

КОЛИЧЕСТВЕН АНАЛИЗ НА НЯКОЛКО МАРКЕРА ОТ БИЛИНГВАЛНА РЕЧ

Владислав Маринов

Социално-икономическите и политическите промени в българското общество през последните две десетилетия оказва влияние и върху развитието на едно от „най-младите“ направления в езикознанието ни – социалингвистиката. След етапа на дефиниране на основни тезиси и концепции и формирането на понятиен апарат българските социалингвисти отделят все повече внимание на конкретни проблеми като езиковите отношения в семейството (К. Алексова, Ю. Стоянова и др.), билингвалността и диглосията (Х. Кючуков, М. Виденов, Н. Николова, В. Маринов, П. Сотиров и др.), проучване на езика на отделни градове (Б. Байчев, М. Виденов, Е. Каневска-Николова, Е. Димитрова и др.) и т.л.

Обект на това изследване е количественият анализ на пет маркера в билингвална реч¹ и техните количествени характеристики. Екскерпиран е лингвистичен материал от аудиозаписи, направени от автора в няколко селища в Северозападна България, чиито жители при успешна комуникация, в зависимост от обстановката и събеседника, използват български език или влашки диалект². Обобщаващите числови характеристики показват отклонението от началната база, която в единия случай е българската книжовна база, а в другия – околните монолингвални западнобългарски териториални диалекти. Наличието на две неутрални величини паложни и приложението на кръстосания анализ при интерпретацията на лингвистичния материал. Окончателните стойности на маркерите определиха като релевантни пет: 1. (А) Рефлекс на **t*; 2. (В) Наличие на средноевропейско л (*l*); 3. (С) Съвпадение на глаголните форми в 3л., ед.ч., сег.вр. и 3л., мн.ч., сег.вр., напр. *тя ходят, опия да не я намери*; 4. (D) Смесване на рода при формите на минало свършено деятелно причастие и минало несвършено деятелно причастие, напр. *тя е видял, той е била*; 5 (Е) Окончание *-м* за 1л., ед.ч., сег.вр. и *-ме* за 1л., мн.ч., сег.вр. при глаголи от I и II спрежение. Представителните стойности (вариационната средна аритметична величина \bar{x})

на тези маркери бяха изчислени чрез математическите форми на осредняване. Наред с прилаганите в социолингвистичните изследвания формули за средна величина: $F = R/P \cdot 100$, за непретеглената средна аритметична: $\bar{X} = \sum x_i / n$ и за непретеглено средно квадратично отклонение: $\sigma = [\sum (x_i - \bar{x})^2 / (n-1)]^{0.5}$ тук бяха използвани и формулите за позиционната средна величина (медиана): $Me = L_{Me} + [(f\sum + 1)/2 - C_{Me-1}] \cdot h / f_{Me}$ и средната честота в група (мода): $Mo = L_{Mo} + [(f_{Mo} - f_{Mo-1}) \cdot h] / [(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})]$. С тяхна помощ бе постигната по-голяма точност при изчисляването на резултатите от трансформациите на изходните данни. Част от етапите при този процес са: *анализ на идиолектите на всеки информатор, класификация на идиолектите във формациите, дефиниране на контрастни и неконтрастни речевни маркери, спрямо „тежестта им“ в билингвалната общност, образуване на нови класификационни схеми, които са съобразени с билингвалните типодии и т.н.*

На първия етап от изследването бяха подбрани две групи от информатори. В първата група бяха включени билингви със средно, полувисше или висше образование, а във втората – билингви с основно, начално или без образование. Тези две групи са представителни извадки от двата компонента на лингвистичната триада – *a*- и *c*-формациите. Третият компонент (*b*-формациите) от това триединство тук е игнориран заради специфичните условия за проявата му в конкретния билингвален микроареал.

В генеричната съвкупност от целия информаторски блок в *a*-формациите попадат идиолектите на 37 информатори. За целите на това изследване бе подбрана разговорна реч на 13 информатори, за която при съставяне на нулевата хипотеза (H_0) бяха заложили ниски стойности на проява на наблюдаваните маркери. Използваната статистическа методология налага дефинирането и на алтернативна хипотеза (H_1), която да се противопостави на основната нулева хипотеза. Водещо в алтернативната хипотеза бе предположението, че общественото мнение е надделяло и речевият филтър на билингвения индивид е по-пропусклив към отклоненията от книжовната норма, които са неконтрастни за останалите членове на групата.

Като представителни за *c*-формациите бяха избрани идиолектите на 17 информатори. Тук също бяха съставени две хипотези – нулева и алтернативна. В този случай за H_0 бе допуснато, че ще има по-голяма фреквентност на прояви на контрастните маркери, както и по-широк толеранс при оценката и самооценката от страна на билингвите. Алтернативната хипотеза H_1 се основаваше на обратното предположение – спорадична проява на контрастни

маркери, мотивирана от явен микрофон и присъствието на непознати лица, когато бяха направени някои от записите.

В началото на диагностичния анализ бяха изчислени средните стойности на всеки от маркерите, без да включим в тях билингвалния атипод. Тук по технически причини са дадени най-ниските (x_1), средните (Me) и най-високите стойности (x_n) за всяко от изследваните множества, но при изчисляването на окончателните резултати бяха включени всички налични данни. Средните стойности (медианите), се различават от употребяваните досега средни аритметични стойности по това, че в първия случай се борави с реалните прояви на явлениято. Например при анализ на идиолектите на трима информатори: I_1 , I_2 , I_3 , ако търсеният маркер има стойности на реализация 0,2 (20 прояви от 100 възможни) за I_1 ; 0,3 за I_2 и 0,7 за I_3 , неговата средна аритметична стойност е $(0,2+0,3+0,7)/3=0,4$. Докато според статистическия метод в този пример медианата е 0,3. Вторият резултат е подостоверен, защото е получил потвърждение чрез данните от събрания материал, докато първият (0,4) е по-скоро в сферата на футурологията и за да се потвърди, трябва да се приложи принципът за достатъчния обем на генералната съвкупност.

Количествените стойности на проява на наблюдавани маркери при информаторите от α - и ϵ -формациите са следните:

	x_1	Me'	Me	Me''	x_n
A	0,05	0,2	0,35	0,48	0,6
B	0,01	0,35	0,67	0,90	1
C	—	—	—	—	—
D	—	—	—	—	—
E	0,09	0,31	0,56	0,8	1

Табл. 1 Числови стойности за идиолекти от α -формациите.

	x_1	Me'	Me	Me''	x_n
A	0,3	0,46	0,6	0,73	0,94
B	0,8	0,84	0,89	0,97	1
C	0,28	0,3	0,51	0,69	0,88
D	0,11	0,19	0,26	0,3	0,47
E	0,54	0,63	0,7	0,9	1

Табл. 2 Числови стойности за идиолекти от с-формациите.

На втория етап от изследването бе проведен социолингвистичен експеримент, чиято цел бе проучване на оценката и самооценката на билингвите за наблюдаваните маркери. Всеки информатор слушаше записи на своя или чужда реч, след което трябваше да постави оценка с 0, 1, 2, 3. От социолингвистично гледище тези цифри означаваха: 0 = не мога да преценя, 1 = минимално отклонение от българската книжовна норма; 2 = отклонение от книжовната норма; 3 = значително отклонение от книжовната норма. За нуждите на статистиката цифрите бяха заменени със стойности: 0=0; 1=0,1; 2=0,3; 3=0,6. С помощта на тези коефициенти бе изчислена претеглената средна аритметична величина \bar{x} на всеки от наблюдаваните маркери по формулата $\bar{x} = \sum x f_i / N$, където N е броят на оценяващите, f_i – тежестта на маркера според билингвалния атитюд, а $\sum x f_i = x_1 f_1 + x_2 f_2 + x_3 f_3 + x_4 f_4$.

i	A	B	C	D	E	
0 (1)	17	9	3	3	23	x_1
0,1 (2)	8	3	0	0	1	x_2
0,3 (3)	3	8	2	5	2	x_3
0,6 (4)	2	10	25	23	4	x_4
	0,09	0,29	0,52	0,51	0,10	\bar{x}

Табл. 3 Стойности на \bar{x} според билингвалния атитюд

Тези данни показват, че билингвите определят като най-контрастни маркерите С и D, а най-неконтрастни – А и Е.

За информационния блок от 50 информатори от двете формации, с интервал на проява на всеки от петте маркера 0,25 (или 25 реални прояви от 100 възможни), медианата отразява средната честота на маркерна реализация, а модата – тенденцията при реализациите. Предимството на медианата и модата пред средната аритметична величина е, че те игнорират случайните единични минимални и максимални стойности и дават достоверни резултати при изследването на групи.

Компонентите във формулата $Me = L_{Me} + f(f_e + 1)/2 - C_{Me-1} / f_{Me}$ са: L_{Me} – долната граница на медианната група; C_{Me-1} – общият брой на единиците (кумулятивна честота) до групата на медианата; h – ширината на интервала в медианната група и f_{Me} – броят на единиците в медианната група.

Данните за маркерните прояви в общия блок от информатори могат да бъдат обобщени в следната таблица:

	A	C _A	B	C _B	C	C _C	D	C _D	E	C _E
0	2	2	38	38	40	40	45	45	5	5
0,01 - 0,25	5	7	0	38	0	40	0	45	3	8
0,26 - 0,50	20	27	0	38	0	40	0	45	10	18
0,51 - 0,75	8	35	2	40	7	47	0	45	27	45
0,76 - 1	15	50	10	50	3	50	5	50	5	50

Табл. 4 Прояви на маркерите и стойности на кумулативните честоти

Първо се открива номерът на медианния случай по формулата $\sum C/2$. При 50 информатори медианният случай е 25, който е еднакъв петте маркера за горната таблица. Ако се заместят символите от формулата за Me с данните за маркер А, ще се получи: медианна група = 0,26 – 0,50, тъй като 25 за първи път се открива в $C_A = 27$; $L_{Me} = 0,26$; $C_{Me-1} = 7$; $h = 0,25$, а $f_{Me} = 27$.

$$MeA = 0,26 + (25 - 7) \cdot 0,25 / 20 = 0,485$$

От което следва, че средната реализация на маркер А в речта на билингвите е $\approx 48\%$. Стойностите за останалите маркери са: $MeB = 0,164$

($\approx 16\%$), $MeC=0,156$ ($\approx 16\%$), $MeD=0,138$ ($\approx 14\%$), $MeE=0,574$ ($\approx 0,57\%$). Релацията между медианата и модата ще покаже тенденцията при маркерните прояви.

Във формулата на модата

$$Mo = L_{Mo} + [(f_{Mo} - f_{Mo-1}) \cdot h] / [(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})]$$

компонентите са: L_{Mo} – долната граница на модалния интервал (интервала с най-висока честота на признака), f_{Mo} – броят на единиците в модалния интервал, f_{Mo-1} – броят на единиците в предмодалния интервал, f_{Mo+1} – броят на единиците в следмодалния интервал и h – ширината на модалния интервал. Ако бъдат използвани данните от таблица 4, Mo за маркер А ще е:

$$MoA = 0,26 + [(20-5) \cdot 0,25] / [(20-5) + (20-8)] = 0,398 \approx 40\%.$$

Реализациите на останалите маркери са: $MoB = 0,125$ ($\approx 12\%$), $MoC = 0,125$ ($\approx 12\%$), $MoD = 0,125$ ($\approx 12\%$), $MoE = 0,618$ ($\approx 62\%$).

Въз основа на резултатите от сравнителния анализ на данните може да бъде направено заключението, че от петте наблюдавани маркера с най-голям потенциал на развитие е употребата на окончание -м в 1л., ед. ч., сег. вр. и -ме в 1л., мн. ч., сег. вр. за глаголи от I и II спрежение. Това твърдение се потвърждава и от състоянието на този маркер в съвременната българска разговорна норма. Стойностите на модата за маркера Е в разговорната реч на билингвите са по-високи от меридианните, т.е. кривата му на развитие е възходяща. Този факт, съчетан с ниския контрастен коефициент (0,1), илюстрира увеличената честота на маркера Е и вероятно ще става все по-контрастен дори и за информатори от a -формациите. Останалите 4 маркера са с намаляваща честота, като най-активен е еднофонемният рефлекс на $\text{ʃ} > \text{с}$, но и при него се наблюдава тенденция на спад в употребата, въпреки че не се отчита като контрастен от билингвите. Реализациите на средноевропейско л, съвпадение на глаголните форми в 3л., ед. ч., сег. вр. и 3л., мн. ч., сег. вр. и смесване на рода при минало свършено деятелно причастие и минало несвършено деятелно причастие имат по-висок коефициент на контрастност и са свързани с влиянието на първия разговорен код (според диахронната класификация на Р. Бел³).

Днес българският език е първи и първичен за билингвите от Северозападна България, което мотивира отчетените ниски стойности на модата и спорадичните прояви (предимно в речта на възрастните билингви) на маркерите **В**, **С** и **Д**. Ако развитието на българската разговорна норма следва сегашната си посока, окончанията **-м** за 1л., ед. ч., сег. вр. и **-ме** за 1л.,

мн.ч., сег. вр. при глаголи от трите спрежения вероятно ще увеличат фреквентността си в речта на билингвите от крайния български северозапад.

БЕЛЕЖКИ

¹ Тук под *билингвална реч* разбираме 'реч, продуцирана от билингви', а за билингви приемаме лицата, които владеят форми на две езикови формации без значение на йерархията при употребата им.

² За понятието *влашки диалект* вж. Маринов 2008: 11.

³ Вж. повече у Бел 1980: 154–155.

ЛИТЕРАТУРА

Алексова, Алексов 1996: Алексова, К., Г. Алексов. Приложението на метода на главните компоненти при изследването на социалната диференциация на езика – В: Проблеми на социолингвистиката, т. 5, С., 1996.

Алексова 2005: Алексова, К. Сравнително изследване на резултатите от няколко метода за установяване на йерархията на социолингвистичните променливи – В: Научни изследвания в чест на проф. д-р Боян Байчев, В. Търново, 2005, с. 199–215.

Бел 1980: Белл, Р.Т. Социолингвистика. М., 1980.

Гатев 1995: Гатев, К. Въведение в статистиката. С., 1995.

Маринов 2008: Маринов, В. Билингвална интерференция в крайния български северозапад. В. Търново, 2008.

Съйкова, Тодорова 2000: Съйкова, И., С. Тодорова, Статистическото изследване. С., 2000.