PALAEOBULGARICA / СТАРОБЪЛГАРИСТИКА, XX (1996), 1

Лора Тасева, Мария Йовчева (София, България)

КОМПЮТЪРНА ОБРАБОТКА НА СРЕДНОВЕКОВНИ СЛАВЯНСКИ РЪКОПИСИ

От 24 до 28 юли 1995 г. в Благоевград се състоя Първата международна конференция „Компютърна обработка на средновековни славянски ръкописи" (Computer processing of medieval Slavic manuscripts). Тя бе организирана от Института за литература при БАН и Университета в Питсбърг с координатори проф. Дейвид Бирнбаум (Питсбургски университет) и ст. н. с. Анисава Милтенова (Институт за литература при БАН). Сред организаторите бяха също Милена Добрева (Института по математика) и Американският университет в Благоевград. Домакин беше Американският университет, а финансовата помощ се дължеше на Съвета за международни проучвания и обмен (International Research and Exchange Board — IREX), Американската фондация за научни изследвания (American Council of Learned Societies) и Фондация „Отворено общество". Назрялата през последните години необходимост от съвместна работа на учените филолози и специалистите по информатика и от съгласуване на усилията на отделните изследователски центрове, както и изключително добрата предварителна подготовка от страна на инициаторите, привлече над 60 участници от 11 страни: България, Великобритания, Германия, Италия, Канада, Македония, Русия, САЩ, Украина, Холандия, Чехия. На научния форум бяха представени досегашните опити за използуване на компютрите в палеославистични проучвания, както и принципните възможности, проблемите и резултатите от прилагането на различни програмни продукти за решаване на конкретни филологически задачи.

От името на организаторите конференцията бе открита от проф. д-р Дейвид Бирнбаум от Питсбургския университет. Приветствия към участниците и пожелания за успешна работа отправиха проф. д-р Грегъри Найт — декан на Факултета по балканистика при Американския университет в Благоевград, ст.н.с. Климентина Иванова от Института за литература при БАН, проф. д-р Иван Добрев — ръководител на Катедрата по кирилометодиевистика в Софийския университет, ст.н.с. Динчо Кръстев — директор на Централната библиотека при БАН, и д-р Трифон Трифонов — директор на Центъра за изкуства и литература „Сорос" — София.

Деловата порграма включваше 34 доклада, обединени в 10 тематични групи, както и три практически занятия.

I. На пленарното заседание „Състояние на проучванията" (State of the art) бяха изнесени 3 доклада

В своя доклад „Как славистите филолози могат да използуват компютрите" (How slavic philologists should use computers) Дейвид Бирнбаум (Питсбург) се спря върху значението на компютърната текстова обработка, която предоставя възможност за по-лесен ревизионен контрол и по-добра шрифтова поддръжка. Изтъкнати бяха предимствата на електронния текст и преди всичко лекотата, с която той може да бъде мултиплициран, пренасян и съхраняван. В заключение авторът подчерта необходимостта от използуване на стандартизирани системи, за да бъдат научните изследвания широко достъпни и използваеми от колегите и в настояще, и в бъдеще.

Винфрид Бадер (Тюбинген) с доклада си „Библията и компютърът: кратка история за последните десет години" (Bible and computer: a brief history of the last ten years) обобщи търсенията на Международната Асоциация „Библия и информатика" (AIBI). Като направи преглед на нейните четири конференции от 1985 до 1994 г., той

отбеляза проблемите, свързани с кодирането на гръцки и еврейски текстове, първите опити за съставяне на индекси, включването на морфологична информация и възможностите за автоматизиране на граматичния анализ. Според В. Бадер шансовете и перспективите, които съвременната техника дава за библеистиката, са в компютърните издания от типа на сега подготвяните нови публикации на BHS (Cod. Leningradensis) и на кумранските свитъци.

В доклада на Елена Паскалева и Милена Добрева (София) „Нови средства за стария език: компютърна обработка на български текстове" (New tools for old language: computer processing of Bulgarian texts) бяха описани опитите за прилагане на различни платформи и програмни продукти при работа с български текстове, както средновековни, така и съвременни. Анализираха се съществуващите трудности, създавани от извънредно вариативната старобългарска графична система, от една страна, и от стандартизираната в лингвистичен, но не и в компютърен смисъл, съвременна кирилица, от друга. Авторките подчертаха, че при сегашните технически възможности специалистите в областта на математическата лингвистика продължават да търсят модели, подходящи за българския език.

II. Проблеми на кодирането (Encoding problems)

Докладът на Дейвид Бирнбаум (Питсбърг) „Информационни и представителни единици в ранната славянска писменост" (Informational and presentational units in Early Cyrillic writing) постави един от най-острите проблеми за славистите филолози, които използват компютри — как да кодират знаковете на ранните славянски писмени паметници като битови комбинации. Подчертана бе необходимостта да се различават информационни единици (characters) от представителни единици (glyphs) и да се изясни кои единици трябва да се кодират: само единият тип, само другият тип или и двата. Трудността за универсален отговор на този въпрос произтича от факта, че опозиции, значими за графичната система на един ръкопис, се оказват ирелевантни за друг. Поради това се налага кодирането на знаковете да е съобразено винаги с локалните правописни и езикови особености.

Карстен Грюнберг (Хайделберг) в доклада си „Транскрипционни правила за старославянската писменост" (Transcription rules for Old Church Slavonic writing) обосновава необходимостта от възприемане на единна, стандартизирана и общодостъпна система за транслитериране и транскрибиране на старите славянски ръкописи, без която е немислимо използването на компютрите за каталогизиране, факсимилно възпроизвеждане или подготовка на издания от дипломатически тип, както и за статистически анализи и сравнителни проучвания на средновековни текстове. Той предложи за обсъждане набор от основни правила за транскрипция и транслитерация на славянските писмени паметници от Средновековието.

В своя доклад „Използването на средновековно писмо в персонален компютър" (Implementing a medieval script on a personal computer) Себастиан Кемпген (Бамберг) сподели десетгодишния си опит в създаването на стари славянски шрифтове и очерта техническите възможности, ограниченията и перспективите при реализацията на комплексни нестандартни писмени системи в съвременните персонални компютри. Авторът разгледа проблеми като: избор на кодиращата схема, трансформиране на данни през различни системи и платформи, въвеждане на надредните и характерните алографи и др.

III. Текстова обработка: проблеми, методи и средства (Text processing: problems, methods and tools)

В доклада си „Приложението на принципите на Инициативата за текстово кодиране (TEI) при кодирането на един сръбски църковнославянски псалтир от XIV в." (The application of Text encoding initiative (TEI) guidelines to the encoding a 14th-century Serbian Church Slavonic Psalter) Мери МакРобърт (Оксфорд) представи работата си по привеждането в електронна форма на ръкопис, от който могат да се извлекат две успоредни текстови версии — оригинална и последователно редактирана. Специално внимание авторката отдели на транскрипцията на диакритичните знаци и съкращения; на отбелязването на химнографската структура, от една страна, и на разположението на стиховете в кодекса, от друга; на различните типове поправки.

Изнесеният от Александър Молдован (Москва) доклад „За създаване на фонд на писмените извори за руския език през X—XII в. в Института за руски език към РАН" (О создании фонда письменных источников русского языка X—XII вв. в Институте русского языка РАН) очерта двете основни направления в работата на колектива: въвеждане в компютър на най-старите преписи като се запазват езиково

релевантните особености на графиката и правописа и създаване на банка от текстове, съхранени само в по-нови преписи с разночетения на основата на текстологическото им проучване. В заключение авторът изтъкна ролята на съвместната работа между учените слависти за изграждане на единна информационна база за исторически и текстологически изследвания.

IV. Сравнение на преписи (Manuscript collation)

Ирина Азарова (Санкт-Петербург) изнесе съобщение на тема: „Обработка на библейски текстове в Санкт-Петербургския отдел на Библейското общество в Русия" (Обработка библейских текстов в Санкт-Петербургском отделении Библейского общества в России). Тя запозна аудиторията с работата и проблемите при събиране, въвеждане и обработване на огромното количество славянски библейски текстове. Подчертано бе както значението на предварителното филологическо изследване, така и необходимостта да се прилагат съвременните технически средства.

В доклада си „Компютърно сравняване на компютърни преписи от ръкописи" (Computer collation of manuscript transcriptions) Майкъл Бакер (Амстердам) представи своя опит от обработката на новозаветните цитати в Синайския евхологий. Значимите разночетения се изолират от огромното множество фонетични и правописни варианти, така че да бъде конструиран текст, независим от конкретните ръкописи. Отбелязани бяха и практическите потребности, с които се сблъсква филологът при създаване на компютърен траскрибиран текст от конкретен ръкопис. В допълнение М. Бакер се спря на начините за проверка на грешки и за икономия на време.

Колективната разработка на Таня Иванова, Нина Шойлева и Ирина Болчева (София) „Advanced searching" запозна аудиторията с техния опит в използването на компютърна програма за текстологически сравнения на преписи от съчинението на Черноризец Храбър „О писменѣхъ". Те изтъкнаха значението на правилното разчленяване на текстовите равнища и подравнища, както и ролята на грешките, регистрирани в отделните ръкописи. Принципните постановки бяха онагледени и детайлизирани в последователната демонстрация на програмния продукт, изработен от Пламен Димитров от Техническия университет в София.

V. Шрифтове. Графична обработка (Fonts. Image processing)

В доклада на Мимоза Майстороска (Скопие) „Църковнославянска азбука за възпроизвеждане на стари ръкописи чрез микрокомпютър" (A Church Slavonic alphabet for reprinting old manuscrpts using a microcomputer) бе предложен компютърен шрифт, създаден от авторката след предварителен геометричен анализ на формата на всеки знак. Тя показа и досегашните резултати от опитите въз основа на този фонд да се разпознават от компютъра буквите в сканирания текст на ръкописа, както и методите за подготовка на този текст за понататъшна обработка.

Друго решение за набора от графеми и тяхното разпределяне по компютърните кодове поднесе Любчо Митрески (Скопие) в доклада „Шрифтове, азбуки и църковнославянски графеми" (Fonts, character sets and Church Slavonic graphemes). Той мотивира необходимостта комплектът от основните редни и надредни знаци да се отдели от този на главните букви и най-често използваните лигатури.

Ивелин Стоянов (София) в доклада „Оптическо символно разпознаване на исторически документи" (Optical character recognition of historical documents) съобщи резултатите от свързания с подготовката на исторически речник на българския език експеримент за разпознаване на сканиран текст от старопечатни издания. Описани бяха основните етапи — представяне на текста, предварителна обработка, символно разпознаване и заключителна обработка, както и типовете порблеми в процеса на работа.

Уилям Федер (Амстердам) в доклада си „Главни възможности на компютърните издания" (General possibilities of computer supported editions) направи характеристика на традиционните и по-новите начини за публикуване на средновековни писмени паметници: в книжно тяло, фотокопийно или чрез електронен носител на информацията. Като анализира предимствата и недостатъците на всяка от трите технологии с оглед на критериите трайност, удобство при използване, възможност за трансформиране на текста, той обоснова своето предпочитание към електронно обработваните компютърни издания.

VI. Компютърна обработка на данни от средновековни славянски ръкописи чрез използването на стандартизирания генерализиран Mark-up език (SGML)

Анисава Милтенова (София) в своя доклад „Компютърен анализ на макроструктурата и типологията на средновековните славянски сборници" (Computer-aided analysis of the macrostructure and typology of medieval slavic miscellanies) представи резултатите от изследването на структурните модели на различните типове сборници в електронна форма, осъществявано от колектив. Авторката разгледа класификационните принципи, съобразени с типологията на различните ръкописи, като отдели особено внимание на сборниците със смесено съдържание и специфичните проблеми, които те пораждат за компютърната обработка. Тя обобщи, че крайната цел на изследването е изработването на репертоар на преводните и оригиналните съчинения чрез компютърни средства.

Докладът на Андрей Бояджиев (София) „Палеографски и правописни особености при описването на славянски ръкописи" (Palaeographic and orthographic features in the description of Slavic manuscripts) постави проблема за оптималния подбор на кодикологически, палеографски, правописни и езикови данни от писмените паметници и тяхното кодиране. Изхождайки от принципите на стандартизирания генерализиран маркиращ език (SGML) и принципите на Инициативата за текстово кодиране (TEI), авторът разгледа въпроса за мястото на филологическите данни в системата и способите за тяхното представяне в електронна форма.

VII. Приложения на системи за база данни (Applications of database management systems (DBMS))

Съвместният доклад на колектив от Скопие в състав Стана Янкоска, Драган Михайлов и Любчо Йосифоски на тема „База данни за стари славянски ръкописи в Македония" (A database for ancient Slavic manuscripts in Macedonia) разгледа експерименталната библиографска и фактографска информационна банка за славянските ръкописи в Македония. Отделният запис включва културно-исторически, литературно-исторически и филологически данни за идентификация на всеки ръкопис. Колективът предвижда бъдещо включване на информацията в международната система INTERNET и пренасянето ѝ на CD-ROM.

Ползотворното сътрудничество между филологията и информатиката пролича в доклада на Зденка Рибарова (Скопие) и Кирил Рибаров (Прага) „Компютърна обработка на стари църковнославянски ръкописи: резултати и перспективи" (Computer processing of Old Church Slavonic manuscripts: results and prospects). Вниманието на авторите се съсредоточи върху две от приложенията на компютърните системи за палеославистичните изследвания: за представяне на оригиналния документ, неговото езиково сегментиране и подлагането му на лексикален анализ. Създадената от колектива програма STINO дава възможност за изработване на различни видове индекси, за селекция по семантичен, морфологичен или словообразувателен критерий, за автоматизирана нормализация и локализация, за оформяне на корпуси старогръцко-старославянски съответствия за отделен ръкопис или общо.

Юрий Лабинцев (Москва) представи замисъла за международен изследователски проект „Slavia Orthodoxa et Slavia Romana" в доклада „Взаимодействие на славянските светове: Духовната култура на Подлесието" (Взаимодействие. славянских миров: Духовная культура Подлесья). Амбициозната задача на научния колектив от Полша, Беларус, Литва, Русия и Украина предвижда създаването на база данни, отразяващи хилядолетния контакт между западното и източното християнство в Подлесието. Информационният продукт ще бъде основа както за многотомно енциклопедично издание, така и за отделни теоретични разработки.

Със своя доклад „Основни задачи при създаването на база данни на Международната славяноведска програма „История на писмената култура в Подлесието" (Основные задачи построения банка данных Международной славяноведческой программы „История книжной культуры Поляшья" Любов Щавинская (Москва) запозна аудиторията с особеностите при осъществяване на такъв тип програми, чиято изворова база включва няколко хиляди многообразни източници с различни типове данни. Подходяща за обработване на подобна специфична информация авторката посочи програмата KLEIO. В заключение тя изтъкна, че разработваната система ще интегрира историческа и културна информация.

Заседанието завърши със съобщението на Уилям Федер (Амстердам) и Уляна Пикулик (Лвов) за започнатото от тях компютърно описание на ръкописите от Лвовските книгохранилища — една отдавна назряла необходимост поради голямата стойност за медиевистиката на пазените в тях писмени паметници.

VIII. Частни изследвания (Case studies). Типове ръкописи (Type of manuscripts)

В доклада „Псалтирът: от устната към информационната традиция" (The Psalter: from the oral tradition to the informatic ones) Моня Камуля (Пиза) сподели своя опит в прилагането на компютърна система за изследване на стари славянски текстове. Тя представи възможностите за анализ и синтез на филологически особености от 50 псалма по 7 отделни редакции и начините за изработване на различни азбучни и честотни индекси. Специално внимание бе отделено на програмата за вътрешноактивна хипертекстова обработка на корпуса от различни псалтирни версии — query system, чрез която текстът може да се анализира и управлява. Като перспективна задача бе набелязано усъвършенстването на програмния продукт върху базата на SGML за комплексни текстологически съпоставки.

Синтия Вакарелийска (Орегон) в доклад на тема „Средновековни славянски месецослови в режим на работа" (Medieval Slavic menologies on line) изложи аргументи за необходимостта да се анализира месецословния материал поради неговата ценност при датиране и локализиране на ръкописите. Обработени са 75 ръкописа от X до XIV в. според принципите на TEI, като се спазват следните важни условия: 1) транскрипционни норми за палеографските и правописните варианти; 2) програма за ефективно търсене и идентифициране на отделните имена на светци; 3) програма за идентифициране на взаимните връзки между отделните ръкописи на базата на специфични общи лексикални и текстови черти и 4) възможност за бъдещи приноси от отделни изследователи, съвместими с транскрипционните норми на проекта.

Върху компютърната обработка на служебните минеи с помощта на програмата Collate 1.1 привлече вниманието Емилия Гергова (София) с доклада си „Компютърен анализ на химнографските сборници" (Computational analysis of hymnographic compendia). Проучването, което е ориентирано към три различни равнища на служебните минеи — техния макросъстав, структурата на службата и вътрешната сегментация в песнопенията, разчита на възможността програмата да работи на четири кодирани нива. Авторката набеляза направленията, в които трябва да бъде усъвършенстван програмният продукт: разнообразно кодиране на подготвителните файлове, автоматизирана съпоставка на разнородните файлове и синхронизирана изява на данните от трите разреза на изследване.

Използването на компютърната програма IST за изследване на друг тип средновековни книги бе разгледано в доклада на Мария Шнитер (Пловдив) „Система за закодиране на евхологични текстове в параметрите на компютърната програма IST" (A system for encoding euchological text within the parameters of the IST computer program). Тя изнесе принципите за подбор и предаване на филологическата информация, извлечена от 85, предимно южнославянски паметника от IX до XVIII в. Анализирана бе оптималната степен на абстракция при кодирането на данните, която би позволила прецизна идентификация на всеки некалендарен текст според композиционните му особености.

В доклада на Франсис Салтер (Онтарио) „Някои съображения, породени от компютърния анализ на Хрисовула на Свети Стефан" (*Some suggestions arising from a* personal computer analysis of the Svetostefanska hrisovulja) бяха разгледани проблемите на програмирането и структурирането на база данни, необходими за продуциране на документа, и бяха изказани препоръки за бъдещото развитие на компютърната порграма. Постигнатите досега резултати включват възпроизвеждане и обработка на грамотата за изготвяне на няколко типа индекси. Широките перспективи за по-нататъшната работа се изразяват в конвертируемостта на текстовете между различните програми, съставянето на компютърен речник на стари славянски текстове и записване на CD-ROM на ранни славянски документи, което ще позволи четене и анализ на текстовете без принтирането им.

В своя доклад „Проблеми при изготвянето на унифициран пълен каталог на славянски текстове" (Problems involved in setting up a unified comprehensive catalogue of Slavic texts) Райнхард Лампе (Хайделберг) мотивира становището си, че източноевропейската православна писмена култура не може да бъде анализирана със средствата и моделите на западноевропейската католическа писмена култура. Като основен проблем при изготвянето на компютърния каталог той изтъкна разноезичието на научната информация. Според него засега създаването на единна система за закодиране на славянски средновековни текстове остава невъзможно, независимо от наличието на Unicode, тъй като различните букви все още имат собствени уникални компютърни кодове.

IX. Multimedia

Докладът на Велина Братанова, Елена Коцева и Драгомир Петров (София „Проектът „Света София" (The „St. Sofia" project) запозна участниците в конференцията с френско-българския експеримент за създаване на компакт диск (CD-ROM), посветен на българското културно наследство. В него са описани, анализирани и сравнени писмени текстове и художествени паметници в синхронен и диахронен аспект, като се акцентира върху отношенията писмена — неписмена култура. Резултатите от проекта показват широките възможности за съпоставка на разнороден материал: ръкописи, стенна живопис, музикални творби, а изработената методика може да се приложи и при други подобни изследвания върху българското Средновековие.

X. Програмата KLEIO

Докладът на Виктория Тяжелникова (Москва) „Руската версия на програмата за обработка на извори KLEIO: нови възможности за третиране на кирилските източници" (A Russian version of the source-oriented software KLEIO: new opportunities for the treatment of Cyrillic sources) обобщи резултатите от работата по двегодишен съвместен руско-немски проект и се спря на особеностите на руския вариант на програмата. Основното предимство, според авторката, е възможността да се създаде модел, базиран върху контекстуалния анализ на документа, като се отчитат двете равнища на информацията: структурното и семантичното. Значението на подобен подход е изключително голямо поради особения характер на разглеждания материал.

Владимир Тихонов (Москва) в доклада си „Вътрешноактивна пълна текстова обработка чрез KLEIO: критерии при сканиране на основните термини (въз основа на жалбите на т. нар. „лишенци")" (Interactive full-text processing in KLEIO: scanning criteria of basic terms (on appeals by disenfranchised people) представи прилагането на програмата KLEIO за съдържателен анализ на корпус от думи и термини. Материалът е извлечен от жалбите на т. нар. „лишенци" — социална група, съществувала в Съветска Русия през 1920—1930 г. Подчертаха се възможностите за компютърна обработка на кодираните лексикални данни с оглед на различните параметри — честота на употребата, семантични полета, граматични характеристики, съчетаемост на езиковите единици.

Докладът на Игор Юшин (Москва) „Моделът на исторически извори в KLEIO като основа на интегрирана социална класификация" (A Model of historical sources in KLEIO as a basis for integrated social classification) запозна аудиторията с трудностите при съставяне на социалната класификация и открои възможностите на KLEIO в тази насока. Изтъкнати бяха основни предимства като: отчитане на съдържанието на данните при обработката им, така че кодирането да е тясно свързано с техния контекст; развиване на моделите за обработване, основани върху временни параметри, което позволява формулирането на повече потенциално правилни хипотези и ускорява процеса на избор измежду тях и др.

Възможностите на програмата KLEIO за решаване на въпроси на авторството разгледа Надежда Романкова (Смоленск) в доклада „Проблемы атрибуции анонимных старославянских произведений и компютер". Тя проследи етапите на изследването и подчерта значението на обема на сравнявания текст за меродавността на различните честотни регистри и оттук — за отчитания чрез програмния пакет LOTUS коефициент на близост.

Аделина Ангушева (София) в своя доклад „Компютърно изследване на средновековните гадателни книги чрез средствата на KLEIO" (A computer investigation of medieval prognostic books by means of KLEIO) запозна с резултатите от сравнителното поручване на гадателните книги в три средновековни традиции: византийската, латинската и славянската. Програмата KLEIO осигурява класификация на текстовете, сравнение между типовете и версиите на една и съща книга, както и между текстовете на различни книги и анализ на тяхната типология. Крайните статии съдържат изчерпателни данни за ръкописа и пълно описание на прогнозите.

Маргрет Димитрова (София) в доклада си „Гръцки и латински заемки в Нюйоркския мисал и компютърната програма KLEIO" (Greek and Latin loadwords in the New York Missal and the KLEIO computer program) представи някои типове резултати от използването на информационния продукт при изследване на чуждите думи в хърватския глаголически паметник от XV в. Въз основа на масив от 2500 лексикални единици са експериментирани възможностите на програмата за изследване на текстологически проблеми, въпроси на правописа и морфологичната адаптация.

Особено полезни за участниците в конференцията бяха двете работни заседа-

ния за запознаване с принципните постановки на Инициативата за текстово кодиране (TEI) и стандартизирания генерализиран маркиращ език (SGML). За пръв път пред аудиторията от слависти медиевисти се представи една съвременна универсална компютърна платформа, която е предназначена както за моделиране на текста, така и за метатекстова информация. Бяха изнесени лекции от проф. д-р Дейвид Бирнбаум, д-р Хари Гейлорд от Гронинген, д-р Николас Финк от Синсинати и д-р Винфрид Бадер. В отделно заседание д-р Бадер характеризира текстовообработващата програма TUSTEP.

Изнесените доклади и активните дискусии поставиха началото на интердисциплинарен диалог, чиято цел е да се преодолее както изолацията в заниманията на отделните специалисти, така и препятстващото научния обмен многоезичие на осъществяваните с помощта на компютър филологически изследвания. В заключителните изказвания бе подчертано, че различията в критериите между славистите са преодолими чрез приемането на единна система за предаване и обработка на средновековните текстове. Чрез нея би се улеснило запазването на славянското ръкописно наследство и размяната на информация между учените от различни страни. Общият извод бе, че на прага на 21-то столетие успешното проучване на Средновековието изисква универсални методи и гъвкави средства, съобразени с постиженията както на съвременната филология, така и на съвременната информатика. Безспорен стимул за тяхното усъвършенстване даде наситената работна седмица на конференцията в Благоевград.