

Språkbanken

Arne Martinus Lindstad
arne.lindstad@nb.no





Språkbank?

- Infrastruktur for norsk språkteknologi (forskning og næringsliv)
- Lager og distribuerer språkressurser



Språkressurser?

- Innsamlede språkdata – tekst og tale (lyd)
- Brukes til å «lære» datamaskiner språk
- Store datamengder som skal konsumeres av datamaskiner
- I større eller mindre grad lingvistisk analysert på forhånd



N-gram

- Brukes til å lage statistiske språkmodeller
- Trigram (3-gram):
 - *brukes til å*
 - *til å lage*
 - *å lage statistiske*
 - *lage statistiske språkmodeller*
- Kan benyttes f.eks. i mobiltelefoner (prediktiv skriving), talegjenkjenning

N-gram

Google

en svale|



en svale

en svale **gjør ingen sommer**

en svale **gør ingen sommer**

en svale **gir ingen sommer**

Press Enter to search.

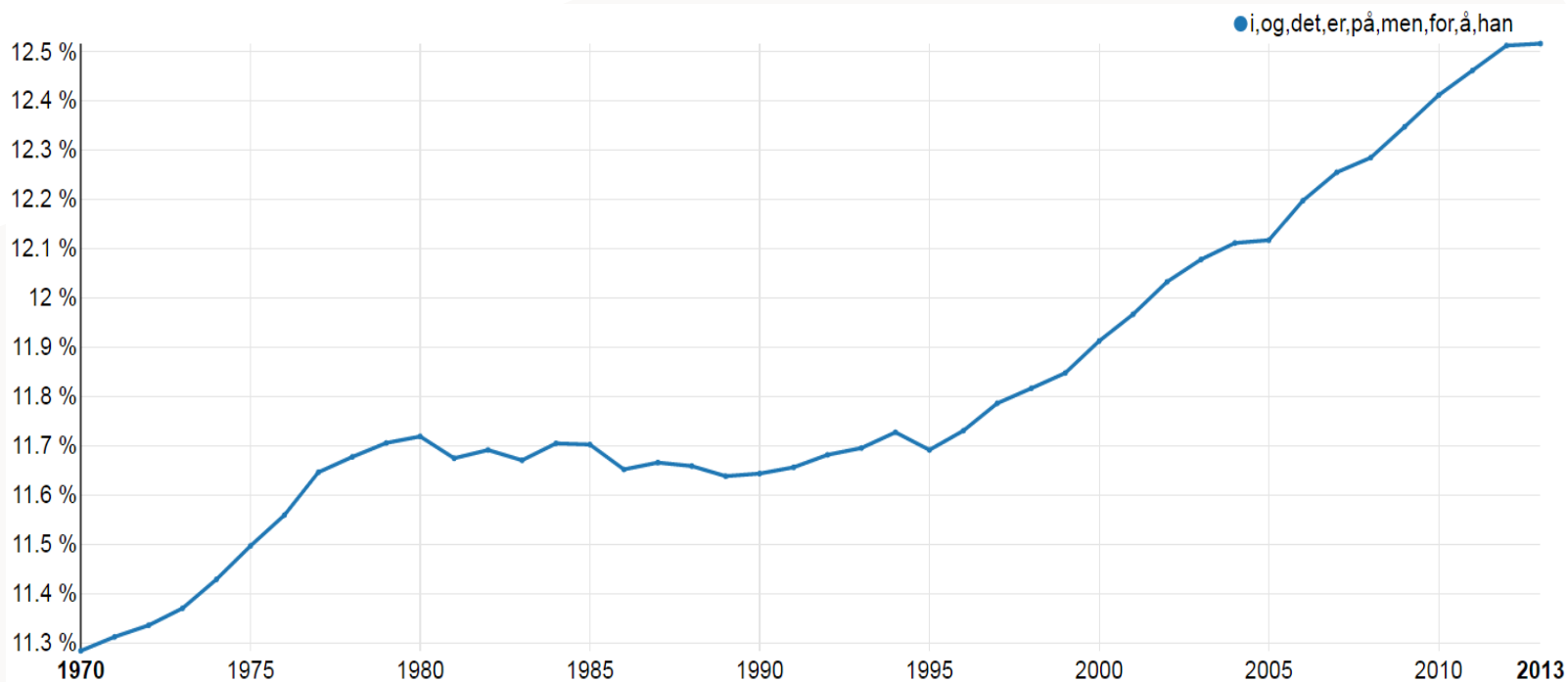




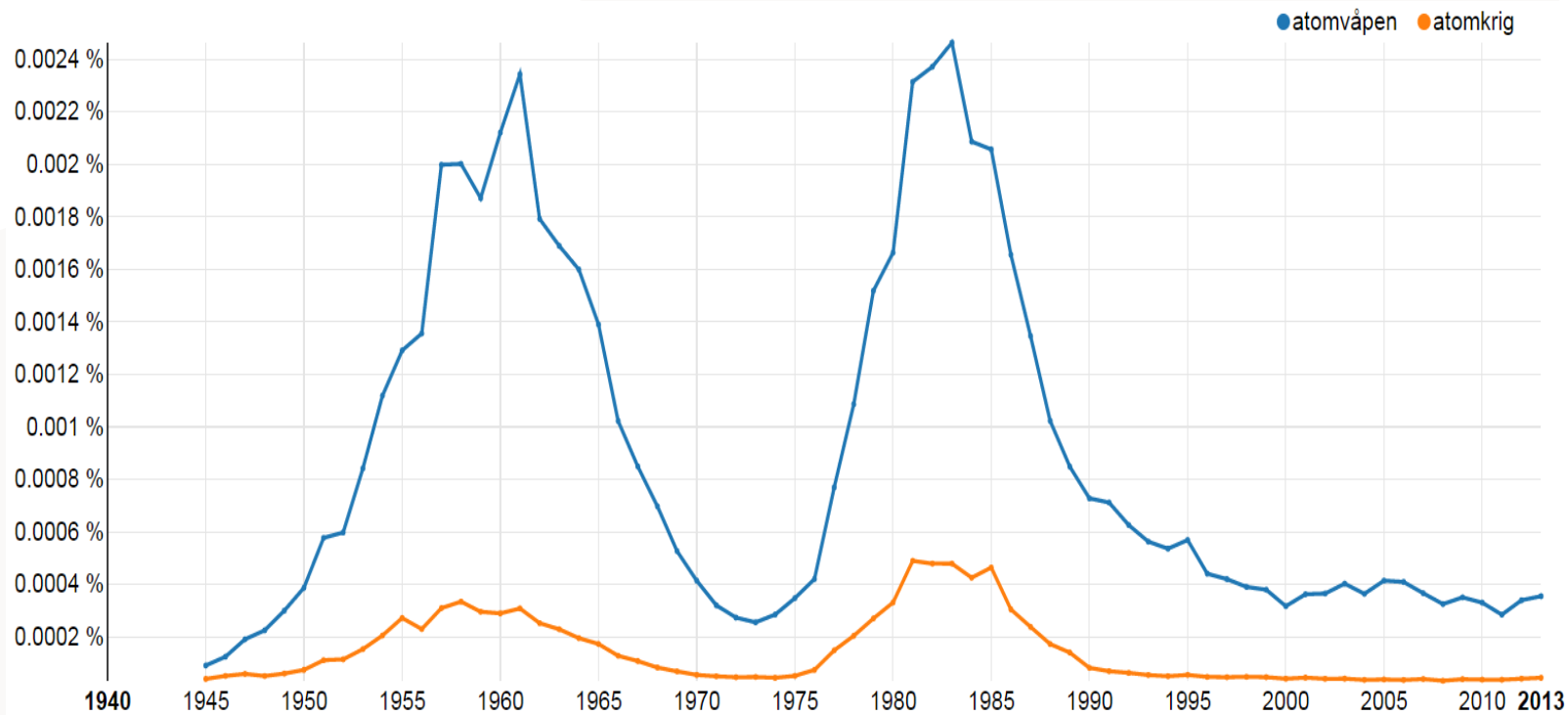
Ordlister (f.eks. frekvenslister)

- Telle alle ord i norske aviser og/eller bøker utgitt i et gitt tidsrom, og rangere dem etter de mest brukte
- Kan være utgangspunkt for mange slags analyser, både språklig og ellers

Hva består tekster av?



Avissøk: historiske hendelser

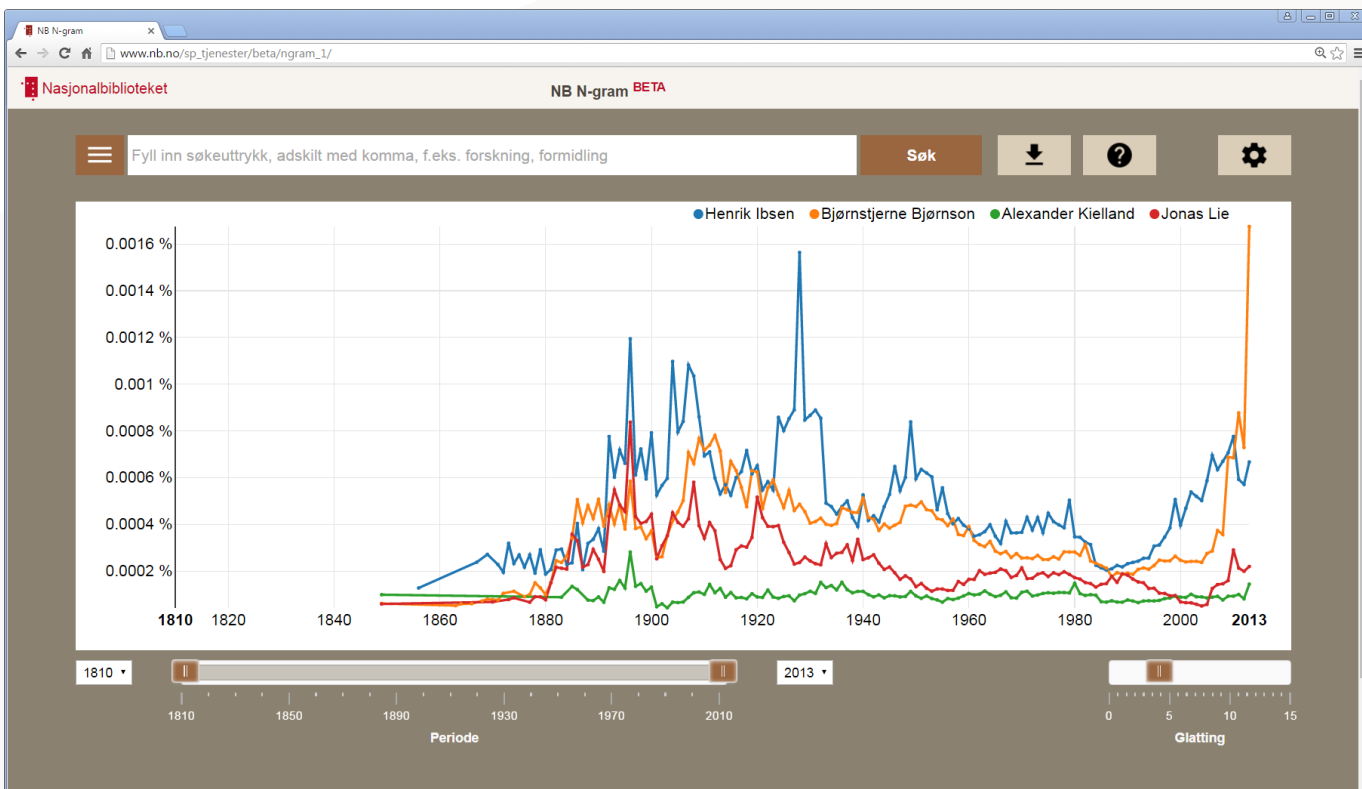




Nasjonalbiblioteket

- Språkbanken profiterer også på ressursene i Nasjonalbiblioteket
- Tjenester basert på digitalisert materiale
 - forskere
 - skoleverket
 - allmennheten

NB N-gram



Språkbankens ressurskatalog

Språkbanken - Nasjonalbiblioteket

Språkbankens ressurskatalog **BETA**

Skriv inn søkeord



















Søk Nullstill

Om katalogen

Viser ressursene: 1 til 40 av totalt 42

Forrige Neste

Språkbanken CLARINO

Tekst	Tale	Tekst	Verktøy	Leksikon	Tekst
Taggede bokmålstekster fra NBdigital (fritt tilgjengelige tekster)	NB Tale - en grunnleggende akustisk-fonetisk taledatabase for norsk	N-grammer fra NBdigital	NB N-gram	Norsk Ordvev - Bokmål	Norsk avis korpus
 07.03.2016	 25.02.2016	 24.02.2016	 24.02.2016	 22.02.2016	 04.02.2016
Tekst	Tekst	Tekst	Leksikon	Leksikon	Tekst
N-gram for bokmål (basert på Norsk avis korpus)	Norske oversettelser av Acquis Communautaire	N-gram for bokmål (basert på nyhetstekst fra NST)	Norsk Ordbank - Bokmål	Ordlister fra Språkrådet	N-gram for nynorsk (basert på Norsk avis korpus og NST)
 04.02.2016		 03.02.2016	 03.02.2016	 03.02.2016	 03.02.2016
Leksikon	Tale	Tale	Tale	Tale	Tale
SCARRIE	Taledatabaser for norsk, opprinnelig produsert av NST (16 kHz)	Taledatabase for norsk, opprinnelig produsert av NST (22 kHz)	NST database for norsk talesyntese (44 kHz)	Taledatabaser for svensk, opprinnelig produsert av NST (16 kHz)	Taledatabaser for svensk, opprinnelig produsert av NST (22 kHz)
 03.02.2016	 03.02.2016	 03.02.2016	 03.02.2016	 03.02.2016	 03.02.2016



Kontaktinformasjon

- nb.no/sprakbanken – katalog
- www.sprakbanken.no – lenker til tjenester
- sprakbanken@nb.no – e-post