

Apertium-omsette tekstar

		Ord / feil*	Prosent	Markert**	Feiltype
1	Omtale fagbok, nettbutikk (Norli)	230 / 5	2,2 %	2 (av totalt 5)	Samansette ord (2) samsvarsbøying (1), feil omgrep (1), kjønn (1)
2	Omtale lærebok, forlag (Univ.forlaget)	148 / 5	3,4 %	2 / 5	Samansette ord (4), ordstilling (1), feil omgrep (1)
3	Informasjon, kommunal heimeside (Oslo)	409 / 13	3,2 %	8 / 13	Samansette ord (9), samsvarbøying (1), kjønn (1), feil omgrep (1), bm (1)
4	Informasjon, kommunal heimeside (Molde)	226 / 2	0,8 %	0 / 2	Feil ordstilling (2)
5	Informasjon, Utdanningsdirektoratet	139 / 1	0,7 %	0 / 1	Samsvarsbøying (1)
6	Elevtekst, eksamensoppgåve vg3(?) (Skrivesenter)	782 / 9	1,2 %	4 / 9	Mange bm-språkfeil. Samansejting (4), feil omgrep/homonym (4), bm (1)
7	Elevtekst, vg3, lærebok (Intertekst 3)	544 / 6	1,1 %	4 / 6	Samansette ord (4), homonym (2)
8	Elevtekst, 10. trinn, nettressurs (norsksidene.no)	637 / 5	0,8 %	4 / 5	Samansette ord (4), homonym (1)
9	Leksikontekst om nasjonen Noreg, utdrag (snl)	2180 / 35	1,6 %	16 / 35	(Ikkje kategorisert feiltypar)
	TOTALT	5295 / 81	1,5 %	40 / 81	

Dei ni tekstane er tilgjengelege i eige vedlegg der feila er markerte.

* Ord/frasar som ikkje er i samsvar med nynorsk rettskriving og ordstilling (manuell språkvask).

** Feilregistrerte ord som er markerte med stjerne i Apertium som «ukjent ord» eller med raud krøllstrek i nynorsk word-korrektur.

Utvalde døme på feil i Apertium-omsette tekstar

Samansette ord	Feil omgrep og/eller ordstilling (bm/nyn)	Homonym (bm – nyn)	Anna
tilkoblingsmekanismer - luftepauser	biologiens betydning for... / tydinga for biologien	å bite – ein bit	stedvis
allmenntilgjengeleg - flettebåndet	samtidig framheves og drøftes / samtidig blir framheva og drøfta	å henge – ein hengar	eit augneblink
forvaltningsrettslig - kallenavnet	kan ikke avholdes / kan ikkje haldast	smalt/smalt – smalt/small	til vanns
hovudtrekkane - inngangsdøren	samlinger frarådes / samlingar blir rådde mot	en by/byer – ei bye/byer	bevilgande
vassdragslovgivningen - påmalt	...og du/dine pårørende deltar aktivt/du, pårørende dine deltar...	bm: ei seter, to (stol)seter	urørlig
læresteder - Golfstrømmen	midt foran sees Brynjulf... / midt foran sest Brynjulf	nyn: ei seter, to (stol)sete	godkjennelse
svømmehaller - treningssentre	tidsrommet [...] betegnes som... / tidsrommet [...] blir kalla som	lenger/lengre	Reg. som særnamn pga. stor bokstav i løpande tekst
innendørsanlegg - smittevernsreglene	landet fikk sitt første felles lovverk / landet fekk dei første felles lovverka sitt	en pike/piken - ein pik/piken	
egentrening/egenorganisert	antallet fylker og kommuner ble redusert... / talet fylkar [...] vart reduserte	hvor/hvor – kvar/der	Oppgave - Julen
eneveldet – innskåret - urørlig	i fjellet forekommer ofte sterk vind / i fjellet hender ofte sterk vind	en regel/ei regle – regler	Fugler - Amfibier
guttekull - forhåndsbestemt	for mi skuld? / for skulda mi?	å se/så – så [og så stort]	Noen
virvelløse – vannflora - ferskvannsfisk	det gjorde meg glad, bunn ulykkelig / det gjorde meg glad, botn ulykkeleg	å leke/leke – å leike/leke	

Oppsummering og vurdering

Omsetjinga frå bokmål til nynorsk er gjort 1.12.2020 ved hjelp av gratisversjonen av Apertium. Denne er fritt tilgjengeleg på nettet.

Grunnlaget for analysen er ni tilfeldig utvalde bokmålstekstar. Litt ulike sjangrar, litt ulik lengde. Seks av tekstane (1-5 og 9) er skrivne av det ein kan kalle «profesjonelle» forfattarar. Dei tre siste (6-8) er skrivne av skuleelevar.

Tekstmengda som ligg til grunn for kartlegginga mi er lite. Eg har heller ikkje gått spesielt grundig gjennom materialet. Tala i tabellen over bør difor ikkje leggest til grunn for meir generelle konklusjonar om kvaliteten på omsetjingsprogrammet Apertium. For å seie noko sikkert om feilprosent og feiltypar må ein både ha eit større og meir variert utval av tekstar og meir systematisk registrering.

Med dette atterhaldet: Apertium pr. desember 2020 ser ut til å vere eit godt hjelpemiddel for å omsetje prosatekstar frå bokmål til nynorsk. Ein del manuell språkvask trengst, men ikkje mykje. I dei fleste høve vil dette ta berre ein brøkdel av tida samanlikna med å omsetje tekstane heilt manuelt.

Feilprosenten eg har registrert i tekstutvalet mitt er svært låg – frå 0,7 til 3,4. Om lag halvparten av feila vart dessutan automatisk markerte i nynorskversjonen, anten med Apertium-generert *stjerne (merking av «ukjende ord» kan veljast i portalen) og/eller korrekturstrek frå word. Dette lettar den manuelle korrekturen.

I tabellen over har eg samla og kategorisert ein del av feila eg fann i dei ni tekstane mine. Nesten alle kan forklarast:

- Gratisversjonen av Apertium omset ikkje samansette ord. Desse finst ikkje i ordlistene og blir difor ståande uendra. Men dei blir merka med stjerne.
- Homonym (dvs. to ord med ulik tyding men med same skrivemåte) gir også ein del feil. Menneske tolkar samanhengen eit ord står i og vel rett variant. Det gjer ikkje datamskina.
- Apertium hoppar over nokre ord dei sjølv om dei står i ordlista. Dette skjer fordi dei byrjar med stor bokstav utan punktum, kolon eller spørjeteikn rett føre. Apertium tolkar dette som særnamn som ikkje skal omsetjast. Slike feiltypar fann eg både i tekstar der stor bokstav er tilsikta (m.a. i snl-artikkelen) og i tekstar med manglande teiknsetjing og tilfeldig bruk av stor bokstav (ein av elevtekstane). Slike feil er elles lette å finne sidan dei blir fanga opp av word-korrekturen.
- Det er ein del døme på «feil» ordstilling eller uttrykksmåte. Her er det ikkje snakk om feil omsetjing ord for ord, men at nokre frasar i det ein kalle «god nynorsk» skil seg frå tilsvarende formuleringar på bokmål. Her det sjølv sagt eit visst rom for skjøn. I denne analysen har eg lagt meg på ei liberal linje, dvs. berre talt «klare» nynorskbrøt.
- Kjønn blir av og til feil når det kjem eitt eller fleire ord mellom artikkelen og substantivet eller når substantivet er samansett og dermed ikkje omsett.

Dei fleste av dei feiltypane som er nemnde over kan reduserast sterkt ved hjelp av manuell programmering eller maskinlæring. Dvs. at databasen lagrar rett staving av samansette ord og rett omsetjing av samansette frasar. Mange feil knytt til homonym og særnamn/allmennord kan truleg også lukast ut ved hjelp av slik teknologi.

Til slutt ei opplagd påminning: At ein maskinomsett nynorsktekst inneheld få eller ingen språkfeil, betyr ikkje at det er ein god tekst. Er utgangspunktet på bokmål dårleg, blir også sluttresultatet på nynorsk dårleg. Av og til endå dårlegare. Nynorsk og bokmål har som kjent litt ulike stilideal og litt ulike måtar å uttrykke meining og samheng på. Det kan ikkje Apertium og anna datateknologi hjelpe til med. Å få ei god nynorsk omsetjing av ein bokmålstekst krev difor ofte litt meir enn berre å finne dei rette orda i ordlista. Menneskeleg språkkompetanse trengst også.

Volda, 3.12.2020/13.1.2021

Torgeir Parr Dimmen