

HUQUQIY MATNLAR YARATISHDA STATISTIK TAHLILLAR VA ULARNING NATIJALARINI TAHLIL QILISH METODLARI.

<https://doi.org/10.5281/zenodo.19478522>

M. Raximov

Buxoro davlat universiteti o'qituvchisi

Annotatsiya

Mazkur tadqiqotda O'zbekiston va inglizabon davlatlar qonunchilik hujjatlarini lingvistik jihatdan solishtirish imkonini beruvchi Corpus of Legislation nomli elektron huquqiy matnlar korpusi yaratilgan. Korpusga O'zbekiston Respublikasi Konstitutsiyasi va Saylov kodeksi, shuningdek AQSh, Buyuk Britaniya va Avstraliyaning konstitutsiyalari hamda saylov qonunchiligi matnlari, ularning inglizcha va o'zbekcha versiyalari kiritilgan. Bu esa ikki tilning grammatik va uslubiy xususiyatlarini taqqoslashga zamin yaratadi.

Tadqiqot doirasida korpusdagi matnlar asosida so'z chastotasi, n-gram modellar (ya'ni ketma-ket kelgan so'zlar), mustaqil va yordamchi so'zlar tahlil qilindi. Natijalar shuni ko'rsatdiki, ingliz tilidagi huquqiy matnlarda eng ko'p uchraydigan birliklar – artikllar ("the", "a") va predloglar ("of", "in") bo'lsa, o'zbek tilida bu vazifani so'zga qo'shiladigan qo'shimchalar ("-gi", "-si", "-ning") bajaradi. Bundan tashqari, o'zbekcha matnlarda "va", "bilan", "uchun" kabi bog'lovchi so'zlar ham faol ishlatilgani kuzatildi.

Umuman olganda, Corpus of Legislation orqali qonun hujjatlari tilining o'ziga xos grammatik va uslubiy xususiyatlarini aniqlash imkoniyati paydo bo'ldi. Ushbu natijalar huquqiy matnlarni tarjima qilish, huquqiy tilni o'rgatish va tilshunoslikda korpus asosida tadqiqot olib borishda amaliy ahamiyatga ega.

Kalit so'zlar

huquqiy matn, statistik tahlil, Corpus of Legislation, metodlar.

Korpus yaratilishining ilk birinchi davri hisoblanmish elektronlashtirishgacha bo'lgan bosqich A. Kruden tomonidan XIII asrda yig'ilgan matnlar word indexing (so'zni indekslash) yoki concordancing (so'zlarning o'zaro munosabati va bog'lanishi) orqali tahlil qilingan [4. 132]. Keyinchalik korpus lingvistikasi taraqqiyoti davomida II va III – bosqichlarda elektronlashgan va gigamatnli korpuslarning yuzaga kelishi bilan statistik tahlillarning qay darajada o'sib borgani namoyon bo'ldi.

2008-yilda zamonaviy Amerika ingliz tili korpusi – COCA (The Corpus of Contemporary American English) tashkil etildi. Bu korpus yildan-yilga 20-30 million yangi soʻzlar bilan boyitilib, doimiy ravishda yangilanib boradi. Uning asosiy tarkibini 1800-yildan hozirgi kungacha chop etilgan yoki aytilgan yozma va ogʻzaki matnlar tashkil etadi. Hozirda COCA ingliz tilidagi eng yirik va nufuzli korpuslardan biri boʻlib, AQShda va butun dunyoda amerika ingliz tilini tahlil qilishda keng qoʻllaniladi.

2009-yilda esa yana bir yirik korpus – Google Books Ngram Viewer yaratilgan boʻlib, u 1500-yildan 2008-yilgacha chop etilgan elektron kitoblar matnlarini oʻz ichiga oladi. Hozirgi kunda ushbu korpusning hajmi 200 milliarddan ortiq soʻzlardan iborat. Bu korpus ingliz tilidan tashqari nemis, fransuz, ispan, rus, italyan va yahudiy tillaridagi matnlarni ham oʻz ichiga oladi.

Ingliz tilining geografik hududlarga xos xususiyatlarini oʻrganish maqsadida 2013-yilda Global Web-based English (GloWbE) korpusi tashkil qilindi. Bu korpusda turli hududlardagi ingliz tili veb sahifalari matnlari jamlangan boʻlib, u 1,9 milliard soʻzdan iborat.[5]

Oʻzbek tilshunosligida korpus lingvistikasi hali toʻliq oʻrganilmagan boʻlsa-da, bir nechta olimlar bu borada izlanishlar olib borgan. Masalan, J. Jumaboyeva BNC korpusidagi semantik gradatsiyalarni aniqlash ustida ishlagan, G. Ergasheva genderga oid konseptual metaforalarni tahlil qilgan, G. Sobirova esa bu korpusdan til oʻqitish jarayonida foydalanish imkoniyatlarini koʻrib chiqqan. Shuningdek, Sh. Hamroyeva mualliflik korpusini yaratish masalalari bilan shugʻullangan, N.B. Ataboyev esa COCA misolida ingliz tilining diaxronik korpuslarining funksional jihatlarini oʻrgangan. [1.5]

Huquqiy terminlarni tanlashda ularning barcha ijtimoiy qatlam vakillariga tushunarli boʻlishi va boshqa matn turlaridan farqlanuvchi uslubiy xususiyatlarga ega boʻlishi zarur. Bunda soʻz yasalishi, feʼlning zamon va mayl shakllari, sintaktik tuzilma kabi grammatik vositalarning toʻgʻri qoʻllanilishi huquqiy matnlarning asosiy tamoyillarini belgilaydi. Grammatik vositalarni aniq qoʻllash bilan birga, huquqiy terminlarni oʻrinli va xatosiz ishlata olish ham matn mualliflarining masʼuliyatli vazifasidir.

H.A. Karimov bu borada shunday fikr bildiradi: “Agar matnda notoʻgʻri yoki joysiz ishlatilgan soʻz yoki ibora boʻlsa, bu fikr bilan uning ifodasi oʻrtasida tafovut paydo boʻlishiga olib keladi va bu ayrim hollarda jiddiy oqibatlarga sabab boʻlishi mumkin, ayniqsa qonunlar bilan bogʻliq hujjatlarda.” [4.123]

Bugungi kunda huquqiy matnlar tilini ifodalashda “qonun tili” tushunchasi shakllangan. Bu borada ham koʻplab tadqiqotlar olib borilgan. Masalan, fransuz huquqshunosi F. Jenya fikricha, qonunchilik faoliyatida tilning maxsus vositalari va

uslublari mavjud bo'lib, ular qonun chiqaruvchi fikrini aniq ifodalash va qonunchilik uslubini adabiy nutqdan ajratish imkonini beradi.

Huquqiy matnlarning o'ziga xos yana bir jihati – bu ularning funksional o'zgarmasligidir, ya'ni ulardagi terminlar faqat bitta ma'noda, aniqlik bilan qo'llanadi.

Mazkur tadqiqot doirasida yaratilgan Corpus of Legislation (<https://corpuslegislation.uz/>) [6.] nomli huquqiy matnlar korpusi O'zbekiston hamda inglizzabon davlatlarning qonunchilik hujjatlarini lingvistik jihatdan tahlil qilish imkoniyatini beradi. Korpus tarkibiga O'zbekiston Respublikasi Konstitutsiyasi, Saylov kodeksi, shuningdek, AQSh, Buyuk Britaniya va Avstraliya kabi davlatlarning konstitutsiyalari hamda saylov qonunchilik hujjatlari kiritilgan.

Til me'yorlarini taqqoslash imkoniyatlarini kengaytirish maqsadida, ushbu matnlarning o'zbekcha va inglizcha variantlari ham korpusga qo'shilgan. (2.3.1 – jadval)

2.3.1 – jadval. *Corpus of legislation korpusi lisoniy ba'zasi statistikasi*

Korpus/corpus	Matn turi/Type of the text	Jami so'zlar yig'indisi /Total count of words	Jami yaxlit so'zlar yig'indisi/Total count of unique words	Matn tili /Language of the text
Corpus of legislation	USA constitution	8544	1779	American English
	USA electoral law	71047	6590	American English
	UK Constitution	8220	2547	British English
	Great Britain electoral law	60607	5087	British English
	Australia Constitution	11936	1979	Australian English
	Australia electoral law	10847	3295	Australian English
	Uzbekistan constitution	15705	2148	British English
	Uzbekistan election code	19222	2226	British English
	O'zbekiston Respublikasi konstitutsiyasi	13174	3323	O'zbek tili
	O'zbekiston	20045	3337	O'zbek

Respublikasi saylov kodeksi			k tili
Jami/total:	239343	32581	

Jami bo'lib tadqiqot davomida yaratilgan lisoniy korpus ba'zasida 239343 (ikki yuz o'ttiz to'qqiz ming uch yuz qirq uchta) so'z o'rin oldi. Mazkur ba'zadagi so'zlarning 32581 (o'ttiz ikki ming besh yuz sakson bir)tasi yaxlit (unique) so'zlarni tashkil etadi. Shulardan 33219 (o'ttiz uch ming ikki yuz o'n to'qqiz)ta so'z hamda 6660 (olti ming olti yuz oltmish)ta yaxlit so'zlardan iborat bo'lgan matnlar shaklida kiritildi. Korpus dasturida kiritilgan matnlarning 2-gram va 3-gram modeli leksemalarning ham statistikasi tadqiq qilindi.

Yaratilgan qonunchilik matni korpusida (Corpus of legislation) so'z turkumlarini tahlil qilish maqsadida matnlar ba'zasidagi so'zlarning ishlatilish chastotasidan kelib chiqqan holda jami dastlabki 2000 ta birlik tahlil qilindi. Ushbu tahlil miqdori har ikkala tilda kiritilgan matn ba'zasida so'z turkumlarining ishlatilish holati va chastotasini o'rganish uchun yetarlicha xulosa chiqarishga imkon yaratadi. Bundan tashqari har ikkala tilda kiritilgan matnlardagi eng kam ishlatilgan birliklarning ham pastki 50 talik o'rinlari ham tahlil qilindi. 3.1.2-jadval

3.1.2 – jadval. *Corpus of legislation*⁷⁰ matn korpusi statistik tahlili

#	Word	Frequency	n-gram	Source	Parts of speech
1	The	4079	1	Great Britain Electoral Law	article
2	The	3796	1	USA Electoral Law	article
3	Of	3487	1	USA Electoral Law	preposition
4	A	2745	1	Great Britain Electoral Law	article
5	Of	2717	1	Great Britain Electoral Law	preposition
6	A	2450	1	USA Electoral Law	article
7	The	2229	1	Uzbekistan Electoral Law	article
8	Or	2115	1	USA Electoral Law	conjunction
9	The	2062	1	Uzbekistan Constitution	article
10	In	2054	1	Great Britain Electoral Law	preposition

Tahlil natijalari shuni ko'rsatadiki, korpus interfeysida eng yuqori o'rinlarda ingliz tili grammatikasiga xos bo'lgan artikl, predlog va bog'lovchilarning ustunligi kuzatiladi. Jumladan, "the" artikli hamda "of" predlogining qo'llanilish chastotasi inglizabon davlatlarning qonun hujjatlari va O'zbekiston qonunchilik hujjatlarining ingliz tilidagi tarjimalarida ayniqsa yuqori bo'lib, bu ularning ingliz tilidagi gap tuzilmasida muhim biriktiruvchi vosita sifatida xizmat qilishini anglatadi.

Buning asosiy sababi shundaki, ingliz tili grammatikasi predlog va artikllar orqali aniqlik hamda tafsilotlarni aniq ifodalashga intiladi. Aksincha, o'zbek tilida bu kabi ma'no ifodasi ko'proq so'zga bevosita qo'shiladigan qo'shimchalar orqali amalga oshiriladi. Natijada, statistik tahlillarda qaror yoki hujjat nomidan keyin ko'pincha qo'shimchalar qo'llanilgani sababli, masalan "-gi", "-si", "-ning" kabi, O'zbekiston Respublikasi Konstitutsiyasi matnida bu birliklar 232-, 1079- va 1168-o'rinlarda uchraydi. 3.1.3- jadval

3.1.3-jadval. (Corpus of legislation) O'zbekiston Respublikasi konstitutsiyasi va saylov kodeksi matni statistik tahlili

#	Word	Frequency	n-gram	Source
232	-gi	144	1	O'zbekiston Respublikasi titutsiyasi
1079	-ning	29	1	O'zbekiston Respublikasi Saylov ksi
1168	-gi	27	1	O'zbekiston Respublikasi Saylov ksi

O'zbek tili morfologiyasining yordamchi so'z turkumlari bo'lmish *bog'lovchi*, *ko'makchi*, *yuklamaga* xos so'zlar korpus matn so'z ba'zasi statistikasining pastki pog'onalaridan o'rin olgan. 3.1.4 –jadval

3.1.4 –jadval. Corpus of legislation O'zbekiston Respublikasi konstitutsiyasi va saylov kodeksi matni statistik tahlili

Statik	So'zla	Chastot	n-am	Manbasi	So'z kumi
48	Va	550	1	O'zbekiston Respublikasi onstitutsiyasi	bog'lovch
57	Va	486	1	O'zbekiston Respublikasi ylov kodeksi	bog'lovch
232	Bilan	144	1	O'zbekiston Respublikasi ylov kodeksi	bog'lovch
300	Uchun	114	1	O'zbekiston Respublikasi ylov kodeksi	bog'lovch

Tahlillar natijasi shuni ko'rsatadiki, o'zbek tilida "va" bog'lovchisi korpus ba'zasida 48 va 50-o'rinlarda joylashgan bo'lib, O'zbekiston Respublikasi konstitutsiyasi matn ba'zasida jami 550 marta va O'zbekiston Respublikasi saylov kodeksi matn ba'zasida 486 marotaba qo'llanilgan. Navbatdagi bog'lovchilar "bilan" hamda "uchun" matn tahlilida 232 va 300 - o'rinlardan joy olgan bo'lib, "bilan" 144 marotaba, "uchun" 114 marotaba ishlatilgan.

Korpusda mustaqil so'z turkumlari ro'yxatini uchinchi o'ntalik statistikasida ko'rish mumkin.

3.1.5 – jadval. *Corpus of legislation morfologik statistic tahlili*

Statistic order	Word	Frequency	n-gram	Source	Parts of speech
21	Any	1102	1	USA Electoral Law	Pronoun
22	By	988	1	USA Electoral Law	preposition
23	Such	949	1	USA Electoral Law	adverb
24	Of	944	1	Australia Constitution	preposition
25	B	876	1	USA Electoral Law	(number sign)
26	Saylov	849	1	O'zbekiston Respublikasi Saylov kodeksi	ot
27	Is	830	1	Great Britain Electoral Law	auxiliary verb
28	And	811	1	Great Britain Electoral Law	Conjunction
29	B	787	1	Great Britain Electoral Law	(number sign)
30	Section	782	1	Great Britain Electoral Law	noun

Yuqoridagi jadvaldan ko'rinib turibdiki, ingliz tili matn ba'zasidan dastlabki o'rinda olmosh so'z turkumi joy olgan bo'lib, mazkur holatning izohi sifatida ingliz tili matn tahlilida olmosh so'z turkumi qo'llanilishining ahamiyati nisbatan yuqori o'rinda turishidir. Statistikaning 21 -o'rnida "*any*" olmoshi joylashgan bo'lib, AQSH saylov qonunchiligi matn ba'zasida 1102 marta ishlatilgan. Keyingi o'rinlarda ingliz tilidagi ravish so'z turkumiga xos bo'lgan "*such*" so'zi 23-qatorda joylashgan. Ushbu so'z manbasi AQSH saylov qonunchiligi matnidan o'rin olgan bo'lib, 949 marotaba qo'llanilgan. O'zbek tilida esa uchinchi o'ntalikning 26-o'rinini ot so'z turkumiga xos "*saylov*" so'zi egallagan bo'lib, O'zbekiston Respublikasi saylov kodeksida 849 marotaba qo'llanilgan. Ushbu o'rinda har ikkala til areali misolida saylov qonunchiligida yuqorida ko'rsatib o'tilgan so'zlarning keng qo'lanishini quyidagicha izohlash mumkin: 1) Matn hajmi va strukturasi nisbatan kattaligi; 2) Qonunchilik hujjatlarining ishlab chiqish usuli va strukturasi; 3) Kontekstual ma'nodan kelib chiqqan holda ushbu so'zlarni ishlatishga talabning ko'pligi va hokazo.

Xulosa qilib aytganda, yaratilgan qonun hujjatlari matn korpusining lisoniy ba'zasini o'rganish borasida qonun tili uslubi va grammatikasi muhim ahamiyat kasb etishi muhim omil sifatida tahlil qilinishi lozim. Shuning uchun *Corpus of legislation* mualliflik matn korpusida dastlabki statistik tahlillar amalga oshirildi. Tahlillar eng ko'p va eng kam chastotaga ega bo'lgan n-gram hamda bigramlarni statistik o'rganish asnosida tadqiq qilind.

N-gramlarni statistik tahlil qilinganida, ingliz tilidagi matnlarda (ayniqsa inglizxabon mamlakatlar qonun hujjatlari matnida) yordamchi so'z turkumi hisoblanmish predlog, artikl, mustaqil so'z turkumidan ot so'z turkumlari yuqori pog'onalarni egallagan bo'lsa, o'zbek tili matnlarida ot va sifat so'z turkumlarining ustunligi aniqlandi. Bu holat esa tillarning lingvistik -tipologik klassifikatsiyasidan kelib chiqadi.

ADABIYOTLAR RO'YXATI:

1. Ataboyev N. B. Ingliz tilining diaxronik korpuslarining funksional xususiyatlari (COCA misolida) /monografiya. Muhammad poligraf. 2023. – b. 5.
2. Djumabayeva J. "Lexical graduonymy in Uzbek and English languages". Monograph. Tashkent. 2014.
3. Керимов. Д. А. Законотетельная техника: науч.-метод. и учеб. пособие – Москва: НОРМАИНФРА-М., 1998. – ст. 123.
4. Солнышкина М.И., Гатиятулина Г.М. История развития корпусной лингвистики (на примере англоязычных корпусов)//Вестник Томского государственного университета № 63, - Томск, 2020. – С. 132. Доступно в: <https://cyberleninka.ru/article/n/istoriya-razvitiya-korpusnoy-lingvistiki-na-primere-angloyazychnyh-korpusov/viewer>
5. GloWbE. URL: <https://corpus.byu.edu/glowbe/>
6. Corpus of legislation <https://corpuslegislation.uz/>