

Vejledning til korrektur af Trykkefrihedens Skrifter

Webapplikationen AltoFaxEdit benyttes til korrekturarbejdet: altofaxedit-test.kb.dk

Se brugervejledningen som er tilgængelig efter login.

Efter login kan man klikke på "Læs om fraktur" og printe den vejledning så den altid er tilgængelig.

I princippet ønskes alle OCR-fejllæsninger rettet i skrifterne, men det må accepteres at perfektion ikke kan opnås. Det er også vigtigt at arbejde med en vis hastighed for at vi kan færdiggøre alle sider.

I det følgende beskrives hvorledes forskellige problemstillinger skal håndteres.

Indhold

1. Generelt	1
2. Sidetal og kustoder	2
3. Tal	2
4. Romertal	2
5. Brøker	2
6. Kapitæler	3
7. Fejlagtigt mellemrum i et ord	3
8. Manglende mellemrum mellem ord	3
9. Anførselstegn (gåseøjne)	3
10. Tabeller	4
11. Skrifter på tysk, fransk og latin	4
12. Fjernelse/indsættelse af orddelingsstreg ved linjeskift	4
13. Tekst i margen	5
14. Markering af noget som skal efterses senere	5

1. Generelt

De originale stavemåder skal beholdes, også hvis der er stavefejl.

Fejl lavet af sætteren skal ikke rettes. Fx beholdes "Riser" hvor sætteren har sat "R" i stedet for "K" i filen ..._1_002_004_010.xml linje 15.

Fejl pga. manglende sværte skal rettes hvis det er muligt at se hvilke bogstaver der mangler.

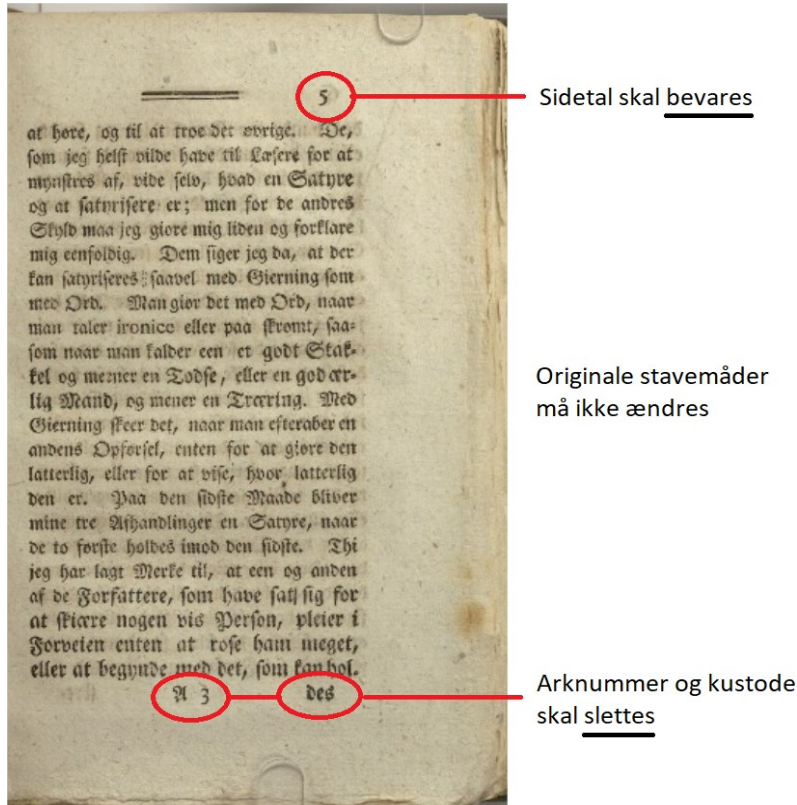
Fx "S kke" rettes til "Sukke" i filen ..._1_002_006_011 linje 12.

Bemærk, at hvis et ord er delt ved linjeskift må man ikke samle ordet. Dvs. at ordene/bogstaverne skal stå på samme linje som man ser i faksimilen (fotografiet af teksten).

2. Sidetal og kustoder

Vi må ikke slette sidetallet der står øverst på mange sider. Hvis sidetallet mangler, skal vi indsætte det. Hvis der ikke findes en tom linje øverst til sidetallet, indsættes en ny linje.

Arknummer og kustode slettes, dvs. hele linjen slettes (ved at klikke på skraldespand-ikonen).



Sidetall skal bevares

Originale stavemåder
må ikke ændres

Arknummer og kustode
skal slettes

NB: Pas på med blå orddelingsstreger ved sletning af en linje med arknummer og kustode. Se afsnit [12](#) nedenfor.

3. Tal

Det ses at tal ofte er skrevet som en blanding af bogstaver og cifre. Eksempelvis bogstav o i stedet for ciffer nul, og bogstav I (stort i) i stedet for ciffer ét. Dette skal rettes så der udelukkende er cifre. Eksempelvis rettes "I5o" til "150".

4. Romertal

Romertal bør principielt beholdes som romertal. Eksempel skrives romertal to normalt med store i'er, altså "II", og romertal fire skrives normalt med et stort i og et stort v, altså "IV".

P.t. (medio april 2020) vides det ikke om romertal kan optræde med små bogstaver i Trykkefrihedens Skrifter, dvs. "i", "ii", "iii", "iv" etc. (som ofte ses i moderne tekster). Hvis det skulle optræde, bør små bogstaver beholdes.

5. Brøker

Brøker kan optræde i teksterne. Det drejer sig fx om $1/5$, $1/4$, $1/2$, $3/4$ etc.

Nogle af disse brøker eksisterer som specialtegn på computere, $\frac{1}{4}$, $\frac{1}{2}$ og $\frac{3}{4}$, men vi må ikke bruge disse specialtegn. I stedet skal vi skrive brøker med skråstreg.

Eksempelvis $3/4$ skrives altså med tre tegn (3 skråstreg 4) uden mellemrum mellem tegnene.

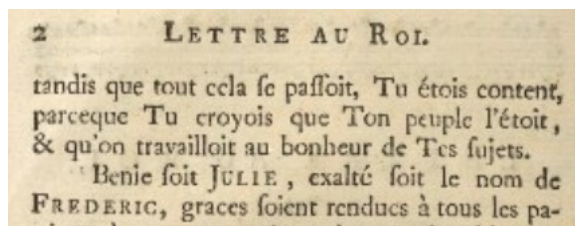
6. Kapitæler

Kapitæler (engelsk: small caps) er særlige skriftsnit opfundet i år 1501 hvor små bogstaver (minuskler) skrives med stort set samme udseende som store bogstaver (versaler), men med samme bredde, stregtykkelse og næsten samme højde som små bogstaver. Eksempel:

MODERNE KAPITÆLER

I Trykkefrihedens Skrifter har OCR-softwaren typisk læst et ord skrevet med kapitæler som et stort bogstav efterfulgt af små bogstaver. Kapitæler kan godt vises på moderne computere, men desværre ikke i Trykkefrihedens Skrifter.

For korrekturen af Trykkefrihedens Skrifter har forskerne valgt at vi skal rette kapitæler til udelukkende store bogstaver. Eksempel fra side ..._2_010_008_002:



Således er det valgt at vi skal rette OCR-resultatet "Lettre au Roi." til "LETTRE AU ROI.", og "Benie soit Julie, exalté soit le nom de Frederic" rettes til "Benie soit JULIE, exalté soit le nom de FREDERIC".

7. Fejlagtigt mellemrum i et ord

Hvis OCR-softwaren fejlagtigt har indsat et mellemrum i et ord som if. faksimilen tydeligvis er sammenskrevet, skal vi fjerne mellemrummet. Det gøres vha. en af de to ordsammenlægningsknapper som ses når man klikker på et ord (i højre/venstre side af indtastningsfeltet).

NB: Det er vigtigt at bruge disse knapper i stedet for at slette tegn i den ene "celle" og indskrive tegnene i "cellen" ved siden af. Ved at bruge knapperne vil den skjulte information om hvor ordet findes på faksimilen nemlig blive justeret korrekt (den grønne ramme i zoom-feltet).

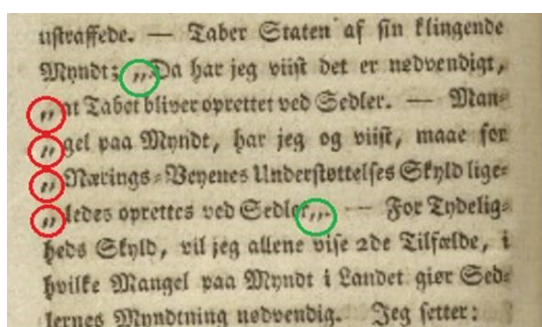
8. Manglende mellemrum mellem ord

Hvis OCR-softwaren fejlagtigt har sammenskrevet to ord som if. faksimilen tydeligvis er adskilt, skal vi indsætte et mellemrum på rette sted ved at indtaste et mellemrum (med mellemrumstangenten).

9. Anførselstegn (gåseøjne)

Anførselstegn anvendes ofte til at markere citater. I Trykkefrihedens Skrifter er der ikke blot anførselstegn i begyndelsen og slutningen af citatet, men længere citater er typisk markeret med et citationstegn i linjestart (og evt. i linjeslut).

Sådanne anførselstegn skal håndteres som forklaret ved dette eksempel:



Ved citater

beholdes anførselteget ved start og slut (grøn markering)

mens de mellemliggende anførselstegn i margen/linjestart slettes (rød markering).

10. Tabeller

Vi ikke skal lave korrektur af tabeller. Så selve tabellerne springer vi simpelthen over. Men tekst over/under tabellerne skal rettes.

Også på sider hvor der er tabeller skal vi indsætte en ny linje øverst med sidetal hvis det mangler, og vi skal også slette den nederste linje med arknummer og kustode hvis de findes.

11. Skrifter på tysk, fransk og latin

Nogle skrifter (hele hæfter) er forfattet på tysk og fransk. Vi skal springe dem over (medmindre man eksplicit er blevet tildelt sådanne skrifter). Der skal altså ikke laves nogen korrektur på dem. Man må dog gerne lave korrektur på forsiden af en skrift (titel) hvis man er i stand til at forstå hvad der er skrevet på det pågældende sprog.

Men hvis et skrift er forfattet på dansk med ord eller små segmenter på tysk, fransk eller latin skal vi også lave korrektur af disse ord/segmenter.

12. Fjernelse/indsættelse af orddelingsstreg ved linjeskift

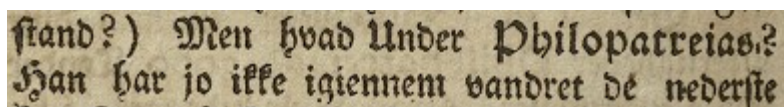
OCR-filerne (XML-filerne i ALTO-format) indeholder information om automatisk identificerede orddelingsstreger ved linjeslut. Det ses ved at en blå bindestreg i OCR-teksten når ordet i linjeslut ikke er i redigeringsmodus:

Aarsagen, hvorfor vore høyløvlige Kon-
ger har gjort alt dette, veed hver Borger og

I nogle tilfælde har OCR-softwaren indsat en blå orddelingsstreg fejlagtigt. P.t. har applikationen AltoFaxEdit ikke en funktion til at slette disse blå orddelingsstreger, så i stedet markerer vi dem som skal slettes med et særligt tegn. Derved kan vi efter korrekturarbejdet er afsluttet anvende en søg-og-erstat-funktion (anden software) til at rette problemet.

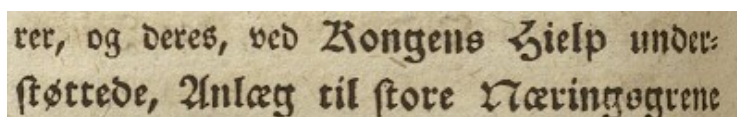
Men pas på! Faktisk slettes en blå orddelingsstreg i den foregående linje hvis man sletter en linje. Det gælder fx når man sletter linjen med arknummer og kustode nederst på siden. I et sådant tilfælde må man så selv indsætte den slettede orddelingsstreg (en bindestreg), hvis den skal være der.

Vi markerer at en blå orddelingsstreg skal slettes ved at indsætte @ (snabel-a) allersidst i linjen. Eksempel fra filen ..._1_002_005_028.xml linje 27:



stand?) Men hvad Under Philopatreias?
Han har jo ikke igiennem vandret de nederste

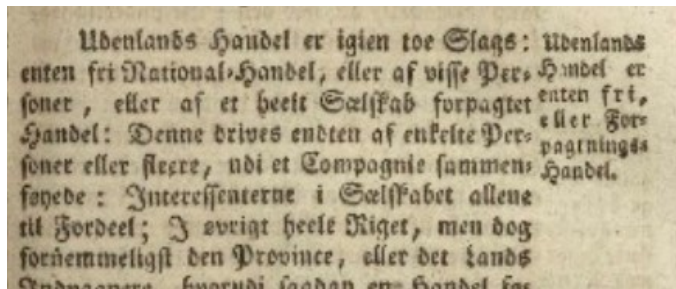
Det modsatte tilfælde ses også. I nogle tilfælde har OCR-softwaren ikke forstået at der var orddeling ved linjeskift. Tilføjelse af orddeling ved linjeskift håndteres ved at indsætte en afsluttende bindestreg. Fx i filen ..._1_003_003_078.xml linje 10 her med vist med rødt hvor der er rettet. Der er manuelt indsat en bindestreg efter "under":



rer, og deres, ved Kongens Hielp **under-**
støttede, Anlæg til store **Næringsgrene**

13. Tekst i margen

Tekst i margen (ikke håndskrift) optræder på visse sider. Eksempel fra side ..._1_007_005_017:



OCR-softwaren vil blande teksten i margen sammen med hovedteksten. Dette bliver uforståeligt i forskellige visninger på web. Derfor er det besluttet at vi skal sørge for at flytte margentekst som det passer bedst til en linje (eller flere linjer) før eller efter den matchende hovedtekst. Margenteksten skal adskilles med kantede parenteser. Her er resultatet fra ovenstående eksempel:

[Udenlands Handel er enten fri, eller Forpagtnings-Handel.]
 Udenlands Handel er igien toe Slags:
 enten fri National-Handel, eller af visse Per-
 soner, eller af et heelt Sælskab forpagtet
 Handel: Denne drives endten af enkelte Per-
 soner ...

Bemærk, at den smalt opsatte margentekst typisk har mange orddelinger ved linjeskift. Sådanne bindestreger er vi nødt til at fjerne efter bedste vurdering når vi flytter margenteksten ind i den brede hovedtekst.

NB: Hvis der skulle optræde tilsyneladende relevant håndskrevet tekst i margen, skal den ikke indskrives i korrektoren. Hvis man mener at håndskriften er vigtig, kan man vælge at notere "håndskrift i margen" som en kommentar i regnearket som vi arbejder ud fra. Der er op til hver medarbejder at vælge hvad der skal noteres som kommentar i regnearket.

14. Markering af noget som skal efterses senere

Fordi der kan være tilfælde som er svære at forklare, og fordi det er nemt at gennemse alle XML-filer for en unik tegnsekvens har vi vedtaget at der kan indskrives en særlig tre-tegnkode til at markere særtilfælde som Bo (boks@kb.dk) og evt. Ulrik Langen (frw104@hum.ku.dk) skal studere efterfølgende. Koden er:

#%=

Koden kan indsættes hvor som helst fordi den senere vil blive fjernet overalt i alle filer med en generel søg-og-erstat-operation

Bemærk, at det er meget vigtigt at der ikke er indsat mellemrum mellem de tre tegn.

OBS: Ved indsættelse af koden beder du hver gang Bo om at finde og åbne den aktuelle fil, og det er tidskrævende. Så vær rar at bruge koden sparsomt og med omhu.