

IVADINIAI DALYKAI: TIKSLIEJI METODAI IR KALBOS TYRINĖJIMAS

I. PAVADINIMO MOTYVACIJA IR BENDRA SITUACIJA

1. METODOLOGINIŲ NUOSTATŲ POLIARIŠKUMAS

Žvelgiant į įvairias humanitarinių mokslų „ideologijas“ apskritai bei metodologines kalbos tyrinėjimų nuostatas konkrečiai, nesunku išvelgti tam tikrą jų poliarizaciją, kurią šiaip jau bene logiška būtų sieti su bendrąja *intuityvizmo* ir *pozityvizmo* priešprieša, paženklusia beveik visą kelių paskutiniųjų amžių kultūrą. Apie intuityvistinės orientacijos polių besitelkiantys kalbos tyrinėjimai daugiausia remiasi tyrinėtojo išvalgumu, nuojauta, jo proto ir sielos galiomis, o tyrinėjimai, vienaip ar kitaip besigrežiantys į pozityvizmą, labiau linkę prie objektyvių duomenų, eksperimentų bei viso to, kas egzistuoja objektyviai ir nelabai tepriklauso nuo žmogaus nuostatų, nuo jo aiškinimų ar spekuliacijų. Būtent šio pobūdžio lingvistiniai tyrinėjimai ir leistų kalbėti apie tikslųjų metodų taikymą kalbotyroje. Pozityvizmas apskritai orientuojasi į *gamtos* mokslus – ir į jų objektą (tai, kas egzistuoja *objektyviai*), ir į metodus. Todėl, nors kalbotyra šiaip jau yra humanitarinis („žmogaus“) mokslas, statistiniai kalbos tyrinėjimai paprastai yra nukreipti į vaizdžiai pasakius, jos „fiziologiją“, o ne į „ideologiją“. Bet savaime suprantama, jog lingvistui būtina pažinti ir kalbos „fiziologiją“...

2. TIKSLIŲJŲ METODŲ PRIGIJIMO PRIEŽASTYS

Tikslųjų metodų skverbimosi į kalbotyrą (paimtą plačiaja prasme – kartu su stilistika, poetika, eilėdara) priežastys šiais laikais būtų dvi pagrindinės:

- a) moderniosios kalbotyros ryšiai su semiotika, o per ją – ir su kitomis „kibernetikos šeimos“ disciplinomis, pvz., su informacijos teorija
- b) kompiuterių technikos vystymasis ir virtimas teksto apdorojimo priemone, vad. *inžinerinės lingvistikos* (automatinė teksto analizė ir referavimas, programavimo kalbos, garsinės kalbos automatinis atpažinimas) atsiradimas .

3. PAVADINIMO PROBLEMA

Kalbos tyrinėjimų, paremtų tiksliaisiais metodais, sfera iki šiol neturi vienareikšmio, galutinai nusistovėjusio pavadinimo: be jau minėto *inžinerinės lingvistikos* termino yra vartotas ir *mašininė lingvistika*, ir kitokie; iš jų plačiausiai paplitęs ir labiausiai žinomas yra terminas *matematinė lingvistika*. Iš pradžių ir šis kursas buvo taip vadinamas. Tačiau „klasikinė“ matematinė lingvistika iš tikrųjų yra kur kas „platesnių horizontų“ disciplina, negu tiksliaisiais metodais besiremiantys mūsų kalbininkų darbai: ji „siekia“ ir aibių teorijos taikymą kalbotyroje, ir kalbos procesų matematinį modeliavimą ir kitas mums gerokai neįprastas sritis. O mūsų kalbininkų darbai daugiausia orientuojasi į įvairius matematinės *statistikos* metodų taikymus. Todėl šio kurso, kurio pagrindinis tikslas ir yra nuosekliau supažindinti su tais metodais, pavadinimas buvo susiaurintas: dabar šis kursas gana sąlygiškai pavadintas *statistinės lingvistikos* kursu.

4. TAI – „STATISTINIŲ INSTRUMENTŲ“ RINKINYS

Savaime suprantama, kad tokia *statistinė lingvistika*, kokia ji turėtų atsiskleisti iš šio kurso, nėra kokia nors nauja, „neregėta“ ar ypatinga disciplina, ji net nėra vientisas „mokslas“ įprastine to žodžio prasme. Greičiau jau tai būtų specifinių, su *matematine statistika* bei prie jos labai glaudžiai prišlijusia *tikimybių teorija* susijusių metodų, pasakytum – instrumentų, tam tikras rinkinys ir kartu – išmanymas, kaip tuos instrumentus pasirinkti ir taikyti vienu ar kitu atveju, kai prisieina spręsti vienokius ar kitokius kalbotyros klausimus.

II. KODĖL KALBININKUI PRAVARTU IŠMANYTI STATISTINIUS METODUS

1. APSTU TOKIO POBŪDŽIO DARBŲ

Dėl to, kad bent jau mūsųose egzistuoja didelė daugybė tais metodais paremtų kalbininkų darbų: kad tuo įsitikintum, pakanka pažvelgti į „Kalbotyros“ ir „Baltistikos“ žurnalus, į didžiumą disertacijų (ypač – prof. A. Girdenio disertantų). Statistinių metodų ar bent jau su tais metodais itin glaudžiai susijusių informacijos pateikimo formų ir pavidalų (lentelių, diagramų, grafikų ir pan.) taikymas kalbos tyrinėjimams tapo kone kasdienybe, norma: verstis be jų būtų kone nepadoru...

2. REIKIA IŠMANYTI, KĄ DARAI

Tačiau dažnokai net patys tų darbų autoriai ne iki galo tesupranta, ką, dėl ko ir kaip jie daro: dažnai jie arba tiesiog kone paradižiui vykdo savo mokslinių vadovų nurodymus, arba pusiau akiai „nusižiūri“ ir formules, ir grafikų ar lentelių pavidalus iš kitų, atseit labiau išmanančių autorių ir net nebemėgina kelti klausimo, ar tie dalykai jų reikmėms

tinka, ar ne. To neturėtų būti: kiekvienas autorius turėtų labai blaiviai ir aiškiai suvokti metodų, kuriuos jis panaudoja, prigimtį, paskirtį ir prasmę, išmanyti įvairius jų niuansus. Šis kursas ir turėtų tam pasitarnauti.

3. KALBA – SFERA, KURIOJE APSTU ATSITIKTINYBĖS IR STATISTIŠKAI APDOROTINŲ FAKTŲ
- Bet visų svariausia priežastis, dėl kurios kalbininkui verta susipažinti su statistiniais metodais, glūdi pačioje kalboje! Kalbai, o ypač jos „performensui“ (kalbėsenos aktams, tekstams) būdinga didelė gausybė įvairiai tarpusavy besisiejiančių elementų, dėmenų, komponentų, kurių jau vien fiksavimas ir registravimas būtų iš esmės statistinio pobūdžio uždavinys. Statistika (kol kas dar net ne „matematinė“, o stačiai statistika, gal daugiau deskriptyvinio pobūdžio, pasakytum, aprašomoji statistika) kaip tik „prasideda“ ten, kur susiduriama su faktų, elementų, dėmenų ir pan. dalykų gausa bei poreikiu juos suregistruoti. Be to, kalbos elementų (komponentų) sekose (grandinėse, srautuose) esama gana daug atsitiktinybės, o ją suvaikyti ir sučiuopti vienokius ar kitokius jos dėsningumus (atsitiktinybei dėsningumai taip pat itin būdingi!) – jau yra matematinės statistikos bei jos „sesers“ tikimybių teorijos „darbas“. Ypač tos atsitiktinybės daug išryškėja, kai imama galvoti apie įvairių kalbos elementų tarpusavy ryšius ir priklausomybes. Tradicinė kalbotyra *atsitiktinius* kalbos elementų ryšius šiaipjau yra linkusi lyg ir atmesti, neskirti jiems dėmesio; net pats toks apibūdinimas *atsitiktinis* joje turi beveik neigiamą niuansą: atseit, čia nieko tikro pasakyti negalima, gali būti ir vienaip, ir kitaip, aiškaus ir vienareikšmiško priklausomumo ar dėsningumo nėra. Tačiau be to aiškaus, vienareikšmio ar, tiksliau pasakius, *funkcinio* priklausomumo (sąryšingumo), dar yra galimas ir universalesnis jo atvejis – *stochastinis* (arba tikimybinis) sąryšingumas, be abejo, siejantis vienus su kitais ir daugely kalbos elementų, ypač – priklausančių skirtingiems struktūriniais lygmenims. Funkcinio ir stochastinio sąryšingumo tarpusavy santykis būtų iš dalie panašus į realiųjų skaičių santykį su kompleksiniais skaičiais. Kita vertus, stochastinio pobūdžio sąsajoms pažinti bei modeliuoti labiausiai ir tinka statistiniai–tikimybiniai metodai. Todėl norėdamas aprėpti kuo gausesnius kalbos faktus ir situacijas, kuriose jie vienaip ar kitaip atsiskleidžia bei funkcionuoja, kalbininkas ir turėtų mokėti pasinaudoti tam darbui šiaipjau pravarčiais „statistiniais instrumentais“.

4. PLAČIAU APIE DVI – FUNKCINĘ IR STOCHASTINĘ – SĄRYŠINGUMO ATMAINAS

Kadangi paminėtieji du sąryšingumo tipai mūsų atveju yra labai svarbūs, juos derėtų pakomentuoti plačiau.

Funkcine (kitaip sakant, vienareikšme, tiesiogine, deterministiška) priklausomybe tarpusavy būna susiję tokie dydžiai, faktai ar reiškiniai, kurių vieno koks nors pokytis vienareikšmiškai sąlygoja atitinkamą kito pokytį. Ši priklausomybė taip ir vadinama dėl to, kad vienas dydis (faktas, reiškinys) yra tarsi *argumentas*, o kitas – jo *funkcija*. Tokias priklausomybes palyginti lengva yra išreikšti vadinamuoju *analitiniu* būdu, t. j. užrašyti, pasitelkus sutartinius žymėjimus, atitinkamomis formulėmis. Puikūs tokių priklausomybių pavyzdžiai galėtų būti iš esmės visi garsieji mokykliniai fizikos bei matematikos dėsniai. Pavyzdys galėtų būti kad ir iš „tolimų“ mokyklos laikų žinoma skritulio ploto formulė $S = \pi \times r^2$, aiškiai nurodanti, kas skritulio plotas S yra jo spindulio r funkcija, ta formule išreiškiamą (aprašoma). Arba garsusis Omo dėsnis, tarpusavy susiejantis elektros grandinės varžą su joje pasiskirstančia įtampa ir tekančia srove). Visur čia galima išvelgti tam tikrą vieno dydžio priklausomybės nuo kito universalumą ar bet absoliutumą. Panašiai būna ir kalboje: jeigu, sakysim, koks nors veiksmožodis valdo tik galininką, tai natūralu, jog rišliame tekste žodis, su tuo veiksmožodžiu sudarantis junginį, bus pavartotas tik galininko linksniu. Matematinei statistikai ar kokiems šiaip tikimybiniais modeliams tokiu atveju tiesiog „nėra kas veikti“.

Stochastinė (t. y. nevienareikšmė, atsitiktinumu paremta, kai kada ryškiai pasireiškianti, o kai kada – niekaip neatsiskleidžianti) priklausomybė paprastai sieja ne tiesiogiai, o *tarpiskai* – per daugelį įvairiopaipersipynusių ir dinamiškai besikeičiančių aplinkybių – vienas su kitu tesusijusius dydžius, faktus bei reiškinius, todėl ji nebūna tokia akivaizdi ar – tuo labiau – absoliuti, ir jos neįmanoma vienareikšmiškai perteikti analitinėmis formulėmis. Todėl ten, kur baigiasi vienareikšmių funkcinių priklausomybių valdos ir įsigali daugiareikšmės stochastinės priklausomybės, ten, norint apibūdinti tos priklausomybės pobūdį, prisiėina išsileisti „netiksliai“ ir, iš pirmo žvilgsnio atrodytų, netgi antimoksliską *tikėjimo* kategoriją: esant tam tikroms sąlygoms galima *tikėti(s)*, jog įvykus vienam įvykiui, įvyks ir kitas (pvz., kai kurie biologijos, genetikos, socialinio gyvenimo ir pan. dėsniai). Grįžtant prie panašaus kaip anksčiau kalbinio pavyzdžio, dabar jau tiktų imti veiksmožodį, galintį valdyti nebe vieną, o du (ar kelis) linksnius. Tad žodis, su tuo veiksmožodžiu sudarantis junginį, rišliame tekste gali būti pavartotas iš principo bet kuriuo iš tų „valdomųjų“ linksnių. Dar daugiau: dabar jau galima kelti klausimą, kokioms „konteksto sąlygoms“ esant *labiau* galima *tikėti*s (ir – kiek labiau) jį būsiant pavartotą vienu ir kokioms – kitu iš tų linksnių. Štai čia ir prasideda tikroji statistinių–tikimybių metodų „stichija“.

III. KO REIKIA NORINT PASIGILINTI Į STATISTINĘ LINGVISTIKĄ

1. TEIGIAMO POŽIŪRIO Į TIKSLIUOSIUS METODUS IR JŲ RAIŠKOS PRIEMONES

Šiame kurse neišvengiamai susidursime su dviem gana specifiskais ir toli gražu ne kiekvienam filologui maloniais bei priimtinais dalykais – su *skaičiais* ir *formulėmis*. Nei statistika apskritai, nei, juo labiau, matematinė statistika apsieiti išvis be jų – niekaip negali. Bet taip pat nėra jokia paslaptis, kad tarp filologų kartais pasitaiko žmonių, kuriems tie dalykai, skaičiai ir formulės, – arba vienas kuris iš jų, arba dažniausiai abu drauge – tiesiog „kelia alergiją“, erzina, pykdo. Esu sutikęs žmonių, kurie tuo net didžiuojasi... Jeigu kam taip jau yra, jeigu skaičiai ir formulės atrodo nepriimtini, atgrasūs ar bjaurūs patys savaiame, tam šito kurso geriau atsisakyti iš karto. Jeigu gi nepaaiškinamo pykčio ir pasibjaurėjimo skaičiams nei formulėms neįaučiate – statistinės lingvistikos imtis galite.

2. PRISIMINTI MOKYKLINĖS MATEMATIKOS PAGRINDUS

Nors skaičių ir formulų apskritai neišvengsime, bus stengiamasi jomis bent jau nepiktinaudžiauti ir verstis mokyklinės matematikos priemonėmis, pasakytum, „išsitemkti jos ribose“. Šį momentą pabrėžti būtina dėl tokios aplinkybės: yra

išleista labai daug matematinės statistikos ir tikimybių teorijos knygų, skirtų visų pirma inžinerinių profesijų žmonėms bei techninių specialybių studentams, kurie yra išsamiai susipažinę su aukštąja matematika, tad jose apie tikimybių teorijos ir matematinės statistikos dalykus paprastai kalbama aukštosios matematikos „kalba“ (tokia yra kad ir garsioji J. Kruopio „Matematinė statistika“). Tačiau lygiai tuos pačius dalykus dažniausiai galima visai suprantamai paaiškinti ir naudojantis elementariosios (mokyklinės) matematikos „kalba“ (sąvokomis, formulių išraiškomis ir kt. raiškos priemonėmis; tik rasti šia „kalba“ paremtų pakankamai turiningų ir dalykiškų knygų yra kiek sunkiau, nors iš principo irgi galima). Šitai čia ir bus stengiamasi daryti, nes aukštosios matematikos „raiškos priemonės“ filologams kažin ar parankios. Tad mokyklinės matematikos pagrindus prisiminti vis dėlto reikia.

3. KOKIOS NORS SKAIČIAVIMO PRIEMONĖS

Be abejo, kuo ta priemonė tobulesnė, tuo geriau. Nors iš bėdos galima verstis ir geru, patikimu kalkuliatoriumi, labai pageidautinas vis dėlto būtų normalus šiuolaikinis personalinis kompiuteris su būtent skaičiavimo darbams pritaikyta programine įranga. Per užsiėmimus naudosisimės kompiuterių klase, bet kompiuterio reikia ir dirbant savarankiškai. Kompiuterį galima (ir reikia) panaudoti ne tik patiems statistiniams–tikimybiniais skaičiavimams, bet taip pat ir *pirminių duomenų rinkimui*, jeigu tik tai iš principo įmanoma ir jei turima tam reikalinga programinė įranga.

Programinė įranga (programa) gali būti tiek universali, skirta iš esmės bet kokiems skaičiavimo darbams (per užsiėmimus viena iš tokių programų ir naudosisimės), tiek ir *specializuota*, t.y. specialiai pritaikyta būtent statistiniams skaičiavimams. Tokių specializuotų programų, skirtų iš esmės profesionalams, irgi yra ir esant reikalui jų galima įsigyti. Paminėtini būtų specializuoti paketai *Statistica* (varianto „for Windows“ v. 5.0 ir aukštesnės) bei *SPSS* (v. 8.0).

Pasinaudojant kompiuteriu, iš tikrųjų nesunkiai galima atlikti išties milžiniškus skaičiavimo darbų kiekius, tūkstančių tūkstančius sudėtingiausių skaičiavimo veiksmų. Statistika yra ta prasme imlus ir iš tiesų didžiai daug „juodo darbo“ reikalaujantis veiklos baras. Iš tikrųjų turbūt yra gerokai pagrindo įtarti, kad 7-ojo dešimtmečio antroje pusėje ir 8-ajame dešimtmetyje pasireiškęs gana ryškus, kone masinis filologų susižavėjimas statistiniais metodais atslūgo gana spėriai kaip tik dėl to, kad didžiuliai susižavėjusiųjų paprasčiausiai neužtekdavo kantrybės atlikti visus reikalingus skaičiavimus (skaičiuoti dažniausiai tekdavo arba „ranka“, t.y. su pieštuku ir popieriumi, arba gremėzdžiškais ir kebliai valdomais tuometiniais skaičiavimo aparatais), „iki galo“ iškęsdavo tikrai tikrieji tų metodų fanatikai... Dabar praktiškai visus skaičiavimus, kad ir kokie sudėtingi ir didžiuliai jie bebūtų, galima išties pavesti kompiuteriui ir jis su jais paprastai susidoroja per kelias ar keliolika sekundžių. Žmogui belieka tik galvoti apie skaičiavimo užduotis ir gautų rezultatų interpretavimą.

IV. PAGRINDINĖS STATISTINIŲ METODŲ PRITAIKYMO KALBOTYROJE SRITYS

1. EKSPERIMENTINĖ FONETIKA

Pirmiausiai paminėtini kalbos garsinio signalo tyrinėjimai, kitaip dar vadinami *eksperimentine kalbotyra*, *eksperimentine fonetika*. Tyrinėtojas čia kokiu nors būdu kokioje nors laikmenoje užsifiksuoja kaip galima daugiau (ar bent jau – pakankamai daug) reprezentatyvių tiriamojo signalo pavyzdžių ir paskui, pakartotinai atkurdamas įrašus, *matuoja* rūpimus signalo *parametrus* (paprastai – trukmę, dažnį ar intensyvumą, rečiau – bendrąjį energijos kiekį, spektro tankį ir pan.). Visi tokie garsinio kalbos signalo parametrai „iš prigimties“ yra ne kas kita, kaip *atsitiktiniai dydžiai*, pasiskirstę pagal vienokius ar kitokius dėsningumus. Aišku, tokiems tyrimams reikalinga speciali įranga, aparatūra, todėl pakenčiami eksperimentiniai kalbos signalo tyrinėjimai būdavo įmanomi tikrai specialiai įrengtose laboratorijose. Dabar dažniausiai naudojamas specialiai tam pritaikytas ir specialią programinę įrangą turinčiais kompiuteriais; tokios paskirties programų taip pat jau yra pakankamai daug. Čia kaip tik ir atsiveria viena iš galimybių pritaikyti kompiuterį pačiam statistiškai apdorosimų *duomenų rinkimui*. Papildomos aparatūros tokiam „kompiuteriniam“ eksperimentinės fonetikos kompleksui tereikia palyginti nedaug: gerų įtaisų signalo įrašams (laikmenoms) skaityti, mikrofono („gyvam“ balsui įvesti) ir pakankamai geros, bet paprastai standartinės garso plokštės.

O statistiniai bei tikimybiniai skaičiavimai šiuo atveju pasitelkiami dviem tikslais: viena, matavimo *rezultatams* bei matavimo *paklaidoms įvertinti* (kalbant apie mūsų kalbininkų darbus čia vis tikrai reikėtų tam tikros išlygos...), ir antra – vienu eksperimentų *rezultatams* deramai *palyginti* su kitų eksperimentų rezultatais, jų *panašumams* bei *skirtumams* įvertinti. Savaiame suprantama, tas įvertinimas būna irgi *tikimybinio* pobūdžio: siekiama nustatyti, *kokiu mastu tikėtina*, kad rezultatų panašumas (arba skirtumas) yra esmingas, ir kokiu – kad jis neesmingas, sąlygotas vien tik „pačios“ atsitiktinybės.

2. TEKSTOMETRINIAI TYRINĖJIMAI

Tekstometriniais tyrinėjimais, kurie kasdieninėje kalboje dažnai vadinami tiesiog *lingvistiniais skaičiavimais* ar netgi stačiai *kalbos statistika*, čia pavadinti tokie tyrinėjimai, kuriais siekiama išsiaiškinti, kaip vieni ar kiti kalbos elementai funkcionuoja vienokios ar kitokios paskirties bei pobūdžio *tekstuose*. Pirminis informacijos šaltinis čia dažniausiai yra konkretus kalbos aktas, tekstas, paprastai – rašytinis. Iš jo kiekvienas tyrinėtojas gali išsirinkti, susiregistruoti ir susiskaičiuoti tai, kas jam rūpi: garsus, raides, skiemenis, žodžius, vienokios ar kitokios sandaros žodžių junginius, sakinius, kalbos dalis, gramatines formas ir t. t. ir pan. Tokie skaičiavimai dažniausiai daromi tekstų aprašymo (deskriptyviniais) ar palyginimo (komparatyviniais) tikslais.

Kalbant apie tekstometriją ir įvairių teksto elementų (komponentų) skaičiavimą, iš pat karto pabrėžtinas toks dalykas: patys tų elementų (raidžių, garsų, žodžių, sakinių, skiemenų ir pan.) *kiekiai* tiesiogiai priklauso nuo tiriamųjų tekstų didumo (ilgumo): kuo tekstas ilgesnis, tuo, savaiame suprantama, visokių elementų jame bus daugiau. Todėl duomenys

apie tokių elementų kiekius *skirtingo* ilgio tekstuose tarpusavy būna „nebendramatiški“ ir lyginti juos vienus su kitais tiesiogiai – būtų nepagrįsta ir stačiai nelogiška. Todėl vietoj kiekių dažniausiai yra imamos tiriamųjų elementų *proporcijos* (t.y. procentais ar vienetu dalimis išreikšti santykiai), kuriomis jie sutinkami tuose tekstuose. Tai gana ryški tekstometrinių duomenų specifika: juos paprastai tenka versti santykiniais dydžiais, procentais. Tik tais retais atvejais, kuomet lyginamieji tekstai yra *vienodo ilgumo*, galima tiesiogiai lyginti ir pačius tiriamųjų elementų *kiekius*.

Kadangi tų ar kitų kalbos elementų proporcijos (kiekiai) konkrečiuose tekstuose irgi yra dideliu mastu *atsitiktiniai* dydžiai, o jų tarpusavio sąsajos – dažniausiai stochastinio pobūdžio, tai norint korektiškai pasiremti tokiais duomenimis ir čia neįmanoma išsiversti be atitinkamų statistinių bei tikimybinių skaičiavimų. Tekstometriniai metodai šiaipjau kokios nors specialios įrangos ar aparatūros kaip ir nereikalauja, vien tik tyrinėtojo kantrybės ir atidos, tačiau norint gauti bent kiek patikimesnius (reprezentatyvesnius) rezultatus bent jau anksčiau, kol rūpimi teksto elementai būdavo išrenkami, registruojami ir skaičiuojami rankomis, jiems reikėdavo palyginti labai daug kruopštaus triūso, ir buvo beveik įsigalėjusi nuomonė, kad jiems reikalingų darbo sąnaudų rezultatų vertė neatperkanti. Padėtis gerokai yra pasikeitusi dabar, nes ir tekstometrinių duomenų rinkimui irgi puikiai gali būti pritaikomi kompiuteriai. Deja, beveik nėra (o gal kur nors ir yra, tik labai sunku rasti?) tiems darbams tinkamos *gatavos* programinės įrangos. Tenka pasitelkti arba sudėtingesnių programų įvairius „pagalbinius efektus“ (sakysim, galimybę kurti savus makrosus), arba tokios paskirties programas ištaisai rašyti patiems. Todėl ir vienu, ir kitu atveju labai praverčia mokėjimas *programuoti*.

Tekstometriniai tyrinėjimai dažniausiai būdavo ir yra daromi deskriptyvinės kalbotyros, kalbinės bei literatūrinės stilistikos, poetikos bei eilėtyros interesais (puikūs tokių eilėdaros veikalų pavyzdžiai būtų J. Girdzijausko „Lietuvių eilėdara: XX amžius“ arba A. Girdenio studija apie K. Donelaičio hegzametą). Tekstometriniais tyrinėjimais gali remtis ir įvairūs *ekspertinio* pobūdžio darbai: nežinomo teksto tikėtiniausios autorystės nustatymas, tekstų chronologijos nustatymas ir pan.

3. ANKETINIAI (SOCIOMETRINIAI) TYRINĖJIMAI

Tai įvairios *respondentų apklausos*, kitaip tariant, *anketiniai* arba *sociometriniai* tyrinėjimai. Jų ypatumas tas, kad čia iš esmės remiamasi *kolektyvine* vienaip ar kitaip parinktų žmonių, vadinamųjų respondentų arba ekspertų, *nuomone* ir *sprendimais* (vox populi – vox dei?). Respondentų vertinimui gali būti pateikiami patys įvairiausi „objektai“, sakysim, garsų panašumas, žodžio keliamos asociacijos, kokio nors posakio skambesio grožis ir t. t. Pavyzdys galėtų būti A. Žuravlio darbuose aprašyti sociometrinio pobūdžio eksperimentai, kuriais buvo siekiama atskleisti ir paliudyti garsų simbolikos egzistavimą. Šie tyrinėjimai bent mūsųose nėra tokie populiarūs kaip pirmosios dvi atmainos ir yra apskritai kiek problemiškesni. Pirma, reikia apgalvotai bei motyvuotai parinkti bent jau teoriškai pakankamą respondentų kiekį, ir iš to kyla savų, specifinių problemų. Antra, anketų klausimus reikia pateikti taip, kad kiekvienas respondentas iš tikrųjų pateiktų būtent *savą* nuomonę ar sprendimą, kuo mažiau įtakotą klausimų formulavimo ar kitų panašių veiksnių. Trečia, anketinių (sociometrinių) apklausų rezultatai visuomet būna kompleksiški ta prasme, kad juose savaip susipina vertinimui pateikiamo *objekto* tie ar kiti ypatumai su pačių vertintojų, respondentų (t.y. vertinančiųjų *subjektų*) ypatumais, ir norint tuos du komponentus pakankamai sąmoningai atsieti, reikia daug apdairumo bei įžvalgos. Beje, tais atvejais, kai anketiniai (sociometriniai) tyrimai daromi tikslingai siekiant įvertinti ne tiek anketose nurodomus objektus, kiek būtent pačius subjektus (vertintojus, respondentus), jie jau tampa nebe tiek lingvistiniais, kiek *psichologiniais* ar bent jau artimais psichologiniams tyrimais. O psichologams tam tikros medžiagos tokie tyrimai duoda visuomet, net ir tada, kai yra daromi tikrai kalbotyros interesais. Beje, sociometriniai tyrimai gana mėgstami ir kitų humanitarinių mokslų: jais naudojasi etnologai, sociologai ir daugelio „gretimų“ sričių specialistai.

Natūralu, kad anketų įrašams apdoroti ir tyrinėtoją dominantiems duomenims iš jų rinkti taip pat tikslinga panaudoti kompiuterius, ir kuo „anksčiau“, tuo geriau. Idealu būtų, jeigu respondentai galėtų anketas matyti ir atsakymus į jiems skirtas vietas įrašyti tiesiog kompiuterio ekrane. Jei tai neįmanoma, anketų įrašus į kompiuterį tikslinga perkelti kaip galima anksčiau, rankomis su jais neatliekant jokių rūšiavimo ar panašių veiksnių: bus mažiau atsitiktinių klaidų. Tiesa, čia vėl gana aštriai iškyla tinkamos programinės įrangos klausimas ir praverčia mokėjimas programuoti...

Statistiniai-tikimybiniai skaičiavimai sociometrinio pobūdžio tyrimuose irgi būtina reikalingi, nes respondentų atsakymai natūraliai įvairuoja ir jų visuma yra ne kas kita, kaip vienu ar kitu būdu pasiskirsčiusių atsitiktinių dydžių sandaupa.

V. KO GALIMA IR KO NEREIKIA TIKĖTIS

1. GALIMA TIKĖTIS: NUOSEKLAUS DĖMESIO FAKTAMS

Kas ryžtasi imtis statistinių eksperimentų ir skaičiavimų, tam neišvengiamai prisieina skirti daug dėmesio pačiai „faktografijai“, faktų, kuriuos norime apdoroti, rinkimui ir suregistravimui. Kaip sakyta, tiems darbams irgi vertėtų pasitelkti kompiuterį, tačiau kompiuteris čia tegali būti tik „aklas juodadarbis“, nors ir labai spartus, o visą tų faktų „ideologiją“ tvarko, kokius faktus rinkti, ką ir kaip registruoti, – sprendžia tik pats žmogus. Kuo tų faktų užfiksuojama daugiau, tuo apskritai geriau. Tokiu būdu jų susitelkia didesnės ar mažesnės sandaupos, tam tikri masyvai, į kuriuos įsižiūrėdamas, tyrinėtojas jau gali daryti vienokias ar kitokias prielaidas ir parinkti „statistinius instrumentus“ tų prielaidų tikėtinumui „pasverti“. Toks galvojimas apie iš didelės faktų sandaupos keltinas įvairias, kartais net ir kiek fantastiškas prielaidas pats savaime yra gana provokuojantis dalykas ir gali mokslinę mintį pakreipti išties kokia nauja ar netikėta linkme.

2. GALIMA TIKĖTIS: OBJEKTYVUMO (GAL NET – OBJEKTYVIZMO)

Statistiniai metodai faktus rikiuoja ir grupuoja tik taip, kaip *nurodyta* atitinkamomis formulėmis: jokie žmogaus, tyrinėtojo, nusistatymai, – prielankumas ar neprielankumas, vienoks ar kitoks požiūris, vienokie ar kitokie siekiai, troškimai ar aistros – tam tiesioginės įtakos nedaro. Tad galima kalbėti ne tik apie gana ryškų nešališkumą ir objektyvumą, bet ir tam tikru mastu – apie iš esmės buką mašinalumą ir aklą faktų sklaidą be jokios atžvalgos. Galimas daiktas, kad taip yra net kiek peržengiama natūrali tikro, pozityvaus objektyvumo riba: jis juk reikalautų ne tik dėmesingai aprėpti visą turimų faktų visumą, bet ir atmesti tuos iš jų, kurie „į apyvertą“ yra patekę aiškiai per klaidą ar neapsižiūrėjimą. O vien tik statistiniai procedūromis tai padaryti vargu ar įmanoma. Todėl ir tenka kelti klausimą, ko daugiau esama įvairių statistinių skaičiavimų rezultatuose – tikro objektyvumo ar mechanistinio objektyvizmo. Viskas turbūt priklauso labiausiai nuo to, kokius ir kiek patikimus pirminius duomenis tiems skaičiavimams paimsi.

ĮSIDĖMĖTINA: norint padaryti *teisingą* išvadą, būtinai reikalingi *teisingi* ir patikimi pirminiai duomenys: nėra *jokių* metodų, formulių ar kokių kitų „statistinių priemonių“, kurios leistų padaryti teisingas išvadas iš klaidingų, neobjektyvių ar falsifikuotų pirminių duomenų. Tai – neįmanoma iš principo.

3. GALIMA TIKĖTIS: NAUJŲ IDĖJŲ

Jų dažnai šauna į galvą tiesiog bedirbant su „statistiniais instrumentais“ – begalvojant, kokie dėsningumai galbūt galėtų išryškėti iš turimų duomenų, kokie metodai ar procedūros jiems labiausiai tiktų, beanalizuojant skaičiavimų rezultatus, ypač – jeigu jie būna pateikti vaizdžiu pavidalu: lentele, brėžiniu, grafiku. Itin daug postūmių naujoms idėjoms galėtų duoti statistiškai tyrinėjami kalbos įvairių elementų tarpusavio ryšiai (koreliacijos) ir jų taksonomijos (grupavimas statistinių duomenų pagrindu į įvairias grupes). Tik toli gražu ne visos tos naujos idėjos pasirodo esančios perspektyvios, bet čia – jau kitas niuansas.

4. NEGALIMA TIKĖTIS: PIRMIAUSIA – FAKTŲ IR REZULTATŲ INTERPRETACIJOS

Tai savaime suprantama: interpretuoti, t.y. vienaip ar kitaip *paaiškinti* surinktus faktus ir jų pagrindu gautus įvairių apskaičiavimų bei procedūrų rezultatus tegali tik žmogus. Patys statistiniai metodai, procedūros ir formulės iš tikrųjų tėra tik *fiksavimo*, o ne interpretavimo „instrumentai“, ne tokia yra jų prigimtis ir paskirtis. O nuo interpretacijos priklauso labai daug kas: paviršutiniškai, neįžvalgiai ar stačiai atmetinai teapžvelgti statistinių tyrimų rezultatai gali net kompromituoti pačią tų tyrimų idėją, sudaryti išpūdį, kad jie neduodą beveik jokios naudos ir esą kone beverčiai. Ir atvirkščiai.

V. LITERATŪRA, TINKANTI STATISTINĖS LINGVISTIKOS KURSUI

- **Aprašomoji** statistika: Mokomoji priemonė. Vilnius: Vilniaus universiteto leidykla, 1998. 136 p.
- **Bitinas B.** Statistiniai metodai pedagogikoje ir psichologijoje. Kaunas: Šviesa, 1974. 208.
- **Čekanavičius V., Murauskas G.** Statistika ir jos taikymai, I. Vilnius: TEV, 2000. 240 p.
- **Čekanavičius V., Murauskas G.** Statistika ir jos taikymai, II. Vilnius: TEV, 2002. 272 p.
- **Januškevičius R.** Statistikos įvadas. Vilnius: Vilniaus pedagoginis universitetas, 2000. 172 p.
- **Kruopis J.** Matematinė statistika.
I leid.: Vilnius: Mokslo, 1977. 362 p.
II leid.: Vilnius: Mokslo ir enciklopedijų leidykla, 1993. 416 p.
- **Matematinė** statistika istorijoje / R. Eidukevičius, D. Juknevičienė, N. Kosareva, S. Pamerneckis. Vilnius: Vilniaus universiteto leidykla, 1998. 280 p.
- **Sakalauskas V.** Statistika su *Statistica*. Vilnius: Margi raštai, 1998. 227 p.
- **Borovikov V.** STATISTICA: iskusstvo analiza dannyh na komputere. Sankt Peterburg, 2001 [rusų k.]
- **Bjulj A., Cjofel P.** SPSS: iskusstvo obrabotki informaciji (analiz statističeskich dannyh i vosstanovlenije skrytych zakonomernostej. Moskva - Sankt Peterburg - Kijev: DiaSoft, 2002 [rusų k.]
- **Piotrovskij R. G., Bektaev K. B., Piotrovskaya A. A.** Matematičeskaya lingvistika. Moskva: Vysšaya škola, 1977. 383 p. [rusų k.]
- **Tiurin J. N., Makarov A. A.** Statističeskij analiz dannyh na komputere / Pod red. W. E. Figurnova. Moskva: Infra-M, 1998. 528 p. [rusų k.]