

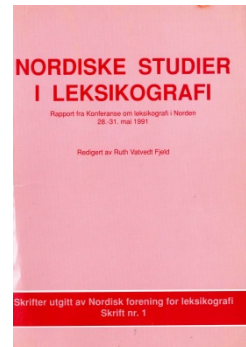
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Leksikografi i IBM Norge

Forfatter: Jan Engh

Kilde: Nordiske Studier i Leksikografi 1, 1992, s. 409-422
Rapport fra Konferanse om leksikografi i Norden, 28.-31. mai 1991

URL: <http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive>



© Nordisk forening for leksikografi

Betingelser for bruk af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Jan Engh

Leksikografi i IBM Norge

IBM Norge driver den største leksikografiske virksomheten her til lands utenfor universitetene. Arbeidet er samlet om databehandlingsterminologi og allmennspråk - både bokmål og nynorsk. Terminologiarbeidet er viktig for oversettervirksomheten i firmaet. Arbeidet med norsk allmennspråk er en forutsetning for utviklingen av språklige funksjoner i dataprogram. Mens de fleste programverktøyene er utviklet i utlandet, er det materialet de blir brukt på, stort sett lagd i Norge. Derfor vil det meste av framstillingen dreie seg om leksikografisk produksjon. Men først vil jeg nevne kort hva slags hjelpemidler vi har for å *bruke* leksikografisk materiale, hvilke ordbøker og orddatabaser vi har tilgang til, og hva slags verktøy vi har til hjelp i terminologiarbeidet. Deretter kommer jeg til å si litt om de allmennspråklige databasene vi har *lagd*, og å drøfte enkelte språklige problem vi har støtt på under arbeidet. Til slutt vil jeg komme inn på hvordan vi *utnytter* disse databasene.

1. Leksikografiske hjelpemidler

Behovet for verktøy til oversetterne¹ har ført til at det er installert maskinleselige ordbøker som en kan slå opp i på skjermen. Når en er inne i en tekst, kan en ved en enkel kommando få opp den ordboksartikkelen en ønsker i et vindu.

Det programmet som er brukt for oppslaget i figur 1, heter WordSmith². Ved hjelp av det kan medarbeiderne slå opp i en rekke maskinleselige ordbøker og ordlister:

- Webster 7. utgave (engelsk ettspråklig ordbok)
- Webster 7. utgave (etymologiske opplysninger)
- *Collins Thesaurus* (engelsk synonymordbok)
- *Roget's Thesaurus 2.* utgave
- IBM (Woods): *Dictionary of Computing*
- IBM (Quade) engelske forkortelser
- Kunnskapsforlaget: *Engelsk-norsk ordbok*
- Kunnskapsforlaget: *Norsk-engelsk ordbok*
- Kunnskapsforlaget: *Fremmedordbok*
- *Bokmålsordboka* (ettspråklig ordbok for norsk bokmål)
- Rommetveit: *På godt norsk* ("fornorskingordbok" med synonymer)
- IBM Norge: *Synonymordbok* (norsk bokmål)
- TFSNOR (engelsk-norsk dataterminologi)
- forskjellige særtermlister: SAA, OS2, AIX osv.

```

WordSmith Dictionary      ENGNOR: (c) IBM Norge AS      IBM Internal Use Only
      gnostic                      hest
      gns                          invoice
      gnu                          kort
+-----+
!
! go (v)                          !
! gå, dra, dra av sted, reise, ta (et sted
!   hen), begi seg, gå av (om skytevåpen)  !
! lyde, ringe (om klokke), slå (om ur), være i !
!   omløp (om rykte)                    !
! anses for                          !
! ha til formål, nå, føre til, finne sted  !
! lykkes, befinne seg                  !
! gå ut på, foreta, ha til hensikt       !
! ta tilflukt til                      !
! være drektig                          !
!
! go (n)                              !
! gang                                !
! hending                              !
! omstendighet, affære                 !
! siste skrik, mote                    !
! energi, mot, pågangsmot, futt        !
+-----+-----More+
go about
go about it
go about your own busi
go a boxing
DICTIONARY: ENGNOR      COMMAND: cms scrncopy
DIMENSION 1: =          DIM 2:          DIM 3: FINDLOG  DIM 4:
1=HELP 2=CMNT 3=QUIT 4=FOUND 5,6=SCRDIM 7,8=SCRBOX 9=SAVE 10=CURS 11=FLIP 12=RET

```

Figur 1. Oppslag på skjermen i Kunnskapsforlagets *Engelsk-norsk ordbok*

Disse bøkene er tilgjengelige for alle eller for visse grupper av medarbeidere - alt etter hvilke konfidensielle opplysninger de inneholder, og hvilke rettigheter IBM Norge har til bruken av hver enkelt ordbok (intern bruk i en enkelt gruppe eller avdeling, eventuelt til bruk i hele organisasjonen).

Det er svært ulike ordbøker vi kan gjøre oppslag i på denne måten, og de er av forskjellig opphav. Noen er kjøpt (for eksempel *Bokmålsordboka* og *På godt norsk*). Andre har blitt lagd i huset. Den siste gruppa omfatter både allmennspråk (synonymordbok for bokmål) og terminologi (forskjellige særtermelister og TFSNOR på rundt 15 000 oppslagsord).

Særskilt verktøy kan brukes for å skille ut og arbeide med terminologi: TIC (Terminology Identification using CRITIQUE³) går igjennom de engelske utgangstekstene, og stiller opp lister over mulig terminologi som mangler norsk oversettelse, både enkeltord og uttrykk med flere ord. TFS (Terminology Filing System) er en "arbeidsbenk" for terminologen. Begge disse programmene er samlet i ett system for styring av termarbeidet, TMS (Terminology Management System). Resultatet lagres i TFSNOR⁴. I tillegg har IBM Norge mer detaljerte databaser som en kan gjøre intelligente søk i. Ett slikt databaseformat

er LDB (Linguistic Database). Blant annet *Bokmålsordboka* finnes i LDB-format⁵. Dette er viktig i samband med utviklingen av funksjoner som virker på norsk.

2. Leksikografisk produksjon

I datalingvistisk forskning og utvikling går IBM gjerne ut i full bredde. Det nytter lite å lage system som når de store analytiske dybder, dersom de bare har et ordforråd på ett hundre ord eller to - både språkvitenskapelig sett og av kommersielle grunner. Alle IBMs funksjoner som virker på naturlige språk, har en høy dekningsgrad.

En nødvendig forutsetning og et grunnlag for utvikling av slike funksjoner er et omfattende leksikon og en fullstendig morfologi. Da IBM Norge begynte å arbeide med datalingvistikk i 1984, fantes det ikke noen maskinleselige ressurser av dette slaget for norsk som var offentlig tilgjengelige. Alt måtte lages fra grunnen av. Resultatet av arbeidet som er lagt ned siden da, er en meget omfattende dokumentasjon av normeringen av moderne norsk allmennspråk, sannsynligvis den mest omfattende som finnes.

2.1 Leksikon og morfologi

Grunnstammen i ordsamlingen er to leksikon, ett for bokmål og ett for nynorsk. Ingen av dem inneholder sideformer⁶. Til hvert leksikon er det knyttet en fullstendig og komplett morfologi; fullstendig i den forstand at hvert lemma i leksikonet er tilordnet ett bøyingsmønster i morfologien, komplett siden den dekker alle bøyingsformene til hvert lemma - også dem som vel bare er mulige reint teoretisk (se figur 2).

rik	finne
rike	finner
rikt	finnes
rikere	fant
rikest	funnet
rikeste	funnet
riks	funnets
rikes	funnets
rikts	funne
rikeres	funnes
rikestes	finnende
	finnendes
	finn

Figur 2. Alle former av adjektivet **rik** og verbet **finne**.

Vi kan når som helst kjøre ut fullstendige lister med alle former av alle ord i hvert leksikon, som i figur 3, eller for eksempel ta ut lister over alle ord som bøyes etter samme mønster som i figur 4.

Bokmålsmorfologien omfatter 629 bøyingsmønster, nynorskmorfologien 576. I dag⁷ omfatter bokmålsleksikonet om lag 79 000 lemma, nynorskleksikonet nærmere 111 000. Selve tallet på bøyingsmønster er litt høyere enn det som er lingvistisk motivert. Det kommer av at formatet i det systemet som skal ta vare på leksikon og morfologi, stiller visse særskilte tekniske krav. Likevel gir tallene en pekepinn om kompleksiteten i moderne norsk morfologi.

gjøremål	800 0	gjøremål
gjøremålet	800 0	gjøremål
gjøremål	800 0	gjøremål
gjøremåla	800 0	gjøremål
gjøremålene	800 0	gjøremål
gjøremåls	800 0	gjøremål
gjøremålets	800 0	gjøremål
gjøremåls	800 0	gjøremål
gjøremålas	800 0	gjøremål
gjøremålenes	800 0	gjøremål

Figur 3. Utsnitt av fil som viser alle former av alle ord i et leksikon. Ordformer i første kolonne, henvisning til bøyingsmønster i andre og oppslagsform i kolonnen lengst til høyre.

```
====>
!...+...1...+...2...+...3...+...4...+...>
```

* * * Top of File * * *		
beltespenne	790	00000
binyre	790	00001
åte	790	00002
nyre	790	00003
sølvspenne	790	00004
skyvelære	790	00005
spenne	790	00006
* * * End of File * * *		
		00007
		00008

Figur 4. Skjerm bilde som viser alle ord i samlingen som bøyes som nyre.

Til grunn for denne ordsamlingen ligger det et omfattende forsknings- og utredningsarbeid som også er utført av og hos IBM Norge, delvis i samarbeid med Norsk språkråd. Dette har vært et nybrottsarbeid - ikke bare fra et kvantitativt, men også fra et kvalitativt synspunkt. Ikke nok med at en omfattende maskinleselig oversikt over norske ord ikke fantes, trykte ordbøker viste seg også å være både upålitelige og ufullstendige. Vi kan snakke om både hvite flekker og gråsoner i trykte ordbøker for norsk⁸. La oss se nærmere på situasjonen for bokmålet. Først forhold som gjelder normering: Ikke noen trykt ordbok gjengav gjeldende rettskrivningsvedtak. Det har vært nødvendig å sjekke alle data mot årsmeldinger og liknende fra Norsk språkråd (i den grad det gjaldt gamle vedtak), over telefonen med medarbeiderne i Norsk språkråd (ved nyere vedtak). Men et hovedproblem er selvsagt å vite hva som er galt, og som bør kontrolleres i ordbøkene.

Det kan by på problem når en skal tolke gjeldende normal. Mens arbeidet ennå omfattet både hoved- og sideformer, var ett slikt problem hva en kunne slutte om tillatte sideformer til sammensatte ord der ett av leddene hadde en sideform. For eksempel **vinne** og **overvinne**:

vinne	vant/[vann]
overvinne	overvant/[?overvann]

Det finnes mange unormerte ord, særlig fremmedord, der vi har vage intuisjoner om bøyningsformer. Hva heter for eksempel flertall bestemt form av **mile** og **royalty**? Etter betenkningstid har Norsk språkråd kommet til at det skal hete:

mile	milen	miles	milene
royalty	royaltyen	royalties/royaltyer	royaltyene

Vi har også oppdaget en rekke forhold som ingen hadde tenkt alvorlig over, før datamaskinen krevde et nøyaktig svar. Hva heter for eksempel perfektum partisipp av **briste**? Eller presens partisipp av **be** og **bede**. Og er "beding" en form av **be**⁹? Slik de er nå, må vi regne med at samlingene er så godt som helt *riktige*.

La oss så se på det vi kan kalle deskriptive forhold; i praksis vil det gjerne si om ord har komplette bøyingsmønstre. Hver ordklasse har visse kategorier som det er vanskelig å si er komplette eller ikke ut fra semantiske forhold. Et godt eksempel er attributive former av verb. Mengden av transitive verb og verb med attributive former er ikke identiske, og i praksis er en nødt til å ta stilling til hvert enkelt verb for seg. Ofte kommer en også opp i normative problem, siden det kan være tvil om hvilket uttrykk en attributiv form skal ha. Hva heter for eksempel de attributivt bøyde formene av perfektum partisipp av **briste** og **sprekke**?

Flertall av substantiv kan være vanskelig. Vi kan stille opp en skala av substantiv, der eksemplene i ytterendene for de fleste språkbrukere enten har eller ikke har flertallsformer¹⁰.

fot	føtter
seneskjedebetennesle	(?)seneskjedebetennelser
godhet	?godheter
lykke	?lykker
rød	*røder
godfot	*godføtter

Et tredje eksempel er gradbøyning av adjektiv. Mellom

snill	snillere	snillest
og		
faglig	*fagligere	*fagligst

finnes det en rekke adjektiv der det er mer eller mindre vanskelig å godta egentlig gradbøyning¹¹. For eksempel

kompleks	(?)kompleksere	?kompleksest
smellfeit	?smellfeitere	?smellfeitest

Men kompletthet er ikke bare et semantisk problem. Et eksempel på det er genitivsformene av adjektiv. Det er tvilsomt om visse adjektiv kan forekomme i genitiv entall intetkjønn positiv, og de kan ikke opptre i genitiv superlativ sterk form entall.

rik riks rikes ?rikts rikeres *rikests rikestes

Under alle omstendigheter er det helt klart at dette er helt marginale ordformer¹², men grunnen til defekten må ha med fonotaktiske/grafotaktiske forhold å gjøre¹³.

Det er alltid en vurderingssak om en skal ta med slike former. Vanligvis lar en være å ta stilling til spørsmål av dette slaget overhodet, men når en har å gjøre med den pedanten som en datamaskin er, må en ta stilling. Og det har vi gjort. Også på dette området er sannsynligvis samlingene våre de mest pålitelige og omfattende som finnes i dag. Men det betyr naturligvis ikke at de ikke kan bli bedre.

2.2 Annen produksjon

Den mest konvensjonelle delen av den leksikografiske virksomheten er redigeringen av synonymordbøker. IBM Norge har lagd en egen synonymordbok for bokmål på rundt 17 500 oppslagsord og en for nynorsk på om lag 27 000 oppslagsord er snart ferdig. Den nynorske synonymordboka har blitt "optimalisert", det vil si at alle ord og uttrykk i definisjonsdelen av artiklene har blitt vurdert som mulige oppslagsord. Alle oppslagsord og ord i definisjonene dekkes av leksikon og morfologi i hver sin målform.

Et eget prosjekt består i å merke alle ord i leksikon etter de syntaktiske og stilistiske egenskapene deres. Dette er opplysninger som kommer til nytte i arbeidet med å lage en syntaks for norsk¹⁴. Figur 5 viser et lite utsnitt av lista over merker som vi til nå har hengt på ord alt etter de syntaktiske eller stilistiske egenskapene deres.

```
NOB      ODTABLE A1 V 80 Trunc=80 Size=431 Line=345 Col=1 Alt=0
====>
!...+...1...+...2...+...3...+...4...+...5...+...6...+...7..
```

420	0	ENEADVPS	adverb eller adjektiv som bare kan stå som ADVP aleine i en setning	00343
427	0	MAAL	ord som uttrykker mål, f eks KILOMETER, KM, GRAM	00346
428	0	MENGDE	uttrykker mengde	00348
429	0	FARGE	ord som viser til en farge, f eks GUL, HVITT, KARMOSIN	00350
430	0	TVGPRP	preposisjon i "transitiv verbalgruppe",	00352
*			"non flippable" dvs kan IKKE sprette over til høyre for	00353
*			verbet, "på" i "tenke på"	00354
*			"Han tenkte på saka."	00355
*			**Han tenkte saka på."	00356
431	0	TVGPRPF	preposisjon i "transitiv verbalgruppe"	00358
*			"flippable" dvs KAN sprette over til høyre fra verbet,	00359
*			"på" i "ta på"	00360
*			"Han tok på frakken"	00361
*			"Han tok frakken på"	00362
432	0	VPART	adverb som verbalpartikkel i løst sammensatt verb,	00364
*			"non flippable" dvs kan ikke sprette over til høyre fra	00365
*			verbet, "an" i "legge an", "over" i "ta over"	00366

Figur 5. Utsnitt av liste over merker som viser til syntaktiske eller stilistiske trekk ved ord.

2.3. Verktøy

Til hjelp i arbeidet med leksikon, morfologi og synonymordbøker har vi verktøy som kontrollerer at den syntaksen som systemet krever¹⁵, blir fulgt (jf figur 6).

```

DIAG2  MESSAGES A1 V 80 Trunc=80 Size=5 Line=0 Col=1 Alt=0
====>
!...+...1...+...2...+...3...+...4...+...5...+...6...+...7..

* * * Top of File * * *
1* ERROR-- NO MATCH FOR FORM ONE.          00000
  CLASS: 335                               00001
  BASE WORD: ettergi                       00002
  FORM ONE: e+%e                           00003
  STEM / FAULTY MATCH: ett ergi            00004
* * * End of File * * *                    00005

```

Figur 6. Feilmelding som viser at verbet **ettergi** i leksikonet ikke kan bøyes etter mønsteret 335. Denne regelen (bøyningsmønsteret) stiller formelle krav (hvilke egenskaper som de siste fire bokstavene i verbet skal ha) som sekvensen "ergi" ikke oppfyller.

For å gjøre enkeltoppslag og se hvilke grammatiske trekk som særmerker en gitt ordform i en ferdigbygd leksikonmodul, kan vi bruke editoren ODL ("Optimised Dictionary") for oppslag i leksikon). Et eksempel er vist i figur 7.

```

hvalfangstens                               OD/morphology codes
====>
!...+...1...+...2...+...3...+...4...+...5...+...6...+...7..

* * * Top of File * * *                    00000
hvalfangst(NOUN HAN ENT BST GEN)          00001
* * * End of File * * *                    00002

```

Figur 7. Editoren ODL. Morfologiske trekk ved "hvalfangstens"

ODL har også en funksjon som gjør at en kan rette, ta bort eller legge til nye trekk. For å gjøre oppslag og se hvilke spesielle syntaktiske trekk som særmerker ett gitt lemma i en ferdigbygd leksikonmodul, kan vi bruke editoren OD (Optimised Dictionary), jf figur 8.


```

ta                                     OD/grammar codes
====>
!...+...1...+...2...+...3...+...4...+...5...+...6...+...7..

* * * Top of File * * *                00000
ta(VERB V2 (TVGPRP over) (TVGPRPF på) (VPART (fram opp))) 00001
* * * End of File * * *                00002

```

Figur 8. Editoren OD. Syntaktiske trekk ved TA (ufullstendig).

På samme måte som ODL har OD også en addendafunksjon, som gjør at en kan rette, ta vekk eller legge til nye trekk, for eksempel mens en er i ferd med å skrive regler for å beskrive et syntaktisk delsystem. Det er naturligvis også mulig å ta ut lister av ord som er kjennetegnet ved samme syntaktiske eller stilistiske trekk (jf figur 9 og 10).

```

                                ODLISTQ version 1.00

For language NOB

File to use as input
====> NOB ODDATA *

Part of speech(s)
====> ADJ

Feature(s)
====> bareattr

Matching by (choose only 1):
====>          all part of speech match and all features match
====>          all part of speech match and any features match
====>          any part of speech match and all features match
====> x        any part of speech match and any features match

ENTER to begin processing   PF1 Help   PF2 Quit   PF3 Quit

CMS command

```

Figur 9. Skjerm bilde som en fyller ut for å lage en liste over alle ord av en bestemt ordklasse med samme syntaktiske eller stilistiske egenskap. Her: Adjektiv som er kjennetegnet ved merket BAREATTR (det vil si at de ikke kan stå i predikativ stilling i en setning).

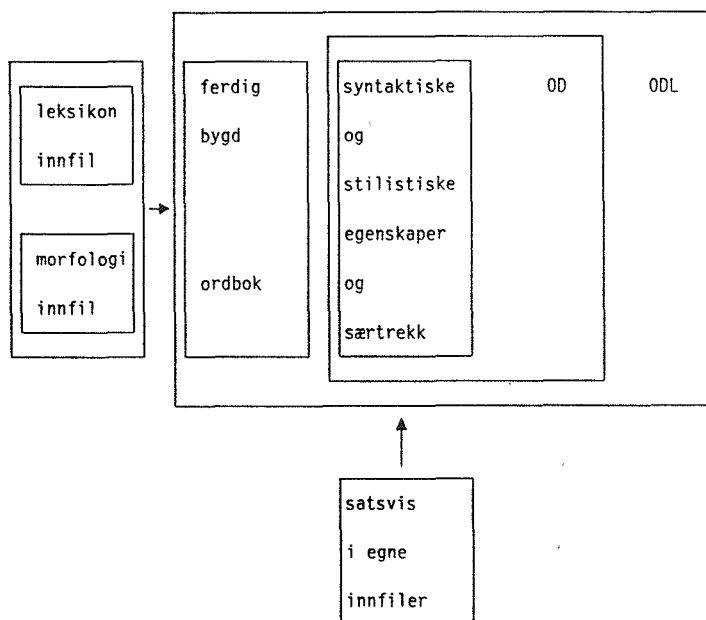
```

NOB      LISTOUT  A1  V 80  Trunc=80 Size=64 Line=3 Col=1 Alt=18
====>
!...+...1...+...2...+...3...+...4...+...5...+...6...+...7..
begge:                                       00009
  begge(ADJ BAREATTR)                       00010
beviselig:                                  00011
  beviselig(ADJ KANADV ENEADVPS BAREATTR)   00012
bevislig:                                    00013
  bevislig(ADJ KANADV ENEADVPS BAREATTR)    00014
bidig:                                       00015
  bidig(ADJ BAREATTR)                       00016
diverse:                                    00017
  diverse(ADJ BAREATTR)                    00018
formentlig:                                 00019
  formentlig(ADJ KANADV ENEADVPS BAREATTR) 00020
godeste:                                    00021
  godeste(ADJ BAREATTR)                   00022
havsens:                                    00025
  havsens(ADJ BAREATTR)                   00026

```

Figur 10. Utsnitt av lista over alle adjektiv med merket BAREATTR.

Dessuten går det an å endre og legge til data satsvis, jamfør skissen i figur 11.



Figur 11. Hvordan en i prinsippet kan legge inn data og kikke på hvilke morfologiske og syntaktiske eller stilistiske trekk som kjennetegner ord(former) i ordsamlingen.

3. Sluttproduktet

Hva brukes så de allmennspråklige ordsamlingene til? Det finnes en rekke bruksområder: Vi kan tenke oss hele program som virker på norsk språk, for eksempel program for automatisk oversettelse til eller fra en norsk målform. Men den viktigste anvendelsen er som grunnlag for *funksjoner* som virker på norsk språk, altså som deler av andre brukerprogram. Og selv om det finnes språklige funksjoner i for eksempel søkesystem, så er anvendelsesområdet for ordsamlingene først og fremst IBMs egne tekstbehandlingsprogram¹⁶. Leksikon og morfologi danner blant annet basis for funksjonen *bøyningsopplysninger* og for grunnordlister som igjen styrer stavekontroll og automatisk orddeling av kjente ord.

```

VISFRAM RFT      A1                      STAVING                      Side 1
** Ordet er feilstavet
==>
<---+---1---+---2---+---3---+---V---+---5---+---6---+---7---+--->
Har du en leiekontrakt som ikke er tidsbegrenset (vanlig leiekontrakt),
og du blir sagt opp, kan du få oppsigelsen vurdert av domstolene.
Hovedregelen er at domstolene kan sette oppsigelsen til side hvis den
mangler saklig grunn, eller vil virke urimelig overfor leietakeren.
Det er den begrunnelsen utleieren gir for oppsigelsen, som blir vurdert
av rette +-----+
      Hvi ! begrunnelsen ! ppsigelsen er urimelig overfor leietakeren,
kan rett ! begrunnelse ! elsen er ugyldig , selv om det er saklig
grunn fo ! begrunnelsene ! opp fra leieavtalen og mener at oppsigelsen
ikke fyl ! grunnens ! gyldighet, må du sende stevning til retten
innen 30 ! grunnens ! ppsigelsen.
+-----+

----- Dokumentslutt -----

PF 1=Hjelp      2=Innskyt      3=Avslutt      4=Koder      5=GSøk      6=Sjekk
PF 7=Tilbake    8=Frem          9=Blokk       10=Bøyning  11=Neste    12=Synonymer

```

Figur 12. Eksempel på bruk av stavehjelp i DW370. Forslag til hva **begrunnelsen** skal være.

```

VISFRAM RFT      A1                      STAVING                      Side 1
** Velg hvilket ord som skal utvides og trykk på ENTER
==>
<---+---1---+---2---+---3---+---V---+---5---+---6---+---7---+--->
Har du en leiekontrakt som ikke er tidsbegrenset (vanlig leiekontrakt),
og du blir sagt opp, kan du få oppsigelsen vurdert av domstolene.
Hovedregelen er at domsto +-----+ sigelsen til side hvis den
mangler saklig grunn, ell ! Verb: få ! lig overfor leietakeren.
Det er den begrunnelsen u ! Adjektiv: få ! ppsigelsen, som blir vurdert
av retten. +-----+
      Hvis retten finner at oppsigelsen er urimelig overfor leietakeren,
kan retten avgjøre at oppsigelsen er ugyldig , selv om det er saklig
grunn for den. Blir du sagt opp fra leieavtalen og mener at oppsigelsen
ikke fyller lovens krav til gyldighet, må du sende stevning til retten
innen 30 dager fra du fikk oppsigelsen.

----- Dokumentslutt -----

PF 1=Hjelp      2=Innskyt      3=Avslutt      4=Koder      5=GSøk      6=Sjekk
PF 7=Tilbake    8=Frem          9=Blokk       10=Bøyning  11=Neste    12=Synonymer

```

Figur 13. Eksempel på bruk av funksjonen bøyning i DW370 (1). Ordformen **få** faller inn under to oppslagsord som hører til to ulike ordklasser.

VISFRAM	RFT	A1	STAVING			Side 1 Linje 0	
====>							
<---+---	1---	+---2---	+---3---	+---V---	+---5---	+---6---	+---7---
Har du en leiekontrakt som ikke er tidsbegrenset (vanlig leiekontrakt), og du blir sagt opp, kan du få oppsigelsen vurdert av domstolene. Hovedregelen er at domsto							
mangler saklig grunn, ell ! Bøyning av ordet "få": !							
Det er den begrunnelsen u ! Verb: !							
av retten. ! aktiv !							
Hvis retten finner a ! infinitiv få !							
kan retten avgjøre at opp ! presens får !							
grunn for den. Blir du sa ! preteritum fikk !							
ikke fyller lovens krav t ! perfektum partisipp !							
innen 30 dager fra du fik ! i perfektum fått !							
! samsvar !							
! intetkjønn fått !							
----- ! han- el. hunkjønn fått !							
! flertall el. svak bøyn. fåtte !							
! presens partisipp fående !							
! S-former !							
! infinitiv fås, fåes !							
! presens fås, fåes !							
! presens partisipp fåendes !							
! imperativ få !							
+-----+							
PF 1=Hjelp	2=Innskyt	3=Avslutt	4=Koder	5=GSøk	6=Sjekk		
PF 7=Tilbake	8=Frem	9=Blokk	10=Bøyning	11=Neste	12=Synonymer		

Figur 14. Eksempel på bruk av funksjonen bøyning i DW370 (2). Bøyningsmønsteret for verbet FÅ som ordformen få tilhører.

```

DEMON   RFT    A1                               STAVING                     Side 1
                                                Linje 2

====>
<---+---1---+---2---+---3---+---V---+---5---+---6---+---7---+-->
og du blir sagt opp, kan du få oppsigelsen vurdert av domstolene.
Hovedregelen er at domstolene kan sette oppsigelsen til side hvis den
mangler +-----+-----+-----+-----+-----+-----+-----+-----+
Det er ! Synonymer for ordet "sette":                                !
av rett ! Verb:                                                       !
      Hv ! - stille, plassere                                        !
kan ret ! - plante                                                  !
grunn f ! - satse, vedde                                           !
ikke fy ! - bringe, føre                                           !
innen 3 ! - lage, danne                                             !
      ! - avtale, beramme                                           !
Vi gjer ! <seg>                                                       !
derfor ! - ta plass                                                 !
at du s ! - si sammen                                               !
halde t ! - feste seg                                               !
set fra ! <etter>                                                    !
busett ! - forfølge                                                 !
gjerast ! <i gang>                                                    !
      ! - begynne, starte, gå i gang, sette i, klemme på, kjøre i vei !
Este ar ! - iverksette, sette i verk, sette ut i livet              !
entre e ! - få til å starte                                         !
de Madr ! <til>                                                       !
un dicc ! - omkomme                                                 !
informa ! <inn>                                                       !
verific ! - gni inn, stryke på                                       !
como he ! - si på, senke seg, falle på                               !
PF 1=Hj ! <seg inn i>                                                !
PF 7=Ti +-----+-----+-----+-----+-----+-----+-----+-----+

```

Figur 15. Eksempel på oppslag i synonymordboka (bokmål) i DW370.

Forskjellige bruksområder stiller ulike krav til det språklige grunnlaget, og dette bringer oss tilbake til leksikografisk produksjon. Ideelt sett er en nødt til å lage egne ordlister til ulike behov. I stedet for å bruke en og samme database må en gjøre utvalg til særskilt bruk. Til grunn for en stavekontroll trenger en for eksempel en liste av bare riktige ordformer. I et system med en språklig komponent for søking i tekst, derimot, trenger en også å ta med visse uriktige ordformer i lista. Det kan dreie seg om tidligere eller regionale varianter. Dette er spesielt viktig når man har med et språk som norsk å gjøre, der rettskrivningsendringene har vært hyppigere og mer radikale enn i de fleste språksamfunn. Et eksempel som viser de forskjellige kravene som ulike brukerprogram krever, er ETTER, som er riktig, og som må være den eneste formen som en stavekontroll bygger på, og EFTER, som også må være med i forbindelse med et søkesystem. Hvis ikke, får en ikke med både "etter" og "efter" i eldre tekster eller tekster som ikke følger bokmålsnormen. Søket blir ikke fullstendig, og kvaliteten på programmet dårlig. Akkurat som når en bygger opp allmennspråklige databaser, er tradisjonelt leksikografisk arbeid en forutsetning for utvikling av språklige funksjoner i brukerprogram.

Noter

1. Opptil 40 oversettere kan være i arbeid på samme tid. I IBM Norge blir det oversatt store mengder av tekniske tekster fra engelsk til norsk: I 1990 hele 4 501 sider med trykt dokumentasjon og 256 362 linjer med maskinkode, "machine-readable information". Altså noe rundt 9 000 sider i alt.
2. Det er et internt systemprodukt som virker på stormaskiner under operativsystemet VM.
3. CRITIQUE er et program for grammatisk og stilistisk analyse av engelsk.
4. Terminologen har også mulighet for å slå opp i IBMs termdatabaser for andre språk, f eks dansk og svensk.
5. Se Baustad 1992.
6. Tidligere opererte vi med sideformer for bokmål, men det har vi sluttet med av praktiske grunner (for å kunne behandle flere lemma som virkelig er forskjellige i stedet), dels av kommersielle grunner: Ordsamlingene er jo utgangspunkt for språklige funksjoner i programvare, og kundene ønsket funksjoner som bygde på læreboknormalen, altså uten sideformer.
7. 30.05.1991.
8. Jf Akø 1992.
9. Dette har vi ryddet opp i med velvillig hjelp fra Norsk språkråd.
10. Og tilsvarende for manglende entallsformer av mange substantiv, naturligvis.
11. Her ser jeg bort fra gradbøyning med "mer" og "mest", som er uten interesse i denne sammenhengen.
12. Jf figur 2.
13. Jf adjektiv som nøye, der en form genitiv entall intetkjønn positiv, "nøyenes", i og for seg er fullt akseptabel.
14. Syntaksen er i sin tur viktig for eksempel for utviklingen av system for automatisk oversettelse med en norsk målform som utgangsspråk og/eller målspråk.
15. Altså ikke syntaks i språkvitenskapelig forstand.
16. "Skriveassistent", DisplayWrite 5 (DW5) for PS-verdenen, DisplayWrite 400 (DW400) for de mellomstore maskinene og DisplayWrite 370 (DW370) for stormaskinene.

Litteratur

- Akø, J.O. 1992. Gråsoner i norske ordbøker. I: Fjeld, R.V. (red.). 1992. *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden, 28.-31. mai 1991*. Oslo.
- Baustad, J. 1992. Automatisk analyse av maskinleselige ordbøker til bruk i en orddatabase. I: Fjeld, R.V. (red.). 1992. *Nordiske studier i leksikografi. Rapport fra Nordisk konferanse i leksikografi i Norden, 28.-31. mai 1991*. Oslo.