

Klikk for å vise/skjule kodeceller

## Nynorsk og bokmål ved Nasjonalbiblioteket

Lars G. Johnsen

[lars.johnsen@nb.no](mailto:lars.johnsen@nb.no)

Nettbiblioteket 52 000 bøker med kode nno, og 460 000 med kode nob

Målfridprosjektet i samarbeid med Språkrådet.

Innhosting av statlige domener på internet. Målt i antall løpende forekomster, er det 290 millioner nynorsk og 2,8 milliarder bokmål.

Norsk Ordbank – Hvor mange fullformer, hvor mange lemnaer?

Dataene er i en tabell der cellene inneholder 1 om ordet finnes i respektive ordbok. Om ordformen er med i begge har felles-cellen også verdien 1.

### Fullformer

	nob	nno	felles
<b>merlespikar</b>	0.0	1.0	0.0
<b>dampkjekkener</b>	1.0	0.0	0.0
<b>skihoppingane</b>	0.0	1.0	0.0
<b>tertiærtider</b>	0.0	1.0	0.0
<b>statiskest</b>	1.0	0.0	0.0
<b>fagfellevurderinga</b>	0.0	1.0	0.0
<b>fiskarbåt</b>	0.0	1.0	0.0
<b>koses</b>	1.0	0.0	0.0
<b>påliteleglaupa</b>	0.0	1.0	0.0
<b>hinderbaner</b>	1.0	1.0	1.0

### Lemnaer

	nob	nno	felles
<b>anestesi</b>	1.0	1.0	1.0
<b>flæe</b>	0.0	1.0	0.0
<b>kommandoraid</b>	1.0	1.0	1.0
<b>helmepleier</b>	1.0	0.0	0.0
<b>etterhogst</b>	1.0	1.0	1.0
<b>Silver</b>	0.0	1.0	0.0
<b>sutte</b>	1.0	1.0	1.0
<b>jippo</b>	1.0	1.0	1.0
<b>in persona</b>	1.0	1.0	1.0
<b>a-våpen</b>	1.0	1.0	1.0

Det er en betydelig forskjell i inventar mellom bokmål og nynorsk

Antall lemma, bokmål har under 75 % av antallet for nynorsk.

**nob:** 74389.0

**nno:** 105042.0

**felles:** 47667.0

Antall fullformer

**nob:** 342977.0

**nno:** 400245.0

**felles:** 127387.0

Ratioen fullform til lemma ligger noe over 4 for bokmål, og noe under 4 for nynorsk.

Fullformer for *beveg, bil, rod, lang*: lengre, lang, rødd, rødes, bevegete, biles, bilene, bevegede, bilende, rødeste, bevegende, langes, rødde, beveg, rødere, langer, beveges, biler, lengst, bevege, rødt, rødende, beveger, langa, rod, bevegd, lengste, bilt, langt, bilte, langete, langende, beveget, bil, lange, røde, bevega, bevegde, bilen, lenger, rødest, langet, langede, røder, bile

45

Fullformer for *beveg, bil, rod, lang*: lengre, lang, rødd, langane, bevegast, beveg, bevegande, rødest, rødande, røda, langast, bevegar, biler, beveg, rødere, langar, bilar, bevege, lengst, langande, beveger, langa, rod, bevegd, lengste, bilt, langt, bila, bilande, bil, lange, røde, beveget, bevega, bilane, bilast, langan, bevegde, bilen, rødt, rødde, røder, bile

42

Hvor mye er likt fra et ordboksperspektiv? Det er verdt å merke seg at det er betydelig overlapp, nynorsken har en større grad av egne ord enn bokmål

Bokmålsformer delt med nynorsk %.0: 37.0

Nynorskformer delt med bokmål %: 32.0

Felles lemma i % av bokmål 64.0

Felles lemma i % av nynorsk % 45.38

Fullformer pr. lemma i bokmål 4.6

Fullformer pr. lemma i nynorsk 3.8

## Distribusjon i tekster

Ser på tekst fra Språkrådets nettsider: <https://www.sprakradet.no/Vi-og-vart/Publicasjoner/Spraaknytt/spraknytt-42020/byggejer-samhald-med-spraklov/>

	nob	nno	tekst
<b>i</b>	1	1	37
<b>det</b>	1	1	33
<b>er</b>	1	1	31
<b>at</b>	1	1	28
<b>og</b>	1	1	28
<b>på</b>	1	1	24
<b>eg</b>	0	1	20
<b>vi</b>	1	1	20
<b>som</b>	1	1	20
<b>å</b>	1	1	15
<b>for</b>	1	1	15
<b>til</b>	1	1	14
<b>har</b>	1	1	14
<b>han</b>	1	1	14
<b>med</b>	1	1	14
<b>ikkje</b>	0	1	13
<b>eit</b>	0	1	13
<b>skal</b>	1	1	12
<b>raja</b>	1	1	11
<b>ein</b>	0	1	10
<b>må</b>	1	1	10
<b>dei</b>	0	1	9
<b>språklova</b>	0	0	9
<b>men</b>	1	1	9
<b>nynorsk</b>	1	1	9
<b>norsk</b>	1	1	9
<b>lova</b>	1	1	9
<b>ei</b>	1	1	9
<b>av</b>	1	1	8
<b>om</b>	1	1	8
<b>også</b>	1	1	7
<b>vil</b>	1	1	7
<b>vart</b>	1	1	7
<b>språk</b>	1	1	7
<b>fram</b>	1	1	7
<b>var</b>	1	1	7
<b>opp</b>	1	1	6
<b>språket</b>	1	1	6
<b>likestilling</b>	1	1	5
<b>språklov</b>	0	0	5
<b>abid</b>	0	0	5
<b>språkrådet</b>	0	1	5
<b>språkpolitikken</b>	1	1	5
<b>bokmål</b>	1	1	5
<b>mellom</b>	1	1	5
<b>frå</b>	1	1	5
<b>panjabi</b>	1	1	5
<b>kan</b>	1	1	4
<b>urdu</b>	1	1	4
<b>nå</b>	1	1	4

	nob	nno	tekst
<b>ansvar</b>	1	1	3
<b>skulen</b>	0	1	3
<b>mykje</b>	0	1	3
<b>korleis</b>	0	1	3
<b>år</b>	1	1	3
<b>vårt</b>	1	1	3
<b>viktig</b>	1	1	3
<b>ut</b>	1	1	3
<b>svært</b>	1	1	3
<b>stortinget</b>	1	1	3
<b>norske</b>	1	1	3
<b>lære</b>	1	1	3
<b>vidaregåande</b>	0	1	3
<b>lagt</b>	1	1	3
<b>gode</b>	1	1	3
<b>gjennom</b>	1	1	3
<b>for</b>	1	1	3
<b>følge</b>	1	1	3
<b>får</b>	1	1	3
<b>få</b>	1	1	3
<b>fire</b>	1	1	3
<b>den</b>	1	1	3
<b>del</b>	1	1	3
<b>da</b>	1	1	3
<b>arbeidet</b>	1	1	3
<b>ha</b>	1	1	3
<b>Innvandrarklasse</b>	0	0	3

Hvor mye tekst er det totalt, merk at **nob** og **nno** bare står for type, ikke forekomst.

**nob:** 319.0

**nno:** 419.0

**tekst:** 1177.0

Hvor mye er bare nynorsk? I teksten er 207 forekomster rene nynorskord så noe under 20 % består av ord som ikke finnes i bokmål.

**nob:** 0.0

**nno:** 4.0

**tekst:** 5.0

Hvor stor del av teksten utgjør topp 50? Omtrent 50 % av totalen - her har vi 586 forekomster av totalt 1177. For mindre tekster er utgjør topp 50 rundt 50 % av ordmengden, som kryper nedover etter hvert som tekstens størrelse øker.

**nob:** 41

**nno:** 47

**tekst:** 586

## En bokmålstekst

Også fra Språkrådet. <https://www.sprakradet.no/Vi-og-vart/Publicasjoner/Spraaknytt/spraknytt-42020/sprakdagen-pa-tv/>

	nob	nno	tekst
<b>og</b>	1	1	36
<b>det</b>	1	1	30
<b>i</b>	1	1	25
<b>er</b>	1	1	20
<b>for</b>	1	1	18
<b>at</b>	1	1	16
<b>om</b>	1	1	15
<b>vi</b>	1	1	15
<b>å</b>	1	1	13
<b>til</b>	1	1	12
<b>har</b>	1	1	12
<b>de</b>	1	1	12
<b>som</b>	1	1	12
<b>på</b>	1	1	11
<b>fra</b>	1	0	10
<b>vaet</b>	0	0	9
<b>solvang</b>	0	0	9
<b>foto</b>	1	1	9
<b>nordskog</b>	0	0	9
<b>johnny</b>	0	0	9
<b>et</b>	1	1	8
<b>med</b>	1	1	8
<b>norge</b>	0	0	7
<b>skal</b>	1	1	7
<b>må</b>	1	1	7
<b>en</b>	1	1	6
<b>ikke</b>	1	0	6
<b>språkloven</b>	0	0	6
<b>språk</b>	1	1	6
<b>hvordan</b>	1	0	6
<b>gjør</b>	1	1	5
<b>språkdagen</b>	0	0	5
<b>være</b>	1	1	5
<b>fredrik</b>	0	0	5
<b>ut</b>	1	1	5
<b>espen</b>	0	0	5
<b>kan</b>	1	1	5
<b>mange</b>	1	1	5
<b>alle</b>	1	1	5
<b>språkene</b>	1	0	5
<b>den</b>	1	1	5
<b>av</b>	1	1	5
<b>nakstad</b>	0	0	4
<b>offentlige</b>	1	0	4
<b>norsk</b>	1	1	4
<b>så</b>	1	1	4
<b>hva</b>	1	0	4
<b>nynorsk</b>	1	1	4
<b>også</b>	1	1	4
<b>rønn</b>	0	0	4

Hvor mye tekst er det totalt, merk at **nob** og **nno** bare står for type, ikke forekomst.

**nob:** 364.0

**nno:** 255.0

**tekst:** 989.0

Hvor mye er bare nynorsk? I teksten er 207 forekomster rene nynorskord så noe under 20 % består av ord som ikke finnes i bokmål.

**nob:** 0.0

**nno:** 4.0

**tekst:** 6.0

De fire nynorskordene i teksten ser ut til å skyldes Norsk Ordbank, ikke teksten selv.

digitalisering, anne, helsesektoren, språkrådet

Men hvor mye er bare bokmål? Antall forekomster utgjør her ca 17 %, med 164 forekomster totalt av 989.

**nob:** 113.0

**nno:** 0.0

**tekst:** 164.0

Hvor stor del av teksten utgjør topp 50? Omtrent 50 % av totalen, samme som for nynorskteksten - her har vi 461 forekomster av totalt 989.

**nob:** 39

**nno:** 33

**tekst:** 461

Legg merke til at både nynorsk og bokmål er høyt oppe i begge tekstene. Nynorskteksten drar mot nynorsk, mens bokmålsteksten drar andre veien. Her tallene for topp 50 for nynorskteksten.

**nob:** 41

**nno:** 47

**tekst:** 586

Hvordan bruke det her i videre arbeid med språknormer?

- Finne korresponderende ord
- Finne register
- Hvor mange hybridtekster er det?
- Kvantisere eksponering av språklig variasjon på et ordnivå