NoTa-taggeren: TAGGEVEILEDNING

Av Åshild Søfteland*

Versjon 15.01.2007

* Taggeveiledninga er skrevet av Åshild Søfteland som et resultatet av diskusjoner mellom flere involverte NoTa-medarbeidere: Kristin Hagen, Fredrik Jørgensen og Janne Bondi Johannessen. Lars Nygaard og Anders Nøklestad har arbeidet med preprosessering og den tekniske delen at NoTa-taggeren.

0. Om NoTa-taggeren og manuell tagging

NoTa-taggeren er en statistisk minnebasert tagger som er trent på materiale fra NoTa-korpuset. Ca 20 % av NoTa-materialet (= ca 190 000 ord) er manuelt tagget med ordklasse og morfologiske trekk. Denne taggeveiledninga beskriver taggsett og ulike valg som ble gjort undervegs i tagginga.

Den manuelle tagginga foregikk slik:
1) Det transkriberte materialet ble preprosessert av Oslo-Bergen-taggeren.
Selv om Oslo-Bergen-taggeren ble laget for tagging av skriftspråk, viste det seg at den med enkle justeringer kunne brukes til å preprosessere talemål, det vil si sette inn forslag til periodegrenser og multitagge hvert enkelt ord med alle mulige tagger. Ulempen med denne løsningen er at NoTa-taggeren dermed får noen av de samme svakhetene som Oslo-Bergen-taggeren når det gjelder f.eks. faste uttrykk og sammensetning av flerleddede navn (se kapittel 6. nedenfor)
2) Det preprosesserte materialet ble disambiguert av Oslo-Bergen-taggeren.
Selv om Oslo-Bergen-taggeren er en skriftsspråktagger, er den såpass robust at den takler talepråksmateriale med pauser, avbrudd osv. uten å gå i stå. Ofte, men langt i fra alltid, gir den riktige analyser også. Derfor fant vi det mest lønnsomt å bruke Oslo-Bergen-taggerens analyser som utgangspunkt for den manuelle tagginga.
3) Manuell tagging ble foretatt ved hjelp av et enkelt Unixprogram der taggeren fikk en og en periode opp på skjermen og ord for ord kunne velge riktig lesning (= tagg) for ordet. Riktig tagg ble valgt ved:

a) Trykke <enter> for Oslo-Bergen-taggerens forslag fra 2)
b) velge en av de andre oppgitte lesningene fra multitagginga 1)
c) skrive inn et eget forslag.

1. Generelle retningslinjer for tagginga

Tagging av materialet i NoTa skjer i hovedsak i samsvar med Oslo-Bergen-taggeren, og de valga som er tatt i implementasjon av denne. Oslo-Bergen-taggeren bygger igjen i hovedsak på Norsk Referansegrammatikk (Faarlund et al 1997) og Bokmålsordboka (BMO). Likevel vil materialet i NoTa i en del tilfeller ikke dekkes av beskrivelsene i Referansegrammatikken og analysene fra Oslo-Bergen-taggeren, fordi talespråk avviker fra skriftspråk på en del områder.

Eksempler i kursiv uten stor bokstav og punktum er hentet fra NoTa-materialet.

1.1 Kongruens

1.1.1 Substantiv får i utgangspunktet det kjønnet taleren bruker, selv om denne bøyningsformen ikke er lov etter BMO.

Sånn var dialekta da jeg vokste opp. (subst appell fem be ent)
Den maskina vi hadde var dårlig. (det dem fem ent, subst appell fem be ent)

Dette gjelder likevel sjelden i ubestemt form entall. (Se neste avsnitt.)

1.1.2 Kongruensfeil: Dersom et substantiv (etter BMO) og tilhørende determinativ eller adjektiv ikke kongruerer i kjønn og tall, tagges vanligvis ordene forskjellig. Noen ganger fordi taleren retter på seg selv etterpå (a), men oftest fordi man må anta et avbrudd – at taleren sannsynligvis begynner på en ny NP eller en ny setning underveis (b).

a) *ja bo- bor i en kollektiv e et kollektiv* (mask, nøyt)
b) (*ja er det en leilighet eller?*) *ja det er nei en hus # rekkehus* (mask, nøyt)
(*er det mange rom?*) *nei men vi har stort stue* (..) (nøyt, mask fem)

1.1.3 Alle ord som kan være hunkjønnsord i BMO kan også være hankjønnsord (tokjønnssystem). Derfor må alle ordene dette gjelder underspesifiseres, dersom konteksten ikke gir indikasjoner om kjønn.

Jeg flytta hit i sjetteklasse. (subst appell mask fem ub ent)
Jentene måtte reise seg. (subst appell mask fem be fl)

Dersom alle ordene i en NP er tvetydige, tagges alle ordene underspesifisert.

Jeg vil ha sånn rosa dokke. (sånn, adj ub ent pos, subst appell mask fem ub ent)
Dere sto på hver deres side av veien. (det mask fem ent kvant, det mask fem ent poss, subst appell mask fem ub ent)

1.1.4. Dersom ett ord er entydig, disambigueres hele konteksten.

Jeg vil ha en sånn rosa dokke. (det mask ent kvant, sånn, adj ub m/f ent pos, subst appell mask ub ent)
De sto på hver sin side av veien. (det mask ent kvant, det mask ent poss, subst appell mask ub ent)

1.2 Manglende kontekst

Noen ord krever en større kontekst for å disambigueres. For eksempel vil en del ord tagges som enten subjunksjon eller preposisjon avhengig av om de etterfølges av en leddsetning eller en nominal leddtype.

Jeg er sterkere enn ... – prep? sbu?

Dersom det ikke er tilstrekkelig kontekst til å disambiguere, tagges ordet med alle mulige alternativer; i dette tilfellet *prep* og *sbu*. Vi velger altså aldri å tolke NP-er som forkorta setninger, selv om det er sannsynlig i mange tilfeller.

Jeg snakka annerledes enn dem. – prep
Jeg snakka annerledes enn dem gjorde. – sbu

1.3 Gjentakelser og avbrudd

Når ord blir sagt flere ganger rett etter hverandre, får alle forekomstene samme tagg.

så så jeg gikk på Hellerud – konj, konj
det det det beste jeg vet – det nøyt ent kvant x3

Det samme gjelder når to (eller flere) ulike ord blir gjentatt som en enhet.

så det så det var e den gang – konj, pron, konj, pron
og det var lenge før lenge før det ble kult – adj pos, sbu, adj pos, sbu

Ellers blir gjentakelser tolket som deler av avbrudd og dobbelttagget.

nei det altså det var helt fryktelig – pron/det, pron
den er den jeg har ikke noe planer – pron, pron/det

Så lenge man kan ane bruddstykker av en setning, blir altså f.eks. *den* over analysert som en del av denne setningen – *den er ...* Andre eksempel:

så det ... – konj/sbu, pron/det

Legg merke til at *det* må dobbelttagges før avbruddstegnet ... . (Se avsnitt 3.1.2)

1.4 Nølelyder, pauser og ikke-språklige lyder

Nølelyder og pauser inni fraser eller inni setninger ignoreres når ordene rundt skal tagges.

det var F1 som e # fikk spørsmål om hun ville sy – sbu (ikke prep/sbu)
vi bor i en em leilighet som tanta mi eier – mask (ikke mask fem)
de e de folka som han kjøpte n av – det fl kvant (ikke pron/det)

Det samme gjelder normalt også for ikke-språklige lyder.

er det ikke det det [latter] det det heter? – pron x4

Men det tas hensyn til avbrutte ord (med bindestrek).

det bl- det var jo greit – pron/det, pron

2. Ordklasser med tilhørende morfologiske tagger

2.1 Ordklasser med morfologi

verb – verb
Tid: inf, pres, pret, perf-part
Ev.: pres inf (f.eks *brukes* i *hva brukes den til?* – verb pres inf

Dobbelttagger: inf/pres, inf/imp, pret/perf-part
Underspesifisering: verb (f.eks. *still going* – verb)

subst – substantiv
Type: appell, prop
Kjønn: mask, fem, nøyt
Bestemthet: ub, be
Tall: ent fl
Ev.: ubøy (f.eks. *skrå* i *på skrå* - subst appell ubøy)

Dobbelttagger: mask fem, mask nøyt, mask fem nøyt, ent fl
Underspesifisering: subst appell (f.eks *drar på linedance* - subst appell)

adj – adjektiv
Kjønn: m/f, nøyt
Bestemthet: ub, be
Tall: ent, fl
Grad: pos, komp, sup
Ev.: (pres-part) (f.eks. *engasjerende* i *en engasjerende lærer* – adj (pres-part))

Underspesifisering: adj ub ent pos, adj pos, adj (pres-part)

det – determinativ
Type: dem, kvant, poss, sp, forst
Kjønn: mask, fem, nøyt
Tall: ent fl
Ev.: gen (f.eks. *andres* i *andres plakater* – det dem fl gen)

Dobbelttagger: mask fem, ent fl
Underspesifisering: det kvant, det dem, det poss
(f.eks. *no affence* – det kvant, *hans problem* – det poss)

pron – pronomen
Kjønn: mask, fem, nøyt
Tall: ent, fl
Person: hum 1, hum 2, hum 3, pers 3
Kasus: nom, akk, nom obj, akk subj
Ev.: sp (f.eks. *hvem* - pron hum sp)

Dobbelttagger: mask fem, ent fl

Tagger til underspesifisering av ordklasse
pron/det (f.eks. *da må vi ta det vekk så det*)
verb/subst (f.eks. *triksing og finte og alt sånt med ball*)
verb/adj (f.eks. *chill* – sagt alene som ikke-replikk)
subst/adj (f.eks. *er ikke det standard sånn ... ?*

2.2 Ordklasser uten morfologi

adv - adverb

prep – preposisjon

interj – interjeksjon
Ev. + fyll, som i *em*: interj fyll)

konj – konjunksjon

sbu – subjunksjon

Tagger til underspesifisering av ordklasse:

konj/sbu/adv (*så*)	adv/sbu (*da, når, enda*)
konj/sbu (*så*)	prep/sbu (*som, fra, til, før, om*)
konj/adv (*så*)	adv/interj (*jo*)
konj/prep/adv (*for*)	

2.3 Tegn

2.3.1 Bindestrek

Alle ord som har en bindestrek til slutt, får taggen ”ukjent”.

sydd noe gardi- hadde sydd noe gardiner – ukjent
så er det sånn # sent- litt sånn liten noe # kakaogreie ved siden av - ukjent
at de ikke to- tok hensyn – ukjent
husker du ikke a- hun jenten som ble # m ranet? – ukjent
Er du kjent i Berg- ... ? – ukjent

2.3.2 Spørsmålstegn og hermetegn

Spørsmåltegn får taggen ”spm”.

?_[”$?” clb (spm)]

Hermetegn som blir tagga aleine får taggen ”anf”.

”_[”$” (anf)] «_[”$«” (anf)]

Oftest står hermetegn sammen med selve ordet: *jeg sier ”sepe”*

2.3.3 Avbrudd og pauser

Avbrudd og pauser blir tagga slik:

..._["..." clb avbrudd]
#_["#" pause] ##_["##" pause] ###_["###" pause]
_["" pause]

3. Problemer med ordklasser og morfologiske trekk

3.1 Ordklasser

3.1.1 verb eller adjektiv?

Ifølge Referansegrammatikken kan kongruens brukes som rettesnor for å avgjøre om et ord er adjektiv eller verb. Kongruerer ordet, er det adjektiv:
Adjektiv: *Han er spent - De er spente.*
Partisipp: *Hun er kommet - De er kommet.*

I Oslo-Bergen-taggeren er denne regelen forenklet slik at ord som kan være både adjektiv og verb, vil få verblesning etter "er", men adjektivlesning i nominale fraser. Denne forenklingen har vi valgt å følge.

er det sånn organisert eller? – verb perf-part
ja spilte på organiserte lag – adj fl pos

Men: *hun er alltid så opptatt* – adj ub m/f ent pos

Vi bruker også Oslo-Bergen-taggeren for å velge mellom adj og adj (pres-part). Taggeren velger adjektivlesningen om denne fins i Bokmålsordboka:

Boka var spennende. (adj ub m/f ent pos)
Boka var engasjerende. (adj (pres-part))

3.1.2 determinativ eller pronomen?

I Oslo-Bergen-taggeren kan *det*, *den*, *de*, *en*, *dette*, *denne*, *disse*, *noen*, *noe*, *alt*, *alle*, *ingen* m.fl. være både pronomen og determinativ. Dette ser også ut til å være intensjonen i Referansegrammatikken, selv om ikke alle ordene er nevnt der. Vi har uansett valgt å følge Oslo-Bergen-taggerens prinsipp; ordene er pronomen dersom de utgjør en frase alene eller dersom de har en leddsetning som utfylling.

Ingen liker meg. – pron
Er disse virkelig gode? – pron
Jeg liker alt. – pron
Det at hun ikke kom, ble et stort problem. – pron
Jeg kjente alle som var der. – pron
Vil du ha noe å bite i? – pron

NB! I eksempler av den siste typen, med utfyllende leddsetning, avviker vår kategorisering fra Referansegrammtikken.
det, den, de, dette, denne og *disse* er ifølge Oslo-Bergen-taggeren også pronomen foran *der, her, derre* og *herre*. Dette har vi så langt fulgt.

Den der er kul. - pron
De derre jentene fra Fagerborg kommer. - pron
Du liker vel ikke disse her? - pron

Andre veiledende eksempler (fra Oslo-Bergen-taggeren):

Og der var vi alle. – pron fl pers hum nom 1 , pron fl pers 3
Alle de skulle komme. – pron fl pers 3, pron fl pers 3 nom
Kan du klare alt det? – det nøyt ent kvant, pron nøyt ent pers 3

sånt noe – pron nøyt ent pers 3
noe sånt – det nøyt ent kvant
noe sånt noe – det nøyt ent kvant, det nøyt ent kvant
noe sånn – det kvant (altså underspesifisert)

ingen må veldig ofte underspesifiseres med *ent fl,* for hvor mange er egentlig ingen?

jeg har ingen hjemme – pron mask fem ent fl pers 3
det er vel ingen av de bilene som tar fem stykker? – pron mask fem ent fl pers 3
man merker jo ingen ting – det mask ent fl kvant

Det samme gjelder delvis også for *noen*.

tror du de kom fordi noen ga dem mat? – pron mask fem ent fl pers 3

En følge av valgene som beskrives i dette avsnittet, er at alle de aktuelle ordene må dobbelttagges ved avbrudd og andre tvilstilfeller: **pron/det**

ja men jeg så jo når det ... – pron/det

3.1.3 substantiv eller interjeksjon?

faen, fader, gud, guri, søren, herlighet o.l. kan bare være substantiv etter BMO. Vi tagger disse som interjeksjon når de har funksjon som interjeksjon.

faen jeg orker ikke mer – interj
jeg jobba som faen - subst

3.1.4 adjektiv eller adverb?

Alle adjektiv i nøytrum som fungerer som adverbial, tagges som adjektiv i nøytrum og ikke som adverb. (I samsvar med Referansegrammatikken.)

Han gikk ikke akkurat helt rett. – adj nøyt ub ent pos (ikke adv)

3.2 Morforlogiske trekk

3.2.1 Kasus

Vi har valgt å tagge personlige pronomen (utenom *den, det, en*, men inkludert *n* og *a* (se 5.1)) med nominativ eller akkusativ kasus, i samsvar med Bokmålsordboka og Oslo-Bergen-taggeren.

BMO gir blant annet at *han* kan være både nominativ og akkusativ, mens *hun*, *ho* og *de* bare er nominativ og *henne* og *dem* bare akkusativ. Som kjent er dette bildet annerledes i talespråk. Vi innfører derfor taggen *nom obj* for tilfeller der nominativsformen (etter BMO) brukes i akkusativposisjon (etter O-B-taggeren), og *akk subj* der akkusativformen brukes i nominativsposisjon. Vi får altså disse valgmulighetene:

hun, ho : pron fem ent pers hum 3 nom / pron fem ent pers 3 nom obj
henne : pron fem ent pers hum 3 akk / pron fem ent pers 3 akk subj
de : pron fl pers 3 nom / pron fl pers 3 nom obj
dem : pron fl pers 3 akk / pron fl pers 3 akk subj

Det er alltid akkusativ etter en preposisjon.

Jeg vil ønske god tur til dere som er her. – pron fl pers hum akk 2
dere hadde snakket med han som sto og fisket – pron mask ent pers hum 3 akk

Kasuseksempler fra NoTa-materialet:

vært på utstillingen og sett på hun Oko Doko – nom obj
ja for nå sendte jo alt opp til hun F1 vet du – nom obj
har du smakt de med lakris? – nom obj
hos noen av de som bor rundt der – nom obj
klarer du og henne å bli som bestevenner – akk subj
dem sender dem til gamlehjem – akk subj
det så ut som dem hadde vaska – akk subj

3.2.2 Tall

Det er ofte vanskelig å bestemme om nøytrumsord i ubestemt form er entall eller flertall. Der bøyningsformene er like, er det bare den semantiske bruken som kan gi signal om det ene eller det andre. Vi har forsøkt å bruke regelen om "tellelige substantiver brukt i ikke-tellelig betydning" (Referansegrammatikken s. 478), men den er ikke alltid enkel å følge i praksis:

er jo så kolossalt med elg og rådyr i Marka - ent
mye barn i området – fl
åssen var det å være barn på Holmlia? – ent fl
hun underviser i språk – ent fl (sannsynligvis *ent*, men ikke sikkert)
det var ikke så masse dataspill og sånn – ent fl
selv om de faktisk bruker dollar i El Salvador nå – ent fl

Det kan også være vanskelig å bestemme seg for om substantivene står i entall eller flertall i mer eller mindre faste uttrykk som disse:

da har man ikke behov for å ha enda en forsikring – ent fl
det er ofte sånn har jeg inntrykk av – ent
åssen er det i forhold til der du vokste opp - ent
vi kan jo skifte tema – ent
etter at de skilte lag – fl

Andre eksempler på tvilstilfeller:

ikke noe problem – ent fl
jeg ga dem tilbud om det – ent fl
vi kan ikke bore hull i veggen – ent fl

De samme problemene gjelder også for noen hankjønnsord, spesielt hyppig er *ting*.

er det noen ting som ikke er så bra med den plassen – ent fl

folk er hyppigst i bruk som flertallsord i vårt materiale

folk er gærne – fl

folk hadde høy utdanning – fl

3.2.3 Kjønn

I utgangspunktet kan alle determinativer, adjektiv og pronomen disambiguere en NP med hensyn på kjønn. Men ofte må man regne med avbrudd der disse ikke samsvarer med det (eller de) kjønnet et substantiv har etter BMO. Som omtalt i 1.1 beholdes substantivets kjønn i flertall og i bestemt form entall, men forholdsvis sjelden i ubestemt form entall. Alle mulige hunkjønnsord må som nevnt dobbelttagges *mask fem* dersom konteksten ikke disambiguerer.

3.2.4 Bestemthet

En del ord må sies å kunne ha bestemt semantisk betydning selv om den morfologiske formen er ubestemt. Vi har valgt å tagge disse som ubestemte (ub) uansett.

har du skrudd på kamera? – subst appell nøyt ub ent

jeg bare skrur på tv – subst appell mask ub ent

jeg skal bare lukke igjen vindu – subst appell nøyt ub ent

er n redd for politi? – subst appell nøyt ub ent

4. Vanskelige ord

4.1 der – subjunksjon eller preposisjon?

Etter Referansegrammatikken og Oslo-Bergen-taggeren er *der* subjunksjon når det innleder en setning, men preposisjon når det følges av *som*.

a) *Åssen var språket der som du vokste opp?* – prep, sbu

Vi antar likevel at *som* er strøket i setninger som b), og tagger *der* som preposisjon uansett om *som* blir uttalt i en setning eller ikke.

b) *Åssen var språket der du vokste opp?* – prep

4.2 som – subjunksjon eller preposisjon?

Hovedregel: *som* er subjunksjon når den etterfølges av en leddsetning, preposisjon når den etterfølges av en NP. Ved avbrudd og i andre tvilstilfeller underspesifiseres det med *prep/sbu*.

puttet den oppi den samme Nesquik-boksen som liksom ... – prep/sbu

det virka som de hadde gjort det med vilje – sbu

det er som et sånt fjellandskap da - prep

De samme reglene gjelder for *enn* (se avsnitt 1.2), men dette ordet er mye mindre brukt i materialet.

4.3 andre – determinativ eller adjektiv?

Formen *andre* kan være både determinativ (av ”annen”) og adjektiv (ordenstall). I mange tilfeller er forskjellen bare semantisk, slik at Oslo-Bergen-taggeren er til liten hjelp. Vanligvis går det likevel greit å velge, men det er en god del tvilstilfeller der vi noterer begge.

andre brukte vel ”du” – det dem fl
men de bodde på andre siden av Akerselva – det dem be ent
vi bodde i det andre huset – det adj

4.4 hvilken/hvilket/hvilke – determinativ eller pronomen?

Disse formene er bare angitt som determinativ i Referansegrammatikken. Vi bruker ikke taggen ”pron sp”, men ”det mask ent sp”, ”det fem ent sp”, ”det nøyt ent sp” og ”det fl sp”.

4.5 jo – interjeksjon eller adverb?

Oslo-Bergen-taggeren er ikke helt klar på når *jo* er interjeksjon og når det er adverb, siden *jo* er mye mindre brukt i skriftspråk enn i talespråk. Vi prøver å følge Referansegrammatikken: ”[*Jo*] kan stå trykklett i midtfeltet og uttrykkje sendarens haldning til innhaldet i setninga (...). [*Jo*] blir [brukt] til å forsterke innhaldet i setninga [og] understrekar at det som blir sagt, verkeleg er i samsvar med røyndommen.” (s. 824). I slik bruk er *jo* adverb.

jeg ser jo at andre kan bli trakkasert – adv
pluss at det også er jo ganske mye trafikk – adv

Jo er også adverb ved en viss type sammenligningskonstruksjoner: ”Ein særskild type samanlikningssetningar er innleidd av [*jo*] (...) pluss eit adjektiv i komparativ.” (s. 1072)

Jo mer du kjøper jo billigere får du dem. – adv

I alle andre forekomster skal *jo* normalt være *interjeksjon*. ”*Jo*|*jau* blir brukt for å uttrykkje at ein er einig i den underforstådde bodskapen når spørsmålet eller eller utsegna inneheld ei nekting. Det kan også brukast når det ligg nøling eller varsemd i svaret.” (s. 968)

så jeg jo jeg jeg er enig – interj
jeg er ikke noe # jo og så kunne jeg ikke – interj

4.6 sånn – adverb, determinativ eller pronomen?

Etter Referansegrammatikken kan *sånn* vere adverb og *sånn*/*sånt*/*sånne* determinativ. Oslo-Bergen-taggeren kategoriserer i tillegg *sånt* som pronomen i visse konstruksjoner. I starten prøvde vi å følge Oslo-Bergen-taggeren så godt som mulig, men det viste seg raskt å være veldig vanskelig å analysere *sånn* konsekvent i det hele tatt.

og når man var ute og sånn – adv? pron?
så jeg tror ikke det var noe sånn voldsomt – det? mask/fem? nøyt?
ordne og styre med anlegg og sånn – pron? ent? fl?
så får lære litt sånn # hvordan bjelke brekte bein – adv?

Vi bestemte oss etter hvert for å ikke prioritere å bruke tid på å analysere de utallige tvilstilfellene, og gir alle forekomster av *sånn, sånt* og *sånne* taggen ”*sånn*”.

Løren Refstad og Risløkka og sånt – sånt_[”sånn” sånn]

4.7 så – konjunksjon, subjunksjon eller adverb?

Her er Oslo-Bergen-taggeren veldig usikker, delvis feil. Vi følger Referansegrammatikken:

4.7.1 Konjunksjon

”Konjunksjonen *så* innleier hovudsetningar. Innhaldet i setninga er normalt ei følgje (konsekvens) av innhaldet i setninga føre.” (s. 1140)

Det ble dårlig vær, så vi avlyste turen.

4.7.2 Subjunksjon

”Hovudsetningar med *så* står nær leddsetningar innleidde av *så*. Slike leddsetningar uttrykkjer føremål (…). *Så* er da subjunksjon, og setninga følgjer skjema B.” (s. 1141)

Jeg må lese mer, så jeg ikke stryker til eksamen.

4.7.3 Adverb

”*Så* brukes som adledd foran adjektiv oftest når adjektivet har en at-setning som utfylling (...)

Han svarer så heftig at hun nesten blir redd.

[og] i nektede setninger, oftest når adjektivet har utfylling med *som* + X.

Hun kom ikke så ofte som før.

Men *så* kan også forekomme uten slike utfyllinger. Det kan da ligge implisitt en sammenlikning, der sammenhengen viser hva en sammenlikner med.” (s. 397)

Han var like stor som deg. / Jaså, var han så stor?

4.7.3.1 Adverb med forsterkende funksjon

”Eller sammenlikningen gjelder en ubestemt grad, slik at *så* får en allment forsterkende funksjon” (s. 397)

er jo så kolossalt med elg og rådyr i Marka

4.7.3.2 Adverb som kontekstbindende adverbial

”Blant kontekstbindande adverbial står *så* i ei særstilling. (…) Vi skal her skilje mellom to hovudfunksjonar for dette adverbet, som vi kan kalle konkret og abstrakt bruk. I den konkrete bruken refererer *så* til eit tidspunkt som følgjer like etter eit anna tidspunkt som alt er etablert, og tyder ’etterpå, deretter’.” (s. 816)

Først bodde vi på Tonsenhagen, så flytta vi til Romsås, og nå bor vi her.

”I den abstrakte bruken er både det semantiske innhaldet og den fonologiske forma svært avsvekka. Her kan *så* berre stå i forfeltet like etter eit ledd i ekstra-posisjon i laust forfelt. Det vanlegaste er at det ekstraponerte leddet er eit fritt adverbial. *Så* fungerer da som pro-ord for det førestilte adverbialet. Dette adverbialet kan vere alle slags frie adverbial (…).” (s. 817)

Da vi hadde gått i flere timer, så kom vi fram til ei hytte.

I London så bodde vi på hotell.

Uten deg så ville jeg ikke overlevd.

4.7.4 Bruk av reglene i praksis

Når en skal utøve disse reglene på det faktiske materialet, oppstår det diverse problem.

Så som **subjunksjon** er sjelden, og kan derfor være vanskelig å få med seg når et tilfelle først dukker opp.

og virkelig sånn så det svinger – sbu

I praksis gir ikke *så* alltid en konsekvens (semantisk) selv om syntaksen tilsier at det skal være **konjunksjon**.

(*hvor i Oslo er du vokst opp hen?*)
det er Oslo altså Skullerud (segm) *så jeg har bodd her i åtte år nå* – konj

Her følger vi syntaksen, slik at *så* uansett tagges som konjunksjon dersom det kommer en helsetning rett etterpå.

I noen av tilfellene der *så* ikke følges av en NP, må man anta at NP-en er utelatt, og at *så* er konjunksjon likevel.

så starta som fjorten men # alltid likt det liksom – konj
så var alene med disse herre to barnebarna – konj

Korte pauser og nølelyder blir normalt ignorert.

så e det var faktisk for det meste fotball og venner – konj
så # de fleste av kameratene og venninnene mine har flytta – konj

Så er også konjunksjon i disse tilfellene:

så du syns det var fint å vokse opp på den måten?
så da måtte vi bare finne oss i det
så jeg bare "ja vi må vel"

Men **adverb** i disse tilfellene:

så når det er nedoverbakke da så hadde vi (…)
så hvor har du gått på skole hen?

I praksis kan dette være en tommelfingerregel: Følges *så* av en NP, er *så* konjunksjon. Ellers er det for det aller meste adverb. (Dessuten: Dersom Oslo-Bergen-taggeren foreslår *konj* i det hele tatt, velger vi det.)

Informantene slutter svært ofte ytringene sine med et enkelt *så* til slutt: *nei da så*. I alle disse tilfellene og ved de fleste avbrudd, underspesifiseres *så* med *konj/sbu/adv*.

Unntak: *ellers så …* – her er *så* pro-ord for *ellers*, og derfor adverb.

5. Nye ord og nye klassifiseringer

5.1 NoTa-ord

Før transkriberinga av NoTa-materialet ble det bestemt at en del ord/lyder som ikke står i BMO, men som er vanlige i talespråk, skulle skrives rett inn uten noen spesiell markering. Disse er:

a – pron fem ent pers 3 nom / pron fem ent pers 3 akk
n – pron mask ent pers 3 nom / pron mask ent pers 3 akk
e, *em*, *m* – interj fyll
aha, eh, ehe, heh, hm, hæ, jaha, m-m, mhm, mm, næ, nja, næhei, ops, u, ææ, å-å, å ja – interj

5.2 Nyord

De fleste andre ord som ikke står i BMO, er markert med ”language =X” i transkripsjonen (samlekategori med all slang, ord med ukjent opphav, utenlandske ord, dialektord). Alle ordene tagges med ordklasse etter den funksjonen de har når de blir brukt. De morfologiske trekkene må ofte underspesifiseres, spesielt for substantiv når de ikke er bøyd.

er ikke det litt risky?	adj nøyt ub ent pos
det der er bare bullshit	subst appell
yes	interj
nummer sjuttiotre	det fl kvant
og bare får skikkelig lættis	subst appell
jeg er hyper fra før	adj ub m/f ent pos
når han driver og battler	verb pres
mange der ramla ta	prep

I tillegg er det en del ikke-tillatte ord som står i hermetegn. Disse er gjerne riksmåls- eller dialektvarianter av ord i BMO, og skrives uttalenært og med hermetegn når talerne snakker om språk og dialektforskjeller.

jeg sier ”sepe” og ”sne” (subst appell, subst appell)
der sier de ”sterinslys” og ”komfemasjon” (subst appell, subst appell)

Noen ord er markert med ”lang” selv om de står i BMO. Det gjelder når ordene har funksjon som en annen ordklasse enn den BMO sier de er. (Se også 3.1.3)

Johnny Depp er konge [lang X] – adj ub m/f ent pos (BMO: substantiv)
han er kaos altså [lang X] – adj ub m/f ent pos (BMO: substantiv)
shit [lang X] – interj (BMO: substantiv)

regnete – (hvordan var det i Bergen?) det var regnete
triksing – triksing og finte og alt sånt med ball
og tilsvarende ord er ikke markert med ”language” fordi produktive ordlagingsmetoder skal være tillatt. Taggeren klarer normalt å analysere denne typen ord riktig fordi den kjenner igjen endinger som *-ete* og *-ing*.

Sammensetninger som *fem-seks* (*jeg var vel fem-seks år*) blir heller ikke markert med ”lang”.

5.2.1 Eksempler fra materialet: Verb

bare kjøre rundt og bare chille - verb inf
fucke for moro skyld – verb inf
du bare klæsjer på noe – verb pres
hun loker så jævlig – verb pres
bøffa brettet til broren min – verb pret
det var han som loka – verb pret
det er noe loka dritt – verb perf-part
da hadde jeg faen meg bæda helt – verb perf-part
fuck det da – verb imp
sms mer ring mindre – verb imp

5.2.2 Eksempler fra materialet: Substantiv

sms – subst appell mask ub ent fl
for å kjøpe sigg – subst appell mask ub ent
hun er jo bimbo da – subst appell mask ub ent
om jeg hadde noe cash – subst appell
vi skal trikse og sånn i halftime – subst appell
vi er enemies – subst appell
har fått masse nye rales – subst appell

5.2.3 Eksempler fra materialet: Adjektiv

det er chill – adj nøyt ub ent pos
det er taz da – adj nøyt ub ent pos
det blir dritsjmø – adj nøyt ub ent pos
det var dritdigg - adj nøyt ub ent pos
jeg skal kjøpe digg bil – adj ub m/f ent pos
hun er litt småchubby – adj ub m/f ent pos
han er rich – adj ub m/f ent pos
jeg er jo så close med fotball – adj ub m/f ent pos
er dem hypp på penger – adj fl pos
mye chillere – adj komp
Norges tjueste høyeste fjelltopper – adj be sup

5.2.4 Eksempler fra materialet: Determinativ

hun måtte be ørti bønner om dagen – det fl kvant
x antall år – det kvant

5.2.5 Eksempler fra materialet: Pronomen

what? – pron sp
å gjør du der – pron sp

5.2.6 Eksempler fra materialet: Adverb

alle de derre løse stolene – adv [fordi *herre* er adverb i BMO]
der er det vestkantspråk de-luxe – adv

5.2.7 Eksempler fra materialet: Preposisjoner

(Ikke markert med ”language” fordi de hører til navn. Se 6.3)
Leonardo di Caprio – prep
Marco van Basten – prep

5.2.8 Eksempler fra materialet: Interjeksjoner

fuck – interj
damn – interj
ållø – interj

mor – interj
jå – interj

5.2.9 Eksempler fra materialet: Lengre fraser

shit happens – subst appell, verb pres
no affence – det kvant, subst appell
still going – adv, verb
good old times – adj, adj, subst appell

5.3 Endringer av / tillegg til ord som står i BMO

fra – sbu (i tillegg til prep)
der bodde vi fra jeg var fem
nær – prep (jf. NRG) (i tillegg til adj)
det er nær Marka
slags – adj (i stedet for subst i konstruksjoner som her)
Hva er dette for slags tull?
rød – adj be ent pos, adj fl pos (i tillegg til adj ub m/f ent pos)
de bor i det rød huset, damene har rød bånd
ferdig, enig – adj fl pos (i tillegg til adj ub m/f ent pos)
vi sier oss ferdig, da er vi jo enig

Ved bruk av taggen "ubøy" markeres ingen morfologiske trekk (≠ BMO og O-B-taggeren):
lov, råd (m.fl.) – subst appell ubøy
Det får du ikke lov til. Det har vi ikke råd til.
norsk, dansk, spansk (m.fl.) – subst appell ubøy
Boken er på dansk.
ingenting – subst appell ubøy
Ingenting er bedre enn det.

6. Sammensatte uttrykk

6.1 Faste uttrykk

Ord som inngår i faste uttrykk, slås ofte sammen i preprosesseringen av Oslo-Bergen-taggeren:

F.eks: *av gårde* (adv prep+subst), ***av veien*** (adv prep+subst), *blant annet* (adv prep+adj), *bortsett fra* (prep), *den dag i dag* (adv det+subst+pret+subst), *den gang* (adv det+subst), *der inne* (prep), *der borte* (prep), *der oppe* (prep), *etter at* (sbu), *fly forbanna* (adj adv+adj), *for eksempel* (adv prep+subst), *for moro skyld* (adv prep+subst+subst), ***for resten*** (adv), *for så vidt* (adv), *for tiden* (adv prep+subst), *fy faen* (interj), *glipp av* (prep subst+prep), *gud hjelpe* (interj), *hvor hen* (prep), *i fjor* (adv prep+subst), *i grunnen* (adv prep+subst), *i går* (adv prep+subst), *i går kveld* (adv prep+subst+subst), *i hvert fall* (adv prep+det+subst), *i kveld* (adv prep+subst), *i morgen* (adv prep+subst), ***i møte*** (adv prep+subst), *i natt* (adv prep+subst), *i ny*

og ne (adv prep+subst+konj+subst), *i stad* (adv prep+subst), *i stedet* (adv), *i stedet for* (prep prep+subst+prep), *i tilfelle* (adv prep+subst eller prep+subst sbu), *ja vel* (interj interj+adv), *ja visst* (adv interj+adv), ***jo da*** (interj interj+adv), *lille julaften* (subst appell mask ub ent), *lite grann* (subst adj+subst), ***m m*** ("adv fork" prep+adj), ***må vite*** (adv verb+verb), ***om enn*** (sbu), *om gangen* (adv prep+subst), *om kapp* (adv prep+subst), ***på si*** (adv prep+subst), *stort sett* (adv adj+verb), ***så som*** (prep adv+prep), *til og med* (prep prep+konj+prep), *til stede* (adv prep+subst), *uff a meg* (interj), *ut over* (prep).

Ordene med fet skrift forekommer sjelden eller aldri som faste uttrykk i vårt talespråklige materiale. Det vil si at de ikke fungerer som den ordklassen som står (først) i parentesene, men som enkeltord med den ordklassen de har / kan ha når de står alene. Ordene burde derfor ikke ha vært slått sammen under tagginga.. Dette er rettet på i det materialet som er manuelt tagget, men i tagginga med den statistiske talemålstaggeren, blir det foreløpig feil.
Manuell tagging:

Og så kjører du bare av veien etter krysset. – prep, subst
Vi kjøpte sjokolade for resten av pengene. – prep, subst
Jeg satt i møte hele dagen. – prep, subst
faren min blir jo da som en onkel – adv, adv
*nei da * m m ja * ja* – interj fyll, interj fyll
Du må vite sånt vet du. – verb, verb
Det vet jeg mer om enn dere. – prep, prep
nei enda bedre holdt jeg på si – prep, verb
det er så som F2 sa – adv, sbu

6.2 Sammensatte fellesnavn

Oslo-Bergen-taggeren klarer under preprosesseringen å slå sammen sammensatte fellesnavn når de er skrevet med bindestrek på formen *xxx- og/eller yyy*.

*ballett- og danselinje*_[”ballett- og danselinje” subst appell mask fem ub ent]
*andre- eller tredjeklasse*_[”andre- eller tredjeklasse” subst appell mask fem ub ent]

I NoTa-transkripsjonen brukes bindestrek også som en avbruddsmarkør, og derfor blir en god del ord slått sammen som skulle stått aleine.

jeg v- og var vant til – ”v- og var”
enda en s- eller forsikring – ”s- eller forsikring”

Konstruksjoner uten bindestrek eller uten *og*/*eller* blir dessverre ikke slått sammen.

seks og trettiåring – det fl kvant, konj, subst appell mask ub ent
femte sjetteklasse – det fl kvant, subst appell mask fem ub ent
West Ham-laget – subst prop, subst appell nøy be ent
elleve- til trettenårsalderen – ukjent, prep, subst appell mask be ent

6.3 Sammensatte egennavn

Alle ord som begynner med stor bokstav får taggen ”prop”, utenom de som er avbrutt med bindestrek, f.eks. *Osl-*, *Holm-*, *M1-*, som er ”ukjent”. Ingen ord som begynner med liten bokstav kan vere ”prop” hvis de står aleine.

*Real*_[”Real” subst prop] *Madrid*_[”Madrid” subst prop]
*Gamle*_[”Gamle” subst prop] *Bislett*_[”Bislett” subst prop]

*Man*_[”Man” subst prop] *City*_[”City” subst prop]
*AC*_[”AC” subst prop] *Milan*_[”Milan” subst prop]
*Boeing*_[”Boeing” subst prop] *747*_[”747” det fl kvant]
*Holmlia*_[”Holmlia” subst prop] *skole*_[”skole” subst appell mask ub ent]
*Norsk*_[”Norsk” subst prop] *to*_[”to” det fl kvant]
*Tegning*_[”Tegning” subst prop] *form*_[”form” subst appell mask fem ub ent]
*og*_[”og” konj] *farge*_[”farge” subst appell mask ub ent]
*van*_[”van” prep] *Nistelroy*_[”Nistelroy” subst prop]
*Leonardo*_[”Leonardo” subst prop] *di*_[”di” prep] *Caprio*_[”Caprio” subst prop]

Taggeren slår bare sammen noen få kjente sammensatte egennavn.
*New York*_[”New York” subst prop]
*North Dakota*_[”North Dakota” subst prop]
*Trinidad og Tobago*_[”Trinidad og Tobago” subst prop]
*Sogn*_[”Sogn” subst prop] *og*_[”og” konj] *Fjordane*_[”Fjordane” subst prop]
*Steen*_[”Steen” subst prop] *og*_[”og” konj] *Strøm*_[”Strøm” subst prop]

De samme prinsippene gjelder for sammensatte egennavn i **hermetegn**.
”Se og Hør” – (anf), subst prop, konj, subst prop, (anf)
”Da Vinci-koden” – (anf), subst prop, subst appell mask be ent, (anf)
”Mannen med den nakne pistol” – (anf), subst prop, prep, det dem mask ent, adj be ent pos, subst appell mask ub ent, (anf)
Men taggeren klarer oftest å slå sammen egennavn i hermetegn dersom alle eller de fleste orda har stor bokstav.
*”Kill Bill”*_[””Kill Bill”” subst prop]
“*Lock Stock and to Smoking Barrels*”_[“”Lock Stock and to Smoking Barrels”” subst prop]

7. Andre forhold

7.1 Ulike grunnformer

Når formen *gamle* dukker opp, må taggeren velge mellom ”gammal” og ”gamle” som grunnformer. Hva den velger vil variere; det blir altså ikke markert at det er to muligheter.

Det samme gjelder for *steder* – [”sted” .. nøyt ub ent] / [”stad” .. mask ub ent]

7.2 Konjunktiv

Bruk av særskilte konjunktivformer er svært sjelden. (Under den manuelle tagginga kom det opp ett eksempel: *frekkheten lenge <u>leve</u>*.) Derfor blir ikke konjunktiv markert i materialet.