

Overture:

“if a person can’t communicate, the very least he can do is to shut up!”

Tom Lehrer.

“Human beings just can’t not communicate.”

Dorthe Duncker.

Ouverture:

Håndtering af sammensatte ord i dansk stavekontrol

Den store danske ordliste (Dsdo.)

Jacob Sparre Andersen,
Henrik Christian Grove og
Flemming Hjernøe

Overview:

- **Status for ordklasseopmrkning**
- **Regler for dannelse af sammensatte ord**
- **OpenOffice.orgs håndtering af sammensatte ord**

Status for ordklasseopmærkning:

- Center for Sprogteknologi (CST) har kørt vores ordliste gennem deres opmærkningssystem. Det er specielt godt til at identificere adjektiver, adverbier og andre irregulært bøjede ord.
- Henrik og jeg har designet og implementeret et opmærkningssystem der bygger på simpel mønstergenkendelse på ords endelser. Det er rimeligt effektivt til navneord og verber.

Status for ordklasseopmærkning:

Eksempel på data fra CST's opmærkningsystem:

sameksisteret/V_PARTC_PAST

sameksistr/N

samen/N

samens/N_GEN

samensunken/N

samensunket/V_PARTC_PAST

samensunkne/N

samensunknes/N_GEN

samer/N

samerne/N

samernes/N_GEN

samers/N_GEN

sames/V_PRES

Status for ordklasseopmærkning:

Eksempel på en kørsel af Henriks og mit opmærkningssystem:

```
word_class_identifier -max_word_length 50 \  
                      -max_word_count `wc -l < ordliste` \  
                      -suffixes +l  +llen  +ller  +llerne  \  
                               +ls  +llens  +llers  +llernes  \  
                      -labels $(NAVNEORDSFORMER) \  
                      -max_mismatches 0 \  
                      -format log \  
  
< ordliste \  
> navneord.flleskn.14.log
```

Status for ordklasseopmærkning:

Eksempel på data fra Henriks og mit opmærkningssystem:

```
word: afgangshal  
root: afgangshal  
conjugation: ubestemt ental  
comment: Found in input dictionary.
```

```
word: afgangshallen  
root: afgangshal  
conjugation: bestemt ental  
comment: Found in input dictionary.
```

```
word: afgangshallens  
root: afgangshal  
conjugation: bestemt ental, ejefald
```

Oversigt:

- Status for ordklasseopmærkning
- Regler for dannelse af sammensatte ord
- OpenOffice.orgs håndtering af sammensatte ord

Regler for dannelse af sammensatte ord:

Eksempler:

- skib+s+lejlighed
- skov+sø
- sammen+sætte
- både+byggeri
- blomster+eng
- ret+s+sag (eller ret+sag, siger jurister)
- møde+lokale

Er der et mønster?

Regler for dannelse af sammensatte ord:

Bud på nogle mønstre:

- navneord i ental+'s'+navneord → navneord
(skibslejlighed, retssag)
- navneord i ental+navneord → navneord
(skovsø, retsag, mødelokale)
- adverbium+verbum → verbum
(sammensætte)
- navneord i flertal+navneord → navneord
(bådebyggeri, blomstereng)

Oversigt:

- Status for ordklasseopmærkning
- Regler for dannelse af sammensatte ord
- **OpenOffice.orgs** håndtering af sammensatte ord

OpenOffice.orgs håndtering af sammensatte ord:

Aspell, Ispell og OpenOffice.org, har alle tre ^a samme tre muligheder for håndtering af sammensatte ord:

- Ingen sammensætninger er tilladte
- Alle ord mærket med et bestemt flag kan bruges som første del af et sammensat ord.
- Alle sammensætninger er tilladte

^atjek lige den pstand

OpenOffice.orgs håndtering af sammensatte ord:

- Oprindeligt har vi argumenteret for at køre i “alle sammensætninger er tilladte”-modus.
- Efter en del kritik af den løsning har Flemming og jeg eksperimenteret med kun at tillade en begrænset del af ordene i ordlisten i sammensætninger.

OpenOffice.orgs håndtering af sammensatte ord:

Da vi rettede op på OpenOffice.orgs håndtering af sammensatte ord valgte vi at lave et hurtigt hack, hvor jeg valgte i første omgang at sørge for at sammensatte ord med et navneord i ubestemt ental (eller ubestemt ental ejefald – for at få fuge-s'et med) kan bruges som første del af et sammensat ord.

Denne udgave af ordlisten til OpenOffice.org kan nu hentes fra:

<http://da.spelling.org/>

Konklusion:

Er der nogen konklusion på det her?
Andet end at vi skal arbejde videre?

Videre arbejde:

- Vi skal gøre en indsats for at få styr på hvilke ord (streng) der reelt bruges som første del af sammensatte ord.
- Lave en grammatik til at beskrive dannelse af sammensatte ord.
- Formulere de ordsammensætningsregler vi kender i den grammatik.
- Implementere en stavekontrol der kan udnytte disse regler.