

# Utvida rettleiing til søk i nynorskkorpuset

## 1 Søk på ordform

For å søkje på ordform, set ein hermeteikn rundt søkjeordet:

```
"bil"
```

Gløymer ein hermeteikna, får ein ikkje tilslag.

## 2 Trunkerte søk

Ein kan også gjere trunkerte søk. Trunkeringsteiknet er punktum følgd av stjerne, slik:

```
"bil.*"
```

Her får ein tilslag i alle ordformer som byrjar med "bil", irekna den nakne ordforma.

Dersom ein vil ha samansetjingar med "bil" som etterledd, set ein trunkeringsteiknet fyrst:

```
".*bil"
```

Ein kan også trunkere i bae endar av ei ordform, eller inni ordforma.

## 3 Søk på fleire ordformer

Ofte kan ein spare tid på søkje etter fleire ordformer i rekkjefølgje. Døme:

```
"kjem" "til" "å" "bli"
```

```
"i" "og" "for" "seg"
```

```
"til" "og" "med"
```

Rekkjefølgja på søkjeargumenta gjev rekkjefølgja i søkeresultatet.

Det må vera eitt mellomrom mellom kvart av orda.

## 4 Søk etter alternativ

Dersom ein vil søkje på alternative former, t.d. ha med ulike rettskrivingsvariantar, kan ein føre spesifiserte alternativ i søkjeargumentet. Søkjealternativ står innanfor parentes og er skilde med rett, vertikal strek. Det kan til dømes sjå slik ut:

```
"xxx (SKRIVEMÅTE1 | SKRIVEMÅTE2) xxx"
```

Vil ein til dømes ha med alle former av "følgja/fylgja" kan ein skriva det slik:

```
[word="f(ø|y)lg.*"]
```

Ein kan òg føre meir enn to alternativ.

Dette dømet:

```
[word="(m|d|s)eg"]
```

vil gje tilslag for *meg*, *deg*, *seg*.

## 5 Kombinerte søkekriterium

Ein kan avgrense søk ved å kombinere ulike søkekriterium. Da må ein føre søkjeargumenta innanfor hakeparentes. For "og" skriv ein "&".

Formelen blir då slik:

```
[ATTRIBUTT1="xxx" & ATTRIBUTT2="yyy"]
```

I avsnittet under er det vist døme på kombinert søk på ordform frå ei viss kjelde. Ein kan kombinere søkekriterium med alle attributt som er brukte i korpuset (sjå øvste tabell på hovudsida).

## 6 Søk på kjelde

I Nynorskkorpuset kan ein avgrense søk etter belegg til ein viss tekst eller ei viss gruppe tekstar. Ein søkjer då på ordform eller lemma kombinert med kjeldesignaturen verket har i Norsk Ordbok. Attributtet for kjelde er

"ref".

Søk på kjeldesignatur kan trunkerast på same måten som søk på ordform ("word"). Søk på kjeldesignatur er definert som å søkja på ein eigenskap ved eit ord. Søkjer ein berre på kjeldesignatur, får ein feilmelding. Merk at ein må ta med trunkeringsteiknet sist i alle kjeldesignaturar for å få treff (dette fordi kvar side i alle kjeldene er tagga med ein unik sidetag).

### Døme:

```
[word="koffert.*" & ref="SS.1990.*"]
```

Ein vil her få treff på alle samansetjingar med "koffert" (i tillegg til den nakne ordforma) i tidsskriftet Syn og Segn årgang 1990.

I Norsk Ordboks kjeldeliste er det mellomrom i kjeldene, slik:

```
"SS 1990"
```

Når ein søkjer på kjelde i korpuset må ein bytte dette mellomrommet med punktum, slik:

```
"SS.1990"
```

Ei komplett liste over kjeldeforkortingar, med fullskrivningar, finn du [her](#).

## 7 Søk på lemma og ordklassetagg

Ein kan òg avgrense søk i korpuset ved å søkje fram bøygde former av ord. Merk at søk på oppslagsform (attributtet "lemma") og ordklassetagg (attributtet "msd") gjev tilslag berre i tekstar frå 1938 eller seinare. Tilslag ein får i tidlegare tekstar, er ikkje pålitelege.

Om ein til dømes søkjer slik:

```
[lemma="båt" & msd="SAM.*"]
```

vil ein få treff på alle bøygde former av hannkjønnsst substantivet "båt" i tillegg til grunnforma. Bokstavane i "msd"-taggen står her for "substantiv", "appellativ" og "maskulinum". Ein kan avgrense slike søk enda meir enn i dømet ovanfor. Sjå neste punkt for å lese meir om ordklassetaggar.

## 8 Ordklassetaggar

Fire ordklassar er merkte med ein bokstav:

- S substantiv
- A adjektiv
- P partisipp
- D determinativ (artiklar, eigedomspronomen, og pronomen som *all*, *kvar*, *ingen*, *nokon*)

Desse ordklassane har eit eigenskapsskjema knytt til seg:

### Taggeskjema for substantiv:

1 ordklasse	2 type	3 kjønn	4 tal	5 kasus	6 binding
S = substantiv	A = appellativ P = proprium	F = femininum M = maskulinum N = nøytrum	E = eintal F = fleirtal	G = genitiv O = alt anna	U = ubunden B = bunden

### Taggeskjema for adjektiv:

1 ordklasse	2 type	3 grad	4 kjønn	5 tal	6 kasus	7 binding
A = adjektiv	Q = kvalitativ	P = positiv K = komparativ S = superlativ	M = maskulin N = nøytrum	E = eintal F = fleirtal	G = genitiv O = alt anna	U = ubunden B = bunden

### Taggeskjema for partisipp:

1 ordklasse	2 type	3 grad	4 kjønn	5 tal	6 kasus	7 binding*
-------------	--------	--------	---------	-------	---------	------------

P = partisipp	F = perfektum P = presens	0	2 = feminin og maskulin N = nøytrum 0 = syner ikkje genus	E = eintal F = fleirtal	0	U = ubunden B = bunden
---------------	------------------------------	---	---	-------------------------------	---	---------------------------

\*partisipp som står predikativt er alltid tagga med 0 for binding

### Taggeskjema for determinativ:

1 ordklasse	2 type	3 kjønn	4 tal	5 kasus	6 binding
D = determinativ	D = demonstrativ K = talord P = eigedomspronomen	M = maskulinum F = femininum N = nøytrum	E = eintal F = fleirtal	G = genitiv 0	0 = ubunden / uviss B = bunden

Om du til dømes ønskjer å søkje på lemmaet "hus", men berre vil ha treff i bunden form eintal, ser altså søkjeargumentet slik ut:

```
[lemma="hus" & msd="SANE0B.*"]
```

Merk at du i 5. ledd i ordklasseattributtet (kasus) må fylle ut "0" om du ønskjer treff på anna enn genitivsform. Om du gløymer å fylle inn noko i dette leddet, får du ikkje treff. Ein kan altså ikkje ha med berre siste eigeskapsledd i eit slikt søkjeargument utan å også fylle ut dei føregåande. Dette gjeld også for dei andre ordklassane i skjemaet over.

Dei andre ordklassane har meir eksplisitte (intuitive) taggar:

ADV	adverb
INF-M	infinitivsmerke
INTERJ	interjeksjon
KONJ	konjunksjon
PREP	preposisjon
PRON-	(pluss særdrag) pronomen
SBU	(subordinasjon - underordning)
TALL	(romartal)
V-	(pluss formdrag)
X	ikkje-nynorsk ord

Verb med tilleggstagga for verbform er:

V	verb
V-INF	verbform i infinitiv
V-PRES	verbform i presens
V-PRET	verbform i preteritum
V-INF-PRES-ST-FORM	verbform på -st i infinitiv eller presens

## 10 Nokre mykje brukte kombinerte søk

Vanlege søk på lemma + ordklasse er:

[lemma="xxx" & msd="SAM.\*"] (substantiv i maskulinum)  
[lemma="xxx" & msd="SAF.\*"] (substantiv i femininum)  
[lemma="xxx" & msd="SAN.\*"] (substantiv i nøytrum)  
[lemma="xxx" & msd="V.\*"] (verb)  
[lemma="xxx" & msd="A.\*"] (adjektiv)  
[lemma="xxx" & msd="P.\*"] (partisipp)

Eit par vanlege fleirordige søk er

søk på "verb + partikkel", som vil gje belegg med t.d. *halda på, held fram, heldt til*:

```
[lemma="hald(e|a)" & msd="V.*"] [msd="PREP"]
```

søk på nominalfraser som t.d. *"ein liten hund / to små hundar / den vesle hunden"*:

```
[msd="D.*"] [msd="A.*"] [lemma="hund" & msd="S.*"]
```

Klipp ut søkjeargumenta, lim dei inn i søkjefeltet og prøv sjølv!

## 11 Søk etter element som ikkje står rett etter kvarandre

Dersom ein vil søkje etter ord som står på litt avstand frå kvarandre, set ein inn eit sett tomme hakeparentesar, fylgde av eit sett sløyfeparentesar (utan mellomrom mellom). Inni sløyfeparentesane skriv ein inn det talet med ord som ein vil ha imellom, t.d. slik

```
[ ]{2}
```

Då søker ein etter tilslag som har to uspesifiserte ord mellom dei som er spesifiserte. Vil ein ha frå så til så mange ord imellom, skriv ein inn baa verdiar, skilde med komma:

```
[ ]{0,5} = 'skild med frå null til fem ord'
```

"0,5" tyder "frå null til fem ord imellom". Tala kan ein sjølvsagt endra etter behov.

**Døme:**

```
[lemma="halde" & msd="V.*"] [ ]{0,5} [word="seg"]
```

Da får ein tilslag som *"halda seg, held seg, halda fram med å kalla seg, heldt på å ta livet av seg"*.

## 12 Søk etter teikn i teksten

Ein kan ha bruk for å søkja etter teikn som ikkje er vanlege bokstavar, som til dømes punktum (".") eller komma (","). Det kan t.d. vere at ein berre er interessert i apposisjonar. Da er det to ting ein må ta omsyn til:

1. Alle teikn i teksten som ikkje er bokstavar, er skilde ut som eigne "ord" (tokens). Dei har mellomrom både før seg og etter seg. Når ein søker etter eit teikn, søker ein derfor etter attributtet word.

2. Mange av teikna er del av søkjespråket CQP. For å markera at søket gjeld sjølve teiknet, set ein omvend skråstrek (backslash) rett framfor.

Dersom ein vil søkja på teikn, til dømes punktum, skriv ein altså:

```
[word="\."]
```

eller

```
"\."
```

**Ei åtvaring:** søk etter berre teikn vil truleg ikkje gje treff lenger ut i korpuset enn til Firda eller Dag og Tid, så søk etter vanlege teikn bør kombinerast med søk på vanlege ord.

**Døme:**

Vil ein ha preposisjonen "med" framfor punktum, ser søkjestrengen slik ut

```
[word="med"] [word="\."]
```

eller

```
"med" "\."
```