

Enspråklig ordbok for fremmedspråklige

Øivin Andersen

In this paper I describe some basic principles of editing the monolingual dictionary for refugees and immigrants in Norway (*Norsk ordbok for fremmedspråklige*). The basic structure of a dictionary entry is illustrated, both as a working document and a finished product. The application of prototype theory and lexical field theory are discussed. Basic principles of pronunciation coding and coding of grammatical information are also described.

1. Bakgrunn

Utgangspunktet for denne enspråklige ordboken, *Norsk ordbok for fremmedspråklige*¹, er den norsk-vietnamesiske ordboken (heretter NVO) som kom ut på Universitetsforlaget i 1983 (Khánh & Tràn 1983). Arbeidet ble påbegynt i 1979. Det mest spesielle med denne ordboken er dens skreddersydde tilpasning til de vietnamesiske flyktningenes behov i Norge. Vietnameserne selv tok initiativet til ordboken, men innså raskt at de trengte hjelp. Universitetet i Bergen bevilget driftsmidler til prosjektet og det ble etablert en arbeidsgruppe bestående av norske og vietnamesiske fagfolk.

Det første problemet i NVO-prosjektet var å etablere et ordtilfang tilpasset vietnamesernes behov. Et frekvensbasert ordutvalg ble presentert for lærere i norsk som andrespråk og fremmedspråk hvor de ble bedt om å krysse av for viktige ord og supplere i ordlisten. Ord fra ulike innvandrerbrosjyrer og lærebøker i norsk for innvandrere ble ekserpert og innlemmet i listen. Det ble også arbeidet terminologisk med fagtermer av særlig betydning for flyktninger og innvandrere. Det ble særlig fokusert på områdene medisin og jus. På disse to områdene ble det nedsatt oversettelsesgrupper bestående av sakkyndige personer. Norske medisinstudenter arbeidet sammen med vietnamesiske leger som i hovedsak var utdannet i Frankrike. Tilsvarende gruppe ble også etablert innen fagområdet jus.

For hele det norske ordtilfanget ble det laget synonymer og definisjoner som i etterkant ble oversatt til vietnamesisk. Oversettelsesprosessen ble utført ved etablering av norske forklaringer og definisjoner av oppslagsord og underoppslag, inkludert idiomatiske konstruksjoner. Disse ble diskutert med våre vietnamesiske medarbeidere med henblikk på vietnamesiske nærekvivalenter. Det er først og fremst dette som dannet utgangspunktet for videreføring av den norske delen

¹ Ordboken er fremdeles under utarbeidelse i skivende stund.

av NVO til en enspråklig ordbok for fremmedspråklige som er bosatt eller skal bosettes i Norge generelt. Dette gjelder særlig flyktninger som har fått permanent opphold i Norge.

Disse personene får et obligatorisk norskkurs og det blir stilt tøffe krav til norskkunnskaper når de skal integreres i det norske arbeidslivet. Derfor har de behov for en tilpasset ordbok både i kursopplæringen og senere ved den sosiolekttilpasningen som skjer når de blir plassert i forskjellige deler av Norge. Dette innebærer også en integrering i norsk kultur- og arbeidsliv.

Dette arbeidet har foregått siden siste halvdel av 80-tallet med diverse avbrudd, men fra 2011 ble revisjonsarbeidet og oppdateringsarbeidet med ordboken tatt opp og har vært drevet kontinuerlig frem til i dag. I det følgende skal jeg redegjøre for dette redigeringsarbeidet.

2. Prinsipper for redigering

Først vil jeg illustrere ordboksartikkelens oppbygging som arbeidsdokument. Som illustrasjon brukes substantivet **hagl**.

```
S HAGL
100 hagl
120 h_agel
140 s.n.
150 (hagl^et, -, -a/-ene)
312 nedbør som består av kompakte, runde iskorn
000 312 skal koordineres for alle ordene REGN, YR, SNØ,
HAGL, SLUDD. Dette er gjort per 6.7.15.
314 Haglene slo blomstene ned.
314 I Syden kan hagl være store som fugleegg.
316 hagl+byge s.fm.
322 liten blykule som brukes i ladning for haglgevær
324 Man skyter fugl og hare med hagl.
327 hagle s.fm.
327:haglgevær
327 hagl+gevær s.n.
327:våpen som lades med hagl
```

Figur 1: Ordboksartikkel som arbeidsdokument.

I arbeidsdokumentet i figur 1 har hver informasjonsdel sin egen separate tallkode: 100 er oppslagsordet, 120 er uttaleangivelse, 140 er ordklasseinformasjon, 150 er bøyningsparadigmet, 312 er oppslagets leksikalske definisjon for betydning 1. Koden 000 er reservert for interne kommentarer for ordboksmedarbeiderne. 314 er reservert for eksempelsetninger og fraser som ikke trenger videre forklaring. I

figur 1 er det tatt med to 314 eksempelsetninger. 315 brukes til ordtak (ordtak er ikke representert i oppslaget **hagl**) og idiomatiske uttrykk som er ikke-komposisjonelle, og som derfor forklares. Eksempel er oppslag som **snø**: 315: *snøen som falt i fjor*: 'noe som tilhører fortiden; noe som er avsluttet/uaktuelt/glemt'. 316 brukes for komposisjonelle sammensetninger som angis uten forklaring, mens 317 brukes for ikke-komposisjonelle sammensetninger med forklaring. 317 brukes også for å angi underoppslag som er avledninger dannet på hovedoppslaget, med forklaring. 322 angir oppslaget betydning 2, og 324, 325, 326, 327 har samme funksjon for betydning 2 som tilsvarende felt for betydning 1.

Hvis dette gjøres på en entydig og konsistent måte skal det ferdige produktet kunne genereres maskinelt som vist i figur 2.

<p>hagl [hagel] s.n.(hagl et, -, -a/-ene)</p> <ol style="list-style-type: none"> nedbør som består av kompakte, runde iskorn <i>Haglene slo blomstene ned. I Syden kan hagl være store som fugleegg.</i> hagl+bygge s.fm. liten blykule som brukes i ladning for haglgevær <i>Man skyter fugl og hare med hagl.</i> hagle s.f.m. haglgevær hagl+gevær s.n. våpen som lades med hagl hagle s.fm. haglgevær hag l+gevær s.n. våpen som lades med hagl
--

Figur 2: Ordboksartikkel som ferdig dokument.

2.1. Bruk av prototypeteori

I den internasjonale litteraturen har man diskutert bruken av prototypeteori i pedagogiske ordbøker (for en oversikt se Jiang & Chen 2015). Potensielt kan man tenke seg flere områder i ordartikkelens oppbygging hvor dette kan anvendes. Her skal jeg drøfte et sentralt anvendelsesområde: betydningsrekkefølge ved polysemi. Som illustrasjonseksempel vises substantivet *land*. Oppslaget har fire betydninger:

- del av jorden som ikke er hav/vann; fastland; jord
Den som har reist lenge til sjøs, synes det er godt å kjenne land under føttene igjen.
- område som utgjør en politisk og oftest geografisk enhet
Norden består av fem land: Danmark, Finland, Island, Norge og Sverige.
- stykke jord, særlig område som skal brukes til dyrking eller bygging
Staten kjøpte opp land til å bygge ny flyplass.
- (til forskjell fra 'by') område med spredt bebyggelse
Mange nordmenn har leilighet i byen og hytte på landet.

Intuitivt vil betydning 1 være den mest generelle og grunnleggende, og derfor klar kandidat for å komme først. Betydning 2 og 3 har et element av institusjonalitet i tillegg til det rent geografiske og betydning 4 er mer spesialisert enn 1. En slik rekkefølge gjør ordartikkelen mer motivert for brukerne. Men hvis betydning 1 er mest prototypisk vil man forvente at den også er den mest frekvente. For å sjekke dette brukte jeg det norske aviskorpuset. Dette korpuset er et monitorkorpus (McEnery & Wilson 2001:30f). Det går tilbake til 1998 og drives kontinuerlig av språk og språkteknologi-gruppen ved Uni Research i Bergen. Prosjektet heter Corpuscle. De tekniske redskapene er utviklet ved Clarinosenteret i Bergen (clarino.uib.no/korpuskel/page).

I desember 2019 hadde dette korpuset over 1,88 milliarder løpende ord-eksemplarer. Ordet *land* forekommer 546 203 ganger i korpuset. Jeg analyserte de 180 første forekomstene i kwik-konkordanser med hensyn til betydningsfordeling. Tidligere undersøkelser har vist at en manuell analyse av de første 180–200 forekomstene gir en tilstrekkelig pekepinn for de øvrige forekomstene, særlig ved høyfrekvente fenomener (cf. Andersen 2007). Betydning 1 forekom bare 13 ganger, betydning 2 forekom hele 161 ganger. Betydning 3 hadde 6 forekomster mens betydning 4 hadde 0 forekomster. En ren frekvensanalyse skulle derfor tilsi at betydning 2 skulle ha kommet først.

Frekvens og intuitiv prototypanalyse motsier hverandre, og det fins ingen enkle regler for hvordan man skal løse problemet.

Jiang & Chen (2015) undersøkte fem store enspråklige pedagogiske ordbøker for engelsk, blant annet med henblikk på struktureringen av prototypiske og ikke-prototypiske betydninger av oppslag. De kom frem til at prototyppløsninger kun ble sporadisk og intuitivt brukt. Forfatterne påpeker at årsaken til dette er at prototypeteorien mangler klare retningslinjer for leksikografisk anvendelse. Enda mer problematisk blir det å velge et prototypisk eksempel på bruk.

For fremmedspråklige innlærere vil sentral encyklopedisk informasjon kanskje være mer nyttig, som i betydning 2 under oppslaget **land** (se ovenfor). Selv når det er klart hva som er grunnbetydningen kan det være problematisk å rangere de andre variantene fordi de går i forskjellige retninger ut fra det sentrale. Under oppslaget **land** er det vanskelig å si om 2 eller 3 bør komme nærmest etter 1. De er omtrent like nær, men går i ulik retning. 3 er mer konkret enn 2, og knytter kanskje litt nærmere til betydningen av «jord, ikke sjø», men er ikke så klart avgrenset i et område som 2 er, og 1 inneholder noe om «del», altså avgrensing.

2.2. Arbeid med leksikalske felt

Et leksikon kan deles inn i en rekke leksikalske felt, også kalt semantiske felt (Bergenholtz et al. 1997:231). Bergenholtz et al. definerer et semantisk felt som en «samling av leksikalske enheter som er forent gjennom en viss betydnings-

messig samhörighet» (for oversikt over utvikling av teorien om leksikalske felt, se Öhman (1953)). Et eksempel på dette i vår ordbok er feltet «nedbør». Oppslaget **hagl** i figur 1 (ovenfor) inneholder en opplysningslinje som kun er til internt bruk for leksikografene. Feltet er merket med 000 og påpeker at definisjonsfeltet 312 skal koordineres for alle oppslagene *regn*, *yr*, *snø*, *hagl* og *sludd*. Dette feltet kan også mer presist defineres som et hierarkisk begrepssystem hvor begrepsrelasjonene utgjøres av overordning (genus proximum) og underordning (differentia specifica) (cf. Bergeholtz et al. 1997).

Genus proximum i vårt tilfelle er oppslaget **nedbør**. Definisjonsfeltet inneholder først en intensjonal definisjon: «vann (i en av sine mulige former) som faller fra skyer ned på jorden». Deretter følger en ekstensjonal definisjon som angir de relevante differentia specifica: *regn*, *sludd*, *snø* og *hagl*. De underordnede begrepene kan da defineres etter et bestemt mønster hvor genus proximum innleder definisjonen og differentia specifica utformes som en relativsetning som modifierer genus proximum. *sludd* defineres slik: «nedbør som består av både regn og snø». *snø* defineres som «nedbør som består av lette, hvite krystaller av is». *regn* defineres som «nedbør som består av vanndråper» og *hagl* defineres som «nedbør som består av kompakte, runde iskorn».

Slike leksikalsk/semantiske felt bidrar til å gi ordboken mer systematikk og vil være motiverende og klargjørende for brukergruppen. Men i praksis er det ikke alle områder av ordboken som ligger til rette for en slik feltanalyse. Derfor har vi ikke satset på å gjøre ordboken til en begrepsordbok, som beskrevet i Bergeholtz et al. (1997:75).

2.3. Prinsipper for uttalemarkering

I Norge har det vært utgitt noen uttaleordbøker for norsk. Ifølge Vanvik (1985:9) har disse brukt «en norsk lydskrift som en ikke kan regne med blir forstått utenfor landets grenser». I stedet brukte han det internasjonale uttalesystemet IPA. Det er flere grunner til at denne løsningen ikke treffer den intenderte brukergruppen for vår ordbok. For det første baserte Vanvik seg på det han kalte standard østnorsk. Det har skjedd mye siden 1980-tallet. I stedet for å satse på en uttalenormering basert på et begrenset sosialt og geografisk område har tendensen i massemedia vært å stimulere våre ulike dialekter. Dermed blir en uttaleangivelse bortimot umulig å forsvare. For fremmedspråklige blir derfor våre betydelige dialekt- og sosiolektforskjeller en stor utfordring.

I fravær av aksepterte uttalenormer har vi utarbeidet et sett av prinsipper som vil være nyttige for vår brukergruppe. Hovedmålet vårt er å få frem én uttale som har stor utbredelse og som trygt kan brukes på tvers av dialektgrenser. For det første, aksentuering (ordtrykk) skal angis, også ved enstavelsesoppslag hvis oppslaget potensielt sett kan aksentueres, for eksempel [huus]. Her ser vi også

at lang vokal angis med dobbelskriving. For det andre, det skal i hovedsak være én uttale for hvert oppslag. I visse tilfeller kan man tillate to alternativer, som i oppslaget **annen** som dels uttales [an^hen] og dels [aa^hen]. For det tredje skal uttalen være så bokstavnær som mulig, og ingen fonetiske tegn skal brukes. Eksempelvis vil oppslag som **kirke** og **kino** transkriberes [kjirke] og [kjiino], siden bokstavkombinasjonen <kj> typisk uttales [ç], en palatovelar spirant. Da får man noen unntak hvor kombinasjonen <kj> skilles ved morfemgrense, som i *brekkjern* [br^hek-jern]. Dette markeres ved å innføre en bindestrek i uttaleangivelsen. Den velare nasalen [ŋ] angis som <ng> som i [sang]. I en del norske dialekter uttales [g] i denne posisjonen, dvs. i stavelseskoda. Et tostavellesord som *kongress* angis som [kång-gress].

Prinsippet om bokstavnær uttale tilsier også at vi unngår markering av de østnorske retro-fleksene, som i *ert* [æf]. Generelt bygger vi på de vestnorske konsonantene og de østnorske vokalene. Verbet *erte* uttalemarkeres derfor med [æ^herte]. For nærmere opplysninger om dette henvises til innledningen i Khánh & Tràn (1983:10f).

2.4. Grammatiske opplysninger

Når det gjelder ordklassebetegnelser følger vi i hovedtrekk Bokmålsordboken (ordbok.uib.no). Men i noen tilfeller avviker vi fra det. I vår ordbok har bøyning alltid forrang foran syntaktisk funksjon. For eksempel oppslaget **delvis**. Dette er registrert som adverb i Bokmålsordboken mens vi har registrert det som adjektiv fordi oppslaget har adjektivmorfologi, som i *delvise tilståelser*. Hvis oppslaget ikke har bøyning anvendes syntaktiske kriterier, som i oppslaget **dagevis**, for eksempel i frasen *i dagevis* hvor oppslaget opptrer som styring for en preposisjon, en typisk substantivisk funksjon.

Et annet problem er representasjonen av oppslag som refererer til ulike språk. For eksempel er oppslaget adjektivet **tysk** definert som «som har sammenheng med, kommer fra eller tilhører Tyskland». *Han studerer tysk språk og litteratur. Det fins mange kjente tyske komponister*. Denne typen oppslag blir typisk registrert som adjektiver. Men disse kan også opptre som substantiver: *Han snakker et utmerket tysk*. Den mest nærliggende forklaringen er å analysere dette som en elliptisk konstruksjon hvor substantivet *språk* er utelatt. Men hva da med konstruksjoner som *Tysken hans er dårlig?* Her kan ellipseløsningen ikke anvendes. Jeg har ikke funnet noen norske enspråklige ordbøker hvor dette er representert. Vår nødløsning er å registrere disse hovedopplagene som adjektiver, og to substantiver som underoppslag, som for eksempel under *tysk* adjektiv: *tysk* s.m. tysk språk, *tysk* s.n. (bare i ubestemt form) tysk språk. Dette representasjonsproblemet gjelder alle oppslag som refererer til ulike lands kulturelle og politiske forhold. Som adjektiver oppfører de seg regelmessig som andre adjektiver, men når

de refererer til ulike lands språk oppstår representasjonsproblemen. Det gjelder adjektiver som **polsk**, **engelsk**, **fransk**, etc. Andre refererer kun til språk, som **latin** og **svahili**. For eksempel *Han snakker et godt svahili/polsk/fransk, men engelsken/tysken/russisken hans er dårlig*. Dette er altså et representasjonsproblem som omfatter en hel gruppe av ord i ordboken.

Når man skal gi bøyningsparadigmer i en ordbok brukes lineær segmentering så langt det er mulig. Dette anvender vi også i slike tilfeller hvor det bryter med morfologisk analyse, som i oppslaget **møbel**: møb|e|et, -ler, -la/-lene. Hvis lineær segmentering ikke er mulig angis formene fullt ut, som i synge, sang, sunget. De norske tonelagene angis ikke. Mange fremmedspråklige snakker fullt funksjonelt norsk uten tonelag. I flere norske dialekter mangler tonelagsmotsetning, og fenomenet er forholdsvis begrenset i norsk til forskjell fra tonespråk som vietnamesisk (hvor tonemarkeringen uttrykkes ortografisk).

3. Konklusjon

I denne artikkelen har jeg vist noen grunnleggende aspekter ved ordboksarbeid som har som mål å tilpasse ordtilfang, strukturering av ordboksartikler, bruk av prototypeteori og leksikalsk feltteori, prinsipper for uttalemarkering, og grammatiske opplysninger. Alle disse aspektene viser at leksikografer alltid må ha en analyse av intendert brukergruppe i tankene ved arbeidet. Dette gjelder særlig det skreddersydde ordutvalget, som beskrevet i seksjon 1 om bakgrunnen for dette arbeidet. NVO-prosjektet var tilpasset vietnamesernes behov i Norge på 1980-tallet. Siden den gang har norsk språk endret seg, særlig når det gjelder oppkomsten av nye ord. Dette har innvirket på ordtilfanget i den enspråklige ordboken for fremmedspråklige.

Det har vært viktig å unngå vanskelige ord og vendinger i synonymangivelse og i definisjonsutforminger. Standard uttalemarkering er viktig, og det er lagt stor vekt på å finne frem til uttalemarkeringer som er gangbare over hele Norge, og som unngår uttalevarianter som er vanskelig å kode på en enkel måte.

Litteratur

Ordbøker

- Bergenholtz, Henning, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén (red.) (1997): *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- Khánh, Nguyễn Quốc & M.M. Trần Ly San (1983): *Norsk-Vietnamesisk Ordbok*. Elektronisk utgave 2008: <www.uib.no/lle/20667/norsk-vietnamesisk-ordbok-p%C3%A5-nett>
- Kjelsvik, Bjørghild, Terje Svardal & Christian-Emil Ore: *Bokmålsordboka*. Nettutgave 2016. <ordbok.uib.no>
- Norwegian Newspaper Corpus Bokmål*. Created by Norsk aviskorpus. Distributed by the CLARINO UiB Portal: hdl:11495/D9B5-0349-4330-0 <clarino.uib.no/korpuskel/clarino-metadata?identifier=avis-plain>
- Vanvik, Arne (1985): *Norsk uttaleordbok*. Oslo: Fonetisk institutt, Universitetet i Oslo.

Annen litteratur

- Andersen, Øivin (2007): Deverbal Nouns, Lexicalization and syntactic Change. I: *Nordic Journal of Linguistics*. 30.1, s. 55–86.
- Jiang, Guiying & Qiaoyon Chen (2015): A Micro Exploration into Learners's Dictionaries: A Prototype Theoretical Perspective. I: *International Journal of Lexicography*. Vol. 30. No. 1. s. 108–139. Oxford: Oxford University Press.
- McEnery, Tony & Andrew Wilson (2001): *Corpus Linguistics*. 2.utgave. Edinburgh University Press.
- Öhman, Suzanne (1953): Theories of the "linguistic field". I: *Word* 9.2, s. 123–134.

Øivin Andersen
 professor, allmenn lingvistikk
 Institutt for lingvistiske, litterære og estetiske studier
 Universitetet i Bergen
 oivin.andersen@lle.uib.no