

Эдуард ФОМИН

СТАТИСТИЧЕСКАЯ ХАРАКТЕРИСТИКА ЧУВАШСКОГО ЯЗЫКА  
КОНЦА XIX – НАЧАЛА XX ВЕКОВ

Мақалада XIX ғасырдың соңы мен XX ғасырдың басындағы чуваш мәтіндерін статистикалық талдаудың нәтижелері берілген.

Лингвостатистика – чуваш тілтану ғылымындағы жүйелі түрде зерттелмей отырған сала. Сондықтан оның кең ауқымда өркендеуіне байланысты чуваш материалдары жалпы лингвистикалық квантитативті ізденістерде жеткілікті дәрежеде қарастырылмай отыр.

Бұл жұмыс осы бос кеңістікті толтыруға арналған және чуваш лингвистикасындағы алғашқы зерттеулердің бірі.

Çalışmada XIX. yy. sonu ve XX. yy. başlangıcı Çuvaş metinlerinin istatistik inceleme sonuçları kapsanmıştır.

Dilbilim istatistiği, Çuvaş dilbiliminde geleneksel olmayan dallardandır.

Gelişmemişliği nedeniyle Çuvaş dilbilim malzemeleri genel dilbilim sayısal araştırma çalışmalarında yer almaz. Bu çalışma sözü edilen boşluğu doldurma amacını taşır.

Лингвостатистика относится к нетрадиционным в чувашском языкознании дисциплинам. По причине ее неразвитости чувашский материал в общелингвистических квантитативных разысканиях чаще всего не привлекается. Настоящая работа отчасти призвана восполнить этот пробел.

Материалом изучения послужили научные и богослужebные тексты периода зарождения и становления современной чувашской письменности и литературного языка (1872 – 1917). Здесь в табличной форме приводятся некоторые результаты статистических исследований.

Представленная работа – один из первых подобных опытов в чувашской лингвистике. Возможны и необходимы исследования более широкого круга языковых признаков с применением специальных машинных программ<sup>1</sup>, охватывающие также иные стилевые разновидности и принадлежащие различным временным срезам. Есть основания считать, что в ближайшее время лингвостатистика займет достойное место в системе чувашеведческих наук: при изучении чувашского языка ученые все чаще обращаются к количественным методам.

Качественному анализу результатов исследования посвящена монография «Становление научного стиля чувашского литературного языка», в настоящее время готовящаяся к изданию.

<sup>1</sup> Современное чувашеведение располагает ограниченными возможностями машинного анализа чувашского текста, и каждое такого рода исследование представляет собой трудоемкое занятие.

Таблица 1

Количественные показатели лексических единиц

Стили	Количество словоупотреблений	Ср. длина слова в буквах	Ср. длина слова в слогах
Научный	89754	5,47	2,47
Богослужебный	28952	5,38	2,36
Итого	118706	5,45	2,42

Таблица 2

Статистические параметры последовательности закона сингармонизма

Стили	Проявление закона сингармонизма					
	слова с заднеязычной огласовкой		слова с переднеязычной огласовкой		слова с нарушением сингармонизма	
	Всего	доля	всего	доля	всего	доля
Научный	45214	0,50	29921	0,33	14619	0,16
Богослужебный	13897	0,48	9554	0,33	5501	0,19
Итого	59086	0,51	36421	0,31	20199	0,17

Таблица 3

Ранговое распределение частот словообразовательных аффиксов (на материале научной литературы)

Ранг	Аффикс	Частота	Доля	Количество слов с данным аффиксом	Ср. повторяемость слов	Примеры
1	-ла	2522	0,24433	93	27,12	пусла, сарла, ёсле
2	-ан	1781	0,17254	90	19,79	пусан, хёсен, юсан
3	-(т)гар	1021	0,09891	126	8,10	тулгар, сёмёрттер
4	-лă	822	0,07964	42	19,57	вăйлă, тирпейлĕ
5	-(а)т	705	0,06828	60	11,75	сурат, тăрат, ёрчет
6	-лан	531	0,05143	126	4021	хаваслан, пурлен
7	-(ă)л	395	0,03826	23	17,17	татал, сёмсел
8	-лăх	391	0,03787	33	11,85	патшалăх, ирĕклĕх
9	-ăм	308	0,02983	15	20,53	ярăм, пурлĕм
10	-ăç	159	0,01540	17	9,35	анăç, тухăçлă
11	-ал	158	0,01530	10	15,80	сывал, кĕскел
12	-лат	132	0,01278	20	6,60	начарлат, турлет
13	-ар	131	0,01269	5	26,20	пусар, пĕтер
14	-ă	118	0,01143	15	7,87	кулă, сутă, чикĕ
15	-ăш	104	0,01007	20	5,20	тупăш, сивчĕш
16	-(ă)ç	94	0,00910	8	11,75	явăç, калăç, сурăç
17	-ке	85	0,00823	1	85,00	Йĕрке
18	-у	75	0,00726	13	5,77	каю, пуху, кĕтĕ
19	-аш	74	0,00717	2	24,67	Малаш, тулаш
20	-çă	66	0,00639	8	8,25	Тарçă, севĕс, эмелçĕ

Ра нг	Аффикс	Частота	Доля	Количество слов с данным аффиксом	Ср. повторяе- мость слов	Примеры
21	-ака	62	0,00600	5	12,40	сапака, саврака
22	-ушă	60	0,00581	1	60,00	хăрушă
23	-лак	59	0,00571	3	19,67	тайлак, шăналăк, çителĕк(лĕ)
24	-ав	56	0,00542	8	7,00	мухтав, пăлхав
25	-ма	48	0,00465	3	16,00	пусма, сурма сĕме
26	-чăк	43	0,00416	16	2,69	хуçланчăк, хырчăк, шанчăк
27	-ал	40	0,00387	12	3,33	сайрал, тасал, сĕнел
28	-шка	33	0,00320	5	6,60	какрашка, лупашка, урлашка
29	-ма	19	0,00184	1	19,00	сырма
30	-еç	15	0,00145	1	15,00	сĕмеç
31	-а	14	0,00136	3	4,67	мăя, суя, турта
31	-кĕ	14	0,00136	1	14,00	еçкĕ
31	-не	14	0,00136	1	14,00	парне
32	-тат	13	0,00126	8	1,62	йăлтăртат, пакăртат, шанкăртат
33	-аç	12	0,00116	1	12,00	хутаç
33	-т	12	0,00116	1	12,00	инçет
34	-ал	10	0,00097	2	5,00	хыçал, тĕпел
34	-х	10	0,00097	2	5,00	пăчах, йуçех
34	-чĕ	10	0,00097	1	10,00	пĕрчĕ
35	-ăр	9	0,00087	3	3,00	палăр, хуйхăр
35	-лаш	9	0,00087	3	3,00	тавлаш, пĕрлеш
36	-ала	7	0,00068	2	3,50	сапала, явала
36	-тăк	7	0,00068	3	3,50	карантăк, лаптăк, мăшкăлтăк
37	-к	6	0,00058	1	6,00	чăмлак
38	-ăнкă	5	0,00048	2	2,50	тытăнкă, йуçĕнкĕ
38	-ăр	5	0,00048	1	5,00	типшĕр
38	-ам	5	0,00048	2	2,50	лапам, тўрем
38	-кка	5	0,00048	3	1,67	муранкка, чăмăркка, чăпăркка
39	-кă	4	0,00039	1	4,00	лăпкă
39	-кăч	4	0,00039	1	4,00	ĕмкĕч
39	-лак	4	0,00039	1	4,00	тăварлак
39	-мăр	4	0,00039	1	4,00	пусмăр
39	-ш	4	0,00039	1	4,00	тиркеш
40	-ăш	3	0,00029	1	3,00	чикĕш
40	-çă	3	0,00029	1	3,00	пĕчĕкçĕ
40	-чăт	3	0,00029	1	3,00	лапчăт
41	-амас	2	0,00019	1	2,00	çарамас
41	-ăм	2	0,00019	1	2,00	сывлăм
41	-ка	2	0,00019	1	2,00	каска

42	-ал	1	0,00010	1	1,00	сиял (лит. сиел)
42	-ак	1	0,00010	1	1,00	сётек
42	-каш	1	0,00010	1	1,00	йӓлкаш
42	-каш	1	0,00010	1	1,00	турткаш
42	-лес	1	0,00010	1	1,00	силлес
42	-мак	1	0,00010	1	1,00	тӓсмак
42	-мӓш	1	0,00010	1	1,00	пусламӓш
42	-та	1	0,00010	1	1,00	пӓчӓрта
42	-так	1	0,00010	1	1,00	лаптак
42	-ча	1	0,00010	1	1,00	хупӓлча
42	-чӓн	1	0,00010	1	1,00	лапчӓн
42	-чен	1	0,00010	1	1,00	ёсчен
42	-хе	1	0,00010	1	1,00	силхе
Итого		10322	1,00000	846	12,20	

Таблица 4

Ранговое распределение частот словообразовательных аффиксов  
(на материале богослужебной литературы)

Ранг	Аффикс	Частота	Доля	Количество слов с данным аффиксом	Ср. повторяемость слов	Примеры
1	-ат	569	0,15871	29	19,62	шурат, вӓрент
2	-ла	375	0,10463	33	11,36	айӓпла, хиспеле
3	-(т)тар	318	0,08873	49	6,49	вырттар, ирттер
4	-лан	258	0,07197	42	6,24	асаплан, тумлан
5	-лӓх	247	0,06892	39	6,33	канӓслӓх, чеелӓх
6	-ӓн	239	0,06668	40	5,97	парӓн, пӓлӓн
7	-лӓ	225	0,06278	22	10,23	хӓватлӓ, ирӓклӓ
8	-ӓҫ	168	0,04687	11	15,27	савӓнӓҫ, шанӓҫ
9	-ар	153	0,04269	12	12,75	хӓтар, пӓтер
10	-ал	136	0,03795	9	15,11	сывал, сӓхал
11	-ӓл	124	0,03460	13	9,54	ватӓл, сӓтӓл
12	-ҫӓ	119	0,03320	13	8,42	тарҫӓ, кӓнекеҫӓ
13	-у	88	0,02455	31	2,84	каҫару, хӓрхенӓ
14	-ав	82	0,02288	4	20,50	мухтав, ултав
15	-кӓн	64	0,01786	3	21,33	сӓпкӓн, тыткӓн
16	-ӓм	61	0,01702	10	6,10	парӓм, пӓлӓм
17	-аҫ	56	0,01562	2	28,00	вырнаҫ, калаҫ
18	-ӓ	45	0,01256	10	4,50	выҫӓ, чикӓ
19	-ӓк	35	0,00977	6	5,83	сыпӓк, сӓрӓк
20	-лат	30	0,00837	8	3,75	мӓнтӓрлат, тӓрлет
21	-н	27	0,00772	6	4,50	ван, вӓрен
22	-а	23	0,00657	3	7,67	ҫӓра, виҫе
22	-не	23	0,00657	1	23,00	парне
23	-аш	15	0,00429	2	7,50	малаш, тулаш
24	-аҫ	13	0,00372	1	13,00	хутаҫ
25	-аш	8	0,00229	1	8,00	тавраш
26	-ӓр	7	0,00200	1	7,00	хуйхӓр
27	-ат	6	0,00171	1	6,00	Анат

27	-әрха	6	0,00171	1	6,00	Асәрха
27	-ка	6	0,00171	1	6,00	авка
28	-аҫ	5	0,00143	2	2,50	ҫапәҫ, пётёҫ
28	-тат	5	0,00143	3	1,67	йәлтәртат
28	-әх	5	0,00143	1	5,00	выҫәх
29	-ал	4	0,00114	2	2,00	хыҫал, ҫиел
29	-ка	4	0,00114	1	4,00	шакка
30	-амас	3	0,00086	1	3,00	ҫарамас
30	-кә	3	0,00086	2	1,50	ләпкә, черченкә
30	-ла	3	0,00086	2	1,50	кәнтәрла, ҫәрле
31	-әш	2	0,00057	2	1,00	хутәш, пелеш
31	-әш	2	0,00057	1	2,00	ҫивчеш
31	-кә	2	0,00057	1	2,00	ёскә
31	-кёҫ	2	0,00057	1	2,00	тётёркёҫ
31	-ам	2	0,00057	1	2,00	ҫөклем
31	-чәк	2	0,00057	2	1,00	шанчәк, ыйкалчәк
32	-ам	1	0,00029	1	1,00	түрем
32	-ёк	1	0,00029	1	1,00	чепёк
32	-кал	1	0,00029	1	1,00	таваткал
32	-каш	1	0,00029	1	1,00	йәлкәш
Итого		3574	1,00000	430	8,31	

Таблица 5

Распределение лексики по классам слов

Классы слов	Частотность, в %			
	общая	в богослужбном стиле	в публицистическом стиле	в научном стиле
Существительное	32,47	33,93	31,87	30,33
Прилагательное	7,78	7,13	6,87	11,67
Числительное	2,44	1,33	2,93	5,00
Местоимение	13,36	17,33	11,20	8,83
Глагол	29,33	28,53	30,20	29,17
в том числе				
инфинитив	2,36	2,13	3,00	1,33
причастие	11,78	13,73	10,13	11,00
деепричастие	7,97	7,80	7,87	8,87
Наречие	4,78	2,87	6,00	6,50
Послелог	2,64	3,47	1,80	2,17
Союз	2,30	1,07	3,67	2,50
Частица	4,64	3,93	5,67	3,83
Междометие	0,17	0,40	—	—
Звукоподражание	—	—	—	—

Таблица 6

Распределение частот словоизменительных аффиксов

Ранг	Научный текст			Богослужебный текст		
	Аффикс	Значение	Частота абс./ относит.	Аффикс	Значение	Частота абс./ относит.
1	-а	дательный падеж	162 (0,162)	-а	дательный падеж	175 (0,175)
2	-ть	3 л. ед. ч. настоящего времени	140 (0,140)	-нă	причастие прошедшего времени	160 (0,160)
3	-нă	причастие прошедшего времени	63 (0,063)	-са	деепричастие	89 (0,089)
4	-са	деепричастие	55 (0,055)	-ăн	родительный падеж	86 (0,086)
5	-ççĕ	3 л. мн. ч. настоящего времени	52 (0,052)	-ĕ	3 л. категории принадлежности	65 (0,065)
6	-сем	мн. ч.	48 (0,048)	-сем	мн. ч.	63 (0,063)
7	-па	творительный падеж	44 (0,044)	-ран	исходный падеж	54 (0,054)
8	-ăн	родительный падеж	43 (0,043)	-ра	местный падеж	49 (0,049)
8				-и	форма выделения	49 (0,049)
9	-сан	деепричастие	42 (0,042)	-па	творительный падеж	29 (0,029)
10	-и	форма выделения	40 (0,040)	-сан	деепричастие	25 (0,025)
11	-ра	местный падеж	34 (0,034)	-ма	инфинитив	16 (0,016)
11	-ран	исходный падеж	34 (0,034)			
12	-ас	причастие будущего времени	31 (0,031)	-акан	причастие настоящего времени	14 (0,014)
12	-ĕ	3 л. категории принадлежности	31 (0,031)			
13	-ма	инфинитив	29 (0,029)	-ть	3 л. ед. ч. настоящего времени	12 (0,012)
14	-ах	усилительная форма	26 (0,026)	-ах	усилительная форма	10 (0,010)
14				-ăр	2 л. ед. ч. настоящего времени	10 (0,010)
14				-лă	форма наличия	10 (0,010)
15	-ма	отрицательная форма	23 (0,023)	-тă	прошедшее время	9 (0,009)
16	-и	3 л. ед. ч. категории принадлежности	17 (0,017)	-ĕ	будущее время	8 (0,008)
17	-лă	форма наличия	11 (0,011)	-ăм	1 л. ед. ч. категории принадлежности	7 (0,007)

Ранг	Научный текст			Богослужебный текст		
	Аффикс	Значение	Частота абс./ относит.	Аффикс	Значение	Частота абс./ относит.
17	-рах	сравнительная форма	10 (0,010)	-џ	3 л. мн. ч. глагола	7 (0,007)
17				-шан	причинный падеж	7 (0,007)
18	-тарах	сравнительная форма	10 (0,010)	-ас	причастие будущего времени	6 (0,006)
18				-ма	отрицательная форма	6 (0,006)
19	-акан	причастие настоящего времени	8 (0,008)	-ап	1 л. ед. ч. настоящего времени	4 (0,004)
19				-сар	лишительный падеж	4 (0,004)
20	-меш	порядковая форма числительных	6 (0,006)	-ан	2 л. ед. ч. будущего времени	3 (0,003)
20	-р	2 л. мн. ч. повелительного наклонения	6 (0,006)	-и	причастие возможности	3 (0,003)
20				-сассан	деепричастие	3 (0,003)
20				-у	2 л. ед. ч. категории принадлежности	3 (0,003)
20				-че	3 л. прошедшего времени	3 (0,003)
20				-ччен	предельная форма	3 (0,003)
21	-а	деепричастие	5 (0,005)	-алла	направительная форма	2 (0,002)
21	-хи	выделительная форма наречий	5 (0,005)	-малла	причастие долженствования	2 (0,002)
22	-ар	2 л. мн. ч. глагола	4 (0,004)	-ай	аспект возможности	1 (0,001)
22	-сар	лишительный падеж	4 (0,004)	-ар	1 л. мн. ч. глагола	1 (0,001)
22				-ан	2 л. ед. ч. глагола	1 (0,001)
22				-џџе	3 л. мн. ч. настоящего времени	1 (0,001)
23	-малла	причастие долженствования	3 (0,003)			
24	-ла	форма уподобления	2 (0,002)			
24	-ан	реликтовый инструментальный аффикс	2 (0,002)			
24	-лах	форма достаточности	2 (0,002)			

Ранг	Научный текст			Богослужебный текст		
	Аффикс	Значение	Частота абс./ относит.	Аффикс	Значение	Частота абс./ относит.
24	-ччен	предельная форма	2 (0,002)			
24	-ай	аспект возможности	2 (0,002)			
25	-ччӑр	3 л. повелительного наклонения	1 (0,001)			
25	-сем	деепричастие	1 (0,001)			
25	-н	2 л. ед. ч. настоящего времени	1 (0,001)			
25	-шар	разделительная форма числительных	1 (0,001)			

Таблица 7

Соотношение простых предложений и сложных синтаксических единиц

Стили	Объем, в предл.	Простые предложения		Сложные синтаксические единицы	
		всего	доля	всего	доля
Научный	700	253	0,36	447	0,64
Богослужебный	600	117	0,19	483	0,81
Итого	1300	370	0,28	930	0,72

Таблица 8

Количественные параметры сложных синтаксических единиц

Стили	Сложные предл.		Сложноподчи- ненные предл.		Сложносочи- ненные предл.		Бессоюзные предл.		Предл. с прямой речью		Однородные предл.	
	всего	доля	всего	доля	всего	доля	всего	доля	всего	доля	всего	доля
Научный	374	0,53	120	0,17	55	0,08	199	0,28	42	0,06	31	0,04
Богослужебный	307	0,51	120	0,20	24	0,04	193	0,32	138	0,23	8	0,01
Итого	681	0,32	240	0,18	79	0,06	392	0,30	180	0,14	39	0,03

Таблица 9

Средние величины синтаксических единиц

Стили	Ср. длина предл. в словах	Ср. длина простого предл. в словах	Ср. длина многопредл. конструкций в словах	Ср. длина предл. в предл.
Научный	13,69	8,85	18,53	2,26
Богослужебный	15,02	9,95	20,10	3,09
Итого	14,35	9,40	19,31	2,67



Таблица 10

Распределение предложений на основе словесной длины

Длина предл.	Доля в научном тексте	Длина предл	Доля в богослужебном тексте	Длина предл.	Доля в научном тексте	Длина предл.	Доля в богослужебном тексте
11	0,066	9	0,060	25	0,013	26	0,013
13	0,054	11	0,060	30	0,011	28	0,012
6	0,051	12	0,055	27	0,009	31	0,012
7	0,048	13	0,047	28	0,009	35	0,012
9	0,048	16	0,047	29	0,009	3	0,010
12	0,048	17	0,047	40	0,009	5	0,010
14	0,048	14	0,043	35	0,008	33	0,010
20	0,044	15	0,043	37	0,008	29	0,008
8	0,043	18	0,040	31	0,006	36	0,008
10	0,043	8	0,037	34	0,006	42	0,008
17	0,040	10	0,037	38	0,006	38	0,007
4	0,038	21	0,037	39	0,006	39	0,007
15	0,037	50	0,037	48	0,006	1	0,005
16	0,034	19	0,032	50	0,006	40	0,005
19	0,026	22	0,028	32	0,005	47	0,005
22	0,026	7	0,027	36	0,005	41	0,003
3	0,023	6	0,025	1	0,003	43	0,003
18	0,021	20	0,023	33	0,003	44	0,003
21	0,015	24	0,022	43	0,003	49	0,003
25	0,015	23	0,018	46	0,003	2	0,002
23	0,014	32	0,018	44	0,001	48	0,002
26	0,014	4	0,017	45	0,001	45	0,000
2	0,012	27	0,015	47	0,001	46	0,000
24	0,012	30	0,015	49	0,001		

Таблица 11

Распределение предложений по цели высказывания

Стили	Общее количество предл.	Повествовательные		Восклицательные		Вопросительные	
		количество предл.	доля, в %	количество во предл.	доля, в %	количество предл.	доля, в %
Научный	6028	5884	97,61	51	0,85	93	1,54
Богослужебный	2058	1739	84,50	73	3,55	246	11,95

Таблица 12

Статистические величины чувашского абзаца

Средняя длина абзаца	Стили	
	научный	богослужебный
в предложениях	5,54	10,18
в словах	75,98	153,99

**RESUME**

**Fomin E. (Cheboksary)**

**Statistical Characteristics of the Chuvash Language in Late  
19<sup>th</sup> and early 20<sup>th</sup> centuries**

The author analyses statistical studies of Chuvash texts Late 19<sup>th</sup> and early 20<sup>th</sup> centuries. Linguistics concerns non traditional branch of science in Chuvash. According to its undevelopment Chuvash material in general Linguistic quantitative researches is not often treated. The present research work is devoted to fulfill this gap partly.

The given work is one the first similar experiences in Chuvash Linguistics.