

ТЕКСТ, СТАТИСТИКА, КОМПЮТЪР

(Статистически методи за анализ на текст и възможностите за използването им при атрибуиране на средновековни творби)

СТОЯНКА БАБАЛИЕВСКА

Изкушението датира отдавна. В областта на текстологията на средновековната литература то е формулирано като примамлива и осъществима задача на бъдещето от Д. С. Лихачов: „По подсчету слов, по исчислению коэффициента употребляемости того или иного оборота речи в будущем можно будет математически обосновать принадлежность того или иного произведения определенному автору.“¹ В същата година (1961) излиза и книгата на В. В. Виноградов „Проблема авторства и теория стилей“, в която се утвърждава необходимостта от използването и на лингво-статистически методи на изследване² в литературознанието. Стремещът на филологическата наука да преодолее субективността в анализите си и особено при атрибуирането на художествени творби е безспорен. 30 години по-късно формалните методи за анализ и атрибуиране на текст все още се приемат като неизбежна и необяснима филологическа екзотика, а опитите за съвместяването им с традиционния съдържателен филологически анализ са по-скоро изключение, отколкото правило³.

С убеждението, че най-добри резултати могат да бъдат постигнати само при комплексното прилагане на различните методи за атрибуция⁴, тук ще се опитам да представя познатите ми опити за статистически анализ на текстове и възможностите, които те предполагат за атрибуиране на изследваните творби.

Кое впрочем определя стила? Формалните елементи, които имат стойност на белег, които са неговият репертоар, речник

¹ Лихачов, Д. С. Вопросы атрибуции произведений древнерусской литературы. — ТОДРЛ, 17, 1961, с. 35.

² Виноградов, В. В. Проблема авторства и теория стилей. М., 1961, с. 159.

³ Ермоленко, Г. В. Анонимные произведения и их авторы. Минск, 1988, с. 7. В предговора си към книгата акад. Р. Г. Пиотровски пише: „Такое комплексное применение формальных и семантических приемов атрибуции, проведенное на семиотико-коммуникативной основе, не только значительно повышает достоверность предлагаемых Г. В. Ермоленко решений, но и закладывает основы новой методики атрибуции анонимных текстов, соответствующей современной методологии семиотики, языкознания и литературоведения.“

⁴ Да припомним още веднъж казаното от Лихачов: „Но ни в коем случае нельзя ограничивать установление авторства только одним каким-нибудь рядом аргументов: палеографических, языковых, биографических, идеологических и т. п. Взаимодействие различных методов атрибуции — необходимое условие ее достоверности.“ — Лихачов, Д. С. Вопросы атрибуции. . . , с. 41.

и понякога мощно оръдие. И нещо повече, но не така очевидно: една поредица от взаимоотношения, един синтаксис...

Анри Фосийон⁵

В една поредица от статии с неизменното (и очевидно решаващо) съв­торство на Л. И. Бородин с помощта на теория на графите⁶ се изследва стилът на някои староруски писатели и съчинения чрез статистически анализ на синтаксиса на текстовете. Авторите изхождат от предположението, че „стилът на автора се проявява в неговото „пристрастие“ към определени граматически връзки, чиято честота на поява в текста е висока“⁷. За целите на своя анализ те избират достатъчно представителни откъси от съответните текстове (1000 значещи думи от средата и края на съответното съчинение), в които описват граматическите категории и връзки. Получените резултати се пренасят в таблица, където всеки граматически клас се отбелязва с някакъв кодов номер, необходим за по-нататъшния статистически анализ. Резултатите от оформените 92 показателя се подлагат на тривиални в статисти­ката операции: за всеки текст се изготвя матрица на честотите, от която се отделят само граматическите връзки с високи честоти. При сравняване на получените съвкупности от граматически връзки с високи честоти се отделя г. нар. *общоезиково ядро*, съдържащо се във всеки текст и неносещо информа­ция за стила. Тук се отчитат (и съответно отстраняват) тези съвкупности от връзки, за които може да се предположи, че се дължат на редакторска на­меса или някакви екстралингвистични фактори — това е нивото, на което се отчита спецификата на всеки средновековен текст. За окончателния ана­лиз остават съвкупностите от *съществени* граматически връзки, носещи най-значима информация за стила на автора и стила на произведението. Целта на учените в първата им работа⁸ е да характеризират с избрания от тях метод за анализ *стила на определен автор*, а също и *стила на определен жанр*. Затова те избират за анализ текстове с установени и общоприети ав­тори: три произведения на Зиновий Отенски („Истини показание“, „Похваль­ное слово на откритие мощей епископа Никиты“ и „Слово на откритие мощей архиепископа Ионы“) и две на Йосиф Волоцки („Послание епископу Нифонту Суздальскому“ и „Послание Иосифа И. И. Третьякову“). Резултатите от из­вършения анализ потвърждават авторството на Зиновий Отенски и на Йосиф Волоцки на съответните произведения и показват серия от характерни за единия (и обикновено отсъстващи у другия) автор съществени граматически връзки, изразяващи стилите особености на писателя (за Зиновий Отенски това са връзките: съществително + наречие, глагол + наречие, местоимение + съществително; а също и връзките: съществително в именителен падеж + съ­ществително в родителен падеж, местоимение + съществително във винителен падеж). Учените констатира­т и съществени разлики между произведенията на един писател в различни жанрове. Така жанровите различия между „Ис­тинь: показание“ на Зиновий Отенски и двете му слова за откриване на мощи могат да се проследят и в разликите между предпочетените съществени граматически връзки в съответните текстове. Налага се изводът, че откриването

⁵ Ф о с и й о н, А. Животът на формите. С., 1984, с. 50.

⁶ Представа за теория на графите и за това, какво от нея е използвано в случая дава книгата на Б о л о б а ш, Б. Теория на графите. С., 1989.

⁷ Б о р о д к и н, Л. И., Л. В. М и л о в, Л. Е. М о р о з о в а. К вопросу о формальном анализе авторских особенностей стиля в произведениях Древней Руси. — В: *Математические методы в историко-экономических и историко-культурных исследованиях*. М., 1977, с. 303.

⁸ Пак там, 298—325.

на индивидуалния авторски стил може да стане след отхвърляне на съвкупността от съществени граматически връзки, които характеризират *ядрото на жанра*. Тази стъпка в анализа учените правят в следващата си статия, посветена на количествените методи за изследване стила на староруските писатели⁹. Те се спират на двете най-стари житийни творби в руската литература — „Житие Феодосия Печерского“ и „Сказание о князях Борисе и Глебе“¹⁰. В желанието си за по-голяма прецизност извършват два пъти своя анализ, като първия път включват и служебните думи в оформянето на таблицата с показатели. Резултатите и в двата случая са достатъчно убедителни и недвусмислено показват, че двете творби принадлежат на различни автори. Опитът е повторен с още два житийни текста — „Житие Сергия Радонежского“ на Елифаний Премъдри (края на XIV — началото на XV в.) и „Житие Кирилла Белозерского“ на Пахомий Серб (XV в.). Резултатите показват изключителна близост, която може да говори и за един автор. Обяснението е в значителната редакторска намеса на Пахомий Серб в творчеството на Елифаний Премъдри, а науката — в невъзможността да се прилагат *само* количествени методи в изследването на средновековни творби. Отделянето обаче на *ядрото на жанра* в последния етап от анализа дава възможност да се установят определени, макар и не особено значителни разлики в стила на двамата писатели.

Интересен, макар и трудно приложим, е опитът на авторите¹¹ да използват метода на групите за установяване отношенията между преписите на един текст. Този метод е предложен от френския текстолог Жак Фроже¹² и се основава на хипотезата, че всеки апограф носи всички грешки на своя антиграф, което би дало възможност анализът на разночетенията да възстанови процесите на преписване на текста. Задължителните условия, за да бъде приложим методът на групите, са: всеки препис да има само един пряк предшественик и да повтаря всички негови грешки, да няма еднакви грешки в преписи, възхождащи към различни протографи. Методът се прилага от авторите на „езика“ на теория на графите и теория на множествата върху 54 преписа на Закон за съдене на хората (произведение, написано според тях от Константин-Кирил Философ през 60-те години на IX век в пределите на Велика Моравия). Получените резултати изглеждат убедителни, но по принцип методът на групите поради изложените по-горе задължителни условия е почти неизползваем в областта на текстологията на средновековната литература.

Вместо коментар към изложените опити на групата на Бородкин ще си позволя да цитирам пасаж от една статия с обобщителен характер на Б. М. Клос: „Нельзя определить достаточно простой универсальный алгоритм анализа стиля (алгоритм задается системой характеристик) в задаче атрибуции анонимных текстов, потому что решение ее требует перебора и анализа огромного множества параметров. По крайней мере не существует универ-

⁹ Б о р о д к и н, Л. И., Л. В. М и л о в. Некоторые аспекты применения количественных методов и ЭВМ в изучении нарративных источников. — В: *Количественные методы в советской и американской историографии*. М., 1983, 363—393.

¹⁰ Учените отнасят „Житие Феодосия Печерского“ към творчеството на знаменития руски агиограф и летописец Нестор, а „Сказание о князях Борисе и Глебе“ се определя като анонимно произведение, чийто най-стар препис, както впрочем и Житието на Феодосий Печорски намираме в Успенския сборник от XII—XIII в. Вж. С л о в а р ь книжников и книжности Древней Руси. Вып. 1. Л., 1987, 274—275; 398—401.

¹¹ Б о р о д к и н, Л. И., Л. В. М и л о в. Некоторые аспекты..., 383—393; а също и: Б о р о д к и н, Л. И. ЭВМ ищет авторов средневековых текстов. — В: *Число и мысль*. Вып. 9, 1986, 129—140.

¹² F r o g e r, D. J. La critique des textes et son automatisation. P., 1968.

сального алгоритма с ограниченным числом характеристик. . .¹³ Същият автор поставя още един важен проблем, свързан с прилагането на статистически методи, в различни научни области на хуманитарното познание: „Статистическите методи изследования органически сочетаются с методами той науки, предмет которой изучается, а исследователь, как правило, совмещает знание нескольких научных дисциплин.“¹⁴ Въпросът за свързването на статистическите методи с тези на съответната научна област е поставен по-скоро като пожелание, отколкото като фиксиране на действителното състояние на нещата, а що се отнася до владенето на няколко научни дисциплини, то Клос очертава една реална, но неефективна тенденция. Опитът показва, че в тази сложна област на взаимодействие най-добре работят екипи от учени, представители на различни научни области. Изложените по-горе изследвания са дело на екип, включващ поне един математик и един историк. В момента в Ленинград работят два такива екипа и резултат от работата на единия е книгата „Язык русской агиографии XVI века (опыт автоматического анализа)“¹⁵.

Опитите за прилагане на статистически методи в българската литературна наука са малко и самотни¹⁶. Първият опит за използване на точни методи в решаването на най-дискутирания въпрос в българската медиевистика — авторството на пространните жития на Кирил и Методий, датира от 1969 г.¹⁷ След обстоен преглед на всички възможни възгледи за авторството на двете жития, В. П. Василев прилага идеята за устойчивост на честотите на употреба¹⁸ върху лексическия фонд на двете жития и на две Климентови похвални слова — Похвално слово за Михаил и Гавриил и Похвално слово за Кирил: т. е. ученият си поставя задачата да провери най-вероятната и най-често дискутираната хипотеза — за Климентовото авторство на двете пространни жития. Тъй като окончателните изводи на автора се основават не само на статистическите данни, тук ще отбележа само отрицателните резултати от статистически анализ: „цифрите изключват с повече от 99% единоавторството по показателите № 1 (съществителни имена) и № 9 (прилагателни имена)“¹⁹ на Житието на Кирил и Похвалното слово за Кирил; а също и заключението: „Климент остава най-възможен, но недоказан автор на ЖК.“²⁰ Днес разполагаме с пълен речник на Климент Охридски, който включва всички сигурно негови слова и в който се посочва честотата на всяка словоформа във всеки текст²¹. Идеята на В. П. Василев може да бъде осъществена отново въз основа на този речник и в задължителна съпоставка с речника на поне още един старобългарски писател от IX—X век.

¹³ К л о с с, Б. М. О статистических методах исследования текстов исторических источников. — В: *Математические методы в историко-экономических и историко-культурных исследованиях*. М., 1977, с. 329.

¹⁴ Пак там, с. 330.

¹⁵ Герд, А. С., Е. Л. Кузнецова, С. А. Аверина, И. В. Азарова и др. *Язык русской агиографии XVI века*. Л., 1990.

¹⁶ Те отразяват по-скоро лични интереси и пристрастия, отколкото някаква тенденция в науката. Вж. и: Тодоров, И. *Електронноизчислителните машини в текстологичните изследвания*. — Помощни исторически дисциплини. Т. 4, 1986, 216—226.

¹⁷ Василев, В. П. Климент Охридски и авторството на Панонските легенди. — *Български език*, 19, 1969, № 3, 229—239.

¹⁸ За да постигне по-голяма прецизност в наблюденията си, авторът използва класически метод от статистиката: *t*-критерия на Стюдънт. Вж.: Юл. Дж. Э., М. Дж. Кендэл. *Теория статистики*. М., 1960; Каллоша, В. К., С. И. Лобко, Т. С. Чикова. *Математическая обработка результатов эксперимента*. Минск, 1982, с. 27, 58.

¹⁹ Василев, В. П. Климент Охридски. . . , с. 236.

²⁰ Пак там, с. 237.

²¹ Христова, И. Книжовната старобългарска норма в словата на Климент Охридски. Автореферат. С., 1985, с. 22.

Интересен, екзотичен и най-вече непознат сред филолозите е опитът на Л. Бонева за хронологическо подреждане на Платоновите съчинения, а също и на Йовковите разкази²². През 1971 г. Бонева се обръща към идеята на немския филолог W. Kaluscha за хронологическо разполагане на Платоновите диалози чрез изследване краищата на изреченията (или клаузите) върху признака дължина — краткост на сричката (изследва се т. нар. *клаузула*, която обхваща последните 5 срички от изречението или синтактичната цялост. Възможни са 32 типа комбинации по признака дължина — краткост и изследването на всеки текст започва с определянето на клаузулите и съответно разполагането им в един от възможните 32 класа). Самият Kaluscha изготвя таблици на така класифицираните Платоновы творби, без обаче да посочи изданието на Платоновите съчинения, което е ползвал. Направеното от него е коментирано от Brandwood, който прилага и свои таблици. Въз основа на всичко това Л. Бонева успява да направи една убедителна хронологическа подредба на Платоновите диалози, като за статистическата обработка на данните използва вариант на метода на многомерното скалиране²³. Безспорно предимство на предлагания от авторката алгоритъм е възможността предварително да се установи дали получените данни са подходящи за изготвяне на хронологическа таблица (сериация) или не. Предоставена е и възможност за ръчно изчисление на данните, когато материалът е в ограничен обем (например 16 Йовкови разказа²⁴). Цел на подобни изследвания е определяне степента на подобие между двойки текстове въз основа на набор от признаци и разполагането им в определено пространство. (Интересен аналог на тази задача е изготвянето на географска карта въз основа на данни за разстоянията между двойки градове. Възниква въпросът, как да определим последователността в това пространство — на географската карта или хронологическата схема. Ако си представим картата на Средновековна Европа и един странник, чиято цел е да обиколи всички градове, отбелязани на тази карта, то намирането на най-краткия път за обхождането им би дало представа за начина, по който се установява последователността в написването на произведенията от един автор. Математиците наричат това задача за търговския пътник.) Опитът с диалозите на Платон насърчава авторката да приложи постигнатото и по отношение на разказите на Йордан Йовков. Тя извежда свои критерии за сегментация на текста²⁵, като при изследването на краища-

²² Boneva, L. A new approach to a problem of chronological seriation associated with the works of Plato. — In: *Proc. of the Anglo-Roman Conf. of Math.* . . . Edinburg, 1971, 173—158;

Chronological seriation applied in literature. — In: *Proc. of the 39-th Session of I. S. I.*, 1973, 199-209;

Chronological seriation of the works of an author by means of computer. — In: *Использование математических моделей и электронных вычислительных машин в лингвистике*. С., 1976, 319-322;

Seriation with applications in philology. — In: *Math. Statist. Banach Center Publications*, 1980, N₀6, 73-82;

Seriation and Some Applications. — In: *Math. Operations for Sch. Statist.*, Ser. Statistics, vol. 13, 1982, N₀3, 473—485.

²³ Boneva, L. Seriation with application. . . ; Seriation and Some. . .

²⁴ Boneva, L. Chronological seriation. . . ; Chronological seriation of the works. . .

²⁵ Boneva, L. Chronological seriation. . . , 201-202: „Беше ясно, че принципът трябва да бъде фонетико-синтактичен, обхващащ по такъв начин мелодията и цялата хармония на езика. В съответствие с това сегментацията бе направена по следните изисквания:

а) всяка пауза в произношението, да кажем, в случая на запетая, точка и запетая, двоеточие, точка, тире или просто естествена пауза, трябва да бъде считана за край на клаузула и сегментирана;

б) синтактични единици от следните два типа трябва да бъдат взети предвид:

та (*клаузулите*) на синтактичните сегменти опозицията дължина — краткост на сричката е подменена с опозицията удареност—неудареност. Първоначално с 16, а след това и с 28 разказа на Йордан Йовков Л. Бонева потвърждава 100% установеното от филолозите мнение за реда на написване на всеки отделен разказ.

И така, можем да отбележим, че почти всички нива на изграждане на езика са били обект на статистически анализ — и нещо по-важно, че изследванията дори върху средновековен материал дават обнадеждаващи резултати. Проблемите? Те започват от факта, че филолозите не са единни в мнението си за това, кое езиково ниво се оказва най-устойчиво в границите на многовековната ръкописна традиция. Аргументи „за“ и „против“ могат да се намерят за всяка изразена позиция. Едно интересно проучване на акцентната система на два преписа (от 1448 г. и от 1615 г.) от едно съчинение (проучване, направено от Яцимирски в началото на века²⁶) показва, че промените в акцентирането на двата паметника са значителни. . . Съществува мнение, че лексиката и синтаксисът на средновековните творби се променят по-бавно от фонетиката и морфологията²⁷. Но има проучвания на честотните речници на определени съчинения, които ни карат да се съмняваме във възможността да се основаваме на разликите, отбелязани при сравнението им. . .²⁸ Разпадането на падежната система и разсеят на езика към аналитизъм се отразяват в синтаксиса на творбите. А в повечето случаи на произведения с неустановено авторство разполагаме предимно с късни преписи и с недостатъчно добре проучена ръкописна традиция. Задоволителните критически издания на паметници и творби са все още рядкост, но разполагаме с добри езиковедски изследвания, които биха могли да легнат в основата на бъдещите опити за статистически анализи на средновековни творби²⁹. Най-мал-

61) прости изречения — самостоятелни или в рамките на сложното — трябва да бъдат сегментирани и в двата случая;

62) фразални единици — определителни и причастни — трябва също да бъдат сегментирани;

с) сегментът трябва да има поне 5 срички.

²⁶ Яцимирски, А. И. Из истории славянской проповеди в Молдавии. — В: *Памятники древней письменности и искусства*. CLXIII, 1906, с. LIX.

²⁷ Вж. Христова, И. Книжовната старобългарска норма. . . , с. 3.

²⁸ Бородкин, Л. И., Л. В. Милов. Некоторые аспекты. . . , с. 368. Авторите съобщават за сравнително изследване на честотните речници на Житието на Феодосий Печорски, Чтението за Борис и Глеб и Сказание за Борис и Глеб (Вялкина, Л. В., Лукина, Г. Н. Опыт применения некоторых методов математической статистики к изучению лексикки древнерусских текстов. — В: *Исследования по исторической лексикологии древнерусского языка*. М., 1964, 298—304). Резултатите показват, че по количеството обща лексика Чтението е по-близо до Сказанието, отколкото до Житието. Това е причина те да отхвърлят статистическото проучване на лексиката като нерезултатно. Ако по подобен начин се подхождаше към техните данни от анализа на синтаксиса, резултатите също биха били необнадеждаващи. По-коректно би било към събрания лексически материал да се приложат операциите, с които те отделят значимите синтактични връзки, а именно — отделяне на *общото езиково ядро*, отделяне *ядрото на жанра* и анализ на специфичното. В конкретния случай вероятно би възникнал проблем с обема на лексическия материал (възможно е да се окаже недостатъчен за надежден статистически анализ), но идеята би могла да се приложи към по-значителни по обем творби.

²⁹ Вих искала да отбележа особено речниците на определени автори и творби. Вж.: Христова, И. Книжовната старобългарска норма. . . ; Христова, В. Индекс на словоформите от пространните жития на Кирил и Методий. — *Годишник на СУ „Кл. Охридски“*. Т. 77. С., 1983, № 1, 261—330; изследването на Пенкова-Педерсен, П. Старобългарският молитвеник и електронната машина. — В: *Първи международен конгрес по българистика. Доклади. Симпозиум кирило-методиевистика и старобългаристика С.*, 1982, 132—149.

ко е направеното в областта на акцентологията³⁰ — поради липсата на достатъчно и системно акцентирани паметници, поради лошото състояние на някои от тях³¹, поради липсата на традиция в този тип изследвания³². Затова и вероятно най-трудно ще бъде използвано направеното от Л. Бонева.

Интересният опит на Л. Бонева и най-вече фактът, че този тип статистически анализ дава определени резултати и върху български езиков материал (т. е. подмяната на опозицията дължина—краткост на сричката с опозицията удареност—неудареност не се отразява съществено върху характера на предварителния анализ), бе причина да се опитам да приложа този метод за статистически анализ върху материал от старобългарската литература. Бяха изследвани откъси, включващи около 300 клаузули от четири творби на старата българска литература: Пространно житие на Кирил (по препис в ръкопис № 19 на московската Духовна академия — средата на XV век³³ — от първа до четвърта глава включително); Похвално слово за Кирил (първичната версия от Севастияновия препис в ръкопис от XIII в. и вторичната версия — по Хлудовия препис от XIV в.³⁴); Успение Кирилово (по препис в ръкопис № 58 от сбирката на Гилфердинг³⁵ и по Втори Белградски препис в ръкопис № РСII от началото на XV в.³⁶) и началото на Слово за четвъртия ден от Шестоднева на Йоан Екзарх³⁷.

Както отдавна е отбелязал Гьорнер³⁸, най-ранният паметник с почти последователно прокарана фонетическа акцентуация е доста кратък и доста късен в сравнение с интересуващите ме творби, затова ударенията на изследваните откъси не бяха съобразени с данните за акцентните системи на старобългарските паметници, а бяха поставени в съответствие с акцентологическите проучвания на Кодов. Така тук, на първото ниво в анализа на текстовете бе заложен първият фактор за известна некоректност на изводите — тъй като очевидно при този подход в най-добрия случай се реконструират не друго, а *нормата*. Специфичните колебания в акцентната система на съответните писатели при това състояние на ръкописната традиция са естествено невъзстановими. В синтактичната сегментация на текстовете бяха следвани основно принципите, изложени от Бонева³⁹. В определен смисъл още на това

³⁰ Освен класическите работи на К о д о в, Хр. Подвижното българско ударение и неговото отношение към праславянското ударение. С., 1929; Ударението в българския книжовен език. С., 1966; и статията на Г ъ о р н е р, Ф. За надредните знаци в старобългарските паметници до XIII в. — *Език и литература*, 21, 1966, № 6, 53—58; тук ще посоча само обстойния библиографски преглед в : G a r d e, P. Histoire de l'accentuation slave. P., 1976, v. 1, 2.

³¹ Например Евтимиевият служебник — виж К о ц е в а, Е. Евтимиев служебник. С., 1985.

³² Интересна тенденция очертават изследванията на акцентната система на определени ръкописи (вж.: С т е н с л а н д, Л. Акцентировка и акцент. Акцентологически анализ Служебника XV в. Chil. 323. Стокхолм, 1990; Л и н д г р е н, Н. Акцентологическа характеристика севернорусской рукописи XVII века. Стокхолм, 1990); поместването на акцентологически коментар при издаването на паметници (вж.: Н о р о в с к а я псалтирь. Среднеболгарская рукопись XIV века. С., 1989 — в т. 1 са поместени Описание на акцентната система на паметника от В. А. Дьбо и Указател на думите с ударение от И. К. Бунина) или речници (вж.: Я з ы к русской агиографии XVI века. Л., 1990, 315—330).

³³ Л а в р о в, П. А. Материалы по истории возникновения древнейшей славянской письменности. Л., 1930, 1—6.

³⁴ И в а н о в, Й. Български старини из Македония. С., 1931, 328—333.

³⁵ Л а в р о в, П. А. Материалы. . . , 154—157.

³⁶ А н г е л о в, Б. Ст. Из старата българска, руска и сръбска литература. Т. 3. С., 1978, 7—10.

³⁷ A i t z e t m ü l l e r, R. Das Hexaemeron des Exarchen Johannes, VI, 1966. Graz, 3—53.

³⁸ Г ъ о р н е р, Ф. За надредните знаци. . . , с. 58.

³⁹ Вж., бел. № 25.

ниво на анализа можеше да се предвиди, че крайните резултати едва ли ще бъдат подходящи за статистически анализ. Логиката на синтаксиса предполага в края на синтактичната цялост самостоятелна дума, която в старобългарските текстове често е многосрична — и може да носи най-много две ударения. За разлика от езика на художествената реч на XX век езикът на старобългарските писатели предлага по-малко възможности за натрупване на едносрични и двусрични думи в края на синтактичната цялост, а това на практика означава, че не всички възможни комбинации по признака удареност — неудареност на последните пет срички могат да бъдат представени с примери. Наистина от 32 възможни класа 16 останаха празни⁴⁰. Това бе причина опитът да бъде повторен, като се изследват последните четири срички от сегмента — така възможните класове от комбинации по признака удареност — неудареност на сричката са 16 и от тях непредставени с примери се оказаха само два. Намалените наполовина класове за анализ обаче намаляват и възможностите за извличане на информация. В крайна сметка получените резултати се оказаха неподходящи за статистически анализ (по критериите, изложени от Бонева) и дори без много изчисления показваха твърде голяма близост помежду си. Всъщност близостта между отделните преписи на едно произведение не беше по-голяма от близостта между две различни произволно сравнени творби.

Защо все пак, въпреки ясно очертаващите се трудности на всеки етап от анализа, аз продължих работата по обработката на текстовете, а сега смятам за нужно да споделя резултатите. Първо, защото мисля, че е важно този тип изследвания да продължат с по-голяма интензивност от средно едно проучване на десетилетие — тъй като обработката на текстовете за статистически анализ предполага един доста сериозен и често нетрадиционен прочит на текстовете (т. е. независимо от резултатите статистическият анализ може да бъде разглеждан и като етап от проучването на творбата и ръкописната ѝ традиция). И второ — защото съществуването на информационни технологии, които все още не са ни изцяло подвластни, няма да ни позволи да снемем идеята за статистически анализи като предстоящ и пред медиевистиката проблем.

Изпреварвайки възможните възражения, бих искала да коментирам причините, поради които изложените тук опити за статистически анализ невинаги са свързани с конкретна задача за атрибуиране на текстове. Нито един от представените автори не крие амбициите си за евентуално използване на прилагания от него метод в решаването на спорен въпрос за авторство и това личи дори от заглавия като: „ЭБМ ищет авторов средневековых текстов“⁴¹. Но „въпросът за атрибуцията на произведението е частен въпрос от историята на текста на това произведение“⁴² и неговото решаване обикновено е ключителен етап в изследването на творбата. . .

Накрая искам да споделя обезпокоителното си впечатление за липсата на траен и системен интерес от страна на съвременната медиевистика към възможностите за използване постиженията на математическата мисъл. И

⁴⁰ За сравнение ще приведа примери от Йовковия разказ „Бели рози“ (Й о в к о в, Й. Разкази. С., 1970, 155—161) за някои от класовете, в които изобщо не бяха регистрирани примери от старобългарските текстове (с плюс са означени сричките под ударение):

Каж и сега, ар-ка-даш, ка-кво и-ма.	+ - + + -
Няма да ме-теш там, ти каз-вам.	+ + - + -
Ах, Ми- тьо, Ми- тьо.	+ + - + -
Не мо- же тъй, как та- ка.	- + + - +

⁴¹ Б о р о д к и н, Л. И. ЭБМ ищет авторов. . .

⁴² Л и х а ч о в, Д. С. Вопросы атрибуции. . ., с. 18.

за да бъде по-точна, ще си позволя да отбележа, че това е общ проблем на филологията днес. Отдавна отбелязаният, поостарял уж проблем за двете култури⁴³ отново ни застига. Изложените дотук опити за преодоляване на дистанцията изненадват с констатацията, че творческата активност принадлежи преди всичко на математиците⁴⁴. Това е до голяма степен естествено, тъй като те отлично познават възможностите на съвременната изчислителна техника, както и развитието на математическата статистика, често са изкушени с трайни литературни пристрастия и по-малко познават конкретните трудности, обезсилващата неподатливост на литературния текст. Това ги прави по-големи оптимисти по отношение на крайните резултати, а и в тяхната наука идеята за експеримента е по-естествена. Те са по-малко консервативни. . . . Какво им пречи? Може би преди всичко липсата на сетива за специфичното в другата, непознатата култура. Във всички познати ми публикации проблясва белегът на известна незаинтересованост от детайлите, от дреболиите; пренебрегване на известни условия, на определена вариативност в изследваните граматически категории — незначителна от гледна точка на статистиката, по обект на не един тънък филологически анализ. Особено притеснява поднасянето на резултатите: „Натрупващите се суми обаче са категорични — разликата е незначителна, но ясно различима.“⁴⁵ Дори ако следват страници, изпълнени с цифри и обяснения, нормалният филолог вече е безразличен към тях. Кой смее да изрази разликата между Толстой и Достоевски с десетична дроб?

Една от важните психологически предпоставки за традиционната консервативност на филолозите по отношение на предлаганите математически облекчения е дълго утаяваното съкровено разбиране, че дълбоката същност на художествения текст е неизмерима количествено, че не думите, не свързването им, а нещо цялостно и неуловимо изразява авторовия облик и художествената идея. Така изкрystalизира убеждението, че художествената творба не може да бъде разбрана с анализ на граматиката на езика ѝ:

Объяснить грамматикой странность литературного произведения безнадежно.

В. Шкловский⁴⁶

Разбира се, филологията отдавна не е единна в това разбиране — не-малко съвременни литературоведи подчертават „дълбинното единство между езиковата дейност и разказа“⁴⁷ и настояват за създаването именно на „граматика на разказа“. Не можем да подминем и факта, че тъкмо математическата логика оформя понятийния апарат на семиотиката например: „Способность различных математических дисциплин выступать в качестве метаязыков также и при описании явлений искусства очевидна.“⁴⁸

Но, изглежда, съвременната математика е точно толкова безразлична към опита на семиотиката, колкото и съвременната литературна наука към опитите на математиката да мери количествено художествени произведения. Доколкото има проникване в границите на другата култура, то крехките нишки, бележещи това проникване, никъде не се преплитат.

⁴³ Сп о у, Ч. П. *Двете култури* и още веднъж за *двете култури*. С., 1983.

⁴⁴ Изключително продуктивни са учените от факултета по компютърни науки при Единбургския университет.

⁴⁵ Стоянов, В. Литературен детектив. — Орбита, 1990, № 49, с. 15.

⁴⁶ Шкловский, В. Энергия заблуждения. М., 1981, с. 108.

⁴⁷ Тодоров, Цв. Поетика на прозата. С., 1985, с. 70.

⁴⁸ Лотман, Ю. М. Искусствознание и „точные методы“ в современных зарубежных исследованиях. — В: *Семиотика и искусствознание*. М., 1972, с. 6.

Всъщност пред нас са два вида действителност — едната на непосредственото артистично възприятие, и другата — на скритата научна обусловеност, и те не си подхождат, и не си приличат, и всъщност нямат много общо една с друга.

Р. М. Пърсиг⁴⁹

Ще припомня, че известната книга на Пърсиг (откъдето е взет горният цитат) поставя отново проблема за двете култури във връзка с търсенето на качеството и налага на читателя извода, че качеството на бъдещия ни живот (а също и на бъдещите ни научни изследвания) зависи твърде много от възможностите за взаимопроникване на двете култури. Когато говорим за количествените методи във филологическия анализ, изглежда, че са нужни малко неща — математическата статистика да се опира преди всичко на формалния литературоведски анализ, а не толкова на изброяване на граматическите единици в текста; комплексността при оформянето на критериите и показателите да се осигурява не от механичния сбор на колкото се може повече показатели, а от стратегически избор на същностни характеристики; и най-важното — и най-трудно — да се пробуди стремеж за проникване в мисловната система на другата култура — или поне на техниката, с която се борави ежедневно.

Но има ли какво да даде хуманитарното познание на точните науки? Ще се позова на лорд Халсбъри (председател на английското кралско философско дружество): „Аристотел и Аквински могат да ни научат как да мислим за една Вселена, която все пак може да се окаже и неразбираема. . .“⁵⁰

⁴⁹ Пърсиг, Р. М. Зен и изкуството да се поддържа мотоциклет. С., 1980, 67—68.

⁵⁰ Стефанов, А., Ив. Райнова. XVIII световен конгрес по философия — многоплановост и проблемност. — Философска мисъл, 1989, № 1, 31—32.