

Anne Golden
 Norskundervisningen for utenlandske studenter
 Universitetet i Oslo

PRESENTASJON AV PROSJEKTET LÆREBOKSPRÅK

Når de fremmedspråklige elevene og de utenlandske studentene begynner henholdsvis på skolen og ved universitetet, starter de med generelle norskkurs. Disse kursene er generelle i den forstand at de bygger opp elevenes/studentenes norske syntaks ved hjelp av tekster som omhandler forskjellige dagligdagse situasjoner og hendelser. Bøkene som brukes er gjerne generelle innføringsbøker i norsk, dvs de er ikke beregnet på de enkelte elevers videre behov for spesiell norskopplæring. Når så elevene/studentene er blitt gode nok i norsk etter de forskjellige klasses/trinns kriterier, blir de integrert i vanlige norske klasser, dvs de får undervisning på norsk sammen med norske elever/studenter. Men erfaringen viser at de får store vanskeligheter når de skal begynne å lese fagbøker. I grunnskolen er det bl a O-fagsbøkene 1) som volder problemer, også for de elevene som gjorde det bra i det generelle norskkurset. I ekstratimene i norsk blir det ofte til at lærerene leser O-fagsleksene sammen med elevene og forklarer det som står der. Flere lærere har klaget over at det ikke finnes noe materiell som elevene kan arbeide med slik at overgangen til de forskjellige fagene blir lettere for elevene.

Med prosjektet LÆREBOKSPRÅK vil vi prøve å lage et slikt støtte-materiell for fagene fysikk, geografi og historie, slik at det blir bygget en bro mellom de generelle norskkursene og norsken i disse fagene i grunnskolen. Vi må derfor finne fram til hva det er som karakteriserer disse fagtekstene og hvorfor de er så vanskelige for de fremmedspråklige elevene. Spørsmålet blir da på hvilken måte disse tekstene skiller seg fra de tekstene som elevene har arbeidet med i de generelle språkkursene. For å gi et svar på dette må vi undersøke følgende:

1. Ordforrådet
2. Syntaksen i setningene
3. Oppbygningen av tekstene utover setningsplanet

Sannsynligvis er det flere faktorer som er med på å vanskeliggjøre overgangen til "faglekser". Selve læringssituasjonen blir forandret i og med at fagtekstene formidler kunnskap som skal læres. I lærebøkene i norsk som fremmedspråk er det språket selv som skal læres, men i fagbøkene er språket et middel til kunnskapstilegnelse. Med andre ord, når elevene begynner å "lese fag", får de en ny oppgave lagt til den de hadde da de bare arbeidet med det nye språket, nemlig at de skal kunne forstå og gjengi med egne ord den nye kunnskapen som teksten gir. For at elevene skal kunne mestre den nye oppgaven, slik at de kan

1) Orienteringsfag er historie naturfag og geografi integrert

lære fagene, er det nødvendig at de allerede så langt det er mulig behersker den type språk de finner i lærebøkene.

I første omgang har vi tatt for oss ordforrådet i O-fagsbøkene. I de generelle norskkursene er tekstene ofte bygd opp omkring sentrale sosiale situasjoner som 'på postkontoret', 'på trygdekontoret', 'på restaurant' osv og ordforrådet som blir presentert er følgelig det som brukes i slike situasjoner. Når vi sammenlikner dette ordstoffet med det i O-fagsbøker er det to typer av ordforråd som lett kan skape problemer for elevene:

Ordtype 1: Fagtermer

Ordtype 2: Ord som tilhører det allmenne ordforrådet, men som forekommer sjelden i de generelle språkkursene

Med ordtype 1 mener vi ord og uttrykk som inngår i selve faget og som enten er nye eller får et nytt og skarpere definert innhold også for de norske elevene i klassen. Eksempler fra en fysikkbok for 6. klasse: oppdrift, friksjon. Disse fagtermene er vanskelige, men fordi de er vanskelige også for de norske elevene, vil de oftest bli gjennomgått og forklart av faglæreren.

Med ordtype 2 mener vi ord og uttrykk som tilhører det allmenne ordforrådet, men som har høyere frekvens i spesielle fagtekster enn i generelle språkkurs. Eksempler fra samme fysikkbok: legeme, avta, gnidning, rivning. Dette er ord som vil være vanligere i fysikkbøker enn f eks historiebøker. Disse ordene vil være kjent for de fleste norske elevene, men fordi de ikke blir behandlet i undervisningen, vil de ofte være spesielt vanskelige for de fremmedspråklige elevene. Ofte vil fagtermene bli forklart ved hjelp av nettopp slike ord - i den fysikkboka vi hentet eksemplene fra ble friksjon forklart ved gnidning og rivning - og dette fører til at de fremmedspråklige elevene ikke kan nytte den hjelpen faglæreren gir når han/hun gjennomgår det nye stoffet.

Vi vil derfor kartlegge disse to ordtypene. For å skille ut fagtermene har vi søkt assistanse hos faglærerene. På alfabetiske ordlister har vi bedt dem krysse av de ordene de regner for fagord i en klasseromssituasjon.

Vi har tatt for oss et utvalg av fysikk-, geografi- og historiebøker fra 4. til 9. klasse. Årsaken til at vi har valgt akkurat disse fagene er at de presenterer stoffet på forskjellige måter. På den ene siden står fysikkfaget som i stor utstrekning presenterer fakta med klare årsaksrelasjoner. På den andre siden står historiefaget som er langt mer beskrivende i formen. Geografifaget blir en mellomting av disse idet naturgeografien ligger nærmere fysikken i form, mens kulturgeografien kan sammenliknes med historiefaget. Vi vil derfor undersøke i hvor stor utstrekning dette gir seg utslag i ordforråd og syntaks.

Vi har lagt vekt på å finne fram til bøker som er mye brukt, men som samtidig er forholdsvis nye, slik at vi kan regne med at de vil være aktuelle i en del år framover. På barnetrinnet finnes disse fagene i O-fagsbøker, men kapitteinndelingen følger svært ofte de tradisjonelle faggrensene, så det har vært forholdsvis enkelt å skille ut de enkelte fagene. På hvert klassetrinn har

vi valgt ut mellom tre og seks bøker fra forskjellige forlag innen hvert fag. På ungdomstrinnet er derimot den tradisjonelle faggrensen opprettholdt og vi har valgt ut to til tre hele bøker fra forskjellige forlag innen hvert fag. Vi har fått kjørt ut alfabetisk sorterte ordlister med frekvens. Programmet gir oss også antall løpende ord, antall ulike ord og T/T-ratio. Når alle tekstene er skrevet inn, vil vi få i alt 18 frekvensordlister.

Ordlisteprogrammet vil regne alle identiske tegnsekvenser avgrenset av skilletegn eller mellomrom for samme ord, og i frekvensordlistene er homografene følgelig ikke skilt ut.

Eks fra fysikk 9. klasse: hell (1)
 helle (1)
 heller (19)
 helt (44)

Vi har derfor gått gjennom alle listene og plukket ut alle ord som kunne være homografer (se nedenfor ang kriteriene for homografer). Deretter har vi kjørt ut konkordanser på disse ordene. Fra konkordanslistene kan vi skille ut de forskjellige homografkomponentene som finnes i teksten og hvilken frekvens de har.

Eks fra fysikk 9.klasse: heller (adv) (8)
 heller (verb, 'tømmer')(10)
 heller (verb, 'skrår') (1)
 heller (subst) (0)

Et ord som har valgfri ortografi (eks likne og ligne) vil bli ført opp som to forskjellige ord i frekvensordlisten siden ordlisteprogrammet regner to ord for ulike idet ett av tegnene varierer. Feilstavinger vil også bli ført opp som egne ord, og ikke minst vil bøyde former av ett og samme grunnord bli regnet for forskjellige ord.

Vi har derfor funnet det nødvendig å lemmatisere ordene, og dette arbeidet er blitt ganske omfattende. I følge Allén (1970) er et lemma "en grupp ordformer inom en ordklass vilka kan hänföras till antigen en och samma flexionsserie eller flera i tal och/ eller skrift konvergerande serier vars divergenser visar rent fakultativ variation" (bind 2, side XVIII). Et lemmatiseringsarbeid der de forskjellige ordenes lemmatilhørighet bestemmes, vil følgelig holdes innenfor ordklassene. Vi har derimot funnet det hensiktsmessig å gå ut over ordklassene i vårt lemmatiseringsarbeid. Vårt utgangspunkt er jo å kartlegge frekvente, men sannsynligvis ukjente ord for elevene. Vi vil da måtte sette en nedre grense for hva vi regner for frekvente på de forskjellige klasseserier. Et ord med mange forskjellige avledningsformer vil da ikke komme med blant disse hvis hver av formene har lav frekvens. Det er med andre ord ikke de enkelte ordenes frekvens vi vil fange opp, men snare finne ut hvor frekvent en semantisk gruppe med lik rot er. Hvis elevene har arbeidet med ordlagning og segmentering, må man kunne anta at så snart ett av

ordene innen en slik semantisk gruppe er kjent, vil de andre ordene som dannes ved at man legger til eller trekker fra suffikser også være forståelige for elevene. Et eksempel: hvis en elev kjenner til ordet oppfinne vil han eller hun også forstå en oppfinner og en oppfinnelse ved hjelp av et visst kjenskap til ordlagingsregler. I vårt lemmatiseringsarbeid har vi derfor samlet et ord og alle avledningene som er dannet av dette ordet ved hjelp av suffikser og infikser. Vi har valgt å si at disse ordene tilhører samme rotlemma for å skille det fra lemmatisering av ord innen samme ordklasse, og prosessen kaller vi rotlemmatisering.

Det har vært nødvendig med visse restriksjoner og modifiseringer for å bestemme rotlemmatilhørigheten til ordene.

1. Når rotlemmatiseringen går ut over ordklassegrensen er det en forutsetning at uttalen eller skriftbildet til en av de bøyde formene av ordet i den ene ordklassen naturlig kjeder seg sammen med minst en av formene fra den andre ordklassen. Eks: verbet bryte og substantivet brudd vil tilhøre samme rotlemma fordi substantivformen kjeder seg naturlig sammen med partisippformen av verbet, brutt. Hvis derimot skriftbildet eller uttalen av de to ordene spriker for mye, vil vi la de tilhøre to forskjellige rotlemma. Eks: verbet gå og substantivet gange vil tilhøre to forskjellige rotlemma.
2. Avledningen må ikke forandre innholdet i noen større grad. Eks: dag og daglig vil tilhøre samme rotlemma, men ikke tid og tidlig.
3. En bøyingsrekke med suppletiv bøyning vil tilhøre forskjellige rotlemma. Eks: god vil tilhøre et rotlemma, mens bedre og best vil tilhøre et annet.
4. I forbindelse med ord som betegner nasjon eller nasjonalitet vil vi også la sammensatte ord tilhøre samme rotlemma som usammensatte. Eks: fransk, franskmann og Frankrike vil tilhøre samme rotlemma.

Som oppslagsord har vi valgt den formen som har færrest bokstaver av ordklassenes grunnform selv om denne formen ikke er belagt i tekstene. Hvis flere former har samme antall bokstaver, har vi valgt substantivformen. Det er imidlertid et unntak: hvis den korteste formen er relativt sjelden og ikke belagt i tekstene, blir den nest korteste formen oppslagsord.

Eks: tikke er oppslagsord i stedet for tikk og strekke i stedet for strek.

Ved nasjon og nasjonalitetsbetegnelser er nasjonens navn oppslagsord, uansett lengde. Det samme gjelder andre proprier.

For å spare arbeid har vi bare ført opp oppslagsord i de tilfellene der minst to forskjellige former med samme rotlemmatilhørighet er belagt i tekstene. Hvis et ord står alene innenfor sitt rotlemma, vil det ikke få eget oppslagsord.

Kriteriene for bestemmelsen av homografer må sees i sammenheng med den rotlemmatiseringen vi utfører. Fordi vi skal fram til

et rotlemmanivå, er det ikke nødvendig å skille mellom homografkomponenter som tilhører samme rotlemma selv om de tilhører ulike ordklasser.

Eks: arbeider (verb) og arbeider (substantiv) tilhører samme rotlemma og har oppslagsordet arbeid. Vi regner derfor ikke arbeider som homograf. Derimot regner vi et ord som homograf hvis det har flere betydningavarianter, selv om komponentene vil tilhøre samme ordklasse og ha identiske bøyingsformer.

Eks: bestå har forskjellig betydning i uttrykkene 'bestå eksamen' og 'bestå av noe'.

Her er det nødvendig med spesifiseringer:

1. Hvis et ord har en abstrakt og en konkret betydning og den abstrakte er en ren overføring av den konkrete, regnes ikke ordet for homograf.
Eks: framkalle vil ikke bli delt etter betydningene 'fotografisk prosess' og 'tankeprosess'. Derimot vil vi ta med eksempler både fra den konkrete og den abstrakte varianten hvis det blir aktuelt å lage øvingsoppgaver til tilsvarende ord.
2. Hvis ordet har flere betydningsnyanser som det er vanskelig å skille, regnes ikke ordet for homograf. I noen tilfeller skiller vi ut en betydning og lar resten stå sammen.
Eks: side vil ha en komponent 'side i en bok', mens andre betydninger står sammen.

Fordi homografsepareringen utføres manuelt og fordi siktemålet er å lage øvingsmateriell, har vi ikke tatt ut konkordansen til homografer hvor det er lite sannsynlig at den ene komponenten vil forekomme i fagtekstene.

Eks: skål som verb i imperativ.

Andre homografer har vi regnet som uaktuelle i et fag, men aktuelle i de andre.

Eks: pil hvor betydningen 'treslag' regnes for uaktuell i fysikktekstene.

Etter at homografsepareringen er utført og rotlemmatilhørigheten bestemt, kjører vi et program som slår sammen alle ordene innenfor samme rotlemma. Vi får ut lister som oppgir samlet frekvens for rotlemmaet og som samtidig spesifiserer de formene som er belagt i teksten og frekvensen til disse.

Eks: bor (12): bore (3), borer (1), bores (1), boring (5)
boringene (2)

Når vi har fått ut alle de 18 rotlemmatiserte listene, vil vi ta stilling til hvor høy den relative frekvensen bør være for at rotlemmaene kan regnes for frekvente. Vi vil så lage øvingsoppgaver med de aktuelle rotlemmaene, et hefte for hvert fag og hver klasse.

I O-fagsbøkene er det en del uekte sammensatte verb, eks gi opp gi ut. Fordi vi ikke får ut frekvensen av disse formene på frekvensordlistene, vil vi ta ut konkordansen til en del adverb og preposisjoner som ofte blir brukt som verbalpartilker, slik at

vi også kan fange opp de løse sammensetningene.

Når vi er ferdige med behandlingen av ordforrådet, vil vi undersøke deler av syntaksen og oppbygningen av teksten utover setningsplanet. Vi ønsker spesielt å sammenlikne tekster på forskjellige klassetrinn som behandler like eller beslektede emner. Vi har foreløpig ikke tatt stilling til i hvilken grad vi skal bruke EDB til dette formålet.

Prosjektet LÆREBOKSPRÅK er finansiert av KUD (skoleforskningsmidler). Prosjektledere er Anne Hvenekilde og Anne Golden. Addressen er : Nordisk institutt, pb 1013, Blindern, Universitetet i Oslo

Programmene vi har kjørt er beskrevet i

Ivar Fønnes: Tekstanalyseprogrammer på DEC-10
stensil, UiO, 1980

S.Allén: Nusvensk frekvensordbok 1, Göteborg 1970