

Tatyana Ilieva

(Sofia, Cyrillo-Methodian Research Centre, Bulgarian Academy of Sciences)

Statistical Methods in Old Bulgarian Lexicography

IN DEDICATION TO THE 90TH ANNIVERSARY
OF PROF. MIROSLAV YANAKIEV WHO PIONEERED THE APPLICATION OF STATISTICAL
MATHEMATICS METHODS TO BULGARIAN
LINGUISTICS

Abstract: The paper discusses the current state of glottometric research and the registration in dictionaries of quantitative features of the lexis in the study of historical Slavic lexicology and lexicography. The author argues in favour of supplying every written monument of the Middle Ages with frequency information and outlines the prospects for this branch of Slavic Medieval studies which will benefit not only applied areas, such as teaching Old Bulgarian and Church Slavonic, but also theoretical research in the field of Paleo-Slavic studies.

Татяна Илиева

(София, БАН, КМНЦ)

Статистическата методика и старобългарската лексикография

ПОСВЕЩАВА СЕ НА 90 ГОДИНИ ОТ
РОЖДЕНИЕТО НА ПРОФ. МИРОСЛАВ ЯНАКИЕВ,
КОЙТО ПРЪВ ПРИЛАГА МЕТОДИТЕ НА
МАТЕМАТИЧЕСКАТА СТАТИСТИКА В
БЪЛГАРСКОТО ЕЗИКОЗНАНИЕ

Статистическата характеристика е важен аспект на проучване лексиката, в това число и на средновековните паметници¹. В чуждестранната научна литература вече има добре разработена методология за изучаване количествените характеристики на текста². От изследванията с отношение към историческата лексикология на славянските

¹ Цейтлин, Р.М. Количественная характеристика словарного состава древнеболгарских рукописей. // Лексика древнеболгарских рукописей X-XI вв., С. 1986, с. 34-35.

² Yule, G.U. The statistical study of literary vocabulary. Cambridge, 1944, p. 78; Zipf, G. K. Human behavior and the principle of least effort. Cambridge (Mass.), 1949; Фрумкина, Р. М. Статистические ме-

езици ще спомена работите на Л. В. Вялкина и Г. Н. Лукина, А. Давидов, Н. Романков, О. Творогов³. Статистически сведения за лексикалния състав на старобългарските писмени паметници има още в трудовете на А. Дорич, А. Достал⁴, а в по-ново време на Р. М. Цейтлин, Р. Станков, В. Ефимова, Л. Тасева и други изследователи⁵. Проблемите на глотометричния анализ на старобългарската книжнина обаче все още чакат своето монографично изследване. На същото равнище – частична разработка, е и лексикографското описание на количествената характеристика на старобългарската лексика. Освен числови данни в словарните статии на редица старобългарски индекси и речници са съставени ранглисти⁶ само към отделни текстове и общи речници⁷. Опит за последова-

тоды изучения лексики. Москва, Наука, 1964, с. 42-43; Янакиев, М. Стилистиката и езиковото обучение. С., 1977, 194-200.

³ **Вялкина** Л. В., Г. Н. **Лукина**. Опыт применения некоторых методов математической статистики к изучению древнерусских текстов. // Исследования по исторической лексикологии древнерусского языка, М., 1964; **Давидов**, А. Статистические наблюдения над "Беседой" Козмы Пресвитера. // Вопросы словообразования и лексикологии древнерусского языка. Издательство "Наука", М. 1974, 285-304; **Творогов**, О. Материалы к частотному словарю среднеболгарского перевода „Хроники Константина Манасии“ // *Palaeobulgarica*, 7, 1983, 1, 75-81; **Романков**, Н. В. Формальные признаки авторского стиля Климента Охридского и Житие Константина-Кирилла. // Советское славяноведение, 1986, №2, 89-96; **Христова**, И. Речник на словата на Климент Охридски. Университетско издателство "Климент Охридски", София, 1994, с. 7-40; **Давидов**, А. Използване на статистическите наблюдения в лексиколожките изследвания. // Старобългарска лексикология. В.Т. 1996, 28-4; **Radovich, N.** Analisi insiemistica del lessico slavoecclesiastico antico, Istituto di Filologia Slava, Universita di Padova. 1974, 213.

⁴ **Doritsch**, A. Gebrauch der altbulgarischen Adverbia. Leipzig, 1910; **Dostál**, A. Studie o víidovém systému v staroslověštině. Praha, 1954.

⁵ **Цейтлин**, Р.М. Цит съч.; **Станков**, Р. Лексика Исторической палеи. В. Т., 1994 г., 311 с.; **Ефимова**, В. С. Старославянская словообразовательная морфемика. Москва, 2006, 366 с.; **Тасева**, Л. Анонимный сербский переводчик XIV века и его современники (Лексикална иновативност как признак индивидуального стиля). – In: *Iter philologicum. Festschrift für Helmut Keipert zum 65. Geburtstag.* Hrsg. von D. Bunčić und N. Trunte. (= Die Welt der Slaven, Sammelbände/Сборники, Band 28). München, 2006, 209-221; **Taseva, L.** Wortbildungstendenzen in den südslavischen Versionen der Synaxarien zum Triodion. – In: *Darъ slovesny. Festschrift für Christoph Koch zum 65. Geburtstag.* Hrsg. von Wolfgang Hock und Michael Meier-Brügger. München, 2007, 277-287; **Тасева**, Л. Езикът на някои оригинални химнографски творби от IX-X век в статистически ракурс. – В: Сборник на трудови од Меѓународниот научен собир „Свети Наум Охридски и словенската духовна, културна и писмена традиција“ (организиран по повод на 1100-годишнината от смртта на св. Наум Охридски). Охрид, 4–7 ноември 2010. Скопје, 2011, 175-183.

⁶ В този вид лексикографски пособия думите са разположени не по азбучен ред, а в зависимост от това колко често се срещат в текста. В списъка има групи думи с еднаква честота – в такъв случай те се разполагат по азбучен ред.

⁷ **Вялкина**, Л. В., Г. Н. **Лукина**. Материалы к частотному словарю древнерусских текстов. // Лексикология и словообразование древнерусского языка, М., 1966, 263-292; **Давидов**, А. Честотен речник на „Беседа против богомилите“ от Презвитер Козма. // Трудове на ВТУ „Кирил и Методий“. Т. XV, кн. 2. С., 1980, 197-219; **Творогов**, О. Лексический состав „Повести временных лет“ (Словоуказатели и частотный словарь). Киев, 1984; **Христова**, И. Цит. Съч., с. 287-319; **Ribarova, Z.** Indexu k Staroslověšnému slovníku/Indexes to the Old Church Slavonic Dictionary/Словоуказатели к Старословянскому словарю. Praha 2003; **Илиева**, Т. Лексиката в химнографското творчество на Климент Охридски – глотометричен профил. Доклад от Националната конференция по случай 1100 години от смъртта на Наум Охридски, проведена в София на 24 и 25 май 2010 г.; **Илиева**, Т. Лексиката в Йоан-Екзарховия превод на Богословието – количествени показатели // *Palaeobulgarica*, XXXV(2011), 2, 14-36. **Dietze, J.** Frequenzwörterbuch zur Synodalhandschrift der Ersten Novgoroder Chronik. Halle (Saale), 1977; **Dietze, J.**

телно прилагане на статистическия подход към проучване лексиката на отделен старобългарски текст едновременно във всички нейни най-важни аспекти е направен в подготвяните за печат в КМНЦ словоуказатели към Книга на пророк Иезекиил по ръкопис F.I.461 от РНБ. Приложени към отделните дялове на лексикографския комплекс, съставен от старобългарско-гръцки словоуказател, гръцко-старобългарски показалец, обратен индекс и рангов списък, количествените показатели разкриват точни и сигурни съотношения на всички равнища на представяне речника на ИезF.I.461. Например, приведени в обратния индекс, те указват сумата на словоупотребите в отделните лексико-семантични групи спрямо дължината на техния списък и позволяват съпоставимост на отделните членове в групата и между групите по показателя фреквентност. Нововъведение в статистическите данни на двуезичен индекс представлява практиката на същите словоуказатели в справочния дял на речниковите статии от старобългарско-гръцкия словоуказател към текста след всяко гръцко съответствие на заглавната единица на първо място да се отбелязва фреквенцията на дадената дума *x* от гръцки език като съответствие на лемата *y* от старобългарски език и след знак / тази на всички словоупотребии на същата дума *x* в текста на подложката. Аналогично се процедира и в гръцко-старобългарския показалец. Това позволява проследяване на процентните отношения между превеждащи и превеждани единици в текста на подложката и превода⁸.

Въпреки направеното през последните десетилетия, броят на този тип лексикографски помагала е крайно недостатъчен. В качеството им на необходим инструмент при проучване глотометричните параметри на текста (статистическа структура на речника, пропорция между количеството лексеми и словоупотребии, дял на единичните употреби, индекс на повторемост, съотношенията между групите думи с определена честота, разпределението им по части на речта, процента на чужда и домашна лексика и пр.), такива фреквенции трябва да се изготвят към всеки лексикографиран паметник предвид факта, че обективността на статистическия метод и честотната оценка нарастват в зависимост от обема и разнообразието на анализирани текстове. В бъдеще един своден честотен речник, обединяващ ранглистите към отделни ръкописи и паметници, ще позволи съпоставителни наблюдения:

Frequenzwörterbuch zur jüngeren Redaction der der Ersten Novgoroder Chronik. München, 1984; **Dietze, J.** Frequenzwörterbuch zur vierten Novgoroder Chronik. Halle (Saale) 1984.

⁸**Илиева, Т.** Словоуказатели към Книга на пророк Иезекиил по ръкопис F.I.461 (РНБ). С., 2009 /хабилитация/.

- на съвременни помежду си текстове от един, респективно от различни жанрове, за да се открият езиковите особености на отделните жанрове и автори в синхрония;
- на разновременни помежду си текстове от един, респективно от различни жанрове, за да се установят приликите и отликите в диахрония;
- на текстове от различни редакции, за да се уточнят ареалните специфики в словообразователните тенденции при лексиката на средновековни славянски паметници, възникнали в различна диалектна среда (с оглед сравнителната лексикология на славянските езици от началния период на писмеността);
- спрямо съвременното състояние на езика по отношение честотата на употребяваната лексика.

Отчитането на числените данни на фона на други сведения ще способства за изясняването на недостатъчно фактологично обосновани феномени в старобългарски език, както и за установяване типологията му. Въз основа на относителната честота на появата на лексемите в различните текстове могат да бъдат отделени книжовните словообразователни модели, представителни както за включените в свода словни масиви поотделно, така и за старобългарската книжовна норма като цяло (микро- и макроравнище); да се удостовери принадлежността, респ. непринадлежността, на даден текст към старобългарската книжовна лексика; да се презентира характерната за ранните етапи на установяване на книжовната норма конкуренция на деривационни модели в езика от средновековния период; да се определи активният и пасивният фонд на езика от съответната епоха на развитието му. Значимостта на този лексикографски подход към старобългарски текстове нараства предвид факта, че увеличението на словния ресурс на езика се развива толкова по-интензивно, колкото дадена епоха от историята на езика е по-близо до неговото съвременно състояние, ще рече че общото число на известните ни стб думи, отнесено към речниковия състав на старобългарски език като цяло, се оказва несравнимо по-голяма величина, отколкото същите количествени данни за съвременен език.

Един свод на честотните индекси има практическо значение и за изготвяне на учебни речници по старобългарски и църковнославянски език, в които е целесъобразно да се отрази речевият фонд с най-висока честота в изучаваната група текстове.

Изработването на своден рангов списък обаче би се затруднило от различните принципи на съставяне на ранглисти към настоящия момент. В зависимост от базата си – общ речник към определена група паметници или индекс към отделен ръкопис или паметник, те отразяват количествените показатели съответно на макро- и микро- езиково ниво. В тях могат да влизат всички или само част от думите, включени в изходния речник (например, някои фреквенции изключват от състава си собствените имена и географските названия). Ранглистата може да се подрежда в низходящ ред (от най-честата дума към единично употребените – в българската лексикография такива са честотните речник на А. Давидов към Беседа на Презвитер Козма⁹ и на И. Христова към Словата на Климент Охридски¹⁰) или във възходящ ред (от единично употребените думи към най-честата – фреквентните списъци на Т. Илиева¹¹ към Йоан-Екзарховия превод на Богословието и химнографските творби на Климент Охридски)¹². Честотата в листата е съответно абсолютна (Беседа на Презвитер Козма, Словата на Климент Охридски) или относителна величина (СС).

Този разнобой налага изработването на общи принципи за оформяне на честотните речници, които тепърва ще бъдат съставяни: въвеждане на единна ортография, единна норма на лематизиране и сегментиране на текста на отделни думи, единни правила за обема на списъка, уеднаквяване на базата (методологично неправилно е да се обединяват данните от отделен ръкопис и паметник), посоката на подреждане, уравняване на величините. Само така ще бъде възможно ползването на компютърни програми при обработката на материала и получаване на необходимите данни, върху които изследователите да градят своите наблюдения и заключения. (Този тип лексикографски продукти са ярко доказателство за това какви резултати могат да се постигнат, когато методите на класическата лексикография се обединят с методите на статистиката и информатиката.) Работата е много и за улеснението ѝ е наложително създаването на програма за директна обработка на старобългарския лексикален материал без необходимостта от многобройни процедури по кодирането и декодирането на кирилския текст с латиница.

⁹ Давидов, А. Честотен речник...

¹⁰ Христова, И. Цит. Съч., с. 287-319

¹¹ Илиева, Т. Лексиката в химнографското творчество на Климент Охридски...; Илиева, Т. Лексиката в Йоан-Екзарховия превод на Богословието...

¹² Думите могат евентуално да се номерират от №1 до № *n* по низходяща, респ. възходяща честота (практиката на ПрезвК).