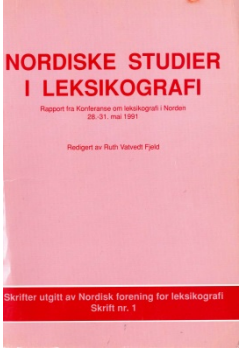


# NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Redigeringsprogrammet EDI	
Forfatter:	Rik Schutz	
Kilde:	Nordiske Studier i Leksikografi 1, 1992, s. 441-446 Rapport fra Konferanse om leksikografi i Norden, 28.-31. mai 1991	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive">http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive</a>	

© Nordisk forening for leksikografi

## Betingelser for brug 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.

Rik Schutz

## Redigeringsprogrammet EDI

Programvaren EDI er spesielt laget for redigering av ordbøker. Jeg vil ta utgangspunkt i Van Dales svensk-nederlandske ordbok for å forklare hvordan programmet arbeider. Deretter vil jeg sammenligne ordboksartiklenes struktur og innhold med den redaktørene tidligere måtte arbeide med. Til slutt vil jeg vise hvordan en ferdig ordboksartikkel ser ut når redaktøren arbeider med EDI.

### Tradisjonell ordboksredigering

Til tross for de mange fordelene med den tradisjonelle redigeringsmetoden, er det store ulemper med den. Det største problemet er systemets kompleksitet og den uforholdsmessig store oppmerksomheten en redaktør/oversetter må vie nokså tilfeldige overenskomster. For eksempel kan punktum eller ikke punktum etter en forkortelse, eller bruk av vanlig parentes i stedet for hakeparentes, gi alvorlige feil. En redaktør som må konsentrere seg om slike ting, kan i mindre grad konsentrere seg om det leksikografiske oversettelsesarbeidet. Det tar svært lang tid å lære konvensjonene. En annen ulempe er risikoen for at tekststykker utilsiktet blir ødelagt eller helt mistes på grunn av uerfarenhet eller skjødesløshet fra medarbeiderens side.

Et problem som fremfor alt forleggeren føler, er tidstapet som etterkontrollen medfører. Tidsmessig går det mer enn et halvt år mellom fullføringen av manuskriptet og den endelige utgivelsen av boken. Mer enn halvparten av denne tiden går med til kontroll og korrigering av manuskriptets struktur.

### EDI

Det er derfor vi har forsøkt å utarbeide en arbeidsmåte som ikke har disse ulempene. Redaktøren må kunne konsentrere seg om sitt egentlige arbeid. Kunnskap om de forkortete navnene på byggestenene og de nøyaktige reglene for rekkefølgen av dem, ligger i programmet. Programmet inviterer til å arbeide i riktig rekkefølge og gir beskjed hvis noe galt hender.

Programmet er i stand til å lese tekster som har den tradisjonelle Van Dale-strukturen, slik at disse kan behandles med EDI. EDIs lagringsstruktur har karakter av en relasjonell database. En direkte kopling mellom EDI og fotosettemaskinen er for øvrig tenkelig, men den brukes ikke ennå.

EDI fastlegger dessuten av hvem og til hvilket tidspunkt det er arbeidet med en fil, noe som har klare administrative fordeler. Det gjør det også enklere å gi feedback.

## Oppbygningen av en artikkel

Vi skiller mellom oppslagsord på grunnlag av deres form, altså på grunnlag av staving. Dette kaller vi det høyeste eller det første nivå. Ord som har felles form, såkalte homografer, kan i flere henseender skille seg fra hverandre. Avhengig av disse forskjellene velges en bestemt typografisk gjengivelse i ordboken. Denne gjengivelsen er nokså tilfeldig. Vi gjør dette slik:

- Hvis ordklassen (grammatisk kategori) er forskjellig, blir homografer trykt med nummer, slik som for eksempel hetta som henholdsvis substantiv og verb (jf figur 1).
- Hvis den grammatisk subkategorien er forskjellig (kjønn ved substantiv, eller transitive versus intransitive verb), blir en inndeling med romertall brukt, slik som for eksempel ved hetta som verb.

Det at Esseltes svensk-engelsk ordbok, som vår baserer seg på med hensyn til den svenske delen, gjør dette helt motsatt, viser hvor vilkårlig gjengivelsen av denne type forhold er. For EDI er det nivået som ordklasseinformasjonen legges på (såvel kategori som subkategori) viktig; vi kaller det for det andre nivå, eller for lemma-nivå.

På dette nivået kan det også angis markeringer av typen fagterm-etiketter, anvisninger av stilregister og morfologiske særegenheter som for eksempel verbbøyning.

Det tredje nivå er nivået der betydningsforskjeller markeres. Jeg vil her ikke gå inn på de spesielle problemene som oppstår i den forbindelse.

For hver betydning kan det gis eksempler på kontekster som oppslagsordet forekommer i. Den vanlige termen for dette er eksempel. Ofte dreier det seg om kollokasjoner og om idiomer. Dette kaller vi det fjerde nivå. I de fleste av Van Dales ordbøker er det fjerde nivå også grafisk skilt fra det tredje. Alle nummererte betydninger av oppslagsordet følger umiddelbart etter hverandre, og først deretter, etter et iøynefallende tegn (en ruter), kommer eksemplene. Men dette systemet er ikke relevant for EDI, og jeg går derfor ikke nærmere inn på det her.

## Byggestener

På hvert av disse fire nivåene forekommer i nesten alle artiklene tilbakevendende serier av byggestener. Hver av disse byggestenene har et navn forkortet til fem bokstaver. For en nederlender er denne forkortelsen tilstrekkelig til at man kan avlede ordets betydning av den.

Eksempler:

TREFW	= TREFWOORD	= oppslagsord
STILA	= STIJLLABEL	= stilistisk etikett
HVERT	= HOOFDVERTALING	= hovedoversettelse

Det er fastlagt i detaljerte skjemaer på hvilke måter disse byggestenene kan sammenføres til en artikkel (se figur 2). Av dette skjema kan man se at bestemte elementer er obligatoriske (de liggende blokkene), mens andre er valgfrie (de stående blokkene med stiplet linje), og at rekkefølgen elementene forekommer i, er fastlagt. Det er også angitt hvilke elementer som kan forekomme mer enn en gang.

Den leksikalske informasjonens lagringsform er en enkel, sekvensiell ASCII-tekst. På basis av rekkefølgen av og navnene på byggestenene angir settedatamaskinen de korrekte typografiske verdiene. Oppslagsord blir automatisk uthevet, og tegn som komma og parentes plasseres automatisk. Før denne transformasjonen kan finne sted i fotosettemaskinen, må manuskriptet være 100% rent. Kontrollen av at det formelle i teksten er helt riktig (for eksempel om betydningene står i riktig rekkefølge, og om det er gitt eksempler på ikke eksisterende betydninger), skjer med spesielle kontrollprogram. Det første lemmaet hetta ser slik ut i denne strukturen:

```

<TREFW.>  hetta
<GRAMT.>  m/v
<GRMAZ.>  ~n
<NULCO.>  0.1
<RESUM.>  sterk v\ "arme
<HVERT.>  hitte
<NULCO.>  0.2
<LABEL.>  fig.
<HVERT.>  vuur
<HVERT.>  enthousiasme
<CIJCO.>  1.1
<ILLUS.>  @solens ~
<VVERT.>  de hitte van de zon
<OMKWD.>  hitte
<CIJCO.>  1.2
<VASTV.>  i @stridens ~
<VVERT.>  in het vuur van de strijd
<OMKWD.>  vuur

```

Den skisserte arbeidsmåten har flere fordeler. Lagringen i tekstform gjør desentralisert bearbeiding enklere. Mer enn førti redaktører fra tre forskjellige land har samtidig arbeidet med en spansk-nederlandsk ordbok, hver på sin PC. Kontrollen etterpå skjedde sentralt.

### Virkemåte

De fire nivåene i artikkelen som jeg nettopp nevnte, ser i EDI ut som i figur 3. Oppe til venstre er det første nivå gjengitt, altså listen over oppslagsord. Deretter følger lemma-nivået, altså de grammatikalske kategoriene. Til venstre, i andre blokk ovenfra, står betydningene med oversettelsene i blokken ved siden av. Under disse til venstre står eksemplene, og til høyre ved siden av oversettelsene.

Når vi starter EDI, ser vi det første ordet fra filen vi arbeider i. Det er med andre ord A hvis vi behandler hele ordboken under ett, eller et annet oppslagsord hvis vi arbeider med mindre omfangsrike filer. I den svensk-nederlandske ordboken disponerer vi nå over hele bokstaven H, og alle eksemplene i denne sammenhengen er dermed tatt derfra. Vi kan søke etter ordet vi vil se på to måter: gjennomgå listen med piltastene eller med søkefunksjonen. Det siste virker absolutt raskere. Hvis vi nå har søkt oppslagsordet hem, ser vi i blokken oppe til venstre en del av oppslagsordlisten som vårt oppslagsord er markert i. I figuren er markeringen gjengitt med understreking av det valgte elementet. Vi ser at det finnes to

grammatikalske kategorier av ordet vi har søkt fram. Adverbet, på samme måte markert med understreking, har tydeligvis bare en betydning. Til betydningen hører en oversettelse og de fem eksemplene fra blokken under til venstre.

Redigeringen av en artikkel foregår på følgende måte: Redaktøren velger nivået han vil arbeide på og flytter markøren dit. Dette kan enten gjøres ved å gå til ønsket nivå (blokk) med piltastene, eller ved å taste inn tallet som svarer til blokken. Ved å taste inn 5 går markøren rett til den femte blokken, altså den med setningseksempler. Informasjonen i en slik blokk kan forandres ved at man taster inn bokstaven E for EDIT, eller ved å trykke på ENTER-knappen.

### Tilføyning av eksempler

La oss gå ut fra at vi vil føye et setningseksempel til den første betydningen for hem som substantiv: nemlig "ett andra hem". Vi sørger for at markøren befinner seg ved den første kategori (substantiv) i den andre blokken (grammatisk kategori) og ved den første betydningen i den tredje blokken (betydninger). I feltet nedenfor vises de tidligere redigerte eksemplene for denne betydningen. Ved å gå til dette feltet og taste inn T for Tilføy, kommer det fram en tom redigerings skjerm nederst, som har skjema-layout. Samtidig vises alle ulike former for informasjon som kan forekomme her. Der skriver jeg eksemplet (se figur 4). Hvis et bestemt felt ikke er aktuelt, for eksempel fordi oppslagsordets ordklasse utelukker det, angir EDI dette. I de feltene hvor det kun er tillatt å fylle ut ett element fra en avgrenset liste, en forkortelse eller en standardisert kode som angivelse av flertall, viser EDI de tillatte mulighetene nederst, så snart markøren befinner seg i det aktuelle feltet.

Jeg kan for eksempel velge mellom fire slags eksempler: 1 idiom, 2 illustrerende eksempler, 3 kollokasjoner (Vaste Verbindingen på Nederlandsk) og 4 grammatikalske kollokasjoner. Etter at jeg har valgt en av dem, kan jeg velge riktige eksempler ved å oppgi tallet. Det er selvsagt også mulig å taste inn den riktige forkortelsen. Hvis jeg taster inn en markering som ikke finnes, nekter EDI å forlate dette feltet. Etter at hele skjemaet er fylt ut, kan jeg gå tilbake til et av de andre nivåene.

Hvis jeg ønsker en oversikt over hele artikkelen, kan EDI vise lemmaet i ordbok-layout ved hjelp av ett enkelt tastetrykk (V for Vis). Da får jeg en oversikt over alle betydninger og alle eksempler, sammen med de tilhørende oversettelsene.

EDI kontrollerer inndata ved forskjellige handlinger. I bestemte felt aksepterer programmet kun det foreskrevne innholdet, slik som for eksempel ved ordklasser og markeringer. Når et skjema er utfyllt, kontrollerer EDI om innholdet er korrekt. Og når artikkelen forlates, altså ved neste oppslagsord, sjekker EDI hele artikkelens struktur.

Etter redaksjonens ønske er det mulig å velge å forlate en artikkel før den er perfekt. EDI kan gjenkjenne og registrere tyve forskjellige slags feil. Når redaktøren forlater en artikkel spør EDI om han er klar over at det registreres en feil, for eksempel at det ikke er fylt ut noen oversettelse. Hvis redaktøren svarer bekræftende, skriver EDI en kode foran oppslagsordet i oppslagsordlisten, som angir den alvorligste feilen. Etter at feilen er løst av redaksjonen, blir feilmeldingen erstattet av en melding om den neste store feilen. Først etter at hele lemmaet er strukturelt korrekt, vises oppslagsordet uten koder på skjermen. Det er selvfølgelig mulig å lete etter en spesiell type feil, slik at en redaktør kan konsentrere seg om å løse liknende redaksjonelle problemer.

## Man ser hva man gjør

Etter min mening er det på en måte tatt et skritt tilbake, men ikke i negativ mening; det er et skritt bakover som likevel fører til fremgang. I urtiden, for ca. ti, femten år siden, arbeidet redaktørene med blanke papirark når de skulle lage en ny ordbok, eller med forstørrede kopier av en eksisterende ordbok. Foran seg kunne de se nøyaktig hvordan manuskriptet ville se ut i satsen, og de skrev den ønskede eller forbedrede teksten, med alle typografiske særegenheter.

Forlagenes ønske om å kunne behandle manuskriptet på datamaskin, har i løpet av siste tiår ført til at våre redaktører har måttet arbeide med et manuskript som ikke lenger så ut som selve ordboken, og som omfattet mange forstyrrende elementer. Diakritiske tegn, slik som ä i värme, måtte koderes og de typografiske særegenhetene ble eliminert i manuskriptet. Etter at manuskriptet var ferdig, kunne forleggeren nyttiggjøre seg kontrollmulighetene som en datamaskin har, og dermed kunne satsen lages billigere enn før; ved å bruke en fotosettemaskin i stedet for å sette i bly. Siden manuskriptet kun inneholdt koder og ikke typografisk informasjon, kunne typografien, og dermed også omfanget, fastslås først i siste øyeblikk.

Tilbakegangen med EDI er at nå kan redaktørene på nytt arbeide på den tradisjonelle måten med en artikkel som viser verken mer eller mindre enn det de fører inn i den. De diakritiske tegnene gjengis slik de skal være, og kodene er ikke lenger synlige. Fremskrittet er at forleggeren fremdeles kan nyttiggjøre seg alle fordelene med et maskinleselig manuskript og i tillegg har garanti for at strukturen i manuskript er riktig.

## Sluttord

EDI har mange funksjoner og finesser for praktisk ordboksarbeid, og jeg håper med disse eksemplene å ha videreført noe av den entusiasmen som vi og våre redaktører har for dette verktøyet.

Til slutt vil jeg kort nevne de gunstige egenskapene ved programmet:

- \* Leksikografen trenger ikke å ha detaljkunnskaper om det valgte systemet.
- \* Leksikografen trenger ikke å huske symboler eller koder; mulighetene vises stadig i en hjelpeskjerm.
- \* Redaktøren ser de spesielle tegnene slik som de ser ut i satsen, selv om de blir lagret med en koding som utelukkende arbeider med standard ASCII-tegn.
- \* Programmet aksepterer ikke strukturelle feil.
- \* EDI aksepterer imidlertid et manuskript som ikke er komplett, men registrerer hva som er feil, slik at det på grunnlag av meldingene raskt kan utføres forbedringer eller tilføyelser på et senere tidspunkt.
- \* Programmet muliggjør svært hurtig søking i andre artikler, og fremmer dermed redaksjonell uniformitet i semantisk sammenhengende artikler som ikke følger hverandre alfabetisk.
- \* EDI gjør det enkelt å arbeide tematisk. Det er svært enkelt å søke på kjennetegn som ordklasse, en bestemt type markering, om det i det hele tatt forekommer visse elementer (en bestemt kode for fagområde) i en artikkel.

- \* En satssimulator kan til enhver tid gi redaktøren en oversikt over artikkelen i samme layout som i den endelige ordboken. Denne satssimulatoren er i stand til å vise den endelige bokens totale sideomfang. Når det gjelder et nytt produkt, er det svært viktig å vite omfanget på den ferdige delen av manuskriptet.

### Litteratur

- Stora svensk-engelska ordboken*. 1988. V. Petti and I. Rider (eds). Solna.  
*Van Dale handwoordenboek nederländsk-svensk/Nederlands-Zweeds*. Under utgivelse. G.L.M. Laureys & J. de Groot et al. Utrecht.  
*Van Dale handwoordenboek svensk-nederländsk/Zweeds-Nederlands*. Under utgivelse. G.L.M. Laureys & J. de Groot et al. Utrecht  
*Van Dale handwoordenboek Nederlands-Spaans*. 1992. P.J. Slagter et al. Utrecht.  
*Van Dale handwoordenboek Spaans-Nederlands*. 1992. P.J. Slagter et al. Utrecht