<i>en million kroner til <u>mig</u></i>
PP2CPN-NU	pronoun personal 2nd-person common plural nominative nonreflexive unmarked-style	PRON_PERS	<i>hvad siger <u>I</u> til det</i>
PP2CPU-.U	pronoun personal 2nd-person common plural unmarked-case unmarked-style	PRON_PERS	<i>men hvad så med <u>jer</u></i>
PP2CSN-NU	pronoun personal 2nd-person common singular nominative nonreflexive unmarked-style	PRON_PERS	<i>bruger <u>du</u> sukker</i>
PP2CSU-.U	pronoun personal 2nd-person common singular unmarked-case unmarked-style	PRON_PERS	<i>det er fuldstændig op til <u>dig</u> selv</i>
PP2C.N-NP	pronoun personal 2nd-person common nominative nonreflexive polite	PRON_PERS	<i>gør <u>De</u> bare det</i>
PP2C.U-.P	pronoun personal 2nd-person common unmarked-case polite	PRON_PERS	<i>jeg lagde det på bordet for <u>Dem</u></i>
PP3CSN-NU	pronoun personal 3rd-person common singular nominative nonreflexive unmarked-style	PRON_PERS	<i>og <u>hun/han</u> var ydermere flink</i>
PP3CSU-NU	pronoun personal 3rd-person common singular unmarked-case nonreflexive unmarked-style	PRON_PERS	<i><u>den</u> er fra Færøerne</i>
PP3NSU-NU	pronoun personal 3rd-person neuter singular unmarked-case nonreflexive unmarked-style	PRON_PERS	<i><u>det</u> er rigtigt</i>
PP3.PN-NU	pronoun personal 3rd-person plural nominative nonreflexive unmarked-style	PRON_PERS	<i>hvis <u>de</u> kan skaffe mig en lejlighed</i>
PP3.PU-NU	pronoun personal 3rd-person plural unmarked-case nonreflexive unmarked-style	PRON_PERS	<i><u>dem</u> jeg har kendt</i>
PP3..U-YU	pronoun personal 3rd-person unmarked-case reflexive unmarked-style	PRON_PERS	<i>man brokker <u>sig</u> over huslejen</i>
PT-CSU--U	pronoun interrogative-relative common singular unmarked-case unmarked-style	PRON_INTER_REL	<i><u>hvilken</u> skole gik du på</i>

PT-C.U--U	pronoun interrogative-relative common unmarked-case unmarked-style	PRON_INTER_REL	<i>hvem skal så bo her</i>
PT-NSU--U	pronoun interrogative-relative neuter singular unmarked-case unmarked-style	PRON_INTER_REL	<i>hvilket år er du født</i>
PT-PU--U	pronoun interrogative-relative plural unmarked-case unmarked-style	PRON_INTER_REL	<i>hvilke lærere synes du er de fede</i>
PT-.SU--U	pronoun interrogative-relative singular unmarked-case unmarked-style	PRON_INTER_REL	<i>heller ikke fløde eller hvad</i>
PT-.G--U	pronoun interrogative-relative genitive unmarked-style	PRON_INTER_REL_GEN	[<i>hvis er det</i>] (alle forekomster af <i>hvis</i> er markeret som konjunktioner)
RGA	adverb general absolute-superlative	ADV	<i>hvor vil du <u>allerhelst</u> bo</i>
RGC	adverb general comparative	ADV	<i>det var <u>længere</u> nede</i>
RGP	adverb general positive	ADV	<i>så <u>længe</u> du bor i stuen</i>
RGS	adverb general superlative	ADV	<i>dem der bor <u>længst</u> væk</i>
RGU	adverb general unmarked-comparison	ADV	<i>det fik jeg <u>ikke</u> gjort</i>
SP	adposition preposition	PRÆP	<i>ude <u>ved</u> Strandvejen</i>
U=	unique.	UNIK	<i>dem <u>der</u> ikke kunne</i>
VADA=----A-	verb main indicative past active. Alt: VADA	V_PAST	<i>det <u>vidste</u> jeg ikke</i>
VADA=----P-	verb main indicative past passive. Alt: VADA	V_PAST	<i>vi <u>skiftedes</u> lidt</i>
VADR=----A-	verb main indicative present active. Alt: VADR	V_PRES	<i>jeg <u>er</u> født i Århus</i>
VADR=----P-	verb main indicative present passive. Alt: VADR	V_PRES	<i>det <u>staves</u> ikke som det <u>siges</u></i>
VAF=----A-	verb main infinitive active. Alt: VAF, VAF-	V_INF	<i>jeg skal lige <u>have</u> fat i dem</i>
VAF=----P-	verb main infinitive passive. Alt: VAF, VAF-	V_INF	<i>den skal <u>hjælpes</u></i>
VAG-=SCI--U	verb main gerund singular common indefinite unmarked-case. Alt: VAG-	V_GERUND	<i>den fysiske <u>for-</u> <u>måen</u></i>
VAM=-----	verb main imperative. Alt: VAM-	V_IMP	<i><u>lad</u> dem da få lov</i>
VAPA (VAPA=....-.)	verb main participle past	V_PARTC_PAST	<i>jeg har <u>fået</u> lov</i>
VAPA=---R--	verb main participle past adverbial-use. Alt: VAPA	V_PARTC_PAST	[<i>han sprang <u>for-</u> <u>skrækket</u> op</i>]
VAPA=P..A-G	verb main participle past plural genitive	V_PARTC_PAST	[<i>de/nogle <u>skræm-</u> <u>tes</u> historie</i>]

VAPA=P..A-U	verb main participle past plural unmarked-case. Alt: VAPA	V_PARTC_PAST	[<i>de/nogle <u>røgede</u> sild</i>]
VAPA=SCDA-U	verb main participle past singular common definite unmarked-case	V_PARTC_PAST	[<i>alle de andre er <u>gåen</u> nu</i>]
VAPA=S.DA-G	verb main participle past singular definite genitive	V_PARTC_PAST	[<i>den <u>skræmtes</u> historie</i>]
VAPA=S.DA-U	verb main participle past singular definite unmarkedcase. Alt: VAPA	V_PARTC_PAST	[<i>den <u>røgede</u> sild</i>]
VAPA=S.IA-U	verb main participle past singular indefinite unmarkedcase. Alt: VAPA	V_PARTC_PAST	[<i>en <u>røget</u> sild</i>]
VAPA=S.I.-U	verb main participle past singular indefinite unmarkedcase	V_PARTC_PAST	[<i>en <u>forskrækket</u> mand</i>] [<i>han sprang <u>for-</u> <u>skrækket</u> op</i>]
VAPR=---R--	verb main participle present adverbial-use. Alt: VAPR	V_PARTC_PRES	[<i>det har været <u>for-</u> <u>bløffende</u> nemt</i>]
VAPR=...A-U	verb main participle present unmarked-case. Alt: VAPR	V_PARTC_PRES	[<i>det <u>flyvende</u> tæppe</i>]
VAPR=....-U	verb main participle present unmarked-case. Alt: VAPR	V_PARTC_PRES	[<i>man blev bare <u>stå-</u> <u>ende</u></i>] [<i>han sov <u>stående</u></i>]
VEDA=----A-	verb medial indicative past active. Alt: VEDA	V_MED_PAST	[<i>det <u>lykkedes</u></i>]
VEDR=----A-	verb medial indicative present active. Alt: VEDR	V_MED_PRES	[<i>det <u>synes</u> jeg er meget</i>]
VEF=-----A-	verb medial infinitive active. Alt: VEF-	V_MED_INF	[<i>vi <u>mødes</u> på et tidspunkt</i>]
VEPA=....-U	verb medial participle past unmarked-case. Alt: VEPA	V_MED_PARTC_PAST	[<i>du har <u>skændtes</u> med din mand</i>]

PoS-tags fra kategorien *Residual* (koden X i PAROLE-tagsættet) beskrives for sig i tabellen nedenfor eftersom disse tags i korpus ikke er benyttet som beskrevet i manualerne. Ifølge manualerne burde de dække over forkortelser (XA), fremmedord (XF), formler (XR) og andet (XX). De er dog ikke benyttet konsekvent, og derudover er kategorien XZ, der ikke er beskrevet i manualerne, blevet brugt flittigt. Tabellen nedenfor giver en oversigt over hvordan de faktisk benyttes i korpus.

Tabel over PoS-tags fra kategorien *Residual*

PoS (PAROLE)	Description	RedPoS	Example
XA	(Egl. "residual abbreviation")	FORK	<i>ottende <u>B</u></i>

	Bogstaver i materialiter ("A", "B", "C"), udbrud ("uh"), sammensætninger og lign. hvor der indgår et bogstav materialiter (snabel_A) samt en række umiddelbart tilfældige ord (eksempelvis <i>nærigt</i> , <i>steppet</i> og <i>grovt</i>)		
XF	(Egl. "residual foreign") Dette tag er benyttet ved fremmedord (dog ikke alle), nogle danske ord, der ortografisk ligner fremmedord (fx <i>and</i>), samt en række umiddelbart tilfældige ord.	UL	<i>det er sådan <u>what</u></i>
XR	(Egl. "residual formulae") Dette tag er kun benyttet 1 gang i korpus, ved ordet <i>OBS</i> .	FORM	<i>Bilka og og <u>OBS</u></i>
XX	(Egl. "residual other") Dette tag ser ud til at være benyttet mere eller mindre tilfældigt.	XX	<i>tysk <u>fremfor</u> fransk</i>
XZ	(Ikke i PAROLE-tagsættet) Proprier (ord der begynder med stort), selvfølgelig, ord med underscore, xxx (uforståelig tale), oe (amerikansk Øh)	[diverse]	<i>i <u>ha-</u> i starten af halvfyrtserne</i>
U	(Ikke i PAROLE-tagsættet) Ordene <i>at</i> , <i>som</i> og <i>der</i> .	UNIK	<i><u>at</u> lære <u>at</u> stave</i>
U	Ord med underscore – men ikke alle ord med underscore er annoteret med dette tag (det er p.t. kun godt 100 hits i alt).	*U*	<i>den der <u>T</u> shirt</i>

2.1.9 Simpel ordklasse (RedPoS)

- Ordkodet

Dette tier indeholder forsimplede (Reduced Parts of Speech) PAROLE-tags. I tabellen ovenfor kan den forsimplede ordklasse for hvert PAROLE-tag findes i kolonnen RedPoS. I tabellen nedenfor er der beskrivelser samt eksempler for hvert af disse reducerede tags. Eksemplerne er som udgangspunkt taget fra korpus. I tilfælde hvor der ikke var nogen gode eksempler i korpus, er nye eksempler blevet konstrueret. Disse er skrevet i kantede parenteser. Tvivlsomme tags (jf. beskrivelsen af kategorien *Residual* på s. 15) er markeret med gult.

Tabel over reducerede PAROLE-tags (RedPoS)

RedPoS	Beskrivelse	Eksempel
ADJ	Adjektiv (tillægsord)	<i>en <u>halv</u> dag</i>
ADJ_GEN	Adjektiv i genitiv	<i>for <u>manges</u> vedkommende</i>
ADV	Adverbium (biord)	<i>det fik jeg <u>ikke</u> gjort</i>
EGEN	Egennavn	<i>jeg er født i <u>Århus</u></i>

EGEN_GEN	Egennavn i genitiv	<i>Danmarks bedste jord</i>
FORK	[Forkortelse]	
FORM	[Formel]	
INTERJ	Interjektion (udråbsord)	<i>ja</i>
N	Substantiv	<i>så skal vi i <u>biografen</u></i>
N_GEN	Substantiv i genitiv	<i>to <u>minutters</u> udvisning</i>
NUM	Numerale (talord)	<i><u>to</u> personer</i>
NUM_GEN	Numerale i genitiv	<i>[Det er de <u>tos</u> projekt]</i>
NUM_ORD	Ordinaltal	<i><u>syvende</u> klasse</i>
NUM_ORD_GEN	Ordinaltal i genitiv	<i>Christian den <u>fjertes</u> oprindelige byggestil</i>
PRON_DEMO	Demonstrativt pronomen (stedord)	<i>lejligheder som <u>denne</u> her</i>
PRON_DEMO_GEN	Demonstrativt pronomen i genitiv	<i>[beskriv <u>dennes</u> form]</i>
PRON_INTER_REL	Interrogativt pronomen	<i>[<u>hvis</u> er det]</i>
PRON_INTER_REL_GEN	Interrogativt pronomen i genitiv	<i><u>hvem</u> skal så bo her</i>
PRON_PERS	Personligt pronomen	<i><u>jeg</u> har fået lov</i>
PRON_POSS	Possessivt pronomen	<i>det er <u>mit</u> arbejde</i>
PRON_REC	Reciprokt pronomen	<i>vi kender <u>hinanden</u></i>
PRON_REC_GEN	Reciprokt pronomen i genitiv	<i>vi skrev <u>hinandens</u> navne op</i>
PRON_UBST	Indefinit (ubestemt) pronomen	<i><u>nogle</u> dejlige naboer</i>
PRON_UBST_GEN	Indefinit pronomen i genitiv	<i>for <u>nogens</u> vedkommende</i>
PRÆP	Præposition (forholdsord)	<i>ude <u>ved</u> Strandvejen</i>
SKONJ	Sideordnende konjunktion	<i>og, eller, men</i>
UKONJ	Underordnende konjunktion	<i>fordi, hvis, mens</i>
UL	[Fremmedord]	<i>det er sådan <u>what</u></i>
UNIK	Unik – forskellige partikler	<i>som, der, at</i>
V_GERUND	Verbum (udsangsord), gerundium	<i>den fysiske <u>formåen</u></i>
V_IMP	Verbum i imperativ (bydemåde)	<i><u>lad</u> dem da få lov</i>
V_INF	Verbum i infinitiv	<i>jeg skal lige <u>have</u> fat i dem</i>
V_MED_INF	Medialt verbum i infinitiv	<i>vi <u>mødes</u> på et tidspunkt</i>
V_MED_PARTC_PAST	Medialt verbum i perfektum (førnutid)	<i>du har <u>skændtes</u> med din mand</i>
V_MED_PAST	Medialt verbum i præteritum (datid)	<i>det <u>lykkedes</u></i>
V_MED_PRES	Medialt verbum i præsens (nutid)	<i>det <u>synes</u> jeg er meget</i>
V_PARTC_PAST	Verbum i perfektum	<i>jeg har <u>fået</u> lov</i>

V_PARTC_PRES	Verbum i præsens participium (pluskvamperfektum eller førdatid)	det <i>flyvende</i> tæppe
V_PAST	Verbum i præteritum	det <i>vidste</i> jeg ikke
V_PRES	Verbum i præsens	jeg <i>er</i> født i Århus
XX	Andet	
XZ	Uklar funktion	

2.1.10 Events (events)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Talerproducerede, betydningsbærende lyde som f.eks. *hoster*, *nyser*, *smasker* eller *rømmer sig*. OBS: Visse af annotationerne påpeger i stedet fejl og mangler.

Eksempel på sekvenskodning: Når eksempelvis en token er annoteret med "rømmen", vil det token få følgende tag: 1_rømmen_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.1.11 Globale events (global events)

- Ordkodet

Bruges ved støj der ikke kan tilskrives en bestemt taler som f.eks. "Hunden gør".

OBS: Denne bør i virkeligheden sekvenskodes da der er forekomster af flerordssekvenser opmærket med denne kode.

2.1.12 Emfase (emphasis)

- Ordkodet

Et ord udtale med emfatisk tryk markeres med et udråbstegn, <!>. Dette tier indeholder altså blot udråbstegn i samme tidsintervaller som de ord der er udtalt med emfatisk tryk.

2.1.13 Phonetic (phonetic)

- Ordkodet

Kontaktperson: Marie Maegaard, mamae@hum.ku.dk

I dette tier markeres både emfase og forlængelse. Emfase markeres med et udråbstegn, <!>, efter ordet og forlængelse markeres med et kolon, <:>, efter det tegn der repræsenterer den lyd der forlænges.

Tieret indeholder derfor kun de ord der er udtalt med emfatisk tryk eller med forlængelse.

Eksempler:

afprøver!

der:

du!

ps:ykotisk

2.2 IIV-tiers

- Kontaktpersoner: Torben Juel Jensen, tjuelj@hum.ku.dk; Astrid Ag, astridag@hum.ku.dk; Frans Greger-sen, fg@hum.ku.dk
- Status: Åben. Alle tiers er åbne for alle.

IIV står for Intra Individual Variation. På niveauerne Aktivitetstype, Samtalestype og Makro-sproghandling er der udført **helkodning**, dvs. alt er kodet og ingen del af en samtale falder uden for en kategori. På niveau-erne Interaktionsstruktur, Genre og Udsigelse er der udført **delkodning**, dvs. kun visse strukturer, genrese-kvenser, fraser eller ord er kodet, og der er således dele af samtalen, der ikke har en kategori. De enkelte koder er ikke talerspecifikke, men alle talere kodes. Dvs. at hvert niveau inden for IIV kun får et tier i hver enkelt samtale uanset antallet af deltagere i samtalen.

2.2.1 IIV Aktivitetstype (Aktivitetstype)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Helkodning.

Alle samtaler inddeles i aktivitetstyper, alt efter hvilken aktivitet eller "fase" deltagerne er engageret i. De seks aktivitetstyper, vi inddeler samtalerne i, er:

Abi	baggrundsinterview
Asa	samtale
Asi	samtale med ikke-deltager
Ael	eliciteret tale
Asp	sprogholdningsundersøgelse
Ati	tilsagnserklæring

Disse kategorier kan kombineres hvis ønsket.

Eksempel på sekvenskodning: Når eksempelvis 70 tokens er annoteret med *Abi*, vil de 70 tokens få disse 70 tags: 1_Abi_I, 2_Abi_I, 3_Abi_I, [...], 69_Abi_I, 70_Abi_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.2.2 IIV Interaktionsstruktur (Interaktionsstruktur)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Delkodning

Interaktionsstrukturen defineres på baggrund af graden af styring i samtalen, som gives af den symmetri eller asymmetri, der findes i den lokale relation mellem deltageres interaktionelle roller. I enkeltinterview og grupeinterview kodes for fem forskellige interaktionsstrukturer: I4, I5, I6, I7 og I8.

Tabel over de forskellige typer af interaktionsstruktur

I1	Interviewer tager stærkt initiativ og informanten svarer kort
I2	Interviewer tager stærkt initiativ og informanten svarer langt
I3	Interviewer tager initiativet og formulerer sig langt og informanten svarer kort
I4	Ingen styring
I5	Informant tager stærkt initiativ, og interviewer svarer
I6	Kamp om ordet
I7	Informant tager stærkt initiativ, og anden informant svarer
I8	Monolog
I9	Andet

Gruppesamtaler uden interviewer kodes for I6, I7 og I8.

Jf. fodnote 28 i kodningsmanualen (side 26) benyttes I1, I2, I3 og I9 ikke længere. De optræder dog i korpus og deres beskrivelser er derfor taget med i dette dokument.

Eksempel på sekvenskodning: Når eksempelvis 39 tokens er annoteret med I8, vil de 39 tokens få disse 39 tags: 1_I8_I, 2_I8_I, 3_I8_I, [...], 38_I8_I, 39_I8_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.2.3 IIV Genre (Genre)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Delkodning.

Der skelnes mellem otte forskellige genrer:

Gna	Narrativ
Gsr	Specifik redegørelse
Ggr	Generel redegørelse
Gsb	Soap box
Gsl	Sladder og nedsættende omtale
Gbe	Betroelse
Gre	Refleksion
Gvi	Vittighed

Nogle af disse har en klar specifik funktion og kan derfor ses som hørende ind under en af kategorierne på makroproghandlingsniveauet, men vi har valgt at placere dem samlet på genreniveauet for systematikken

skyld. Andre talesprogsgenrer, fx skænderi og rådgivning, som meget vel kan optræde ofte i andre samtale typer, er meget sjældne, nærmende sig det ikke-eksisterende, i de sociolingvistiske interviews og gruppesamtaler i vores data, og de er derfor ikke med her.

De otte genrer kan kombineres i kodningen hvis ønsket.

Eksempel på sekvenskodning: Når eksempelvis 97 tokens er annoteret med *Gsl*, vil de 97 tokens få disse 97 tags: 1_Gsl_I, 2_Gsl_I, 3_Gsl_I, [...], 96_Gsl_I, 97_Gsl_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.2.4 IIV Samtale type (Samtale type)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Helkodning.

Hver samtale tildeles en overordnet kode, afhængig af deltagers sammensætningen. Der skelnes mellem fem samtale typer, og disse fem typer dækker spredningen i korpus mht. interaktionstype (interview eller ikke-interview), antal deltagere (en, to eller flere), og bekendthed mellem interviewer(e) og informant(er). Ved længerevarende skift i deltagers sammensætningen undervejs i samtalen gives den pågældende sekvens en anden kode end den overordnede. De fem samtale typer er:

Siek	enkeltinterview bekendt
Sieu	enkeltinterview ukendt
Sifk	gruppeinterview bekendt
Sifu	gruppeinterview ukendt
Sgfk	gruppesamtale uden interviewer

Eksempel på sekvenskodning: Når eksempelvis 136 tokens er annoteret med *Sifk*, vil de 136 tokens få disse 136 tags: 1_Sifk_I, 2_Sifk_I, 3_Sifk_I, [...], 135_Sifk_I, 136_Sifk_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.2.5 IIV Makro-sproghandling (Makro-sproghandling)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Helkodning.

Samtalerne inddeles på dette niveau på baggrund af, hvilken pragmatisk funktion sproget i samtalen på et givent tidspunkt udfylder – med andre ord, hvilken "makrosproghandling" der finder sted. Dette defineres på baggrund af, hvilken type indhold, der udveksles mellem samtaledelegerne. Der skelnes mellem følgende fem overordnede makrosproghandlinger:

Mvi	vidensudveksling
Mho	holdningsudveksling

Mfø	følelsesudveksling
Mha	handlingsbundet tale
Mfi	fiktion

De fem kategorier kan kombineres i kodningen hvis ønsket.

Eksempel på sekvenskodning: Når eksempelvis 682 tokens er annoteret med *Mvi*, vil de 682 tokens få disse 682 tags: 1_Mvi_I, 2_Mvi_I, 3_Mvi_I, [...], 681_Mvi_I, 682_Mvi_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.2.6 IIV Udsigelse (Udsigelse)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Delkodning.

Vi koder samlet for citat, imitation og illustrative lyde, uanset om det er taleren selv eller en anden (person, et dyr eller en ting), der bliver citeret, imiteret eller illustreret.

Koden for dette er *Ucil*.

Eksempel på sekvenskodning: Når eksempelvis seks tokens er annoteret med *Ucil*, vil de seks tokens få disse sels tags: 1_Ucil_I, 2_Ucil_I, 3_Ucil_I, 4_Ucil_I, 5_Ucil_I, 6_Ucil_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.2.7 Leksis og fraser

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Findes i BySoc, Køge, ModSjæl, Næstved og Odder.

OBS: Det har indtil videre ikke været muligt at tilvejebringe en fyldestgørende beskrivelse af denne kategori. Men "Ltot/ltot" og "Like/like" bør formentlig harmoniseres mht. store/små bogstaver, og det følgende inventar bør tjekkes, kompletteres og forklares:

Ltot/ltot/ÂˆLtot	"Kontaktpåkaldelser" (ikke, vel, hallo)
Like	(hvad pokker, javel, aha, er du rigtig klog, herregud, det må jeg love dig, uha, hov, uh)
like	(fucked up, eller sådan et eller andet, eller hvordan det, skidt med det, saftsus-, det er helt vildt, gud)
ludl	(yes yes yes)
Lsl	(ikke en junk)
s#	(halvtreds)

Isk	(funker)
Lsk	(tøsen)

Eksempel på sekvenskodning: Når eksempelvis en token er annoteret med *Ltot*, vil dette token få følgende tag: 1_Ltot_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.2.8 IIV AIG-kommentarer (IIV AIG Kommentarer)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Eventuelle kommentarer til Aktivitetstype-, Interaktionsstruktur- og Genre-kodningen. OBS: Indeholder til dels registreringer af fejl.

Eksempel på sekvenskodning: Når eksempelvis en token er annoteret med "Afbrud i interview pga. båndskift", vil dette token få følgende tags 1_Afbrud_i_interview_pga._båndskift_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.2.9 IIV SMU-kommentarer (IIV SMU kommentarer)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Eventuelle kommentarer til Samtalestype-, Makrosproghandling- og Udsigelse-kodningen. OBS: Indeholder til dels registreringer af fejl.

Eksempel på sekvenskodning: Når eksempelvis en token er annoteret med "Ucil tilskrives HCM", vil dette token få følgende tags 1_Ucil_tilskrives_HCM_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.2.10 IIV SMUL-kommentarer (IIV SMUL Kommentarer)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Måske eventuelle kommentarer til Samtalestype-, Makrosproghandling-, Udsigelse og Leksis og fraser-kodningen. OBS: Tieret er p.t. – uvist af hvilken grund – ikke tilgængeligt. Dog er det også næsten tomt, så der er under alle omstændigheder ikke meget at finde.

2.2.11 IIV-kommentarer (IIV Kommentarer)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Eventuelle kommentarer til IIV-kodningen. OBS: Indeholder fortrinsvis registreringer af fejl.

2.3 Fonetiktiers

- Kontaktperson: Marie Maegaard, mamae@hum.ku.dk
- Status: Åben. Alle tiers er åbne for alle.

Mere information kan findes i artiklen Fonetisk variantsætning i DGCSS' wiki samt manualer i Sprogforandringscentrets fonetikmappe på N-drevet – kontakt Marie hvis det lyder relevant.

Disse tiers indeholder kodninger af udtale af enkeltlyde i forskellige kontekster. Som udgangspunkt kodes alle udtaler med et tal, der har en specifik betydning for den givne kategori. I tvivlstilfælde er der benyttet forskellige metoder til at angive dette. Enten er der benyttet de generelle symboler (se nedenfor) eller også er der eksempelvis skrevet 12 hvis der er tvivl om lyden er af kvalitet 1 eller 2.

Generelle symboler:

*	Tvivl, umulig afgørelse grundet efterfølgende segment (=schwaD og participier)
* K	Usikkerhed om kvalitet
* T	Usikkerhed om tryk
* L	Usikkerhed om forlængelse
/	Adskillelse af forskellige bud på afgørelse
æ	Kvalitet der falder uden for variabelens konventionelle varianter
.&	Forlængelse grundet tøven
.§	Forlængelse grundet emfase
:	Tidligere markering af forlængelse
.!	Tidligere markering af forlængelse grundet latter eller andre ekstralingvistiske faktorer
.	Forlænget vokal. <.> (= punktum) skrives efter talsymboler (se kategorierne nedenfor)

2.3.1 AN-markering (ANmarkering)

- Ordkodet

AN kodes sammen med AM, men de to ligger i hver sit tier. AM/AN symboliserer betonet kort A i andre kontekster end før eller efter R og før halvvokal. Det drejer sig altså om fonologisk formodet kort A, som ikke indgår i faldende diftonger og ikke er nabolyd til R.

AN-kategorien bruges til a-lyde der forventeligt udtales som /a/ før /n/.

Følgende koder anvendes:

1	den sædvanlige vokalkvalitet i københavnsk langt 'a' i fx 'mase'.
2	den sædvanlige vokalkvalitet i rigsdansk kort 'a' i fx 'kast'.
3	den sædvanlige vokalkvalitet i rigsdansk kort 'a' i fx 'gammel'.
4	den vokalkvalitet, der er længere tilbagetrukket end 3, og som undertiden høres i københavnsk i fx 'kaffe'.
0	den vokalkvalitet, der er højere end 1, og som derfor lyder som 'e' i 'meld' (eller endnu højere).

2.3.2 AM-markering (AMmarkering)

- Ordkodet

AM kodes sammen med AN, men de to ligger i hver sit tier. AM/AN symboliserer betonet kort A i andre kontekster end før eller efter R og før halvvokal. Det drejer sig altså om fonologisk formodet kort A, som ikke indgår i faldende diftonger og ikke er nabolyd til R.

AM-kategorien bruges til a-lyde der forventeligt udtales som /a/ før /m/.

Der anvendes de samme koder som for ANmarkering.

2.3.3 ENG-markering (ENGmarkering)

- Ordkodet

ENG-variablen henviser til variationen af udtalen af den trykstærke vokal i ord som tænke, engelsk, enkel, penge mv.

Følgende koder anvendes:

1	symboliserer en e-udtale, der er den sædvanlige i ord som 'vinde', 'dele'
2	symboliserer en æ-udtale, der er den sædvanlige i ord som 'æske', 'sæbe'.

2.3.4 AJ-markering (AJmarkering)

- Ordkodet

AJ symboliserer den faldende diftong, der findes i eksempelvis *dig*, *vej* og *leg* i andre kontekster end efter R. AJ kodes sammen med ANR. Et evt. bortfald af halvvokal J i AJ an gives vha. 0 - (fx 30), når A-kvaliteten bedømmes som 3, og der efterfølgende er halvvokalbortfald.

Følgende koder anvendes:

1	den sædvanlige vokalkvalitet i københavnsk langt 'a' i fx 'mase'.
2	den sædvanlige vokalkvalitet i rigsdansk kort 'a' i fx 'kast'.
3	den sædvanlige vokalkvalitet i rigsdansk kort 'a' i fx 'gammel'.
4	den vokalkvalitet, der er længere tilbagetrukket end 3, og som undertiden høres i københavnsk i fx 'kaffe'.
0	den vokalkvalitet, der er højere end 1, og som derfor lyder som 'e' i 'meld' (eller endnu højere).
x0	halvvokalbortfald (efter vokal med kvalitet x – der står altså et tal i stedet for 'x')

2.3.5 ANR-markering (ANRmarkering)

- Ordkodet

ANR symboliserer enkelte særord med monoftongen A, hvor der indgår et R i den umiddelbare kontekst - fx *andre* og *anderledes*. ANR kodes sammen med AJ.

Der anvendes de samme koder som for ANmarkering.

2.3.6 RU-markering (RUmarkering)

- Ordkodet

RU symboliserer udtalevariationen "ru/ro" i ord som fx rulle, rugbrød, russisk og bruge, rude, rune.

Følgende koder anvendes:

1	for 'u'-udtale
2	for 'o'-udtale

2.3.7 W-markering (Wmarkering)

- Ordkodet

W henviser til udtaler af ord som eksempelvis skrevet, prøvet og lovet som kan udtales uden W-diftongen - altså som "skredd", "prødd" og "lådd".

Følgende koder anvendes:

1	for 'w'-udtale
0	for 'w'-bortfald

2.3.8 D-schwa-D-markering (DschwaDmarkering)

- Ordkodet

SchwaD-variablen er endelsen -et, som på dansk som regel realiseres enten som "schwa-blødt d", eller som "schwa-hårdt d". Det er altså variationen i ord som boet, løbet, siddet, huset, vandet, brødet. Variablen er inddelt i to og variantsættes i to tiers: et tier der hedder XschwaDmarkering og et der hedder DschwaDmarkering. Pointen er at der er en del danske sprogforskere der gennem tiderne har ment at variablen oftere (nogle mener: altid) udtales med hårdt d, når den kommer efter et blødt d. Altså at ord som siddet og brødet ovenfor typisk har hårdt d, selv hvis taleren ellers bruger blødt d. Det skal vi efterprøve, og derfor opdeler vi altså variablen i to.

Følgende koder anvendes:

1	for blødt d
2	for hårdt d
0	hvis endelsen ikke udtales

2.3.9 X-schwa-D-markering (XschwaDmarkering)

- Ordkodet

SchwaD-variablen er endelsen -et, som på dansk som regel realiseres enten som "schwa-blødt d", eller som "schwa-hårdt d". Det er altså variationen i ord som boet, løbet, siddet, huset, vandet, brødet. Variablen er inddelt i to og variantsættes i to tiers: et tier der hedder XschwaDmarkering og et der hedder DschwaDmarkering. Pointen er at der er en del danske sprogforskere der gennem tiderne har ment at variablen oftere (nogle mener: altid) udtales med hårdt d, når den kommer efter et blødt d. Altså at ord som siddet og brødet ovenfor typisk har hårdt d, selv hvis taleren ellers bruger blødt d. Det skal vi efterprøve, og derfor opdeler vi altså variablen i to.

Følgende koder anvendes:

1	for blødt d
2	for hårdt d
0	hvis endelsen ikke udtales

2.3.10 Kontekst fonetik (kontekst fonetik)

- Ordkodet

En fyldestgørende beskrivelse af dette tier har indtil videre ikke kunnet tilvejebringes.

2.3.11 Variant fonetik (variant fonetik)

- Ordkodet

Tieret findes kun i de eksplorative filer (jf. afsnittet "Metadata på samtaleniveau").

I dette tier angives vokalkvaliteten for A, E og U. Tieret er tæt forbundet med tierne *variant fonetik R*, *variant fonetik kontekst forventet* og *variant fonetik kontekst realiseret*.

I *variant fonetik R* angives det om der kommer et /r/ før eller efter vokalen. I *variant fonetik kontekst forventet* og *variant fonetik kontekst realiseret* lykskrives det der kommer efter vokalen – både det forventede og det realiserede.

A:

Her benyttes fire varianter:

A1	fladt katte-a
A2	katte-a
A3	kappe-a
A4	tilbagetrukket kappe-a
A0	et a hævet så meget at det lyder som [æ]

Hvis vokalen er lang, sættes et kolon efter tallet, hvis den er overlang pga. tøven sættes et "Ambers and", &, efter kolonet.

AJ:

Diftongen /aj/ er klassificeret for vokalens kvalitet med samme værdier som A i øvrigt, dvs. AJ2 for den lidt gammeldags, konservative udtale og AJ3 for den udtale den har i moderne københavnsk rigsmål (og AJ1 og AJ4 hvis det bliver nødvendigt, fx i 'bage' og 'meget')

Desuden tilføjes et 0, hvis [j]'et er helt væk, fx i ord som 'jeg, mig, sig, dig'. Eksempelvis AJ30 for et A med kvaliteten 3 og et [j] der er helt væk.

EN:

Vokalkvaliteten for E er fordelt på 2 kategorier:

EN0	æble-æ
-----	--------

EN1	mele-e
-----	--------

R:

Sænkningen fra [a] til [ɑ] efter /r/ som i 'ret, frem, græde'. Her skelnes 3 kvaliteter:

RA0	forekomst af snævert [a] efter /r/.
RA3	samme kvalitet som ved A
RA4	samme kvalitet som ved A

Her angives vokalkvaliteten af den vokal der i ortografien er repræsenteret med <u> når den optræder umiddelbart efter /r/, som i 'rulle, rugbrød'. Her skelnes mellem 2 kvaliteter:

RU0	vokalkvaliteten [u]
RU1	vokalkvaliteten [o]

Eksempler:

Tier	<i>bagved</i>	<i>Fælleparken</i>	<i>ja</i>	<i>grund</i>	<i>tænkt</i>
variant fonetik	A1:	A3	A2	RU1	EN0
variant fonetik R		_R		R_	
variant fonetik kontekst forventet	?+veD#	+g@n#	#	n?#	N?gd#
variant fonetik kontekst realiseret	+ve#	g#	#	n#	N#

2.3.12 Variant fonetik R (variant fonetik R)

- Ordkodet

Tieret findes kun i de eksplorative filer (jf. afsnittet "Metadata på samtaleniveau").

Variant fonetik R angiver om der kom et /r/ eller efter forekomsten af varianten beskrevet i variant fonetik. Er der ikke noget /r/ i umiddelbart naboskab er intervallet tomt.

Inventar:

_R	/r/ forekommer efter varianten
R_	/r/ forekommer før varianten
	/r/ forekommer ikke

2.3.13 Variant fonetik kontekst forventet (variant fonetik kontekst forventet)

- Ordkodet

Dette tier er tæt forbundet med tierne variant fonetik og variant fonetik kontekst realiseret.

I dette tier lydskrives det der forventes at komme efter varianten angivet i *variant fonetik*. Der er altså tale om en lydskrift af distinkt udtale af det givne ord.

Den følgende lydskriftskonvention benyttes (midterste kolonne):

IPA (grov)	DGCSS konvention	eksempel
i	i	mit
e	e	lidt
ɛ	E	mæt
æ	æ	bade
a	a	ladt, dreng
ɑ	A	lak
y	y	tyst
ø	ø	øst
œ	Ø	skøn
œ̃	rØ	grynt
œ̃	ö	grønt
U	u	husk
O	o	foto
ɔ	å	bund
ɒ	År	orne
ʌ	Å	bånd
ə	@	sætte
p ^h	p	pas
t ^s	t	tal
k ^h	k	kat
b	b	bip
d	d	dit
g	g	gik

m	m	mit
n	n	nat
ŋ	N	lang
l	l	lidt
ʁ	r	rod
w	w	hav
ð	D	med
j	j	jeg

Derudover benyttes < # > for ordgrænser og < + > for stavelsesgrænser.

2.3.14 Variant fonetik kontekst realiseret (variant fonetik kontekst realiseret)

- Ordkodet

Dette tier er tæt forbundet med tierne *variant fonetik* og *variant fonetik kontekst forventet*.

I dette tier lydskrives det der kommer efter varianten angivet i *variant fonetik*. Der er altså tale om en lydskrift af den givne udtale af ordet.

Der benyttes samme lydskriftskonvention som i tieret *variant fonetik kontekst forventet* (se ovenfor).

2.4 Grammatiktiers

- Kontaktperson: Varierende, se de enkelte tiers.
- Status: Varierende, se de enkelte tiers.

2.4.1 Grammatik generisk pronomener (grammatik)

- Kontaktperson: Torben Juel Jensen: tjuelj@hum.ku.dk.
- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Status: Forskerniveau.

I dette tier er der kodet for generiske pronomener i nominativform (*du*, *man*) uden at tage højde for generiske forekomster af *en* og *den*. Dette er gjort efter analyseapparatet "generisk" (jf. tieret "Generisk", se s. 33). Alle koder fra Generisk-tieret findes også i Grammatiktieret.

Grammatiktieret indeholder desuden kodning af alle reflektive pronomener, kodet efter analyseapparatet "refleksiv", der er beskrevet nedenfor.

Refleksiv

Yderligere information findes i dokumentet "Analyseapparat (refleksiv)". (Se mappen Korpusgruppe/_oversigt/Oversigt over tiers og tags/Kildedokumenter/Grammatiske analyseapparater fra TJJ/).

Alle forekomster kategoriseres i forhold til "anvendelse"; reflektivt anvendte forekomster kategoriseres tillige i forhold til "domæne" og "reference".

Kodning:

Alle forekomster af søgestrengene, der repræsenterer pronomener, markeres med et "R" samt som minimum delkoden for "anvendelse", dvs. "R_AR", "R_AN", "R_AA" eller "R_AI". I forbindelse med de forekomster, der kategoriseres som reflektive, tilføjes tillige delkoderne for "domæne" og "reference", fx "R_AR_DSE_RP".

Hvis man er i tvivl om en kodning anvendes bogstavet "X" på den pågældende plads, fx "R_AX" eller "R_AR_DX_RP". Hvis man mener, at en forekomst er tvetydig, kan der anvendes to koder (i alfabetisk rækkefølge), fx "R_AR_DSE_RPT". NB: denne kodningsmulighed skal begrænses mest muligt.

Følgende er en oversigt over de koder der anvendes:

Anvendelse		Domæne		Reference	
Tag	Beskrivelse	Tag*	Beskrivelse	Tag	Beskrivelse
AR	Refleksiv anvendelse	DSE	Pronominet forekommer efter subjektet	RP	Person (ental) , inklusive stillinger/roller, som kan indtages af personer, og dyr eller

					ting, hvortil der kan refereres med han eller hun
AN	Ikke-refleksiv anatorisk , kataforisk eller exoforisk (deiktisk) anvendelse	DSS	Pronominet forekommer i et ikke-sætningsformet sammenligningsled indledt med "end" eller "som"	RT	Dyr eller ting (ental)
AA	Anden anvendelse	DSF	Pronominet forekommer før subjektet	RG	Generisk
AI	Ikke-fuldført konstruktion	DLA	koreference inden for " akkusativ med (at-)infinitiv"	RU	Ubestemt
		DLI	koreference mellem direkte og indirekte objekt	RF	Flertal
		DLO	koreference mellem direkte objekt og prædikativ	RA	Anden
		DLP	koreference mellem direkte objekt og en til sætningen hørende præpositionsforbindelse		
		DLN	koreference mellem adled og kerneled i nominalhypotagme eller mellem nominalled i apposition		
		DLØ	koreference mellem pronomen og anden form for "logisk subjekt"		
		DIF	Pronomenet indgår i en fast vending		
		DII	Pronomenet er koreferentielt med et implicit (oftest generisk) "subjekt" i en infinitiv		
		DBO	Pronominet er koreferentielt med både det grammatiske subjekt og en af de ovenfor nævnte typer af logiske subjekter		

*Domænetags kan inddeles i tre kategorier: grammatisk subjekt (S), logisk subjekt (L), implicit (I)

Eksempel på sekvenskodning: Når eksempelvis en token er annoteret med *G_AI*, vil dette token få følgende tag: *1_G_AI_E*. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.4.2 Gramma_II (gramma_II)

- Kontaktperson: Torben Juel Jensen: tjuelj@hum.ku.dk.
- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Status: Forskerniveau.

I dette tier er der kodet for generiske pronomener *den/en* samt alle forekomster af *ens, dens, dig, din, dit* og *dine*. Dette er gjort efter analyseapparatet "generisk" der er beskrevet i afsnittet "Generisk" nedenfor. Alle generisk-koder fra dette tier findes også i tieret "Generisk".

Eksempel på sekvenskodning: Når eksempelvis en token er annoteret med *G_AB_DI*, vil dette token få følgende tag: 1_G_AB_DI_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

Generisk (generisk)

- Kontaktperson: Torben Juel Jensen: tjuelj@hum.ku.dk.
- Ordkodet
- Status: Forskerniveau.

Dette tier indeholder alle generisk-koder fra tiersne "grammatik" og "gramma_II".

Analyseapparat for forekomster af pronominer med generisk betydning

Alle forekomster af søgestrengene, der repræsenterer pronominer, er markeret med et "G" samt som minimum delkoden for "anvendelse", dvs. "G_AS", "G_AA", "G_AI" osv. I forbindelse med de forekomster, der kategoriseres som andet end "Anden anvendelse" og "Ikke-fuldført", er delkoden for "diskursfunktion", dvs. "G_AS_DM", "G_AS_DS" eller "G_AS_DI" desuden tilføjet. For de to førstnævnte typer er der derefter tilføjet delkoderne for "syntaks" og "reference", fx "G_AS_DS_SB_RA".

Ved tvivlstilfælde anvendes bogstavet "X" på den pågældende plads, fx "G_AX" eller "G_AS_DX_SA_RX". Ved tvetydige forekomster, anvendes der to koder (i alfabetisk rækkefølge), fx "G_AS_DS_SA_RAJ" eller "G_AS_DMS_SA_RA".

Kategorier

Anvendelse		Diskursfunktion	
Tag	Beskrivelse	Tag	Beskrivelse
AS	Pronomen fungerer som grammatisk subjekt	DM	Truisme eller morale
AO	Pronomen fungerer som grammatisk objekt (indirekte eller direkte) eller prædikativ	DS	Generalisering af situation
AP	Pronomen fungerer som styrelse i en præpositionsforbindelse (også middelbare objekter)	DI	Ikke-generaliserende anvendelse
AB	Pronomen fungerer som bestemmerled i et nominalled		
AA	Anden anvendelse		

AI	Ikke-fuldført konstruktion
----	----------------------------

Syntaks		Reference	
Tag	Beskrivelse	Tag	Beskrivelse
SB	I betingelseskonstruktion	RA	Alle mennesker eller en afgrænset gruppe af mennesker, som indbefatter både taler og adressat
SP	I komplementsætning (ofte en substantivisk ledsætning) til en projicerende sætning	RJ	Afgrænset gruppe af mennesker, som indbefatter taler (jeg), men ikke adressat
SA	Andet	RD	Afgrænset gruppe af mennesker, som indbefatter adressat (du), men ikke taler
		RH	Afgrænset gruppe af mennesker, som hverken indbefatter taler eller adressat

2.4.3 Ordstilling (ordstil)

- Kontaktperson: Torben Juel Jensen: tjuelj@hum.ku.dk.
- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Status: Forskerniveau.

Analyseapparat for ordstilling i underordninger (ledsætninger)

Kategorier:

L: ledsætning, F: Funktion, O: Ordstilling, M: matrixsætning, _: Andet.

Kodning

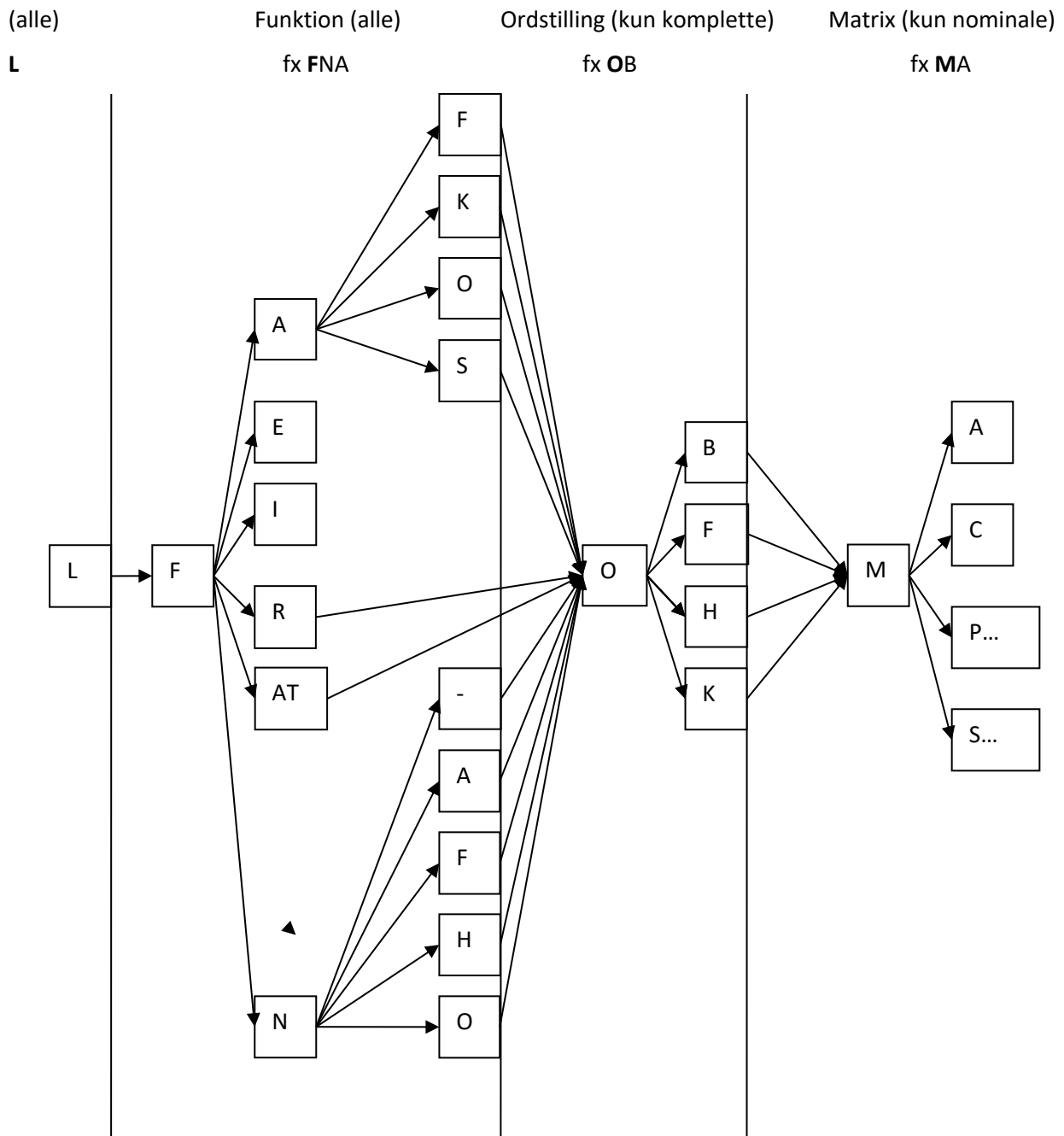
Alle forekomster af underordninger med sætningsadverbier kodes med et "L" samt som minimum koden for **Funktion**, dvs. som minimum "L FE" eller "L FI". Funktionerne "ellipse" og "ikke fuldført" kodes ikke yderligere.

Ved tvivlstilfælde anvendes bogstavet "X" på den pågældende plads, fx "L FX", "L FNA OX MA PI Aikke" eller "L FNA OH MX PI Aikke". X-koder bør dog ikke findes i de færdigkodede filer. Hvis en forekomst udtrykker to ting på en gang, anvendes der to koder i alfabetisk rækkefølge, fx "L FAO OBF PI Aikke". Der anvendes dog kun dobbeltkoder i forbindelse med ordstilling og matrixsætning.

Ved underordninger som er indlejrede i andre underordninger, sættes parenteser om koderne, også om koden for den "overordnede" underordning, fx "(L FAK OB PI Aikke(L FNA OB MA PI Aikke) (LFN- OB MA PI Aikke))", jf. s. **Error! Bookmark not defined.**

En komplet beskrivelse af de koder der anvendes kan findes i dokumentet "Analyseapparat(ordstilling)".

Følgende er en oversigt over de koder de anvendes:



Eksempler:

(L FATNN\at OB A\ikke SL3P\de PRL\glemme FPS PF D0 L9 II FVN SPN))

(L FNA OH MA PF Aaltså\bare PRhave\brug\for TN S3Pde AMrentfaktisk SYST

L FAK\fordi OH A\jo SL3P\alle\maskiner PRL\være\styre FPS PF D0 L25 II FVF SPN
 L FAK\fordi\at OH A\egentlig SL1S PRL\tænke FPS PF DI L44 IU FVF SPN
 L FAO\hvis OB A\bare\lige SL1S PRL\blive\sætte\ind\i FPS PF D0 L18 II FVN SPN
 L FAO\nÅr OB A\ikke SL3S\hun PRL\vÅ're FPS PU D0 L6 II FVF SPN
 L FI
 L FN- OH MA PF Abare PRbestemme TTFD S1P AMegentlig SYO
 L FN- OH MCA PF Aikke
 L FN- OH MCA PF Ajo\ikke
 L FN- OH MSA PF Ajo
 L FN- OH MSA PF Asgu\da\fÅ_rst
 L FN- OK MSA PK Anok
 L FNA OH MA PF Agerne PRkigge\på TN S3Sman AMegentlig SYST
 L FNA OH MA PF Agodt PRhave\ngt TTD S1P AM0 SYO
 L FNA OH MA PF Aikke PRsige TTD S1S AMbare SYO
 L FNH OB MA PF Aikke PRsige TN S3Pde AM0 SYO

Eksempel på sekvenskodning: Når eksempelvis ni tokens er annoteret med *L FAK\fordi OH A\jo SL3P\de PRL\have FPS PF PR\0 S\0 AM\0 DD L9 KU IU FVF SPN*, vil de ni tokens få disse ni tags:

1_L_FAK\fordi_OH_A\jo_SL3P\de_PRL\have_FPS_PF_PR\0_S\0_AM\0_DD_L9_KU_IU_FVF_SPN_I,
 2_L_FAK\fordi_OH_A\jo_SL3P\de_PRL\have_FPS_PF_PR\0_S\0_AM\0_DD_L9_KU_IU_FVF_SPN_I,
 3_L_FAK\fordi_OH_A\jo_SL3P\de_PRL\have_FPS_PF_PR\0_S\0_AM\0_DD_L9_KU_IU_FVF_SPN_I, [...],
 8_L_FAK\fordi_OH_A\jo_SL3P\de_PRL\have_FPS_PF_PR\0_S\0_AM\0_DD_L9_KU_IU_FVF_SPN_I,
 9_L_FAK\fordi_OH_A\jo_SL3P\de_PRL\have_FPS_PF_PR\0_S\0_AM\0_DD_L9_KU_IU_FVF_SPN_E.

Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.4.4 Participium (Part)

- Kontaktperson: Torben Juel Jensen: tjuelj@hum.ku.dk.
- Ordkodet
- Status: Forskerniveau.

Undersøgelsen er et samarbejde mellem grammatikgruppen og fonetikgruppen. Det, der undersøges, er brugen af -en i præteritum participium.

For den fonetiske del af undersøgelsen: se artiklen Fonetisk variantsætning for yderligere oplysninger. For den grammatisk del af undersøgelsen: se artiklen Analyseapparat(participier).

Følgende koder anvendes:

Hver kode består af en kombination af to bogstaver, der beskriver funktionen og et tal (eller U) der beskriver udtalen:

FV	verbal funktion
FU	uafgørelig (det kan ikke entydigt afgøres om participiet har verbal funktion, dvs. om det betegner en handling/begivenhed eller en beskaftenhed)
FA	anden funktion end verbal (dvs. adjektivisk eller substantivisk)
FI	forekomster i ikke-fuldførte (dvs. uanalyserbare) konstruktioner

1	symboliserer 'blødt d'
2	symboliserer 'hårdt d'
3	symboliserer '-en'
0	symboliserer schwa eller ingen endelse
U	ikke angivet

Eksempler:

FA U

FI U

FU 0

FU 1

2.4.5 Epistemisk sætning (epistsætn)

- Ordkodet
- Kontaktperson: Tanya Karoli Christensen: tkaroli@hum.ku.dk
- Status: Forskerniveau

I dette tier beskrives de epistemiske (inklusive evidentielle) verber og prædikative konstruktioner med epistemiske adjektiver: *tro, tænke, mene, vide og givet, klart, sikkert, tydeligt, indlysende, oplagt, muligt, sandsynligt, uklart, usikkert, umuligt, usandsynligt*.

For hvert af disse ord kodes forekomsten (F), konstruktionstypen (K), Ledstilling (L), Placering af den epistemiske ledsætning (P), Tempus (T), Subjekt (S), Sætningsadverbial (A), Betydning (B) samt Forhold i den evt. tilknyttede sætning (C)

Alle kodninger starter med "ES".

Forekomst

Her skrives det givne ord. Ved faste udtryk skrives hele udtrykket. Eksempelvis *Ftro, Fdet/vil/sige*

Konstruktionstype

Der anvendes følgende koder:

KN	Prædikat med NP-objekt
KS	Prædikat med sætningsobjekt
KSK	Prædikat med sætningsobjekt i form af sætningsknode
KSE	Prædikat med elliptisk sætningsobjekt
KO	Prædikat der ikke har sætningsobjekt
KX	Tvivl

Ledstilling

Der anvendes følgende koder:

LOVS	Objekt (NP), finit verbal, subjekt
LSVO	Subjekt, finit verbal, objekt (NP)
LSV	Subjekt, finit verbal (uden NP-objekt og ved KX)
LVS	Finit verbal, subjekt (uden NP-objekt og ved KX)
LA	Andet
LX	Tvivl

Placering af den epistemiske ledsætning

Der anvendes følgende koder:

PI*	Placering initialt ift. den styrede/modificerede sætning
PM	Placering medialt ift. den styrede/modificerede sætning
PF	Placering finalt ift. den styrede/modificerede sætning
PA	Placering alene (dvs. styrer/modificerer ikke en anden sætning)
PX	Tvivl

*PI har en række underkodninger:

PIk	initial placering og underordningskonjunktion
PIØ	initial placering og udeladt underordningskonjunktion
PIO	initial placering uden mulighed for underordningskonjunktion
PIX	tvivl

Tempus

Der anvendes følgende koder:

TN	Nutid
TD	Datid
TFN	Førnutid
TFD	Førdatid
TI	Infinitiv
TX	Tvivl

Subjekt

Der anvendes følgende koder:

SP[formen]	Pronomener, fx SPjeg og SPman
SL	Appellativ
SA	Andet, fx infinitiv som <i>At rejse er at leve.</i>
S0	Intet
SX	Tvivl

Sætningsadverbial (i sætningen med det epistemiske prædikat)

Der anvendes følgende koder:

A[form]	Skriv formen, fx <i>Aikke</i> eller <i>Aslet/ikke</i>
A0	Intet sætningsadverbial
AX	Tvivl

Betydning

Der anvendes følgende koder:

BU	epistemisk usikkerhed
BS	epistemisk sikkerhed
BA	Andre slags betydning end epistemisk
BX	Bruges i tilfælde af KSE hvor det ikke er muligt at afgøre om der er tale om BU eller BS.

Forhold i den evt. tilknyttede sætning

Der anvendes følgende koder:

CP	Prædikat i den tilknyttede sætning, dvs CP[lemma]. Hvis der er både en ledsætning og en helsætning, skrives helsætningens prædikat først, og ledsætningens i en parentes derefter.
----	--

CT	Tempus, med underkoderne TN, TD, TFN, TFD. Evt. ledsætnings tempus skrives i parentes. TI + SO= imperativ T0 + verbal CP + SO = infinitiv
CS	Subjekt, med underkoderne SP, SL, SA, SO, SX (fx CSPjeg). Evt. ledsætnings subjekt skrives i parentes.
CA	Evt. adverbialer i den tilknyttede sætning, fx CAjo/ikke. Evt. ledsætnings adverbialer skrives i parentes.
CL	Ledstilling i den tilknyttede sætning (de to-tre første pladser i sætningsskemaet), fx CLSV
	CLØVS: Hvor der er tomt forfelt
	CLHVS: Hv-led i forfeltet
	CLXVS: Kodes når der er noget andet end subjektet i forfeltet
	CLOVS: Ved knuder angives den del af ledstillingen der ikke allerede fremgår af L-koden for den epistemiske sætning

For yderligere information se kodningsmanualen *Kodningsmanual for epistemiske sætninger 23 02 15.doc*

Eksempler:

Det ved jeg ikke hvad han havde set

ES Fvide KS LOVS Plk TN SPjeg Aikke BU CPse/have CTFD CSPhan CA0 CLSV

(dobbelt repræsentation af argument)

Majbritt tror jeg hun hed.

ES Ftro KSK LVS PIØ TN SPjeg A0 BU CPhedde CTD CSPhun CA0 CLXSV

Det er svært at sige, fordi ..

ES Fsvært/at/sige K0 LSV PA TN SPdet A0 BU

2.4.6 semvar epistemicitet (semvar)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Kontaktperson: Tanya Karoli Christensen: tkaroli@hum.ku.dk
- Status: Forskerniveau.

I dette tier kodes sætningsadverbier for epistemicitet. En del af disse epistemiske sætningsadverbier er polyseme og/eller homografe med andre ord. Sådanne ord er det kodet med tagget "LA".

Sætningsadverbierne kodes for kategorierne Forekomst (F), Tempus (T), Verbum (V), Subjekt (S), Negation (N), Ledtype (L), Betydning (B) og Placering (P)

Alle koder starter med "E".

Forekomst

Her skrives det fremsøgte ord, eksempelvis *Fbestemt*, *Fgangskegivet*

Tempus

Der anvendes følgende koder:

TN	nutid
TD	datid
TFN	førnutid
TFD	førdatid
TX	tvivl

Verbum

Hovedverbets lemma skrives efter koden V. Ved sideordningskonstruktioner (fx "går og synger") skrives begge lemmaer med og imellem (*Vgåogsynge*). Eventuelle modalverber skrives efter hovedverbet, eksempelvis *Vsekunne* (fra sætningen *det kan de vel godt se*).

Subjekt

Subjektet i den sætning som søgeordet forekommer i. Der anvendes følgende koder:

SP[form]	Pronomener. Pronoments form angives
SL	Appellativer (proprier og substantiver)
SA	Andet
S0	Intet subjekt
SX	Tvivl

Negation

Der anvendes følgende koder:

N[form]	Negationens form angives
N0	Ingen negation

Ledtype

Der anvendes følgende koder:

LP	Modalpartikel af formen <i>nok, vel</i> eller <i>vist</i>
LS	Sætningsadverbium
LA	Andet (alle ikke-epistemiske sætningsadverbier)

Betydning

Der anvendes følgende koder:

BS	Vidensmæssig sikkerhed
BU	Vidensmæssig usikkerhed
BX	Tvivl

Placering

Der anvendes følgende koder:

PF	Fundamentfelt
PA	Sætningsadverbialplads
PH	Ekstraposition til højre
PV	Ekstraposition til venstre
PU	Uden finit sætning
PI	Infinit sætning
PX	Tvivl

Eksempler på kodninger:

ja **klart nok** (bysoc1ny-05-JKS)

E Fklartnok LP BS PU

nu er det godt **nok** noget siden jeg sidst har været derovre

E Fnok LA (koncessiv betydning)

men altså min mor havde **tilsyneladende** planlagt det

E Ftilsyneladende TFD Vplanlægge SA N0 LS BU PA

det er sgu ikke **sikkert**

E Fsikkert TN Vvære SPdet Nikke LS BU PA

og **sandsynligvis** så er der ikke noget om det
E Fsandsynligvis TN Vvære SA Nikke LS BU PV

Eksempel på sekvenskodning: Når eksempelvis en token er annoteret E Foverhovedet V T S N LA B P, vil dette token få følgende tag: 1_E_Foverhovedet_V_T_S_N_LA_B_P_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.4.7 Ledsætning (ledsaet)

- Kontaktperson: Torben Juel Jensen: tjuelj@hum.ku.dk.
- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Status: Forskerniveau.

Kodning:

Alle forekomster af underordninger kodes med et "L" samt som minimum koden for **Funktion**, dvs. som minimum "L FE" (for "ellipse") eller "L FI" (for "ikke fuldført"). Funktionerne "ellipse" og "ikke fuldført" kodes ikke yderligere.

Ved forekomster af alle andre funktioner, dvs. "adverbiel": "L FA"; "relativ": "L FR", "samordning": "FS" eller "nominal": "L FN", markeres tillige disses underkoder, fx "L FAS", "L FRE", "L FNH".

Derudover tilføjes koden for **Ordstilling**, fx "L FRE ON".

I forbindelse med de underordninger der kategoriseres som "nominale", tilføjes tillige koden for **Matrix-sætning**, fx "L FNA OH MP".

Det er vigtigt at sætte mellemrum mellem delkoderne som angivet ovenfor!

Ved tvivlstilfælde anvendes bogstavet "X" på den pågældende plads, fx "L FX", "L FNT OX MA" eller "L FNA OH MX". X-koder bør dog ikke findes i de færdigkodede filer. Hvis en forekomst udtrykker to ting på en gang, anvendes der to koder (i alfabetisk rækkefølge), fx "L FAO OHU".

Ved underordninger som er indlejrede i andre underordninger, sættes parenteser om koderne, også om koden for den "overordnede" underordning, fx "(L FAK ON (L FNA ON MA) (LFN- ON MA))".

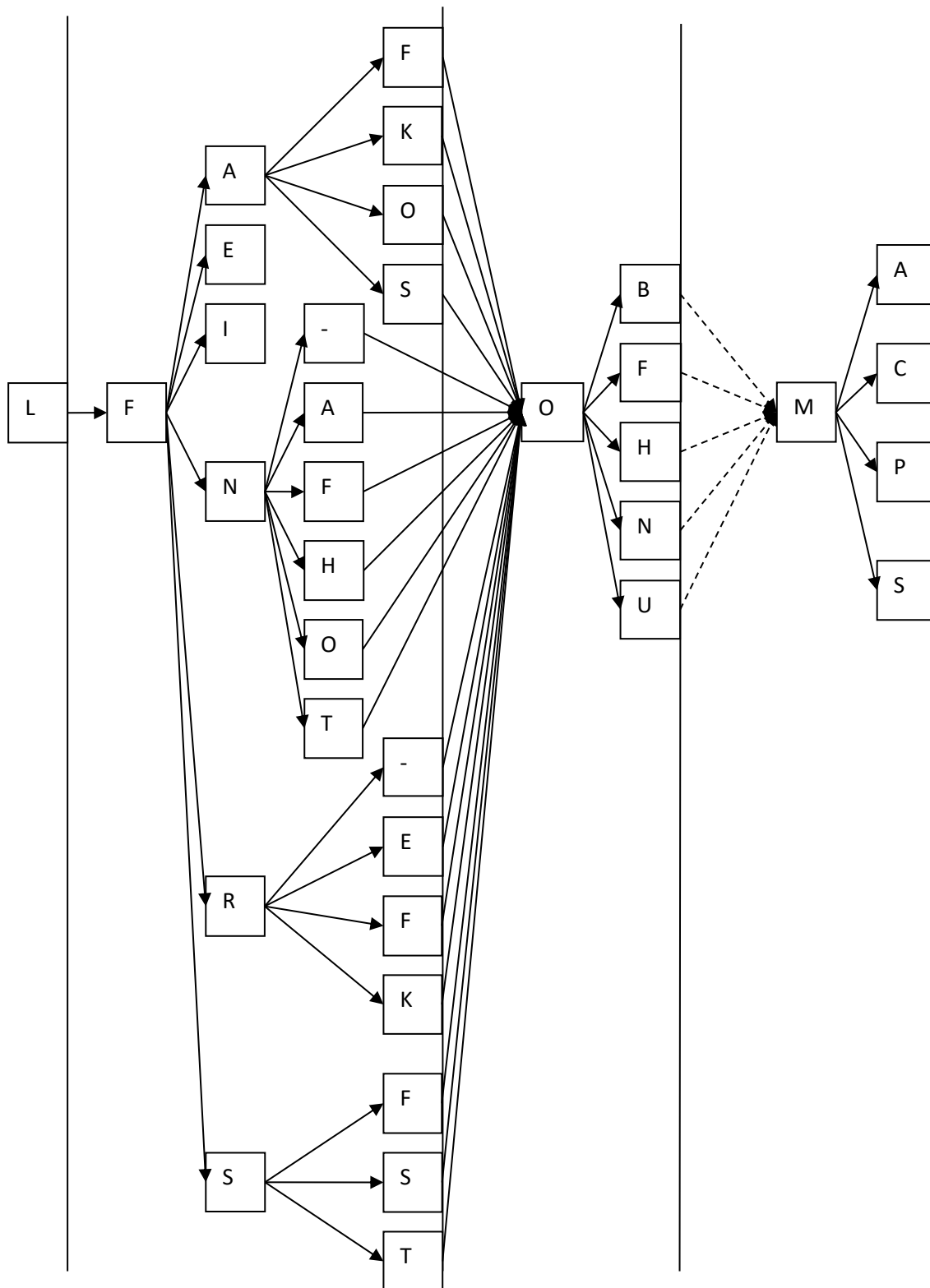
Tieret findes kun i de eksplorative filer (jf. afsnittet "Metadata på samtaleniveau").

En komplet oversigt over de forskellige koder kan findes i dokumentet "Analyseapparat(ledsætninger) 18 04 07"

Følgende er en oversigt over de koder der anvendes:

(alle) Funktion (alle) Ordstilling (kun komplette) Matrix (kun nominale)

L fx FNA fx ON fx MP



Analyseapparat for forekomster af pronominer med generisk betydning.

Alle forekomster af søgestrengene, der repræsenterer pronominer, er markeret med et "G" samt som minimum delkoden for "anvendelse", dvs. "G AS", "GAA", "G AI" osv. I forbindelse med de forekomster, der kategoriseres som andet end "Anden anvendelse" og "Ikke-fuldført", er delkoden for "diskursfunktion", dvs. "G AS DM", "G AS DS" eller "G AS DI" desuden tilføjet. For de to førstnævnte typer er der derefter tilføjet delkoderne for "syntaks" og "reference", fx "G AS DS SB RA".

Ved tvivlstilfælde anvendes bogstavet "X" på den pågældende plads, fx "G AX" eller "G AS DX SA RX". Ved tvetydige forekomster, anvendes der to koder (i alfabetisk rækkefølge), fx "G AS DS SA RAJ" eller "G AS DMS SA RA".

Kategorier:

Anvendelse		Diskursfunktion	
Tag	Beskrivelse	Tag	Beskrivelse
AS	Pronomen fungerer som grammatisk subjekt	DM	Truisme eller morale
AO	Pronomen fungerer som grammatisk objekt (indirekte eller direkte) eller prædikativ	DS	Generalisering af situation
AP	Pronomen fungerer som styrelse i en præpositionsforbindelse (også middelbare objekter)	DI	Ikke-generaliserende anvendelse
AB	Pronomen fungerer som bestemmerled i et nominalled		
AA	Anden anvendelse		
AI	Ikke-fuldført konstruktion		

Syntaks		Reference	
Tag	Beskrivelse	Tag	Beskrivelse
SB	I betingelseskonstruktion	RA	Alle mennesker eller en afgrænset gruppe af mennesker, som indbefatter både taler og adressat
SP	I komplementsætning (ofte en substantivisk ledsætning) til en projicerende sætning	RJ	Afgrænset gruppe af mennesker, som indbefatter taler (jeg), men ikke adressat
SA	Andet	RD	Afgrænset gruppe af mennesker, som indbefatter adressat (du), men ikke taler
		RH	Afgrænset gruppe af mennesker, som hverken indbefatter taler eller adressat

Eksempel på sekvenskodning: Når eksempelvis ti tokens er annoteret (*L FNA ON MA*, vil de ti tokens få disse ti tags: 1_(L_FNA_ON_MA_I, 2_(L_FNA_ON_MA_I, 3_(L_FNA_ON_MA_I, [...], 9_(L_FNA_ON_MA_I, 10_(L_FNA_ON_MA_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.4.8 GEX Generel udvider (GEX)

- Kontaktperson: Torben Juel Jensen: tjuelj@hum.ku.dk.
- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Status: Forskerniveau.

Se dokumentet "Analyseapparat for general extenders" for yderligere informationer

General extenders er udtryk der prototypisk består af en sideordningskonjunktion (*og/eller*) og et generisk nominalsyntaxme der refererer til noget i det umiddelbart forudgående, og i udvidet forstand udtryk der udfylder samme funktion: at fremkalde forestillingen om en større mængde/kategori ud fra tilhørers for-modede viden om det der refereres til. Under alle omstændigheder er udtrykket efterstillet.

Kodning

Alle forekomster af udtryk der fungerer som general extenders, er opmærket med et interval i GEX-tieret. I labelen angives som minimum delkoden for "udtryk", fx "UI" eller "UNN\og\alt\sådan\noget". Medmindre forekomsten kategoriseres som "UI", er delkoderne for "øvrige led i syntagme (L*)", "placering i ytring (Y*)" og "turskifte (T*)" desuden tilføjet, fx "UNA\og\hele\lortet LNIF\2 YF T0".

Ved tvivlstilfælde anvendes bogstavet "X" på den pågældende plads, fx "UX" eller "UNA\og\hele\lortet LNIF\2 YX T0".

Kategorier:

U: Udtryk (type og ortografisk repræsentation af ordene med ' \' i stedet for mellemrum)

L: Øvrige led i syntagme (type og antallet af forudgående led adskilt af " \")

Y: Placering i ytring

T: Turskifte

Udtryk		Øvrige led i syntagme	
Tag	Beskrivelse	Tag	Beskrivelse
UNN	Nominalsyntaxme med noget/nogen/nogle som kerne	LNP(E\F)	Syntaxmets øvrige led er nominaler eller nominalsyntaxmer og refererer til personer eller andre agentive størrelser
UNT	Nominalsyntaxme med ting/alt-ting som kerne	LNI(E\F)	Syntaxmets øvrige led er nominaler eller nominalsyntaxmer og refererer til individuerede størrelse
UNA	Nominalsyntaxme med anden kerne	LNМ	Syntaxmets øvrige led er nominaler eller nominalsyntaxmer og refererer til ikke-individuerede masser

UA	Adverbialsyntagme	LPAP	Syntagmets øvrige led er andre former for prædikativer end nominalprædikativer
UP	Præpositionsforbindelse	LPØ	Syntagmets øvrige led udgøres af øvrige typer af prædikater end prædikativkonstruktioner
UØ	Øvrige udtryk	LC	Syntagmets øvrige led udgøres af selvstændige ytringer der fungerer som egentligt eller tænkt citat
UI	Ikke fuldført, eller forekommer i en ytring der afbrydes før en mening er udtrykt	LB	Syntagmets øvrige led udgøres af flere forskellige af ovennævnte typer (blandet)
		LA	Syntagmets øvrige led er af anden type end de ovennævnte
		LU	Syntagmets øvrige led er uidentificerbare, eller kan ikke afgrænses entydigt

Placering i ytring		Turskifte	
Tag	Beskrivelse	Tag	Beskrivelse
Y0	Syntagmet og GEX'en udgør i sig selv en ytring (nul)	T0	Taleren fortsætter uden bidrag fra tilhører, dvs. nul turskifte
YI	GEX'en er placeret initialt i ytringen, dvs. før det finitte verbal, men efter det syntagme det lægger sig til	TP	Taleren fortsætter uden bidrag fra tilhører, men efter en pause
YM	GEX'en er placeret medialt i ytringen	TM	Taleren beholder turen men der er et minimalbidrag fra tilhører
YF	GEX'en er placeret finalt i ytringen	TOV	Taleren fortsætter men overlappes et stykke tid af tilhøreren
YU	Placeringen i ytringen er uafgørlig	TS	Turskifte
		TU	Syntagmets placering i forhold til turtagning er uafgørligt

Eksempler:

UA\eller\hvad LNPE\1 Y0 T0

UA\og\så\videre LPAP\1 YF T0

UA\og\så\videre LPØ\1 YF TM

UA\og\sådan LPØ\1 YFI TP

UNA\eller\et\eller\andet LB\2 YF T0

UNA\eller\sådan\et\eller\andet LNPE\1 YF T0

UNN\eller\noget LA\1 YF TM
UNN\eller\sådan\noget LA\2 YF TM
UNN\eller\sådan\noget LC\1 YF TP
UNN\og\alt\sådan\noget\underligt\noget LPØ\1 YF TS
UNN\og\sådan\noget LA\1 YI TO
UNN\og\sådan\noget LNIF\1 YF TO
UNN\og\sådan\noget LNIF\1 YF TM

Eksempel på sekvenskodning: Når eksempelvis to tokens er annoteret *UA\og\sådan LPØ\1 YF TP*, vil de to tokens få disse to tags: 1_UA\og\sådan_LPØ\1_YF_TP_I, 2_UA\og\sådan_LPØ\1_YF_TP_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.4.9 at-tab-...

- Kontaktperson: Kasper Boye: boye@hum.ku.dk

Der er et antal forskellige tirs der har at gøre med tab af underordningskonjunktionen *at*. Det har indtil videre ikke været muligt at tilvejebringe en fyldestgørende beskrivelse af disse. Tirsne er:

- at-tab-ekspl
- at-tab-ekspl-komm
- at-tab-vide
- at-tab-vide-komm
- at-tab-mene
- at-tab-sige
- at-tab-sige-komm
- at-tab-kunnehuske
- at-tab-adj
- at-tab-nogetmed
- at-tab-tro
- at-tab-tro-komm

2.4.10 Case

- Ordkodet
- Status: Åben
- Kontaktperson: Jeffrey Parrott

Omhandler kasusvariation i personlige pronomener der forekommer i samordnede substantivfraser. Både pronomener og konjunktioner kodes.

Der kodes for person (P), numerus (N), køn (G), kasus (Ca), *conjunct* (Co) og *coordinate head* (C0) (selve konjunktionen):

Person

Der anvendes følgende koder:

P1	første person
P2	anden person
P3	tredje person

Numerus

Der anvendes følgende koder:

Nsg	ental
Npl	flertal

Køn

Der anvendes følgende koder:

Gfe	female (hunkøn)
Gma	male (hankøn)

Kasus

Der anvendes følgende koder:

CaSF	subjekt-form
CaOF	oblik form

Conjunct

Der anvendes følgende koder:

Co1	første led i den samordnede substantivfrase
Co2	andet led i den samordnede substantivfrase
Co3	tredje led i den samordnede substantivfrase
Con	led nummer <i>n</i> i den samordnede substantivfrase

Coordinate head (selve konjunktionen)

Her angives substantivfrasens funktion i sætningen. Denne kategori har en lang række underkategorier. Den samlede liste kan ses i kodningsmanualen *Skelton key 2.doc*.

Eksempel:

<i>hende</i>	<i>og</i>	<i>jeg</i>	<i>drikker</i>	<i>øl</i>
P3 Nsg Gfe CaOF Co1	C0 SM	P1 Nsg CaSF Co2		

2.4.11 GIDDY Gender in Danish and Dutch (GIDDY)

- Ordkodet
- Kontaktperson: Kontaktperson: Torben Juel Jensen: tjuelj@hum.ku.dk (oprindelig Frans Gregersen)
- Projekt: GIDDY (Gender in Danish and Dutch)

I dette tier beskrives det grammatiske køn anvendt i substantiver, adjektiver, pronomener og i konstruktioner. Hvert tag viser ordklasse samt det/de givne ord. Derudover angives værdier for relevante kategorier – for en stor del af tag-sættet indebærer det det forventede og realiserede køn.

Der er anvendt følgende tags:

Substantiver

- **Determinatorer** (indefinit, singularis)
 - DUU[lemma]: standard **U**trum, usus **U**trum fx DUUstol = *en stol*
 - DNN[lemma]: std. **N**eutrum, us. **N**eutrum fx DNNbord = *et bord*
 - DUN[lemma]: std. **U**, us. **N** fx DUNstol = *et stol*
 - DNU[lemma]: std. **N**, us. **U** fx DNUbord = *en bord*
 - DT[evt. lemma]: tvivlstilfælde
- **Bøjningssuffikser** (definit, singularis)
 - BUU[lemma]: std. **U**, us. **U** fx BUUstol = *stolen*
 - BNN[lemma]: std. **N**, us. **N** fx BNNbord = *bordet*
 - BUN[lemma]: std. **U**, us. **N** fx BUNstol = *stolet*
 - BNU[lemma]: std. **N**, us. **U** fx BNUbord = *borden*
 - BIO[lemma]: nøgen form, +forvent. genus fx BIOmåde = *[på sjov] måde*
 - BI1[lemma]: nøgen form, ÷forvent. genus fx BI1skole = *[gå i] skole*
 - BT[evt. lemma]: tvivlstilfælde

Adjektiver (singularis)

- AUUA(*attributiv*)[lemma_referent]: std. **U**, us. **U** fx AUUAgod_stol = *[en] god [stol]*
- AUUAA[lemma_referent]: std. **U**, us. **U**, afvigende artikel fx *et god stol*
- AUUUA[lemma_referent]: std. **U**, us. **U**, uden art fx *gå med rød vest*
- AUUP(*prædikativ*)[lemma_referent]: std. **U**, us. **U** fx AUUPgod_stol = *[stolen er] god*
- ANNA[lemma_referent]: std. **N**, us. **N** fx ANNAgod_bord = *[et] godt [bord]*
- ANNAA[lemma_referent]: std. **N**, us. **N**, afv. Art. fx *en godt bord*
- ANNUA[lemma_referent]: std. **N**, us. **N**, u art fx *gå i sjovt tøj*

- ANNP[lemma_referent]: std. N, us. N
 - AUNA[lemma_referent]: std. U, us. N
 - AUNAA[lemma_referent]: std. U, us. N, afv. Art.
 - AUNUA[lemma_referent]: std. U, us. N, u art
 - AUNP[lemma_referent]: std. U, us. N
 - ANUA[lemma_referent]: std. N, us. U
 - ANUAA[lemma_referent]: std. N, us. U afv. Art.
 - ANUUA[lemma_referent]: std. N, us. U u art
 - ANUP[lemma_referent]: std. N, us. U
 - AIA [item_referent]: attributivt, ÷genusmærke
 - AIAA: attributivt, -genusmærke, afvigende artikel
 - AIP[item_referent]: prædikativt, ÷genusmærke
 - AØA[lemma_referent]: ikke afgørlig grundet kontekst
 - AT[evt. lemma]: tvivlstilfælde
- fx ANNPgod_bord = *[bordet er] godt*
fx AUNAgod_stol = *[en] godt [stol]*
fx et godt stol
fx gå med rødt vest
fx AUNPgod_stol = *[stolen er] godt*
fx ANUAgod_bord = *[et] god [bord]*
fx en god bord
fx gå med sjov tøj
fx ANUPgod_bord = *[bordet er] god*
fx AIAlette_bord = *[det] lette [bord]*
fx AIPhele_ABSTRAKT= *[det] hele*
fx en halv time

Pronominer (singularis)

- Pronominelle referencer til nominaler
 - PUUA[lemma_referent]: std. U, us. U
 - PUUP[lemma_referent]: std. U, us. U
 - PNNA[lemma_referent]: std. N, us. N
 - PNNP[lemma_referent]: std. N, us. N
 - PUNA[lemma_referent]: std. U, us. N
 - PUNP[lemma_referent]: std. U, us. N
 - PNUA[lemma_referent]: std. N, us. U
 - PNUP[lemma_referent]: std. N, us. U
 - Særlige pronomine referencer
 - PF[item]: foreløbigt subjekt/pladsholder
 - PHU[lemma]: højereordensreference, us. U
 - PHN[lemma]: højereordensreference, us. N
 - PIU[lemma]: betinget reference, us. U
 - PIN[lemma]: betinget reference, us. N
 - PVU[lemma]: vag/uklar reference, us. U
 - PVN[lemma]: vag/uklar reference, us. N
 - PT[evt. lemma]: tvivlstilfælde
- fx PUUAmin_stol = *min [stol]*
fx PUUPmin_stol = *[stolen er] min*
fx PNNamin_bord = *mit [bord]*
fx PNNPmin_bord = *[bordet er] mit*
fx PUNamin_stol = *mit [stol]*
fx PUNPmin_stol = *[stolen er] mit*
fx PNUamin_bord = *min [bord]*
fx PNUPmin_bord = *[bordet er] min*
fx PFdet = *det [er i orden]*
fx PHUden = *den [er fin med dig]*
fx PHNden = *[jeg skal nok gøre] det*
fx PIUdenne = *[brug] denne [her]*
fx PINdet = *[giv mig] det [der]*

Konstruktioner (flerordsforbindelser) med genusmarkering

- K[hele konstruktionen/alle ord]
- fx Kenafvennerne = *en af vennerne*
fx Ketellerandet = *et eller andet*

Eksempler:

ANNPnormal_arabisk_SPROG

BI1ingeniør

BI1lejrskole
 BNNhold
 Knogetaf
 PNNPden_tørklæde
 PUUAden_liga
 PVNden

2.4.12 Hjælpeverber (AUX)

- Ordkodet
- Ansvarlig: Anu Laenemets
- Status: Åben

I dette tier er det sidste perfektum participium i verbalforbindelser (hovedverber) kodet. Disse verber kodes for kategorierne Hjælpeverbum (AU), Diatese (D), Argumentstruktur (R), Adverbialer (A), Type of event (E), Subjekt (S), Koordinering og ellipse (K), Afstand (O) og Udtale (U).

Hjælpeverbum:

Inventar:

AUH/_	Hjælpeverbet er 'have'. Formen skrives efter skråstregen. Flerordsforbindelser skrives med skråstreger mellem hvert ord, eks: AUH/ville/have
AUV/_	Hjælpeverbet er 'være'. Formen skrives efter skråstregen som beskrevet ovenfor.
AUB/_	Hjælpeverbet er 'blive'. Formen skrives efter skråstregen som beskrevet ovenfor.
AUF/_	Hjælpeverbet er 'få'. Formen skrives efter skråstregen som beskrevet ovenfor.
AUX	Tvivi
AUZ	Denne koder indikerer at det ikke kunne afgøres om der blev sagt 'har' eller 'er'
AUQ	Ikke-fuldført konstruktion. Hvis det er indlysende hvilket ord der mangler, kodes det pågældende ord med Q.
AUY	Hvis ordet er tagget forkert (i tieret PoS)

Diatese:

Inventar:

DP	Passiv
DA	Aktiv

Argumentstruktur:

Inventar:

RI	intransitive
RT	Transitiv
RO	Præpositionsobjekt
RP	Prædikativkonstruktion
RX	Tvivl

Adverbialer:

Note: sætningsadverbialer noteres ikke, bortset fra negationer (*ikke, aldrig*)

For hvert tag skal alle underkategorier for adverbialer skrives. Optræder en bestemt type adverbial ikke, tilføjes <Q> til tagget for at vise dette. Består et adverbial af flere ord, skrives disse uden mellemrum mellem dem.

Inventar:

AK/_	Inkorporering og overført/metaforisk betydning. Formen skrives efter skråstregen.
AR/_	Retningsadverbialer. Formen skrives efter skråstregen.
AS/_	Stedsadverbialer. Formen skrives efter skråstregen.
AT/_	Tidsadverbialer. Formen skrives efter skråstregen.
AM/_	Mådesadverbialer. Formen skrives efter skråstregen.
AZ/_	Andre typer af adverbialer. Formen skrives efter skråstregen.
AX	Tvivl
A_/Q	Uafsluttede adverbialer hvor man kan se hvilken type der er tale om, men ikke hvilke(t) ord. Q står da i stedet for det ord. Eks AK/iQ. Bemærk at denne kode er forskellig fra AKQ der indikerer at der ikke nogen adverbialer af typen AK.

Type of event:

Inventar:

ET	Telisk
EA	Atelisk
ES	State (statisk)
EX	Tvivl

Subjekt:

Subjektskoden består af to positioner: S12. Position 1 beskriver om subjektet er animat (A) eller inanimat (I), og position 2 beskriver subjektets form (substantiv eller pronomen). S vil altså være efterfulgt af enten A eller I, og derefter en kode der beskriver subjektets form. Substantiver er kodet med S og pronomener med

P. Personlige pronomener er kodet således: *jeg* = P1, *du* = P2, *han/hun* = P3, *vi* = P4, *I* = P5, *de* = P6. Upersonlige pronomener (*man*, *nogen*) er kodet med U, og relative pronomener er kodet med R.

Formelt subjekt er kodet med F i begge positioner (*SFF*), mens foreløbigt og reelt subjekt kodes med F i anden position (*SIF*).

X benyttes ved tvivlstilfælde.

SQ indikerer at der mangler et subjekt

Inventar:

Første position	
SA_	animat subjekt
SI_	inanimat subjekt
SX_	tvivl
SQ	intet subjekt
Anden position	
S_S	subjekt = substantiv
S_P1	subjekt = <i>jeg</i>
S_P2	subjekt = <i>du</i>
S_P3	subjekt = <i>han/hun</i>
S_P4	subjekt = <i>vi</i>
S_P5	subjekt = <i>I</i>
S_P6	subjekt = <i>de</i>
S_U	upersonligt pronomen
S_R	relativt pronomen
SFF	formelt subjekt
S_F	foreløbigt/reelt subjekt
S_P	subjekt = andet pronomen

Koordinering og ellipse:

Ved koordinering er fuldverberne markeret med K1 og K2 (og evt. K3) for at vise at de hører sammen. Ved ellipser er koderne i stedet KE1 og KE2 (og evt. KE3) i stedet.

Hvis der ikke er tale om koordinering/ellipse er KQ anvendt.

Afstand:

Denne kode angiver antallet af ord der er mellem hjælpeverbet og hovedverbet.

O0	nul ord mellem hjælpeverbum og hovedverbum
O1	et ord mellem hjælpeverbum og hovedverbum
O2	to ord mellem hjælpeverbum og hovedverbum
On	n ord mellem hjælpeverbum og hovedverbum

Udtale:

Her er det noteret hvilket ord der kommer umiddelbart før *er/har*

U/_	Efter skråstregen står det ord der kommer før <i>er/har</i>
UP	Dette indikerer at der er en pause før <i>er/har</i>
UQ	Dette indikerer at hjælpeverbet er noget andet end <i>er/har</i>

Eksempler:

AUB/blev D R AK AR AS AT AM AZ E S K O U

AUH/har DA RT AK AR AS AT AM AZ E S K O U

AUV/er DA RI AK/påefterløn ARQ ASQ ATQ AMQ AZQ ET SAP3 KQ O3 U/han

AUV/var DA RI AK/nediomsætning ARQ ASQ ATQ AM/fyrreprocent AZQ ET SIS KQ O1 UQ

2.4.13 Kløvning 1 (KL)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Kontaktperson: Marie Herget Christensen: wzr818@hum.ku.dk
- Projekt: Kløvninger

OBS: Det har indtil videre ikke været muligt at tilvejebringe en helt fyldestgørende beskrivelse af dette tier. Bemærk i øvrigt at det kun findes i filer med encodingfejl.

- KLdet Fdet KPA LA0 KON- FOK0 PRÆS4 K-
- KLdet Fdet KPA LAikke KONim FOK1 PRÆS4 Khvor
- KLdet Fdet KPO LAjo KON- FOKal PRÆS4 Kat
- KLdet FsÅ KPA LAfÅ_rst KONim FOKal PRÆS4 Kat

1. ((L, (L, F, KL0, KLder, KLdet, KLel, KLo, KLpeu, KLpseu, KLpseudo, L
2. FAK, FAO, Fdengang, Fder, Fdet, Fhv-, Fhvd, Fhvdad, Fhvem, Fhvor, Fhvornår, FI, Fmåske, FN-, FNA, Fnogle, Fnu, FØ, FRE, FRF, FRK, Fså, FSS
3. KMO, KP0, KPA, KPMO, KPO, KPS, KPSP, LA, meget, OB, OH, OHU, ON, ON)))
4. LA, LA0, LA1, LAalligevel, LAaltid, LAaltså, LAaltså\ikke, LAbare, LAda, LAda\bare, LAda\også, LAegentlig, LAgetnlig, LAellers, LAfaktisk, LAfaktisk\ikke, LAførst, LAgodt, osv.
5. KOim-, KON, KON-, KONek, KONex, KONi, KONim

6. FOK, FOK0, FOK1, FOK2, FOKal, FOKO, FOKØ
7. PRÆS, PRÆS2, PRÆS3, PRÆS3S, PRÆS4
8. K, K-, K., K0, Kat, Ksom, Ksom\der

Eksempel på sekvenskodning: Når eksempelvis otte tokens er *annoteret* *KLdet Fdet KPMO LA0 KON- FOK1 PRÆS2 K0*, vil de otte tokens få disse otte tags: 1_KLdet_Fdet_KPMO_LA0_KON-_FOK1_PRÆS2_K0_I, 2_KLdet_Fdet_KPMO_LA0_KON-_FOK1_PRÆS2_K0_I, 3_KLdet_Fdet_KPMO_LA0_KON-_FOK1_PRÆS2_K0_I, [...], 7_KLdet_Fdet_KPMO_LA0_KON-_FOK1_PRÆS2_K0_I, 8_KLdet_Fdet_KPMO_LA0_KON-_FOK1_PRÆS2_K0_E. Tallet før tagget angiver positionen, og bogstavet efter angiver, om det er det sidste tag i rækken eller ej.

2.4.14 Kløvning 2 (kl)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Kontaktperson: Marie Herget Christensen: wzr818@hum.ku.dk
- Projekt: Kløvninger

OBS: Det har indtil videre ikke været muligt at tilvejebringe en helt fyldestgørende beskrivelse af dette tier. Lille "kl" bør formentlig slå sammen med store "KL".

KL0X
KLel
KLdet_Fdet_KPMO_LA0_KON-_FOK1_PRÆS4(KLdet_TMP_TLSP_Fdet_KPO_NP_AN0_DEFI_PAT_TING_Vvære_KØ_PRÆS1_OU)
KLdet_Fdet_KPO_LA0_KON-_FOK2_PRÆS4(KLdet_TMPR_TLSPR_Fdet_KPO_AP_Vvære_KØ_PRÆS1_OU)___
KLder
KLdet_Fdet_KPO_LA0_KON-_FOK1_PRÆS4(KLdet_TMPR_TLSPR_Fdet_KPO_NP_AN0_DEFI_PAT_TING_Vvære_KØ_PRÆS1_OU)
KLdet_Fdet_KPMO_LA0_KON-_FOK0_PRÆS4(KLdet_TMPART_TLSPART_Fdet_KPMO_NP_AN0_DEFI_ØV_TING_Vvære_KØ_PRÆS1_OU)
KL0an
KLdet_Fhv-_KPO_LA0_KON-_FOKØ_PRÆS2(KLdet_TMP_TLSP_Fhv_KPO_XP_Vvære_KØ_Præs1_OU)
KLdet_Fdet_KPMO_LA0_KONex_FOK1_PRÆS3(KLdet_TMP_TLSP_Fdet_KPMO_NP_AN0_DEFD_PAT_AB_Vvære_KØ_PRÆS1_OU)_
KLdet_Fhv-_KPO_LA0_KON-_FOKØ_PRÆS2(KLdet_TMP_TLSP_Fhv_KPO_XP_Vvære_KØ_Præs1_OU)
KLdet_Fdet_KPO_LA0_KON-_FOK1_PRÆS4(KLdet_TMPR_TLSPR_Fdet_KPO_NP_AN0_DEFI_PAT_IND_Vvære_KØ_PRÆS1_OU)

Pladser i opmærkningen:

1. KL (kløvningstype?): 0X, el, det, der, 0an
2. F (fokus?): det, hv-
3. KP (komplement?): MO, O, 0
4. LA: jo, 0
5. KON-:
6. FOK: 0, 1, 2, Ø
7. PRÆS: 2, 3, 4

I parentes:

1. KL: det
2. TM: P, PR, PART,
3. TLS: P, PR, PART
4. F: det, hv
5. KP: O_NP, O_AP, O_XP, MO_NP ..
6. V: være
7. K: Ø
8. PRÆS: 1
9. O: U

2.5 Dialekttiers

- Kontaktperson: Malene Monka: monka@hum.ku.dk
- Status: Forskerniveau

NB: I Odder hedder de tiers, hvor der er markeret 32 forekomster endelser_32, ordmarkering_32 og blodt_d_32, da der for nogle informanternes vedkommende også ligger fuldkodede interviews

2.5.1 Artikel (artikel)

- Ordkodet

Undersøgelse af foransat artikel.

Inventar:

- F = at der er oprettet en automatisk kodning af et script, nogle af ordene egner sig ikke eller falder uden for den ratio, der kodes indenfor.
- 0 = standard
- 1 = dialekt
- F* = uafgørligt
- F0* = uafgørligt

2.5.2 Blødt D Odder (blodt_d)

- Ordkodet

Undersøgelse af dialekt- kontra standardudtale af blødt d.

Inventar:

- D0 = standard
- D1 = dialekt
- D2 = neutrale
- D3 = regionale
- D4 = uafgørligt
- D = at der er oprettet en automatisk kodning af et script, nogle af ordene egner sig ikke eller falder uden for den ratio, der kodes indenfor.

2.5.3 Blødt_D_32 Odder (blodt_d_32)

- Ordkodet

Undersøgelse af dialekt- kontra standardudtale af blødt d, hvor kun de første 32 forekomster efter 900 sekunder er opmærket.

Inventar:

- D = at der er oprettet en automatisk kodning af et script, nogle af ordene egner sig ikke eller falder uden for den ratio, der kodes indenfor.
- D0 = standard
- D1 = dialekt
- D4 = uafgørlig

2.5.4 Udtale IKKE (ikke)

- Ordkodet

Undersøgelse af dialekt- kontra standardudtale af *ikke*.

Inventar:

- I = at der er oprettet en automatisk kodning af et script, nogle af ordene egner sig ikke eller falder uden for den ratio, der kodes indenfor.
- I1 = dialekt
- I0 = standard

2.5.5 Udtale JEG (jeg)

- Ordkodet

Undersøgelse af dialekt- kontra standardudtale af *jeg*.

Inventar:

- JA = at der er oprettet en automatisk kodning af et script, nogle af ordene egner sig ikke eller falder uden for den ratio, der kodes indenfor.
- JA0 = standard
- JA1 = dialekt

2.5.6 Endelser (endelser)

- Ordkodet

1. Dialekt- eller standardudtale af -et i participium, bestemt form, m.v. (ordet "meget" er opmærket særskilt). Trækket defineres hos Nielsen og Nyberg således: R = [E6], -R = [Et, Ed] (Nielsen og Nyberg 1988:46).

2. Dialekt- eller standardudtale af -ede i præteritum (her er opmærket de første 32 forekomster efter 900 sekunder)

Inventar:

- END0 = standard
- END1 = dialekt
- END2 = regional
- END3 = -en bøjning af stærke verber
- END4 = uafgørlig

2.5.7 Endelser_32 (endelser_32)

- Ordkodet (hvert token (ord) er annoteret separat).

Se endelser. De første 32 forekomster efter 900 sekunder er opmærket.

Inventar:

- END0 = standard
- END1 = dialekt
- END2 = regional
- END3 = -en bøjning af stærke verber
- END4 = uafgørlig

2.5.8 Skriftens OR (skriftens_or)

- Ordkodet

Dialekt- eller standardudtale af -or i følgende ord: gjorde, gjort, fjorten, historie, horn, jordbær, lort, skjorte, sort, torden, fjortende, fjortenårig, lortet og tror.

Inventar:

- OR0 = standard
- OR1 = dialekt
- OR4 = uafgørligt

2.5.9 Langt O-vokalisme (OOvokalisme)

- Ordkodet (hvert token (ord) er annoteret separat).

Undersøgelse af diftongering af langt 'o'.

OBS: I visse kodninger er koden (fx OO) ikke slettet foran tallet, dvs. langt o udtalt med dialekt er OO1.

Inventar:

- OO = at der er oprettet en automatisk kodning af et script, nogle af ordene egner sig ikke eller falder uden for den ratio, der kodes indenfor.
- 0 = standard
- 1 = dialekt
- 3 = regional
- 4 = uafgørligt
- 5 = gjort udtalt uden finalt t

2.5.10 Langt E-vokalisme (EEvokalisme)

- Ordkodet (hvert token (ord) er annoteret separat).

Undersøgelse af diftongering af gammelt langt 'e'.

Inventar:

- EE = at der er oprettet en automatisk kodning af et script, nogle af ordene egner sig ikke eller falder uden for den ratio, der kodes indenfor.
- 1 = standard
- 2 = dialekt
- 3 = uafgørligt

2.5.11 Ordmarkering (ordmarkering)

- Ordkodet (hvert token (ord) er annoteret separat).

Markering af ikke og jeg i Vinderup (og sådan og også i Odder?).

Inventar:

- DS0 = standard
- DS1 = dialekt
- DS2 = neutral
- DS3 = regional
- DS4 = ufgørligt

2.5.12 Ordmarkering_32 (ordmarkering_32)

- Ordkodet (hvert token (ord) er annoteret separat).

Inventar:

- ORD = at der er oprettet en automatisk kodning af et script, nogle af ordene egner sig ikke eller falder uden for den ratio, der kodes indenfor.
- ORD0 = standard
- ORD1 = dialekt
- ORD4 = afgørligt
- ORD5 = regionalt

2.5.13 Vestjysk stød (V)

- Ordkodet (hvert token (ord) er annoteret separat).

Undersøgelse af vestjysk stød.

Inventar:

- V= at der er oprettet en automatisk kodning af et script, nogle af ordene egner sig ikke eller falder uden for den ratio, der kodes indenfor.
- V0 = standard
- V1 = dialekt

2.5.14 Klusilspring (K)

- Ordkodet (hvert token (ord) er annoteret separat).

Undersøgelse af klusilspring.

Inventar:

- K = at der er oprettet en automatisk kodning af et script, nogle af ordene egner sig ikke eller falder uden for den ratio, der kodes indenfor.
- K0 = standard
- K1 = dialekt

2.6 Køgeprojektet

- Kontaktperson: Janus Spindler Møller: janus@hum.ku.dk
- Status: Åben

2.6.1 Tyrkisk-dansk (TR-DK)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Oversættelser fra tyrkisk til dansk

Eksempel på sekvenskodning: Når eksempelvis tre tokens er annoteret med *hun_snakkede_aldrig*, vil de tre tokens få disse tre tags: 1_hun_snakkede_aldrig_I, 2_hun_snakkede_aldrig_I, 3_hun_snakkede_aldrig_E. Tallet før tagget angiver positionen og bogstavet efter angiver om det er det sidste tag i rækken eller ej.

2.7 Sprogholdningstiers

Kontaktperson: Tore Kristiansen: tk@hum.ku.dk

2.7.1 SH Sprogholdning (SH)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Status: Åben

I dette tier markeres hvornår der tales om sprogholdning. De anvendte tags benyttes altså blot til at markere hvilke passager der er relevante i forhold til sprogholdning.

Der er tre typer af relevante passager:

SHsp	passager, der handler om sprogholdningsundersøgelsen
SHyt	passager med direkte ytringer af eller om sprogholdninger
SHin*	passager, der er interessante i forbindelse med sprogholdninger (avis-vaner, tv-vaner, radio-vaner m.fl.)

*Denne kategori blev dog aldrig anvendt og kan derfor ikke findes i korpus

Eksempel på sekvenskodning: Når eksempelvis tre tokens er annoteret med *SHsp*, vil de tre tokens få disse tre tags: 1_SHsp_I, 2_SHsp_I, 3_SHsp_E. Tallet før tagget angiver positionen og bogstavet efter angiver om det er det sidste tag i rækken eller ej.

Se kodningsmanualen *kodningsmanualSH.doc* for yderligere information.

2.8 Hanne Sæderup

- Kontaktperson: Hanne Sæderup, sbc846@hum.ku.dk
- Status: Privat (gælder alle tiers såvel som lydfiler).

2.8.1 DP Discourse Presentation (DP)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Se kapitel 4 i dokumentet HANNE_SÆDERUP_PHD.pdf

Discourse Presentation

"På basis af en kodningsmanual, jeg har afstemt til mine data, opmærkes korpusset på tre niveauer: ift. gengivelsesmodus, ift. tale-, skrift- og tankegengivelseskategorier og endelig ift. grammatiske realiseringer af disse kategorier"

Plads 1: Discourse presentation

SIS	Speech presentation, indirect speech
SFIS	Speech presentation, free indirect speech
SDS	Speech presentation, direct speech
SRV	Speech presentation, representation of voice
SRSA	Speech presentation, representation of speech act
WRN	Writing presentation, representation of writing
WRWA	Writing presentation, representation of writing act
WIW	Writing presentation, indirect writing
WFIW	Writing presentation, free indirect writing
WDW	Writing presentation, direct writing
TRT	Thought presentation, representation of thought
TRTA	Thought presentation, representation of thought act
TIT	Thought presentation, indirect thought
TFIT	Thought presentation, free indirect thought
TDT	Thought presentation, direct thought
RU	Report of language use

Plads 2 (modifikatorer til plads 1):

h (hp, hn, hf)	Hypothetical (proper, negation, future)
----------------	---

i	Non-specific discourse presentation (generic and iterative)
s	Specific discourse presentation
e	Embedded discourse presentation
p	RV/RW/RN or RSA/RWA/RTA with topic (topic = p)
m	Metonymic discourse presentation
inel	Interactional, elicited
in	Interactional discourse presentation
u	Unfinished discourse presentation
s1sing, s3sing, etc.	Speaker Voice (relevant for all instances of discourse presentation)

Plads 3: Taler. 3 underpladser:

Person (S1, S2, S3)	Speaker 1st person, speaker 2nd person, speaker 3rd person
Tal (s, p)	Singular, plural
Rolle (d, p, dy, dg, ...)	Doctor, patient, doctor (ps y chiatrist), doctor (g e neral), ..

Plads 4: Genericity

Rs	Specific
Ri	Non-specific

Plads 5: Hypothetical discourse presentation

Hp	Hypothetical, proper
Hn	Hypothetical, negated
Hf	Hypothetical, future

2.9 Vollsmose ... (LaPUR)

- Kontaktperson: Pia Quist: pqj@hum.ku.dk
- Status: NORS-niveau

Se dokumentet "Tiers til kodning af Vollsmose" for yderligere informationer.

Se desuden Astrid Ravn Skovses ph.d.-afhandling (Danske Talesprog 18), der omhandler disse data samt Bylderup-dataene.

2.9.1 Blødt D Vollsmose (Blødt d)

- Ordkodet (hvert token (ord) er annoteret separat).

ø-bortfald inde i ord

Inventar:

u d	for "uden d"
-----	--------------

2.9.2 Udtale -EDE (EDE)

- Ordkodet (hvert token (ord) er annoteret separat).

Realisering af endelse, svage verber i præteritum med endelse på -ede, hhv. standard [əðə] og fynsk [əd].
Fx *hoppedede*.

Inventar:

0	uafgørligt
1	standard
2	dialekt
EDE	Grovopmærkning til at identificere potentielt interessante steder
u d	Muligvis for "uden d"

2.9.3 Udtale -ET (ET)

- Ordkodet (hvert token (ord) er annoteret separat).

Realisering af endelse i diverse ordformer på -et, hhv. standard [əð] og fynsk [əd]. Fx *huset, fjollet*, (har) *hoppet*.

Inventar:

0	uafgørligt
1	standard
2	dialekt
ET	Grovopmærkning til at identificere potentielt interessante steder

2.9.4 Fortsætterintonation (Fortsætterint)

- Ordkodet (hvert token (ord) er annoteret separat).

Fortsætterintonation

fort (eller f eller 1)	fortsætterintonation
------------------------	----------------------

2.9.5 Fynsk intonation (Fynsk int)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Særlige intonationsmønstre.

Tieret beskrevet i kodningsmanualen hedder "Intonation".

Inventar:

fynsk	Fynsk intonation
ME	Multietnolekt

Eksempel på sekvenskodning: Når eksempelvis tre tokens er annoteret med *fynsk*, vil de tre tokens få disse tre tags: 1_fynsk_I, 2_fynsk_I, 3_fynsk_E. Tallet før tagget angiver positionen og bogstavet efter angiver om det er det sidste tag i rækken eller ej.

2.9.6 Grammatisk køn (Gramm. køn)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Variation i grammatisk køn. Opmærkningen spænder over et givet ord og dets artikel.

Inventar:

i	intetkøn
f	fælleskøn

Eksempel på sekvenskodning: Når eksempelvis tre tokens er annoteret med *i*, vil de tre tokens få disse tre tags: 1_i_l, 2_i_l, 3_i_E. Tallet før tagget angiver positionen og bogstavet efter angiver om det er det sidste tag i rækken eller ej.

2.9.7 Kommentarer

Kommentartier.

Eksempler: "kraftigt"" i tænkte og tiden; "thorn""-lyd; "under"" minder om Ebru; holder f0 næsten konstant meget længe; ME; ME-agtigt; obs speciel intonation

2.9.8 Leksis

- Ordkodet (hvert token (ord) er annoteret separat).

Bemærkelsesværdige ting i ordforrådet. Fx *walla*.

Kun 3 fyldte intervaller.

2.9.9 Manglende vokalkontrast (Mgl vokalkontrast)

- Ordkodet (hvert token (ord) er annoteret separat).

Forkortelse af langvokal, fx [sgolə] i stedet for [sgo:lə].

Inventar:

f	(for "forkortelse")
---	---------------------

2.9.10 Palatalisering

- Ordkodet (hvert token (ord) er annoteret separat).

Palatalisering: /t/ → [tj]

Inventar:

0, 1, 2, R

0	uafgørligt
1	standard
2	dialekt
TJ	Grovopmærkning til at identificere potentielt interessante steder

2.9.11 Præposition (Præp)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Afvigende brug af præpositioner. Opmærkningen spænder over flere ord da man ellers ikke kan opmærke en manglende præposition.

Inventar:

i/af/på... + d	drop af præposition (fx "vi går skole" i stedet for "vi går i skole")
i/af/på... + a	afvigelse i præposition (fx "jeg skal på skole" i stedet for "jeg skal i skole")
f	?

Eksempel på sekvenskodning: Når eksempelvis tre tokens er annoteret med p_+_a , vil de tre tokens få disse tre tags: 1_p+_a_I, 2_p+_a_I, 3_p+_a_E. Tallet før tagget angiver positionen og bogstavet efter angiver om det er det sidste tag i rækken eller ej.

2.9.12 Stødafvigelse (Stødafv)

- Ordkodet (hvert token (ord) er annoteret separat).

Stødafvigelse – hhv. manglende stød eller "ekstra" stød, hvor stødsløshed var forventet.

Inventar:

m s	minus stød
p s	plus stød
u s	uden stød

2.9.13 Ustemt R (Ustemt r)

- Ordkodet (hvert token (ord) er annoteret separat).

En særlig kvalitet af konsonant + r der er svær at skelne fra én lang stemt klusil.

Inventar:

0	uafgørligt
1	standard
2	dialekt
R	Grovopmærkning til at identificere potentielt interessante steder

2.9.14 V2->V3 (V2->V3)

- Ordkodet (hvert token (ord) er annoteret separat).

Ordstillingsalternation hvor verbet ender på 3. plads, fx ”men ellers jeg ved godt hun har gode venner.

Inventar:

v3	Verbet på 3. plads
----	--------------------

2.10 AMDA (Amerikadansk: AmDa, ArgDa, CanDa)

2.10.1 Language-stated

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).
- Kontaktperson: Jan Heegård Petersen: janhp@hum.ku.dk (tidl. Gert Foget Hansen)
- Status: Forskerniveau

'Language-stated' beskriver de sprog som udskriverne har annoteret ordene i ortografien med. Står der fx *spansk*, er det fordi udskriveren har ment at det givne ord er et spansk ord (sagt på spansk). Derudover er det noteret når det ikke er entydigt hvilket sprog der er benyttet (pga. udtale), samt når der er benyttet en hybrid af to eller flere sprog. Et givent ord kan kun have ét af disse tags tilknyttet til sig.

OBS: Der ser ud til at være kommet language stated-data ind i comments_x.

Tags benyttet i dette tier:

Sprog:

- dansk
- engelsk
- norsk
- spansk
- svensk
- tysk

Andre notationer:

- ambig
 - De ord som i kraft af deres udtale ikke kan tilskrives et sprog, er markeret med tagget 'ambig'.
- hybrid
 - Ord der er sammensat af ord eller et ord og et affiks fra mere end ét sprog, er markeret med tagget 'hybrid' – uanset hvilke sprog der er tale om.
- ambig, hybrid
 - En kombination af de to tags.
- non-autonom
 - Hvis informanterne læser op, synger eller reciterer (dvs. producerer ikke-selvstændig sprogbrug), er det markeret med tagget 'non-autonom'.
- discuss
 - En note til korrekturlæseren om at der er et tvivlstilfælde. Dette tag forekommer kun i kombination med andre tags (fx 'ambig, discuss').

2.10.2 Language revised

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

- Kontaktperson: Jan Heegård Petersen: janhp@hum.ku.dk (tidl. Gert Foget Hansen)
- Status: Åben

'Language-revised' beskriver de sprog som udskriverne har annoteret ordene i ortografien med. Står der fx *spansk*, er det fordi udskriveren har ment at det givne ord er et spansk ord (sagt på spansk). Det er også noteret når det ikke er entydigt hvilket sprog der er benyttet (pga. udtale), samt når der er benyttet en hybrid af to eller flere sprog. Derudover er der tags til alle ikke-sprogspecifikke lyde såsom ved tøyen, latter o. lign. Et givent ord kan kun have ét af disse tags tilknyttet til sig.

Tags benyttet i dette tier:

Sprog:

- dansk
- engelsk
- norsk
- spansk
- svensk
- tysk

Andre notationer:

- ambig
 - De ord som i kraft af deres udtale ikke kan tilskrives et sprog, er markeret med tagget 'ambig'.
- hybrid
 - Ord der er sammensat af ord eller et ord og et affiks fra mere end ét sprog, er markeret med tagget 'hybrid' – uanset hvilke sprog der er tale om.
- ambig, hybrid
 - En kombination af de to tags
- non-autonom
 - Hvis informanterne læser op, synger eller reciterer (dvs. producerer ikke-selvstændig sprogbrug), er det markeret med tagget 'non-autonom'.
- egennavn
 - Proprier tilskrives ikke specifikke sprog men er derimod annoteret med 'egennavn'
- fejl
 - Er der sket en fejl ortografiteret, tilskrives ordet ikke noget specifikt sprog men er derimod annoteret med 'fejl'
- latter
 - Latter (ordet 'ha' i ortografiteret) er annoteret med ordet 'latter'
- minimalrespons
 - Ved minimalrespons (ordet 'mm' i ortografiteret) er ordet 'minimalrespons' anvendt
- selvafbrydelse
 - Hvis taleren afbryder sig selv (halve ord der slutter med bindestreg i ortografiteret), er det annoteret med 'selvafbrydelse'
- tøyen
 - Tøyen (ordet 'oe' i ortografiteret) er annoteret med ordet 'tøyen'
- udbrud
 - Ved ikke-sprogspecifikke udbrud (såsom 'uh', 'arh' eller 'årh') er ordet 'udbrud' anvendt
- uforståelig
 - Ved uforståelige ord (annoteret med 'xxx' i ortografiteret) er ordet 'uforståelig' anvendt
- zzz

- 'zzz' er anvendt når der står 'zzz' i ortografitieret. Dette tag er anvendt når der er passager i interviewet der af forskellige årsager bevidst ikke er transskriberet.

Derudover kan alle sprognavne samt 'ambig', 'hybrid', 'ambig, hybrid' og 'non-autonom' have fået tilføjet endelsen '-farm'. Dette betyder at det pågældende ord har stammen 'farm'. Dette er blevet gjort eftersom ord med stammen 'farm' er meget hyppigt forekommende, og man gerne ville kunne analysere disse ord separat.

2.10.3 AmDa Syntax (Syntax)

- Sekvenskodet (BIO-opmærket). Oprindelige annotationssekvenser "X X X" er omkodet til "1_X_I 2_X_I 3_X_E" (jf. beskrivelsen af BIO-opmærkning ovenfor).

Jf. HUM-NORS-L-Danske_Stemmer_i_USA_og_Argentina/Danske_Stemmer_kodning/Jessie og Dittes manual for Amda og Argda_Syntakskorrektur_20161212.docx

Inventar

Danske stemmer-projektets syntakskodning koder helsætninger. Hver syntaksannotation indeholder således typisk kodning af type m.m. for én helsætning og derudover kodning af type m.m. for hver evt. ledsætning. Kodningerne for hver sætning (hel- hhv. ledsætning) er adskilt af "_" (oprindeligt af " ").

Kodningen for en enkelt sætning består af to dele adskilt af skråstreg: sætningstype og ledstillingskategori. De beskrives i det følgende.

Sætningstype

Hvis der er flere sætninger i helsætningen, tilføjes et løbenummer direkte efter sætningstypekoden, fx H/.. L1/.. L2/.. Hd3/.. Ld4/..

H	Helsætning
Hkoo	Helsætning med kongruenskonstruktion
Hd	Helsætning i direkte tale
L	Ledsætning
Lkoo	Ledsætning med kongruenskonstruktion
Ld	Ledsætning i direkte tale
Xeng	Engelsk helsætning (uden yderligere kodning)
Xspan	Spansk helsætning (uden yderligere kodning)
PU	Sætningsformet påhængsudtryk (står altid som sidste element i den samlede annotation)

Ledstilling

Delen efter skråstregen angiver rækkefølgen af subjekt, finit verbum og eventuelle sætningsadverbialer.

A	Sætningsadverbialer
Vfin	Finit verbum; eventuelt infinitte dele af verbalet kodes <u>ikke</u> .
S	Subjekt
Sf	Formelt/foreløbigt subjekt (inkl. <i>som/der</i> i relativsætninger)
Se	Elliptisk subjekt
0	Udeladt subjekt

2.11 Ukendte tiers

Her anføres tiers som det ikke har været muligt at finde fyldestgørende information om, og hvis status derfor er uklar.

2.11.1 kommentarer

Forekommer i: Odder2; BySoc1; Odder1; Næstved1.

LANCHART-korpusset indeholder diverse kommentartiers med navne der ligner hinanden meget. Dette vil der blive ryddet op i ved lejlighed.

2.11.2 Kommentar

Kun 3 intervaller. I Odder2. Endnu et af de mange kommentartiers, jf. ovenfor.

2.12 Tiers der er skjult

Tiers i dette afsnit er p.t. taget ud af korpus da de tilsyneladende ikke bruges af nogen. Kan sættes ind igen ved evt. behov.

2.12.1 variant leksikalsk

808 fyldte intervaller. Findes kun i filen bysoc1gl-87-KJN.TextGrid.

Det er uklart hvad det er for en type kategorier, men der er tilsyneladende tale om en slags ordklasser, dog ikke i traditionel forstand.

Inventar:

D	?
S	Slang?
EE	Engelske ord
B	Bandeord
d	?
L	?
ED	?
EO	?

2.12.2 global turn

Viser antagelig hvilken taler der har turen. Findes kun i Bylderup3 og Modsjæl2.

2.12.3 global sync

Kun tomme intervaller. Findes kun i Bylderup3 og Modsjæl2.

2.12.4 sync

Findes i: HanneSæderup; ModSjæl2; AmDa-dana; ArgDa; CanDa-ch; Køge3; Amager3; Vollsmose3; AmDa-tk; Vinderup1; Bylderup3; Næstved2; Køge25; Bornholm; AmDa-kbl; Vinderup0; CanDa; Munkebjerg3.

1,4 mio. intervaller, ingen fyldte.

Det vides p.t. ikke hvad det bruges til.

Tieret matcher tidsstemplerne i den oprindelige transcriber-fil.

2.13 Tiers der er slettet

Tiersne i denne sektion er enten tomme (nul fund i den gamle søgemaskine) eller på anden måde ubrugelige.

De er derfor slettet helt fra LANCHART-korpusset og er kun med her som historisk dokumentation.

2.13.1 Comments global

Det må vel være jf. udskrivningsmanualen: "Kommentarer som drejer sig om baggrundslyde eller andre ting som ikke kan tilskrives en bestemt taler skal udskrives som en global kommentar, eksempelvis *{global: CMP forlader lokalet}*, *{global: chatlyd}*."

Kun 2 tomme intervaller. Ud.

Hvorfor er der kun 2 tomme intervaller?? Det kan da ikke passe at der ikke er nogen globale kommentarer?? Svar: Mon ikke det er dem der er i "global events".

2.13.2 dialekt comments

Kommentarer ifm. dialekttiers (flyttet ind i mere generelt kommentartier)

2.13.3 grammatik2DUPLICATE

Ansvarlig: Torben Juel Jensen

Kun 99 fyldte intervaller. Kun Næstved3.

2.13.4 intervaltier

?

2.13.5 lydskrift

Gammelt lydskrifttier?

2.13.6 parole_PoS

Gammel version af ordklasse?

2.13.7 parole_PoSRED

Gammel version af reduceret ordklasse?

2.13.8 style

Gammel version af IIV?

2.13.9 Syntax edit

Med Syntax-annotation fra Danske Stemmer-projektet. Alle hits var i Testfiler der nu er slettet.

2.13.10 translation

Tomt tier.

3 Metadata

I LANCHART-korpusset er diverse metadata på korpus-, samtale- og informantniveau søgbare parallelt med annotationerne i de forskellige tiers. Disse metadata beskrives i det følgende.

Benævnelser i parentes er de rå variabelnavne der bruges internt i søgesystemet – og som optræder som kolonnenavne i dataeksporten.

NB: Tomme værdier bliver i Korp til den særlige værdi "__UNDEF__". Hvis man fx vil specificere at et givet ord IKKE er annoteret med noget i kategorien "Socialklasse", skal man lave en søgning på tokens hvor "Socialklasse er __UNDEF__".

3.1 Særlige metadata – segmentering og varighed

De følgende kategorier er tids- og varighedsvariable der er genereret ud fra det enkelte TextGrid samt fra turdelsegmenteringen. Hvad er turdelsegmenteringen? Samtalerne er segmenteret i noget der kan minde om taleture – som dog ofte ikke svarer til hele taleture. Derfor bruger vi termen "turdel" om de enkelte segmenter.

- **Talers ord nr.** (text_enum): Løbenummer for den givne talers ord i samtalen.
- **Ord nr. i samtale** (ordnummer): Løbenummer for det givne ord i hele samtalen.
- **Ord nr. i turdel** (turn_enum): Ordets løbenummer i den givne turdel.
- **Ord start** (xmin): Tidspunkt for ordets begyndelse (i sekunder fra TextGrid-start).
- **Ord slut** (xmax): Tidspunkt for ordets slutning (i sekunder fra TextGrid-slut).
- **Ordlængde** (xlength): Ordintervallets længde i sekunder if. TextGrid.
- **Turdel nr.** (turnummer): Turdelens løbenummer i samtalen.
- **Talekilde** (talekilde): "deltager", "baand" (= stemmeprøve) eller "andet" (forbipasserende osv.).
- **Turdel start** (turnmin): Tidspunkt for turdelens begyndelse (i sekunder fra TextGrid-start).
- **Turdel slut** (turnmax): Tidspunkt for turdelens slutning (i sekunder fra TextGrid-start).
- **Turdelslængde** (turnduration): Turdelens længde i sekunder.

3.2 Traditionelle metadata

Traditionelle metadata er kategorier som køn, fødselsår, socialklasse osv.

Variable med præfikset "samtaler_" kommer fra informantbasens "samtaler"-tabel, og tilsvarende med "informanter_" og "projekter_". Hvis en variabel ikke kommer direkte fra informantbasen men er genereret af koden, angives dette.

3.2.1 Metadata på projektkorpusniveau

P.t. er den eneste metadatakategori på korpusniveau de enkelte projektkorpussers id-kode i korpusdatabase:

- **CWB-korpus-id** (label): Fx LANCHART_AMAGER, LANCHART_OKSBOEL osv.

3.2.2 Metadata på samtaleniveau

Oplysninger der gælder for en hel samtale/optagelse.

- **Filnavn** (filename): Udskriftens filnavn, fx "oksboel3gl-11-CMP+JLY+JRI+KAK+MLB+SJU_selvoptagelse.TextGrid"
- **Optagedato** (samtaler_dato): Dato for lydoptagelsen, fx 2006-01-31
- **Ny/gammel optagelse** (oldnew, genereret ud fra filnavn): "gl" (første optagelse med en informant, OBS: kun relevant for visse projekter); "ny" (anden optagelse med en informant – optaget en år-række efter den første optagelse, OBS: også kun relevant for visse projekter).
- **Samtaletype** (samtaler_samtaletype): "Interview" (samtaler mellem en interviewer og én informant. OBS: I visse projekter (Vissenbjerg, Odder 1, Næstved 1) inkluderer kategorien også samtaler med flere informanter, som rettelig burde kategoriseres som "Gruppeinterview"); "Gruppeinterview" (samtaler mellem en interviewer og flere informanter); "Gruppesamtale" (samtaler mellem to eller flere informanter); "Selvoptagelse" (samtaler som informanten selv har optaget); "Andet" (andre typer samtaler, fx radiointerview og vejledningssamtaler).
- **Eksplorativ**: Indgår samtalen i "det eksplorative korpus" – ja eller nej (1/0). Det eksplorative korpus er et mindre delkorpus beregnet til eksplorativt og hypotesegenererende arbejde.
- **Korrektur**: Er udskriften korrekturlæst – ja eller nej (1/0).
- **Prioriteret samtale**: Er samtalen en del af samlingen af prioriterede samtaler – ja eller nej (1/0).
- **Prioriteret samtale, ekstra**: Er samtalen en af de ekstra samtaler der blev kategoriseret som prioriteret – og altså ikke en del af de oprindelige prioriterede samtaler. Ja eller nej (1/0).
- **Projekt** (projekter_name): Projektkode, fx "Odder3". Find mere information om de enkelte projekter på Sprogforandringscentrets hjemmeside.
- **Projekt-id** (samtaler_projekt): Projektets løbenummer.
- **Indspilning længde** (textduration, genereret): Samtalens længde i sekunder if. TextGrid.

3.2.3 Metadata på informantniveau

Oplysninger der gælder for de enkelte informanter.

- **Taler** (speaker): Talerkode. Typisk tre store bogstaver, fx "ADA". Der er dog også visse andre koder: UK1, UK2 osv. er 'ukendte talere', dvs. talere der ikke hører til interviewet ("forbiasserende"). (OBS: UK-koden er udfaset, så her er der en rettelse på vej). S01, S02 osv. er stemmeprøver, dvs. båndoptagelser som informanterne skal forholde sig til.
- **Køn** (informanter_koen): "m" eller "f".
- **Fødselsår** (informanter_foedselsaar): Fire cifre.
- **Generation** (informanter_generation): "0" (født før 1942 (OBS: AmDa-dataenes underopdeling i generation "0", "00", "000", "0000" er ikke registreret på informantniveau i informantbasen)); "1" (født 1942-1963); "2" (født 1964-1973); "2.5" (født 1981-1986); "3" (født 1987-1996); "4" (født 1997 og frem). Læs mere om generationskategorierne: <https://dgcss.hum.ku.dk/online-ressourcer/hvad-ved-vi-nu-om-danske-alesprog/2015/hvad-ved-vi-nu-alle-filer/Indledning.pdf>, <https://www-tandfonline-com.ep.fjernad-gang.kb.dk/doi/full/10.1080/03740460903364003>).
- **Socialklasse** (informanter_socialklasse): "AK" for 'arbejderklasse' og "MK" for 'middelklasse'.
- **Rolle** (rolle): "meddeler" (informant), "interviewer", "forbiasserende", "andet".