

МЕТОДИ ЗА ОБРАБОТКА НА ЕМПИРИЧНИ СОЦИОЛИНГВИСТИЧНИ ДАННИ

КРАСИМИРА АЛЕКСОВА

СОФИЙСКИ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

krasimiraaleksova@gmail.com

METHODS FOR PROCESSING EMPIRICAL SOCIOLINGUISTIC DATA

KRASIMIRA ALEKSOVA

SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”

krasimiraaleksova@gmail.com

The paper presents and discusses some of the main methods of statistical processing of empirical sociolinguistic data, which have often been applied in Bulgarian sociolinguistic research. The focus is not on the logic and the calculations of each statistical method but on the use of an appropriate method of proving or rejecting those research hypotheses which are related to three main relationships: 1) social stratification Vs speech variation, 2) the relationship between variants of sociolinguistic variables or socio-demographic features which are taken into account, and 3) linguistic attitudes related to sociolinguistic markers Vs sociodemographic characteristics of respondents. In the first part, the article deals with various procedures of calculating average values of sociolinguistic variables' variants applied in different studies on this topic. In the following part, the paper gradually presents the applicability of dispersion, regression and correlation analysis, the chi-square method, as well as the principal components analysis in sociolinguistic and psycholinguistic studies in Bulgaria.

Keywords: sociolinguistic variable, dispersion analysis, correlation and regression analysis, chi-square analysis, principal components analysis, language attitudes

I. Въведение

Основната цел на настоящата работа е да представи методите за обработка на социолингвистични емпирични данни, използвани за доказване на връзката между социалната стратификация и речевата вариантност в трудове на български изследователи. По-общата задача – преглед на статистическите методи в социолингвистични изследвания, не е постижима в рамките на една статия. Акцентът тук попада не върху описанието на статистическите методи, нито върху изчисленията и формулите, а върху това как тези методи служат за потвърждение или отхвърляне на хипотези за връзките между социално-демографските и стиловите фактори, от една страна, и от друга, социолингвистичните променливи.

В проучванията, принадлежащи към т.нар. вариационна социолингвистика, основните хипотези са свързани с наличието на обвързаност между речевите особености в идиолектите на отделни информатори или на социолингвистично значими групи изследвани лица със социално-демографските характеристики на тези индивиди и/или със спецификата на текста, по-точно с т.нар. от У. Лабов контекстуални стилове. Те се отличават с различна степен на внимание, грижливост на речта, напр. четене на списък от думи, четене на двойки думи, неформален разговор, официален разговор) (Labov 1972, вж. също Trudgill 1974 и Ангелов/Angelov 1999).

Първо е необходимо да се изясни понятието *социолингвистична променлива* – основната единица, която дава възможност да се операционализира изследването на езиковата вариантност във връзка със социалната нееднородност и ситуационното многообразие. Тук насочвам към рецепцията у нас на най-цитираното определение за лингвистичната променлива – това на У. Лабов: „Променливата е „непоследователност или разлика, изразена чрез конкретна езикова форма (форма на съществуването на езика, формация – бел. М.В.), видима в сравнение с някакъв абстрактен стандарт“ (Labov 1966: 35). Под ‘вариант’ авторът разбира „специфично значение на променливата“, от което се разбира, че променливата се реализира чрез своите варианти.“ (Виденев/Videnov 1998: 78). Не всеки езиков елемент, реализиращ се чрез варианти, придобива статус на лингвистична променлива, тъй като някои варианти имат случаен или временен характер. Затова едно от същностните изисквания към лингвистичната променлива е варирането да бъде системно. Освен това варирането не е свободно, а ограничено в определени парадигми и модели на езиковата система. Към критериите, особено когато става дума за социолингвистичните променливи, някои учени добавят изискванията за висока честота, известна неподатливост към съзнателно потискане, лесна изчисляемост по линейна скала (Labov 1966: 49), висока степен на стратификация и асиметрична дистрибуция в широк кръг от възрастови групи или социални слоеве в обществото (Labov 1972: 8).

Верифицирането на хипотези в социолингвистичните проучвания се постига най-често чрез проверка на наличието на връзки между вариантите на подбрани от изследователите лингвистични променливи, дефинирани като зависими променливи (признаци резултати), и параметри от социалната константа на индивида, възприемани като независими променливи (признаци фактори). Проверката се извършва чрез различни статистически методи, като изборът на метод за анализ зависи от типа на признаците според възможността те да бъдат метрирани. От статистическа гледна точка признаците биват: количествени (метрирани, вариационни) и качествени (неметрирани, категорийни). Количествените признаци могат да се измерват с числа (стойности), напр. възраст, месечна заплата и др., а качествените се характеризират описателно, с категории; те не могат да се мерят, напр. пол, професионален статус и др. От своя страна качествените при-

наци се поделят на подредими (квантифицируеми) и неподредими (класификационни). Качествен неподредим признак е този, чиито разновидности „не отразяват никакви количествени различия, поради което не могат да се подреждат „от малко към голямо“, „от слабо към силно“ и пр. (например пол, семейно положение)“ (Венедиков/Venediktov 1993: 106). Качествени подредими признаци са онези, чиито разновидности „могат да бъдат подредени, на тях могат да се припишат числа, съответни на количественото различие, което изразяват“ (Венедиков/Venediktov 1993: 106), напр. степен на образование.

Определянето на вида на признаците е необходимо и за да се опише видът на скалата, по която те се измерват. Количествените признаци се измерват по интервална скала, качествените неподредими – по номинална скала, а качествените подредими – по ординална скала. Интервалната скала се прилага, ако значенията на признака имат числов израз. Номиналната скала „съдържа възможните и необходими определения на категориите (атрибутивните) признаци, за да се получи съответната групировка на единиците, която е нужна на анализа“. Ординалната скала „съдържа такива поделения, които не са числови, но по съдържание изразяват някаква степен, известно различие според признака на единиците“ (Гатев, Спасов, Радилов/Gatev, Spasov, Radilov 1989: 24–25).

По-нататък в настоящия текст се представя как се използват различни процедури по изчисляване на средни величини на вариантите на социолингвистичните променливи в няколко проучвания. След това последователно се разглежда приложението на дисперсионния, регресионния и корелационния анализ, хи-квадрат метода, както и на факторния анализ по метода на главните компоненти в социолингвистични и социалнопсихологически проучвания у нас.

II. Средни величини на варианти на социолингвистичните променливи

Изчисляването на средните величини на вариантите на социолингвистичните маркери е необходимо, за да се провери дали избраните речеви особености са действително значими, като и ниските, и високите стойности са показателни за изследователя. Те дават информация за състоянието на изборния за проучване вариант на дадена променлива в речта на всеки един от информаторите. За изчисляването на средните величини се използва формулата $F = R/P \cdot 100$, където F е величината на вариант x на дадена социолингвистична променлива, представена в проценти; P представлява сумата от потенциално съществуващите случаи, в които би могъл да се реализира вариантът x , а R = реализираните случаи на поява на варианта x . Авторите проследяват в извадки от транскрибирани записи на спонтанна реч с дължина, еднаква за отделните информатори, потенциално съществуващите и реално реализираните случаи на поява на некодифицирания вариант, за да изчислят средните величини по дадената формула (вж. напр.

у Алексова/Aleksova 2000).

Получените стойности дават възможност да се съпостави кои от избраните в изследването варианти на социолингвистични променливи имат относително по-високи средни величини, а кои по-ниски, което говори за разпространението на некодифицирания вариант в речта на различни информатори, в дадена социолингвистично значима група или в макро-общността като цяло. Средните величини разрешават да се съпоставят идиолекти, да се формират групи информатори, да се изведат изводи както за мястото на дадена социолингвистична променлива сред другите наблюдавани променливи, така и за причините, довели до наблюдаваната стойност на некодифицирания вариант.

Стойностите на некодифицирания вариант на променливата могат да бъдат показателни и за владенето на местната норма, стига избраните от изследователя социолингвистични променливи да имат варианти, характерологични за съответния диалект. Подобен подход може да се открие в работата на П. Кънчева, изследваща речта на преселници от Западните покрайнини в София. Честотата на варианта на всяка от променливите, който е част от диалектната норма, е в основата на изчисляването на диалектния индекс на речта на всеки информатор (Кънчева/Kancheva 2008: 99–108). За тази цел авторката използва формулата: $I = \frac{Fn}{n}$, в която с **I** е означен диалектният индекс, **Fn** представлява сумата от стойностите на диалектния вариант за всички променливи, а **n** е броят на променливите – 12 (Кънчева/Kancheva 2008: 99). Получените диалектни индекси са ранжирани по ординална скала и речта на всеки от информаторите е получила своя рангов номер. Подреждането след това се съпоставя в таблици с възрастта, пола и образователното равнище, за да се провери хипотезата дали еднакви или близки по диалектно съдържание идиолекти са свързани с определени социално-демографски характеристики на информаторите. Диалектният индекс и средните стойности на диалектните варианти служат на авторката да вземе решение за статуса на всяка от социолингвистичните променливи в изследването, т.е. дали става въпрос за индикатори, или за маркери според делението на У. Лабов.

В немалко от социолингвистичните изследвания у нас при проучване на социалната детерминираност на речевите особености се работи със социолингвистични променливи, чийто базисен вариант е книжовноезиковият. Измерват се средни величини на некодифицирания вариант. Но в работата на Е. Димитрова като изходен вариант за съпоставка се използват както книжовната норма, така и местната норма. Това е продиктувано от виждането на Е. Димитрова, че съществуващата конкуренция между двете норми в гр. Криводол може да бъде изследвана чрез отчитане на варирането спрямо всеки от двата еталона. Авторката е подбрала 12 социолингвистични маркера от фонетичен, морфофонетичен, морфологичен и морфосинтактичен характер. Изчисляването на стойностите на социолингвистичните променливи се прави за всяка от трите избрани групи информа-

тори: ученици, представители на местната интелигенция и представители на земеделците от гр. Криводол. Е. Димитрова използва формулата

$$X = \frac{(n_1 + n_2 + \dots + n_7) \times 100}{N_{\text{максимално}} (140/280)} \quad \text{Димитрова 2004: 77),}$$

като X е всяка социолингвистична променлива X във всяка една от групите, n са отклоненията от съответната норма (книжовната или местната), а индексът на n посочва номера на информатора от всяка група (по 7 души във всяка от трите групи). Сумата $(n_1 + n_2 + \dots + n_7)$ се отбелязва с N и представлява общия брой на отклоненията от съответната норма. От своя страна $N_{\text{максимално}}$ е равно на 140 или на 280 (280 е само в текст 10), като сумата от 140 е произведението от броя информатори в групата и 20-те експериментални единици във всеки тест. $N_{\text{максимално}}$ сочи 100% владееене на съответната норма (книжовната или местната) (Димитрова/Dimitrova 2004: 76–77). По този начин се получават процентни стойности за степен на владееене на съответната норма по отношение на дадена променлива за всяка една от групите.

III. Приложение на дисперсионния анализ

Дисперсионният анализ се използва, ако се изследва връзката между количествен признак резултат и качествен признак фактор (Харалампиев/Haralampiev 2007: 15). В социолингвистичните проучвания дисперсионният анализ се прилага, когато се търси отговор на въпроса дали варирането на социолингвистичните променливи, възприемани като зависими променливи (или още – признаци резултати), е резултат от системното въздействие на избраните от изследователя социални или ситуативни фактори, приети като независими променливи (признаци фактори), или пък това е резултат от влиянието на случайни фактори. У нас във вариационни изследвания е използван както еднофакторен, така и двуфакторен дисперсионен анализ (напр. у Алексова/Aleksova 2000; Кънчева/Kancheva 2008 и съответно Ангелов/Angelov 1999). Изискването за приложението на дисперсионния анализ е свързано с характера на признаците фактори и признаците резултати – вариантите на социолингвистичните променливи като признаци резултати трябва да са количествени, т.е. трябва да са метрирани, да се измерват с числови стойности, а признаците фактори да са качествени (неметрирани, категорийни).

При дисперсионния анализ се тръгва от твърдението, че върху резултативния признак, напр. избрания от изследователя вариант на социолингвистичната променлива, действат случайни и систематични фактори, които поради своята комплексност не могат да се разделят видимо. Затова се предполага, че ако се оформят групи от социолингвистични зависими променливи, обособени по съответния факторен признак (чието влияние се изследва), би могло да се съди за наличието или отсъствието на някакво

факторно влияние върху зависимата променлива по разликите на средните величини на всяка група. Дали тези разлики на средните величини на групите са случайни, или са статистически значими, се проверява чрез статистическа проверка на хипотезата, че средните величини за всяка група са равни – H_0 (нулева хипотеза), срещу алтернативната хипотеза, че средните величини се различават (т.е. разликата на средните величини на зависимата променлива се дължи на систематичното въздействие на независимия фактор) – H_1 . При вземане на решение за приемане или отхвърляне на H_0 винаги съществува риск да се направи грешка. Грешка от първи род има, когато се отхвърли нулевата хипотеза, при положение че тя е вярна. Обикновено при лингвистическите и социологическите изследвания се работи с вероятност за грешка от първи род (т.нар. ниво на значимост) = 0.05 (Венедиков/Venediktov 1992: 16). В дисперсионния анализ проверката на хипотезата H_0 срещу H_1 се извършва чрез статистическия F критерий (тест) на Стюдънт-Фишер (Гатев, Спасов, Радилов/Gatev, Spasov, Radilov 1989: 165–188).

Резултатите от еднофакторния дисперсионен анализ помагат на лингвистите да приемат или отхвърлят своите предварителни хипотези за наличието на системна детерминираност на вариацията на лингвистичните променливи от дадени социално-демографски фактори. Дисперсионните анализи дават важни данни не само с положителните, но и с негативните резултати, които сочат коя независима променлива на коя зависима лингвистична променлива не влияе. Напр. дисперсионният анализ в изследването на 28 софийски семейства показва, че факторът ‘място на оформяне на първоначалния езиков опит’ на изток или на запад от ятовата изоглоса оказва системно влияние върху всички основни зависими (лингвистични) променливи, избрани в изследването, с изключение на окончанието *-ме* за 1 л. мн. число, сег. време при глаголи от I и II спрежение. Факторът ‘тип професия’ (с две равнища: професия, незадължаваща спазване на книжовната норма, или професия, задължаваща нейното спазване) влияе върху всички основни западни по локален произход социолингвистични променливи с изключение на непалаталността на съгласната пред глаголното окончание в 1 л., ед. число и 3 л., мн. число, сег. време.

Степента на образование системно влияе върху всички изследвани от авторката западнобългарски по произход променливи без непалаталността на съгласната при: глаголи от II спр. (тип *спя, спят*), глаголи от III спр. (тип *вечерям, вечеряш* и т.н.), при съществителни имена, завършващи на *-алня*, при членувани съществителни със суфикс *-ар, -яр* и *-тел* и др. Степента на образование не оказва системно влияние и върху окончанието *-ме* за 1 л. мн. число, сег. време при глаголи от I и II спрежение (Алексова/Aleksova 2000). Тези данни са важни, тъй като доказват, че непалаталността на предфлексийната съгласна в сегашни глаголни форми (*вървъ, вървът, вечерам, вечераш* и т.н.) в речта на изследваните лица не зависи системно от императива да се спазва книжовната норма в процеса на про-

фесионална дейност. Съществува ясна тенденция т.нар. мекане (*четеме, пишем*) все повече да не зависи от степента на образование и от мястото на оформяне на първоначалния езиков опит на изток или на запад от ятовата изоглоса.

В социолингвистичното проучване на речта на преселници от Западните покрайнини в София П. Кънчева доказва чрез еднофакторни дисперсионни анализи между некодифицирани варианти на подбраните социолингвистични променливи и всеки от трите значими според изследователката независими социално-демографски фактори, че „независимите променливи степен на образование, пол и възраст оказват системни въздействия върху социолингвистичните променливи от тип В“ (Кънчева/Kancheva 2008: 98). В тип В авторката включва екането, премета на ударението в аорист и в аористното причастие, отмета на ударението в императив, отмета на ударението при някои глаголи от второ спрежение и производни от тях форми, глаголното окончание *-м* за 1 л. ед.ч. и частицата за бъдеще време *че*. От тези резултати произтича изводът, че вариантността, системно обусловена в речта на тези преселници от образованието, пола и възрастта, засяга предимно маркери, които са широко разпространени именно в западнобългарските територии. Дисперсионните анализи доказват, че същите тези три фактора не оказват системни влияния върху вариантността на лингвистичните променливи, които са специфичен белег на традиционния териториален диалект на преселници (Кънчева/Kancheva 2008: 98).

П. Кънчева използва еднофакторния дисперсионен анализ и за да установи влиянието на образованието, пола и възрастта върху диалектната компетентност на информаторите. Данните за нивото на диалектната компетентност са събрани чрез анкета, включваща освен въпроси за социално-демографската идентификация на преселниците от Западните покрайнини и тестови въпроси за избор между 4 алтернативни отговора (Кънчева/Kancheva 2008: 149–157). Диалектната компетентност има три нива: правилен отговор, неправилен отговор и колеблив отговор. Дисперсионният анализ показва, че степента на образование (с две равнища: а) основно и б) средно, полувисше, висше) статистически значимо определя различията в диалектната вещина на преселниците. Факторът пол не влияе статистически значимо върху разпределението на отговорите в трите групи на диалектната компетентност. Възрастта в статистически значима степен определя различието в разпределението на тези три вида отговори на анкетираните лица от трите възрастови групи (до 29 г., между 30 и 59 г., над 60 г.) (Кънчева/Kancheva 2008: 157).

Дисперсионният анализ се използва и в изследването на П. Бокова (Бокова/Вокоча 2000). Чрез него авторката търси статистически значимите зависимости между избраните шест социално-демографски фактора и варианти на социолингвистични променливи (фонетични, лексикални и граматически заемки) в речта на различни групи сред българите от с. Суворово,

Украйна. Оказва се, че възрастта влияе върху вариациите на лексикалните заемки и върху граматическата интерференция. Видът на упражняваната професия влияе върху лексикалните заемки, калките и полукалките, но не и върху семантичната интерференция. Типът на брака (смесени и еднородни бракове) оказва влияние върху лексикалните заемки. Дисперсионните анализи сочат, че вариантността на променливата 'граматична интерференция' е най-слабо зависима от социално-демографската нееднородност на шекерлийските българи. Това подкрепя според авторката предварителната ѝ хипотезата, че граматичната система на шекерлийския диалект изпитва най-слабо влиянието на руския език в сравнение с лексикалното и фонетичното равнище и че именно граматичните особености на този говор са се запазили в значителна степен (Бокова/Vokova 2000).

В социалнопсихологическия труд на Ст. Абазова (Абазова/Abazova 2014) се изследват нагласи към писмена реч, съдържаща жаргонизми, два вида вулгаризми и професионална лексика. Авторката използва дисперсионния анализ, като изчисленията са направени със статистическия пакет SPSS и е използван методът MANOVA (Multivariate analysis of variance). Ст. Абазова формулира една независима променлива – пол на говорещия, и осем зависими променливи – осем ситуации, в които биха могли да се появят изброените четири вида лексикални маркери. Респондентите (ученици, студенти, учители и университетски преподаватели) оценяват в анкетата доколко уместни са тези лексикални средства във всяка една от следните осем ситуации, различаващи се по степента на формалност на общуването: в интервю по телевизията, в разговор с приятели, в семейна среда, в разговор с непознати в автобуса, при разговор с колеги, във фейсбук или друга социална мрежа, в интернет форум, в собствен блог. Скалата, по която се оценява уместността, е петстепенна ординална скала: „напълно уместно“, „по-скоро уместно“, „не мога да определя“, „по-скоро неуместно“, „напълно неуместно“. Приложен е методът на семантичния диференциал. В изследването признакът фактор (независимата променлива) е качествен неподредим, а оценките за уместност в ролята на признаци резултати (зависими променливи) са качествени подредими признаци. В дисертационния труд се изследва и влиянието на факторите *местожителство* на респондентите и *учещ/завършил* върху оценката за уместност на изказването, съдържащо всеки от четирите вида лексикални маркери. Потърсени са статистически значими разлики в оценката на респондентите и според осемте ситуации на общуване. Авторката проверява дали има корелация между възрастта на анкетирания и възрастовата група, към която той отнася говорещия според използваните в изказването му жаргонизми, два вида вулгаризми и професионална лексика (като анализите се правят за всеки от четирите маркера поотделно). От гледна точка на социалната идентификация и на личностната оценка на говорещия по появата в речта му на четирите вида лексикални маркери особено ценни са анализите на Ст. Абазова върху данните, показващи дали има статистически значима

зависимост на оценката за говорещия от пола, местожителството и статуса учещ/завършил на анкетиранията лица (Абазова/Abazova 2014).

В социолингвистиката е познат и друг подход към дисперсионния анализ. Той може да бъде открит в работите на У. Лабов (Labov 1966), П. Тръдгил (Trudgill 1974) и А. Ангелов (Ангелов/Angelov 1999), които прилагат непълен двуфакторен дисперсионен анализ. При него се търсят връзки между релевантни за изследваната общност варианти на социолингвистични променливи като зависими променливи, от една страна, а от друга – социални и контекстуални фактори като независими променливи. Дисперсионният анализ е определен като непълен, защото не се правят изчисления за приемането или отхвърлянето на нулевата хипотеза, а се приема, че и социално-демографските, и контекстуалните фактори (т.нар. от Лабов контекстуални стилове) влияят върху вариацията на зависимите социолингвистични маркери. В българската социолингвистика А. Ангелов използва този подход и оформя статистическите наблюдения от двуфакторния дисперсионен анализ в таблици, в които се комбинират всяко от равнищата на съответния социално-демографски фактор с всяко от равнищата на контекстуалния фактор за всяка от избраните лингвистични променливи. В таблицата се записват средните стойности на вариантите на променливите, които се движат между 0.00%, което сочи последователна употреба на книжовния вариант, и 100% – последователна поява на не книжовния вариант. Таблица 1 представя данните на А. Ангелов за ковариациите на променливата ‘депалатализация при глаголни окончания в сегашно време’. Първият независим фактор (фактор А) е дефиниран като контекстуален стил с равнища, различаващи се по степента на внимание на говорещия към собствената му реч: четене на списък от думи, четене на текстов пасаж, официална реч и небрежна реч. За независим фактор Б е из-

Таблица 1. Променлива ‘депалатализация при глаголни окончания в сегашно време’ – средни процентни стойности за всяка социална група в различните контекстуални стилове (Ангелов/Angelov 1999: 149)

СОЦИАЛНИ ГРУПИ	КОНТЕКСТУАЛЕН СТИЛ			
	Четене на списък от думи	Четене на текстов пасаж	Официална реч	Небрежна реч
Родени в Гевгели или околността	40.00	60.00	90.00	100.00
Първо поколение потомци	20.00	31.00	80.00	100.00
Второ поколение потомци	0.00	0.00	100.00	100.00
Пришълци по брачна линия	0.00	40.00	60.00	80.00
Пришълци по съседство	20.00	46.00	76.00	90.00

брана принадлежността на живеещите в Гевгелийския квартал, изследван от автора, към една от петте социолингвистично значими групи информатори: родени в Гевгели или околността, първо и второ поколение потомци, както и пришълци по брачна линия или по съседство.

Прилагането на двуфакторния дисперсионен анализ онагледява чрез резултатите си системното въздействие както на социално-демографски, така и на ситуативно-контекстуални фактори. В този смисъл той действително отчита варирането на социолингвистичните променливи от типа *маркери* според разбирането, което влага в понятието *маркери* У. Лабов – лингвистични променливи, вариращи както спрямо равнищата на социалните фактори, така и спрямо равнищата на контекстуалния стил.

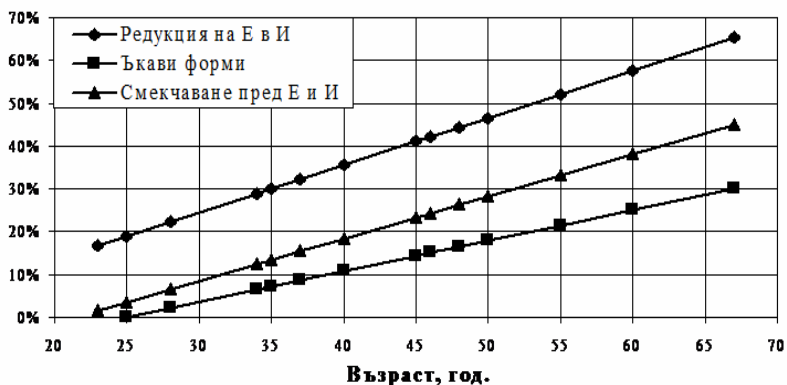
IV. Приложението на регресионния анализ

Когато се изследва връзката между количествен признак резултат, в случая варианти на социолингвистични променливи като метрирани (измерващи се с числа), и количествен признак *фактор*, в случая количествени социални или демографски фактори (напр. възраст, продължителност на местоживеенето в дадено населено място и под.), може да бъде използван регресионен анализ. В българската социолингвистика регресионен анализ е прилаган при установяване на връзката между възрастта на информатори от София с източнобългарски произход и източнобългарските маркери в тяхната реч (Алексова/Aleksova 2000). В проучването на речевата акомодация в 28 столични семейства се установява в резултат на дисперсионния анализ, че съществува връзка между демографския фактор възраст и източнобългарските маркери в речта на 21 изследвани лица. В изследването се поставят въпросите дали между възрастта на тези лица и допусканията от тях източнобългарски маркери съществува корелация и какъв е типът на тази зависимост – положителна (дали с нарастването на възрастта расте и величината на източнобългарските маркери) или отрицателна (дали с нарастването на възрастта намалява величината на тези маркери).

За да бъде изследвана връзката между признаците резултати (основните източнобългарски променливи в изследването) и признака фактор (възрастта), е приложен линеен модел за еднофакторен регресионен анализ от типа $y = a + bx$, като y е източнобългарският вариант на всяка от 6 избрани променливи, а x е социално-демографският фактор възраст. При линейната зависимост от вида $y = a + bx$ a е свободният член (константа), която показва в графиката точката, в която регресионната линия пресича ординатната ос, а b се нарича *регресионен коефициент*. Той показва с колко процента се изменя y при изменение на x с една година. За всяка от 6-те избрани източнобългарски променливи е направен отделен регресионен линеен еднофакторен анализ. В получените графики (тук е представена на фиг. 1 само една от тях) линията, която изразява зависимостта между приз-

нака резултат (социолингвистичната зависима променлива) и признака фактор (независимата променлива възраст), се нарича *регресионна линия* и „отговаря на едно условие: да минава най-близо до всички точки така, че сумата от квадратите на разликите между емпиричните стойности на y и техните оценки \hat{y} , намерени за всяка стойност на x , които биха се намирили на линията, да е минимум, т.е. $(y-\hat{y})^2$ minimum“ (Гатев, Спасов, Радилков/Getev, Spasov, Radilov 1989: 193).

**Стойности на източнoбългарските лингвистични
променливи в зависимост от възрастта**



Фигура 1. Стойности на източнoбългарските варианти на три социолингвистични променливи в зависимост от възрастта (Алексова/Aleksova 2000)

Чрез функциите на електронна таблица е извършен регресионният анализ и са изчислени стандартните грешки и *коэффициентът на корелация* (коэффициент на Пирсън) r^2 . Най-голямо познавателно значение за анализа има *коэффициентът на корелация* r^2 и *регресионният коэффициент* пред независимата променлива – b . „Корелационният коэффициент измерва теснотата на зависимостта. Той може да приема стойности от 0 до 1... Колкото е по-близък до 1, толкова зависимостта е по-силна“ (Гатев, Спасов, Радилков Getev, Spasov, Radilov 1989: 198). Зависимостта между признака резултат (всеки от 6-те източнoбългарски варианта) и признака фактор (възрастта) се смята за положителна (т.е. с нарастване на възрастта x нараства и признакът резултат y , ако знакът пред коэффициента b е плюс), или отрицателна (с нарастването на x да намаляват стойностите на y), ако знакът пред коэффициента b е минус. Между признака фактор възраст и всеки от шестте източнoбългарски варианта на социолингвистичните променливи в речта на 21 информатори съществува според направените регресионни анализи положителна слаба корелационна зависимост – до 0.3. Това означава, че с нарастването на възрастта се увеличава средната величина (в %)

на източнобългарските по тип отклонения от книжовната норма на изследваните столичани с източнобългарски произход.

V. Приложение на корелационния анализ

Ако с регресионния анализ се изследва наличие и посока на връзката между количествени признаци резултати (напр. средни величини на варианти на социолингвистични променливи) и количествени признаци фактори (напр. възраст, продължителност на столичното местожителство и др.), то корелационният анализ изследва силата на връзката, т.е. чрез него се установява степента на свързаност между две променливи. Най-често използваният индекс за линейна корелация е този на Пийърсън, отбелязван с **R**. Той заема стойности от -1 до 1, а когато **R** = 0, то тогава липсва линейна връзка (липсва корелация) между признака резултат и признака фактор. Знакът пред стойността на **R** – плюс или минус, сочи права или съответно обратна зависимост между X и Y. Известни са няколко мнения за тълкуване на силата на връзката според стойностите на коефициента на корелация. Според едно от тях, ако **R** е до 0.3, има слаба корелация между X и Y, ако **R** е между 0.3 и 0.5, има умерена корелация, между 0.5 и 0.7 – значителна корелационна зависимост, между 0.7 и 0.9 – висока корелационна зависимост, между 0.9 и 1 – много висока корелационна зависимост. Друга скала предлага следните тълкувания: между 0 и 0.2 – слаба корелация, от 0.2 до 0.4 – умерена корелация, от 0.4 до 0.6 – значителна корелационна зависимост, от 0.6 до 0.8 – висока по степен корелация, от 0.8 до 1 – много висока корелационна зависимост.

В социолингвистичните изследвания целта на корелационния анализ е да се идентифицира степента на свързаност между количествен признак резултат (зависима променлива, означавана с X) и количествен признак фактор (независима променлива, означавана с Y). Като признаци резултати и признаци фактори могат да бъдат приети напр. средните величини на варианти на социолингвистични променливи и метрирани социално-демографски характеристики и др. Например в изследването на речевата акомодация в 28 столични семейства (Алексова/Aleksova 2000) се доказва чрез корелационен анализ, че между факторите продължителност на брака, продължителност на столичното местожителство и възраст има силна по степен корелационна зависимост, т.е. коефициентът на корелацията е от 0.7 до 0.9 (вж. Таблица 2).

Изследването на връзките между социално-демографските характеристики на информаторите е важен етап от вземането на решение кои независими (социални) променливи (дефинирани също като признаци фактори) трябва да бъдат подбрани в конкретното социолингвистично проучване. Може да се окаже необходимо да се уедрят предварително набелязаните независими променливи (признаците фактори) поради логическа обвързаност и импликации между тях, които водят до несъществени различия в начина, по който те влияят върху речта на информаторите.

Таблица 2. Коефициенти на линейната корелация между три социално-демографски фактора (Алексова/Aleksova 2000)

Социално-демографски характеристики	Продължителност на брака	Продължителност на столичното местожителство	Възраст
Продължителност на брака	1.0000	0.7140	0.8567
Продължителност на столичното местожителство	0.7140	1.0000	0.7252
Възраст	0.8567	0.7252	1.0000

V. Приложение на хи-квадрат метода

Когато и признакът резултат, и признакът фактор са качествени признаци, най-подходящ метод за изследване на връзките между тях е **хи-квадрат** методът (вж. у Харалампиев/Haralampiev 2007: 10). В българските социалнопсихологически изследвания този метод е прилаган при изследване на въпроса съществува ли връзка между езикови нагласи (като признаци резултати) и социално-демографски характеристики на респондентите (като признаци фактори). При проверката за наличие или отсъствие на връзка се формулира нулева хипотеза, според която двете променливи (признакът фактор и признакът резултат) са независими. Алтернативната хипотеза гласи обратното, т.е. че между тях съществува закономерна връзка. Логиката и математическите изчисления при хи-квадрат анализа могат да се видят в множество учебници и помагала по статистика, затова тук няма да се разглеждат, а само ще бъде подчертано, че важна стъпка при проверката на статистическа хипотеза е определянето на риска за грешка α . В социологически изследвания, респективно при социопсихолингвистичните, се работи с грешка $\alpha = 0.05$. За да се вземе решение дали да се приеме, или да се отхвърли нулевата хипотеза, трябва да се сравни равнището на значимост с риска за грешка α . Ако равнището на значимост е по-малко от α , се отхвърля нулевата хипотеза и се приема алтернативната. В конкретния случай тя гласи, че между езиковите нагласи и социалните фактори, избрани от изследователя, съществува значима от статистическа гледна точка връзка.

При положение че е доказано наличие на връзка между признаците, може да се измери и силата на връзката. За тази цел се използва коефициентът на Крамер (V), който е нормиран в граници между нула и единица и се отчита по следната скала, сочеща степента на зависимост при дадена стойност на коефициента на корелация: слаба – до 0,3; умерена – над 0,3 до 0,5; значителна – над 0,5 до 0,7; голяма – над 0,7 до 0,9; много голяма – над 0,9 (Мишев, Цветков 1998). Познати са и други скали: от 0 до 0.3 –

връзката е слаба, от 0.3 до 0.7 – средна, и над 0.7 – силна (вж. Гоев/Goev 1996: 128–129); от 0 до 0.20 – слаба връзка, от 0.20 до 0.35 – умерена връзка, от 0.35 до 1 – силна връзка.

Ако признакът фактор е качествен подредим, може да се изследва и посоката на връзката, която показва как се променя признакът резултат, ако признакът фактор се увеличава: факторът се увеличава и резултатът се увеличава (наличие на права връзка); факторът се увеличава, а резултатът намалява (обратна връзка); факторът се увеличава, а резултатът няма строга посока.

Приложението на хи-квадрат анализа може да бъде видяно в два вида социалнопсихологически проучвания у нас. В първия вид се изследва връзката между признаците, по които се оценяват идиолекти, различаващи се както по броя, така и по характера на некодифицираните варианти на социолингвистичните променливи (Алексова/Aleksova 2002). В основата на това проучване са езиковите нагласи към 5 идиолекта (един представител на книжовната устна реч, два западнобългарски идиолекта с различен брой и характер на правоговорните отклонения от западен тип и два източни идиолекта с разлика в спецификата и броя на източнобългарските маркери), като респондентите изслушват на аудиозапис извадка от идиолекта. Идиолектите се оценяват по признаците правилна : неправилна, престижна : непрестижна, книжовна : диалектна, западнобългарска : източнобългарска, градска : селска, приятна : неприятна, столична : провинциална, благозвучна : неблагозвучна реч. Тези признаци са обособени в три групи: демографски, нормативистки и субективнооценъчни, като всички признаци са качествени подредими (т.е. равнищата на тези признаци може да бъдат подредени и да им се припишат числа, съответстващи на различията им). Скалата е седемстепенна и се използва методът на семантичния диференциал.

Целта, с която се прилага хи-квадрат методът, е да се изследват връзки между самите признаци, по които се оценяват идиолектите. Търси се отговор на въпроса съществуват ли такива връзки между тях, които да разрешат обединяване на признаци в макрофактори. Установено бе, че между признаците ‘правилна : неправилна реч’ и ‘престижна : непрестижна реч’ според хи-квадрат анализа съществува зависимост ($\alpha < 0.05$), а коефициентът на Крамер, който се движи между 0.44 и 0.61 при петте идиолекта, сочи, че връзката е средна по степен. При три от идиолектите между посочените два признака и признака ‘книжовна : диалектна реч’ съществува слаба връзка. Установена е и средна степен на сила на връзката между признака ‘книжовна : диалектна реч’ и признака ‘западнобългарска : източнобългарска реч’. Пак хи-квадрат методът показва, че има връзка и при петте идиолекта между признаците ‘градска : селска реч’ и ‘книжовна : диалектна реч’, а според коефициента на Крамер тя е средна по сила. Слаба по степен връзка имат признаците ‘градска : селска’ и ‘западнобългарска : източнобългарска реч’. При четири идиолекта с изключение на информа-

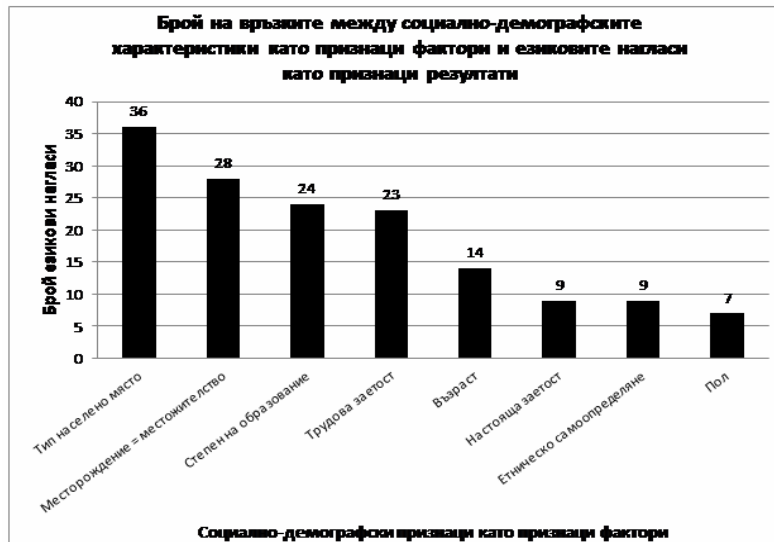
торката без отклонения от правоговора съществува средна сила на връзката между признаците 'столична : провинциална реч' и 'книжовна : диалектна реч'. Между признаците 'приятна : неприятна реч' (т.нар. естетически критерий) и 'благозвучна : неблагозвучна реч' (еуфоничния критерий) при всички говорителки е налице средна сила на връзката според критерия на Крамър (Алексова/Aleksova 2002).

Тези анализи на връзките дават основание да бъдат обединени признаците 'правилна : неправилна реч' и 'престижна : непрестижна реч' (средна по сила връзка) в един – 'правилна : неправилна реч', тъй като той е доминиращият. Признакът 'книжовна : диалектна реч' е във връзка (средна по сила), от една страна, с признака 'западнобългарска : източнобългарска реч', а от друга страна, с признака 'градска : селска реч'. Тези три признака могат да бъдат обединени в един, който да бъде дефиниран като 'книжовна : диалектна реч'. Естетическият признак ('приятна : неприятна реч') и еуфоничният ('благозвучна : неблагозвучна реч') също могат да бъдат обединени. Така от изходните 8 са получени 3 обобщени критерия.

Резултатите от хи-квадрат метода на изследване на връзки между оценъчни признаци, макар и приложен в непредставително изследване (Алексова/Aleksova 2002), дават основания да се подберат признаци, които са достатъчно обобщени според натрупания вече опит и могат да се използват при изследване на езикови нагласи към некодифицирани варианти в едно национално представително проучване (Алексова/Aleksova 2016). Именно това е вторият тип социалнопсихологически изследвания, прилагащи хи-квадрат метода, но не между оценъчни признаци, а между езикови нагласи към 12 некодифицирани варианта, оценявани по 3 критерия, и социално-демографските характеристики на респондентите. Трите оценъчни признака са 'правилно : неправилно', 'западнобългарско : източнобългарско явление', 'звучи добре : не звучи добре'. Използвана е десетстепенна скала. Оpozитивните стойности на всеки признак формират скала, чиито значения (подстойности) от 1 до 5 се движат в рамките на едната опозитивна величина (напр. правилност, западнобългарска локализация, добре звучащо явление), а от 6 до 10 в рамките на другата опозитивна величина (напр. неправилност, източнобългарска локализация, недобре звучащо явление).

За да бъде онагледена съпоставката на влиянието, което имат изследваните 8 социално-демографски фактора (в ролята им на признаци фактори) върху нагласите към 12-те правоговорни отклонения (в ролята им на признаци резултати), на фиг. 2 е нанесен броят на нагласите, които са в статистически значима връзка с всеки от осемте социално-демографски признака.

От фиг. 2 се вижда, че факторът 'тип населено място' (столица, общински град, малък град, село) има максимален брой статистически значими връзки с изследваните 36 типа езикови нагласи. След него се нарежда месторождението, като се оказва, че между него и актуалното местожителство



Фигура 2. Брой на връзките между социално-демографските характеристики на респондентите като признаци фактори и езиковите нагласи към 12-те правоговорни отклонения като признаци резултати

липсват различия в броя на връзките с езиковите нагласи и в начина, по който влияят двата фактора върху оценките за правоговорните отклонения. Степента на завършено образование и трудовата заетост са фактори, които са във връзка с 2/3 от 36-те езикови нагласи. За разлика от това възрастта определя малко над 1/3 от нагласите. Настоящата професионална заетост и етническото самоопределяне са във връзка с 9 от 36-те езикови нагласи. Полът е най-слабо влияещият върху изследваните 36 вида езикови нагласи.

В коментираното проучване хи-квадрат методът се използва и за изследване на връзки между идентификацията на три социално-демографски характеристики на говорещия (възраст, образование и материално положение) по едно правоговорно отклонение (последователно за т.нар. мякане, непалаталността на съгласната в сегашни глаголни форми и отмета на ударението в аористни форми) и социално-демографските характеристики на респондентите (Алексова/Aleksova 2016). Същият метод е използван и за изследване на връзки между самите езикови нагласи (във Втора част на Алексова/Aleksova 2016). Тъй като 12-те некодифицирани варианта на социолингвистични променливи се оценяват по три признака, се получават 36 вида нагласи. Тъй като нагласите се изследват по десетстепенна скала и представляват качествени подредими признаци, то е възможно да се изследва и посоката и силата на връзката. Резултатите от това приложение на хи-квадрат метода могат да бъдат видени у Алексова 2016.

VI. Приложение на факторния анализ по метода на главните компоненти

Факторният анализ по метода на главните компоненти се прилага, когато изследователят разполага с голям обем от данни и търси да открие на базата на взаимната корелация на признаци дали съществуват латентни (ненаблюдаеми) признаци, които са по-малко от изходните, но ги детерминират адекватно. Именно методът на главните компоненти (по-нататък МГК) е статистически евристичен метод, който се прилага за генериране на хипотези за наличие на причинно-следствени връзки. При МГК целта е да се намали размерността на пространството от променливи без съществена загуба на информация. В социолингвистиката той е прилаган в две проучвания – в изследване на речевата акомодация в столични семейства (Алексова/Aleksova 2000) и в анализа на резултатите от национално представителна анкета за изследване на езиковите нагласи към 12 некодифицирани варианта на социолингвистични променливи, наблюдавани в устната комуникация (Алексова/Aleksova 2016). Тук се представя резюмирано второто проучване.

Като изходни величини в изследването функционират 36 вида нагласи, които са получени при оценката на 12 ненормативни варианта в устната реч според 3 оценъчни признака. Имено тези 36 вида нагласи са наблюдаваните метрични величини x . Задачата е да се определят известен брой латентни (ненаблюдаеми) фактори F ($F_1, F_2, F_3, \dots, F_k$), които да са по-малко от първоначалните x и които да обясняват високата корелация между изходните променливи, обединени във формиращите се групи. От гледна точка на социалната психология на езика тази информация е нужна, за да разберем по какъв начин се групират нагласите, какъв признак лежи в основата на това групиране, без да може той да се наблюдава пряко.

Алгоритъмът на факторния анализ по метода на главните компоненти е следният: изчислява се корелационна матрица на изходните променливи и техните нива на значимост; прави се проверка на тестовете за адекватност на факторния анализ по МГК – спазено ли е изискването изходните променливи да имат случаен характер и да бъдат метрирани, а също да корелират помежду си в различна степен, изчислява се Бартлет текстът за сферичност и неговото ниво на значимост, което трябва да е по-малко от критичното ниво на риска 0.05, проверява се и индексът на Кайзер-Майер-Олкин (КМО), който трябва да бъде по-голям от 0.5 (вж. у Гочева-Илиева/Gocheva-Ilieva 2013: 294–306). Следващата стъпка е извличане на факторите F ($F_1, F_2, F_3, \dots, F_k$). Изчисляват се факторните тегла на всяка изходна величина x в избрания брой главни компоненти. „Факторният модел се определя най-вече от факторните тегла (loadings) [...], които представляват стандартизираните коефициенти на регресия на наблюдаваните променливи върху групата фактори. [...] Факторните тегла представляват същевременно и корелационните коефициенти между съответните наблюдавани променливи и фактори, понеже при метода на главните

компоненти факторите се избират с нулеви корелации (независими)“ (Цветков/Tsvetkov 2013). Вземат се предвид онези главни компоненти, които обясняват с натрупване значима част от общата дисперсия. Следващата стъпка е преизчисляване чрез ротация по метода Varimax, което улеснява интерпретацията на резултатите. Задача на изследователя е да проследи в ротираната матрица в кой от получените компоненти факторното тегло на всяка от изходните променливи е най-голямо. Изходната променлива участва в изграждането на онзи от главните компоненти, в който тя е с най-голямото си факторно тегло.

Чрез факторния анализ по метода на главните компоненти в изследването на нагласите към 12 некодифицирани варианта в устната реч, оценени по 3 признака, се оформиха 10 нови променливи. Първият главен компонент бе назван ‘най-неправилните речеве отклонения’ според оценките на респондентите, защото според високите си факторни тегла в него се групират изходни променливи, които са свързани с най-висок процент оценки за неправилност на 5 правоговорни отклонения: *Н’е, дит’ету ни спи; главъ/главѣтъ; тоа/таа/онаа; вървъ/вървѣтъ; песентъ/радосттѣ*. Вторият главен компонент се оформя от изходни променливи, представляващи оценки по скалата ‘звучи добре : не звучи добре’ на форми като *бел, бехме; вървъ, учителъ; тоа, таа, онаа*. Той е назван ‘неблагозвучни западнобългарски речеве явления’. Третият главен компонент е определен като ‘проблематична мезолектна (локална) маркираност’, защото е оформен от изходни променливи, представляващи оценки по скалата за мезолектна маркираност на форми от типа *четах/четахме, вървъ/вървѣтъ, учителъ/лекаръ, песентъ/радосттѣ, главъ/главѣтъ*. Обединяващото при тези явления е високият процент отговори „не знам“ на въпроса за идентификация на мезолектната принадлежност, разсейване на отговорите по скалата ‘западнобългарско – източнобългарско’ и липсата на категорична оценка за мезолектна маркираност. Четвърти главен компонент получава названието ‘модел на членуване по принципа на аналогията’, защото е оформен от изходни променливи, представляващи оценки по скалата ‘правилно : неправилно’ на форми като *учителъ/лекаръ*, както и оценки по скалите ‘правилност : неправилност’ и ‘звучи добре : не звучи добре’ на форми като *песентъ/радосттѣ*. Хипотезата за обединяването им в отделен главен компонент е свързана със съществуването на аналогия в менталната репрезентация на двете явления, която води до обединяването на двата речеве варианта в общ модел ‘съществително, завършващо с твърда съгласна в ед.ч., членувано с определителен член, който съдържа звук ъ, вкл. и под ударение’. Петият главен компонент е наречен ‘езикови тенденции’, защото в него влизат четири изходни променливи – оценките по скалите ‘правилно : неправилно’ и ‘звучи добре : не звучи добре’ на свърхякането и на т.нар. мекане. Това са двете речеве явления, които са получили сравнително най-малък процент оценки за неправилност – около 70% за свърхякането и 48.7% за мекането. Шестият главен компонент е назван ‘силно

маркирани западни речеви явления’, защото той е оформен от четири изходни променливи, представляващи оценки за мезолектната маркираност на: екането в глаголни и именни форми, показателни местоимения от типа *таа, тоа, онаа* и на аористни форми с крайно ударение (*правѝх, правѝл*). Седмият главен компонент носи названието ‘негативни нагласи към екането’, защото в него най-високи факторни тегла имат оценките по скалата ‘правилно : неправилно’ на екането в именни и в глаголни форми. Според защитаваното от авторката виждане стигматизирането на екането е оформен културен стереотип у нас. Осмият главен компонент може да се нарече ‘генерализиран формален аористен модел’, защото е оформен от две променливи с високи факторни тегла – оценките по скалите ‘правилно : неправилно’ и ‘звучи добре : не звучи добре’ на аористни форми от типа *четах, четахме*. Деветият главен компонент носи названието ‘некодифициран аористен акцентен модел’, защото се оформя от оценките по скалата ‘правилно : неправилно’ и съответно ‘звучи добре : не звучи добре’ на аористни форми с крайно ударение. Макар че не е разпространено само в западните говори, некодифицираното отмятане на ударението в аористни форми най-често се идентифицира при речева перцепция като показател за западнобългарска мезолектна реч. Десетият главен компонент е определен като ‘контрастни правоговорни отклонения’ – в него влизат нагласите към локалната маркираност на мекането (силно разколебана) и нагласите към териториалната ограниченост на мекостта на съгласните пред *е* и *и* + редуцията на неударена гласна *и*. Факторните тегла на двете променливи са с противоположни знаци, което отговаря на факта, че мекането е с най-малко негативни нагласи, а мекостта на съгласните пред *е* и *и* + редуцията на неударена гласна *и* са едни от най-високо негативно оценените. Тъй като първият компонент (‘най-неправилните речеви отклонения’) обяснява относително най-голяма част от наблюдаваната вариация и следователно е най-значим, може да се каже, че доминиращата скала, която формира езиковите нагласи, е именно ‘правилност : неправилност’. Важен латентен фактор, който обединява изходните променливи, е липсата на категорична локална маркираност на няколко речеви маркера (III главен компонент), а също така и факторът, който корелира с нагласите към правоговорни отклонения, които поради относително по-ниския процент на негативните оценки имат статута на езикови тенденции (V главен компонент). Извлечените 10 латентни фактора (скрити признаци) биха могли да послужат като основа на множество изследвания в областта на социолингвистиката, социалната психология на езика, прагматиката, устната комуникация и др. (вж. повече у Алексова/Aleksova 2016).

VII. Кратко заключение

Краткият обзор на прилаганите в българската социолингвистика и социална психология на езика статистически методи дава основание да се

каже, че тези методи за обработка на емпирични социолингвистични данни дават възможност за доказване на зависимости между социални и езикови променливи. Някои от тях, като напр. факторният анализ по метода на главните компоненти, имат евристичен характер – доказват зависимости, ненаблюдаеми чрез други методи. Статистическите методи несъмнено са мощно средство за доказване или отхвърляне на изследователски хипотези. Те са и средство за постигане на обективност на оценките, а също така дават възможност за възпроизвеждане, верифициране и съпоставимост на резултатите.

ЛИТЕРАТУРА / REFERENCES

Абазова 2014: *Абазова, Ст.* Езикови нагласи към жаргонизми, вулгаризми и професионална лексика. (Дисертация за присъждане на образователната и научна степен „доктор“ – ръкопис), София. [Abazova 2014: *Abazova, St.* Ezikovi naglasi kam zhargonizmi, vulgarizmi i profesionalna leksika. (Disertatsia za prisazhdane na obrazovatelnata i nauchna stepen „doktor“ – rakopis), Sofia.]

Алексова 2000: *Алексова, Кр.* Езикът и семейството. София, Интервю ПРЕС. [Aleksova 2000: *Aleksova, Kr.* Ezikat i semeystvoto. Sofia, Intervyu PRES.]

Алексова 2002: *Алексова, Кр.* Езикови атитюди, нормативни статуси и социолингвистични маркери. – В: Проблеми на социолингвистиката. Билингвизъм и диглосия – съвременни проблеми. Т. 7. София, Международно социолингвистическо дружество, 127–135. [Aleksova 2002: *Aleksova, Kr.* Ezikovi atityudi, normativni statusi i sotsiolingvistichni markeri. – In: Problemi na sotsiolingvistikata. Bilingvizam i diglosiya – savremenni problemi. T. 7. Sofia, Mezhdunarodno sotsiolingvisticheskoto druzhestvo, 127–135.]

Алексова 2016: *Алексова, Кр.* Социолингвистична перцепция, езикови нагласи и социална идентификация по речта. София, Парадигма. [Aleksova 2016: *Aleksova, Kr.* Sotsiolingvistichna pertsepsiya, ezikovi naglasi i sotsialna identifikatsiya po rechta. Sofia, Paradigma.]

Ангелов 1999: *Ангелов, А.* Правилата на езика в столичния квартал. София, Международно социолингвистическо дружество. [Angelov 1999: *Angelov, A.* Pravilata na ezika v stolichniya kvartal. Sofia, Mezhdunarodno sotsiolingvisticheskoto druzhestvo.]

Бокова 2000: *Бокова, П.* Лексикограматична интерференция в условията на българско-руското двуезичие (върху материал от с. Суворово (Шекерлик-Китай). (Автораферат за присъждане на научната и образователна степен „доктор“). София. [Bokova 2000: *Bokova, P.* Leksikogramatichna interferentsiya v usloviyata na balgarsko-ruskoto dvuezichie (varhu material ot s. Suvorovo (Shekerlik-Kitay). (Avtoreferat za prisazhdane na nauchnata i obrazovatelna stepen „doctor“). Sofia.]

Венедиков 1992: *Венедиков, Й.* Статистика, социология и още нещо... София, Информационно обслужване. [Venedikov 1992: *Venedikov, Y.* Statistika, sotsiologiya i oshte neshto... Sofia, Informatsionno obsluzhvanе.]

Венедиков 1993: *Венедиков, Й.* Обществено мнение. Епистемологични проблеми. София, Университетско издателство „Св. Климент Охридски“. [Venedikov 1993: *Venedikov, Y.* Obshtestvenoto mnenie. Epistemologichni problemi. Sofia, Universitetsko izdatelstvo „Sv. Kliment Ohridski“.]

Виденов 1998: *Виденов, М.* Социолингвистическият маркер. Към теорията и практиката на теренните изследвания. София, Делфи. [Videnov 1998: *Videnov, M.* Sotsiolingvisticheskiyat marker. Kam teoriyata i praktikata na terennite izsledvaniya. Sofia, Delfi.]

Гатев, Спасов, Радилев 1989: *Гатев, К., Ас. Спасов, Д. Радилев.* Обща теория на статистиката и икономическа статистика. София, Наука и изкуство. [Gatev, Spasov, Radilov 1989: *Gatev, K., As. Spasov, D. Radilov.* Obshta teoriya na statistikata i ikonomicheska statistika. Sofia, Nauka i izkustvo.]

Гоев 1996: *Гоев, В.* Статистическа обработка и анализ на информацията от социологически, маркетингови и политически изследвания със SPSS. София, Университетско издателство „Стопанство“. [Goev 1996: *Goev, V.* Statisticheska obrabotka i analiz na informatsiyata ot sotsiologicheski, marketingovi i politicheski izsledvaniya sas SPSS. Sofia, Universitetsko izdatelstvo „Stopanstvo“.]

Гочева-Илиева 2013: *Гочева-Илиева, Сн.* Приложна математика. [онлайн], Пловдив, Университетско издателство „Паисий Хилендарски“. [прегледан 01.11.2017]. <http://fmi.uni-plovdiv.bg/gocheva/PM-lekcii.pdf> [Gocheva-Ilieva 2013: *Gocheva-Ilieva, Sn.* Prilozhna matematika. [onlayn], Plovdiv, Universitetsko izdatelstvo „Paisiy Hilendarski“.]

Димитрова 2004: *Димитрова, Е.* Диглосията в гр. Криводол. София, Херон Прес. [Dimitrova 2004: *Dimitrova, E.* Diglosiyata v gr. Krivodol. Sofia, Heron Pres.]

Кънчева 2008: *Кънчева, П.* Софийският език на преселниците от Западните покрайнини. София, Академично издателство „Марин Дринов“. [Kancheva 2008: *Kancheva, P.* Sofiyskiyat ezik na preselnitsite ot Zapadnite pokraynini. Sofia, Akademichno izdatelstvo „Marin Drinov“.]

Мишев, Цветков 1998: *Мишев, Г., Ст. Цветков.* Статистика за икономисти. София, Университетско издателство „Стопанство“. [Mishev, Tsvetkov 1998: *Mishev, G., St. Tsvetkov.* Statistika za ikonomisti. Sofia, Universitetsko izdatelstvo „Stopanstvo“.]

Харалампиев 2007: *Харалампиев, К.* SPSS за напреднали. София, Университетско издателство „Св. Климент Охридски“. [Haralampiev 2007: *Haralampiev, K.* SPSS za naprednali. Sofia, Universitetsko izdatelstvo „Sv. Kliment Ohridski“.]

Цветков 2013: *Цветков, Д.* Лекции по статистика за психолози. § 10. Факторен анализ. [онлайн], [прегледан 01.11.2017]. <http://arsmath.org/msl/VTULlectures/PsiStatDPT/PSYSTAT-lecture10.htm> [Tsvetkov 2013: *Tsvetkov, D.* Lektsii po statistika za psiholoji. § 10. Faktoren analiz.]

Лабов 1966: *Лабов, W.* The Social Stratification of English in New York City. 1st edition 1966, Washington: Center for Applied Linguistics, 2nd edition 2006,

Cambridge, Cambridge University Press.

Labov 1972: *Labov, W.* Sociolinguistic Patterns. Oxford, Basil Blackwell.

Trudgill 1974: *Trudgill, P.* The Social Differentiation of English in Norwich (Cambridge Studies in Linguistics). Cambridge, Cambridge University Press.

✉ Проф. д-р *Красимира Алексова*, Катедра по български език
Факултет по славянски филологии, СУ „Св. Климент Охридски“
бул. „Цар Освободител“ 15, 1504 София, България

✉ *Prof. Krasimira Aleksova, PhD*, Department of Bulgarian Language
Faculty of Slavic Studies, Sofia University St. Kliment Ohridski
15 Tsar Osvoboditel, Sofia 1504, Bulgaria