

LINGVISTIK TADQIQOTLARDA STATISTIK METODNING O‘RNI

Sirojiddinova Muxarram Rahimjon qizi

O‘zbekiston Milliy universiteti Kompyuter lingvistikasi
Mutaxassisligi 1- kurs magistranti.

E-mail: muharramxonsirojiddinova@gmail.com

Mirzarahimov Bahodir Shavkat o‘g‘li

O‘zbekiston Milliy universiteti Kompyuter lingvistikasi
mutaxassisligi 1- kurs magistranti.

E-mail: Bahodirmirzarahimov70@gmail.com

Annotatsiya: XXI asr boshlarida axborot texnologiyalarining jadal rivojlanganligi sababli nafaqat globallashuv jarayoni, balki ta’lim tizimi ham yangi pog‘onalarga ko‘tarildi. Ushbu bosqichda barcha sohalar, xususan, ilm-fan tezkorlikni, tez sur‘atlar bilan rivojlanishni talab qilmoqda. Shu sababli, inson omilidan ko‘ra kompyuter, texnika imkoniyatiga tayanish muhim ahamiyatga ega. Natijada, tilshunoslik va lingvistika birlashib, kompyuter lingvistikasi sohasi shakllandi va rivojlanmoqda. Ushbu tezisda kompyuter lingvistikasi uchun statistik metodning o‘rni va taraqqiyotiga doir ayrim mulohazalar bildirilgan.

Kalit so‘zlar: statistika, statistik metod, chastotali lug‘at, SAS (Statistical Analysis Software).

Abstract: At the beginning of the 21st century, due to the rapid development of information technologies, not only the process of globalization, but also the education system rose to new levels. At this stage, all fields, especially science, require rapid development. Therefore, it is important to rely on the computer and technical capabilities rather than the human factor. As a result, linguistics and linguistics merged, and the field of computational linguistics was formed and is developing. This thesis presents some comments on the role and development of the statistical method for computer linguistics.

Keywords: statistics, statistical method, frequency dictionary, SAS (Statistical Analysis Software).

Аннотация: благодаря бурному развитию информационных технологий в начале XXI века не только процесс глобализации, но и система образования поднялись на новую высоту. На данном этапе все области, особенно наука, требуют быстрого развития. Поэтому важно полагаться на компьютер и технические возможности, а не на человеческий фактор. В результате лингвистика и лингвистика объединились, и образовалась и развивается область компьютерной лингвистики. В этой диссертации представлены некоторые комментарии о роли и развитии статистического метода в компьютерной лингвистике.

Ключевые слова: статистика, статистический метод, частотный словарь, SAS (программное обеспечение для статистического анализа).

Statistika soʻzi lotincha “status”, yaʼni ahvol, holat degan soʻzning italyanacha “state” - davlat degan maʼnoni beradi. Statistika soʻzi fanga XVIII asrda kirib keldi. Dastlab, savdo va moliya kapitali hamda pul munosabatlari taraqqiyoti natijasida vujudga kelgan ehtiyojni qondirish maqsadida toʻplangan mamlakat aholisi, iqtisodiy va siyosiy ahvoli haqidagi maʼlumotlar statistik axborotlar deb yuritilgan [A.A.Mirisaev, 2021].

Bundan 3500 yil oldin Qadimgi Misrda oʻtkazilgan aholi hisobi, Rossiyada Pyotr I davridagi «aholi taftishi», Ajigliyadagi «Dahshatli sud kitobi» kabilar dastlabki statistik maʼlumotlardir. Lekin statistika fan sifatida XVII asrda Angliyada paydo boʻlgan. Siyosiy arifmetika maktabining namoyondalari Jon Graunt (1620-1674) va Uilyam Petti (1623-1687) statistika asoschilari hisoblanadilar. Chunki birinchi boʻlib ijtimoiy-iqtisodiy tekshirishlarda statistikani qoʻllash fikri U. Pettida tugʻilgan edi.

U. Petti oʻz asarlarida (“Soliqlar va yigʻimlar”, 1662-yil; “Donolarga soʻz”, 1664-yil; “Siyosiy arifmetika”, 1676-yil) oʻz tengdoshlarini qiziqtirgan sotsial va iqtisodiy savollarga miqdoriy hisob-kitoblarga suyangan holda javob berishga harakat qiladi [Prof. X.Shadiyev, I.Xabibullayev, 2013].

Keyinchalik statistika barcha fanlar uchun umumiy metod sifatida boshqa fanlarda ham keng qoʻllanila boshladi. Xususan, tilshunoslikda, lingvistik tadqiqotlarda ushbu metodning oʻrni beqiyosdir. Koʻplab olimlar statistik metodning qachon va qay tarzda tilshunosligimizga kirib kelishi, lingvistik tadqiqotlardagi oʻrni haqida turlicha fikr yuritganlar. Masalan, Xolmatova Z oʻzining “Tilshunoslik nazariyasi” darsligida u haqida qisqa va loʻnda shunday fikr bildirgan: “Statistik metod tilning barcha sathlarida qoʻllanadi, shuning uchun universal hisoblanadi. Bu metod til birliklari va til hodisalarining nutqda ishlatilish darajasini aniq koʻrsatib beradi. U natijalarni matematik asosda hisoblablash usulidir”.

Filologik tadqiqotlarda statistik metodni tilshunoslikka olib kirgan olim V.Ya.Bunyakovskiy hisoblanadi. Matematik olim Yu.I.Levin fikriga koʻra, XX asrning boshlarida til hodisalarini oʻrganishda ehtimollik nazariyasi va statistikani qoʻllash boshlandi. Bu soha bilan dastlab shugʻullangan shoir Andrey Belin va taniqli matematik A.A.Markovlar boʻlgan. Keyinchalik A.A.Chuprov, A.M.Peshkovskiy, V.F.Chistyakov, B.K.Kramorenko, A.A.Lyapidoyskiy, E.A.Shteynfeldt, Yu.Markov, I.I.Revzin, V.A.Artyomov, R.G.Piotrovskiy, Ye.V.Paducheva, I.A.Malchuk, R.M.Frumkina kabi oʻnlab shu sohaning mutaxassislari tilshunoslikning turli birliklarini statistik metodda oʻrganib tadqiq qilishdi. Oʻzbek tilshunosligida esa statistik metodlardan foydalanish 1930-yillarga toʻgʻri kelib, I.Kissen lugʻatlari bilan boshlangan. Ushbu lugʻatda hozirgi oʻzbek tilidagi soʻzlarning qoʻllanilishi (chastotasi) hisobga olingan holda tuzilgan. Masalan, 1941-yilda bir necha oy ichida I.A.Kissen rahbarligi va ishtiroki bilan 17000 soʻzdan iborat katta oʻzbekcha-ruscha lugʻat tuzilgan. Shuningdek, I.A.Kissen “Hozirgi oʻzbek adabiy tilida faol qoʻllanuvchi soʻzlar lugʻati”ni yaratgan. Buning uchun qirqta oʻzbek roman va qissalardan 101,044 ta soʻz shaklini tanlab, shundan 5,4 foizini atoqli otlar (5494 ta soʻz), qolgan 95,550 ta soʻz mustaqil va yordamchi soʻzlar boʻlib, 94,6 foizni tashkil etishini aniqlagan. Darvoqe, olim V.V.Reshetov oʻzbek tili matbuotining lugʻaviy tarkibini birinchilardan boʻlib tekshirgan. U oʻz nomzodlik dissertatsiyasida va shu asosda eʼlon qilingan maqolalarida oʻzbek tili matbuotidagi soʻzlarning kelib chiqishi (oʻzbekcha, arabcha, forsha, ruscha va boshqalar) va soʻz turkumlari boʻyicha chastotasini aniqlagan. Xuddi, shuningdek, gazeta materiallari boʻyicha statistik tekshirish ishini 1940-yilda A.K. Borovkov ham olib borgan edi [Ризаев С, 1972; 101-102]. Keyinchalik S.Rizayev “Minsk22” elektron-hisoblash mashinasida

“O‘zbek bolalar adabiyoti tilining chastotali lug‘ati”ni yaratgan. R.Qo‘ng‘urov va S.Karimovlarning “Zulfiya poeziyasi tilining lug‘ati”ida uning asarlarida keltirilgan nafaqat adabiy tildagi so‘zlar, balki umumxalq leksikasida keng foydalaniladigan *alqamoq, ardog‘, arslonto‘sh, baxmalto‘sh, betaxlit, betuman, bexalol, biyuda, boybeka, betuman, betus, burkimoq* kabi so‘zlar ham kiritilgan. Shuningdek, S.Muhamedovning gazeta matnlari asosida “O‘zbek tilining alfavitli-chastotali lug‘ati”ni alohida ta‘kidlash kerak. Olim alfavitli-chastotali lug‘atlar tuzish uchun gazeta mavzularidan tekslarning to‘g‘ri tanlanilishiga alohida e‘tibor bergan. Bu ish R.G.Piotovski rahbarligidagi umumittifoq “Statistika nutqi” gruppasi tomonidan ishlab chiqilgan prinsplarga asoslanib amalga oshirilgan. Olim 1972-1973 yillarga oid “Sovet O‘zbekistoni”, “Yosh Leninchi”, “O‘zbekiston madaniyati”, “Toshkent haqiqati” gazetalaridan oltita katta mavzudan iborat hamda ma‘lum bir proporsiya bo‘yicha hammasi bo‘lib, 200 mingdan ortiq so‘z qo‘llanishi olindi.

Ushbu matnlar o‘zbek tili gazeta matnlarining lingvostatistik modeli bo‘lib, shu model asosida maxsus programma yordamida “Minsk-32” elektron-hisoblash mashinasida alfavitli-chastotali lug‘at tuzildi.

Shu bilan birgalikda, Rizayev, N.Bo‘ronovlarning “Abdulla Qahhor “Sinchalak” povesti tilining chastotali lug‘ati” kabi lug‘atlari statistik metod asosida yaratilgan. D.O‘rinboyeva, U.Umrzoqov “O‘zbek xalq dostonlari tilining alfavitli chastotali lug‘ati”, S.Karimov, A.Qarshiyev, G.Isroilovalarning ” A.Qahhor asarlari tilining lug‘ati”, N.Xadjimusayevaning “Orzigul” dostoni tilining alfavitli, chastotali, chappa lug‘ati”, D.O‘rinboyevaning “O‘zbek xalq og‘zaki ijodi asarlarining qiyosiy chastotali lug‘ati”, S.Umirovaning “Usmon Azim asarlari tilining lug‘ati” kabi qator lug‘atlar ham statistik metod asosida yaratilgan lug‘atlardir.

Kompyuter lingvistikasi fani XX asrning 50-yillarida (1952-yilda) tilshunoslikning alohida yo‘nalishi sifatida yuzaga kelgandan so‘ng, statistik metodning o‘rni va ahamiyati yanada kengaydi. Aniqroq qilib aytganda, statistik metod kompyuter lingvistikasining asosini tashkil qildi. Kompyuter lingvistikasi doim statistik metodga tayanadi, u bilan birgalikda ish olib boradi. Bu metod til va til birliklari miqdoriy parametrlar asosida tahlil qiladi. Bunda statistika hodisalarning turli omillar ta‘siridagi o‘zgaruvchanlik xususiyatini – variatsiyasini e‘tiborga oladi. Shu jihatdan statistik ma‘lumotlar hodisalar rivojining dinamikasini aks ettira oladi. Statistikaning til hodisalariga tatbiq etilishi natijasida lingvostatistika sohasi shakllandi, unda til birliklari, lisoniy hodisalar, matnlar statistik tahlil etiladi. Matnning statistik tahlili - tilni ehtimollik belgilariga ko‘ra tahlil qilish, til hodisalarining statistik xarakteristikalarini asosida baholash demakdir. Til birligining gapda yoki matnda qo‘llanish chastotasi (sur‘ati) ehtimollik deyiladi. Til hodisalarini haqidagi haqiqatlar lingvostatistik tadqiqotlar yordamida aniqlanadi. Masalan, o‘zbek tili uchun qaysi bo‘g‘in strukturasi tipik ekanligini, gap modellaridan qaysi biri eng asosiy konstruktsiya ekanligini, bosh kelishikdagi otlarning, asosan, qaysi gap bo‘lagi vazifasida kelishini, tanlangan matnda so‘z turkumlarining qay darajada ishlatilishini aniqlash uchun lingvostatistik tahlillarga ehtiyoj seziladi. Til hodisalarini statistik metod asosida tahlil qilish uchun doim ma‘lum bir matn yoki matnlar to‘plami olinadi. Ular badiiy adabiyotdan, gazeta va jumallardan, publitsistik maqolalardan, ilmiy asarlardan, dialektologlarning jonli nutq orqali yozib olgan materiallaridan tanlanadi.

Matnlarni tahlil qiladigan ko‘plab kompyuter dasturlari yaratilgan. Shular jumlasidan, SAS (Statistical Analysis Software), Microsoft Excel, StatSoft (STATISTICA), Minitab, SPSS (IBM), STATA, XL STAT, Wizard Mac kabilarni keltirish mumkin. Ushbu dasturlarda

matnlarning har tomonlama statistik tahlil qiladi. SPSS (IBM). Kompleks statistik tahlil, rejalashtirish, biznes hisoboti uchun dasturiy ta'minot. SPSS Statistikasi dasturlashni talab qilmaydigan qulay interfeysga ega. U turli xil boshqaruv funktsiyalariga, statistik buyruqlar to'plamiga va hisobotlarni yaratish vositalariga ega. Barcha IBM SPSS mahsulotlari izchil tizimga integratsiya qilingan, shuning uchun siz bir dasturdan ikkinchisiga osongina o'tishingiz mumkin. O'rnatilgan modullar analitik imkoniyatlarni oshiradi. IBM SPSS barcha operatsion tizimlarda ishlaydi - Windows, MacOS, Linux, Android. Ko'pincha ushbu dastur tibbiyot, marketing, davlat va ta'lim muassasalarida qo'llaniladi. Dasturiy ta'minot to'plamida hujjatlar bilan ishlashni soddalashtiradigan metama'lumotlar lug'ati mavjud.

Xulosa qilib aytganda, statistik metod ma'lumotlar to'plash, tartibga solish, tavsiflash, tahlil qilish va talqin qilish va qaror qabul qilishda ulardan foydalanish usullarini ta'minlaydigan amaliy matematikaning bir qismi bo'lib, nafaqat tilshunoslik, balki kompyuter lingvistikasining ham tadqiq qilinishida muhim ahamiyatga ega. Eng asosiysi, statistik methodsiz kompyuter lingvistikasini tasavvur qilish imkonsizdir.

Foydalanilgan adabiyotlar:

1. A.A.Mirisaev. Statistika. O'quv-uslubiy qo'llanma. - T.: TDIU, 2021.
2. Prof. X.Shadiyev, I.Xabibullayev "Statistika" darslik. T.: Tafakkur bo'stoni, 2013.
3. Ризаев С. Тестни автоматик қайта ишлаш ҳақида. // Ўзбек тили ва адабиёти. 1972, №1. 101-102 -Б.
4. D.O'rinboyeva, U.Umrzoqov . O'zbek xalq dostonlari tilining alfavitli chastotali lug'ati.2006.
5. S.Muhamedov. O'zbek tilining alfavitli-chastotali lug'ati. -T.: 2021.
6. R.Qo'ng'urov S.Karmov. Zulfiya poeziyasi tilining lug'ati - T.: O'qituvchi, 1981.
7. A.Rahimov. Kompyuter lingvistika. - T.: Akademnashr, 2011.
8. Z.Xolmanova. Kompyuter lingvistikasi. - T.: Asian book house, 2020.