

Studiar i bokmålsnaturen  
via *Bokmålsordboka*,  
*Norsk referansegrammatikk* og HPSG

Leiv Hellebø

Hovudfagsoppgåve  
Institutt for lingvistiske fag  
Universitetet i Oslo

15. desember 2003



# Mange takk

Takk til rettleiaren min, Jan Tore Lønning, for både oppmuntring og kritikk.

Takk til Liv Ellingsen for samveret vi hadde då ho enno heldt på med hovudfagsoppgåva si.

Takk til Erik Velldal for det same då han var i sin hovudfagsinnsput.

Takk til Janne Torkildsen for grundig lesing av nokre kapittel.



# Innhald

<b>1</b>	<b>Innleiing</b>	<b>11</b>
<b>2</b>	<b><i>Bokmålsordboka</i></b>	<b>13</b>
2.1	Oppslaget . . . . .	13
2.2	Bøyingsendingar og -former . . . . .	15
2.3	Bøyingskodane . . . . .	15
2.4	Variantformer og leksikonstruktur . . . . .	18
2.5	Radikale og konservative former . . . . .	22
<b>3</b>	<b><i>Tabellmaterialet</i></b>	<b>27</b>
3.1	Lemmalista . . . . .	27
3.2	Paradigmelista . . . . .	28
3.3	Lemma-paradigme-lista . . . . .	31
3.4	Bøyingskategorilista . . . . .	31
3.5	Bøyingssendingslista . . . . .	31
<b>4</b>	<b>Bokmålstreet</b>	<b>35</b>
4.1	Eit augekast på paradigma . . . . .	36
4.2	Formendingstabell . . . . .	37
4.3	Paradigmetypen . . . . .	40
4.4	Paradigmegrupper og RRAG-ar . . . . .	42
4.5	Paradigmetre . . . . .	44
4.6	Klammeparadigme? . . . . .	46
4.7	Svake bokmålsverb . . . . .	48
4.8	Bokmålstreet . . . . .	57
4.8.1	Andre paradigmegrupper . . . . .	58
4.8.2	Konservative og radikale paradigmegrupper . . . . .	58
4.8.3	Kor mange paradigmegrupper finst? . . . . .	61
4.8.4	Paradigmegrupper for både bokmål og nynorsk . . . . .	61

<b>5</b>	<b><i>Norsk referansegrammatikk</i></b>	<b>63</b>
5.1	Første kritikk: Fleirtytipe bøyingsgrupper . . . . .	64
5.2	Andre kritikk: Infinitivstammen . . . . .	65
5.3	Tredje kritikk: Gruppeinndeling . . . . .	68
5.4	Ein tankekross . . . . .	71
<b>6</b>	<b>Bokmål i HPSG</b>	<b>73</b>
6.1	Rammene . . . . .	74
6.2	Leksikonbyggjeplan . . . . .	75
6.3	HPSG-typar for bokmål . . . . .	76
6.3.1	Paradigmehierarkiet . . . . .	76
6.3.2	Typar for grammatiske kategoriar . . . . .	79
6.3.3	Leksemtypar . . . . .	81
6.3.4	Ordtypar . . . . .	83
6.3.5	Maskingenererte tradisjonelle reglar . . . . .	84
6.4	Kompakt leksikon . . . . .	86
6.5	Til sist . . . . .	86
6.5.1	Sjølvskrøyt . . . . .	86
6.5.2	Men noko står att . . . . .	86
6.5.3	Om vedlagte filer . . . . .	87

# Figurar

4.1	Paradigmetypene <i>001</i> , <i>030</i> og <i>001-030</i> . . . . .	44
4.2	Paradigmetypene <i>001</i> , <i>030</i> og <i>001-030</i> . . . . .	45
4.3	Forenkla tre over svake verb . . . . .	51
4.4	Paradigmetre over dei svake verba . . . . .	54
4.5	Alternativt paradigmetre over dei svake verba . . . . .	56
5.1	Referansegrammatikkens paradigmetre over dei svake verba .	70
6.1	Utsnitt av paradigmehierarkiet . . . . .	77
6.2	Trekkstruktur for kjønn med meir . . . . .	79





# Tabellar

2.1	Ulike oppslag for bøyelege ord frå <i>Bokmålsordboka</i> . . . . .	14
2.2	Utdrag frå kodenøkkelene frå <i>Bokmålsordboka</i> . . . . .	16
2.3	Bøyingsstypar i bøyingskodane . . . . .	16
2.4	Bøyingsstypar med utgangspunkt i ordstammen. . . . .	17
2.5	Bøyingsstypar med variantformer . . . . .	19
2.6	Variantformer og obligatoriske former . . . . .	20
2.7	Variantendingar og obligatoriske endingar . . . . .	20
2.8	Kjønnsfordeling på substantiv . . . . .	23
2.9	Oppslag der bøyingsstypane inneheld fleire kjønn . . . . .	24
2.10	Fullformene frå førre tabell . . . . .	24
3.1	Oversyn over <i>Tabellmaterialet</i> . . . . .	27
3.2	Utdrag frå <i>Paradigmelista</i> . . . . .	29
3.3	Bøyingsgrupper i <i>Tabellmaterialet</i> . . . . .	30
3.4	Bøyingskategorilista . . . . .	31
3.5	Bøyingsendingslista . . . . .	32
3.6	Formendingstabell for to adjektivparadigme . . . . .	33
4.1	Formendingsliste . . . . .	36
4.2	Formendingstabell . . . . .	37
4.3	Klammer og bøyingsgrupper . . . . .	46
4.4	Formendingstabell for førre tabell . . . . .	47
4.5	Bøyingsgrupper for svake verb (ikkje alle) . . . . .	49
4.6	Formendingstabell for svake verb (ikkje alle) . . . . .	49
4.7	Resten av førre formendingstabell . . . . .	53
4.8	Utdrag frå <i>Paradigmelista</i> . . . . .	58
5.1	Infinitivsstammen som leieform . . . . .	67
5.2	Infinitiven som leieform . . . . .	67
5.3	Verbbøyingskodar i <i>Norsk referansegrammatikk</i> . . . . .	71

5.4	Verbbøyingskodar i <i>Bokmålsordboka</i> . . . . .	72
6.1	Leksikonfiler . . . . .	76

# Kapittel 1

## Innleiing

Eg har hatt tilgang til bokmålsmaterialet frå *Norsk Ordbank* (heretter *Tabellmaterialet*). Dette har leidd meg følgjande:

Eg granskar nøye bøyingsregimet i *Bokmålsordboka*.

Eg knyt bokmålsleksema til bøyingsgrupper avleidde frå *Tabellmaterialet*. Eg gjev ein instrumentell definisjon av ein fleksibel leksemmodell som høver språk med mange variantformer. Til eit leksem høyrer ei *leieform* og eit varierende tal *formendringar* som gjev fullformene.

Eg skisserer det modellteoretiske grunnlaget for oppgåva.

Bøyingsgruppene høver saman med leksemmodellen, og med bakgrunn i desse definerer eg ein *morfologisk konsekvent bokmålstekst*. Ved å skissere nokre algoritmer, legg eg grunnlaget for stavesjekkarakar som kan teste for morfologisk konsekvens. Eg skisserer seinare korleis ein også kan sørgje for stilistisk eller dialektal konsekvens.

Eg diskuterer bruken av klammer i representasjonen av bokmålsmorfologien.

Eg byggjer ein større modell som rommar alle verba *Norsk referansegrammatikk* kallar *svake verb*.

Modellen høver godt til morfologiske granskingar i bokmålet, så eg ønskjer meg dataverkty som kan vise fram ein større struktur som husar både bokmål og nynorsk. Modellen høver også til tekstkorpusundersøkjingar.

Eg kritiserer *Norsk referansegrammatikk* for ikkje å gje ei fullgod skildring av for eksempel verba som har fleire tilhøyrande preteritumsformer.

Eg kritiserer referansegrammatikkens leksemmodell, for eg synest ikkje han

høver til språk med både langformsinfinitivar (*roe*) og kortformsinfinitivar (*ro*).

Eg kritiserer referansegrammatikkens gruppeinndeling av dei svake verba. Ei betre forståing av bokmåls morfologien kan vi få om vi delar eine gruppa i to.

Eg samanliknar bøyingskodar frå *Bokmålsordboka* med tilsvarande frå *Norsk referansegrammatikk*.

Eg implementerer bokmålet i formalismen HPSG. Eg lagar HPSG-typar for fundamentale grammatiske kategoriar og for leksem og fullformer. Med datamaskin byggjer eg tre regelsett med varierende storleik som representerer bokmåls morfologien som han er skildra tidlegare i oppgåva.

## Kapittel 2

# *Bokmålsordboka*

*Tabellmaterialet* som ligg til grunn for min implementasjon, er igjen ein implementasjon av *Bokmålsordboka*. Det vert følgjeleg nødvendig å forstå *Bokmålsordboka* for å forstå *Tabellmaterialet*.

Dette kapitlet inneheld mange refleksjonar eg har gjort meg med nasen nedi *Bokmålsordboka*. Eg har særleg vore oppteken av bøyingsregimet, representasjonen av variantformer og representasjonen av kjønn.

### 2.1 Oppslaget

Den fremste delen av ordboksartikkelen, det som står framføre definisjonane, og ordhistorikken om den er skildra, skal eg i det følgjande kalle *oppslag*. Eit oppslag har to delar. Det har eit *oppslagsord*, og det har det eg skal kalle ein *bøyingstype*.

Eg har valt ut nokre av dei mest komplekse oppslaga, samt nokre få enkle, for framsyning. Dei er å finne i tabell 2.1. Merk at fleire av linene inneheld meir enn eitt oppslag.

Her ser vi at bøyingstypane vert førte dels i kodar, dels i fri tekst. Særleg er bokføringa av bokmålets variasjonsrikdom komplisert.<sup>1</sup>

Merk at nokre av linene inneheld *to* oppslag, ikkje eitt. Einskilde artiklar har nemleg to tillatne bøyingsformer som oppslagsord og får då eit dobbelt-oppslag. Både dei to første linene inneheld døme på dette, men det kjem til syne på ulikt vis: Oppslagsorda i første lina skal bøyast forskjellig, og dei får ulike bøyingstypar. Oppslagsorda i neste lina vert derimot likt bøyde,

---

<sup>1</sup>Om det er ønskjeleg kan ein finne systematiske måtar å føre oppslaga på ved ein jamføring av denne lista (og kanskje fleire oppslag) med bøyingsgruppene eg utleier i seksjon 4.3.

**I bein** *-et, -, -a [-ene]* el. **I ben** n1  
**II bein** el. **II ben** a1  
**bru** fl [m1]  
**II bære** v ubøy  
**dreie** v1 el. v3 [v4]  
**I dy** *-et*  
**feit** adj *fe(i)t, -e*  
**II grave [-a]** v3 el. *grov* i pret  
**inngangspenger** m fl  
**jubel** *-en*  
**kader** *-en, -drer*  
**kaie** fl el. m1  
**II kakk** *-en, -* el. n1  
**kalendarium** *-iet, -ier, -ia* el. *-iene*  
**kaste** v1  
**I katalansk** m1 el. n ubøy  
**klær** subst fl, bf *-ne*  
**I mil** *-a* el. *-en, -, -ene*  
**I min** eiendomspronomen *mi* el. *min, mitt, mine*  
**less** el. **lest** v, bare i pr og pret  
**II mark** *-a* el. *-en, merker*  
**skride** el. **I skri** *skred* el. *skrei[d], skredet*  
**II stemme** *stemme, stemt*

Tabell 2.1: Ulike oppslag for bøyelige ord frå *Bokmålsordboka*

dei får derfor same bøyingstypen, og oppslaga vert samantrekte, sikkert for å spare plass. Enno meir samantrekkjing finn vi for klammeforma **grava**.

Sjå så på ein kode- og alternativfri bøyingstype, la oss ta **kader** først.<sup>2</sup> Bøyingstypen fortel som sagt korleis oppslagsordets ulike bøyingsformer skal vere, og substantivbøyingstypar kan då i alt oppgje fire ulike bøyingsformer (eintal/fleirtal og bunden form/ubunden form). Den eine bøyingsforma er gjeve gjennom oppslagsforma, og for substantiv er dette normalt ubunden form eintal, sjølv om unntaka florerer (**klær**, **inngangspenger** og **tjukka** til dømes). For **kader** får vi altså spesifisert tre av fire moglege bøyingsformer, og bunden form fleirtal (*kadrene*) må lesaren sjølv tenkje seg fram til ut frå dei gjevne formene. Legg også merke til at **less** og **lest** ikkje er infini-

<sup>2</sup>Faktisk har *Tabellmaterialet* to variantformer til ubunden form fleirtal av *kader*: *kadre* og *kadrer*. Grunnen for tillegget er at forma er mykje brukt.

tivsformer, som er den vanlege oppslagsforma for verb (infinitivsforma finst ikkje).

Frå desse eksempla ser vi at ordboka ikkje spesifiserer alle bøyingsformene som høyrer til oppslagsordet, ho gjev berre ein peikepinn om bøyingstypen. Lesaren må derfor sjølv vite kva og kor mange bøyingsformer som høyrer til ordklassetypen, samt kunne følgje bøyingsmønsteret for ordklassen. Normalt er vel ikkje dette for mykje forlangt av ordboksbrukaren; eg har for eksempel ikkje høyrte om nokon som mislukkast i å uttrykkje seg med imperativskonstruksjonar av di imperativa ikkje står oppførte i *Bokmålsordboka*. For folk som lærer norsk, er det verre, vil eg tru.

## 2.2 Bøyingsendingar og -former

Dei oppførte bøyingsformene i bøyingstypene i tabell 2.1 står anten som fulle former eller som bøyingsendingar. For å finne dei fulle formene for dei siste, må lesaren setje saman oppslagsordet med endinga. I samband med **kader** får vi bunden form eintal ved å leggje til bøyingsendinga *-en*. Skal den ubundne fleirtalsforma tre fram, nyttar det ikkje med så enkle middel. I staden må vi bakke opp tilbake i oppslagsordet til nett framføre siste “d” og stryke resten av ordet, før vi legg til endinga *-dre*. Igjen ser vi at lesaren sjølv må nytte sin språklege intuisjon for å finne at *kaderen* og *kadre* er meir sannsynlege bøyingsformer enn *kaden* og *kaderdre*.<sup>3</sup>

Oppbakkingsendingar inneber rimeleg nok ein meir komplisert morfologisk prosess enn ei enkel suffiksering. Det kan vere tale om konsonantforenklingar, men for *kadrer* var det vokalbortfall, samt eit tillegg av fleirtalssuffikset “-er”.

No er vi igang med å tolke bokstavrekkeendringane (strengoperasjonane) som morfologiske prosessar utførte på oppslagsordet. Det treng ikkje vere så dumt for ei ordbok, og det skal i det følgjande vise seg at *Bokmålsordboka* *krav* ei slik tolking. Årsaka finn vi i bøyingskodane.

## 2.3 Bøyingskodane

For bøyingstypar som er så vanlege at det finst meir enn omlag eit halvt tusen av dei, nyttar ordboka bøyingskodar. Til kodane finst ein kodenøkkel, og tabell 2.2 syner eit utdrag frå denne. I kodenøkkelene står paradigmatiske

<sup>3</sup> *Tanums store* har løyst problemet for lesaren med litt ekstra teiknsetjing: To punktum i ei bøyingsending gjev signal om at lesaren må leite bakover i oppslagsordet for å finne ønskt bøyingsform: **kader**, -en; ..drer, ..drene

bøyingar, og i rettleiinga [*Bokmålsordboka*, s. IX] vert lesaren oppfordra til å følgje eksempla (med høvelege endringar) for å finne ønskt bøyingsform.

f1	bru, brua, bruer, bruene pumpe, pumpa, pumper, pumpene
m1	stol, stolen, stoler, stolene bakke, bakken, bakker, bakkene pumpe, pumpen, pumper, pumpene
v1	kaste, kaster, kasta el. kastet, kasta el. kastet
v2	lyse, lyser, lyste, lyst

Tabell 2.2: Utdrag frå kodenøkkelen frå *Bokmålsordboka*

Ein bøyingstype kan innehalde ingen, ein eller fleire bøyingskodar. Kvar bøyingskode inngår på si side i fleire bøyingstypar. Det er rimeleg å tenkje seg at kvar bøyingskode åleine kan utgjere ein bøyingstype, men slik er det faktisk ikkje. Samtlege substantiv som i bunden form eintal kan få hokjønnsending, kan òg få anten hankjønnsending eller inkjekjønnsending, eller begge deler. Dette er nok grunnen til at nøkkelen inneheld to liner for oppslaget “**pumpe** f1 el. m1”. Det verkar kanskje forvirrande om ordet berre står under f1-nøkkelen, slik det gjorde i 1986-utgåva av ordboka. Med **bru** vert det litt annleis, for forma *bruen* er ei klammeform.

I denne seksjonen ønskjer eg å undersøke nærare tilhøvet mellom bøyingstypar med og utan kodar. Byrjar vi med litt reverskonstruering kan vi i første omgang finne tilbake til bøyingstypane som skjuler seg i nøklane. Resultatet er å finne i tabell 2.3.

<b>bru</b>	-a, -er, -ene
<b>pumpe</b>	-pa, -r, -ne
<b>stol</b>	-en, er, -ene
<b>bakke</b>	-n, -r, -ne
<b>pumpe</b>	-n, -r, -ne
<b>kaste</b>	-r, -ta el. -t, -ta el. -t

Tabell 2.3: Bøyingstypar i bøyingskodane

Slåande er det at dei to linene som høyrer til *f1* er ulike, og det same gjeld *m1*-linene. Men merk at med dei rette morfologibrillene på, er dei ikkje det likevel! Tabell 2.4 viser bøyingstypane som dei ville vore, om dei tok utgangspunkt i ordstammen i staden for oppslagsforma.



<b>bru</b>	-a, -er, -ene
<b>pump</b>	-a, -er, -ene
<b>stol</b>	-en, er, -ene
<b>bakk</b>	-en, -er, -ene
<b>pump</b>	-en, -er, -ene
<b>kast</b>	-er, -a el. -et, -a el. -et

Tabell 2.4: Bøyingstypar med utgangspunkt i ordstammen.

Så ryddig og fint at! Nyttar vi stammen, ser vi korleis dei to eksempela i nøkkelen *f1* tilhøyrer ein og den same bøyingstypen. Litt underleg vert det då å finne følgjande bøyingsinstruks i rettleiinga i *Bokmålsordboka*:

“*hinne* har koden *f1* el. *m1* og skal bøyes som *pumpe* eller *bakke*: *hinna*, *hinner*, *hinnene* eller *hinnen*, *hinner*, *hinnene*”

[*Bokmålsordboka*, s. IX]

Dette er forvirrande av di det heiter at to bøyingskodar til saman utgjer *koden* til *hinne* (eg nyttar termen *bøyingstype* for å unngå problemet). Misvisande er det òg, ikkje vert det fortalt at *hinne* også er som *bru* og *stol*, for ikkje å snakke om *tang*, *tapperhet* og *tatovering* og alle andre *f1/m1*-ord. Dermed ser det mest ut til at ordboka ønskjer å prente inn (meir overflatiske?) trekk som utsjånad og rytmeliskskap som vesentlege for lesaren. Ho leier såleis lesaren inn i sitt morfologiske teorigrunnlag gjennom ei bakkdør.

Kanskje ligg forklaringa på det tvetydige forholdet til eigen bøyingspraksis i at ordstammen er ein lingvistisk abstraksjon og i grunnen lite eigna som oppslagsform? Merk for eksempel at vi i i tabellen med ordstammene mista informasjon om den opprinnelege oppslagsforma, og vi måtta hatt ein kolonne til for å få den tilbake. Denne kolonnen ville igjen gjort linene innad i bøyingskodane ulike og dermed vist at til sjuande og sist er linene like berre så lenge vi vektlegg det som er likt. Dette har sjølvsagt kodemakarane sett, men dei har likevel valt å halde fast ved bøyingskodane, og det kan vel ikkje botne i anna enn eit djupt ønskje om å gjere desse bøyingsmønstra tydelege for lesaren.

Det er verdt å merkje seg at ein bonus ved denne føringsmåten, er at ordboka også får fram substantivas kjønn på eit vis som andre norske ordbøker ikkje maktar. Slik rettleier ordboka brukaren ikkje berre til å bøye oppslagsordet, men også til å velje høvande artikkel og egedomspronomen til ordet, alt i henhold til ordbokas kjønnssystem.

Før eg seier meg ferdig med gjennomgangen min av bøyingskodane, er det nokre detaljar eg vil gjere merksam på. Dei går alle på problem ein

støyter på når ein byrjar føre rekneskap på to ulike måtar, den eine med lingvistiske kategoriar, den andre utan.

Ordboka nyttar eit morfologisk informert tenkjesett for dei vanlegaste, produktive og svake bøyingsmønstra. Men ho gjer det ikkje heilt konsekvent. Ta til dømes **snill** *snilt, -e*. Dette adjektivet følgjer same bøyingstypen som *a1*-adjektiva, men står ikkje ført opp slik sidan vi har ortografiske reglar som krev forenkling av dobbelkonsonant føre “t”. Dermed ser **snill** ut som eit meir avvikande ord enn det reint morfologisk er.

Likeleis finst mange verb der stammen endar på dobbel-m (*brumme, stamme*). Sidan imperativforma er lik med stammen<sup>4</sup>, og sidan vi har ein ortografisk regel som ikkje tillet dobbel-m i slutten av ord, skulle ein kanskje tru at desse vart ført opp som mønsteravvikarar. Så er ikkje tilfellet for dei nemnte døma, dei har både bøyingstypen *v1*. Grunnen er at ordboka ikkje fører opp imperativsformene, og dermed kjem ikkje konsonantforenklinga til syne. **stamme** er visst ikkje like avvikande som det **snill** er.

Akkurat like avvikande som **snill** er derimot **stemme**, som følgjer *v2*-mønsteret, men har bøyingstypen “*stemte, stemt*”. Og dermed kunne ein for **temme** no vente seg ei blanding av desse to føringsmåtane, men det verbet er utstyrt med bøyingstypen “*v1* el. *v2*”.

Restriksjonen på dobbel-m i slutten av ord fører til enno fleire unntak. Når oppslagsordet sluttar på *m*, må det “fordobling” til for å finne einskilde bøyingsformer: Adjektivet **lam** og substantivet **program** følgjer høvesvis *a1* og *n3*, men får bøyingstypane “*-t, -me*” og “*-met, - el. -mer, -ma* el. *-mene*”.

Til sist vil eg peike på følgjande doble ordboksoppslag: “**kvi** el. **kvie** *v1* el. *v4*”. Dersom ein les ordboka bokstavleg, som eg har gjort, får **kvi** bøyingstypen “*v1* el. *v4*”. Men **kvi** høver ikkje med bøyingskoden *v1*, for i så fall skulle vi fått presensforma *kvier* for *v1*-koden.

## 2.4 Variantformer og leksikonstruktur

Eit sentralt ord i oppslaga i tabell 2.1 er konjunksjonen “el.”, og vi finn han i nesten halvparten av bøyingskodane òg. Det meste av bokmålets variasjonsrikdom vert forvalta med hjelp av dette vesle ordet, og det er på tide å sjå korleis det vert gjort.

No såg vi først i seksjonen om ordboksoppslaga (2.1) korleis konjunksjonen vert brukt mellom oppslag, så det har vi alt dekt. Vidare såg vi i førre

<sup>4</sup>Dei einaste unntaka er ord som ikkje har imperativsformer; ein del modale hjelpeverb, samt *less/lest*

seksjon korleis vi kan finne bøyingsstypene til bøyingskodane, så vi treng ikkje gå inn i dei. Dermed sit vi att med bruken inne i bøyingsstypene. Tabell 2.5 inneheld nokre døme. (Bøyingsstypar for ord som kan ha fleire kjønn utset eg til neste seksjon.)

**håpe** v1 el. v2

**skride** *skred* el. *skrei[d]*, *skredet*

Tabell 2.5: Bøyingsstypar med variantformer

Oppslaget til **skride** er ein bra stad å byrje. Dette er eit ord med tre alternative fortidsformer, den eine ei klammeform (var *d*-en i bogeparentes ville den vore ei sideform). Dersom verb vanlegvis har fem bøyingsformer — infinitiv, presens, preteritum, perfektum partisipp og imperativ — då har **skride** sju: *skride*, *skrider*, *skred*, *skrei*, *skreid*, *skredet*, *skrid*.

Verre er det å finne ut kor mange bøyingsformer **håpe** skal ha. Følgjer vi mekanisk det som står i bøyingsstypen ender vi opp med to presensformer, for både v1 og v2 har ei, men korfor skal vi det? Ser vi på bøyingsstypene til **kaie** og **I mil** i tabell 2.9, finn vi at **I mil** berre har ei form for ubunden fleirtal, medan hankjønnsbøyinga og hokjønnsbøyinga til **kaie** har fått kvar si. Det ser såleis ikkje ut til å vere noko gale i å reformulere bøyingsstypen til **håpe**. Ved å låse ut bøyingsformene som ligg i bøyingskodane får vi følgjande:

**håpe** *håpa* el. *håpet* el. *håpte*, *håpa* el. *håpet* el. *håpt*

Legg vi til ei presensform og ei for imperativ endar vi opp med ni former, og vi slepp uroe oss over kva som er skilnaden på identiske presensformer.

**håpe** har altså tre alternative former i preteritum og tre i perfektum partisipp. Gangar ein ut dei moglege kombinasjonane får ein ni stykke. Men er det rimeleg å lese bøyingsstypen som om det var så mange? Kan den som seier “*han hadde håpt...*” elles velje mellom tre preteritumsformer? Forpliktar ikkje brukarar av radikale *håpa* seg til å unngå konservatismen i perfektum partisipp?

For den einskilde bokmålsbrukar gjev svaret seg sjølv. Han klarar seg stort sett med ei variantform for kvar bøyning, men litt vakling er vanskeleg å unngå når ein har så mange former å velje mellom. Di lenger frå kvarandre dei aktuelle formene står, i ulike tekstar, i samtalar med ulike personar, di vanskelegare er det å halde sin stilistikk rein. Deskriptivt kan ein vente seg at dei fleste kombinasjonane finst.

Kva seier normeringsinstansane til dette? Det er snart lov å skrive *tred-*

*vesyv* på bokmål, så kanskje vert også dette spørsmålet overlata til bokmålsbrukaren og hans stilkjensle. I så fall får vi ein tydeleg og lettfatteleg modell: Utan band på kva former som høyrer saman, vil kvar “bøyning” verte ei mengd moglege bøyingsformer.

Alternativet er å normere bort stilforvirra kombinasjonar. Då vert ein til slutt sittande att med eitt eller fleire sett med samhörande former.

Det siste alternativet er det eg meiner eg finn i *Bokmålsordboka*. Når bøyingskoden v1 inneheld “kasta el. kastet, kasta el. kastet”, har eg alltid lese det som “kasta og kasta” *eller* “kastet og kastet”, og *Tabellmaterialet* seier seg einig i denne lesinga. For **håpe** spesifiserer ordboka i så fall tre normerte *sett*.

Men så langt er setta til **håpe** i ein viss forstand ufullstendige; dei inneheld berre to bøyingsformer. Og i kvart av dei tre normerte setta for **skride** finst berre ei form. Kva med dei andre? Er ikkje dei like tett knytte saman som formene i kvart av dei normerte variantsetta, og utgjer dei ikkje like mykje eit normert sett? Jo, det må vi vel seie at dei gjer.

Lat oss prøve å skildre tilhøvet mellom det obligatoriske og dei valfrie setta og korleis dei til saman deler oppgåvene segimellom med ein tabell (2.6).

inf	pres	pret	perf part	imp
		håpa	håpa	
håpe	håper	håpet	håpet	håp
		håpte	håpt	

Tabell 2.6: Variantformer og obligatoriske former

Dette er ikkje akkurat atomfysikk, men eg presenterer likevel ei modifisert utgåve av den same tabellen i 2.7. Skilnaden er at eg har bytt ut fullformene med bøyingsendingar.

inf	pres	pret	perf part	imp
		a	a	<b>1</b>
e	er	et	et	<b>2</b>
		te	t	<b>3</b>

Tabell 2.7: Variantendingar og obligatoriske endingar

Visuelt og forståingsmessig er dette reine barnematen. Vi har formene/endingane spreidde over tre rader som gjev dei normerte variantsetta,

og fem kolonner som gjev bøyingsinformasjonen (som ikkje har med sjølve bokstavrekkeoperasjonen å gjere). Litt verre er det å sjå korleis dette kan late seg gjere implementasjonsmessig; vi har tre liner i tabellen, men berre ei celle i nokre av kolonnene. Sjå det slik:

Bakanfor tabell 2.7 ligg, kan vi tenkje oss, alle oppslagsorda i *Bokmålsordboka* i ei klynge. Delmengder av desse vert bøyde etter mønster knytt til ei eller to eller alle tre radene i tabellen, og **håpe** er eitt av dei som er knytt til alle tre.

Leitar vi etter presensforma *håper*, ser vi først etter om det finst ei celle som dekkjer alle tre radene i **pres**-kolonnen. Det gjer det. Like heldige er vi ikkje om vi ser etter preteritumsformer. Når ikkje ei celle dekkjer alle tre radene, leitar vi vidare i alle moglege subsett av radene etter bøyingsformer: Vi kan sjå etter i radene 1+2, 2+3, 1+3, 1, 2 eller 3 for bøyingsformer. Og i dei tre siste tilfella finn vi ei normert form.

Slik kan vi altså førestille oss korleis vi finn fram til både obligatoriske og valfrie former. Lesaren skjønner nok at det har sin pris: Skal vi handsome data som er både like og ulike på same stund, krev det ei investering i meir kompleks struktur enn ein ønskte seg i utgangspunktet.

For nokre vert prisen for høg, kanskje er det ikkje absolutt nødvendig å få med eit korrekt tal bøyingsformer. Eit døme på dette er *Bokmålsordboka*, som ikkje akkurat engstar seg for at folk skal finne rette presensforma til **håpe**. Eit anna døme er *Tabellmaterialet*. Eg skal sjølv sagt seie meir om det om litt, men eg synest det høver godt å nemne her korleis dei har løyst problemet med variantformene.

I *Tabellmaterialet* finst ein ryddig og tydeleg modell, i tabellform naturleg nok, og vi kan godt illustrere den med dei to føregåande tabellane (2.6, 2.7). Hovudskilnaden mellom det vi finn i *Tabellmaterialet* og desse går på at det ikkje finst noko bakanforliggende omgrep om celler som vert delt mellom radene. Ein kan gjere oppslag på (**pres** × **1**) og (**pres** × **3**), og begge returnerer *håper*. Det gjer dei fordi kvar rad i andre kolonnen er fylt ut (med det same innhaldet). Løysinga har sine egne problem: Med på kjøpet får ein to ekstra presensformer til **håpe**, *håper* og *håper*, i tillegg til den ein kunne rekne med å få (*håper*).<sup>5</sup> Det er ikkje dermed sagt at *Tabellmaterialet* borgar for eksistensen av alle desse like bøyingsformene.

---

<sup>5</sup>Faktisk er det fire ulike like presensformer for **håpe** i *Tabellmaterialet*, og utrulege seks for **skride**.

## 2.5 Radikale og konservative former

“Bokmål er eit trekjønna språk (som nynorsk), og det er eit tokjønna språk (som dansk).” Setninga let seg knapt tenkje eller uttrykkje, sjølvmotseiande som ho er, og vil ein sleppe fri frå sjølvnektinga må ho skrivast om. Betre går det å seie at bokmål er ei samling norske dialekter der nokre har tre kjønn og andre to, men då står ein i fare for å misse bokmålet som *ein* storleik.

Det er vanskeleg å fange bokmålet som ei eining i fritext, og situasjonen er ikkje mindre komplisert for den som prøver fange det i ein robust, ein-skapleg og heilskapleg struktur. Nøkkelordet i *Bokmålsordboka* er igjen “el.”, og vi såg i førre seksjon korleis det vert brukt for å strukturere oppslag med både *-et* og *-a*-endingar.

Når det gjeld substantiva, vert kjønnsidentiteten ein ytterlegare faktor som må takast med. I bøyingskodane for han- og hokjønnsst substantiva (tabell 2.2) ser vi at det ikkje lenger nyttar å ha radikale og konservative former i same kodane.

Ordboka opererer med eitt kjønnssystem, eit trekjønnsystem, og det er minst tre svært gode grunnar for det. Først og fremst er det nokså lett å representere to kjønn om ein rår over tre, det omvendte er umogleg.

Dernest, og om lag like viktig, er det faktiske trekjønnsystemet og det tenkte tokjønnsystemet svært like kvarandre. Vi ser dette best i hankjønnsorda frå trekjønnsystemet: “en fold” er “en fold” for heile bokmålsfolket. Det finst ulikskapar også for inkjekjønnsformene, men dei let seg flette like fint inn i bøyingskodane som radikale og konservative verbformer.

Meir iaugnefallande vert ulikskapen når vi ser på orda som kan stå til substantiva, for artikkel og possessiv som *min*, *mi*, *en* og *ei* kan avsløre kjønnnet.<sup>6</sup> Som oppslagsord i ordboka står likevel desse orda åleine, og oppslaga deira er rommelege nok for både to- og trekjønnsformål.

Vanskane med valet av trekjønnsystemet og med inkorporeringa av radikale og konservative dialekter i same språket, dukkar først opp i samband med femininumsubstantiva. Men dei er ikkje større enn at lesaren nok alt veit korleis dei kan løysast: Hokjønnsorda finst jo òg i tokjønnsystemet, dei vert berre bøyge som hankjønnsorda (frå trekjønnsystemet).

Eg har lagt vekt på kor godt systema høver saman i trekjønnsystemet. Den tredje grunnen for å velje eit trekjønnsystem er eigentleg ei anna side

---

<sup>6</sup>Men dei treng ikkje gjere det. I følge *Bokmålsordboka* og *Norsk referansegrammatikk* kan både *min* og *en* vere hokjønnsord. Dette har ikkje *Norsk Ordbank* fått med seg. Den som søkjer etter bøyge former for “en” i ordboka på nettet får berre sjå hankjønnsforma uttrekt. For å sjå at det også kan vere ei hokjønnsform må ein lese ordboksopplaget.

av same saka: I ei ordbok vert det alltid trangt om plassen, og kortfatta uttrykksmåtar vert følgjeleg ettertrakta.

Med ordboka sine tre moglege kjønn kan kvart substantiv få ein av i alt sju moglege kjønnskombinasjonar. Desse har eg sett opp i tabell 2.8. Her fortel andre kolonnen kor mange bøyingsstypar det finst for denne kjønnskombinasjonen, og tredje kolonnen er meint å gje eit inntrykk av kor mange oppslagsord kombinasjonen har. Tala i den siste er ikkje til å stole på, det har med kodinga av ordboka i *Tabellmaterialet* å gjere, men mengdeforholda mellom dei er ikkje av vegen, og dei er også interessante. For dei som finn sjeldne unntak interessante, vil eg opplyse om at det er **gamla** og **smukka** som er hokjønnsst substantiv, og **III flo** som tar nøytrum eller femininum.

kjønn	bøyingsstypar	oppslagsord
mask	90	34500
nøyt	67	14890
fem	2	2
mask-nøyt	27	980
mask-fem	33	12034
mask-nøyt-fem	9	100
nøyt-fem	1	1
totalt	231	62508

Tabell 2.8: Fordeling av kjønn på substantivbøyingsstypar (inkludert typar for propria) i *Bokmålsordboka*. Fleirsifra oppslagsordstal er mykje for høge, medan dei einsifra er for låge.

Tabellen viser at kjønnsidentitetsforvirringa ikkje berre gjeld dei nemnde hokjønnsformene og radikal kontra konservativ. Det er i grunnen eit ålment problem: Sjølv om det eine eller andre systemet over tid skulle verte eignerådande vil det fortsatt vere rom for substantiv som ikkje heilt klarar avgjere kva kjønn dei høyrer til.<sup>7</sup>

Dei substantiva som har funne seg til rette i ei av dei tre første radene i tabellen er uproblematiske. Finst det variantformer i desse, oppfører dei seg som andre bøyingsstypar med variantformer, og slike tok eg for meg i førre seksjon.

Eksempel på fleirkjønna bøyingsstypar har eg sett i tabell 2.9. For betre

<sup>7</sup>Sjølv lir eg av ein lei tendens i nynorsken min til å velje hokjønnsformene for ubunden form fleirtal av ord som *klasse* og *stamme* (begge har bøyingsstipe f2/m1), sjølv om eg utmerka godt veit, dersom eg tenkjer meg om, at dei eigentleg er hankjønnsord.

å skildre problema vi får i samband med desse, skal eg gjere nesten som i førre seksjon. Eg skal setje dei ulike formene direkte i tabell 2.10, men utan å først dekode og trekkje saman bøyingsstypene. (Det vil seie: Eg slettar ikkje former med like skriftbilette.)

**kaie** f1 el. m1  
**I mil** -a el. -en, -, -ene

Tabell 2.9: Oppslag der bøyingsstypene inneheld fleire kjønn

ub ein	b ein	ub fl	b fl	
kaie	kaia	kaier	kaiene	<b>1</b>
kaie	kaien	kaier	kaiene	<b>2</b>
mil	mila	mil	milene	<b>3</b>
	milen		<b>4</b>	

Tabell 2.10: Fullformene frå førre tabell

Vi ser med ein gong at einskilde former vert delte mellom nokre rader, medan andre like former ikkje vert delte. Les vi dette bokstavleg — og det er heilt sikkert ikkje meint at vi skal gjere det, forstår vi at det finst ei *kaie* i hankjønn og enno ei i hokjønn, men berre ei *mil*. Kva kjønn har den siste?, om den i det heile har noko?

Tabellradene er her berarar av grammatisk relevant informasjon. Første og tredje rada inneheld hokjønnsformer, andre og fjerde rada har hankjønnsformer. Slik var det ikkje for tabellradene i førre seksjonen, dei inneheldt berre stilistisk, eller dialektal, informasjon. For tabell 2.6 kunne vi følgjeleg trekkje saman like former straks vi såg dei utan å misse noko, det kan vi ikkje her.

I staden burde vi kanskje gjere det motsette, leggje inn kjønna former der ordboka (av plassomsyn?) tilsynelatande har valt ei elliptisk føring. (Vi kan sjølvsagt fortsatt slette like former dersom dei har same kjønnnet.) Følgjande sitat frå *Norsk referansegrammatikk* understøttar ei slik tolking:

Genus [...] er en inherent egenskap ved substantivet. [...] Genus er knyttet til stammen i ordet, og normalt vil et substantiv høre til ett av disse tre genus. [...] Substantivet skifter normalt heller ikke genus ved bøyning, og genus er derfor ikke en bøyingskategori ved substantivet på linje med tall- og bestemthetskategorien.

[*Norsk referansegrammatikk*, s. 149-150]



Følgjer vi det norske språkets (ei enno vanskelegare eining!) fremste autoritet vert vi leidde til å ta med eit dobbelt sett kjønna former for kvart femte substantivoppslagsord. Det er nok ein god idé for ei ordbok som skal vere supplementsbind til brukarens eige mentale leksikon, og *Tabellmaterialet* har fulgt opp denne tenkjinga. Resultatet vert mellom anna fleirtalsformene *kai-er* (hankjønn) og *kaier* (hokjønn), som elles tyder det same og vert brukt likt. I min implementasjon leier det til at parsing av bokstavrekkeja “kaier” fører til fem strukturar: Fleirtal hankjønn og hokjønn for leksema *kai* og *kaie*, samt presensforma til verbet *kaie*.



## Kapittel 3

# *Tabellmaterialet*

Her går eg stort sett berre gjennom *Tabellmaterialet*, så lesaren bør kanskje heller gå seg ein tur enn å lese dette. Eg refererer seinare til nokre av tabellane, så for dokumentasjonen sin del må det med.

Nokre av tabellane eg har arbeidd med har eit format som eignar seg dårleg for auget. Eg vil derfor presisere at det ikkje er slik i *Norsk Ordbank*, for det er ein database, og formatet vert då mykje meir fleksibelt. Ingen fornuftige personar vil for eksempel stire særleg lenge på tabellen eg kallar *Boyingslista* (sjå tabell 3.6).

Tabellnamn	kolonneinnhald
Lemmalista	lemma-id, lemma-grunnform
Paradigmelista	paradigme-id, bøyingsgruppe, ordklasse, utdju- pande informasjon om ordklassen, forklaring, døme <sup>1</sup>
Lemma-par-lista	lemma-id, paradigme-id
Bøyingsendingslista	paradigme-id, bøyingsnr, -gruppe og -ending
Bøyingskategorilista	bøyingsnr, -gruppe, grammatisk info

Tabell 3.1: Oversyn over *Tabellmaterialet*

### 3.1 Lemmalista

I *Tabellmaterialet* finst ei nummerert liste med alle oppslagsorda frå ordboka. Dette inneber at *Tabellmaterialet* følgjer ordboka i å la ei bøyingsform stå som ein slags innfallsport til og representant for alle bøyingsformene som

høyrrer til eit oppslag. For substantiv er det skriftbiletet til den ubundne eintalsforma som står som oppslagsord, for verb er det infinitivsforma, og så frametter.

## 3.2 Paradigmelista

Den neste tabellen er ei nummerert liste over alle *paradigma* som finst. I alt er det 644 paradigme, og dei er koda i tresifra talkombinasjonar frå 001 til 999 i eit 26-talssystem!<sup>2</sup> Eit paradigme fortel kva ordklasse oppslagsordet tilhøyrrer og korleis oppslagetts ulike bøyingsformer skal vere.

Samanhengen mellom det eg i 2.1 kalla bøyingstypen og paradigma er som følgjer: Lemmaet får i regelen like mange paradigme som det finst normerte sett bøyingsformer til oppslagsordet. Eg seier *i regelen*, for på same måten som lemmalista inkluderer unormerte, men vanlege skrivemåtar, vert unormerte bøyingsformer tillagt gjennom at lemma vert knytt til fleire paradigme enn det bøyingstypen skulle tilseie.

I tabell 3.2 finst eit utdrag frå paradigmatelista. Paradigmatet kodar ord som tilhøyrrer ordklassen gjeve gjennom det første ordet i siste kolonnen. Resten av denne kolonnen kodar grammatisk informasjon som gjeld nett det paradigmatet. Ettersom ikkje alle ord i same klassen har likt tal bøyingsformer (for eksempel skil *LITE<sub>adjektiv</sub>* seg frå andre adjektiv ved å ha ei hokjønnsform, *lita*), vert ordklassen knytt til ei eller fleire bøyingsgrupper, alt etter korleis dei tilhøyrrande orda skal bøyast.

Kvart paradigme høyrrer til ei, og berre ei, bøyingsgruppe av i alt nitten, og berre ein ordklasse av i alt elleve (infinitivsmerket har fått sin eigen). Eg vil gjere merksam på at nokre bøyingsgrupper med “ubøy” i namnet ikkje er teke med i HPSG-leksikonet. Tabell 3.3 viser alle bøyingsgruppene, dei tilhøyrrande paradigma, og kor mange bøyingsformer gruppa har. Bøyingsgruppas ordklasse finn vi i første ordet i bøyingsgruppenamnet.

---

<sup>2</sup>Eg trur det starta som eit heksadesimalt system, og at det etterkvart trongst nokre fleire bokstavar. Grunnen er at dei ulike ordklassane har fått avgrensa område, og nokre har vore for små til å omfatte alle paradigma som etterkvart måtte innlemmast. Løysinga var å utvide området ved å innføre fleire siffer.

PARADIGME-ID	BØY-GRUPPE	ORDKLASSE
001	verb, normal	verb
020	verb, normal	verb
450	pronomen, normal	pron pers 1 ent hum
459	pronomen, normal	pron refl
47F	determinativ	det poss res
500	adjektiv	adj
515	adjektiv	adj <ordenstall>
691	preposisjon	prep
695	adverb	adv
696	infinitivmerke	inf-merke
700	substantiv	subst mask appell
902	substantiv	subst fem appell
995	substantiv, ubøy	subst fem prop

Tabell 3.2: Utdrag frå *Paradigmelista*.

bøyingsgruppe	tal paradigme	tal bøyingar
adjektiv	51	7
adjektiv, adverbtype	10	3
adjektiv, ulik m/f	2	8
adverb	1	ub
determinativ	32	4
determinativ, <adj>	5	5
determinativ, enkel	14	ub
infinitivsmærke	1	ub
interjeksjon	1	ub
konjunksjon	3	ub
preposisjon	1	ub
pronomen, enkel	11	ub
pronomen, normal	14	2
subjunksjon	2	ub
substantiv	188	4
substantiv, ubøy	10	ub
verb, normal	267	11
verb, s-verb	30	5
verb, ubøy	1	ub
	644	49

Tabell 3.3: Bøyingsgruppene og kor mange paradigme dei omfattar. “ub” i tredje kolonnen fortel at orda som tilhøyrer denne gruppa ikkje bøyast.

### 3.3 Lemma-paradigme-lista

Denne tabellen er absolutt nødvendig for å få *Tabellmaterialet* til å hengje saman, her vert tilhøvet mellom paradigma og lemma bokført. Ettersom mange oppslagsord har same skriftbiletet, er det lemmanummera som vert para med paradigmanummera. Kvar line i lista inneheld ei parvis oppføring, eit lemmanummer og eit paradigmenummer, så dersom eit lemma er knytt til fleire paradigme, vert det ført over fleire liner. (Det motsette er sjølvstøtt og så tilfelle: Dersom eit paradigme er knytt til fleire lemma, vert det å finne i fleire liner.)

### 3.4 Bøyingskategorilista

Meir ekstern informasjon enn i lista frå førre seksjon, er å finne i *bøyingskategorilista*. Ho fortener derfor å få synt fram nokre liner, noko som er gjort i tabell 3.4. Her ser vi alle linene som gjeld ei bøyingsgruppe, og seinare skal vi sjå korleis dei kjem til nytte når vi byggjer formendringstabellar. For adjektiv som høyrer til bøyingsgruppa *adjektiv*, er første bøyingsforma ei positiv ubunden eintalsform, og den kan stå til maskuline eller feminine kompliment. Og så frametter.

BØY-NUMMER	BØY-GRUPPE	BØY-TEKST
1	adjektiv	pos m/f ub ent
2	adjektiv	pos fl
3	adjektiv	pos be ent
4	adjektiv	pos nøyt ub ent
5	adjektiv	komp
6	adjektiv	sup ub
7	adjektiv	sup be

Tabell 3.4: Bøyingskategorilista

### 3.5 Bøyingsendingslista

Tabell 3.5 inneheld alle bøyingsendingar for alle bøyingsnummer for alle paradigme. I og med at paradigmenummeret via paradigmelista (seksjon 3.2) fortel kva bøyingsgruppe paradigmet tilhøyrer, er tredje kolonnen berre

for tabulatorisk sukker å rekne: Den sparar lesaren for eit ekstra oppslag i nemnde tabell.

Har ein no eit lemma til paradigme 500, *adaptiv* til dømes, kan ein plugge inn dette for bøyingsnummer ein (oppslagsorda tilhøyrrer mest alltid første bøyingsnummeret, unntak er berre *klær* og nokre andre). Dersom feltet for endinga (i sistekolonnen) er tomt, tyder det at dei andre bøyingsendingane til same paradigmet er reine suffiks. Inneheld feltet ein eller fleire bokstavar, har ikkje dette noko å seie for bøyingsnummer ein. Men for dei andre fortel dette at det same talet bokstavar skal hukast av bokstavrekkja til fordel for dei nye endingane. Når eit ord ikkje kan bøyast i alle kategoriane frå *Bøyingskategorilista* står det ein bindestrek i endingsfeltet. Tabellen har eitt døme på dette, frå paradigmet for nøytrumsfleirtalsord som *klær*.

PARADIGME-ID	BØY-NUMMER	BØY-GRUPPE	BØY-ENDING
500	1	adjektiv	
500	2	adjektiv	e
500	3	adjektiv	e
500	4	adjektiv	t
500	5	adjektiv	ere
500	6	adjektiv	est
500	7	adjektiv	este
520	1	adjektiv	e
520	2	adjektiv	e
520	3	adjektiv	e
520	4	adjektiv	t
520	5	adjektiv	ere
520	6	adjektiv	est
520	7	adjektiv	este
880	1	substantiv	-

Tabell 3.5: Bøyingsendingslista

Når eg held meg til berre adjektiv frå denne lista, og eg brukar informasjonen om bøyingskategoriane frå førre seksjon, er det ikkje vanskeleg å lage tabellar som viser kva formendingar paradigmet spesifiserer for lemmaet, sjå tabell 3.6.



par	p m/f ub ein	p fl	p b ein	p n ub ein	komp	sup ub	sup b
500		e	e	t	ere	est	este
520	e	e	e	t	ere	est	este

Tabell 3.6: Bøyingsendingstabell for to adjektivparadigme. *p* står for *positiv* og *n* for *nøyt*. Til liks med kodinga frå *Tabellmaterialet*, blandar eg latinske namn med norske for å sleppe å skrive “han/ho” o.s.f.



## Kapittel 4

# Bokmålstreet

I denne seksjonen skal vi gjere mykje forskjellig medan vi stegvis byggjer ein stor modell av heile *Tabellmaterialet*. Modellen eignar seg for datamaskinell implementasjon, og dermed er ikkje storleiken eit praktisk problem.

Undervegs skal vi definere fleire omgrep som høver til representasjonen av språk med mange variantformer. I seksjon 4.3 definerer vi *bøyingsgrupper*, og i seksjonen etterpå lagar vi ein leksemmodell. Vi gjev ein definisjon av ein morfologisk konsekvent bokmålstekst same stad, og i seksjon 4.8.2 ser vi korleis vi i tillegg kan definere ein stilistisk (eller dialektal) og morfologisk konsekvent bokmålstekst.

I seksjon 4.7 teiknar vi ein modell som representerer alle verba som *Norsk referansegrammatikk* identifiserer som svake verb. Denne modellen høver også godt til å forklare fonetiske/fonologiske samanhengar for språk med variantformer. I norsk samanheng vil eg i tillegg tru han eignar seg for å tenkje kring normeringsspørsmål.

Seinare, i seksjon 4.8.4, føreslår eg å byggje ein enno større modell som inkluderer nynorsk. Til bruk for eit felleleksikon i HPSG for bokmål og nynorsk vil modellen for eksempel eigne seg for generering av både nynorsk- og bokmålssetningar frå (meir eller mindre) same semantiske representasjonen.

Eg føreslår også å inkludere svensk og dansk i ein kjempemodell som vil mogleggjere systematiske studiar av morfologiske likskapar og ulikskapar mellom dei skandinaviske språka.

*Modellen* som mogleggjer alt dette, er ein såkalla rota og retta asyklisk graf, og for å forklare denne nærare, tek vi ein snartur innom litt grafteori i seksjon 4.4.

## 4.1 Eit augekast på paradigma

Frå *Tabellmaterialet* har vi paradigmet som lim mellom lemmaa og deira bøyingsformer: Kwart lemma er knytt til eitt eller fleire paradigme, som til saman gjev dei normerte fullformsetta til leksetet. Dersom ordet ikkje er bøyeleg, det kan vere eit adverb, så er det knytt til eit paradigme som ikkje spesifiserer bøyingar.

Vi såg i seksjonen om *Tabellmaterialet* korleis vi kunne lage formendringstabellar ut av *Bøyingsendingslista* og *Bøyingskategorilista*. I tabell 4.1 gjev eg eit nytt døme på ei slik liste, utvida med ein kolonne for døme, slik at vi kan studere nærare kva eit paradigme er. Lista må lesast på særlege vis, i likskap med alle slike vi kan lage frå dei nemnte listene: Leieforma er ferdig bøygd. Kolonnen for leieforma inneheld derfor ikkje ei formending, men bokstavane som må takast frå forma, slik at det står att ein slags stamme som dei andre kolonnene berre kan leggje sine endingar til. I neste seksjon skal vi sjå på ein betre representasjon av slike tabellar.

døme	inf	pres	pret	perf part	imp
kaste	e	er	a	a	<b>001</b>
koke	e	er	te	t	<b>020</b>
bøye	e	er	de	d	<b>030</b>
flå		r	dde	dd	<b>040</b>
legge	egge	egger	a	agt	egg <b>104</b>
rekke	e+++e	e+++er	a+te	a+t	e++ <b>109</b>

Tabell 4.1: Formendingsliste

Plussteiknet i denne tabellen er ein variabel. *Tabellmaterialet* inkluderer ikkje eksplisitte opplysingar om kva teikn som kan stå for variablane, men i dette tilfellet er det *k* og *l*. Eitt paradigme kan nytte opp til to ulike variabelteikn, pluss- og prosentteiknet. Plussen kan finnast opp til to gongar i same forma, og kodar då alltid ein dobbelkonsonant.

Paradigme 109 kodar fullformssettet til *rekke*, i tydinga å *strekke fram* eller *opp* eller *ut*, samt det eine av dei to fullformssetta til *telle*. Legg merke til at paradigmet spesifiserer både konsonantforenkling og omlyd.

**Leieform** *Tabellmaterialet* sine leieformer er *Bokmålsordboka* oppslagsord. Dette er eit naturleg val, for vi er alle godt kjende med leieformene frå før. Dessutan er oppslagsorda utmerka leksemrepresentantar til bruk for samanlikningar mellom ulike leksikon blant anna.

Likevel er det ikkje eit sjølvsgagt val, for det er ikkje gjeve at det beste valet for ei morfologisk skildring av dei ulike ordklassane alltid vil vere oppslagsordet.<sup>1</sup>

Eg har merka meg at *Norsk referansegrammatikk* sin oversikt over “de viktigste bøyingsmønstrene” til bokmålssubstantiva (s. 182-183) tek utgangspunkt i ubunden form fleirtal, men det vert ikkje fortalt korfor. For ei nynorskdialekt frå Sør-austlandet viser Endresen [1996] elegant korleis presensforma til verba inneheld mengder med informasjon om kva bøyingar verba kan undergå.

Det ville vore teknisk semiinteressant med ei studie i korleis ulike maskinelle transformasjonar av *Tabellmaterialet* kan ta seg ut. Ei slik studie kan ta utgangspunkt i at, for det første, formendingane i tabellar som den ovanfor gjev alle dei normerte fullformene til leksema, slik at det er lett å utleie nye leieformer. For det andre er formendingane reversible, så når vi har generert eit sett nye leieformer, kan vi nytte dei same formendingane til å få alle fullformer frå det nye settet.

Dette burde kunne gje fornuftige mål for kva vi bør forvente av leieformene til fullstendige leksikonrepresentasjonar av bokmålet.

## 4.2 Formendingstabell

Eg set tabell 4.1 oppatt på ny i tabell 4.2, men no er han vorten ein genuin og ein meir lettlesn *formendingstabell*.

døme	inf	pres	pret	perf part	imp	
kaste	like	suff-r	mod-e-a	mod-e-a	stripp-e	<b>001</b>
koke	like	suff-r	mod-e-te	mod-e-t	stripp-e	<b>020</b>
bøye	like	suff-r	mod-e-de	mod-e-d	stripp-e	<b>030</b>
flå	like	suff-r	suff-dde	suff-dd	like	<b>040</b>
legge	like	suff-r	mod-egge-a	mod-egge-agt	stripp-e	<b>104</b>
rekke	like	suff-r	mod-e!n!ne-a!nte	mod-e!n!ne-a!nt	stripp-e	<b>109</b>

Tabell 4.2: Formendingstabell

Felta i den nye tabellen inneheld namn som fortel kva formendingsoperasjon dei står for. *like* står for at vi får den aktuelle forma utan å endre på leieforma. *suff-r* fortel at vi får presensforma ved å leggje til suffikset *-r*,

<sup>1</sup>Det let seg sjølvsgagt gjere å ha oppslagsorda som leksemrepresentantar, og likevel nytte ei anna form som leieform, men det vil resultere i mykje større leksikon.

og *mod-egge-a* seier at vi får fortidsforma ved å bytte ut bokstavane “egge” med “a”. *stripp-e* fortel at vi får imperativsformer for ein del paradigme ved å ta bort ein *-e*.

I staden for pluss- og prosentteikn, nyttar eg ein toteiknskombinasjon for å representere variablar. Grunnen er at LKB gjer det på dette viset. “!n” står i mine formendingar for ein av bokstavane i følgjande bokstavrekke: “dfgklmnpst”, så han er vidare enn han treng vere for dette paradigmet som berre treng bokstavane “k” og “l”, men eg nyttar same variabelen også for andre. I alt nytta eg meg av elleve ulike variablar, og eg rekna meg på førehand fram til at desse dekte alt eg trong, og at eg ikkje ville få nokre uheldige konsekvensar ved å blande dei slik eg ville blande dei. Eg kunne ha brukt anten fleire eller færre, men skjønnsmessig fann eg storleiken på dei elleve til å vere høvande.

Ser vi på tabellen og samanliknar han med den opprinnelege, ser vi at dersom vi skulle overteke direkte dei formendingane som *Tabellmaterialet* spesifiserer, ville vi endt opp med *mod-egge-egger* i staden for *suff-r* for presensformendinga til siste paradigmet. Mine formendingar vert ikkje berre enklare, dei tydeleggjer også kva morfologiske operasjonar som er involverte, og dei gjer det lettare å sjå kva som er regulære samt kva som ikkje er det. Her kan det vere to relevante innvendingar å komme med, og dei vil eg svare på med ein gong.

For det første reagerer lesaren kanskje på ordvalet mitt: Når vart vel pluss- og utropsteikna relevante for morfologien? Til dette har eg ikkje eit fullgodt svar, for ettersom *Tabellmaterialet* er så stort har eg ikkje handsama alt i detalj. Eg vil svare ved å seie at etter mitt skjønn er bruken av variablar i *Tabellmaterialet* verken for sparsommeleg eller for overdriven. I tilfellet ovanfor såg vi korleis ein vanleg norsk ortografisk operasjon, konsonantforenkling generelt, kunne kodast, og eg kunne ha gjeve døme på fordobling òg. Til andre tider vil eg tru variablane kan stå som generaliseringar over fonologiske einingar, men sjølv veit eg ikkje nok om fonologi til alltid å kunne seie kva einingar det er tale om.

Når eg ikkje engstar meg for variabelbruken eg ikkje kjenner til, har det sin bakgrunn i at eg, når eg treng det, kan finne grunngjeving for kodinga i *Norsk referansegrammatikk*. Eit døme:

Til denne gruppa [gruppe *a* frå klasse 2 svake verb, v2-verb o.fl. i *Bokmålsordboka*] hører også en del verb der stammen ender på dobbeltkonsonantene -kk, -ll, -mm[...]

[*Norsk referansegrammatikk*, s. 484]

No er det faktisk slik at bøyinga av bokmålsverba, uavhengig av klasse og

bøyingstype, avheng av dei siste bokstavane. Verbet *felle* får preteritumsforma *felte* og så frametter. Dette er sjølv sagt berre eit ortografisk unntak, og heldigvis for alle bokmålsbrukarar er det ikkje mange slike. I *Tabellmaterialet* må desse unntaka handsamast separat, og paradigme 021 kodar *2a*-verb på *-kk* og *-ll* (og fleire), medan verb på *-mm* igjen får eit eige paradigme, 022, for dei har i tillegg avvikande imperativsform.<sup>2</sup>

Den neste innvendinga går på kor lur samantrekkjinga er, står vi ikkje i fare for å misse informasjon når vi forenkler formendingane? Det gjer vi, men ettersom eg i min bruk av formendingane overtek *Tabellmaterialet* sine samband mellom leksem og paradigme, risikerer eg aldri at formendingane kan brukast for leksem dei ikkje allereie er eigna for. Følgjande observasjon er relevant her: Kwart paradigme plukkar ut ei delmengd av alle lemmaa frå *Lemmalista*, *dermed tek alle paradigma utgangspunkt i leieformene for leksema*. Når ein trekk saman formendingane som eg har gjort, finn ein at for mest alle paradigma finst ei formending som kan skildrast enklare enn det paradigmet sjølv gjer, utan at vi fundamentalt endrar formendinga, og heller ikkje effekten av ho.

Med innvendingane sette til sides, kan vi gle oss over følgjande konsekvens av samantrekkjinga: Vi får fram fleire *like* formendingar mellom paradigma enn det som fanst i *Tabellmaterialet* opprinneleg. *Tabellmaterialet* spesifiserer i alt 3824 formendingar dersom vi utelet leieformkolonnen.<sup>3</sup> Av desse er berre 1733 unike. Dette er ikkje det minste overraskande, for dei fleste verbparadigma til dømes, får ein presensforma ved å leggje til ein *r*.

Etter at eg har forenkla formendingane, sit eg att med 999 unike formendingar. Dette talet er det igjen mogleg å forminske ved å dele formendingsoperasjonen i to, ein del for bokstavrekkjeoperasjonen på skriftbiletet, og ein annan for den grammatiske delen. Då sit ein att med om lag 540 reine bokstavrekkjeoperasjonar for bokmål.<sup>4</sup>

<sup>2</sup>Alt dette er sjølv sagt frykteleg trøyttande og heilt sjølv sagt stoff. Kanskje vert det nett derfor så forbløffande lett å oversjå. Den som held seg til kodinga i *Tabellmaterialet* ser der slike unntak med ein gong (og vi skal møte meir interessante eksempel seinare), men sjølv eg som har full tilgang til kodinga og inngåande studert delar av den, måtte lese setninga mange gongar før eg la merkje til utelatinga.

<sup>3</sup>Nokre av desse er forbodne formendingar, det vil seie at den aktuelle fullforma ikkje er tilleten, til dømes har ikkje modale hjelpeverb imperativformer.

<sup>4</sup>Eg har implementert eitt leksikon der det til dømes fanst berre ein *suff-r*-bøyingsregel. Denne gav presens til dei fleste verba, laga den ubundne fleirtalsforma til ein del substantiv, samt komparativbøyinga til adverbet *lenge*. Leksikonet er spennande nok, men som skildring av bokmåls morfologien, er det så uvant at det ikkje tener til stort i denne oppgåva.

### 4.3 Paradigmetypen

Dersom eit oppslag i ordboka ikkje spesifiserer variantformer for oppslagsordet, klarar vi oss med ein einrada fullformstabell for å gje alle formene. Vi får eitt normert fullformssett til oppslagsordet, og i *Tabellmaterialet* vert det motsvarande lemmaet assosiert med eitt paradigmenummer,<sup>5</sup> og følgjeleg vert leksemet hjå oss assosiert med ei rad i ein formendingstabell.

Kvart leksem og kvar rad i formendingstabellane er knytt til eit paradigmenummer.<sup>6</sup> Det vil seie at kvart leksem og kvar formending i tabellane er knytte til eit paradigmenummer. Dersom ein vil studere korleis dei ulike leksema, formendingane og kategoriane står i høve til kvarandre, kan dette såleis uttrykkjast ved hjelp av paradigmenummera.

Til dette formålet introduserer vi no eininga *paradigmetypen*: Kvart paradigmenummer er, seier vi, ein paradigmetype. Då kan vi seie at det og det leksemet, den og den formendinga og den og den kategorien, er knytte til ein og same paradigmetype (eller ikkje).<sup>7</sup>

**Paradigmegrupper** Men dette er ikkje alt vi treng, for til dømes er litt over halvparten av leksema knytte til *fleire* paradigmetypar. Eit sett med paradigmetypar skal vi kalle ei *paradigmegruppe*. Alle paradigmatypar inneheld minst ein paradigmetype.

Vi ordnar typene i gruppene etter nummeret på paradigmetypane.<sup>8</sup> Ferdig ordna grupper gjev vi namn etter nummera på paradigmetypane som inngår i dei: Vi limar nummera saman med bindestrekar, og får til dømes paradigmatypene *020-030* og *020*. Det vert no klart ut frå namna på gruppene at dei er nærskyldige, og dermed slepp vi gå vegen om dei tilhøyrande

<sup>5</sup>I praksis er det ikkje alltid slik. Nokre ord som ofte vert bøygde utan omsyn til normeringa har fått fleire paradigmenummer enn dei får rader i fullformstabellar. Denne oppgåva handlar ikkje om normeringsspørsmål, så unntaka har ingenting å seie her.

<sup>6</sup>Det same gjeld kvar rad i *Paradigmelista* sjå seksjon 3.2, så det som skildrast i dette avsnittet gjeld også for dei grammatiske kategoriane i denne lista.

<sup>7</sup>I skrivande stund (desember 2003) er det straks eit halvt år sidan rettleiaren min, Jan Tore Lønning, spurde meg kva eit paradigmatyp er. Her gjev eg eit *instrumentelt* svar på spørsmålet. Dersom definisjonane eg gjev i denne oppgåva skal vere av meir enn teknisk interesse, er det nødvendig å finne ei lingvistisk tolking av paradigmatypen. (I grunnen er det eit lite kledeleg namn, så det ville gle meg om det kan byttast ut: Eg brukar det berre av di eg tidleg i arbeidet med oppgåva jobba med HPSG-typar.) Sjølv har eg hatt så mykje moro av å utforske praktiske følgjer av den instrumentelle definisjonen, at eg dessverre ikkje har tenkt noko særleg vidare over dette kjernes spørsmålet.

<sup>8</sup>Dette er strengt tatt ikkje nødvendig, men for programmeraren som skal skrive algoritmer som kan operere på mengda av typar, er det fint at dei er ordna i sett.



setta med formendingar for å finne at det eine settet er eit subsett av det andre.

**Bøyingsgrupper** Alle dei paradigmegruppene som er tilknytte til eitt (eller fleire) leksem, skal vi no kalle for *bøyingsgrupper*.<sup>9</sup> Verbet *lyse* er knytt til berre ei paradigmegruppe, *020*, så vi kallar *020* for ei bøyingsgruppe. Verbet *føde* er knytt til eit sett med to paradigmatypar, *020* og *030*, så paradigmegruppa *020-030* vert også ei bøyingsgruppe.

Bøyingsgrupper som omfattar fleire paradigmatypar skal vi kalle for *fleirtypige bøyingsgrupper*. Bøyingsgrupper med berre ein paradigmatype skal vi kalle *eintypige* eller *enkle*. For dei siste kan vi lage einrada formendingstabellar, for dei første kan vi lage tabellar med like mange rader som det er paradigmatypar i bøyingsgruppa.

**Kva har hendt?** Eg kjenner meg viss på at termen *bøyingsgruppe* gjev eit bra verkty for representasjonen av språk med mange variantformer. Med dei grepa vi har teke i denne seksjonen, trillar det ut lange rekkjer av flunkande nye opplysingar om bokmål frå *Tabellmaterialet*.

I alt får vi om lag 600 bøyingsgrupper som eit utgangspunkt for bokmålsrepresentasjonen.<sup>10</sup> Det kan verke overraskande at det er færre bøyingstypar enn paradigme. Årsaka ligg først og fremst i bøyingsgrupper som *700-900*,<sup>11</sup> som kodar appellativa som kan vere både hokjønn og hankjønn (eller felleskjønn i tokjønnsystemet). Det finst knapt noko hokjønnsbøyingsgruppe.

Vi får vidare ein ny innfallsvinkel til kva det vil seie å vere eit regulært bøyingsmønster: Morfologisk gjev det god mening å utpeike paradigme 126, som kodar verbet *gjøre*, som irregulært. Men *Lemma-paradigme-lista* inneheld heile 79 forbindingar mellom paradigmenummeret og lemmaidentifikasjonsnummer.<sup>12</sup> Bøyingsgruppa *126* har såleis mange fleire tilknytte leksem enn gruppa *001-010-011-020-030* for eksempel, som berre omfattar *love* og *utlove*, og bøyingsmønsteret til *126* er mykje meir utbreidd enn den siste gruppas bøyingsmønster.

---

<sup>9</sup>Lesaren kan merkje seg at nemninga ikkje er heilt velvalt. Preposisjonar og så frametter vert ikkje bøyde, men dei tilhøyrer like fullt bøyingsgrupper.

<sup>10</sup>Dette talet er det mogleg å redusere, men kor mykje veit eg ikkje. Reduksjonen avheng av måten formendingane er ført på, og eg har ikkje hatt mekanismar for, eller tid til, å utforske dette i særleg grad.

<sup>11</sup>*700-900* har 20706 tilhøyrande substantivleksem, deriblant *skje*.

<sup>12</sup>Det er verdt å merkje at mest alle desse er bokført i *Bokmålsordboka*, så vi talar stort sett om det leksikografar kallar *ikkje-føreseielege* samansetjingar.

Eit liknande døme som viser at bøyingsgruppene er nyttige for dei som forvaltar bokmålsorda: Mykje over 4000 verb i *Tabellmaterialet* tilhøyrer bøyingsgrupper som står i direkte relasjon til paradigmatypen *001*. I alt er det 27 bøyingsgrupper der typen *001* finst. Blant desse finn vi at også *001* utgjer ei bøyingsgruppe for seg sjølv, og berre er knytt til eitt verb.<sup>13</sup> Den som har koda dette verbet har i effekt utvida språkrepresentasjonen med ei ny bøyingsgruppe. Herfrå kan vi slutte at anten har vedkommande hatt noko heilt spesielt i tankane, eller så er det feilkoda.

#### 4.4 Paradigmegrupper og RRAG-ar

I neste seksjon skal vi teikne (oppned) tre der nodane representerer paradigmegrupper. Vi skal vidare skissere korleis eit kjempetre, *Bokmålstreet*, kan byggjast frå *Tabellmaterialet*.

Eigentleg er det ikkje tale om eit tre, men om ein rota og retta asyklisk graf (pluss pluss), men for å sleppe unna nemninga *BokmålsRRAGen*, skal vi likevel kalle modellen for eit tre. Når det krevst, kan vi presisere nærare kva vi meiner på bakgrunn av det modellteoretiske grunnlaget som vi skildrar her.

Ein rota og retta graf er ein graf som i likskap med eit tre har ein toppnode øvst, og ein eller fleire bladnodar nedst. Ein graf der den einaste bladnoden også er toppnode, er også eit tre.

Skilnaden på ein rotretta graf og eit tre ligg i talet på forgreiningar vi kan ha *oppover*. Frå kvar trenode går det maks ei grein oppover, så vi får forgreining berre nedover. I grafen kan det vere forgreiningar både oppover og nedover.

Med forgreining oppover får vi òg syklar i grafen. Det er uheldig, for det gjev ikkje særleg god mening å tale om at ein frå ei paradigmatgruppe i grafen kan ta seg ein rundtur nedover og oppover (eller omvendt) via null eller fleire nodar for andre paradigmatgrupper, og så komme tilbake til same noden. Derfor seier vi eksplisitt at grafen skal forståast som ein graf med stiar som berre går nedover (eller oppover); då får vi ein asyklisk graf. (Og ettersom vi talar om ein asyklisk graf, kan det ikkje gå ei grein frå ein node til den same.)

Vi har altså ein *RRAG*, ein rota og retta asyklisk graf. Her er ei presisering til som er vel verdt å merkje seg: For grafar er det generelt slik at nodar som har andre nodar under seg (altså ikkje bladnoder), kan ha ein eller fleire nodar under seg.

<sup>13</sup>Dette verbet er det unormerte *smæsje* frå *Tabellmaterialet* sitt tillegg til ordboka.

Slik er det ikkje i *BokmålsRRAGen* (eller RRAGane) vi seinare skal teikne. Alle nodar som har nodar under seg i *BokmålsRRAGen* har minst to nodar under seg.

Dette er ein direkte konsekvens av at vi byggjer fleirtypige grupper av grupper med færre typar, og det har såleis ei fornuftig lingvistisk tolking: Ei eintypig bøyingsgruppe spesifiserer korleis vi kan få eitt normert fullformssett for tilhøyrande leksem. Ei totypig bøyingsgruppe spesifiserer korleis vi kan få to normerte fullformssett, og så frametter. Det ville gje dårleg mening om det for ei fleirtypig gruppe berre kunne gjevast eitt fullformssett. For *BokmålsRRAGen* ville det gje dårleg mening dersom det frå ein ikkje-bladnode berre gjekk ein sti nedover.<sup>14</sup>

**Høgdenivå i *BokmålsRRAGen*** Ein RRAG-bladnodes høgdenivå er 1. For dei andre nodane er talet på høgdenivået lik talet på bladnodar han har under seg i treet, forutan bladnodane som høyrer til nivå 1. Det er ein føresetnad for at denne definisjonen skal gje mening at vi har alle dei enkle paradigmegruppene som bladnodar i RRAG-en.

Dersom vi let ein node stå for paradigmegruppa og dei paradigmatypene som finst i gruppa, kan vi alternativt definere nodenivået som den tilhøyrande gruppa sitt kardinaltal (det er: talet på paradigmatypar i gruppa). For våre RRAG-ar gjev definisjonane same resultatet: Ei paradigmegruppe med tre bladnodar under seg kan assosierast med tre paradigmatypar (og tre sett med formendingar). Gruppa ligg på nivå 3, og det fortel oss at det er maksimalt to steg ned til første nivået, men i praksis kan det vere at alle greinene nedover leier til første nivået i eitt steg.<sup>15</sup>

For arbeidet med denne oppgåva har gruppenes kardinaltal vore særst nyttige i algoritmeskrivinga og -tenkjinga. Høgdenivået til ein node kan fortelje oss ulike ting, alt etter kva kategoriar i *BokmålsRRAGen* vi vel å fokusere på.

---

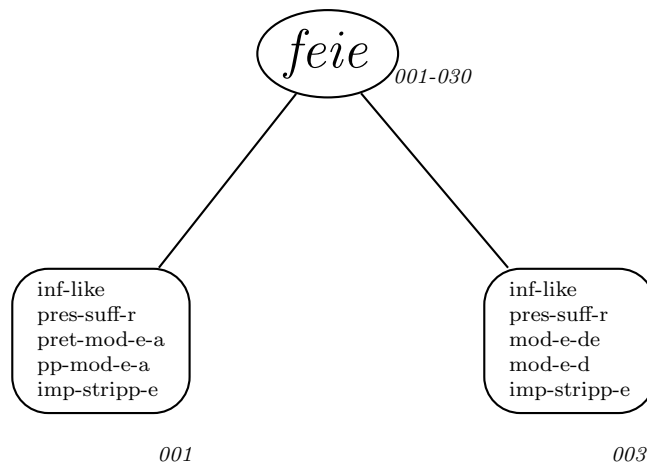
<sup>14</sup>Vi skal også byggje paradigmegrupper høgt i treet, og med desse kan vi uttrykkje generaliseringar over bøyingsgrupper. Den lingvistiske tolkinga vi her gjev er såleis berre eit særtilfelle av følgjande: Det gjev ikkje god mening at ein kategori frå ei fleirtypig gruppe (ei formending eller ein grammatisk kategori til dømes) berre skal gjelde for ei eintypig gruppe.

<sup>15</sup>Ein kontraintuitiv konsekvens av å definere høgdenivået slik for RRAG-ar generelt, er at toppnoden i ein RRAG kan hamne på eit nivå på fleire tusen (faktisk eit *uendelig* høgt nivå), men at det samstundes berre tek maksimalt to steg å gå ned til botnen. I *BokmålsRRAGen* er det likevel ikkje slik, så eg nøyer meg med desse definisjonane som ikkje høver RRAGar generelt.

## 4.5 Paradigmetre

Frå paradigmegruppene kan vi byggje modellar som hjelper oss å visualisere bokmålet, eller dei aspekta ved bokmålet vi vil fokusere på. Då kan vi stort sett leggje til sides førestellinga om paradigmegruppene som sett av paradigmatypar, og heller sjå på korleis tilhøvet mellom paradigmegruppene er. Men vi skal ikkje gløyme paradigmesetta heilt, for det er nyttig å kunne uttrykkje algoritmer ved hjelp av dei.

Det er soleklart at bøyingsgruppene har mykje til felles, det ser vi ut frå namna, og i figur 4.1 set eg opp ein modell som kan hjelpe oss å handtere likskapane. I rein iver over å støtte dei radikale formendingane frå den radikale paradigmatypen *001*,<sup>16</sup> utelet eg både no og seinare dei konservative endringane og typane, for slik å kunne lage så enkle modellar som mogleg.



Figur 4.1: Paradigmatypane *001*, *030* og *001-030*

Nedst finn vi dei eintypige paradigmegruppene, desse vert bladnodar i paradigmatreet, og over finn vi dei paradigmegruppene vi vel å innføre.

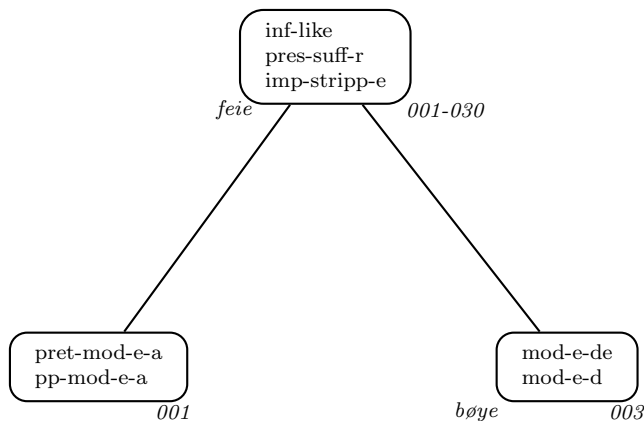
I treet finn vi tre paradigmegrupper. I øvste noden har vi bøyingsgruppa *001-030*, og i dei nedste finn vi paradigmegruppene *001* og *030*.

**Nok eit paradigmatre** Det er uheldig at det er fleire like formendingar i paradigmatreet. Derfor skal vi kvar gong to like formendingar høyrer til

<sup>16</sup>Eg vil i den samanhengen oppmode lesaren om å lese følgjande andre språkarbeidarar: Thure Erik Lund og Per Petterson.

same paradigmegruppene, og gruppene i tillegg er knytte til same fleirtypige gruppa, drage formendinga opp til den siste gruppas nivå i treet.<sup>17</sup>

For figur 4.2 har eg gjort nett dette. For betre å fokusere på formendingane, flyttar eg *bøyingsgrupperepresentanten feie* ut av trenoden til bøyingsgruppa 001-030. I tillegg har eg lagt til ein representant for bøyingsgruppa 030, for paradigmegruppe 030 er også ei bøyingsgruppe. Legg merke til korleis representantane kan bøyast ut frå alle formendingar vi kan finne ved å gå anten oppover eller nedover i treet. Det er forbode å gå opp ei grein og så nedover i ei anna grein (eller omvendt).



Figur 4.2: Paradigmetypene 001, 030 og 001-030

**Kva har hendt** Vi har no ressursane til ein leksemmodell som høver for språk med alternativformer. Eit leksem er knytt til dei formendingane som vi finn ved å gå opp i alle greinene frå bøyingstypens node, og ned i alle greinene frå den same noden.<sup>18</sup>

Dersom det frå ein node går fleire greiner oppover, vil alle greinene innehalde dei same formendingane. (Dersom det er fleire greiner oppover vil dei møtast høgare oppe, og i knutenodar vil formendingane liggje.) Nedover derimot, vil vi finne ulike formendingar. Eit normert fullformssett til leksemet finn vi ved å plukke med oss formendingar i kva som helst grein

<sup>17</sup>Vi skal modifisere denne skildringa når vi seinare byggjer meir innfløkte tre. For no er det heilt greitt å tenkje slik.

<sup>18</sup>Seinare skal vi forby å gå opp i treet og heller tale om arv av formendingar, men for no kan vi tenkje som her.

oppover, men maks *ei* grein nedover. Det er såleis ikkje hipp som happ kva nedtur vi vel.

No kan vi definere ein *morfologisk konsekvent tekst* som ein tekst der skribenten er konsekvent i sitt val av nedturar. Har han valt fortidsforma *feia* ein stad, kan han ikkje ein annan stad velje fortidsforma (eller perfektum partisipp-forma) *feide*. Tenkjer vi at skribenten for kvart ord han skal nytte, finn det aktuelle leksemets bøyingsgruppe i treet, må han halde seg til same greina når han vil nytte formendingar for leksemet som ligg i nodar lenger nede. På dette viset kan ei morfologisk skildring av eit språk med mange variantformer høve for den einskilde som berre treng delmengder av orda som finst i språket.

Ved hjelp av datalingvistens medtenkjar, elektronhjernen, kan vi no nytte settforståinga av paradigmegruppene til å lage algoritmer som kan brukast i praksis: For formendingar som kjem frå paradigmegrupper med like høgt eller høgare kardinaltal enn leksemets eiga bøyingsgruppe, er alt i orden. Men for formendingar som kjem frå paradigmegrupper med lågare kardinaltal, må dei ulike paradigmegruppene vere subsett av kvarandre. Slik kan vi maskinelt finne ut om ein tekst er morfologisk konsekvent. I seksjon 4.8.2 skal vi sjå korleis vi kan definere ein *stilistisk* (eller dialektal) og morfologisk konsekvent tekst.<sup>19</sup>

## 4.6 Klammeparadigme?

Her finst ein kompliserande faktor som eg vil diskutere før eg går vidare. Sjå på bøyingsgruppene i tabell 4.3.

Døme	Bokmålsordboka	Bøyingsgruppe
bie	v1 el. v4	001-043
greie	v3 [v4]	030-043
føye	v3 [v4]	030-043
bøye	v3	030
skrue	v4	043

Tabell 4.3: Klammer og bøyingsgrupper

Følgjande framgår ikkje av tabellen, men vi skal likevel ta det med oss før vi går vidare: *Samtlege* verb som endar på *-eie*, og som i tillegg har

<sup>19</sup>Det er sjølvsagt ein kompliserande faktor at skriftbileta for ulike fullformer som høyrer til ulike leksem kan vere like, men det treng vi ikkje bry oss om her.

ordboksbøyingstype med “[v4]”, er feilkoda. I alt dreier det seg om 21 stykke inkludert nokre få samansetjingar, men i tillegg kjem verba *føye*, *innføre* og *snøye*. Alle desse 24 er feilkoda i *Tabellmaterialet* med 040 i staden for 043.<sup>20</sup> Som ein konsekvens får vi preteritumsforma *dreiedde* og perfektum partisippforma *føyedd*. I tabellen kjem ikkje dette til syne, for eg har retta mistaka.<sup>21</sup>

Derimot fortel tabellen at nokre *v4*-former er sideformer (*greie*), og nokre er jamstelde former, heller bør nemnast hovudformer (*skrue*).<sup>22</sup>

I tillegg synleggjer dei to øvste linene at høgare normeringsmakter også er involvert: Her grip dei tydeleg inn og gjev påbod om at nokre *v4*-former skal vere klammeformer (*greie*). Paradigme 043 har i *Tabellmaterialet* fått arbeidet med å kode alle desse.

Dette er problematisk, for rolla *043* spelar i *001-043* ser ut til å vere annleis enn rolla i *030-043*. Men er den det? I tabell 4.4 set eg opp paradigma i ein formendingstabell slik at vi betre kan sjå kva som står på spel.

inf	pres	pret	perf part	imp	
like	suff-r	mod-e-a	mod-e-a	stripp-e	<b>001</b>
like	suff-r	suff-dde	suff-dd	like	<b>030</b>
like	suff-r	mod-e-dde	mod-e-dd	stripp-e	<b>043</b>

Tabell 4.4: Formendingstabell for førre tabell

Spørsmålet no vert om eg ikkje må splitte paradigme 043 opp i to nye paradigme, eitt for klammeformendinga, og eitt for sideformendinga? Dette er eit fundamentalt spørsmål som best kan svarast når vi kjenner svaret på eit anna spørsmål: Kva ønskjer vi av ein representasjon av bokmåls morfologien?

*Norsk referansegrammatikk* har sitt svar på dette, han skal “tjene som et oppslagsverk for alle dem som av ulike grunner har bruk for kunnskap om

<sup>20</sup> *dreie*, *kadreie*, *kleie*, *meie*, *pleie* samt *forføre* høyrer til ordbokas *v1*, *v3* [v4]. *Tabellmaterialet* har ikkje fått med seg at *pleie* også kan bøyast *pleia*. Likeleis finst konservative former i *Tabellmaterialet*, men fortsatt ikkje radikale ditto, til verba *breie*, *leie*, *sneie*, *ugreie*, trass i at desse ikkje har *v1* i bøyingstypen. Eg legg så mykje vekt på dette, for feilkodinga av slike problemkompleks i *Tabellmaterialet* er så mykje meir omfattande, enn det ho er for enklare område. I tillegg er feilkodinga systematisk, noko som fortel at det er ei eller anna form for orden i rotet.

<sup>21</sup> Eg meiner å hugse at eg for over eitt år sidan fortalde dette til ansvarlege på *Tekstlaboratoriet*, så kanskje er dette alt endra på.

<sup>22</sup> I *Norsk referansegrammatikk* står det å lese at *skrue* også kan bøyast etter klasse 1, så det er kanskje eit lite velvalt døme. Om lesaren ønskjer det, kan han bytte det ut med eitt av dei sjeldne verba som elles har denne bøyingstypen, for eksempel *lye* (*lytte*) eller *løe* (*stable*).

norsk språk” [*Norsk referansegrammatikk*, s. 1]. Det same svaret gjev *Bokmålsordboka* i og med undertittelen “Definisjons- og rettskrivningsordbok”. Slik underlegg dei seg tre herrar både to; Røyndommen, Pedagogikken og Normeringsmaktene, og det er knapt tenkjeleg at dei skulle kunne gjere noko anna.

Med tanke på inkonsekvensane vi fann i *Bokmålsordboka* (sjå seksjon 2.3), vil eg seie at ordboka lukkast betre i å følgje Normeringsmaktene sine dekret, enn i å følgje Røyndommen. At det same kan seiast om referansegrammatikken kan vi lese ut av følgjande:

Verbene *skade*, *skygge*, *tygge*, *love*, *dreie*, *feie* kan også bøyes etter 1. klasse (*dreie* og *greie* har også preteritum [-dde] og perfektum partisipp [-dd]).

[*Norsk referansegrammatikk*, s. 485]

Det kan vere tilfeldig, så eg er kanskje i ferd med å strekkje fortolkingsstrikken for langt, men likevel: *Dreie* kan altså bøyast etter 1. klasse. Korfor vert ikkje klammeformene med *-dde* og *-dd* identifiserte som medlem av klasse 2c? Er klammeformendingar ulike sideformendingar?

Svaret på det siste spørsmålet må verte nei. Ei formending er ei formending, og klammene er instrument som vert brukte for å leie Røyndommen inn i ei særskilt lei, slik at einskilde leksem kanskje ein dag ikkje lenger kan assosierast med einskilde formendingar.

Konklusjonen vert at i reine skildringar av Røyndommen høyrer ikkje klammene inn nokon stad. I høgda kan dei vere med som tilleggsopplysingar, eller kanskje som ønskjer for framtida.

## 4.7 Svake bokmålsverb

Lat oss byggje eit tre<sup>23</sup> som kan huse dei verba som *Norsk referansegrammatikk* kallar *svake verb*. Vi skal utelate dei som krev konsonantforenkling (*stemme*) og bortfall av *j* (*svelgje*, *følgje*), vi utelet òg dei verba som har alternativ imperativsform identisk med infinitivsforma (*boltre*). Vidare utelet vi verba som alternativt kan få *-dte* (*kle*) eller *-gde* (*vie*) i fortid, og vi utelet dei svake verba som også kan bøyast sterkt (*suge*, *tie*).<sup>24</sup> I første omgang skal vi i tillegg utelate ei gruppe verb med kortformsinfinitiv (*bo/bu*, *nå*).

<sup>23</sup>Sjå andre fotnota til seksjon 4.5.

<sup>24</sup>I effekt er dette å gå ut frå ei hypotese om at dei uteletne verbformendingane er morfologisk mindre relevante for representasjonen av dei svake verba enn formendingane vi sit att med. Dette kan forsvarast i alle fall for konsonantforenklingane, som berre er *ortografisk* ulike dei vi tek med. Eg er meir usikker når det gjeld dei andre.



Då sit vi att med bøyingsgruppene i tabell 4.5. Ein tilsvarande formendingstabell har eg sett opp i 4.6. Blant bøyingsgruppene som berre inneheld ein paradigmatype finn vi representantar for alle dei svake verbgruppene som *Norsk referansegrammatikk* opererer med.

Døme	Bokmålsordboka	Bøyingsgruppe
love	v1, v2 el. v3	001-020-030
buse	v1 el. v2	001-020
føde	v2 el. v3	020-030
dreie	v1, v3 [v4]	001-030-043
feie	v1 el. v3	001-030
føye	v3 [v4]	030-043
grue	v1 el. v4	001-043
kaste	v1	001
koke	v2	020
bøye	v3	030
løe	v4	043

Tabell 4.5: Bøyingsgrupper for svake verb (ikkje alle)

døme	inf	pres	pret	perf part	imp	
kaste	like	suff-r	mod-e-a	mod-e-a	stripp-e	<b>001</b>
koke	like	suff-r	mod-e-te	mod-e-t	stripp-e	<b>020</b>
bøye	like	suff-r	mod-e-de	mod-e-d	stripp-e	<b>030</b>
løe	like	suff-r	mod-e-dde	mod-e-dd	stripp-e	<b>043</b>

Tabell 4.6: Formendingstabell for svake verb (ikkje alle)

**Første modell** Eg sa ovanfor, i seksjon 4.5 då vi laga våre første paradigmatre, at vi kunne flytte opp ei formending eitt nivå i paradigmatreet når to formendingar finst for paradigmatregrupper som har ei sams fleirtypig paradigmatregruppe rett over.

Dersom dette var alt, ville vi ikkje kunne byggje eit paradigmatre over dei svake verba. Ut frå bøyingstypene i tabell 4.5, finn vi nemleg at det vert to paradigmatregrupper øvst i treet, *001-020-030* og *001-030-043*.

Vi må ta eit kunstgrep, men sidan vi skal teikne fleire store tre på dei neste sidene, vil eg her berre fortelje korleis vi kan gå fram for å finne ei ny toppgruppe som kan knyte saman heile treet.

Frå formendingstabell 4.6 plukkar vi ut formendingane som er like i infinitivskolonnen (i dette tilfellet er det alle, men eg formulerer meg generelt), og så merkjer vi oss kva *paradigmatypar* vi henta dei frå. Frå desse (fire) typane lagar vi så ei hittil ikkje-eksisterande paradigmegruppe *001-020-030-043*, og vi knyt formendinga *inf-like* til denne. Ettersom *001-020-030-043* ikkje er ei bøyingsgruppe, har vi ingen bøyingsgrupperepresentant vi kan knytte han til.

Dernest observerer vi at *001-020-030-043* er høveleg som toppgruppe over både *001-020-030* og *001-030-043*. *001-020-030* er knytt til paradigmatypane *001*, *020*, *030*, og det er sanneleg *001-020-030-043* òg, men denne er i tillegg knytt til *043*. Kvar gong vi finn slike samanhengar mellom to paradigmegrupper, skal vi teikne liner mellom dei.<sup>25</sup> Vi teiknar såleis ei line opp frå *001-020-030* til *001-020-030-043*, og vi teiknar ei anna frå *001-030-043*.

Så går vi laus på presenskolonnen. Igjen viser det seg at alle felta i kolonnen er like, og vi finn igjen paradigmatypa *001-020-030-043*, men når vi kik etter i treet vårt, finn vi at denne har vi alt. Følgjeleg legg vi enkelt og greitt til *pres-suff-r*-formendinga i toppnoden. Det same gjer vi for imperativkolonnen.

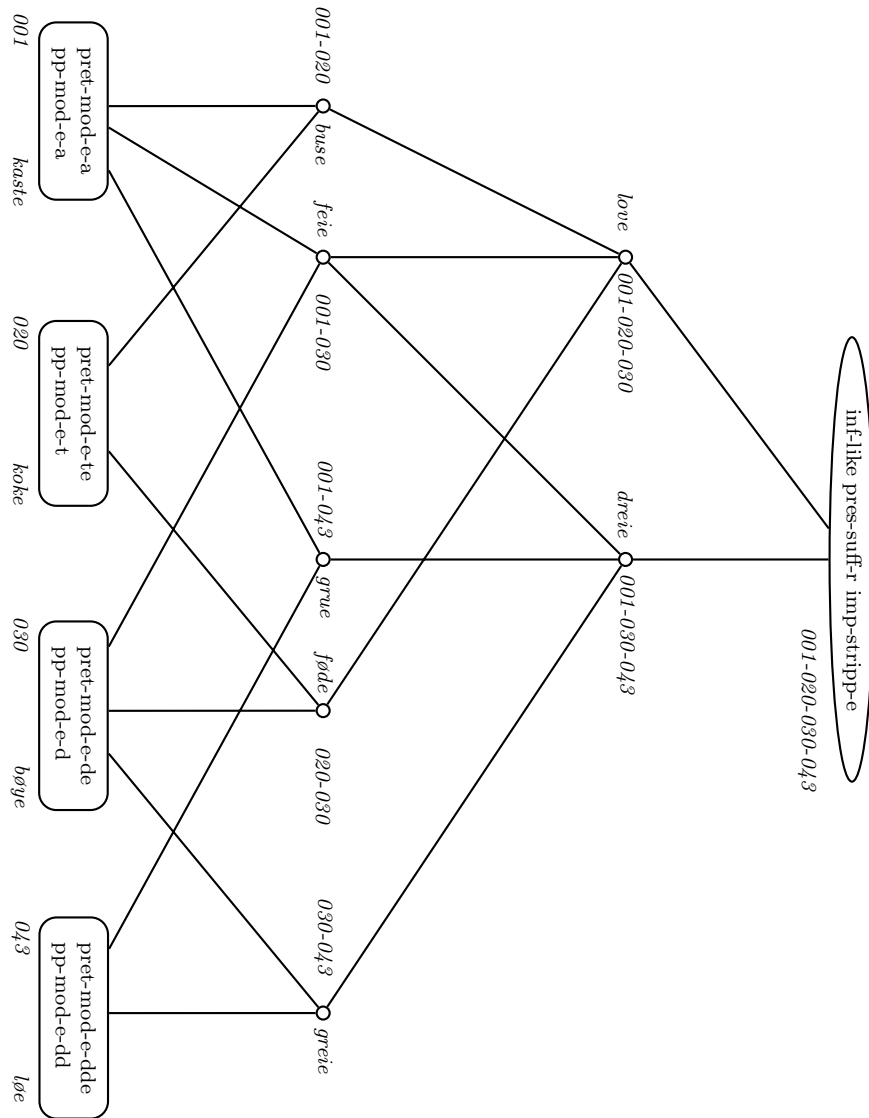
Så går vi laus på preteritumskolonnen, men her viser det seg at alle felta inneheld ulike formendingar. Følgjeleg kan vi ikkje flytte nokon av dei opp på eit høgare trenivå; dei vert knytte til bladnodar i treet, det vil seie til paradigmegrupper med berre ein paradigmatype. Tilsvarende gjer vi for perfektum partisipp-kolonnen. Det heile er gjort mekaniske og enkle operasjonar, og resultatet er å finne i figur 4.3.

---

<sup>25</sup>Lesaren kan merkje seg at det er umogleg for ein paradigmegruppe å verte samanbunden med ei anna gruppe *horisontalt* i treet, det vil seie på same høgdenivået (sjå seksjon 4.4). Grunnen er at dei ikkje er knytte til dei same paradigmatypane.

Merk også at for innflokke tre vil vi etterkvart få eit presentasjonsproblem. For når vi teiknar todimensjonale tre vil det verte mange liner mellom dei ulike nivåa i treet, og for å gjere trea meir lettlesne vert det så freistande å teikne paradigmegrupper med same kardinaltalet over og under kvarandre, for slik å hindre at ein får for mange kryssande liner.

Heldigvis kan vi la elektronhjernen byggje og halde reie på alle gruppene og linene mellom dei. Det har eg gjort, og eg har vore så heldig at eg har hatt program (sjå seksjon 6.1) som så kunne vise heile det ferdige treet på skjermen slik at ein kan følgje liner opp og ned i treet. Ein alvorleg restriksjon har allikevel vore at paradigmegruppene i treet vert vist fram utan den grammatiske informasjonen (formendingar, bøyingstypar og alle andre opplysingar vi måtte ønskje å fylle treet med!) gruppene er tilknytt, så programmet oppfyller slett ikkje alle ønskene morfologar og leksikografar har.



Figur 4.3: Forenkla paradigmatre over dei svake verba og deira formendingar

**Arv** For å lette forståinga av treet er det nyttig å tenkje seg at formendingar (og andre einingar vi måtte fylle treet med), høgt oppe i treet arvast av paradigmatypar lenger nede. Dette vil vi sjå svært tydeleg i neste tre. Det som ikkje arvast, er eigenskapen *å-vere-bøyingsgruppe*, og det gjev heller ikkje så god meining å la leksema arvast.

**Spekulasjonar** Vi merkjer oss at eit steg nedover i paradigmatreet inneber ei bortveljing av ei formending, så vi kan spekulere i om bøyingsgruppenodar i dei øvre skikta representerer identitetsforvirra leksem: Har dei enno ikkje funne ut av kva enkelgruppe dei helst vil høyre til? Di lenger ned vi går, ser vi i så fall at leksema vert tryggare på seg sjølv, og for *bøye* er det ikkje tvil. Det verbet er knytt til berre eitt normert fullformsett og er eit *030*-verb rakt inn i ryggmergen.

Men det er slett ikkje rett å seie at til dømes *001-020-030*-verba er heilt surrete, for dei har som dei andre verba frå dei andre fleirty pige bøyingsgruppene funne seg til rette med ein *r* som presenssuffiks og med imperativsformer like med verbstammane. Valfrie er preteritum- og perfektum partisipp-formendingane.

Spekulasjonane vert til sist misleiande, for dei synast også å innebere at vi må tvinne tommeltottar medan vi ventar på at verba frå dei fleirty pige bøyingsgruppene skal falle til ro i ein bladnode. Dette har vi sjølv sagt ingen garanti for, så det er betre å nøye seg med at per i dag er *feie* (normert som) eit *001-030*-verb.

**Slektskapstilhøve mellom bøyingsgrupper** Eg kjenner meg viss på at paradigmatre er nyttige for systematiske studiar i morfologiske likskapar og ulikskapar mellom verb, av di dei gjev eit så klart bilete av samanhengen mellom bøyingsgruppene og formendingane.

Vi ser frå treet at dersom *001-020-030-043* fanst som bøyingsgruppe, ville han omfatte berre svake verb, og for desse ville preteritum- og perfektum partisipp-formene vere valfrie. Men korfor finst ho ikkje som bøyingsgruppe? Dette er eit spørsmål eg ikkje kan svare fullt ut på, men det er moro å prøve seg:

Merk at vi i paradigmatreet må heilt opp til paradigmatregruppa *001-020-030-043* for å finne noko som er sams for bøyingsgruppene *020* og *043*.

Alle verb frå bøyingsgruppe *043* er tostavingsord som endar på trykklett sistestaving, *-e*. Nest siste bokstaven, fortel *Tabellmaterialet* oss, er alltid ein vokal. Det same gjeld, fortel *Tabellmaterialet* oss, for alle bøyingstypar som er over han i paradigmatreet: *breie*, *dreie*, *føye*, *kadreie*, *kleie*, *leie*, *meie*,

*pleie, reie, sneie, snøye* eller *spreie* kan ta 030- eller 043-endingar, og nokre av *-eie*-orda tek også 001-endingar. Alle desse orda fungerer dårleg med *pret-mod-e-te*-formendinga.

I motsetnad har vi dei reine 020-verba, 2577 stykke i *Tabellmaterialet*. Desse er alle fleirstavingsord som også endar på trykklett *e*, men nest siste bokstav er *alltid* ein konsonant. Igjen viser *Tabellmaterialet* at det same kan seiast for alle bøyingsgruppene over 020 i paradigmatreet.

Det er derfor kanskje ikkje så rart at bokmål ikkje har ord som vert bøygd etter både 020-paradigmet og 043-paradigmet. Og det finst då heller ingen ordboksbøyingsstype “v2 el. v4”.

Nokre ord kjem likevel forunderleg nære. For *sprede* og *føde* har nemleg bokmål også kortformene *spre* og *fø*<sup>26</sup> I *Bokmålsordboka* finst nett desse formene som leieformer, koda som v4-verb, side om side med sine meir bokstavrike søsken! *spre* og *sprede* er alternativoppslag for same tydinga, medan *fø* og *føde* er knytte til ulike tydingar. Kortformene til (alle to) 020-030-verba har bøyingsstypen *v4*!

Men i motsetnad til *v4*-verba vi så langt har hatt i treet vårt, høyrer dei ikkje til bøyingsgruppe 043. Dei høyrer til bøyingsgruppe 040, og dei utgjer om lag fire femdelar av *v4*-verba. Alle verb knytt til denne paradigmatypen er ord der rota berre inneheld ei staving: *blø, bre, flå, gry, kle* (og hertil høyrer også det forbodne *sne*). Desse tek berre *-dde* i preteritum. I tabell 4.7 set eg opp formendingane som gjeld for paradigmatypene 040 og 043. Etterpå skal vi utvide treet vårt.<sup>27</sup>

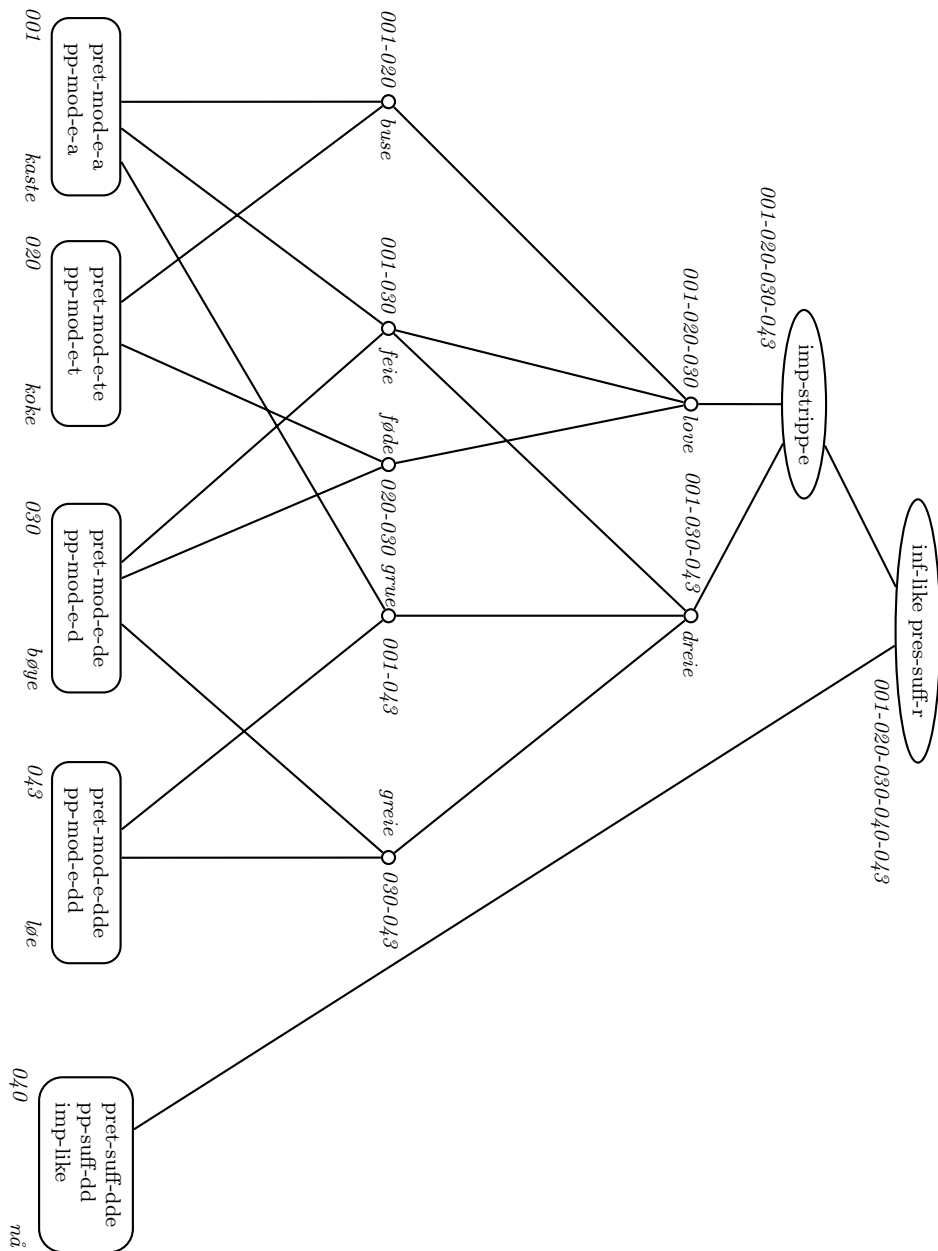
Det er så absolutt verdt å merkje seg at *alle* verba som er knytte til 001, 020, 030 eller 043 har infinitiv på (trykklett) *-e*. Mellom *v4*-verba går det såleis eit fundamentalt skilje mellom dei med einstavingsrot, *bre, flå, sy*, og dei andre, *bie, kvie*.

døme	inf	pres	pret	perf part	imp	
nå		like	suff-r	suff-dde	suff-dd	like 040
grue		like	suff-r	mod-e-dde	mod-e-dd	stripp-e 043

Tabell 4.7: Resten av formendingstabellen for dei svake verba

<sup>26</sup> *sprede, føde, love* og *utlove* er alle orda som er knytte til formendingane frå *paradigmegruppene* 020 og 030, og *love* og *utlove* er også knytt til formendingar for 001.

<sup>27</sup> Det vert eit penare tre om ein byttar om plassane til paradigmatypene 040 og 043, så det gjer eg. Merk at feilkodinga nok ikkje var så dum likevel!



Figur 4.4: Paradigmetre over dei svake verba

**Eit alternativt tre** Det kan synast uheldig at vi må dele opp verb som endar på *-dde* i to grupper. I 4.5 set eg derfor opp eit alternativt tre der eg har utvida representasjonen av formendingane.

*jafs* skal i modellen lesast som at den trykklette sistestavinga skal bort, om den finst. Eg har altså lagt til formendingar som tek bort endingane frå langformene, men let kortformene vere, før eit suffiks kan leggjast til om det er ønskjeleg.

No får vi at dei same fem formendingane gjeld for både *040-* og *043-*verba, og resultatet er ein modell der vi kunne slått saman dei to verbgruppene var det ikkje for dei fleirtypige bøyingsgruppene. Modellen er nærast *for* smart: Formendingane er blinde for skiljet mellom kortformer og langformer, og i modellen kjem fundamentale fonetiske kjensgjerningar for bokmålet dårleg fram. Til sitt bruk kan det kanskje vere nyttig.

I treet kan vi elles merkje oss først at gruppa *001-020-030-043* har vorte overflødig, så den er borte. Til gjengjeld får vi *040-043*, som heller ikkje er ei bøyingsgruppe. Denne gruppa fangar inn likskapen mellom verba som kan bøyast ut frå paradigme *040* og *043*.

Dernest merkjer vi at får problem om vi ikkje nyttar eit arveomgrep. Dersom vi tenkjer oss i ståstaden til *greie*, og ønskjer oss *043*-preteritumsbøyinga, må vi gå ned til bladnoden *043* og så opp att til *040-043* for å finne denne om vi ikkje let formendingane gå i arv.





## 4.8 Bokmålstreet

Det er fullt mogleg å inkludere all informasjonen i *Tabellmaterialet* i eit stort paradigmetre over bokmål, og eg trur eit slikt tre kan vere nyttig til så mangt. Først og fremst vil det kunne hjelpe oss å få eit meir presist bilete av bokmålet. Mellom anna vil det vere nyttig å sjå i praksis kva formendingar som delast mellom dei svake og dei sterke verba. I denne seksjonen vil eg spekulere litt i andre saker ein kan bruke paradigmegruppene til, og eg vonar det kan vere med å setje fantasien i sving.

Til no har vi sett paradigmegrupper for dei enkle paradigmatypene, bøyingsgruppene og for formendingane. I tillegg kan vi vriste ein del andre grupper ut av representasjonen i *Tabellmaterialet*. Mange av dei er overraskande nyttige for den som vil byggje eit HPSG-leksikon ut av *Tabellmaterialet*. Ved å la maskina finne paradigmatypar for alle kategoriane ein treng, får ein vite kva kategoriar som står i relasjonar til kvarandre, og ein får vite kor høgt dei skal plasserast i HPSG-typesystemet.<sup>28</sup>

I seksjonen om *Tabellmaterialet* viste eg eit utdrag frå *Paradigmelista*, og her har eg sett det opp att i tabell 4.8. Denne lista inneheld mykje av informasjonen vi er ute etter.<sup>29</sup> Alle paradigmegruppene vi kan finne herfrå, kan vi hente fram med maskinell prosessering: For kvar kategori som finst, samlar vi dei ulike paradigmatypene han er knytt til og lagar ei paradigmegruppe av det. Mellom anna finn vi då ei gruppe med alle verbleksemsbøyingsgruppene under seg. Denne typen har eit grueleg langt namn (*001-002-003-004-007-...-427-428-429*), men det treng vi ikkje engste oss for sidan datamaskina kan vise han fram til oss med eit meir meningsfullt namn.

Ein interessant ting som no hender er at kategorien *mask* viser seg å vere knytt både til substantiv, pronomen og adjektiv. Det vil seie, for substantiv finst det hankjønns-, hokjønns- og inkjekjønnsparadigme, så han er berre knytt til ei delmengd av substantivparadigma.<sup>30</sup> Vi finn altså paradigme-grupper som har under seg heile ordklassar. Ved første augekast verkar det kanskje litt skummelt, men nærare ettertanke syner at alt dei einskilde paradigma samlar potensielt motstridande informasjon: *020* samlar til dømes informasjon om dei semantisk svært så ulike verba *gjære* og *gjøne*.

<sup>28</sup>Eg skal komme tilbake til HPSG-typar i seksjon 6.

<sup>29</sup>Meir kan finnast i *Bøyingskategorilista*.

<sup>30</sup>*mask* er òg knytt til determinativ, og ettersom vi kan få adjektivavleiingar frå verb, er han også knytt til verba. I tillegg er han berre knytt til delmengder av pronomenparadigma, så her har eg gjort mitt ytste for å forenkle skildringa.

PARADIGME-ID	BØY-GRUPPE	ORDKLASSE
001	verb, normal	verb
020	verb, normal	verb
450	pronomen, normal	pron pers 1 ent hum
459	pronomen, normal	pron refl
47F	determinativ	det poss res
500	adjektiv	adj
515	adjektiv	adj <ordenstall>
691	preposisjon	prep
695	adverb	adv
696	infinitivsmärke	inf-merke
702	substantiv	subst mask appell
902	substantiv	subst fem appell
995	substantiv, ubøy	subst fem prop

Tabell 4.8: Utdrag frå *Paradigmelista*.

#### 4.8.1 Andre paradigmegrupper

Vi kan lage ein paradigmatype som har under seg alle leksema som *ikkje* får presens ved å leggje til ein *r*. Denne typen heiter *123-126-151-152-314-315-335-380-381-382-383-384-385-38A-38B-38C*, og han omfattar verbleksema *burde, gjøre kunne, måtte, si skulle, spørre, tore, ville, vite* og *være*.

Og vidare kan vi lage ein type for, sei dei leksikalske orda frå ordklassane adjektiv, verb, substantiv og adverb. Frå tabellen ovanfor kan vi jamvel blande ei subgruppe av dei leksikalske orda med andre og setje saman paradigmatypen *P691-P695-P696*. Dette er ein type som står for alle preposisjonane, alle adverbane og infinitivsmærket, og dei delar alle eigenskapen *det-å-ikkje-kunne-bøyast*. Om han kan brukast til noko, det er ei anna sak.

#### 4.8.2 Konservative og radikale paradigmegrupper

Dei som kjenner *Tabellmaterialet* godt, vil no kanskje ane sjansen til å finne ei gruppe som har under seg formendingane frå dei mest konservative bokmålsdialektene: Dette vil i så fall vere ei gruppe som omfattar dei fleste einskildparadigma, men utelet hokjønnsparadigma, samt ein del verbparadigma (*001*) med meir.<sup>31</sup>

<sup>31</sup>Eg uttrykkjer meg omtrentleg, for eg kjenner ikkje så godt til dei ulike dialektene og paradigmatypingane.

Ei slik gruppe kan nyttast til å avgrense talet på paradigmegrupper vi plukkar formendingar frå, slik at vi berre vel konservative formendingar til leksema: Vi plukkar berre formendingar frå greiner (frå topp til botn) som ender i konservative nodar.

Tenkjer vi paradigmegruppene som sett, kan vi uttrykkje det same slik: Vi nyttar formendingar berre frå paradigmegrupper som har dei konservative paradigmegruppene som subsett.

Verre er det å lage ei gruppe som berre inkluderer radikale paradigme, for her har uheldigvis *Tabellmaterialet* vidareført *Bokmålsordboka* sitt system med å inkludere tokjønnsystemet i trekjønnsystemet. Det som var plassparande i *Bokmålsordboka*, og plassparande i *Tabellmaterialet*, om enn ikkje på same viset, vert no ei stengsle for den lingvistiske fantasien: Sjølv om vi nyttar all informasjon i *Tabellmaterialet*, vil fortsatt tokjønnsystemets “fold” hamne i same falden som “fold” i trekjønnsystemet (sjå seksjon 2.5).

Lesaren undrar kanskje på korfor det er noko problem. Eg vil forklare det med å vise til det eine leksikonet eg har bygt (seksjon 6): Her hadde eg leksemet representert med ei leieform og ein bøyingsstype, og eg hadde ein slags formendingar som verka på leksema og gav fullformene.<sup>32</sup> Med ei paradigmegruppe for konservative formendingar, kunne eg ha skrive ein funksjon som under generering av fullformer til leieforma avgrensa valet mellom moglege formendingar. For leksemet med leieforma *budeie* og bøyingsstypen *702-902*, ville eg då kunna godkjent berre forma *budeien*. Men eg ville ikkje klart å godkjenne berre forma *budeia*, for formendinga *bunden-eintal-mask-suff-n* tilhøyrt paradigmegruppa *702-716-748-760-772-773*, og den måtte eg ha, for ho inneheldt hankjønnsformendingar for trekjønnsystemet.

For å bøte på dette kan *Norsk Ordbank* endre på kjønnsystemet sitt, slik at ein kan skilje mellom felleskjønnsord frå tokjønnsystemet og hankjønnsord frå trekjønnsystemet. Med nærare 53000 leksem tilknytt paradigmatypen *700* er dette enklare sagt enn gjort: Ein kan ikkje overlate arbeidet til datamaskina, for vi har alt fortalt ho at *en fold* er *en fold*, og ho har klisterhjerne.

Vil ein fortsatt late representasjonen av trekjønnsystemet innehalde både hankjønns- og hokjønnsformer for substantiv som *gruppe*, kan desse orda knytast til tre paradigme: Eit hokjønns-, eit hankjønns- og eit felleskjønnsparadigme. Og vil ein ha full kjønnsfridom (*budeien* i trekjønnsystemet), kan ein late systemet vere som det er.

Det eksisterande systemet er kanskje ikkje så ille. Dersom ein trekjønna

---

<sup>32</sup>Formendingane verka òg i revers, slik at eg kunne finne ut kva leksem ei fullform kunne høyre til.

tekst er morfologisk konsekvent (sjå seksjon 4.5), og han inneheld *budeien*, ligg det gjerne ein tanke bak det.

Men likevel: Eit femkjønnsystem ville tillate ein rikare representasjon av bokmålet, så det ville nok vore det aller beste.

**Morfologisk inkonsekvente dialektgrupper** Bokmålstreet er i utgangspunktet koda etter “forvirringa” *det kollektive bokmålet* er i. Det let seg gjerne å utforske denne nærare via statistiske korpusanalyser og jamføring med paradigmatreet. Kanskje kan ein finne vanlege stilistiske eller morfologiske inkonsekvensar?

Dersom ein i tillegg og for eksempel utvidar representasjonen med den høgst unormerte formendinga *hankjønn-bu-ein-mod-e-a*,<sup>33</sup> vil ein også kunne representere tokjønnsbokmålsfolkets veksling mellom *hytten* og *hytta*.

**Ei utfordring** Oppslaget til *min* er i ordboka formulert slik: “**min** eienomspronomen *mi* el. *min, mitt, mine*”, og dette vert stadfesta av *Norsk referansegrammatikk*. I *Tabellmaterialet* vert pronomenet koda med ei eintypig bøyingsgruppe (for possessivt determinativ) korfrå det kan utleiest ei hankjønnsform, (berre) ei hokjønnsform, ei inkjekjønnsform samt ei fleirtalsform. Som for artikkelen *en*, har såleis ikkje kodarane av *Tabellmaterialet* fått med seg at *min* også er eit hokjønnsord.

Eg har merka meg at det finst dei som kjem nær å hevde at referansegrammatikkens analyse er feil:

Man kunne naturlegvis påstå at noen av oss opererer med to sett av hankjønnspronomen, hvorav det ene faller sammen med hankjønnssettet. En slik hypotese kan imidlertid lett avvises. For at den skulle tas alvorlig, måtte vi være i stand til å si *jakka min* så vel som *min jakke* [...] Lånet av visse språklige former som opprinnelig forutsatte et grammatisk system med tre kjønn, førte ikke til at tokjønn-systemet gikk i oppløsning. [...] Til tross for at det historisk sett utvilsomt må karakteriseres som en sammenblanding av to normer, representerer det i dag en egen norm.

[Fretheim, 1985, s. 100-101]

Eg trur ei korpusanalyse vil kunne hjelpe oss til ei betre forståing av dette og nærskyldte fenomen, men ettersom her er grammatikk involvert, duger det ikkje berre å sjå på einskildorda.

<sup>33</sup>Eller kanskje *felleskjønn-bu-ein-mod-e-a*.

### 4.8.3 Kor mange paradigmegrupper finst?

Eg vil tru ein kjem langt i bokmålsrepresentasjonen med mellom 1200 og 1300 paradigmegrupper. Men for moro skuld set eg opp ein generell formel for kor mange grupper vi kan finne til eit vilkårleg tal paradigme. Den er gjeve ved likning 4.1.

$$s = 2^n - 1 \quad (4.1)$$

Tek vi maksimalt i og puttar inn alle dei 644 paradigma frå *Paradigme-lista*, finn vi alle moglege typenamn,  $2^{644} - 1$  stykke i talet.<sup>34</sup> Mengda av alle desse paradigmegruppene inneheld reiskapar for å uttrykkje eit ufatteleg stort tal generaliseringar over bokmålet, men, som nemnt, ikkje alle.

### 4.8.4 Paradigmegrupper for både bokmål og nynorsk

Dersom vi ikkje let oss avskrekke av det store talet i førre avsnittet, kan vi også sjå korleis vi kan la bokmål og nynorsk møtast i same modellen, om enn dei ikkje vil nærme seg i praksis: Legg til ein *B* (for *bokmål*) i bokmålsparadigma og ein *N* (for *nynorsk*) til nynorskparadigma frå *Norsk Ordbank*, og bygg så paradigmatreet på nytt.<sup>35</sup>

Ofte er definisjonsartiklane i *Nynorskordboka* og *Bokmålsordboka* ganske så ulike, men til omsetjingsformål kunne målføra kanskje nytte same semantikken? Det meste er jo elles likt når det gjeld dei fundamentale grammatiske kategoriane, og dei deler også dei fleste grammatiske reglane. I så fall vil ein med rett val av paradigmegruppe lettvis kunne velje om ein vil omsetje til nynorsk eller til bokmål.

---

<sup>34</sup>Dette er ein bortimot  $3,65 \times 10^{194}$  stykke, utan at det fortel noko særleg meir. Til samanlikning kan nemnast at estimatet for kor mange atom det finst i universet ligg mellom  $10^{78}$  til  $10^{81}$ . Sjå til dømes <http://pages.prodigy.net/jhonig/bignum/qauniver.html>. Kan det verkeleg vere så mange? Ja, men det er svært vanskeleg å førestille seg det. Her er mitt haltande forsøk: Tenk deg at du har ei lang rad med konsollar med ein av-på brytar for kvart paradigme. Vi har ein konsoll der berre ein brytar er på, vi har ein annan konsoll der berre ein annan brytar er på, vi har ein tredje der begge desse brytarane er på, men ingen andre. Og så bortetter. I alt er det 644 brytarar som kan vere anten på eller av, og trekkjer vi ifrå den tomme konsollen, den konsollen der alle brytarane er av, står vi att med dei nemnte tala.

<sup>35</sup>Og medan ein er i gang: Kva er eigentleg skilnaden på *den* norske *presens-suff-r*-formendinga og den danske og den svenske? Med eit enormt paradigmat vil vi kunne systematisk undersøkje morfologiske skilnadar og likskapar mellom alle dei skandinaviske språka.



## Kapittel 5

# *Norsk referansegrammatikk*

*Pirk*  
*må til. Utan*  
*stoggar ein, stend.*  
*Det kan vere eit bos*  
*i klokka, det må burt*  
*skal ho koma på løs*  
*att, ho tikkar og gjeng.*  
*Sume må tergast, til bust*  
*veks på gamalt skinn.*

Olav H. Hauge, *Pirk* frå samlinga *Dropar i austavind*

Denne seksjonen inneheld først og fremst tre grunnleggjande kritikkar av *Norsk referansegrammatikk*. Undervegs vil eg både referere til mine egne analyser, og eg vil samanlikne dei med referansegrammatikken sine. Det er fint å få denne samanlikninga, men eg vil understreke at det ikkje er nødvendig for kritikane: Dei kan reformulerast uavhengig av mitt arbeid.

Først kritiserer eg *Norsk referansegrammatikk* for handsaminga, eller snarare mangelen på handsaminga, han gjev det eg har kalla fleirtypige bøyingsgrupper. Generelt vil eg hevde at ein ikkje kan forklare morfologien til eit språk med mange variantformer utan (noko som liknar) fleirtypige bøyingsgrupper.

Dernest tek eg for meg, og forkastar, referansegrammatikens morfologi-modell. Etter mi meining eignar ikkje infinitivsstammen seg for å skildre bokmålet, ettersom han gjer einskilde ord fonetisk og skriftbiletmessig likare enn dei i røynda er.

Etter det viser eg korleis referansegrammatikken mislukkast i å klassifisere verba, og at forfattarane i tillegg ikkje klarar drage dei rette konklusjonane

av sin eigen sin eigen leksemmodell i gruppeoppdelinga.

Eg vil understreke at sjølv om infinitivstammeanalysen heng nøye saman med verbklassifiseringa, er kritikane av dei same innhaldsmessig uavhengige. Referansegrammatikkens klassifisering kan ein få ved å utestenge fleirtypige bøyingsgrupper<sup>1</sup> og elles nytte for eksempel modellen eg skisserte sist i seksjon 4.7.

## 5.1 Første kritikk: Fleirtypige bøyingsgrupper

Referansegrammatikken delar bokmålsverba i to klassar, ein for dei svake og ein for dei sterke. Den svake er igjen delt i to klassar. *Klasse 1* omfattar verb som i preteritum tek endingane *-a* eller *-et*. Den andre klassen er igjen delt i tre *grupper*, *2a*, *2b* og *2c*, og desse er kvar for seg meint å romme verb som i fortid får endingar på *-te*, *-de* og *-dde*.<sup>2</sup>

Verb som kan ta meir enn ei preteritumsending vert medlem i fleire klassar/grupper, utan at referansegrammatikken ser grunn til å forklare dette høgst i augnefallande faktum. Tek vi utgangspunkt i verbet *feie*, som i min analyse tilhøyrrer bøyingsgruppa *001-030* (sjå seksjon 4.7) kan vi sjå korleis referansegrammatikken handsamar det. Sistesetninga frå skildringa av klasse *1* går slik:

Legg spesielt merke til følgende verb: *herme* (som bøyes alternativt etter 1. klasse eller klasse 2b), *vokse* som bøyes etter klasse 2a, og verb på *-ls* som kan bøyes etter 1. klasse (*kulse*), men som oftere går etter klasse 2a (*hilse/helse, frelse*).

[*Norsk referansegrammatikk*, s. 483]

Vi merkjer oss at *feie* ikkje høver i skildringa,<sup>3</sup> men vi kan finne verbet nemnt eksplisitt i skildringa av gruppe *2b*:

Verbene *skade, tygge, love, dreie, feie* kan også bøyes etter 1. klasse [...]

[*Norsk referansegrammatikk*, s. 485]

<sup>1</sup>Ved å utelate fleirtypige bøyingsgrupper får vi ikkje lenger ein språkmodell, men noko som liknar meir ein dialektmodell.

<sup>2</sup>I *Tabellmaterialet* finst om lag 4300 *1*-verb og 3350 *2a*-verb, nesten 300 *2b*-verb og over 210 *2c*-verb. Hugs at ein del er samansetjingar, som ikkje finst i ordboka. I tillegg finst også nokre unormerte ord, som sjølv sagt heller ikkje er å finne i ordboka. Eit eksempel er *juge*, og det er det einaste *2b*-unntaket.

<sup>3</sup>Skildringa nemner forøvrig heller ikkje at det finst *1*-verb som kan bøyast etter klasse *2c*. Som døme kan vi ta fram verbet *grue*. Om dette verbet står det til gjengjeld eksplisitt i skildringa av *2c* at det også kan bøyast etter klasse *1*.



Slik spring verb frå fleirty pige bøyingsgrupper opp som brysame sær-syn; dei står gjerne i parentesar, og gjerne til sist i skildringa av dei ulike gruppene og klassane. Sjølv teiknsetjinga ber bod om kor vanskeleg det er å presse desse verba inn i referansegrammatikkens kategoriar: Skildringa av gruppe 2a vert avslutta med ein sluttparentes (forutan det obligatoriske punktumet), utan at det finst ein tilsvarande byrjingsparentes.

For lesaren vert effekten av særsyna kvar gong den same, skissa han hadde byrja teikne seg i hovudet vert full av rot. Dersom han ikkje gjev opp, vert han i beste fall sitjande att med ei grovskisse samt ei mengd tilsynelatande vilkårlege unntak. Konklusjonen vert følgeleg at *Norsk referansegrammatikk* langt frå *forklarar* bokmåls morfologien.

Ettersom referansegrammatikken heller ikkje kan svare på følgjande spørsmål, vil eg òg påstå at han heller ikkje *forstår* morfologien: Kva klasse, eller gruppe, høyrer verbet *feie eigentleg* til? Klasse 1? Gruppe 2b? Er dei unntak?, kva klasse eller gruppe er dei i så fall unntak til? Kva med dei andre 74382 orda, i alt litt over halvparten av orda i *Tabellmaterialet*, som har alternativformer?

## 5.2 Andre kritikk: Infinitivsstammen

“Former dannet til infinitivsstammen” er tittelen på seksjon 7.1.3.1 (sidene 475-478) i referansegrammatikken. For bokmål gjeld dette infinitiv, presens og imperativ (s. 474). Her følgjer ein kort gjennomgang.

**Infinitiv** Følgjande sitat introduserer lesaren for infinitivsstammen:

I de fleste verb ender infinitiv på -e:

kast-e, lys-e, lev-e, vær-e, finn-e

Verb med stamme på enkel vokal får ingen endelse i infinitiv:

le, spe, gro, ro, bu, tru, få, stå

Unntak er enkelte verb der stammen ender på -i eller -u:

*bie, tie, vie, grue, kue, skrue* (også *skru*), *true*.

Vi merkjer oss at forfattarane har dei opplysingane dei treng for å kunne slutte at for å få infinitiven *skrue* frå infinitivsstammen *skru*, og infinitiven *bu* frå infinitivsstammen *bu*, treng ein ulike formendingar. Her finn vi att skiljet mellom verba som er knytte til paradigmegruppene 040 og 043.<sup>4</sup>

<sup>4</sup>I følge *Bokmålsordboka* er forøvrig ikkje *tru* eit bokmålsord.

Vidare følgjer ein gjennomgang av verb som til liks med *skrue/skru* har både kortformsinfinitiv og langformsinfinitiv (*late/la, stride/stri* med mange fleire).

**Imperativ** Etter infinitiv går turen til imperativ, kanskje av di referansegrammatikken ikkje utleier nynorskpresensen<sup>5</sup> frå infinitivsstammen, og her får infinitivsstammen vist kva han er god for: “Imperativ er vanligvis lik infinitivsstammen”. Så enkelt er det . . . nesten. Forutan dei (stort sett valfrie) unntaka som referansegrammatikken nemner, der imperativen er lik infinitiven, finst ortografiske unntak som *stem* til *stemme* og så frametter.<sup>6</sup>

(Kanskje er det på grunn av denne analysen, og no tenkjer eg ikkje på arbeidsuhellet, at *Bokmålsordboka* ikkje nemner imperativa med eit ord, ikkje ein gong unntaka?)

### Presens

Presens blir i bokmål dannet av infinitivsstammen + *-er*. Ender stammen på enkel vokal, dannes presens ved bare å legge til endelsen *-r*. Det fører til at presens i bokmål som hovedregel kan sies å være lik infinitiv + *-r*

[*Norsk referansegrammatikk*, s. 478]

Formuleringa “kan sies” viser at referansegrammatikken ser dei to første setningane som den grunnleggjande skildringa av kva som hender. Sistesetninga vert meir ein hendig hugseregulering for dei som ikkje fulgte med i timen om infinitivsstammen. Vi merkjer oss at hugseregelen er så god at han vert verdiga omtalen *hovudregel*.

<sup>5</sup>Og eg klarar ikkje dy meg: Nynorsk har ein ekstra svakverbsklasse, og svært mange sterke verb, der imperativen er lik presensen. Der bokmål har det avmålte og avklara *synger*, har nynorsk det kontante og tvetydige(!) *syng*. I hovudet på lesaren av nynorske tekstar vert det spent ein bogestreg mellom presenslesinga av verbet og av imperativlesinga, og strengen *dirrar* heilt til hjernen klarar avgjere kva som er den grammatisk korrekte. Følgjeleg kan ein få subtile nynorskleseopplevingar om ein varmar opp med å lese seg gjennom ein del majoritetsspråksprosa for å “avlære” hjernen om dette faktum: Då varar dirringa lenger.

Når ein så har fått røska opp i kategoriane sine, høver det kanskje med vidare lesing av det morfologisk meir tempererte bokmålet. . .

<sup>6</sup>Eg må sei eg vart paff når eg las imperativsanalysen for minst fjerde gongen, og det gjekk opp for meg at ortografiske unntak ikkje vart nemnte her heller. Det tyder på at eg, og ikkje berre eg, tenkjer forsvinnande lite på imperativen og dessutan har ortografien internalisert ganske så bra. Kanskje er det imperativsforma som er leieforma for verba i mitt mentale leksikon? Med ortografien rett koda?

**Infinitivsstammen som leieform** Den som vil implementere bokmålsleksikonet, finn i sum at det er enkelt å utleie dei tre formene frå infinitivsstammen. Spørsmålet eg tumla med under min implementasjon var om det var bryet verdt å grave fram infinitivsstammen frå infinitiva frå *Tabellmaterialet*. Dersom vi for augneblinken held oss til desse tre formene kan vi setje opp og samanlikne formendringstabellane 5.1 og 5.2.<sup>7</sup> I første kolonnen har eg sett opp døme på leieform, dei andre kolonnene lagar eg formendringar for. Skilnaden mellom *ro* og *roe* i første tabellen er om leieforma representerer verbet med infinitiven *ro* eller verbet med infinitiven *roe*. Skiljet mellom dei tilsvarande bøyingsgruppene i *Tabellmaterialet* vil gå på om verbet har kortformsinfinitiv eller langformsinfinitiv.

leieformsdøme	inf	pres	imp
ro	like	suff-r	like
ro	suff-e	suff-er	like

Tabell 5.1: Infinitivsstammen som leieform

leieformsdøme	inf	pres	imp
ro	like	suff-r	like
roe	like	suff-r	stripp-e

Tabell 5.2: Infinitiven som leieform

Av tabellane framgår det at vi får fram meir likskap innad i kolonnene dersom vi *ikkje* går omvegen om infinitivsstammen: Med infinitiven som leieform klarar vi oss følgeleg med ein *inf-like* og ein *pres-suff-r*, men vi må ha to formendringar for imperativsforma. Referansegrammatikken må derimot ha fem formendringar for å gjere same arbeidet, og gjer følgeleg seg sjølv (og lesaren) ei bjørneteneste ved å presentere bokmålet som han gjer.

Med innføringa av infinitivsstammen, vert også heile analysen mindre realistisk. Dersom infinitivsstammen er del av det mentale leksikonet, må hjernen så og seie analysere seg fram til stammen *ro* for kvar gong han møter infinitiven *ro*. Dersom hjernen nyttar stammen når han skal uttrykkje seg med infinitivskonstruksjonar, må han bruke ei slags formendring, *inf-like*, for

<sup>7</sup>Eg ser bort frå ortografiske unntak og imperativsunntaka referansegrammatikken bokfører.

å hente fram kortformsinfinitiven. Når han vil ha fram langformsinfinitiven frå den andre stammen *ro*, må han bruke ei slags *inf-suff-e*-formending.

Bokmålshjernen gjer klokt i å utnytte den grunnleggjande lydlege ulikskapen mellom langformene og kortformene i *sin* representasjon av bokmålet.<sup>8</sup> Korfor held så referansegrammatikken seg til ein forvanskande, ikkje-realistisk morfologisk modell som ikkje harmonerer med fonologien?

### 5.3 Tredje kritikk: Gruppeinndeling

Vi såg i førre seksjon at grammatikken handsamar alle verba under eitt i infinitivsanalysen, gruppeinndelinga vert først gjort etterpå. Forfattarane tek her eit kunstgrep som synast å innebære at gruppeinndelinga ikkje vedkjem infinitivs-, presens- og imperativsanalysen. Dersom dette grepet skal vere noko verdt, måtte det vere fordi det gjer dei nemnte analysane betre på eitt eller anna vis. Den einaste vinsten eg finn er at analysane (seksjon 7.1.3) kan gjerast før enn, og uavhengig av, preteritums- og perfektum partisipp-analysen (seksjon 7.1.4).<sup>9</sup>

Det vil seie, heilt uavhengig er den ikkje, for forfattarane refererer til bøyingsklassane i imperativsanalysen (s. 477-478). Eg skulle for øvrig gjerne ha sett det gjort fleire gongar. Vile ikkje infinitivsanalysen vore sterkare om ein der fekk vite kva klassar dei tilhøyrrer, dei verba som ikkje får infinitivs-ending?<sup>10</sup>

I tillegg har det liten effekt først å handsame verba som del av same usorterte masse, for gruppeinndelinga har sjølvstekt tilbakeverkande kraft i og med at alle verb med fortidsform (det har dei alle) også har ei presensform. Dersom inndelinga vart gjort først kunne ein heller seie at infinitivsanalysen og så frametter er lik for alle klassane. Då kunne ein òg etter skjønn referere til klassane for unntaka, og boka ville vore betre som oppslagsbok.

Når forfattarane endeleg delar opp verba i grupper, gjer dei det ut frå kva endingar *infinitivsstammen* tek for å gje preteritum, og slett ikkje, som dei seier, ut frå måten “verbene danner preteritum på”. Dei klassifiserar

<sup>8</sup>Lesaren kan merkje seg at *føde* ikkje berre kan vere eit langformsverb, men også eit langformssubstantiv. *ro* kan vere eit kortformsverb og eit kortformssubstantiv. Følgjeleg talar vi om eit heilt grunnleggjande skilje som gjeld både substantiv og verb.

<sup>9</sup>7.1.3 har tittelen “Oversyn over bøyning”. Analysane her vert unnagjort før siste-seksjonen 7.1.3.2 “Inndeling i bøyingsklasser”. Deretter kjem 7.1.4 med den uforstålege tittelen “Bøyingsmønster”, som om ikkje infinitiv, presens og imperativ var del av noko bøyingsmønster.

<sup>10</sup>Det gjeld berre verb frå svakklasse *2c* (040-verb, ikkje 043-verb), men det gjeld verb frå mest alle sterkklassane.

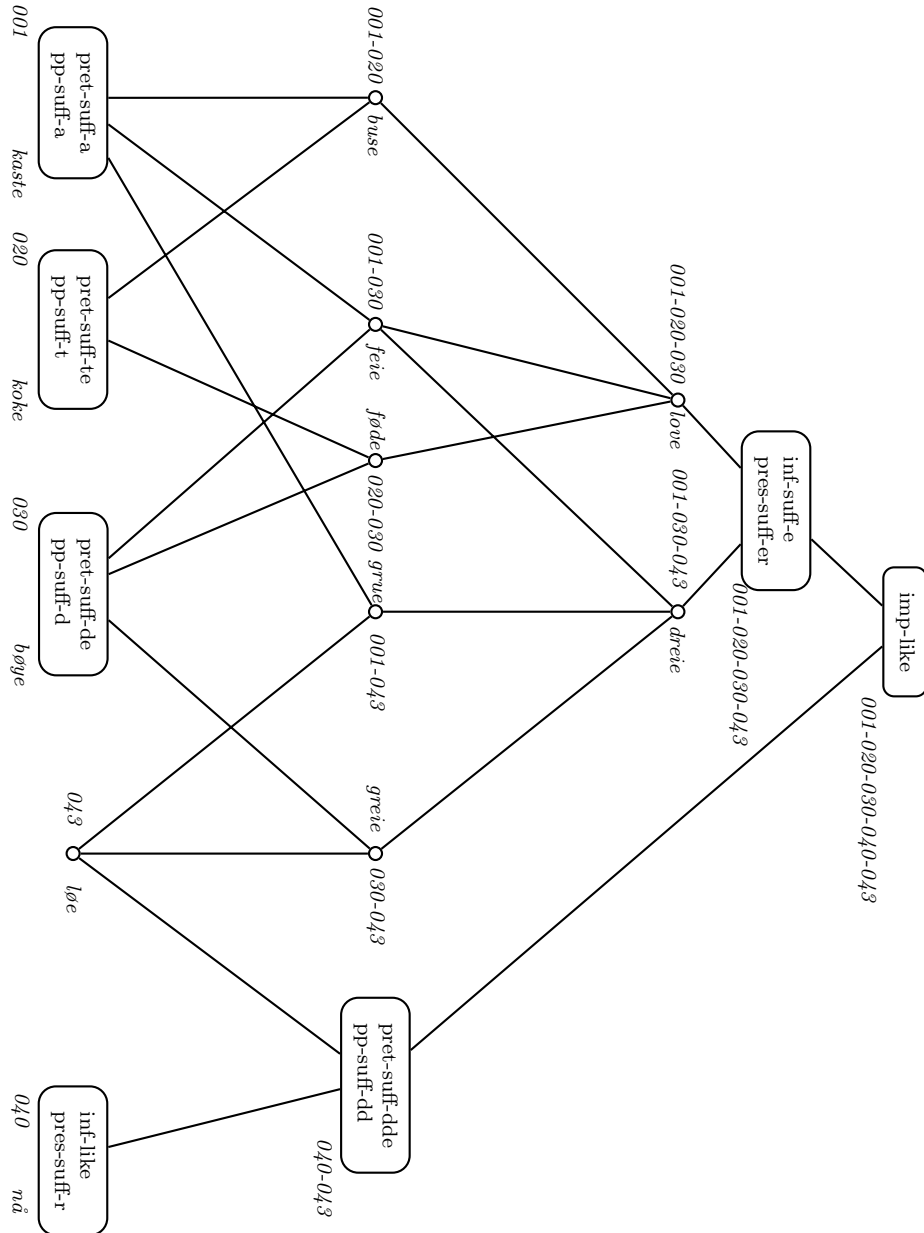
verba etter abstrakte og ikkje naturlege eigenskapar. Enden på visa vert at *Norsk referansegrammatikk* ikkje klassifiserar verba i det heile, men derimot *dei fire preteritumsendingane* som infinitivstammene kan ta.

**Referansegrammatikkens modell** Forfattarane går aktivt inn for å blinde sin modell for fundamentale fonetiske kjensgjerningar for bokmål (sjå førre seksjon og 4.7). Men dei klarar det ikkje.

I figur 5.1 har eg sett opp det treet<sup>11</sup> referansegrammatikken ville fått om dei hadde hatt eit omgrep om fleirtypige bøyingsgrupper. Vi ser at det er umogleg å slå i lag gruppene *040* og *043*, for dei delar ikkje alle formendingane.

---

<sup>11</sup>Sjå seksjon 4.4 for ei nærare skildring av denne strukturen.



Figur 5.1: Referansegrammatikkens paradigmatre over dei svake verba

## 5.4 Ein tankekross

Verbbøyingskodane frå *Bokmålsordboka* finn vi att i ei meir kompakt form på side 481 i referansegrammatikken. Tabell 5.3 viser desse. For meg ser det ut til at tabellmakarane ikkje heilt veit kva dei gjer her, så eg skal forklare lina for andre klassen i detalj: Gruppeinndelinga ser vi att i dei to siste felta i andre lina, dei er skilde med komma.

I infinitivsfeltet finn vi att infinitivsanalysen, og bak kommaet står det som i nokre morfologiske teoriar vert kalla ei *null-ending* [Matthews, 1991, s. 123] (noko referansegrammatikken ikkje opplyser lesaren om). Denne svarar til mi formending *inf-like*. Kommaet i første feltet svarer altså til skiljet som berre finst internt i gruppe *2c*, og følgjeleg må kommaet lesast både som dette skiljet, og skiljet mellom *2c* og *2a/2b*. Dette medfører sjølvstøtt rot i lesarens forståing av tabellen, og det vert ikkje noko betre av presensfeltet.

For meg såg parentesbruken først ut som dobbelføring av både hugse-regelen (sjå 5.2) og den tilgrunnleggande presensregelen, men ettersom den siste krev eit komma (som infinitivsfeltet), må lesaren for å gje mening til feltet forstå parentesen på same tvesynte viset han forstod kommaet i første feltet.

	inf	pres	pret	perf part
1. klasse	-e	-er	-et/-a	-et/-a
2. klasse	-e, -0	-(e)r	-te, -de, -dde	-t, -d, -dd

Tabell 5.3: Verbbøyingskodar i *Norsk referansegrammatikk*

Vi hugsar frå seksjonen om *Bokmålsordboka* at ho hadde eit uavklara tilhøve til sitt eige morfologigrunnlag. Følg med på kva som hender dersom vi stiller ovanståande tabell opp mot svakverbskodane frå *Bokmålsordboka*. Dei kan nemleg lesast som ei fortetta utgåve av dei to siste kritikkane: Vi finn infinitiven som leieform, og vi finn fem liner som svarar til fem paradigmatypar, og grunnen til det, er at ordboka er tvungen til å representere alle formene i eit godt heilskapleg grep.

Vi har no hamna i den heilt særeigne posisjonen kor vi ser at for å forstå *Bokmålsordboka* må vi forstå *Norsk referansegrammatikk*, og at vi for å forstå denne igjen må forstå spekulative morfologiske modellar utvikla av engelsktalande morfologar. Desse modellane høver bokmålet dårleg, og derfor ser vi endeleg at *Bokmålsordboka* av lagnaden er dømt til å utføre eit tabulatorisk fadermord, slik at ho kan leve i nærleik til, og kjærleik til, bokmålsorda som *Norsk referansegrammatikk* sviktar.

	<b>inf</b>	<b>pres</b>	<b>pret</b>	<b>perf part</b>
v1	kaste	kaster	kasta el. kastet	kasta el. kastet
v2	lyse	lyser	lyste	lyst
v3	leve	lever	levde	levd
v4	nå	når	nådde	nådd
	bie	bier	bidde	bidd

Tabell 5.4: Verbbøyingskodar i *Bokmålsordboka*



## Kapittel 6

# Bokmål i HPSG

Eg har implementert i alt tre leksikonversjonar av *Tabellmaterialet* i formalismen *Hovuddriven frasestrukturgrammatikk*, HPSG (sjå Sag and Wasow [1999]). Felles for dei alle er HPSG-typene som representerer grammatiske kategoriar (ordklassar, kjønn o.s.f.) for bokmål, og bøyingsstypene. Dei skil lag først og fremst i talet på formendingar som vert nytta.

For den første versjonen, den *enkle*, slo eg ikkje i lag formendingane, så eg endte opp med over 3600 bøyingsreglar, deriblant 266 *pres-suff-r*-formendingar. For den andre, den *tradisjonelle* versjonen, slo eg i lag alle like formendingar, slik at eg sat att med berre ei *pres-suff-r*-formending. For den siste, den *kompakte* versjonen, delte eg formendingane opp i to delar; *pres-suff-r* vart til ei *pret*-formending, og ein bokstavrekkjeoperasjon *suff-r*. Bokstavrekkjeoperasjonen kunne såleis fungere for andre enn verba.

I det følgjande skal eg først seie litt om verktya eg har brukt til å representere bokmål, og så litt generelt om leksikonbyggjing. Deretter fortel eg om HPSG-typene eg nytta,<sup>1</sup> og eg skal forsøke å klargjere typeteoritenkjinga undervegs. Fokuset fell etterkvart på den andre leksikonversjonen, den *tradisjonelle*, ettersom denne liknar mest det lesaren vil finne i skildringar av bokmåls morfologien. Når det er gjort, vil eg forklare korleis ein kan byggje det tredje leksikonet. Til sist vil eg vise med praktiske døme korleis paradigmatypen fungerer slik at bøyingsreglar kan gje fullformer til leksema dei skal verke på.

---

<sup>1</sup>Eg fortel meir enn nok til at lesaren kan skjønne korleis eg har gått fram, men eg fortel slett ikkje alt. Eg har gjort mange val undervegs (nokre av dei er heilt sikkert uheldige), og alle dei relevante typene som eg ikkje her kjem inn på, kan finnast att i *bokmål.tdl* og *ord-og-leksemtypar.tdl*. Sjå 6.5.3.

## 6.1 Rammene

Eg har arbeidd med eit dataprogram som heiter *the Linguistic Knowledge Building system (LKB)*.<sup>2,3</sup> Programmet har ikkje dei best tenkjelege føresetnadane for å representere bokmåls morfologien: Mekanismane for å bruke regulære uttrykk er alt for svake og omstendelige til å vere hjelpsame i skildringa av til dømes vokalvekslinga for dei sterke verba.<sup>4,5</sup> Til gjengjeld utmerkar programmet seg med suverene verkty for å samle all leksikalsk informasjon (grammatisk som semantisk), og i tillegg syntaktisk informasjon,<sup>6</sup> for heile språk i ein stor datastruktur. Kransekaketoppen er reiskapane ein får til å inspisere dei leksikalske og syntaktiske strukturane med.

Som sagt er det HPSG eg har nytta for å skildre bokmål, men det er ikkje *LKB* som krev dette. Eg har teke utgangspunkt i ein meir generell

---

<sup>2</sup>Sjå Copestake [2002] og <http://lingo.stanford.edu>.

<sup>3</sup>*LKB* er, for å vere eit så stort program, utstyrt med ein sjeldant tilbaketent *open kjeldekodelisens*: Dersom ein brukar sel eller gjev vidare sine endringar på eller tillegg til programmet, er det nok at lisensen gjeld for “a substantial portion of the Software”. Kva som er ein substansiell del, eller kven som avgjer det, står det ingenting om.

<sup>4</sup>Den teknisk interesserte lesaren kan merkje seg at referansegrammatikken mange stadar nyttar skildringar som kan representerast med regulære uttrykk. Sjå til dømes skildringa av klasse *1*-verba, der det vert lista opp om lag 70 relevante konsonantgrupper.

<sup>5</sup>Eg vil tru det er fleire djuptliggjande årsaker bak denne mangelen. Den eine ligg, vil eg tippe, i den engelske morfologien, og måten den skil seg frå den norske. Her tenkjer eg ikkje berre på meir eller mindre regulære vokalvekslingar, men også at tilhøvet mellom skriftbilete og uttale for dei engelske orda er meir innfløkt enn norsk på sitt enklaste (*genre* på engelsk og norsk, kontra den norske genistreken *sjanger*, det engelske substantivet *rope* kontra det norske verbet *rope*). Som ein direkte konsekvens vert det meir interessant å utforske regulære uttrykk som eit mogleg bindeledd mellom skrift og lyd for norsk enn for engelsk.

Dernest kan vi kanskje også finne ein grunn i at hovudmotivasjonen for notidas datalingvistiske arbeid er å forstå syntaks og få færrest mogleg semantiske lesingar av setningar. Følgjeleg er ikkje morfologi like spennande.

Til sist finst også brutale tekniske restriksjonar. *LKB* er implementert i programmeringsspråket Common Lisp. Den internasjonalt aksepterte standarden for dette språket (ANSI/X3.226, ANSI står for American National Standards Institute) legg ikkje føringar for lispimplementasjonen av regulære uttrykk. Dette har leidd til at ulike implementasjonar, deriblant fleire store kommersielle, har valt ulike løysingar med ulike brukargrensesnitt. I tillegg finst plattformuavhengige implementasjonar med andre grensesnitt. Blant desse finn vi ein som *LKB* nyttar (og som eg sjølv nyttar), så her kan *LKB* verte betre. Likevel: *LKB* vert parallellutvikla på fleire plattformer samstundes, nokre kommersielle, andre ikkje, og det er ein relativt stor jobb å få alt til å harmonere. Ettersom det alt finst fleire svært store grammatikkar som er implementert i *LKB*, kan ein ikkje ta det for gjeve at programutviklarane finn det interessant å forbetre desse mekanismane.

<sup>6</sup>livreferanse

HPSG-struktur med det tidsspeglande namnet *Matrix*.<sup>7</sup> Denne strukturen er utleidd av arbeid med store HPSG-grammatikkar for mellom anna tysk, engelsk og japansk!

HPSG set oss i stand til å strukturere all informasjonen i *Tabellmaterialet*. Vi lagar mellom anna ein HPSG-type som samlar og representerer informasjonen om eit ords kjønn, person, tal og om ordet er ei bunden form eller ikkje. Den same typen inkluderer vi mellom anna i både substantiv og adjektiv, og vi lagar

Tidlegare i oppgåva (seksjon 4.8) har vi handsama alle dei grammatiske kategoriane som om dei var meir eller mindre like. HPSG gjev oss mekanismar til å teikne grove skiljer mellom ordklassar og finare skiljer mellom kjønn, mekanismar til å samle all informasjonen i ein stor struktur, og mekanismar til å trekke ut nett dei naturlege strukturane vi ønskjer å fokusere på, til dømes ord.

## 6.2 Leksikonbyggjeplan

Eg har spreidd informasjonen frå *Tabellmaterialet* over filene som er gjevne i tabell 6.1. Alle filene er koda i *Type Definition Language* (Copestake [2002, s. 99-107]), og filnamna er utstyrte med det konvensjonelle suffikset “.tdl”. Ein “h” i andre kolonnen fortel at eg måtte kode fila for hand, ein “d” seier at elektronhjernen kunne gjere arbeidet.

Det er svært mange måtar å byggje heile HPSG-strukturen på. Strategien eg etter mykje om og enno meire men snubla meg fram til, tek utgangspunkt i dei HPSG-typene som trengs for å byggje sjølve leksikonoppføringane og formendingane.

Alle leksikonoppføringane er utstyrte med ein bøyingsstype, og til kvar av bøyingsstypene laga eg ein HPSG-type som eg la i *leksem-og-ordtypar.tdl*. Bøyelege ord fekk leksemoppføringar og ubøyelege ord fekk ordoppføringar. Oppføringane fortalde vidare kva ordklasse dei tilhøyrande orda var av (og ofte enno meir). Til dette arbeidet var ei maskinell prosessering av *Paradigmelista* svært nyttig.

For formendingane var først og fremst informasjonen frå *Bøyingskategorilista* viktig. Det var om å gjere å kunne lage formendingane enkelt, for

---

<sup>7</sup>Sjå <http://lingo.stanford.edu> ein gong til. Nettsida eksisterte i alle fall medio desember 2003. For *Matrix* er koden tilgjengeleg på nettet, men brukslisensen er per i dag ikkje ein open kjeldekodelisens. Eg har forhøyr meg med ein av *LKB*-utviklarane, som fortel at *Matrix* er meint å vere fritt, men at det tydelegvis i skrivande stund ikkje vert reflektert i lisensen. Når lesaren les dette kan det alt vere endra på.

filnamn	koding	filinnhald
bokmål.tdl	h	Informasjonen frå <i>Paradigmelista</i> og <i>Bøyingskategorilista</i>
leksem-og-ordtypar.tdl	h	leksem- og ordtypar
btypar-leksikontypar	h	hjelpetil som knyt saman ty- par frå dei to førre listene med bøyingstypene i leksikonet
leksikon.tdl	d	leksemoppføringane
bformer.tdl	d	formendringar med bøying
ubformer.tdl	d	formendringar utan bøying
ptypar.tdl	d	paradigmatreet

Tabell 6.1: Liste over filene som utgjør leksikonet

di færre spesifiseringar eg trong for å få fullformer frå leieforma, di enklare ville koden som byggjer formendringane verte. Følgjeleg var det hendig med HPSG-typar som direkte svarer til “nøyt ubunden ein” og dei andre oppføringane i den nemnte lista. Desse typene la eg i *bokmål.tdl*. Somme tider la eg likevel til litt fleire opplysingar. For eksempel får ein adjektivavleingingar frå verbleksema, men *Bøyingskategorilista* fortalde ikkje noko om kva gradbøying adjektivformene skulle ha, så det la eg til.

Først når ein har byrja få seg eit bilete av kva typar som trengs til formendringane og leksikonoppføringane, er det på tide å fylle ut med typar som spesifiserer dei meir fundamentale strukturane. Då ser ein betre på kva måte (meir eller mindre grunnleggjande) typar er innvevde i kvarandre, og ein kan betre avgjere om ein skal bruke *trekk* eller *typar* eller både deler for å lage dei. (Meir om trekk seinare.) Til sist er det om å gjere å hekte alt saman fast i *Matrix*-typene.

I praksis gjorde eg dette litt framover og litt bakover, knekte ei nøtt her, og løyste eit problem der.

## 6.3 HPSG-typar for bokmål

### 6.3.1 Paradigmehierarkiet

Tidlegare i oppgåva har vi skilt mellom paradigme grupper og paradigmatypar (sjå seksjon 4.3), og vi har gjort med dei nett det vi har lysta. Det kan vi fortsatt gjere i HPSG, men i *LKB* støyter vi på ein fundamental restriksjon: Programmet skil ut bokstavrekkeoperasjonane som kodar morfologien i ein

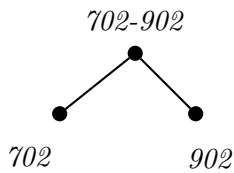
eigen del. Måten dette er implementert på, gjer det meningslaust å tale om arv av formendingar frå ei paradigmegruppe til ei anna.

I staden vert det lagt opp til at *bøyingsreglar* så og seie tek leksem som argument og gjev ei fullform som resultat. Her kan vi nytte paradigmegruppene som bindeledd mellom reglane og leksema, og lettast kan det gjere om vi lagar *trekkklause* HPSG-typar<sup>8</sup> av paradigmegruppene.

Vi skal kalle desse typane for *paradigmegruppetypar*.<sup>9</sup> No kan vi seie at ein bøyingsregel *har* ein slik og slik paradigmegruppetype, og at eit leksem eller eit ord *har* ein slik-og-slik paradigmegruppetype. For dei siste gjev det også meining å bruke termen *bøyingsgruppetype*. Med litt meir avslappa språkbruk, kan vi også seie at datastrukturane *er* av ein slik og slik *paradigmegruppetype*, og der samanhengen er krystallklar, skal vi snike oss til å tale om *paradigmatypar* når vi eigentleg meiner *paradigmegruppetypar*.

I HPSG-hierarkiet<sup>10</sup> legg vi derfor det vi kan kalle *paradigmegruppetypar*. Til det tradisjonelle leksikonet brukte eg om lag 1200 paradigmegruppetypar, for formendingane, bøyingsgruppene og dei enkle paradigmegruppene. Og ein til: Gruppen som samlar under seg all informasjonen i *Tabellmaterialet*, med namnet *001-...-999*. Denne har eg omdøyppt til det meir hendige *PTYPE*, og sett han øvst i den delen av HPSG-hierarkiet som inneheld paradigmegruppetypene.

**Eit døme** Figur 6.1 inneheld eit utsnitt frå botnen av *Paradigmehierarkiet*.



Figur 6.1: Utsnitt av paradigmehierarkiet

I typehierarkiparlør: Som far til to søner<sup>11</sup> kan vi no seie at supertypen

<sup>8</sup>Når vi kjem til typar med trekk, skal eg forklare det nærare.

<sup>9</sup> Dette omstendelige namnet får fram at omgrepa paradigmegrupper, -typar og -gruppetypar er tett knytte til kvarandre, utan at vi risikerer å blande dei saman.

<sup>10</sup> Ein rota og retta asyklisk graf. Sjå seksjon 4.4 for forklaring.

<sup>11</sup> Ein HPSG-type kan ha ein eller fleire fedre, og han kan ha ein eller fleire søner. Dette resulterer i eit nokså uvanleg slektstre, så eg talar med andre ord om ein biologisk umogleg

er av typen *702-eller-902*.<sup>12</sup> Vi seier at supertypen er underspesifisert. Dette er om lag som å seie at ein person kan vere anten mann eller kvinne, men at vi ikkje kjenner personens kjønn.

Paradigma *702* og *902* arvar eigenskapane faren har. Faren har ingen eigenskapar i seg sjølv, så i praksis tyder ikkje det så mykje. Men farskapet er viktig nok i seg sjølv, for *702* er i kraft av å vere son til sin far også av typen *702-902*, han er berre ikkje underspesifisert, tvert imot: Han er maksimalt spesifisert som både person og mann, ettersom han er nedst i hierarkiet.

Når vi kjem til sønene, ser vi at dei ikkje har nokon søner, og særleg har dei ingen søner i fellesskap. Dermed er det ingen typar som arvar eigenskapane frå både *702* og *902*, og det er ingen typar av typen *702-og-902*: Det finst ingen personar som er både mann og kvinne.

Dermed har eg forklart det sentrale skiljet mellom *disjunktive* typar (*702-eller-902*) og *konjunktive* typar (*702-og-902*).

**Koding** Slik ser eit par av oppføringa i fila *ptypar.tdl* ut:<sup>13</sup>

P695 := ptype.

P702 := P702-P902 & P702-P812 & P702-P773 & P702-P716.

Dette skal lesast som at *P695* er ein undertype til *ptype*, og at *P702* har fire fedre, deriblant *P702-P902*.

**Notabene** *LKB* krev ikkje berre at konjuntive typar skal vere resultatet av ein unifikasjon (sjå nedanfor) av fedrene (pluss kanskje vidare spesifisering). *LKB* krev vidare at fedrene må ha det vi kan kalle ein *minste sams subtype*.<sup>14</sup>

Lispkoden som byggjer parhierarkiet gjer det annleis, for det byggjer ein RRAG med færrest mogleg nodar. Når så *LKB* puttast RRAGen inn

---

tilstand. Motivasjonen for å bruke fedre og søner i staden for mødre og døtre, ligg i at eg dermed kan referere til ein sonetype med *han*.

<sup>12</sup>Merk at dette er noko vi *seier* og *tenkjer*. Typen heiter *702-902*.

<sup>13</sup>“P”-en i typenamna er der av di den eine bøyingsgruppetypen vart heitande *76L-953*. (Hugs at paradigmenummera er i 26-talssystemet! Sjå seksjon 3.2) *LKB* gjer på eitt eller anna vis lispsymbol ut av HPSG-typenamna, også dei numeriske, men det klarar ikkje gjere den nemnte paradigmegruppetypen sitt namn til eit symbol, lispenn oppfattar det som eit tal av typen *double-float*, 76 opphøgd i minus 953! For å omgå dette la eg ein “P” fremst i namna på paradigmetypene.

<sup>14</sup>Copestake [2002, s. 38-43] har ein fin diskusjon av motivasjonen for dette.

i sin RRAG, skuvar han inn mange ekstra nodar (sjå Copestake [2002, s. 111]). For det tradisjonelle paradigmehierarkiet la *LKB* til om lag 250 ekstra nodar.<sup>15</sup> Dette gjer sjølvsagt at oppteikninga av paradigmehierarkiet vert enno mindre oversiktleg enn det det i utgangspunktet var.

### 6.3.2 Typar for grammatiske kategoriar

I *Matrix*-hierarkiet finst ein type som heiter *sort*, (engelsk for *slag*). Under denne legg eg heile paradigmehierarkiet, men eg legg òg ein del bokmålskategoriar under her.

**Typa trekkstruktur** Mellom anna legg eg kjønna her. Kjønna og ein del annan informasjon er det vanleg å samle i ein eigen HPSG-struktur:

<i>nor-png</i>	PERS	pers,
	TAL	tal,
	KJØNN	kjønn,
	BUND	bund

Figur 6.2: Trekkstruktur for kjønn med meir

Denne HPSG-strukturen heiter *nor-png*, og han arvar typen *png* (Person, Number, Gender) frå *Matrix*. Til venstre (PERS osf.) i figuren finn vi *trekka* strukturen inneheld, og til høgre finn vi kva *verdiar* trekka kan ha. Eit trekk kan ha anten ein atomisk type eller ein trekkstruktur som verdi, men i dette dømet er det berre atomiske verditypar. Per konvensjon skrivast trekk med store bokstavar og verdiar med små.

*nor-png* er altså ein type. Til skilnad frå typane i førre delseksjon er han ikkje trekklaus. Her er det verdt å merkje seg at det for ein type kan vere spesifisert trekk sjølv om definisjonen (eller skildringa) av typen ikkje inneheld referansar til (eller skildring av) trekk. Det kan for eksempel vere ein konjunktiv type som arvar trekka frå sine fedre, men som ikkje er vidare spesifisert.

Kanskje er det også verdt å merkje seg termen *atomisk type*. Copestake [2002, s. 68] seier “I will refer to types which have no appropriate features and which have no descendants with appropriate features itself as *atomic*

<sup>15</sup>Det er ikkje umogleg at ein kan finne semiinteressante lingvistiske tolkingar til fleire av desse nodane, men eg vil tru det er vanskeleg. Det dreiar seg stort sett om minste sams subtypar av paradigmegruppetypar for formendingar som i BokmålsRRAGen vil ha same subgraf under seg.

types.” *p*type er ein slik trekklaus type som berre har trekkause subtypar. Kva denne termen er nyttig for veit eg ikkje, for Copestake brukar den ikkje til noko.

**Koding** Slik ser filoppføringa for *nor-png* ut:

```
nor-png := png &
  [ PERS person,
    TAL tal,
    KJØNN kjønn,
    BUND bund ].
```

Vi ser at *nor-png* arvar typen *png* og at han i tillegg spesifiserer fire verdjar for fire trekk.<sup>16</sup>

Når ein type med trekk arvar ein annan type med trekk, kan arvingen leggje til nye trekk til fartypens struktur, eller han kan vidare spesifisere eitt eller fleire trekk. For *nor-png* la eg til fire trekk til ein tom trekkstruktur, *png*, som igjen arva ein annan tom trekkstruktur (utan å leggje til noko informasjon), den generiske trekkstrukturtypen *avm* (Attribute Value Matrix). Typen *png* er såleis informasjonsmessig overflødig, men for bokhaldarens regnskapsføring er det nyttig å kunne hindre at einskilde typar får uoversiktleg mange subtypar.

**Kjønn** Vi ser i strukturen i figur 6.2 at trekket *KJØNN* må ha verdi av typen *kjønn* (eller ein subtype av *kjønn*). Det er rimeleg å tenkje seg at med tre kjønn i *Tabellmaterialets* trekjønnssystem får vi tre undertypar av *kjønn*. Men sidan vi skal ha leksem i leksikon korfrå vi kan få ord av både to og tre kjønn, er det betre å gjere det på ein annan måte.

Vi vil til dømes hindre at inkjekjønnsbøyingar kan fungere for *mask-fem-leksema*. Det kan vi gjere ved å lage ein (disjunktiv) supertype *mask-fem*, som arvar *kjønn*, og som både *mask* og *fem* arvar frå. I tillegg spesifiserer vi sjølvstøtt at *mask-fem-leksema* sine kjønn skal vere av typen *mask-fem*.

Frå førre delseksjon hugsar vi at subtypar kan seiast å vere av same type som supertypane (dei er berre vidare spesifiserte). Når vi så seier at hankjønnsbøyingar skal fungere for leksem med kjønn *mask*, så vil dei også fungere for leksem med kjønn av ein supertype av *mask*.<sup>17</sup>

<sup>16</sup>I *Matrix* står nokre kommentarar til *PNG* i seksjonen “Basic semantic types”. Søk etter bokstavrekkja “ref-ind”.

<sup>17</sup>Dei vil sjølvstøtt ikkje fungere for *nøyt-leksema* og *nøyt-fem-leksema*.



**Unifikasjon** Her er det viktig å få med seg at den resulterande fullforma får *mask* som kjønn, ikkje *mask-fem*. Mekanismen som her er i sving er *unifikasjon*. Det vil eg kort forklare slik: Dersom to typar ikkje inneheld motstridande informasjon kan vi unifikere dei til ein type.<sup>18</sup> Dersom nokre verditypar er meir spesifiserte for den eine typen enn for den andre, får den resulterande typen dei mest spesifikke verditypane. (Dersom den *andre* typen i tillegg inneheld nokre meir spesifikke verditypar enn den første, får den resulterande typen også desse meir spesifikke verditypane.)

### 6.3.3 Leksemtypar

*Matrix* omfattar ikkje ein leksemtipe, så eg laga ein type *leksem* og la han under *word-or-lexrule*. *leksem* utstyrte eg med ein disjunktiv type, *lex-or-phrase-synsem*. Dersom ei fullform til eit leksem viser seg å verte hovudet i ein frase, vil det ende opp med ein *phrase-synsem*, viss ikkje skal det ha ein *lex-synsem*.<sup>19</sup>

Frå verba, unnateke dei passive og dei ubøyelege, får vi adjektivavleiingar i tillegg til vanlege verb. Eg laga derfor ein (disjunktiv) leksemtipe som heitte *adj-verb-leksem*. Denne definerte eg slik:

`adj-verb-leksem := leksem & [ SYNSEM.LOCAL.CAT.HEAD adj-verb ]`.

Vi ser at *adj-verb-leksem* arvar typen *leksem*. Trekket *SYNSEM* er ein struktur som samlar all syntaktisk og semantisk informasjon for både leksemtypar og ordtypar (og frasetypar og setningstypar om vi hadde hatt slike med).

*SYNSEM* sitt trekk *LOCAL* er igjen ein trekkstruktur. Her finn vi trekket *CAT* (kategori), som også er ein trekkstruktur. Her finn vi trekket *HEAD*, og via dette trekket er det vanleg å identifisere kva ordklasse leksemet/ordet tilhøyrrer.<sup>20</sup>

<sup>18</sup>Eit døme på motstridande informasjon er typane *nøyt* og *mask-fem*. Desse høyrer til kvar si undergrein av *kjønn*, og dei har ingen felles subtype, så ein kan ikkje unifikere desse typane.

<sup>19</sup>Lesarar som ser nærare på hierarkiet eg har bygd, vil finne at eg ikkje gjer så mykje ut av dette som eg kunne ha gjort. Eg har stort sett nøydd meg med å få på plass kategoriane som finst i *Tabellmaterialet*, så her står det att arbeid til dei som er meir semantisk kyndige enn eg er.

<sup>20</sup>Lesaren undrar kanskje på kva *LOCAL* står for. Det kontrasterande trekket i *SYNSEM* er *NON-LOCAL*. Dette må ein til syntaksen og til koordineringa av fleire strukturar i ein for å forklare, noko som ligg bortanfor denne oppgåva.

Legg merke til kor enkelt det er å fokusere på berre utsnitt av store strukturar!: Vi seier at *adj-verb-leksem* sin *SYNSEM* sin *LOCAL* sin *CAT* sin *HEAD* skal vere av typen *adj-verb*. Når ein har vent seg til denne tenkjinga, og ein er van med å spesifisere ordklasse slik, seier ein at *adj-verb-leksem* sin *HEAD* er av typen *adj-verb*.

Som lesaren sikkert alt har skjønt er *adj-verb* nok ein disjunktiv super-type, og både *adj* og *verb* arvar denne.

**Adjektiv** Slik ser adjektivleksemdefinisjonen ut:

```
adj-leksem := adj-verb-leksem & agr-leksem &
  [ SYNSEM.LOCAL.CAT.HEAD adj ].
```

*adj-leksem* arvar typane *adj-verb-leksem* og *agr-leksem*. Den siste er ein disjunktiv leksemtypen for substantiv-, pronomen-, determinativ- og adjektivleksem. Tilhøyrande fullformer inngår i samsvarsbøygde nomenfraser, så *png-leksem* har eg definert som eit leksem med ein *nor-png*.

**Typetvinging** Her er definisjonen av *agr-leksem*:

```
agr-leksem := leksem & [ SYNSEM.LOCAL.AGR ref-ind ].
```

Trekket *AGR* (Agreement) vert i *Matrix* spesifisert for typen *local-min* (ein verdi som trekket *LOCAL* kan ta). Men når trekket introduserast, får det der ein supertype av *ref-ind*. I *agr-leksem* tvinger vi (ved arv/unifikasjon) verditypen til *AGR* nedover i ei grein som leier til ein høveleg semantisk struktur for *agr-leksema* (dvs. *agr-leksem* og subtypane).

**Verb** Slik ser verbleksemdefinisjonen ut:

```
verb-leksem := adj-verb-leksem &
  [ SYNSEM.LOCAL [ CAT.HEAD verb,
                    AGR event ] ].
```

Verbleksem skal ikkje ha *nor-png*. Eg hindrar dei i å få det ved å type-tvinge deira *AGR* til å vere av eit anna slag enn *agr-leksema* sin.

Dei passive verba som finst i *Tabellmaterialet* har eg enkelt og greitt kalla som *verb-leksem*.

### 6.3.4 Ordtypar

*Matrix* inneheld typar som gjer det mest naturleg å ikkje lage leksem av dei ubøyelege orda (preposisjonane til dømes). Typen *non-affix-bearing* høver som supertype for desse orda.

For dei bøyelege orda tok eg utgangspunkt i typen *lexeme-to-word-rule*.

Det er hendig med ein sams supertype for bøyelege og ubøyelege ord. Under *Matrix*-typen *word* la eg typane *bøyelege-ord* (som også arva *lexeme-to-word-rule*) og *ubøyelege-ord* (som også arva *non-affix-bearing*).<sup>21</sup> I tillegg la eg supertypar, til dømes *substantiv-ord*, som samlar under seg både bøyelege og ubøyelege substantiv, under typen *word*.

Under *substantiv-ord* la eg typane *bøyelege-substantiv* (som også arva *bøyelege-ord*) og *ubøyelege-substantiv* (som også arva *ubøyelege-ord*).

Under *bøyelege-ord* la eg typane *ltf-ord* og *ltfub-ord* (for Leieform-Til-ord og LTF-Utan-Bokstavrekkjeoperasjon-ord).

Under *bøyelege-substantiv* la eg *subst-ltf-ord* (som også arva *ltf-ord*) og *subst-ltfub-ord* (som også arva *ltfub-ord*). Desse typane er supertypar til dei maskinelt genererte bøyingsreglane.

Under *ubøyelege-substantiv* la eg supertypen *proprium-ord*, som under seg igjen fekk typar som kan finnast att i leksikonoppføringane.

**Frå leksem til ord** I *Tabellmaterialet* finst sju bøyelege possessive determinativparadigme for *deres*, *hans*, *hennes*, *hverandres* (og nokre til). Sjølve skriftbiletet til desse orda vert ikkje endra ved hankjønns- og hokjønnsbøyning osf. Desse leksema har eg derfor gjort om til (ubøyelege) ord, og spesifisert at dei skal ha *nor-png*. Her ser vi i praksis kor nyttig *underspesifisering* kan vere, og korleis det kan uttrykkje det som verkar heilt naturleg: At kjønnet for desse orda ikkje ligg i orda sjølve. Kjønn får dei når dei opptrer i nomenfraser, og det vert opp til den som implementerer grammatikken at dette gjerast rett.

**Adverbleksem** For både substantiva og pronomena finst nokre ord som vert bøyge og andre ikkje. I *Tabellmaterialet* finst vidare bøyingsgruppa “adjektiv, adverbtype”. Til denne høyrer orda *fort*, *lenge*, *nødig*, *ofte*, *svint* og *titt*. Desse kan gradbøyast som adjektiv, men liknar likevel meir på adverb, så eg laga *adv-leksem* av dei.

*bøyelege-substantiv*.

---

<sup>21</sup>Takk til Liv Ellingsen for ei oppklarande samtale om korfor dette er ein god idé. Ein generell diskusjon av emnet kan finnast hjå Copestake [2002, s. 38-43] (som eg alt har vist til ein gong før).

*ubøyelege-substantiv* arva både *non-affix-bearing* og *substantiv-ord*. Under *ubøyelege-substantiv* igjen, hamna eigennamna og diverse spesielle ubøyelege appellativ som *folkens* og *vare* m.fl.

Dei siste orddøma er vanskelege,<sup>22</sup> for dei vert vel først og fremst brukt som del av andre uttrykk, men på ulike vis. Her skulle det vere eitt og anna å gripe fat i for semantikarar.

I *Tabellmaterialet* finst også ubøyelege verb (*fordundre*, *nåde*, *vorde*, *vredes*, *bære*), men desse utelot eg frå leksikona. Dei inngår vel først og fremst i ståande uttrykk (som ein måtta hatt særlege grammatiske reglar for), så eg synest ikkje leksikonet vart mykje dårlegare ved utelatinga.

### 6.3.5 Maskingenererte tradisjonelle reglar

Eg nemnte ovanfor typane *ltf-ord* og *ltfub-ord* (Leieform Til Fullform med og Utan Bokstavrekkjeoperasjon). Desse er supertypar til alle ordtypane for dei bøyelege orda.

**ltfub-ord** For eit verb som *gå* er både infinitiv- og imperativforma lik leieforma, så i *ubformer.tdl* hamna blant anna denne definisjonen:

```
imp_rule := verb-ltfub-ord &
  [ SYNSEM.LOCAL.CAT.HEAD imp,
    PTYPE P003-P007-P013-P016-...-P426-P427-P428-P429 ] .
```

Dette er ein type som høver *LKBs* parsermekanismer.<sup>23</sup> Under parsing av einskildord sjekkar *LKB* om det for desse finst bokstavleg like leieformer i leksikon. Dersom i tillegg *HEAD* og *PTYPE* kan unifiserast for regelen og leksikonoppføringa, har vi ein vinnar.<sup>24</sup>

---

<sup>22</sup>Særleg gjeld dette *vare*. *Bokmålsordbok*gjev følgjande bruksdøme: *ta vare på*, *ta seg i vare*.

<sup>23</sup>Og genereringsmekanismer. Ein kan lett leggje til ein simpel semantikk til leksikonoppføringane ved å utvide perlfunksjonen som byggjer *leksikon.tdl*. Frå semantikken kan ein så (med høvelege typar) generere fullformer frå leieformer.

<sup>24</sup>Eg er ikkje sikker på rekkjefølgda og korleis dette vert gjort i praksis, så skildringa bør ikkje takast heilt bokstavleg.

**ltf-ord** Her er reglane som sørgjer for at ei *kake* kan verte til fleire:

```
fem-ubunden-fl-subst-suff-r_irule :=
%suffix (* r)
ub-fl-subst-ord &
  [ SYNSEM.LOCAL.AGR.PNG ubunden-fl-png & [ KJØNN fem ],
    PTYPE P902 ].

mask-bunden-ein-suff-n_irule :=
%suffix (* n)
subst-ltf-ord &
  [ SYNSEM.LOCAL.AGR.PNG bunden-ein-png & [ KJØNN mask ],
    PTYPE P702-P716-P748-P760-P772-P773 ].
```

Dersom eit ord endar på bokstaven *r*, vil *LKBs* parsar ta bort denne, og sjekke om det *r*-lause ordet finst som leieform for ein leksemtipe i leksikon med høveleg *HEAD*, *PNG* og *PTYPE*.<sup>25</sup>

**Morfologisk konsekvens** Etter mi meining burde vi hatt fleire kaker. *Tabellmaterialet* gjev oss “kaker” i ubunden form fleirtal hankjønn og ubunden form fleirtal hokjønn, men det ville vore fint med felleskjønnskaker i tillegg. (Sjå seksjon 4.8.2.) Då ville ein under parsing av tokjønnsbokmålstekstar kunne unifisere ein konservativ paradigmegruppetype med bøyingsregelens paradigmegruppetype, og slik finne høveleg kjønn for kakene.

Ein radikal paradigmegruppetype vil kunne hindre parsing av felleskjønnsformene, men dersom berre fleirtalsformene vert brukt, klarar vi ikkje finne ut kva kjønn dei har. Kjønnnet er berre synleg i eintalsformene, så sjølv om vi sørgjer for morfologisk konsekvens<sup>26</sup> vil vi få at kakene kan vere anten hokjønnskaker eller hankjønnskaker.

Generering av tekstar utgjer ikkje det same problemet, for då bør ein avgjere på førehand kva kjønn dei ulike substantiva har.

<sup>25</sup>Ein gjer ikkje noko gale om ein foresteller seg rekkjefølgda slik, men igjen vil eg seie at eg ikkje er viss på korleis dette er implementert.

<sup>26</sup>Dette kan gjerast på minst to vis: 1) Med ein tekstomgripande struktur som sørgjer for å unifisere alle paradigmegruppetypene for kvart leksem som finst i teksten. 2) Ved (f.eks.) ein lispfunksjon som gjer om lag det same, men med utgangspunkt i settoperasjonane skissert i seksjon 4.5.

## 6.4 Kompakt leksikon

For det kompakte leksikonet brukte eg ikkje subtypane *verb-ltf-ord*, *subst-ltf-ord* med fleire til *ltf-ord*.

Fila *kompakte-ubformer* vart veldig enkel å skrive. Reglane såg slik ut:

```
suff-r_irule :=
%suffix (* r)
ltf-ord &
  [ DTR suff-r-leksem ].
```

Trekket *DTR* fortel kva som er “argumentet” til bøyingsregelen.

*suff-r-leksem* la eg som ein subtype under *kompakt-leksem*, og eg maskin-genererte (med heller innfløkt kode) nye leksemtypar under *suff-r-leksem*, slik at all grammatisk informasjon for fullformene så var å finne for subtypar av *kompakt-leksem*.

Leksemtypane arva også frå leksemtypane frå *leksem-og-ordtypar.tdl*. Lesarar som har vore borti *LKB* vil no skjønne at det fullstendige HPSG-hierarkiet vart temmeleg komplisert.

## 6.5 Til sist

### 6.5.1 Sjølvskryt

Sidan eg ikkje kan levere med ein cdrom med oppgåva som inneheld heile leksikonet med alle formendingane, vil eg be lesaren tru meg når eg seier at det tradisjonelle leksikonet inneheldt over 152415 leksem og ubøyelege ord.<sup>27</sup> Eg fekk i alt 27 bøyingsreglar utan bokstavrekkjeoperasjonar, men eg kunne klart meg med færre dersom eg hadde gjort meir ut av pronomenregelgenereringa i lispkoden.

Vidare fekk eg 919 bøyingsreglar med bokstavrekkjeoperasjonar. Det er færre enn dei 999 formendingane eg tala om i seksjon 4.2, og hovudgrunnen er at eg har utelete ein del paradigme for heller brysame ord.

### 6.5.2 Men noko står att

**Leksem utan leieform** Det finst 25 paradigme i *Tabellmaterialet* der leieforma ikkje er tillaten. Det eine paradigmet kodar det refleksive pronomenet (*seg*), så det gjorde eg om til eit ubøyeleg ord, og spesifiserte at typen *refl-pronomen-ord* hadde *akk* (akkusativ) som verdi til trekket *KASUS*.

<sup>27</sup>13 millionar 633 tusen og 299 teikn finst i *leksikon.tdl*.

Dei andre har eg lete stå att, men den vedlagte lispkoden min skriv ei fil som heiter *problemreglar*, som inneheld informasjon om kva paradigme det gjeld og kva formendingar ein treng.<sup>28</sup>

Mellom anna gjeld det fleirtalssubstantiv som *Bokmålsordboka* seier det ikkje finst eintalsformer for, som “klær”. Dette kan løysast på fleire vis: Ein kan lage egne reglar for hand og leggje dei til den maskingenererte fila, eller i ei eiga fil. Betre er det likevel å skilje ut nokre nye bøyingsgrupper og utvide *Bøyingskategorilista* (seksjon 3.4). Då kan, tenkjer eg spontant, den vedlagte lispkoden nyttast utan endringar for å generere nye reglar. Kanskje må ein file litt meir på regelgenereringskoden.

Men eg spør meg: Kva er gale med *klede* som leieform til klær? Den prinsipielt beste løysinga er å utstyre også desse leksema med eit fullstendig formsett.<sup>29</sup>

**Presens partisipp** Lispkoden som eg har skrive, gjer berre adjektiv utav presens partisipp-fullformene som ein får frå verba. Vidare spesifisering ville vore fint.

**Paradigmesamanslåing** Det er mogleg å få færre bøyingsgrupper dersom ein slår i lag formendingane frå fleire paradigme. Det er kanskje ønskjeleg å slå i lag paradigme som kodar konsonantforenklingar og -fordoblingar for ord som elles er regulære (hankjønnsordet *søm* til *dømes*), med sine fullstendig regulære motpartar. Bortfall av *j* (verba *følgje*, *telgje*) høver også kanskje best saman med dei heilt regulære.

Interessant ville det også vore å undersøkje korleis ein kan slå saman formendingar for omlyd osf. Eg har ikkje sett noko vidare på dette.

I tillegg finst nokre identiske verbparadigme: 230 (for *flyte* m.fl) og 376 (*tryte*), 169 (kodar ingenting) og 180 (*bestri*), 167 (*lide*) og 182 (*bestride*), 033 (*telgje*) og 044 (*svelgje*). Desse er så sjeldne, få og like at det burde vere ei smal sak å rydde opp i dei.

### 6.5.3 Om vedlagte filer

Dei som har den skriftlege utgåva av denne oppgåva, har også fått med ein diskett med programkode og eit lite utdrag av regelfiler og *leksikon.tdl*, nok

<sup>28</sup>Eg tok meg ikkje bryet å finne leieformene, så berre den grammatiske informasjonen og fullformsendingane er der.

<sup>29</sup>Eg synest *Bokmålsordboka* burde gjere det same, korfor ikkje ein tilvisingsartikkel frå *klær* til definisjonsartikkel under *klede*?

til at ein kan utforske dei to siste leksikona i praksis i *LKB*.<sup>30</sup>

For dei som finn oppgåva via nettet, vil eg opplyse om at eg i skrivande stund har ei heimeside ved UiO. På <http://folk.uio.no/leivhe/ptype/> vil lesaren kunne finne den same koden som er på disketten.

Sjølv om denne sida skulle verte borte rimeleg snart, vil eg halde filene tilgjengelege i, tja, lat oss seie eit par år framover. Med ein god søkjemotor skulle det vere mogleg å oppdrive den aktuelle sida med filene.

Det tener ikkje til så mykje å gå gjennom filene her. I staden vil eg vise til fila *LESMEG* (på disketten) og <http://folk.uio.no/leivhe/ptype/LESMEG>.

---

<sup>30</sup>Det er nødvendig med ein versjon som er frå minst august 2003. Først då fekk *LKB* ressursar til å handsame så mange bøyingsreglar som eg trong.



# Litteratur

- A. Copestake. *Implementing Typed Feature Structure Grammars*. CSLI Publications, 2002.
- R. T. Endresen. Kognitiv morfologi i eit faghistorisk perspektiv. *Norsk Lingvistisk Tidsskrift*, 14(?):105–141, 1996.
- T. Fretheim. Er bokmålet tvekjønnet eller trekjønnet? In E. H. Jahr and O. Lorentz, editors, *Morfologi/Morphology*, number 3 in Studier i norsk språkvitenskap/Studies in Norwegian Linguistics, pages 99–101. Novus Forlag, 1985.
- B. F. Jakob Sverdrup, Marius Sandvei. *Tanums store rettskrivningsordbok*. Tanum-Norli, 1983.
- K. I. V. Jan Terje Faarlund, Svein Lie. *Norsk referansegrammatikk*. Universitetsforlaget, 1997.
- M. I. Landrø and B. Wangensteen, editors. *Bokmålsordboka*. Universitetsforlaget, 1984.
- P. H. Matthews. *Morphology*. Cambridge University Press, 2nd edition, 1991.
- I. A. Sag and T. Wasow. *Syntactic Theory - A Formal Introduction*. CSLI Publications, 1999.