

Anne Karin Ro
 Engelsk institutt
 Eirik Lien
 Edb-tjenesten for humanistiske fag
 Universitetet i Trondheim

WYCLIFFES BIBELTEKSTER PÅ RA2

Bakgrunn for prosjektet

Prosjektet som skal beskrives her, datamaskinell registrering av ortografiske variantforekomster i 12 håndskrifter inneholdende The General Prologue to the Wycliffite Bible, med henblikk på datering og lokalisering av håndskriftene, er et ledd i et større forskningsprosjekt som pågår ved Engelsk institutt i Trondheim under ledelse av professor Conrad Lindberg. Det utvidede prosjektet har to hovedsiktemål: (1) en nyutgivelse av The Middle English Bible, også kalt The Wycliffite Bible eller The Lollard Bible, en bibeloversettelse fra latin til senmiddelengelsk; (2) inngående studier av ca. 230 bibelhåndskrifter for om mulig å tilveiebringe mer presis informasjon om oversettelsesarbeidet som tilkjennes den engelske reformator John Wycliffe og hans læresvenner.

En fullstendig utgave av The Wycliffite Bible, basert på ca. 170 håndskrifter, ble utgitt i 1850 av J. Forshall og F. Madden, som hadde brukt 20 år på å undersøke håndskrifter lokalisert i Storbritannia og Irland. Forshall og Madden fant at håndskriftene kunne inndeles i to grupper; den ene gruppen oppviste en nesten ordrett oversettelse fra den latinske originalversjonen, den andre gruppen presenterte en oversettelse der idiom og syntaks samsvarte med engelsk språkbruk rundt 1400. (Bibeloversettelsen antas å være utført i perioden 1370-1395.) Forshall og Madden opererte derfor med to versjoner av bibeloversettelsen; en bokstavelig, tidlig versjon (Early Version) og en friere, senere versjon (Later Version), og deres 1850-utgave presenterer versjonene i parallelle kolonner på samme side.

Forshall og Maddens teorier om tilblivelsen av bibeloversettelsen ble ikke alvorlig utfordret før den svenske forskeren Sven L. Fristedt publiserte sin doktoravhandling om The Wycliffite Bible i 1953. Fristedt fant at håndskriftene på ingen måte lot seg gruppere så entydig i en tidlig og en senere versjon som Forshall og Madden hadde antydnet. Det fantes håndskrifter som delvis samsvarte med den ene versjonen, delvis med den andre, og Fristedt mente at disse håndskriftene kunne representere en mellomliggende gruppe. Fristedts teori

gikk da ut på at det forelå ikke to separate oversettelser, snarere én original oversettelse (Early Version) med stadige revideringer, som i siste instans munnet ut i den senere versjon. Spørsmålet om antall versjoner ville imidlertid bare kunne svares på etter inngående analyser av hvert enkelt håndskrift og detaljerte sammenlikninger av alle håndskriftene. Fristedt påpekte også en del svakheter ved 1850-utgaven, som gjorde den uegnet for slike analyser; bl.a. hadde Forshall og Madden modernisert ortografi og tegnsetting slik at den redigerte teksten til en viss grad avvek fra basehåndskriftet.

Flere av Fristedts synspunkter ble delt av Conrad Lindberg, som siden 1952 har arbeidet med tekststudier av bibelhåndskriftene, primært med håndskrifter som Forshall og Madden klassifiserte som Early Version. I 1970 publiserte Lindberg i Studia Neo-Philologica en oversikt over kjente eksisterende Wycliffe Bible håndskrifter, og kunne vise til et antall på ca. 230, altså ca. 60 flere enn Forshall og Madden hadde undersøkt. Dermed forelå et større, ukjent materiale, som ved nærmere undersøkelse kanskje kunne gi svar på en del av spørsmålene i forbindelse med tilblivelsen av The Middle English Bible, i første rekke om hvor mange versjoner oversettelsen kan inndeles i; datering av de ulike håndskriftene, som muligens kan føre til en angivelse av den innbyrdes rekkefølgen av dem; og dialektal bestemmelse av håndskriftene, som muligens kan føre til geografisk lokalisering av dem og gi en pekepinn om opphavsmennene bak de forskjellige revideringene. I sitt arbeid med bibeltekstene har Lindberg kommet til den konklusjon at bare gjennom en detaljert sammenlikning av alle bibelhåndskriftene når det gjelder ortografi, tegnsetting, morfologi, vokabular og syntaks, i tillegg til en sammenlikning av håndskriftene med den latinske originalversjonen, kan man håpe å komme fram til antakelige løsninger på disse problemene.

The General Prologue: Innhold og betydning

Den delen av The Middle English Bible som er i fokus her, The General Prologue, skiller seg ut ved at dette ikke er en oversatt tekst, men en genuin middelengelsk tekst, på ca. 40 000 ord. Teksten er således spesielt interessant, fordi den både kan kaste lys over bibeloversettelsen og gi verdifull informasjon om senmiddelengelsk fra en språkhistorisk synsvinkel. The General Prologue er oppdelt i 15 kapitler og er i hovedsak en oppsummering av innholdet i de gammeltestamentlige bøker, ispedd argumenter for nødvendigheten av en bibel på morsmålet - et omstridt spørsmål på den tida -, og ender med en relativt detaljert beskrivelse av de prinsipper og metoder som ligger til grunn for det oversettelsesarbeidet som prologforfatteren har vært hovedansvarlig for.

The General Prologue finnes i 12 av de ca. 230 håndskriftene man kjenner til; de fleste av disse håndskriftene tilhører gruppen klassifisert av Forshall og Madden som Later Version,

men noen er knyttet til håndskrifter som er en blanding av Early Version og Later Version. Ett håndskrift inneholder utelukkende The General Prologue. Ikke alle håndskriftene presenterer den fullstendige prologteksten; her følger en oversikt over de håndskrifter som inneholder i det minste ett fullstendig kapittel:

- A. MS W.H.Scheide 12, Princeton, N.J. (fullstendig)
- B. MS CUL Mm.2.15, Cambridge (fullstendig)
- C. MS CUL Kk.I.8, Cambridge (kap.1-9, 11-15)
- D. MS Corpus Christi College 147, Cambridge (fullstendig)
- E. MS University College 96, Oxford (fullstendig)
- F. MS Bodley 277, Oxford (kap.1)
- G. MS Lincoln College Latin 119, Oxford (kap.1-9, kap.11)
- H. MS Royal I.C.VIII, British Library (kap.1)
- I. MS Harley 1666, British Library (kap.1-15, men 15 fragmentarisk)
- J. MS Trinity College Dublin 75, Dublin (fullstendig)

I tillegg finnes fragmenter av The General Prologue i MS Worcester Cathedral F.172 (K) og i MS British Library 10.046 (L).

Skissering av prosjektfaser

I studiet av håndskriftene har jeg funnet det formålstjenlig å dele prosjektet i to faser:

1.fase: Tekstredigering (avsluttet)

Siktemålet i denne fasen var å utarbeide en nyutgave av The General Prologue, basert på MS W.H.Scheide 12 (A), et håndskrift som ikke tidligere er kollatert av Forshall og Madden. Alle håndskriftene er blitt sammenliknet og varierende lesearter innen morfologi (hvor grammatisk signifikant), vokabular og syntaks notert i fotnoter. I tillegg forekommer språklige analyser av varierende lesearter. Det er også foretatt en detaljert sammenlikning med Forshall og Maddens utgave, slik at feilforekomster i 1850-utgaven er blitt rettet opp. En foreløpig glossar fra middelengelsk til moderne engelsk er blitt utarbeidet.

2.fase: Datamaskinell registrering av alle ortografiske variantforekomster i håndskriftene (under arbeid)

Det fantes ingen enhetlig nasjonal ortografisk standard i middelengelsk. Middelengelske håndskrifter reflekterer derfor regionale lingvistiske særdrag og kan således gi verdifull informasjon om dialektale forhold. Av denne grunn kunne bibelteksternes ortografiske og grammatisk insignifikante morfologiske varianter vise seg å være av betydning når det gjaldt å bestemme den dialektale distribusjon av håndskriftene. Ifølge artikler publisert av A. McIntosh (Edinburgh University) og M.L.Samuels (Glasgow University) forefinnes utarbeidet

detaljerte senmiddelengelske dialektkart som muliggjør datering og geografisk lokalisering av senmiddelengelske håndskrifter, ved å bruke den såkalte 'fit-technique'. Dialektkartene er hittil upubliserte, men tilgjengelige for forskere ved henvendelse til The Middle English Dialect Project ved Edinburgh University. Den såkalte 'fit-technique' forutsatte imidlertid at man utarbeidet en 'Linguistic Profile', d.v.s. en, fortrinnsvis fullstendig, analyse av ca. 300 forskjellige grafsekvenser for hvert håndskrift, og det ble derfor besluttet å foreta en registrering og analysering av alle variantforekomstene i The General Prologue. Ettersom tekstmaterialet var såvidt omfattende, over 300 000 ord, ble en datamaskinell registrering foretrukket framfor en manuell bearbeiding av variantene.

Edb-opplegget

To problem måtte løses når vi først hadde valgt å bruke edb:

- hvordan skal dataene overføres til maskinleselig form
- hvilken metode skal brukes for å sammenlikne skrivevarianter av samme ord fordelt på de enkelte manuskriptene

Det er klart at registreringsarbeidet ville bli enormt om vi skulle skrive inn de 12 aktuelle manuskriptene fortløpende og hver for seg. Fordi ikke alle de 12 tekstene er fullstendige og fordi de fullstendige tekstene ikke nødvendigvis behøver å ha samme antall ord, ville vi etterhvert få problemer med å vite om vi er kommet like langt i alle tekstene. Dessuten ville mange ord som - tross alt - er like i de 12 tekstene bli skrevet 12 ganger i stedet for en gang. Her blir det et stort skrivearbeid (40 000 ord x 12 - manglende deler) - la oss si ca. 300 000 ord - og enhver effektivisering vil gi gunstige utslag.

Vi valgte i stedet å bruke en annen metode, å snu det hele "på hodet". Ved å gjøre det oppnår vi at vi kan bruke vår egen innsikt til å si hvilke ord som hører sammen og ikke overlate til maskinen å kontrollere om vi er kommet like langt i alle 12 tekstene til enhver tid.

Men vi har ennå ikke oppnådd noen rasjonaliseringsgevinst. Den får vi hvis vi nå ser nærmere på dette. Her er det nemlig lett å se hvilke ord som skrives likt og hvilke manuskripter de er knyttet til. La oss "nummerere" manuskriptene A,B,...,L. Vi kan da knytte denne koden til de aktuelle skrivevariantene og slippe å skrive flere ord enn nødvendig. Vi har ikke regnet noe på hvor mye vi har spart på dette, men et raskt overslag viser i hvert fall at vi har redusert skrivearbeidet med bortimot 2/3. Det gir jo en betydelig arbeidsbesparelse når innmatingen blir på omkring 100 000 ord i stedet for ca. 300 000!

I tillegg er det knyttet til tagger som bærer av grammatisk informasjon. De er alltid markert i dataene med \$. Disse taggene gjør det også mulig å skille mellom likt skrevne ord som har forskjellig betydning og funksjon (homografer).

For å få oversikt over de forskjellige skrivevariantene av samme ord har vi laget et programsystem som består av disse tre programmene:

delogsort

lager to ordlister

1. Hver eneste skriveform får knyttet til seg et løpenr som forteller i hvilken rekkefølge i teksten det kommer. Alle skrivevariantene av samme ord på samme sted i de 12 tekstene får samme nr. I tillegg blir versjonskoden knyttet til.
2. En alfabetisk liste over alle ord med tilknyttet løpenr.

I tillegg lager programmet en utlisting av alle ordene den fant i den aktuelle teksten, en inventarisering. Den er utgangspunkt for å lete etter ord.

input

mater inn de to listene i databasen,

- den ene slik at det er mulig å starte med et bestemt løpenr. og finne alle skrivevarianter knyttet til det, og med manuskriptkoder koplet til de enkelte skrivevariantene
- den andre slik at det er mulig å starte med en bestemt skrivevariant og se hvilke løpenr. som er knyttet til den.

For å legge opp databasen bruker vi databasesystemet RA2, som er utviklet ved Regnesentret ved Universitetet i Trondheim.

letord

får som inndata de "kanoniske" formene som vi vil ha som oppslag. En kanonisk form er den skrivevarianten av et ord som vi ønsker å oppfatte som felles form. Maskinelt har det ingen betydning hvilken form vi velger, men rent praktisk vil vi vanligvis velge den formen som ligger nærmest opptil skrive-normen idag.

Programmet tar det første ordet i denne lista og slår opp i den alfabetiske delen av databasen. Der finner den alle løpenr. knyttet til den skriveformen, og legger dem på en egen arbeidsliste. Så går programmet til det første løpenummeret på den lista og finner der de aktuelle skrivevariantene som er knyttet til. Disse variantene, samt manuskriptkode legger den på en annen arbeidsliste. Den starter så på toppen av denne lista og går inn i den alfabetiske delen igjen for å se om det er knyttet andre løpenr. til den varianten. Hvis det er det, kompletteres arbeidslista med løpenr. Slik gjør den med alle variantene på denne ordlista.

Hvis det er kommet til nye løpenr., går programmet til løpenummerdelen av databasen, og ser etter om det er nye skrive-

varianter der som ikke er registrert.

Slik veksler den mellom oppslag i de to listene, kontrollerer og eventuelt supplerer arbeidslistene til alle mulighetene er sjekket. Da har den grunnlag for å skrive ut en oversikt som kan se slik ut:

x x x x	ALL		DET								
	A	B	C	D	E	F	G	H	I	J	SUM
ALLE	6	6	6	7			5		7	14	51
ALL					16					2	18
AL	11	11	11	10	1		12		10	1	67
											136

Her er alle skrivevariantene knyttet til versjonskode og summert opp. Det sier seg selv at med store datamengder blir dette tidkrevende, selv med raske og effektive dataanlegg. Mange av oppslagene er dessuten resultatløse, men det vet vi ikke før oppslaget er gjort.

Resultatet må sies å være godt, programmet letord er i stand til å finne de enkelte skrivevariantene av samme ord. Tidsforbruket er akseptabelt, på et UNIVAC 1100/62-anlegg bruker programmet i underkant av 10 minutter cpu-tid til å gå gjennom et "gjennomsnittskapittel" og sette opp tabellene. Et slikt kapittel har ca. 3 500 ord i løpende tekst, ca. 2 500 graford (med tagger inkludert) og ca. 750 kanoniske former.