

Aksel Noesgaards ordstatistiske pionerarbejde

Af HENNING SPANG-HANSEN

Enhver med tilknytning til retskrivningsundervisningen i den danske skole ved, hvordan denne undervisning takket være *Aksel Noesgaards* arbejder over ordhyppighed og fejltyper er blevet et mønstereksempel paa videnskab i pædagogikkens tjeneste. Men det er vist mindre kendt og har vist ikke før været genstand for selvstændig behandling, at Noesgaards ordstatistiske undersøgelser ogsaa fra et sprogvidenskabeligt synspunkt udgør en indsats og et resultat, der fortjener sidestilling med ordstatistiske undersøgelser i andre lande. Vel havde der i udlandet tidligere været foretaget en del undersøgelser af denne art, med Kaedings tælling af 11 millioner løbende ord tysk som det mest impoerende – men samtidigt afskrækkende – eksempel, men da Noesgaard omkring 1930 paa eget initiativ begyndte sine tællinger, kunne han kun i ringe grad drage nytte af disse arbejder, der i de fleste tilfælde forfulgte andre maal, især stenografiens krav. Det var overvejende Noesgaards egen idemæssige indsats, da han – økonomisk og arbejdsmæssigt ogsaa kun med beskeden bistand – iværksatte de undersøgelser, som han fortsatte gennem henved 30 aar, og takket være hvilke dansk sprog og nyere dansk litteratur tør siges at være undersøgt i ordstatistisk henseende saa omhyggeligt som noget andet lands sprog og litteratur.

Den sprogvidenskabelige side af Noesgaards arbejde træder maaske tydeligst frem i de undersøgelser over danske forfatters sprogbug, som han netop havde naaet at afslutte inden sin død i maj 1959. Som en videreførelse af tællinger i tekststykker fra forfatterskaber tilhørende tiden før og omkring aarhundredskiftet, og fra forfatterskaber tilhørende perioden ca. 1900 til 1935, havde Noesgaard nu undersøgt 10 forfatterskaber tilhørende de sidste ca. 20 aar for ved sammenligninger at konstatere eventuelle forskelle mellem de tre perioders sprogbug. Undersøgelsens resultater forelaa i færdigt manuskript fra Aksel Noesgaards haand, og det havde været hans tanke at søge arbejdet udgivet til sin 80-aarsdag, 3. februar 1960, som afslutning paa et mangeaarigt virke inden for dette felt. Med enkelte tilføjelser, foretaget paa grundlag af Noesgaards tællingsresultater, og med et forord af Paul Diderichsen er arbejdet nu posthumt blevet udgivet til samme dag under

titlen: *Hyppighedsundersøgelser over ordforraadet i dansk, III; tre generationer af danske prosaister (1878–1958)*.

Dette sidste arbejde tilsigter som nævnt at supplere og afslutte den stilistisk betonedede side af Noesgaards undersøgelser og slutter sig i det hele taget nøje til hans tidligere offentliggjorte ordstatistiske arbejder: *Hyppighedsundersøgelser over Ordforraadet i Dansk* (København 1934), *Hyppighedsundersøgelser II* (København 1937), *De nødvendige danske ord*, 2. ændrede udg., København 1956, spec. s. 27–35. Det synes derfor rimeligt og berettiget at underkaste disse arbejder en samlet vurdering fra et sprogstatistisk synspunkt.

For en første betragtning kan undersøgelser som Noesgaards tage sig ud som blot og bar rubricering og optælling af foreliggende ordforekomster. Hvad den matematiske side af sagen angaar, rummer Noesgaards undersøgelser, saaledes som de foreligger offentliggjorte fra hans haand, heller ikke for en nøjere betragtning nogen problematik: De matematiske operationer indskrænker sig til addition, udregning af middeltal og procentregning. Selv ikke i den nyeste bog har den i udlandet ret udbredte interesse for at finde matematisk formulerede sammenhænge inden for ordhyppighedens fordeling sat sig spor. Disse bestræbelser, med den saakaldte »Zipf's lov« som et af de første og mest kendte eksempler, har Noesgaard givetvis haft kendskab til; han besøgte, for en stor del gennem personlig kontakt, et førstehaandskendskab til ordstatistisk forskning i USA, hvor disse matematiserende tendenser først og fremmest har haft og stadig har udbredelse (inden for statistisk stilistik maa dog ogsaa fremhæves *Yule* og senere forskere i England).

Men dels var Noesgaard – efter sit eget udsagn i sin erindringsbog (*Fra plogen til asfalten*, Skjern 1953) – mindre interesseret i matematik og har vel derfor taget det standpunkt at lade sine tællingsresultater staa som de er og for hvad de er; dels maa det bemærkes, at de matematiserende tendenser overvejende repræsenteres af forskere med psykologisk eller i den senere tid informationsteoretisk orientering, saaledes at tendenserne selv i USA kun i mindre grad har paavirket de pædagogisk orienterede ordstatistikere, som Noesgaard havde sin nære kontakt med.

Den forestilling om problemløshed i videnskabelig henseende, som en forhaandsvurdering af undersøgelser som Noesgaards kan føre til, har vel ogsaa spillet med ind ved afslagene fra flere videnskabelige fond, som Noesgaard henvendte sig til midt i 30'erne om støtte til videre undersøgelser. I sprogvidenskabelig henseende er problemløs-

heden imidlertid kun tilsyneladende. Enhver, der blot i beskeden maalestok har givet sig i lag med ordtællinger i tekster, vil have erkendt, at opgaven er meget langt fra at kunne løses ved rubricering og optælling af ordforekomsterne, saaledes som de umiddelbart fremtræder. Hvis opgaven var af en saadan triviell art, ville den iøvrigt lade sig bedre, dvs. nøjagtigere løse helt med mekaniske hjælpemidler end med menneskelig arbejdskraft, der dels sløves af gentagelser og dels distraheres af interessante, men en saadan triviell optælling uvedkommende forhold i teksten eller maaske i omgivelserne. Og selv om man bl. a. i USA har benyttet tekniske hjælpemidler (hulkortmaskineri) til ordstatistiske undersøgelser (se f. eks. H. H. Josselson: *The Russian Word Count*, Detroit 1953) og der seriøst arbejdes paa at udvikle elektronisk data-maskineri (»matematikmaskiner«, »elektronhjernener«) specielt til saadanne formaal, er det kun visse af undersøgelsernes operationer man herved mekaniserer. Fuldt mekaniserede ordtællinger i tekster af sædvanlig art er et perspektiv af samme fjerne og maaske principielt urealistiske art som fuldt mekaniseret oversættelse af tekster fra et dagligsprog til et andet.

Ligesom det saaledes, hvad mekaniseringsperspektiver angaar, er berettiget at nævne ordstatistik og oversættelse i sammenhæng, er det, hvad angaar de nuværende forhold, berettiget at sammenligne ordstatistikens problemer med leksikografiens, især saadan som disse optræder ved étsprogordsbøger. Det har næppe interesse at forfølge denne sammenligning i detaljer; det vil være tilstrækkeligt at nævne spørgsmaal som afgrænsning af enheden ord overfor ordsammensætninger, behandlingen af bøjningsformer og adskillelse af homonymer. Samt først og sidst afgrænsning af materialet: Hvilke arter af tekster skal man dække, og i hvilken henseende og grad kan man paaregne, at de tekster, man faktisk kan overkomme at undersøge, giver et repræsentativt materiale.

Ordstatistikerens arbejde er naturligvis i visse henseender mindre problemfyldt end leksikografens eller grammatikerens, bl. a. fordi han som regel netop kan bygge paa foreliggende ordbøger og grammatiske beskrivelser. Men hvor grammatikeren og leksikografen ofte kan nøjes med at beskrive og eksemplificere analyseproblemerne (f. eks. hvor det drejer sig om en syntaktisk begrundet inddeling af forekomsterne af *der* i danske tekster), staar ordstatistikerens i den situation at skulle rubricere (og dermed tælle) hver enkelt ordforekomst her og nu. Han

kan ikke operere med glidende skalaer, men maa overfor alle tvivls-tilfælde – og netop de hyppige ord frembyder ofte tvivls-tilfælde – give sig og sine eventuelle assistenter procedureregler, der saa vidt gørligt forener hensynet til undersøgelsens formaal (f. eks. at fastslaa visse syntaktisk-stilistiske forskelle) med de krav, som tælling i praksis stiller (hurtigt konstaterbare kriterier).

Denne oversigt over ordstatistikens hovedproblemer har dels haft til formaal at begrunde den paastand, at et ordstatistisk værk har mere tilfælles med en ordbog end med Statistisk aarbog, dels at give baggrund for en betragtning af den maade, hvorpaa Aksel Noesgaard har taget problemerne op. Som det allervæsentligste maa det fremhæves, at han klart har formuleret formaalet (formaalene) med sine arbejder, og at han ogsaa til stadighed har haft disse formaal klart for sig :

Primært at finde hvilke ord ethvert barn i skolen og senere faar brug for at skrive og dermed at stave, *sekundært* at finde eventuelle forskelle af nærmere angivet art mellem forskellige slags tekster (herunder specielt forskellige forfatterskaber). I Noesgaards nyeste og sidste arbejde, efter at arbejdet med stavelisterne havde naaet en afslutning, er kun det sidstnævnte formaal paa tale.

Det er først og fremmest i kraft af denne begrænsede, men klare maalsætning, at Noesgaards fremstilling giver læseren en fornemmelse af sikker grund, til forskel fra ikke saa faa udenlandske ordstatistiske arbejder, der sætter sig videre og maaske mere interessante maal saasom at finde sammenhænge mellem teksters ordstatistiske struktur og menneskelig psyke eller termodynamikkens grundlag. Uden at rejse ringeste tvivl om de paagældende forskeres videnskabelige redelighed tør man sige, at selve disse formaals vide perspektiver har skudt interessen for empiriske forhold noget i baggrunden til fordel for betragtninger over, hvad der under inddragen af forskellige forudsætninger kan udledes af et – ofte stærkt begrænset – erfaringsmateriale.

Noesgaards primære formaal har imidlertid, til trods for sit udtalt pædagogiske tilsnit, almen sprogvidenskabelig interesse. For hvis der kan konstateres et inventar af ord, som forekommer i (men alene naturligvis ikke udgør) tekster af alle de arter, som dansk-talende aktivt eller passivt sædvanligt gør brug af, staar man overfor forhold, der maa beskrives som hørende til det danske sprogs *system*, ganske ligesom man udfra de fælles syntaktiske forhold, bøjningsforhold og andre grammatiske forhold i et saadant bredt tekstudsnit har opstillet det danske

sprogs grammatiske system. Dermed er ikke sagt, at et saadant fælles ordinventar indgaar i det danske sprogs *struktur*, naar struktur forstaas som et netværk af relationer af nærmere angivne typer som f. eks. de glossematiske funktioner. Men det samme gælder grammatiske forhold i vanlig forstand, eftersom ikke alle systematiske forhold inden for grammatikken kan beskrives som strukturelle, og det er paa den anden side plausibelt, at omfanget og sammensætningen af et saadant fælles ordinventar i visse henseender kan vises at være betinget af strukturelle forhold; man kan til illustration tænke paa konjunktioners og visse adverbiers rolle i syntaksen.

Men lader et saadant fælles ordinventar sig overhovedet konstatere? Noesgaard synes selv ikke at have været i tvivl om muligheden derfor, og han har her haft visse erfaringer fra undersøgelser i andre sprog at støtte sig til. Det har ved undersøgelser af hyppighedsfordelingen for ord i tekster altid vist sig, at temmelig faa forskellige ord udgør en stor procentdel af tekstens løbende ord, og endvidere, at det i nogen grad er de samme ord, der forekommer hyppigt i enhver undersøgt tekst paa et og samme sprog. Sammenholdes dette med en grundforestilling om, at ethvert ord i sproget har en vis, »sand« hyppighed, hvis værdi lader sig bestemme med stadig bedre tilnærmelse jo større og mere blandet en tekstmængde man optæller, føres man frem til den slutning, at sprogets ord kan ordnes efter vigtighed udfra tilstrækkeligt store og tilstrækkeligt varierede tællinger af ordforekomster i tekster.

Paa denne basis, der først i den seneste tid er blevet genstand for principiel kritik og efterprøvning, startede Noesgaard sin efterforskning af de vigtigste ord for børn i danske skoler. Men uden at fordybe sig i teoretiske betragtninger og principielle undersøgelser naaede han udfra sit eget materiale, som han med sikkert skøn havde behandlet i adskilte grupper til indbyrdes sammenligning, hurtigt frem til et mere nuanceret billede af forholdene: De ord som er fælles for en flerhed af tekster, f. eks. fælles for lige store uddrag af ti forfatters produktion, er – taget som gruppe – langt fra de samme som de hyppigste ord i disse tekster taget under ét. For i hver enkelt sammenhængende tekst eller tekststykke er der som regel visse ord, som er meget hyppige i kraft af tekstens specielle tema eller emne, »hvad teksten handler om«, men som sjældent eller slet ikke genfindes i andre tekster af lignende art (ved art kan man her f. eks. forstaa nutidig fiktionsprosa). Da antallet af mulige forskellige tekstspecifikke ord er uhyre stort, nemlig af samme

størrelsesorden som et sprogs hele registrerede ordinventar, ville en talmæssig udjævning kræve tælling af et uoverkommeligt antal tekster. Resultatet i praksis bliver derfor, at selv meget omfattende tællinger kun giver tekstafhængige hyppighedsværdier for et lille antal meget hyppige ord. »Mammuttællinger« kan derfor ogsaa kun siges at tjene noget formaal, hvis materialet behandles i adskilte dele, der kan betragtes som – og stikprøvevis vises at være – homogene hver for sig.

Hvor det gælder opstillingen af et tekstafhængigt ordinventar, dvs. ord der er fælles for tekster af alle sædvanlige arter, er hyppighedsbestemmelse i vanlig forstand mindre egnet end en direkte sammenligning af ordinventaret i en flerhed af tekster. Dette kan lyde som en selvfølgelighed, men set paa baggrund af ordstatistikens oprindelse og datidige problemstilling var det en overordentlig vigtig erkendelse, Noesgaard i sine »Hyppighedsundersøgelser II« bragte frem ved at skelne mellem de hyppigste ord og de vigtigste ord; karakteristikken *vigtigste* er maaske mindre velvalgt, mere neutralt er det at tale om de mest almene ord, eller om man foretrækker: de mest generelle ord. Men uanset terminologien er der næppe noget sted i den ordstatistiske litteratur, hvor denne vigtige forskel er indført, paavist og eksemplificeret saa klart som hos Noesgaard. Man finder naturligvis i udenlandske værker fra omkring samme tidspunkt eller senere det samme forhold behandlet, idet man (i engelsk terminologi) skelner mellem ords *frequency* og *range*, men i de fleste tilfælde er definitionerne vage og der inddrages skøn til bestemmelsen af ordenes *range* (ogsaa denne betegnelse er lidet vejledende, og den forveksles ydermere let med *rank*, fransk *rang*, der er gængs betegnelse for et ords hyppighedsnummer).

Fra et sprogvidenskabeligt synspunkt er det beklageligt, at Noesgaard ikke i senere arbejder fulgte denne sin erkendelse op med yderligere undersøgelser over ords »vigtighed« eller »generalitet«, men som ovenfor nævnt er undersøgelser af denne art ogsaa i udlandet stadig paa begyndelsesstadiet. Det skal ogsaa bemærkes, at bestemmelser af ords generalitet byder paa noget større matematiske problemer end hyppighedsbestemmelse udfra løbende ord i tekster; det er f. eks. aabenbart, at de sammenholdte teksters længde spiller ind paa en mere indviklet maade end ved udregning af en hyppighedsværdi. Under alle omstændigheder vil det kræve omfattende undersøgelser at fastslaa, om og i hvilken udstrækning et tekstafhængigt ordinventar

kan opstilles, og dermed bl. a. at afgøre, om og hvordan den liste, som Noesgaard naaede til ved sin metode, skal revideres og suppleres. Noesgaard fremhæver selv, at hans resultater kan møde den skæbne at blive kuldkastet af mere omfattende undersøgelser.

Det ville føre for vidt her at studere de lister og tabeller over de vigtigste og hyppigste ord, som Noesgaard har givet i sine bøger. Ordene saavel som deres hyppighedstal kan give anledning til betragtninger af mange arter: Hyppighedsfordelinger kan opstilles og eventuelt sammenholdes med de beskrivelsesformler, som er foreslaaet i den matematiserende ordstatistik, eller ordhyppighederne kan sammenholdes med teksternes arter, emner og særlige grammatiske forhold, saaledes som Noesgaard selv i stor udstrækning gør det i sine bøger; eller ordene i listerne kan betragtes i henseende til deres indhold og til deres udtryk, hvorved man bl. a. ser en udpræget sammenhæng mellem hyppighed (eller »vigtighed«) og udtrykstype. Men i alle henseender gælder det, at Noesgaards resultater, saaledes som han har fremlagt dem, vedrører de bestemte tekster, han har valgt at tælle, og uanset tællingernes betydelige omfang og de valgte tekststarters spændvidde har enhver konklusion vedrørende andre tekster eller sproget som system hypotesens karakter. Noesgaard gør selv opmærksom paa dette og er yderst forsigtig med at drage almene slutninger udfra sine resultater; problemet har naturligvis ogsaa foreligget for Noesgaard under hans opstilling af stavellister paa grundlag af ordtællingerne, men det fremtræder mindre udpræget ved dette pædagogiske sigte end for rent videnskabelige formaal, fordi der ved enhver pædagogisk anvendelse uvægerligt indgaar et normativt moment.

I forbindelse med omtalen af det primære formaal for Noesgaards arbejde er det rimeligt at berøre hans behandling af de tidligere nævnte metodiske spørgsmaal ved ordtællinger. Til ortografisk-pædagogiske anvendelser er det en nødvendighed at betragte bøjningsformer som forskellige tælleenheder, hvad Noesgaard da ogsaa gør, men de sammenstillinger og betragtninger over bøjningsformernes indbyrdes hyppigheder, som han medtager allerede i »Hyppighedsundersøgelser II«, har videre, sprogvidenskabelig værdi.

Derimod er det beklageligt, at Noesgaard veg tilbage fra at holde homonymer adskilt: Verbet *faa* og adjektivet *faa* tælles f. eks. under ét (hvorimod substantivet *Faar* holdes adskilt fra præsens *faar* i kraft af begyndelsesbogstaverne). For naar man saaledes kun tager ordud-

trykket i betragtning, bortfalder ikke blot muligheden for at uddrage de enkelte homonymers hyppighedsværdier, men mulighederne for at sammenligne bøjningsformers hyppighed forringes væsentligt. Naar f. eks. *faa* omfatter baade verbet og adjektivet, kan man ikke af tallene slutte noget om forholdet mellem verbalformerne *faa* og *faar*. Og selv fra et ortografisk-pædagogisk synspunkt er adskillelsen af homonymer (nærmere præciseret homografer) vigtig som følge af forskellen i dansk mellem homofoni og homografi. Det er f. eks. til ringe nytte at kunne stave udtrykkene *bor* og *bord*, hvis man ikke samtidigt behersker disses fordeling paa ordindhold, inklusive den omstændighed at der til hvert af udtrykkene er knyttet mere end ét indhold.

Det er forstaaeligt i betragtning af de teoretiske og praktiske vanskeligheder, at Noesgaard valgte kun at skelne ord efter udtrykket, og takket være den daværende retskrivnings skelnen mellem stort og lille begyndelsesbogstav har han faktisk kunnet opnaa en større diskrimination end i de dengang foreliggende store tællinger for engelsks vedkommende (Thorndike's og Horn's tællinger). Det er imidlertid værd at bemærke, at begyndelsesbogstavet som umiddelbart kriterium jo svinger efter punktum og visse andre steder, og i disse – relativt ikke helt *faa* – tilfælde har Noesgaard alligevel gennemført adskillelsen paa syntaktisk-semantic grundlag. Springet til en videregaaende adskillelse havde derfor ikke været saa stort, og Noesgaards store undersøgelser ville have vundet ikke saa lidt i sprogvidenskabelig værdi, hvis de var blevet udført med adskillelse blot paa det ordklassegrundlag, som Noesgaard har indført i sit sidste arbejde (Hyppighedsundersøgelser III). Det er de store bogstavers bortfald som gav anledning til, at Noesgaard i denne tælling af godt 100.000 løbende ord har rubriceret forekomsterne efter ordklasse, hvorved f. eks. verbet *faa* og adjektivet *faa* bliver to forskellige tælleenheder. Uanset de traditionelle ordklassers uklare videnskabelige status har denne inddelingspraksis vist sig gennemførlig, hvilket under alle omstændigheder understreger ordklassernes pragmatiske værdi. Det er iøvrigt værd at lægge mærke til, at den mangel paa balance, der præger ordklasseinddelingen i grammatisk og leksikalsk henseende: Substantivernes utal overfor f. eks. konjunktionernes faatal, ikke kan siges at være til stede mellem forekomster i tekster: Substantivprocenten er f. eks. af nogenlunde samme størrelse som konjunktionsprocenten. Ogsaa dette forhold bør tages i betragtning ved vurderingen af ordklassernes status og teoretiske basis.

Hvad homonymadskillelse angaar, er Noesgaards sidste arbejde paa denne maade videregaaende end hans tidligere tællinger, selv om der kun er tale om en ordklasseinddeling, som hverken i princip eller i detailler falder sammen med en semantisk homonymadskillelse. Iøvrigt er metodikken uændret i den sidste tælling, hvilket er fordelagtigt ved sammenligninger mellem tællingernes resultater; dog er det i betragtning af den nyeste undersøgelses udtalt stilistiske formaal uheldigt, at egennavne ikke er talt med. I sin første tælling (udelukkende af skolestof) medtog Noesgaard egennavne paa linie med andre ord, med det resultat at *Kefas* saavel som andre mere danskklingende navne optræder blandt tællingens 600 hyppigste ord. For opstillingen af et fælles ordinventar er et saadant resultat aabenbart meningsløst, og i sine følgende tællinger udelod Noesgaard derfor egennavne (personer, steder, maa-neder og ugedage). Hvor det drejer sig om undersøgelser til belysning af en teksts ordstatistiske struktur, anderledes sagt: belysning af hvad der gør en række ord til en tekst, eller om stilistiske undersøgelser, er denne fuldstændige udeladelse af egennavne imidlertid blot den anden grøft. For ikke alene bliver ved en saadan udeladelse alle andre ords hyppighedsprocent større, men samtidig forrykkes syntaktisk og stilistisk væsentlige talforhold, idet f. eks. substantivprocenten bliver mindre i forhold til verbalprocenten, og talforholdet mellem subjekt og finit verbum forrykkes. I tekster af visse arter, f. eks. avisstof og visse typer fiktionsprosa, bliver saadanne forskydninger ved udeladelse af egennavne ganske væsentlige. Da det imidlertid sjældent er selve de individuelle navnes hyppighed, som har interesse, er den fordelagtigste praksis at tælle navneforekomster under ét, evt. specificeret i personnavne, stednavne og andre navne.

Bortset fra de – antagelig smaa – forskydninger, som ville følge ved at medtage egennavne i betragtningen, frembyder Noesgaards nyeste og tidligere tællinger over sproget i danske forfatters produktion et solidt og righoldigt materiale saavel for videregaaende stilistiske sammenligninger mellem forfatterne enkeltvis og periodevis, hvad Noesgaard saa som sit formaal, som for undersøgelser over kvantitative forhold i teksters syntaktiske struktur. Det er ikke her stedet at drøfte Noesgaards resultater, de konklusioner som kan drages derudfra, eller de videre undersøgelser som inspireres deraf. Som et simpelt eksempel kan blot nævnes, at naar Noesgaard for ti forfattere finder nogenlunde samme procentdel verber i finit form (nutidsform + datidsform), tør

man deraf slutte, at ogsaa den gennemsnitlige sætningslængde er nogenlunde den samme for de undersøgte tekster, hvilket igen vækker interessen for undersøgelser over tekstlængdernes fordeling: Hvordan varierer sætningslængden hos den enkelte forfatter, og viser variationen samme billede for forskellige forfatters vedkommende?

Ogsaa forhold vedrørende forfatternes ordvalg belyses ved Noesgaards tællinger, saaledes forskelle i synonymers udbredelse og ikke mindst det almene spørgsmaal om ordrigdom, dvs. hvorvidt der hos visse forfattere og i visse arter af tekster gøres brug af et større antal forskellige ord end hos andre. At naa til holdbare resultater i saa henseende rummer imidlertid betydelige problemer. Dels er den talmæssige sammenhæng mellem tekstlængde og antal forskellige ord i teksten vanskelig at formulere matematisk, hvorfor det er vanskeligt at drage sammenligninger mellem tekster af ulige længde. Dels spiller de undersøgte teksters indholdsmæssige homogenitet en afgørende rolle, hvilket betyder at man faar stærkt afvigende resultater; en samling af flere tekststykker, f. eks. en novellesamling, giver et andet resultat (som regel i retning af flere forskellige ord) end et sammenhængende tekststykke, f. eks. en roman, af samme totale længde. Noesgaards resultater angaaende ordrigdom naar da faktisk heller ikke mere end at belyse sagens problemer. Det kan f. eks. let paavises ud fra Noesgaards opgivelser, at hvor hans udvalg af en forfatters produktion bestaar af flere indbyrdes uafhængige tekststykker finder han gennemgaende flest forskellige ord; den større ordrigdom hos den paagældende forfatter er imidlertid antagelig fiktiv.

Disse og andre uløste problemer er Noesgaards arbejder dog ingenlunde alene om. Spørgsmaalene beskæftiger fremdeles ordstatistikere i mange lande, og ikke helt faa udenlandske arbejder af tidligere dato har i dag temmelig problematisk værdi, fordi deres almene konklusioner er draget paa et ikke blot utilstrækkelig stort, men ogsaa utilstrækkelig defineret grundlag. En saadan skæbne vil ikke overgaa Aksel Noesgaards ordstatistiske arbejder, der i kraft af deres nøgternhed og klare forbindelse mellem de principielle problemer og et velafgrænset formaal staar som forbilledlig forskning, uanset om man skulle naa saa vidt, at mere omfattende undersøgelser gør selve hans resultater overflødige.