

## Problem:

Du står med en tjekkisk tekst i en pdf fil (eller billedfil), du vil gerne oversætte den, men en billedtekst kan ikke læse af en oversætter, den er netop et billede ikke en redigerbar tekst.

*Dyre pdf programmer kan klare pdf tekster, der er billeder af tekst, men det er ikke altid muligt med en gratis pdf læser som Acrobat Reader*

## Billede af tekst til redigerbar tekst

**Optical character recognition** eller **optical character reader** forkortet **OCR**.

OCR er et program, som kan omsættet et billede af en skrift ( enten som billedfil eller som pdf fil) til en redigerbar tekst. Der findes en del betalingsprogrammer men også gratis løsninger. Man skal aldrig forvente en fejlfri tekst

OCR programmet kan levere en tekst i Word format eller lignende eller blot som ren tekst, som man så må kopiere over i ens tekstbehandlingsprogram. Dyre købeprogrammer kan håndtere billeder af tabeller m.m, som afleveres som regnearksfiler.

Her omtales et gratis program **gImage Reader**.

Selve programmet er en grafisk frontdel (betjening) byggende på Tesseract (Google gratis programdel) som normalt ikke kan køre under Windows, men her er backbone. Tesseract er et ældre program, men stadig under opdatering. Tesseract arbejder med et billede af teksten og laver det om til redigerbar tekst, derfor skal man have en fil til hvert sprog, da bogstaver/læsetegn varierer stærkt.

gImageReader hentes på:

<https://sourceforge.net/projects/gimagereader/>

Download og derpå installerer det ved klik. Der kommer en ikke en ikon på skrivebordet, men det ordnes let ved at åbne start med venstreklik (nederste venstre hjørne) her står under "tilføjet for nylig" gImage Reader, træk det ud på skrivebordet. Selve programmet ligger i normal programbibliotek c:\Program Files\gImageReader ... . Tesseract bliver samtidig installeret. Når programmet er installeret er det kun brugbart til engelsk, eksempelvis kender det ikke mønstrene for æ ø å. Man skal selv hente tilføjelser for hvert sprog man vil arbejde med.

## Valg af sprog.



Dette skal gøres som administrator ellers er der ikke adgang til at fylde underbiblioteket med sprogfilerne til programmet. Hos mig ligger disse i:

c:\Program Files\glImageReader\share\tessdata\dan.traineddata. (dansk sprogfil)

I denne situation skal glImageReader startes af **administrator**! (dvs person med ret til at ændre og arbejde i programområdet). Normalt er du som ejer også administrator, men du arbejder ikke ved normal start som administrator. Når du møder et vindue eksempelvis ved indlæsning af nyt program, som spørger, om du vil tillade at læse program ind, så skifter du over til administrator tilstand.

Du starter glImageReader som administrator ved at **højreklikke** på ikonet for programmet og i den fremkomne højre menu vælges .... **Kør som administrator**.

1. PC skal have netkontakt - de supplerende sprog skal hentes fra nettet,
2. Start programmet som administrator
3. Når glImageReader er klar, vælges HJÆLP ikon med værktøj yderst oppe til højre
4. Her vælges Manager language. Der fremkommer et vindue med sprogvalg
5. Der er hak ved eng, sæt hak ved dem som du vil have med selvfølgelig dan
6. Derefter apply og vent, det kan tage tid.

Bemærk at manualen ligger som punkt samme sted

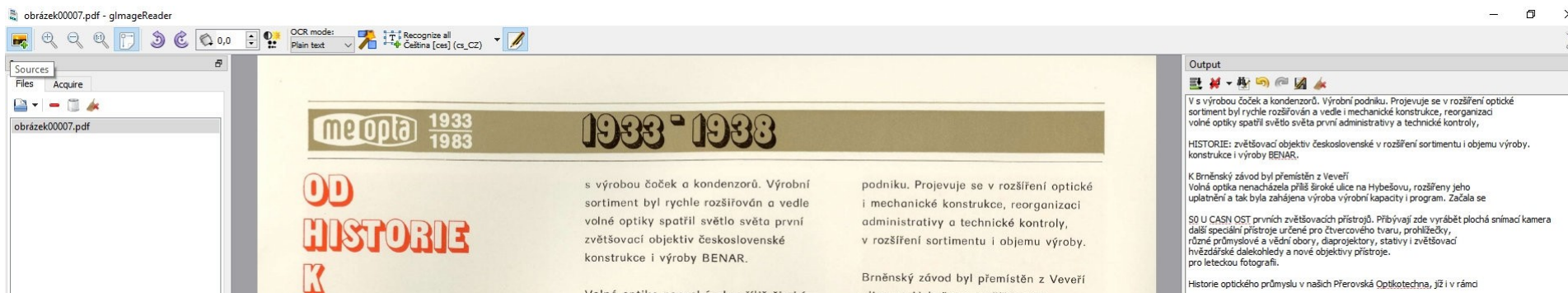
Du kan komme ud for, at programmet ved åbning af et tekstbillede meddeler med en gul bjælke, at der mangler f.eks. et orddelingsbibliotek. Så må du gå tilbage til en administratoråbning gentag åbning af tekstbillede og ved bjælken klikkes på installer.

NB

*Hvis du går frem og tilbage mellem glimageReader og Tekstbehandling (Word) kan du komme ud for at tekstbehandling fryser (blokerer). Du kan dog lukke tekstbehandlingen og husk her at gemme. Det er kendetegn på, at glimageReader egentlig ikke er et Windows program eller måske et amatørprogram.*

Hvis du vil have lavet OCR på din tjekkiske manual skal du under punkt 5 også have hak ved ces og du skal have orddelingsfilerne indlæst.

1. Start glimageReader normalt.
2. Åbne din billedfil/pdf fil
3. Sikre at sprog er korrekt valgt til ces
4. Klik på firkant med T ved sprogvvalg
5. I højre del fremkommer nu den redigerbar tekst som "ren tekst" du kan skubbe med midtergrænsen mellem billedtekst og redigerbar tekst
6. Du kopierer i den redigerbar tekst over i dit tekstbehandling eller gemmer som fil

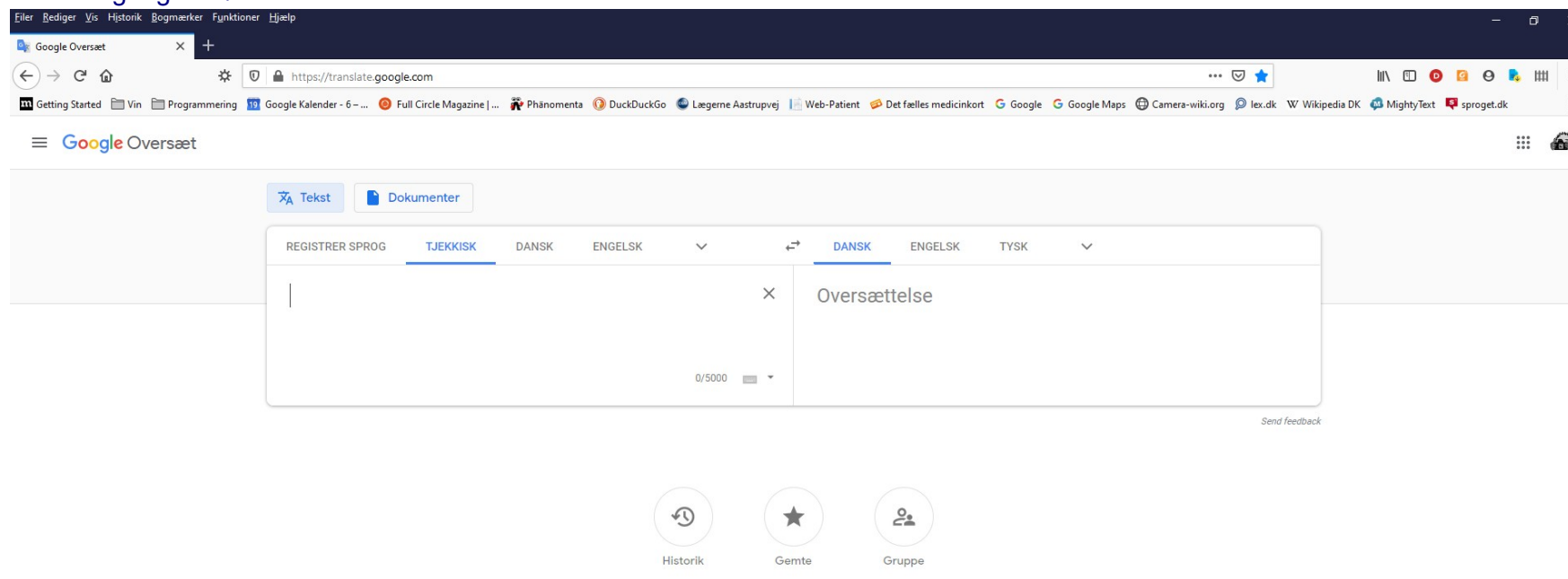


## Oversættelsen

Når du har den redigerbar tekst på tjekkisk, kan du bruge Google Translater til at lave en acceptabel oversættelse til dansk. Den redigerbar tekst skal enten ligge i kopi i clipbord eller som en fil på disken.

I din browser går du nu til:

<https://translate.google.dk/>



Her vælger du sprog til venstre hvorfra du skal oversætte – her tjekkiske  
Til højre vælger du sprog, hvortil du skal oversættes – dansk

Enten:

Har du teksten liggende på clipboard vælges tekst foroven og nu kopieres clipboard direkte ind til venstre (ctrl + v), oversættelsen kommer løbende i en dansk version til højre.

Eller:

Har du teksten i en fil, vælges foroven Dokument, vælg knappen *Find fil på din computer*. Så finder du filen, og vælger den. Klik på *Oversæt* hvorpå den danske tekst kommer på en ny side i browseren.

Den danske tekst kan altid kopieres videre. Marker det hele (ctrl+a) og vælg *ctrl+c* så har du den danske tekst i clipboard, som så kan sættes ind i tekstbehandlingprogram.

i.

Kopiere mellem programmer af tekst (og billeder m.m.) bruger clipboard et ikke synligt lager i PC'en. En tekst der skal kopieres markeres (træk musen over teksten den får en blå baggrund). Nu skal det markerede ind i clipboard. Det kan gøres direkte ved **ctrl+c** eller ved højre klik på det blå område markerede og her vælge kopiere. C står for copy. Man kan også slette teksten samtidig med den gemmes i clipboard. Det gøres med ctrl+x (saks) eller højreklik og klip. Clipboard kan kun rumme et emne ad gangen. Når man så skal have teksten ind på det andet sted eller i et andet program, skal man klikke med musen på stedet, så cursor står der, hvor teksten skal ind. Nu trykkes ctrl+v og ind kommer clipboard indholdet. Man kan også højreklikke på stedet og vælge sæt ind. Clipboardet bliver ikke tømt ved den proces.