

СИМЕОН СТЕФАНОВ

Институт за български език – БАН

✉ simeonste@yahoo.com

ЛЕКСИКАТА В *ŠIBENSKA MOLITVA*

Глотометричен анализ

Анотация

*Настоящото изследване е опит да се приложи методът на статистическата лингвистика ([7]: 42–54; [9]; [8]: 7–40; [1]: 28–43; [2]: 14–36; [3]: 72–86, и др.) върху материал от едно от най-ранните средновековни произведения в хърватската книжнина, познато като *Šibenska molitva*. Тъй като не ми е известно подобни проучвания да са правени върху други ранни хърватски произведения, основната задача е да се даде материал за сравнение с резултатите от бъдещи изследвания на други ранни творби.*

Ключови думи: средновековна хърватска книжнина, историческа лексикология на хърватския език, лингвостатистика

Предмет, цели, задачи и методи на изследване

Настоящото изследване е опит да се приложи методът на статистическата лингвистика ([1]: 28–43; [2]: 14–36; [3]: 72–86; [7]: 42–54; [8]: 7–40; [9] и др.) върху материал от едно от най-ранните средновековни произведения в хърватската книжнина, известно като „*Šibenska molitva*“ или „*Oratio pulcra (et) devota ad B(ea)tam V(irginem) Ma(r)ia(m)*“ (нататък също *ŠM*). Тъй като не ми е известно подобни проучвания да са правени върху други ранни хърватски произведения, основната задача е да се даде материал за сравнение с резултатите от бъдещи изследвания. Все пак за някои от разглежданите категории ще дам сведения за резултатите от работата ми върху други два паметника, писани на народен хърватски език и записани с латински букви. Това са най-ранният известен засега хърватски текст, писан на народен език и записан с латински букви, „*Red i zakon od primljenja na dil dobroga činjenja sestara naših reda svetog otca našega dominika*“ (нататък също *RZ*¹) от 1345 г. и житието на св. Екатерина „*Legenda sv. Katarine divice*“ (нататък също *LSK*²) от XIV–XV в. (по ръкопис I с 6 от Архива на ХАЗУ). За да прецизирам обобщенията и изводите си, се запознах и с резултатите от изследвания върху езика на

¹ За статистическите характеристики на този текст вж. [5].

² За статистическите характеристики на този текст вж. [4], [6].

произведения от старата българска литература, както и с известното за състоянието в съвременни български литературни жанрове. Сведения за състоянието в текстовете от старата и съвременната българска литература заемам от [3] и [9]. Трябва да се има предвид, че разглежданият тук паметник е превод (и то почти дословен) от латински език. Поради това неговите езикови особености са присъщи не само на хърватския текст, но и на латинската подложка. Тъй като съпоставката с латински източници не е част от задачите на настоящото изследване, а и понеже хърватският превод все пак е направен на народен език, смятам, че *ŠM* отразява до голяма степен особеностите на говоримия език от разглеждания период и поради това може да ни послужи като свидетелство и за неговото състояние.

За паметника

ŠM е един от най-ранните известни засега хърватски паметници, писани на народен език и записани с латински букви. Текстът е публикуван за пръв път през 1911 г. в *Starine* от Иван Милчетич и Йосо Милошевич ([12]: 89–190). До днес се пази в библиотеката на францисканския манастир в Шибеник, където е и открит в началото на XX в. от тогавашния игумен на манастира абат Йосо Милошевич. *ŠM* е ритмизиран поетичен текст, написан под формата на молитва към света Богородица. Езикът, на който е написан, е чакавски от Централна Далмация от областта около Шибеник. Датиран е приблизително от началото на XV в. В по-ново време *ŠM* е публикувана в транслитерация и транскрипция от Драгица Малич ([11]: 81–190), която прави и подробен езиков анализ на паметника. Текстът е поместен и в първи том на антологията „*Pet stoljeća hrvatske književnosti*“ ([10]: 528–531).

Лексиката в ŠM

Показатели на словника

Словникът на *ŠM* наброява 214 заглавни единици. От тях по честота на употреба в текста в групата от първата до петдесетата дума влизат 17 съществителни имена, 15 прилагателни имена, 4 глаголни форми или общо 36 лексеми пълнозначна лексика³ срещу 14 служебна лексика. Изчислени в проценти тези стойности показват, че 34% от петдесетте най-често употребявани думи в текста са съществителни имена. С общо 122 словоупотреби те заемат 18,8% от текста, а като 17 съществителни имена – 7,94% от целия словник. Прилагателните имена са 30% от групата. С 86 словоупотреби те заемат 13,25% от текста, а като 15 прилагателни имена – 7% от целия словник. Глаголните форми са 8%. С общо 56 словоупотреби те заемат 8,63% от текста, а като 4 глаголни форми – 1,87% от целия словник. Сред петдесетте най-често

³ В пълнозначната лексика в разглеждания текст влизат съществителни имена, прилагателни имена и глаголи, а в служебната – местоимения, местоименни наречия, съюзи, предлози, частици, междуметия.

употребявани думи в текста пълнозначната лексика е 72%. С 235 словоупотреби тя заема 36,21% от текста, а като 36 лексеми – 16,82% от целия словник. Непълнозначната лексика съставлява 28% от групата. С 224 словоупотреби тя заема 34,51% от текста, а като 14 лексеми – 6,54% от целия словник.

В групата от 51 до 100 най-чести думи в словника на *ŠM* влизат 24 съществителни имена, 11 прилагателни имена, 8 глаголни форми, или общо 43 лексеми пълнозначна и 7 – непълнозначна лексика. Изчислени в проценти, тези стойности показват, че 48% от думите в тази група са съществителни имена. С общо 33 словоупотреби те заемат 5,08% от текста, а като 24 съществителни имена – 11,21% от целия словник. Прилагателните имена са 22% от групата. С общо 16 словоупотреби те заемат 2,47% от текста, като 11 лексеми – 5,14% от целия словник. Глаголните форми са 16% от групата. С общо 14 словоупотреби те заемат 2,16% от текста, а като 8 лексеми – 3,74% от целия словник. Процентът на пълнозначната лексика в групата е 86%. С общо 63 словоупотреби тя заема 9,71% от текста, а като 43 лексеми пълнозначна лексика – 20,09% от целия словник. Непълнозначната лексика съставлява 14% от групата. С общо 11 словоупотреби тя заема 1,69% от текста, а като 7 лексеми – 3,27% от целия словник.

В групата от 101 до 150 по честота на употреба единица в словника влизат 28 съществителни имена, 5 прилагателни имена, 8 глаголни форми, или общо 41 единици пълнозначна лексика и 9 единици непълнозначна лексика. Изчислени в проценти, тези стойности показват, че 56% от тази група са съществителни имена. С общо 31 словоупотреби те заемат 4,78% от текста, а като 28 лексеми – 13,08% от целия словник. Прилагателните имена са 10% от групата. С общо 6 словоупотреби те заемат 0,92% от текста, а като 5 прилагателни имена – 2,34% от целия словник. Глаголните форми са 16% от групата. С общо 8 словоупотреби те заемат 1,23% от текста, а като 8 лексеми – 3,74% от целия словник. Пълнозначната лексика е 82% от групата. С общо 45 словоупотреби тя заема 6,93% от текста, а като 41 лексеми – 19,16% от целия словник. Непълнозначната лексика е 18% от групата. С общо 9 словоупотреби тя заема 1,37% от текста, а като 9 лексеми – 4,21% от целия словник.

В групата от 151 по честота думи до края на словника, която наброява 64 лексеми, влизат 32 съществителни имена, 12 прилагателни имена, 15 глаголни форми. Изчислени в проценти, тези стойности показват, че 50% от тази група са съществителни имена. С общо 58 словоупотреби те заемат 8,94% от текста, а като 32 лексеми – 14,95% от целия словник. Прилагателните имена са 18,75% от групата. С общо 12 словоупотреби те заемат 1,85% от текста, а като 12 лексеми – 5,61% от целия словник. Глаголните форми са 23,44% от групата. С общо 15 словоупотреби те заемат 2,31% от текста, а като 15 лексеми – 7,01% от целия словник. Общият брой на пълнозначната лексика в тази група е

60 лексеми. С 85 словоупотреби те заемат 13,1% от текста, а като 60 единици – 20,04% от целия словник. Непълнозначната лексика в групата е 4 лексеми. С общо 46 словоупотреби те заемат 7,09% от текста и 1,87% от целия словник. Процентът на пълнозначната лексика в тази група е 93,75%. Непълнозначната лексика съставлява 6,25% от групата.

Общото количество на съществителните имена в словника е 101 лексеми или 47,2%, на прилагателните имена – 43 лексеми или 20,09%, на глаголните форми – 35 лексеми или 16,36%. Общото количество на пълнозначната лексика е 182 лексеми или 83,65%, а на непълнозначната – 34 лексеми или 16,35%.

Въз основа на изнесените резултати, въпреки малкия обем на паметника, може да се забележи тенденцията за висок процент на служебната лексика сред най-често употребяваните думи в текста. Колкото повече намалява честотата на употребените единици в текста, намалява и количеството на служебната лексика, а нараства това на пълнозначната. Относително ниският процент на служебната лексика в словника показва, че голяма част от единиците тук са употребени по повече от веднъж в текста.

Отношението *покритие в текста – покритие в словника* показва, че колкото по-висока е честотата на употреба на лексемите, по-голямо е и покритието в текста, а по-малко – в словника. С намаляването на честотата на употреба намалява покритието на групата в текста, а нараства покритието в словника.

Показатели на отношението между словника и текста

Коефициент на лексикално многообразие

Важна категория при глотометричния анализ на един текст е и т.нар. „коефициент на лексикално многообразие“. Чрез него се оценява съотношението между употребените поне веднъж лексеми и всички употреби в текста. Той се изчислява, като разделим броя лексеми в словника на общия брой употреби и умножим полученото число по 100, за да го превърнем в проценти. По този начин се игнорират повторените единици, а се взимат пред вид само първите употреби на всяка лексема. В разглеждания текст коефициентът на лексикално многообразие е 32,97%. Това е доста висока стойност, която може да е повлияна и от малкия обем на паметника. За сведение коефициентът на лексикално многообразие в прозаичните части на *LSK* е 29,05%, а в *RZ* – 40,06 %.

Приведените резултати все пак показват известна стабилност за разглежданата категория в най-ранните хърватски текстове, писани на народен език и записани с латински букви, и са свидетелство за това, че макар и да не се опират на предходна книжовна традиция от подобен характер, хърватските книжовници все пак успяват да създадат текстове, които, дори и преводни, да отразяват особеностите и на говоримия език, но и да са лексикално разнообразни. Все пак това са различни по

обем текстове, така че приведените данни биха могли да служат само като ориентир и отправна точка при бъдещи съпоставки на равни части от тези и други подобни произведения.

Индекс на повторемостта

Друга съществена категория при плотометричния анализ на едно произведение е т.нар. „индекс на повторемостта“ ([1]: 38; [3]: 76; [7]: 42–43, 50–54, и др.). Той се пресмята, като разделим общия брой словоупотреби на броя единици в словника. Индексът на повторемостта в разглежданото произведение е 3,03. За сведение в прозаичните части на *LSK* той е 3,44, а в *RZ* – 2,5.

Това са относително ниски стойности на разглежданата категория в средновековната хърватска литература, писана на народен език с латински букви, което показва, че вероятно и тя е относително стабилна в текстовете на народен език от този период. Все пак ниските стойности може да се дължат и на малкия обем на разглежданите паметници⁴. И тук трябва да уточним, че това са различни по обем текстове, така че приведените данни биха могли да служат само като ориентир и отправна точка при бъдещи съпоставки на равни части от тези и други подобни произведения. Напълно очаквано е колкото по-ниска е стойността на индекса на повторемост, толкова по-голямо да е разнообразието на речника на дадено произведение.

Показатели на текста

Статистически изчисления

С понятието „глаголна температура“ се означава процентът на лични глаголни форми, употребени в даден текст или текстови отрязък. Терминът е въведен от Мирослав Янакиев (вж. [9]), където авторът уточнява, че по този начин студентите му назовават споменатата текстова категория (срв. [9]: 155, 185–190). По модела на този термин бихме могли да въведем (и изчислим) и „номиналната температура“ или процента на съществителните имена, „адективната температура“ или процент на прилагателните имена. Бихме могли да разграничим и процента на личните от този на неличните глаголни форми в общия брой единици от тази категория в даден текст или текстови отрязък.

⁴ В съвременната българска поезия, където този индекс е най-малък, той се движи около 3; в журналистиката – 4,1, а в научната литература – 6,2 ([1]: 38; [3]: 76). В съчинения от старобългарската литература този показател е: в съчиненията на Петър Черноризец – 5,15; в химнографията на Климент Охридски – 6,9; в словата на Климент Охридски е 7,07; в творчеството на Презвитер Козма – 7,48; в творчеството на Черноризец Храбър – 4,07; в „Повесть временных лет“ – 10 (срв. [3]: 76). Причината за толкова висок индекс на повтаремост изобщо в произведенията от старата (българска) литература е честата употреба на служебни думи и абстрактна лексика за фундаментални мирогледни понятия“ ([1]: 38; [3]: 76).

Въз основа на изследванията си върху стила на съвременни български автори М. Янакиев установява, че текст или текстов отрязък с относително висок процент лични глаголни форми (с висока „глаголна температура“) звучи по-живо, по-динамично, по-раздвижено, по-действено, а насищането на разказа със съществителни имена внася предметност ([9]: 155). За относително висок процент лични глаголни форми се смята резултат в рамките на 18–25% в една килолекса текст ([9]: 185–190). В цитираното изследване Янакиев уточнява, че в разговорната практика употребяваме повече лични, отколкото нелични глаголни форми, докато по-честата употреба на нелични глаголни форми е присъща на писаните текстове ([9]: 181–185). Това заключение е направено върху езика и стила на съвременни български автори, но понеже не ми е известен подобен труд в хърватската наука за литературата, обхващаща текстове от периода XIV–XV в., цитирам българския изследовател. Смятам, че наблюденията му биха могли да се окажат валидни и за средновековни (хърватски) текстове, писани на народен език.

Словникът на *ЅМ* наброява 214 лексеми в 649 словоупотреби. От тях 218 са съществителните имена в текста, 120 – прилагателните имена, 93 – глаголните форми. Общото количество на пълнозначната лексика в текста е 431, а на непълнозначната лексика – 248 единици. Това означава, че пълнозначната лексика съставлява 66,41% от текста, а непълнозначната – 33,59%. Съществителните имена, от своя страна, заемат 33,59% от текста, прилагателните – 18,49%, глаголните форми – 14,33%. От приведените резултати се вижда, че процентът на глаголните форми 14,33% е относително нисък. За сведение: в житието на св. Екатерина (*LSK*) е 23,39%, а в *RZ* – 18,18%. Процентът на съществителните имена в разглеждания текст е 33,49%. За сведение: в *LSK* той е 23,11%, а в *RZ* – 30,97%. Процентът на прилагателните имена в разглеждания текст е 18,49%. За сведение: в *LSK* той е 7,98%, а в *RZ* – 10,51%.

Тук отново ще припомня, че това са различни по обем текстове, така че приведените данни биха могли да служат само като ориентир и отправна точка при бъдещи съпоставки на равни части от тези и други подобни произведения. Все пак напълно е възможно някои от изчислените категории да се влияят и от жанровите специфики на произведенията.

Средна аритметична стойност и средно квадратично отклонение от нея за разглежданите категории в текста. Гаранция за 100% надеждност на ексцерптна оценка на средната аритметична стойност⁵.

Тук установявам какво е състоянието на разглежданите категории в *ЅМ* и по килолекси, и за целия текст. Тъй като обемът на паметника е 649 единици, ще приведа информацията за шестте килолекси, но ще използвам първо тази за първите три, а после за вторите три цели кило-

⁵ За изчислението на тези характеристики вж. ([9]: 191–200).

лекси, понеже идеята ми е след това да съпоставя получените резултати с резултатите от *RZ*, а неговият обемът е 349 единици или три цели килолекси и малко повече от половина. В бъдеще тези резултати ще могат да послужат за сравнения и съпоставки и с резултатите от други средновековни хърватски текстове, писани на народен език и записани с латински букви.

В първата килолекса *ŠM* са употребени 39 съществителни имена или 39% от ексцерпта, 22 прилагателни имена или 22% от ексцерпта, 9 глаголни форми или 9% от ексцерпта, общо 70 думи пълнозначна лексика или 70% от ексцерпта и 30 непълнозначна лексика или 30% от ексцерпта.

Във втората килолекса се срещат 39 съществителни имена (39% от ексцерпта), 17 прилагателни имена (17% от ексцерпта), 9 глаголни форми (9% от ексцерпта), общо 65 думи пълнозначна лексика (65% от ексцерпта) и 35 – непълнозначна лексика (35% от ексцерпта).

В третата килолекса от *ŠM* се срещат 40 съществителни имена (40% от ексцерпта), 22 прилагателни имена (22% от ексцерпта), 9 глаголни форми (9% от ексцерпта), общо 71 единици пълнозначна лексика (71% от ексцерпта) и 29 единици непълнозначна лексика (29% от ексцерпта).

В четвъртата килолекса от *ŠM* са употребени 29 съществителни имена (29% от ексцерпта), 23 прилагателни имена (23% от ексцерпта) 10 глаголни форми (10% от ексцерпта), общо 62 думи пълнозначна лексика (62% от ексцерпта) и 38 – непълнозначна лексика (38% от ексцерпта).

В петата килолекса се срещат 31 съществителни имена (31% от ексцерпта), 19 прилагателни имена (19% от ексцерпта), 13 глаголни форми (13% от ексцерпта), общо 63 думи пълнозначна лексика (63% от ексцерпта) и 37 – непълнозначна лексика (37% от ексцерпта).

В шестата килолекса от *ŠM* се срещат 23 съществителни имена (23% от ексцерпта), 21 прилагателни имена (21% от ексцерпта), 23 глаголни форми (23% от ексцерпта), общо 67 единици пълнозначна лексика (67% от ексцерпта) и 33 единици непълнозначна лексика (33% от ексцерпта).

В останалите 49 единици текст се срещат 14 съществителни имена (28,57% от ексцерпта), 7 прилагателни имена (14,29% от ексцерпта), 9 глаголни форми (18,37% от ексцерпта) или общо 30 единици пълнозначна лексика (61,22% от ексцерпта) и 19 – непълнозначна лексика (38,76% от ексцерпта).

Броят на глаголните форми в първите три килолекси от *ŠM* е съответно 9, 9, 9. Броят на съществителните имена – 39, 39, 40, а броят на прилагателните имена – 22, 17, 22. Средното аритметично количество глаголни форми за трите килолекси текст е 9. Оттук, като извадим от средното аритметично реалното количество глаголни форми за всяка килолекса, получените резултати за всеки един ексцерпт вдигнем на квадрат и съ-

берем стойностите помежду им, след това разделим полученото число на две, понеже работим с три килолекси, а накрая сложим под квадратен корен резултата, получаваме 0. Това е средното квадратично отклонение на глаголните форми за всеки ексцерпт в първите три килолекси текст. То ни показва с колко в една килолекса текст техният брой се отклонява от средната стойност за тази категория в n на брой (в нашия случай 3) килолекси текст. С негова помощ бихме могли да изчислим интервала на средната стойност на употребените глаголни форми в първите три килолекси от \check{M} . В нашия случай за глаголните форми той би трябвало да е 9–9, понеже стойностите в трите килолекси са еднакви и са равни на средната стойност. Отново с помощта на средното квадратично отклонение можем да изчислим гаранцията за 100% надеждност на ексцерптна оценка на стойността на тази категория в първите три килолекси от \check{M} . Тя се получава, като умножим по три средното квадратично отклонение.

Средното аритметично на употребените съществителни имена в първите три килолекси текст от \check{M} е 39,33. Средното квадратично отклонение за тази категория в текстовите ексцерпти е 0,58. Интервалът на средната стойност на употребените съществителни имена за първите три килолекси е 38,75–39,91. Гаранцията за 100% надеждност на ексцерптна оценка на стойността на съществителните имена отново само за първите три килолекси текст е 1,74.

Средното аритметично за прилагателните имена в първите три килолекси е 20,33. Средното квадратично отклонение за тази категория е 2,89. Интервалът на средната стойност на употребените прилагателни имена е 17,44–23,22. Гаранцията за 100% надеждност на ексцерптна оценка на стойността на прилагателните имена в първите три килолекси е 8,66.

Средното аритметично на употребените съществителни имена във вторите три килолекси текст от \check{M} е 27,67. Средното квадратично отклонение за тази категория в текстовите ексцерпти е 4,16. Интервалът на средната стойност на употребените съществителни имена във вторите три килолекси от \check{M} е 23,51–31,83. Гаранцията за 100% надеждност на ексцерптна оценка на стойността на съществителните имена във вторите три килолекси текст е 12,48.

Средното аритметично на употребените прилагателни имена във вторите три килолекси текст от \check{M} е 21. Средното квадратично отклонение за тази категория в текстовите ексцерпти е 2. Интервалът на средната стойност на употребените прилагателни имена във вторите три килолекси от \check{M} е 19–23. Гаранцията за 100% надеждност на ексцерптна оценка на стойността на прилагателните имена във вторите три килолекси текст е 6.

Средното аритметично на употребените глаголни форми във вторите три килолекси текст от \check{M} е 15,33. Средното квадратично отклонение за тази категория в текстовите ексцерпти е 6,81. Интервалът на

средната стойност на употребените глаголни форми във вторите три килолекси от *ŠM* е 8,52–22,14. Гаранцията за 100% надеждност на ексерпнтната оценка на стойността на глаголните форми във вторите три килолекси текст е 20,42.

Средното аритметично на употребените съществителни имена за целия текст е 33,5. Средното квадратично отклонение за тази категория в текста е 6,92. Интервалът на средната стойност на употребените съществителни имена в текста е 26,58–40,42. Гаранцията за 100% надеждност на ексерпнтната оценка на стойността на съществителните имена в текста е 20,76.

Средното аритметично на употребените прилагателни имена за целия текст е 20,67. Средното квадратично отклонение за тази категория в текста е 2,25. Интервалът на средната стойност на употребените прилагателни имена в текста е 18,42–22,92. Гаранцията за 100% надеждност на ексерпнтната оценка на стойността на прилагателните имена в текста е 6,75.

Средното аритметично на употребените глаголни форми за целия текст е 12,17. Средното квадратично отклонение за тази категория в текста е 5,53. Интервалът на средната стойност на употребените глаголни форми в текста е 6,64–17,7. Гаранцията за 100% надеждност на ексерпнтната оценка на стойността на глаголните форми в текста е 16,59.

Критерий на Пиърсън⁶

Сега вече бихме могли да изчислим и потенциално колко съществителни имена, прилагателни имена, глаголни форми, пълнозначна и непълнозначна лексика бихме могли да очакваме в останалите 51 единици текст от *ŠM* от последната седма килолекса текст. Така ще получим още една цяла хипотетична седма килолекса. Текстовата корекция до края на текста е 51 единици. За да получим търсеното от нас количество на употребени единици в текстовата корекция, ще разделим гаранцията за 100% надеждност на изчисление на стойността на съответната категория за целия текст на 51 и ще умножим по гаранцията за 100% надеждност на изчисление на съответната категория. След направените изчисления получаваме, че очакваният брой съществителни имена в текстовата корекция е $(20,76/51) \times 20,76 = 8,45 = 8,5 = 9$. Очакваният брой прилагателни имена в текстовата корекция е $(6,75/51) \times 6,75 = 0,89 = 1$. Очакваният брой глаголни форми е $(16,59/51) \times 16,59 = 5,4 = 5$.

Общото количество съществителни имена в хипотетичната цяла седма килолекса *ŠM* е $14 + 9 = 23$ (23%). Общото количество прилагателни имена в хипотетичната цяла седма килолекса *ŠM* е $7 + 1 = 8$ (8%). Общото количество глаголни форми в хипотетичната цяла седма кило-

⁶ В статистическата лингвистика тази процедура се използва за оценка на отношението между теорията и практиката при изчисляване на дадени стойности, вж. ([9]: 220–222).

лекса $\check{S}M$ е $9 + 5 = 14$ (14%). Общото количество пълнозначна лексика в хипотетичната цяла седма килолекса $\check{S}M$ е $23 + 8 + 14 = 45$ (45%), а на непълнозначната – $19 + 36 = 55$ (55%).

Критерий на Колмогоров-Смирнов⁷

В статистическата лингвистика тази процедура се използва за установяване на отношението между стойностите на избрани категории в два вариационни реда от по равен брой килолекси при съпоставка на равни по обем ексцерпти. Тук я прилагам, за да установя отношението между два вариационни реда от по три килолекси глаголна, номинална и адективна температура в $\check{S}M$ и RZ . С Δmax . се отбелязва максималната разлика между стойностите за съответната категория по килолекси в двата вариационни реда. С помощта на λ^2 можем да заключим съществена ли е или не разликата между средните стойности за дадена категория в два вариационни реда от по равен брой килолекси. Изчислява се, като разделим $\Delta^2 max$. на броя килолекси в редицата, умножени по две, или $\lambda = \Delta^2 max./2n$. В лингвостатистиката е установено, че ако λ е по-голямо от 1,84, то е възможно двата съпоставяни ексцерпта да са от различна генерална съвкупност, а ако е по-голямо и от 2,65, то двата съпоставяни ексцерпта със сигурност са от различна генерална съвкупност, или казано по друг начин, принадлежат на различни по тип текстове. Тъй като RZ ни дава три цели килолекси текст, а $\check{S}M$ – 6, ще съпоставя трите килолекси RZ първо с първите три килолекси $\check{S}M$, а после и с вторите.

Във вариационната редица от глаголни температури в общо шест килолекси текст, по три от $\check{S}M$ и RZ , $\Delta max.=3$, а $\lambda=3^2/2 \times 3=9/6=1,5 < 1,84 < 2,65$. Следователно би трябвало двата вариационни реда да са от една генерална съвкупност.

Във вариационната редица от адективни температури в общо шест килолекси текст, по три от $\check{S}M$ и RZ , $\Delta max.=3$, а $\lambda=3^2/2 \times 5=9/10=0,9 < 1,84 < 2,65$. Следователно би трябвало двата вариационни реда да са от една генерална съвкупност.

Във вариационната редица от номинални температури в общо шест килолекси текст, по три от $\check{S}M$ и RZ , $\Delta max.=3$, а $\lambda=3^2/2 \times 4=9/8=1,125 < 1,84 < 2,65$. Следователно би трябвало двата вариационни реда да са от една генерална съвкупност.

Във вариационна редица от глаголни температури в общо шест килолекси текст (трите RZ и вторите три $\check{S}M$) $\Delta max.=1$, а $\lambda=1^2/2 \times 4=1/8=0,125 < 1,84 < 2,65$. Следователно би трябвало двата вариационни реда да са от една генерална съвкупност.

Във вариационната редица от номинални температури в общо шест килолекси текст (трите RZ и вторите три $\check{S}M$) $\Delta max.=3$, а

⁷ По-подробно за Критерия Колмогоров-Смирнов вж. [9]: 202–206.

$\lambda=3^2/2 \times 5=9/10=0,9 < 1,84 < 2,65$. Следователно би трябвало двата вариационни реда да са от една генерална съвкупност.

Във вариационната редица от адекватни температури в общо шест килолекси текст (трите *RZ* и вторите три *ŠM*) $\Delta \max.=3$, а $\lambda=3^2/2 \times 6=9/12=0,75 < 1,84 < 2,65$. Следователно би трябвало двата вариационни реда да са от една генерална съвкупност.

Според направените изчисления по Критерия на Колмогоров-Смирнов двата съпоставяни текста би трябвало да принадлежат с всички свои части към една генерална съвкупност.

*Критерий на Фишер*⁸

Когато съпоставяме ексцерпт, състоящ се от проби, които имат статистически инвариант, с друг ексцерпт, състоящ се от проби, които също имат статистически инвариант, но друг, е възможно критерият λ да не регистрира различие, защото е възможно двата ексцерпта да се различават не по средните си аритметични, а по показателите си за разнообразието на вариантите, от които се състоят вариационните редове, т.е. по сигмите си (или по средните квадратични). За да се регистрира това различие, съществува критерий, който се нарича Критерий на Фишер. Така, ако сме изчислили средните квадратични отклонения на два вариационни реда от една езикова категория и разделим по-голямото на по-малкото, по големината на получената стойност можем да съдим дали двата вариационни реда действително се различават статистически. Ако полученото число е по-малко от 1,69, трябва да се признае наличието на един статистически инвариант за пробите от двата ексцерпта. Ако е по-голямо от 2,11, трябва да се приеме, че разликата между дисперсиите е достатъчно ярка и можем да разглеждаме ексцерптите като принадлежащи на различни от плотометрическо гледище генерални съвкупности (на различни по жанр текстове).

Тук отново правя съпоставка между резултатите от изчислението на вече познатите категории в трите килолекси „*Red i zakon*“ и двете редици от по три килолекси *ŠM*. Средното квадратично отклонение за вариационните редове глаголни форми, съществителни имена и прилагателни имена в „*Red i zakon*“ е 4,04; 0,58; 1,53. Средното квадратично отклонение за вариационните редове глаголни форми, съществителни имена и прилагателни имена в първите три килолекси от *ŠM* е 0; 1,73; 2,89. Стойността на Критерия на Фишер за глаголните форми в съпоставяните ексцерпти е $4,04/0=4,04 > 2,11 > 1,69$, т.е. двата съпоставяни ексцерпта нямат общ статистически инвариант. Стойността на Критерия на Фишер за съществителните имена в съпоставяните ексцерпти е $1,74/0,58=3 > 2,11 > 1,69$, т.е. двата съпоставяни ексцерпта нямат общ статистически инвариант. Стойността на Критерия на Фишер за прилагателните имена в съпоставяните ексцерпти е $2,89/1,53=1,89 > 1,69$, т.е. в „границите на съмнението“.

⁸ По-подробно за Критерия на Фишер вж. [9]: 210–212.

Средното квадратично отклонение за вариационните редове глаголни форми, съществителни имена и прилагателни имена във вторите три килолекси \bar{SM} е 6,81; 4,16; 2. Стойността на Критерия на Фишер за глаголните форми в съпоставяните ексцерпти е $6,81/4,04=1,69=1,69$, т.е. в „границите на съмнението“. Стойността на Критерия на Фишер за съществителните имена в съпоставяните ексцерпти е $4,16/0,58=7,17>2,11>1,69$, т.е. двата съпоставяни ексцерпта нямат общ статистически инвариант. Стойността на Критерия на Фишер за прилагателните имена в съпоставяните ексцерпти е $2/1,53=1,31<1,69<2,11$, т.е. би трябвало двата съпоставяни ексцерпта да имат общ статистически инвариант.

Според направените изчисления по Критерия на Фишер в по-голямата част от ексцерптите двата съпоставяни текста не би трябвало да имат общ статистически инвариант.

Представените данни от плотометричния анализ на „Šibenska molitva“ позволяват да се направят следните обобщения и изводи:

Показатели на словника

Въз основа на изнесените резултати, въпреки малкия обем на паметника, може да се забележи тенденцията да е висок процентът на служебната лексика сред най-често употребяваните думи в текста. Колкото повече намалява честотата на употребените единици в текста, намалява и количеството на служебната лексика, а нараства това на пълнозначната. Относително ниският процент на служебната лексика в словника показва, че голяма част от единиците тук са употребени по повече от веднъж в текста.

Отношението *покрытие в текста – покрытие в словника* показва, че колкото по-висока е честотата на употреба на лексемите, по-голямо е и покритието в текста, а по-малко – в словника. С намаляването на честотата на употреба намалява покритието на групата в текста, а нараства покритието в словника.

Показатели на отношението между словника и текста

Коефициент на лексикално многообразие

Приведените резултати все пак показват известна стабилност за разглежданата категория в най-ранните хърватски текстове, писани на народен език и записани с латински букви, и са свидетелство за това, че макар и да не се опират на предходна книжовна традиция от подобен характер, хърватските книжовници все пак успяват да създадат текстове, които, дори и преводни, да отразяват особеностите и на говоримия език, но и да са лексикално разнообразни. Все пак това са различни по обем текстове, така че приведените данни биха могли да служат само като ориентир и отправна точка при бъдещи съпоставки на равни части от тези и други подобни произведения.

Индекс на повторяемостта

Сведенията за резултатите от други средновековни хърватски текстове, писани на народен език с латински букви, показват относително ниски стойности на разглежданата категория в ранната хърватска литература, което означава, че вероятно и тя е относително стабилна в текстовете на народен език от този период. Все пак ниските стойности може да се дължат и на малкия обем на разглежданите паметници. И тук трябва да уточним, че това са различни по обем текстове, така че приведените данни биха могли да служат само като ориентир и отправна точка при бъдещи съпоставки на равни части от тези и други подобни произведения. Напълно очаквано е колкото по-ниска е стойността на индекса на повторяемост, толкова по-голямо да е разнообразието на речника на дадено произведение.

Показатели на текста

Статистически изчисления

Това са различни по обем текстове, така че приведените данни биха могли да служат само като ориентир и отправна точка при бъдещи съпоставки на равни части от тези и други подобни произведения. Все пак напълно е възможно някои от изчислените категории да се влияят и от жанровите специфики на произведенията.

Съпоставка на ексцерпти от два текста

Според направените изчисления по Критерия Колмогоров-Смирнов би трябвало двата съпоставяни текста да принадлежат с всички свои части към една генерална съвкупност. Според направените изчисления по Критерия на Фишер в по-голямата част от ексцерптите двата съпоставяни текста не би трябвало да имат общ статистически инвариант (три пъти нямат, два пъти са „в границите на съмнението“, веднъж би трябвало да имат). Получените резултати не дават еднозначен отговор. Тъй като в лингвостатистиката се съпоставят на първо място ексцерпти, а чак след това чрез тях и текстовете, на които те принадлежат, трябва да уточним, че направените обобщения в съответните раздели важат преди всичко за ексцерптите. Понеже резултатите от двата Критерия за съпоставка се различават, а Критерият на Фишер все пак отчита с по-голяма точност съпоставяните характеристики, склонен съм да смятам, че двата текста по-скоро нямат общ статистически инвариант. Подобно твърдение звучи по-логично, понеже е напълно нормално различните текстови жанрове да си служат с наличните езикови средства по различен начин, както и да ги използват в различна степен.

Литература

- [1] Давидов, А. Използване на статистическите методи в лексиколожките изследвания. В: Давидов, А. (Ред.) *Старобългарска лексикология*. Велико Търново, УИ „Св. Климент Охридски“, 1996, с. 28–43.

- [2] Илиева, Т. Лексиката в Йоан-Екзарховия превод на Богословието – количествени показатели. // *Старобългаристика*, XXXV, 2011, № 2, с. 14–37.
- [3] Илиева, Т. Лексиката в творчеството на Петър Черноризец – глотометричен профил. // *Език и литература*, 2013, № 1–2, с. 72–86.
- [4] Стефанов, С. Лексика и стил в средновековно житие на св. Екатерина на народен хърватски език. // *Littera et Lingua, Пролет–Лято, 2015*. – <http://slav.unisofia.bg/naum/lilijournal/2015/12/1-2/>
- [5] Стефанов, С. Лексиката в „Red i zakon zadarskih dominikanki“: глотометричен анализ. В: *Стереотипът в славянските езици, литератури и култури* (Сборник от XVI Международни Славистични четения – 26–28.04.2018 г.). София, 2018.
- [6] Лексиката в „Legenda sv. Katarine divice“: глотометричен анализ. // *Дзяло*, 2018.
- [7] Фрумкина, Р. М. *Статистические методы изучения лексики*. Москва, „Наука“, 1964.
- [8] Христова, И. *Речник на словата на Климент Охридски*. София, УИ „Св. Климент Охридски“, 1994, с. 7–40.
- [9] Янакиев, М. *Стилистиката и езиковото обучение*. София, „Народна просвета“, 1977.
- [10] Damjanović, K., Mihaljević, Ž. *Red i zakon zadarskih dominikanki iz 1345. Godine*. In: *Povijest hrvatskoga jezika. Knj. 1: Srednji vijek*. Zagreb, Društvo za promicanje hrvatske kulture i znanosti „Croatica“, 2009, s. 530–531.
- [11] Malić, D. Šibenska molitva. // *Rasprave Instituta za jezik*, 1973, № 3, Zagreb, s. 81–190.
- [12] Milčetić, I., Milošević, J. Šibenska molitva (14. vijek). // *Starine*, 1911, № 33, s. 81–190.