

# Lyrik: Et Matematisk System af Lingvistik

## Lyrikkens Natur

En dag slog det mig: Lyrik er egentligt bare et system af ord.

Ord, der hver især er symbolske repræsentanter, bærere, for én eller flere lyde. Bestemte lyde, ganske vist, der typisk har en indbyrdes akustisk overensstemmelse i form af rim. Dog hører alt dette, analyse af poesien, til humanioraens verden. Men behøver det nødvendigvis være tilfældet? For hvis der forefindes et system i lyrikteksters ord, ordenes lyde, må det kunne opdages. Formaliseres. Kvantificeres. Programmeres. Og hertil kræves matematik, for det er vel indbegrebet af systematisering.

Formålet med dette projekt er at udvikle en metode til at kortlægge og analysere det lingvistiske indhold i et givet tekststykke, navnlig lyrik, ud fra veldefinerede regler. Og netop fordi der er tale om objektive instruktioner, er det nærliggende, at en computer således anvender disse og udfører selve arbejdet i form af et computerprogram. Ved det *lingvistiske indhold* forstås en række veldefinerede parametre, såsom forekomsten af allitteration (også kaldet bogstavrim, eksempelvis ”kat” og ”klud”), assonans (også kaldet vokalrim, eksempelvis ”klud” og ”kulde”) og, vigtigst af alt, rim (eksempelvis ”kat” og ”hat”) graderet efter, hvor komplekse og internt sammenhængende disse er. Et tilfælde af en af disse rimtyper mellem to vilkårlige ord kaldes under ét for en *lingvistisk hændelse*. Hvad der præcis menes med alt dette, uddybes senere, men beskrives og defineres i særdeleshed i selve projektrapporten.

## Lingvistikens Natur

En danskfaglig definition af et rim er ”den akustiske effekt, som frembringes, når der er lydlig overensstemmelse mellem to eller flere ord”<sup>[1]</sup>, men denne udlægning er naturligvis, som lyrikken i sig selv, subjektiv af natur. Derfor må et mere formelt sæt af begreber indføres; ellers kan objektivitet aldrig tilnærmes eller, i sidste ende, opnås. Idet der her er tale om et, så vidt jeg ved, nyt felt i grænsefladen mellem to ellers totaladskilte områder, datalogi og lyrik, sammenkoblet ved anvendt matematik, er det alt for omfattende at definere og beskrive samtlige koncepter og begreber i detaljer i dette projektresume, men det er naturligvis gjort i rapporten. Herunder følger der i stedet en overfladisk beskrivelse af, hvordan og hvorfor computerprogrammet virker med henblik på at opnå en konceptuel, intuitiv forståelse. Hvad angår metodens validitet, er programmets logik åbenbar i selve koden, og matematikkens gyldighed udspringer af formler og ligninger.

Først og fremmest tages der udgangspunkt i ordenes fonetiske lyd, da man alene af stavning ikke kan udtale sig om, hvorvidt to eller flere ord rimer. Hertil benyttes en fonetisk ordbog, der omskriver ord til en række af veldefinerede symboler, kaldet fonemer, der bærer bestemte lyde. Rim er da, groft sagt, overensstemmelse mellem to ords respektive fonemer fra højre mod venstre; jo flere fonemer, der er ens, desto ”bedre” rimer ordene. Således opstilles der stringente kriterier for forekomsten og kvaliteten af rim, og tilsvarende gøres for de øvrige førnævnte rimtyper.

## Matematikkens Natur

Samtlige ord i et tekststykke skal da krydsrefereres med hinanden for at undersøge forekomsten af rim. Dette gøres ved symmetriske matricer, hvor hvert ord i tekststykket flugter ad både matricens rækker og søjler, således at hver indgang da angiver, hvilke, hvis nogen, lingvistiske hændelser, der optræder mellem de to pågældende ord i tekststykket. De forskellige typer af lingvistiske hændelser tildeles hver et bestemt primtal, hvorved en indgang da indeholder et produkt af de til de forekommende typer af lingvistiske hændelser tilsvarende primtal. Ved efterfølgende primfaktoropløsning

---

<sup>[1]</sup> Fibiger, Johannes m.fl.: *Litteraturens veje*, 2. udgave, 6. oplag, Narayana Press, 2006, s. 517

kan man da af matricen udlede, hvor lingvistiske hændelser optræder i tekststykket, og hvilke typer der er tale om. Matricen indeholder udelukkende lyrikkens rå-data, og det viser sig, at visse mønstre, sammenhænge, optræder heri, som induktivt kan generaliseres. Herefter kan man gennem kombinatorik og løsning af ligningssystemer, automatiseret af programmet, ”regne baglæns” for at nå frem til, præcis hvad tekststykket indeholder. Endeligt opsummeres alt dette i fem numeriske indeks, talværdier, der dækker over forskellige lingvistiske aspekter, hvorved forskellige tekststykker kan sammenlignes. Objektivt.

Programmets virkemåde er eksemplificeret i en undersøgelse af udviklingen i det lingvistiske indhold af de mest populære rap-tekster hvert år fra 1990 til 2015, der viser, qua indeksene, at ”kvaliteten” generelt er faldende. Denne undersøgelse bør ikke anses som et forskningsprodukt i sig selv, men snarere som en demonstration af ét af programmets mange anvendelsesaspekter.

### **Datalogiens Natur**

Idet computerprogrammer er logisk konsistente, behandles alle input efter samme spilleregler, og det interessante ved disse spilleregler er, at de udgøres dels af matematiske systemer, som per definition er logiske og objektive, og dels af de førmtalte parametre, der er defineret så stringent, at et program kan læse og anvende dem. Det kan naturligvis diskuteres, hvorvidt definitionerne og metodikken er repræsentative for kvaliteten af et stykke lyrik, men det er udarbejdet på en sådan måde, at programmets resultater i hvert fald er repræsentative for forekomsten af lingvistisk indhold ud fra definitionerne. Det er aksiomer, der medfører, at samtlige følgeslutninger er lige sande. Eksempelvis tages der på ingen måde højde for ordenes semantik, herunder lyrikkens budskab, betydning, og hvilke følelser det vækker i læseren; alt dette hører til menneskets verden. Men idet det i lyrikkens, især rap-musikkens, kredse ofte tilstræbes at rime vel, kan resultaterne give et praj om, hvorvidt et stykke lyrik er velskrevet; rent teknisk. Og dog er det alligevel ikke heri, dette projekts primære værdi består.

### **Videnskabens Natur**

At en computer kan udføre det samme stykke arbejde som et menneske, såsom at analysere rimene i et digt, hvilket typisk hører til det humanistiske fakultet, og sågar gøre det hurtigere og ud fra samme spilleregler hver gang, har potentielt enorme implikationer for eksempelvis kunstig intelligens. Denne abstrakte metode kan anvendes til at optimere computeres kommunikation med mennesker ved at få deres sætninger til at lyde mere flydende. Endvidere ville man kunne bruge metoden ”omvendt”, således at computere i sidste ende kan skrive, skabe, lyrik på lige fod med mennesket; eller snarere med en sådan troværdighedsgrad, at produktet ikke kan skelnes fra den menneskelige pendant: Ud fra en given, eftertrægtet matrice, der anvendes som ”skabelon”, skriver den nu knap så kunstige intelligens en tekst, der vil resultere i selvsamme matrice. Endeligt forefindes der muligvis øvrige anvendelsesaspekter inden for blandt andet generel lingvistik, såsom at man ved statistisk analyse af big data måske numerisk kan fastlægge en kunstners såkaldte idiolekt; forfatterens unikke lingvistiske fingeraftryk. Døren er i hvert fald åben for videre arbejde, for sådan er grundforskning.

Lyrik, kunst, og matematik, videnskab, anses ofte som hinandens modsætninger; eller i hvert fald som forskellige segmenter af menneskets forståelsesområde. At bygge bro mellem disse verdener, at give matematikken et nyt genstandsområde at udforske, udtaler sig ikke blot om menneskets natur, men også videnskabens. Derfor bør dette projekt ikke anses som en direkte demonstration af den naturvidenskabelige metode som sådan, men snarere som en intuitiv undersøgelse af, hvad der ligger bag den:

Den udviklede metode er selve produktet. For kun ved at udfordre grænserne for hvad videnskaben egentligt kan beskrive, opdager man, hvor videnskabens begrænsninger i hvert fald ikke er.