

Meget fjern læsning

Litteraturen, de store data og Google Ngrams

At finde mønstre i en tekst er ikke det samme som at læse. Men det er også svært at benægte, at etableringen af forbindelser i en tekst og til andre tekster er en del af det at læse og analysere tekster. Der ligger ikke mindst en god del af læseglæden gemt i at opdage noget genkomment, som skaber mening gennem variationer.

Begæret efter at finde og fremstille mønstre er også centralt for et af tidens store løsener, *digital humanities*, som dækker over en hel del forskellige måder, hvorpå humanvidenskaber benytter sig af digitale redskaber. Nogle af dem bruges af alle som hjemmesider, nettidsskrifter mv. Det er adgange som nu tages for givet, men som har haft enorm betydning for måden, man kan søge og dele information på. Andre *digital humanities*-tilgange er mere sofistikerede og søger eksempelvis at finde stilistiske træk i tekster, som ellers ville gå den mere eller mindre intuitivt indstillede læser forbi. Eller der eksperimenteres med at visualisere data på nye måder, både for at kunne fremstille resultater og for at kunne hjælpe med at se nye sammenhænge.

Sådanne tilgange forsøger at tage favntag med datamængder, som ligger langt ud over, hvad en menneskelig arbejdsindsats vil kunne klare. Men det er ikke alting, som blot lader sig omsætte til opgaver, der kan sættes på en formel. Der er derfor også en berettiget skepsis over for *digital humanities*, når computere stadig er meget, meget dårlige til mange ting, som mennesker er gode til. En ung studerende kan lave en temmelig god analyse af et digt eller komme med mange præcise betragtninger over en roman, mens en computer ikke kan producere noget lignende (om end der arbejdes på sagen).

I denne artikel vil jeg se nærmere på et værktøj, som Google har stillet til rådighed siden 2011, Google books Ngram Viewer, som kan have en række anvendelser i litteraturvidenskaben. Ud over at være et fascinerende værktøj, der giver indblik i, hvilket stof kulturen er gjort af, og hvordan den udvikler sig, så kan Ngram Viewer være med til at underbygge teser, som man hidtil kun har haft en mere sporadisk dækning for, og måske endnu vigtigere kan det være med til at stille nye forsknings-spørgsmål afledt af de store datas mønstre.

Very big data

Selvom *digital humanities* først har fået momentum i de senere år, og en term som *big data* ikke kunne optages på Wikipedia for blot fire år siden, så har paradigmet en lang forhistorie (Lunenfeld 2012, 8ff). De første parodier på denne tilgang til litteraturstudier er såmænd heller ikke helt nye. David Lodge beskriver i *Small World* fra 1984 en forfatter, der er gået helt i stå, efter en lingvist har kørt hans romaner gennem et program og fundet ud af, at afledte former af “grease” er det mest brugte ord hos ham ud over sprogets helt sædvanlige:

“ Grease. Greasy. Greased. Various forms and applications of the root, literal and metaphorical. I didn’t believe him at first, I laughed in his face. Then he pressed a button and the machine began listing all the phrases in my works in which the word *grease* appears in one form or another. There they were, streaming across the screen in line numbers. (Lodge 1993, 415)

Pointen er jo dog midt i morskaben og sympatien for forfatteren, at der i Lodges scenarie faktisk bliver afsløret indsigter i teksten, men dog også af en karakter, som nemt kan tage fortolkeren på et vildspor.

Beklagelsen over ikke at have tilstrækkelige værktøjer til rådighed for analyse er endnu ældre. I 1958 bemærker Claude Lévi-Strauss i det, der nærmest må læses som en ansøgning om forskningsmidler, at der må IBM-udstyr (dengang hulkortmaskiner) til for at kunne lave de strukturelle sammenligninger, som han lagde op til, og ikke blot de store tavler, som han hidtil havde benyttet sig af (Levi-Strauss 1963, 229).

Værktøjer til at se på frekvenser af ord i tekster er blandt de mest udbredte og simple og kan som i eksemplet ovenfor være med til at afsøge og underbygge hypoteser omkring brugen af ord og begreber i tekster.¹ I Jørn Vosmars disputats om J. P. Jacobsen fra 1984 er det markant, at forskeren med et suk, der kan høres ud gennem siderne, beretter om, hvordan han har talt visse ord og figurers forekomst op i *Niels Lyhne*, noget som allerede Søren Hallar i *Synselementerne i Naturskildringen hos J. P. Jacobsen* fra 1921 havde taget under grundig statistisk behandling på et tidspunkt, hvor drømmen om en computer knapt nok var mulig. Men samtidig bemærker Vosmar, at det jo kræver indsigt at finde ud af, hvordan klynger af beslægtede ord hænger sammen:

“ En væsentlig del af JPJs stilistiske særpræg kan umiddelbart forbindes med omverdensmønstret i hans digtning. En række stilfigurer og ordgrupper, som har særlig affinitet til mønstret, skal i det følgende drages frem. Der er kun i nogle tilfælde – hovedsageligt, når det drejer sig om NL – meddelt eksakte tal for de behandlede træks forekomst. Dette skyldes først og fremmest, at det ikke er muligt at afveje sådanne tals signifikans, når det drejer sig om litterær stil, dels foreligger der ikke – og kan af metodologiske grunde næppe tænkes tilvejebragt – et materiale, der muliggør en tilfredsstillende sammenligning af de enkelte stiltræks “normalfrekvens”. (Vosmar 1984, 23)

Så ingen ansøgning om en computer på dette tidspunkt, og Vosmars påpegning af problemerne med at sætte stilen på formel forekommer stadig at være en udfordring, der ikke bare uden videre lader sig løse. Det betyder ikke, at kvantitative tilgange og tællearbejde er ligeegyldigt – det er jo netop det, både Hallar og Vosmar beskæftiger sig med, og derfor de har siddet og talt – men at det arbejde også kræver en række kompetencer, som den menneskelige hjerne stadig klarer bedre. Derfor kan man dog stadig have behov for at skabe sig et overblik, hvis man ved, hvad man søger efter, og hvad begrænsningerne i søgningerne er.

Mulighederne for at arbejde med en eller flere tekster ved hjælp af digitale værktøjer kan være gode og nyttige, men samtidig sidder man med en fornemmelse af, at det ikke for alvor flytter litteraturvidenskaben. Hvis man arbejder med et korpus af tekster, som man har læst, så har man jo læst dem og har på den baggrund en mere kompleks palet af observationer, som så kan understøttes af optællinger og søgninger, men noget kvantespring synes ikke inden for rækkevidde, især ikke hvis man betænker den umådelige kompleksitet, som litterære værker møder læseren med. Det virkelig interessante opstår derfor især, når man vil arbejde med et korpus, som er umuligt for nogen person eller gruppe at få læst, hvad enten det er umuligt inden for et liv eller en overskuelig periode.

Sådanne korpuser kan man selv generere, igen ud fra eksempelvis offentligt tilgængelige værker, som er skannet og karaktergenkendt. Desværre giver det nogle begrænsninger, ikke mindst i forhold til den store mængde af litteratur fra det tyvende og enogtyvende århundrede, som ikke blot kan downloades. Store mængder af ældre litteratur, også dansk, ligger dog “derude” og kan samles til et korpus, eksempelvis fra Arkiv for Dansk Litteratur.

Google Books Ngram Viewer

Men hvad med endnu større mængder af tekster? Her kommer Googles skanningsprojekt Google Books til hjælp. Projektet har været og er kontroversielt, fordi det måske, måske ikke, krænker forfatternes ophavsret. Google kan derfor ikke gøre al tekst tilgængelig, men de har skannet meget, meget store mængder og ud af det lavet værktøjet Google Books Ngram Viewer.² Et n-gram udtrykker den frekvens, hvormed et eller flere ord forekommer relativt til det samlede indhold i korpus. Unigrammer er de mest brugte, men bi- og tri- og så videre-grammer kan der også søges på. Desværre ikke på dansk, hvor Google formentlig heller ikke har skannet en kritisk masse af tekster, men på en række hovedsprog.

Princippet bag Ngram Viewer er beskrevet i artiklen “Quantitative Analysis of Culture Using Millions of Digitized Books” af Jean-Baptiste Michel m.fl. i tidsskriftet *Science*. Ud over de mere tekniske detaljer bag datasættet og den enorme mængde af bøger, der er skannet, gives der også en række slående eksempler på brugen af Ngrams. Eksempelvis hvor man ser effekten af censur af kunstnere i tysk kultur under Nazismen, eller hvordan begreber inden for både madkultur og videnskab stiger og falder i betydning over tid. Michel og hans gruppe bruger termen “culturomics” til at beskrive dette emergerende forskningsfelt i studier af tendenser i samfundet baseret på enorme datamængder.

De har store forhåbninger til feltet, men de er også bevidste om, at metoderne til at bruge værktøjet i en række forskellige discipliner endnu ikke er modnede. Det er dog et problem, at man ikke kan overskue, hvilke bøger, der gemmer sig bag tallene, hvorfor man i praksis er overladt til at stole på, at mængderne er så enorme, at resultaterne bliver repræsentative. Men det er åbenlyst, at Ngrams er et meget stærkt værktøj til at spore udviklinger over tid. Samtidig kræver det netop også meget store mængder, før det bliver interessant. Man kan sammenligne med andre ngram-visere, som fx British Librarys ngrams for internettet, som viser nogle utroværdigt store udsving over for termer, som formentlig har optrådt med en vis konstans.³

Den fremmeste litterære kritiker, der har beskæftiget sig med *digital humanities*, er uden tvivl Franco Moretti, som er initiativtager til Stanford Literary Lab, og som i bogen *Distant Reading* samler en række bidrag, der forsøger at bringe kvantitative metoder ind i litteraturhistorien. Hans bog fra slutningen af 1990'erne, *Atlas of the European Novel*, er også et vigtigt bidrag i denne sammenhæng, selvom om den ikke som sådan er afhængig af det digitale, men af abstraktioner, som gør det muligt at kvantificere. Moretti er tillige en central teoretiker inden for verdenslitteratur, hvor det netop er udfordringen ved den enorme og for ethvert individ ulæselige mængde af tekster, der står centralt for ham.

I sin seneste bog, *The Bourgeois*, bruger Moretti både et mindre korpus af digitaliserede romaner til at undersøge, hvornår og hvor omfattende ideen om borgerskabet opstår, såvel som Googles omfattende korpus af tekster. Denne tilgang er på mange måder eksemplarisk i kombinationen af anvendelsen af et tekstkorpus med specifik relevans for det område, der forskes i, hvor hver tekst kan udpeges præcist, og så det langt mere omfattende, men også upræcist afgrænsede korpus af indscannede tekster hos Google.

En anden grundlægger af Literary Lab, Matthew Jockers, peger i sit fine og nuancerede studie *Macroanalysis: Digital Methods & Literary History* med rette på, at der er mange faldgruber og muligheder for at lave problematiske fortolkninger baseret på det endnu relativt nye Ngrams (Jockers 2013: 120ff). Han hæfter sig især ved, at kulturelle udtryk ikke blot kan koges ned til enkelte ord, men udtrykkes i klynger af beslægtede termer. Det har han naturligvis ret i, men der er ikke desto mindre en række tilfælde, hvor man med relativ stor sikkerhed kan se betydningen af et eller flere ords forekomst. Nye teknologier, nye litterære genrer og egennavne er eksempelvis forholdsvist entydigt bundet op på enkelte termer, i al fald i en sådan grad, at markante udviklinger kan iagttages.

Jockers peger også på, at skriftsproget er en verden for sig, mens kulturen også udfolder sig oralt. Det er dog et vilkår for alle skriftbaserede korpuser, men her undlader han at pege på, at en af de mest interessante funktioner i Ngrams er muligheden for at søge på fiktion og kontrastere resultaterne med det samlede korpus af tekster på engelsk. Denne funktion vender jeg tilbage til. Først et tilbageblik på et par eksempler, hvor Ngrams ville have været nyttigt at have haft.

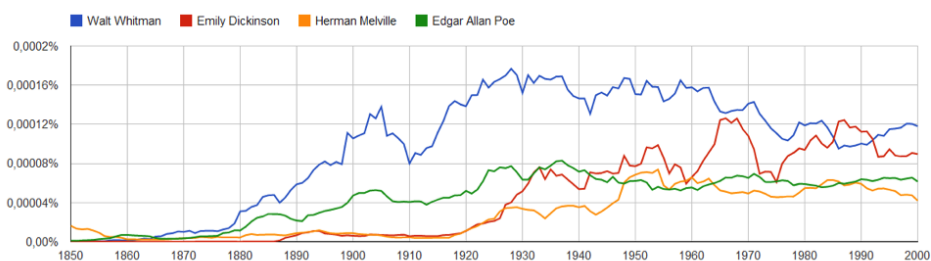
Bedre data til gode intuitioner

I min bog fra 2003, *Kanoniske konstellationer*, er der et kapitel om amerikansk lit-

teratur i 1920'erne, hvor jeg både skriver om, hvordan forfattere som Sinclair Lewis var meget synlige, men senere er forsvundet en del ud af litteraturhistoriens søgelys. I *Mapping World Literature* fra 2008 skriver jeg om genopdagelsen af Herman Melville og Emily Dickinson i 1920'erne, der trak dem frem i søgelyset efter at have været mere eller mindre glemt.

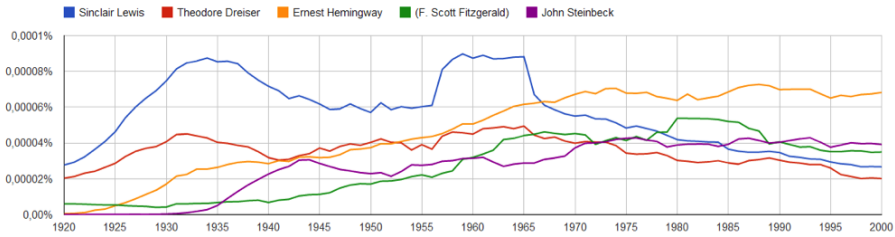
Der var noget utilfredsstillende ved at skrive om dette. For ganske vist kunne jeg henvise til andre, som kom til at stå som garant for udsagnene, ligesom jeg lavede mindre beregninger af oplag og filmatiseringer som udtryk for deres "litterære aktiekurs". Så på trods af, at jeg var overbevist om holdbarheden af beskrivelserne, ville det have været godt med et mere solidt datamateriale til at bakke pointerne op med.

Ngrams viste sig relativt enkelt og overbevisende i stand til at få svagt underbyggede teser gjort mere solide. Dokumentationen for, at Melville og Dickinson først for alvor bliver taget op i 1920'erne, fremgår af denne graf, hvor Walt Whitman og Edgar Allan Poe er medtaget som sammenlignelige forfattere, der var langt større interesse for i midten af 1920'erne.



Fra at være nærmest usynlige sammenlignet med de to andre forfattere, vokser både Melville og Dickinson i statur i 1920'erne og har siden haft den robuste tilstedeværelse i amerikansk litteraturhistorie, som tages for givet i dag. Men forhistorien er ganske forskellig for de fire udvalgte forfattere. Det er også markant, at Whitman og Poe også bliver styrket i den periode, hvor amerikansk litteratur for alvor selvstændiggør sig fra den britiske og finder en fornyet selvtilid. Præmissen er naturligvis, at man ikke finder det sandsynligt, at der er et helt andet kredsløb for omtale af forfatterne end det, som udtrykkes gennem den litterære kultur. Eller at forfatterne kunne have været store i aviser, men ikke i bøger.

Ser man på forfattere, som skrev i 1920'erne, argumenterede jeg i 2003 via udvalgte amerikanske litteraturhistorier for, at den status, som nogle forfattere tildeles i slutningen af 1940'erne, var ganske anderledes end den, de kunne gives i 1980'erne, ud fra den underlæggende præmis, at litteraturhistoriske værker i nogen grad søger at repræsentere den konsensus omkring kanon, der er i samtiden. Andre indikatorer, som jeg anvendte, var antallet af filmatiseringer af værker, antal af akademiske artikler og nogle stikprøver ind i bogmarkedets svært gennemtrængelige verden af oplag og oplagstal. Med en søgning i Ngrams blev billedet i nogen grad bekræftet.



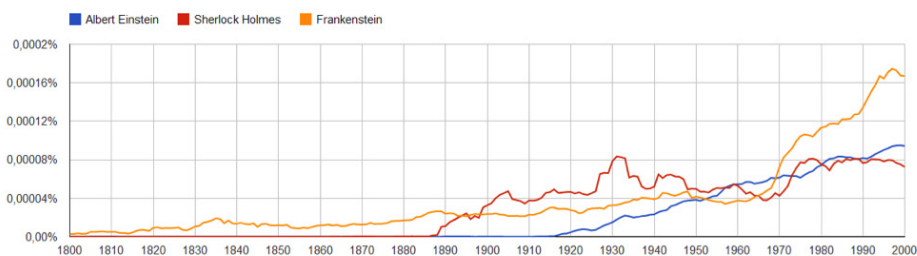
Den unge Hemingway stiger og stiger i betydning, og Fitzgerald opnår også pænt stor betydning, om end han skiller sig overraskende lidt ud fra især Dreiser og Lewis, mens Steinbecks popularitet uden for det akademiske ikke skal undervurderes. Den bratte stigning i Fitzgeralds omtale fra midten af 1970'erne falder sammen med, at *The Great Gatsby* filmatiseres.

I min seneste bog, *The New Human in Literature: Posthuman Visions of Changes in Body, Mind and Society after 1900*, var jeg interesseret i at få undersøgt, om ideen om det posthumane – en tilstand hvor mennesket har udviklet sig til art, som adskiller sig fra mennesket, som vi kender det, måske hjulpet på vej af dets egen teknologi – var en ny term, eller om det havde haft en substantiel interesse tidligere. Det første viste sig at være tilfældet, hvilket også bekræftes af termens fravær i eksempelvis science fiction, både i genrens primær- og sekundærlitteratur.



Det transhumane, en udvikling af mennesket, der dog ikke et brud med arten, har en længere historie, men viger nu i betydning for det posthumane. Samtidig er der et langt spor af rester af anvendelse af begreberne, som altså ikke fandt fodfæste og inspirerede andre, men som dog var.

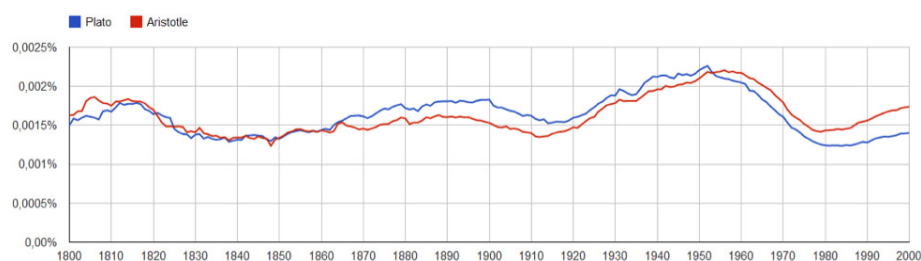
I min bog beskæftiger jeg mig blandt andet med denne forhistorie til det posthumane gennem Mary Shelleys roman fra 1818 *Frankenstein, or the Modern Prometheus*. Google har meget apropos valgt Frankenstein som forsideeksempel på Google Ngram Viewer i en demonstration af, hvordan der er en voldsomt stigende omtale af Frankenstein i modsætning til Sherlock Holmes' stagnation og Albert Einsteins støt voksende inklusion i den kulturelle erindring, der følger det mønster, som Michel m.fl. fremstiller i deres introduktion til emnet (Michel 2011: 179).



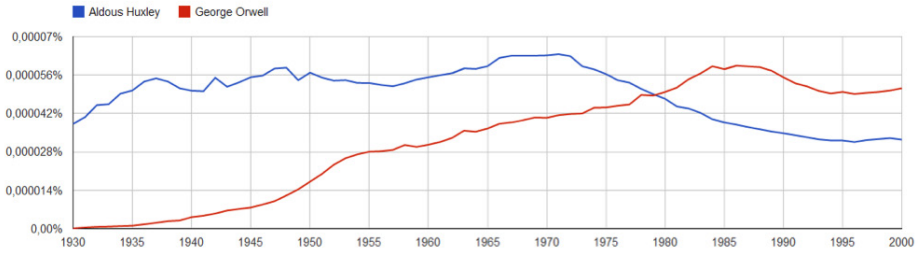
Det forekommer ikke urimeligt at spekulere i, at der er en sammenhæng mellem fremkomsten af genetik, kloning og selve forestillingen om det posthumane på den ene side og den øgede interesse for Shelleys roman på den anden. Men man må også regne med andre faktorer som interessen for litteratur af kvindelige forfattere, for Romantikken, samt filmatiseringer af romanen. Ikke desto mindre er der tale om en markant stigning for romanen om den navnløse kæmpe, og denne udvikling kunne bruges som en lille del i et større puslespil til at afdække interessen for det posthumane.

Litterære kursstigninger og -fald

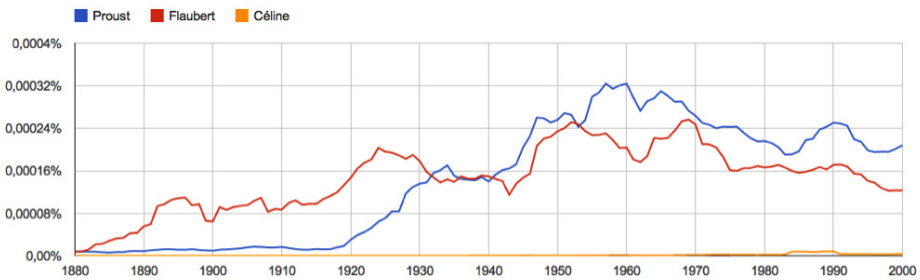
Egennavne og bogtitler har den fordel, at de som regel henviser entydigt til det ønskede studieobjekt, og at de er relativt uafhængige af udviklinger i den øvrige sprogbrug. Ngrams kan derfor bruges til at give et godt fingerpeg om udviklingen i forfatters position i kulturen. Nogle gange kan man få bekræftet, at der findes forfattere, der har kørt parløb i flere århundreder. Som det er tilfældet med Platon og Aristoteles.



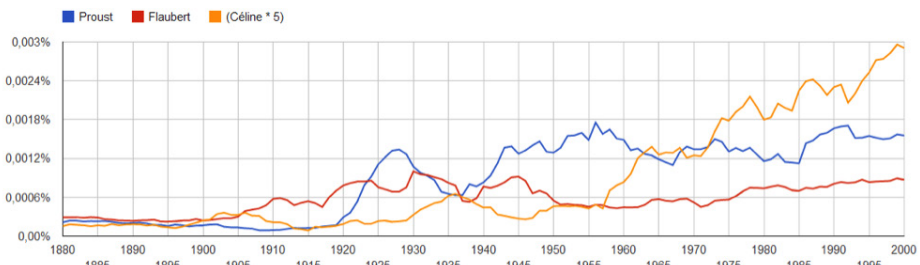
Andre gange kan man få en reminder om, at tingene ikke altid har været, som de ser ud nu. I tråd med de amerikanske forfatters foranderlige position, er betydningen af forfatterne til to dystopier skiftet betragteligt. Fra Aldous Huxleys markante overtag til George Orwells voksende betydning, der ikke overraskende kulminerer i årene efter 1984, hvorpå begge forfattere finder et solidt plateau i den offentlige diskurs.



Tendenserne i en udvikling kan ofte være svære at få øje på, hvis man uden filter måler en stor forfatter op mod en mindre. Louis-Ferdinand Céline fylder langt mindre end Marcel Proust og Gustave Flaubert i de engelsksprogede kulturer, og man ville let kunne se forandringerne i receptionen af forfatterskabet på en graf.



Men ganger man Célines inklusion op med fem, så ser man en markant tendens i udviklingen af interessen for forfatterskabet, som langt fra er så mættet eller selvføl- gelig målt over for to af den franske litteraturs moderne klassikere.



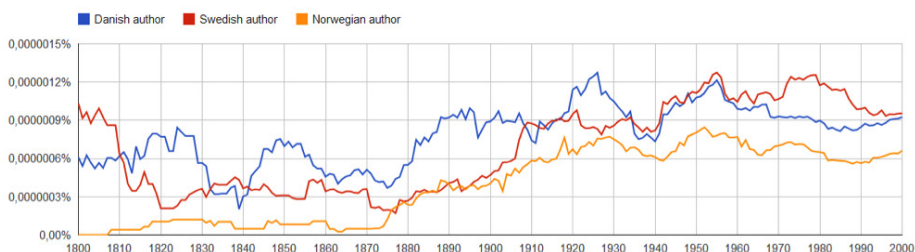
På lignende vis kan man se, at sonetten har en langt mere robust tilstedeværelse i kulturen end haikuet, men at tendensen for haikuet går i en anden retning end sonetten.



Sonnetten er stadig langt mere omtalt, men forandringen i interessen er markant. Sådanne afkodninger af en genres betydning er afgjort en af Ngrams styrker.

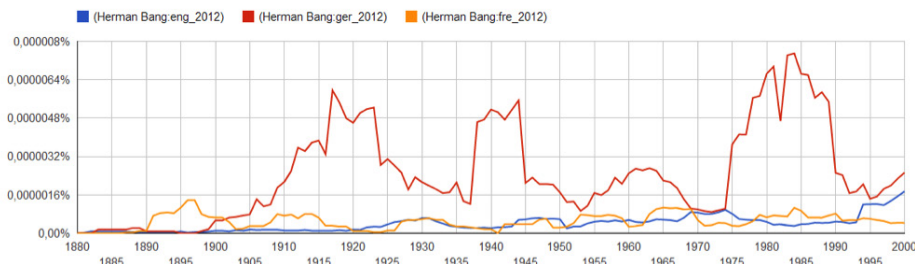
Verdensberømt i Danmark

En anden meget anvendelig brug af Ngrams ligger i at undersøge forskellige nationers tilstedeværelse i andre sprog. Her illustreret ved benævnelsen af forfattere fra Skandinavien på engelsk, som skulle give en idé om tendenser, selvom der kan gå en del tabt uden for de termer, der er søgt på her.

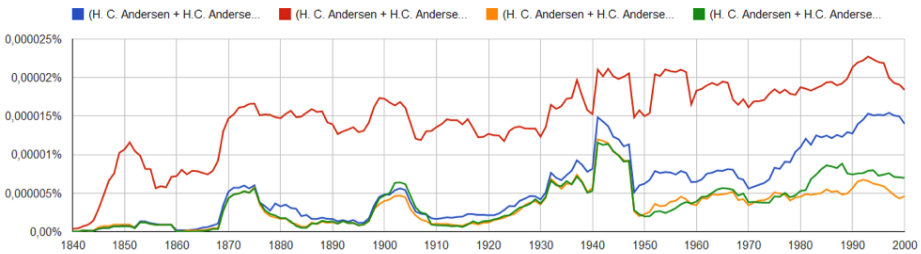


Danmark får et løft efter H. C. Andersens gennembrud på engelsk i slutningen af 1840'erne, mens Georg Brandes formentlig har en stor del af æren for den yderligere opblomstring i 1880'erne. Den norske tilstedeværelse er som forventet beskedent inden Henrik Ibsen, men vokser siden, og de tre nationer fremstår i dag nogenlunde jævnbyrdigt repræsenteret i den engelsksprogede verden.

Man må så også være klar over, at succes et sted ikke er en garanti for succes andre steder. Herman Bang er eksempelvis først og fremmest et navn i den tysksprogede verden, mens han ikke har haft et gennembrud i den fransk- og engelsksprogede, om end interessen for Bang på engelsk har været stigende i de senere år.



Omvendt er H. C. Andersen meget mere markant til stede på engelsk end på andre sprog, især tidligt i forfatterskabet, mens interessen på tysk nærmer sig den på engelsk.



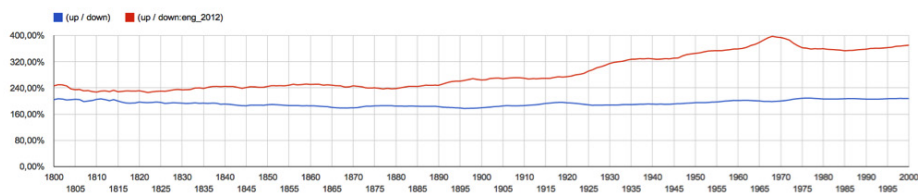
Det er bemærkelsesværdigt, at interessen stiger voldsomt under 2. verdenskrig, mens det er mindre overraskende, at Hollywood-filmen *Hans Christian Andersen* med Danny Kaye i hovedrollen sætter sig igennem på engelsk i 1950'erne. Men hvorfor en interesse under 2. verdenskrig? En tese kunne være, at det er digtet "Soldaten" fra 1829, der fandtes i oversættelse til flere sprog, som har sat sig spor, og ret så markant endda. Her viser værktøjets evne til at pege på oversete elementer i forfatterskabers gennemslagskraft.

Kultur vs. Litteratur

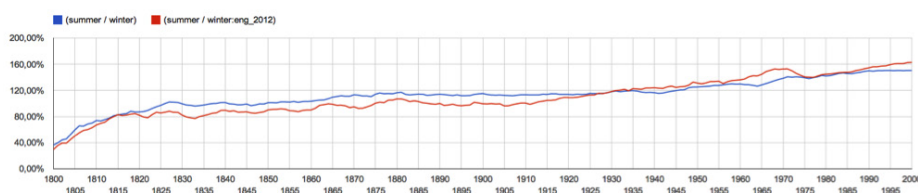
En af de meget interessante funktioner i Ngrams er muligheden for at se relationen mellem to navne eller begreber, fx "Plato" og "Aristotle", "summer" og "winter", eller "up" og "down". Her får man udtrykt, hvordan forholdet mellem termerne udvikler sig. Det tidligere eksempel med Platon og Aristoteles giver således en kurve, der svinger mellem 90 og 120 %, hvor 100 % angiver, at termerne er lige brugte, og kurven er mere eller mindre ret. Til gengæld fylder Orwell ikke meget tidligt i forhold til Huxley, der debuterer før, men forholdet mellem dem forandrer sig indtil det når et plateau i midten af 1980'erne, hvor Orwell er godt 50 % mere omtalt end Huxley.



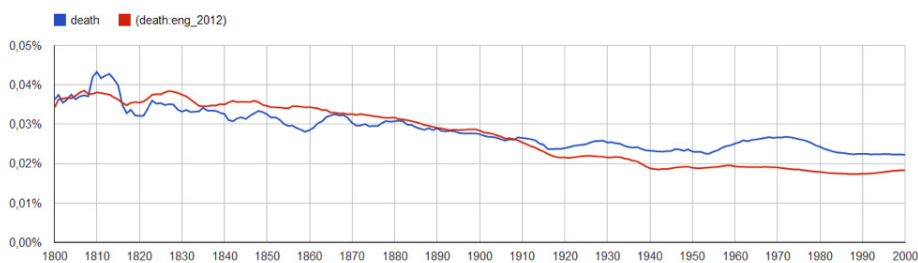
Denne funktion til sammenligning af forholdet mellem termer kan kombineres med at se på brugen af begreber i fiktion til forskel fra den samlede produktion af engelske bøger. Således får man ved "up/down" disse grafer.



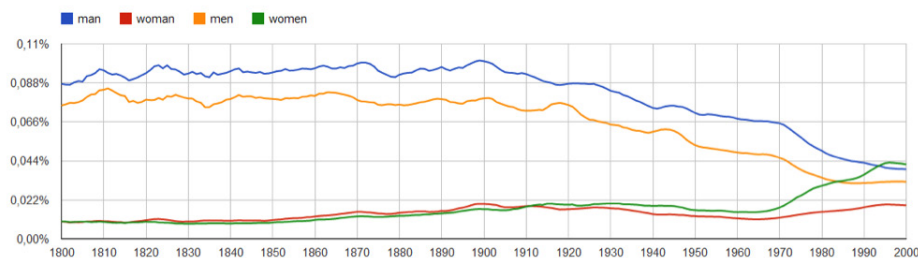
“Up” bruges mere både i fiktion og det samlede korpus, men det er kun uden for fiktionens verden, at der er en markant stigning i brugen af “up”. Er det udtryk for en kultur, der søger det positive, mens litteraturen trives ved en højere grad af balance mellem denne centrale modsætning? Måske, men det er ikke entydigt. Fiktionen følger den generelle tendens, når det kommer til sommer og vinter, endog i en særdeles markant udvikling fra fokus på vinter til sommer.



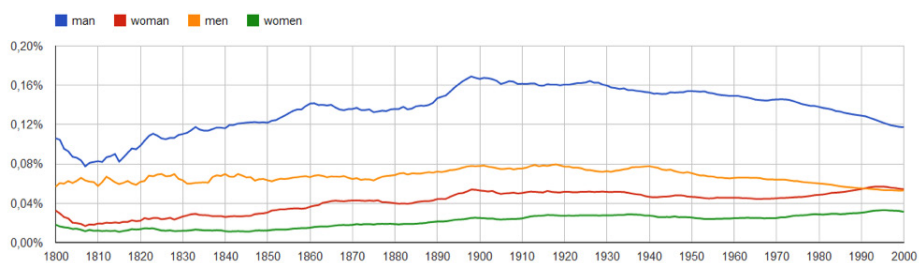
Men hvad betyder det? Ngrams giver ikke noget svar, men hjælper til at udvikle hypoteser om kulturelle forandringer. Set i forhold til dette eksempel, er det måske ikke så overraskende, at vi skriver mindre om døden, om end fiktionen, holder mere fast i temaet end den almene kultur.



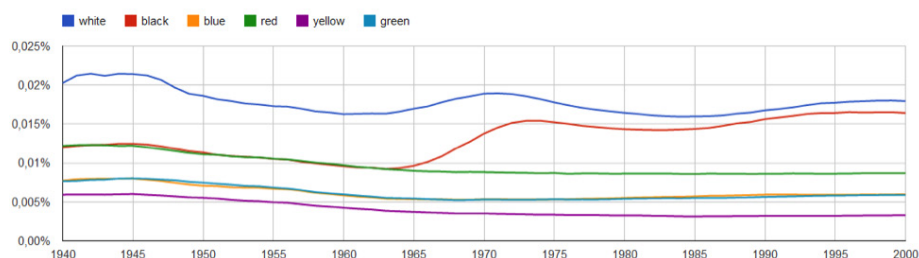
En anden markant udvikling er, at kvinder fylder mere efter at have været meget underrepræsenteret i skriftkulturen.



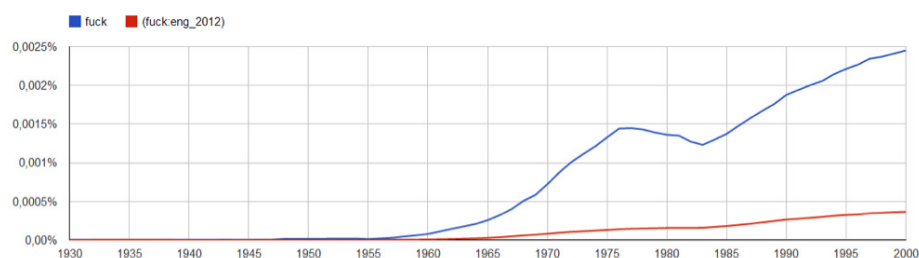
Men i fiktionens verden er billedet anderledes. Her er manden, måske overraskende, stadig langt oftere repræsenteret. Den voldsomme stigning af flertalsformen for det kvindelige køn er ikke markant, mens entalsformen til gengæld er det.



Politiske forandringer, eksempelvis borgerrettighedskampen i USA, kan også spores helt ned på frekvensen af brug af farver. Se blot hvilken graf, der er “sort”, og som signalerer en ny måde at tale om sine medborgere på og en større interesse for afrikansk-amerikaneres vilkår.



Endelig, hvis man skulle være i tvivl om, at litteraturen har bedre fat i meget brugte eder end den generelle skriftkultur, så er der også dokumentation for det.



For let til at være stærkt?

Det burde være klart fra de eksempler, jeg har givet, at man skal være varsom med værktøjet og anvende det kritisk med fuld opmærksom på dets styrker og svagheder. Men Google Books Ngram Viewer er et fascinerende redskab til at kunne underbygge teser om udviklinger i den kulturelle opmærksomhed langt bedre end ved de sporadiske nedslag, som man ellers ville kunne foretage. Dertil kommer, at

Ngrams er glimrende til at generere nye spørgsmål, som eksempelvis hvorfor H. C. Andersen blev markant mere omtalt under 2. verdenskrig.

Der kan nemt opstå en følelse af, at det er for let at bruge Ngrams, og at de svar, man får, er for overvældende og for gode til at være sande. Der kræves dog stadig en god portion kritisk sans og fornemmelse for at lave de rigtige søgninger, hvis man skal have gavn af værktøjet. Og det er også den slags værktøjer, der er brug for, hvis *digital humanities* skal have en bredere appel. Interessen for litteratur, historie, sprog, psykologi, kultur etc. vil hurtigt dø, hvis man begynder at besvare alle spørgsmål med mere eller mindre hårde data, som alligevel har behov for fortolkning, og som fremkommer ved, at man kan stille relevante spørgsmål, der igen er en færdighed, der ikke blot kan sættes på formel. Integrationen af nyttige værktøjer og metoder vil variere efter forskeres temperament og interesse, men der vil i overvejende grad være tale om at tilegne sig et redskab og ikke et paradigme.

Det er ikke let at lave mere avancerede studier med data, og der sker, som Matthew Jockers og andre tydeligt viser, meget inden for stilstudier af tekster. Men det er heller ikke let at blive en god litteraturhistoriker og oparbejde en forståelse for alle de elementer, der hører med til at beskrive en tekst i alle de kontekster, som gør den betydende og værdifuld. Netop forståelsen af konteksten er der, hvor redskaber som Ngrams kan være til endog nogle gange stor nytte.

Noter

- 1 Og alle har værktøjet, hvis man bruger en Chrome-browser. Indlæser man en tekst i html, fx et værk fra en åben base som Gutenberg.org, og søger på et ord i browseren, så laves der en visuel markering i højre side af skærmen af, hvordan søgeordet optræder i teksten. For en generel og glimrende introduktion til *digital humanities* værktøjer, se dette website: <http://toolingup.stanford.edu/>.
- 2 <http://books.google.com/ngrams>.
- 3 <http://www.webarchive.org.uk/ukwa/ngram/>.

Litteratur

- Hallar, Søren (1921): *Synselementerne i Naturskildringen hos J. P. Jacobsen*. København: Levin og Munksgaard.
- Jockers, Matthew (2013): *Macroanalysis: Digital Methods & Literary History*. Urbana-Champaign: University of Illinois Press.
- Lévi-Strauss, Claude (1963): *Structural Anthropology*. New York: Basic Books.
- Lodge, David (1993): *A David Lodge Trilogy: Changing Place – Small World – Nice Work*. London: Penguin.
- Lunenfeld, Peter et al. (2013): *Digital Humanities*. Cambridge: MIT Press.
- Michel, Jean-Baptiste et al. (2012): "Quantitative Analysis of Culture Using Millions of Digitized Books" in *Science* vol. 331, s. 176-182.
- Moretti, Franco (1998): *Atlas of the European Novel – 1800-1900*. London: Verso.
- Moretti, Franco (2000): "Conjectures on World Literature" i *New Left Review* nr. 1, s. 54-68.
- Moretti, Franco (2013a): *The Bourgeois: Between History and Literature*. London: Verso.
- Moretti, Franco (2013b): *Distant Reading*. London: Verso.

Thomsen, Mads Rosendahl (2003): *Kanoniske konstellationer. Om litteraturhistorie, kanonisering og 1920'ernes litteratur*. Odense: Syddansk Universitetsforlag.

Thomsen, Mads Rosendahl (2008): *Mapping World Literature: International Canonization and Transnational Literatures*. London: Continuum.

Thomsen, Mads Rosendahl (2013): *The New Human in Literature: Posthuman Visions of Changes in Body, Mind and Society after 1900*. London: Bloomsbury.

Vosmar, Jørn (1984): *J. P. Jacobsens digtning*. København: Gyldendal.