

# Kvantitative metoder inden for korpuslingvistiske projekter

## Øvelser

Jørg Asmussen

23. august 2005

## 1 Grundlæggende

### 1.1 Metodeforskelle

Overvej, hvorved den kvantitative metode adskiller sig fra den kvalitative mht.

- holdningen til fænomenernes grundlæggende beskaffenhed
- forskerens position i forhold til disse fænomener
- måden, fænomenerne kategoriseres på
- præsentation af resultater
- afprøvning af hypoteser

## 2 Korpuslingvistik

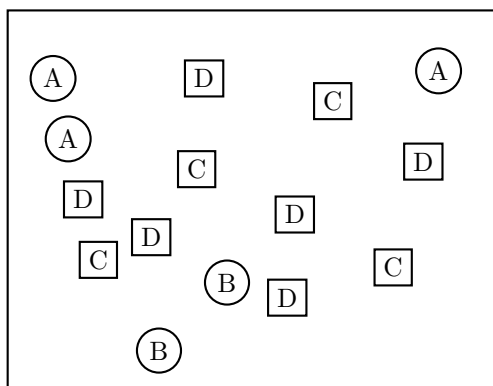
### 2.1 Korpuslingvistisk eller korpusbaseret?

Giv et bud på, hvad man kan (eller bør) forstå ved *korpuslingvistik*. Hvilken forskel gør det, om man kalder sin metode for *korpuslingvistisk* eller *korpusbaseret*?

## 3 Finde, tælle, konkludere

### 3.1 Data-analyse

Diskuter på baggrund af nedenstående figur udsagnet:  
*I materialet er der flere forekomster af C end af B.*



## 4 Sammenligninger

### 4.1 Ordbegrebet

Betragt nedenstående tekstudsnit...

- Hvor mange ord indeholder nedenstående tekstudsnit?
- Hvor mange adjektiver?
- Hvor mange sætninger?
- Hvor mange ord med en 'økonomisk' betydning?

Hvilke problemer er der ved identifikation og afgrænsning af fænomener som *ord*, *adjektiver*, *sætninger*, *betydning*?

Afdeling IT-infrastruktur præsterede årets laveste afkast, idet afdelingen faldt næsten 60 pct. i 2001. Den lavere vækst i verdensøkonomien gik hårdt ud over den type selskaber, denne meget specialiserede afdeling investerer i. De store og populære obligationsafdelinger klarede sig gennemgående tilfredsstillende. Afdelinger som Korte Danske, Dannebrog og Lange Danske, som alle investerer i danske obligationer, gav stabile afkast på omkring 6 pct. og udbetaler pæne udbytter for året. Årets samlede regnskabsmæssige resultat i afdelingerne under Investeringsforeningen Danske Invest og Specialforeningen Danske Invest (opgjort før skat i den akkumulerende afdeling Stabil) blev på ÷3.165,9 mio. kroner i 2001 mod ÷139,8 mio. kroner for 2000.

### 4.2 Tekstrepræsentation

Betragt nedenstående tekstudsnit. Forklar, hvad det viser, og gør rede for metodiske problemstillinger.

```

1> mm $$ S%%
2>
A>tten B%%%% din bror som bor her hva- hvad hedder han mm $ og det
K>
-----
1> otteogtyve jeg har også en anden bro
2>
A> er ham der er syvogtyve ? otteogtyve $$$
K>
-----
1>r der er femogtyve $ men han er~ $$ han er en uge indenfor soldaterne $ han
2>
A> nå
K>
-----
1> er lige blevet ind for soldaterne ik' altså~ $ (fu) hvad hedder det (lat
2>
A> hjnå
K>
-----
1>ter) genindkaldt så til ham~ kan vi ikke (uf) foreløbi
2> (uf) genindkaldt
A> nå~ $$
K>
-----

```

### 4.3 Frekvensprofiler

Sammenlign tabellerne over de hyppigste ord i K90 og K200 (appendiks B.5 og B.6 på side 39 og 40 i noterne).

- Hvilke konklusioner kan man drage på baggrund af disse frekvensprofiler?
- Hvilke metodiske problemstillinger er der forbundet med en sammenligning af profilerne?

## 5 Struktur og kollokation

### 5.1 Kollokater

Betragt nedenstående figurer, der stammer fra Korpus 2000.

- Hvad viser de?
- Hvad er forskellen på dem?
- Hvordan er de fremkommet?

Korpus 2000		Korpus 90	
venstre	højre	venstre	højre
1000 i præp.	1000 og konj.	1000 i præp.	1000 og konj.
1000 en art.	1000 i præp.	1000 en art.	1000 i præp.
1000 og konj.	1000 er vb.	1000 og konj.	1000 der pron.
1000 den art.	1000 der pron.	1000 den art.	1000 er vb.
1000 af præp.	100 med præp.	1000 af præp.	1000 med præp.
1000 med præp.	100 som pron.	1000 med præp.	100 som pron.
100 de art.	100 på præp.	1000 de art.	100 en art.
100 er vb.	100 har vb.	100 er vb.	100 af præp.
100 for præp.	100 en art.	100 denne pron.	100 han pron.
100 at konj.	100 for præp.	100 hans pron.	100 på præp.

Korpus 2000		Korpus 90	
venstre	højre	venstre	højre
<ul style="list-style-type: none"> <li>glidende vb.</li> <li>roterende adj.</li> <li>langsomme adj.</li> <li>blide adj.</li> <li>nazistiske adj.</li> <li>folkelige adj.</li> <li>pludselige adj.</li> <li>rytmiske adj.</li> <li>hurtige adj.</li> <li>folkelig adj.</li> </ul>	<ul style="list-style-type: none"> <li>hånden sb.</li> </ul>	<ul style="list-style-type: none"> <li>nyreligiøse adj.</li> <li>fejende adj.</li> <li>sindene sb.</li> <li>roterende adj.</li> <li>langsomme adj.</li> <li>rytmiske adj.</li> <li>folkelige adj.</li> <li>revolutionære adj.</li> <li>glidende vb.</li> <li>rolige adj.</li> </ul>	<ul style="list-style-type: none"> <li>hurtige adj.</li> </ul>

## 6 Heuristisk tekstklassifikation

### 6.1 Tekstklassifikation

Beskriv, hvad man kan forstå ved *tekstklassifikation*. Prøv herefter at klassificere teksten fra opgave 4.1.