

МИНИСТЕРСТВО НА ОБРАЗОВАНИЕТО И НАУКАТА НА РЕПУБЛИКА БЪЛГАРИЯ
КИРИЛО-МЕТОДИЕВСКИ НАУЧЕН ЦЕНТЪР ПРИ БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. М. Т. КАЛАШНИКОВА
НАУЧНОЕ СООБЩЕСТВО “ПИСЬМЕННОЕ НАСЛЕДИЕ”
DIGITAL MEDIEVALIST SCHOLARLY COMMUNITY
ФОНДАЦИЯ „УСТОЙЧИВО РАЗВИТИЕ НА БЪЛГАРИЯ“

**Писменото наследство
и информационните технологии**

El’Manuscript–2014

Материали от V международна научна конференция
Варна, 15–20 септември 2014 г.

София · Ижевск
2014

Сборникът е издаден с финансовата подкрепа на Министерството на образованието и науката на Република България по процедура за подкрепа на международни научни форуми.

Отговорни редактори: проф. дфн В. А. Баранов
 доц. д-р В. Желязкова
 д-р А. М. Лаврентъев

Редактори: Нели Ганчева, Веселка Желязкова (български текст)
 О. В. Зуга, В. А. Баранов (руски текст)
 Кевин Хокинс (Kevin Hawkins) (английски текст)

Писменото наследство и информационните технологии [Текст] : материали от V международна науч. конф. (Варна, 15–20 септември 2014 г.) / отг. ред. В. А. Баранов, В. Желязкова, А. М. Лаврентъев. — София ; Ижевск, 2014. — 448 с.

Сборникът съдържа материали от конференция, посветена на разработването и създаването на съвременни средства за съхраняване, описване, обработка, анализ и публикуване на ръкописни и старопечатни книжовни паметници и исторически извори, а също и на въпросите за подготвянето на електронни ресурси в областта на хуманитаристиката и тяхното използване в научните изследвания и преподаването.

© Кирило-Методиевски научен център — БАН, 2014
© Ижевский государственный технический университет
им. М. Т. Калашникова, 2014
© Авторски колектив, 2014
© Лилия Тошкова — графичен дизайн на корицата, 2014

ISBN 978–954–9787–25–2

Житието на св. Евстатий Плакида и съвременните технологии

Андрей Бобев, Любка Ненова

Колационен софтуер, тестване, старобългарски език, средновековна литература, Житието на св. Евстатий Плакида

The Vita of St. Eustathius Plakidas and Contemporary Software

Andrej Bobev, Lyubka Nenova

This article presents the results of testing two programs: *Compare It!*, for simple comparison of files, and *Juxta*, a specialized program for textual collation. Both are tested for suitability to medieval Slavic texts on a portion from the Vita of St. Eustathius. The results show some basic shortcomings of *Juxta*. Despite the negative result of the test, it allows the authors to gather enough impressions for creating a list of common *desiderata* for creating collation software for Slavic texts in the future.

1. Проблемът

Ключовият проблем за отношението между теория и практика при издаването на средновековни текстове придобива нови аспекти във време на светкавично развитие на техническите средства. От една страна, съвременният филологически софтуер предоставя възможности, които доскоро бяха в сферата на фантастиката, ала създава и нови зависимости. От друга, възможността за обработване на далеч по-сериозни масиви текст, отколкото е било мислимо в миналото, разкрива пропуски в досегашните методики — бели полета, които поставят под съмнение част от досегашните резултати.

В предишен текст, посветен на текстологията на Богородичния акатист, демонстрирахме слабостите от прилаганата досега текстологическа класификационна схема и аргументирахме нуждата от нова методика [Бобев, Ненова 2012]. В настоящия продължаваме наблюденията си, но този път върху агиографски материал.

Житието на св. Евстатий Плакида е било досега обект на отделни проучвания, от които най-многобройни и обхватни са тези на Олга Гладкова. Тъй като част от нейните публикации не са ни достъпни, поставихме си за задача да анализираме не толкова собствено текстологическата история на Житието, колкото възможността тя да бъде проучвана с помощта на съвременните технически средства.

2. Източниците

Използвани са копия от 13 ръкописа. Южнославянските, изброени у Иванова [Иванова 2008], се посочват с дадените от нея съкращения, а именно: от група *1a* по нейната класификация ркп. РМ4/5, УББһ30, МСПЦ103, Пл98 (базов за колацията), а от група *1в* — ркп. НБКМ1039 (базов за колацията), Деч94, МСПЦ139, Рс59. С ръкописи от група *1б* не разполагаме. От единствения южнославянски ркп. от група 2, посочен у Иванова, разполагаме с непълно копие, затова тук той не е включен. Добавяме за съпоставка няколко източнославянски ръкописа: от група *1a* — ркп. ТСЛ663, ТСЛ666, МДА.І.88, от група *1в* — ркп. ТСЛ745, ТСЛ792, F19–329 (Вилнюс).

За проверка на възможностите на софтуерните продукти беше използван началният откъс от Житието в следния обем: *Въ д/ни цр<с>тва Траіана, идолъстѣи жрътвѣ одръжеціи. бѣ же єдинъ нѣкто стратїлатъ, имене<м> Плакіда... — ... бѣ бѣ храбр'рь вел'ми, и лов'ць худогъ. и по в'се весело лове* (цит. по РМ4_5). Цялостно колационно издание на текста по достъпните преписи ще бъде обект на отделна публикация. Задача на настоящата работа е само да тества възможностите на достъпния софтуер.

3. Софтуерът

Първоначалната проверка на достъпния софтуер, който може да се използва за медиевистична съпоставителна работа с различни текстове, показва, че той може да се групира в два типа: 1) *инструменти за проста съпоставка* и 2) *специализирани инструменти за колация*.

Към първия тип спадат буквално десетки софтуерни продукти, разпространявани безплатно чрез Световната мрежа. Измежду тях се спряхме на програмата *Compare It!*, която се отличава с много олекотен изглед и интуитивно управление на функциите.

По-сложен е проблемът с подбора на специализирана програма. Повечето от тях са основани на употребата на програмния език XML и използването им изисква специфични компютърни умения. Това са напр. *CollateX*, *TXSTEP* и др., които въпреки уверенията на създателите им, че са ориентирани към ползвател неспециалист, всъщност са неизползваеми от лица без познание за XML. На страницата на програмата *TUSTEP* дори е упоменато, че работата с нея изисква предварителен период на обучение, продължаващ до няколко месеца. Други програми, напр. *Collate*, не само се основават на HTML/SGML парсиране, но имат и неправдоподобно висока цена. Затова ние се спряхме на програмата *Juxta*, която според описанието е подходяща за потребители неспециалисти, разпространява се безплатно и към нея се предлага упътване за работа. Предстои оглед на възможностите, които предоставя програмата DV-COLL.

Всяка от програмите изисква пълен наборен текст на използвания източник; нито една не предлага използване на отделни въведени варианти (напр. изработване на стема по варианти).

4. Резултатите

Пълен анализ на работата с двата софтуера с презентация както на основните им преимущества, така и на ключовите проблеми, затрудняващи тяхното използване за славистични нужди, ще бъде изнесен при представянето на доклада. Представени накратко, трудностите, които възникват, са следните.

4.1. Представяне на специализираните шрифтове. Въпреки използването на нови шрифтове по стандарта Unicode (*CyrillicaOhrid10U* и *CyrillicaBulgarian10U*), специфичните им знаци като цяло се отчитат от софтуера, но не се изобразяват, което силно затруднява наблюдението на процеса — отстраняването на грешки, четимостта. Напр. формите $\widehat{m\alpha}$ и $\widehat{m\lambda}$ биха изглеждали еднакво: $m\overline{\alpha}$. Наложил се употребата на *Arial Unicode MS*, който съдържа по-важните от специфичните знаци на средновековната кирилица. Наложил се също прилагането на условен модел за предаване на надписаните знаци (буквени и небуквени). Допълнителна трудност възникнала при *Juxta* — при съставянето на критическия апарат специалните знаци се отчитат като словоделител и програмата съпоставя само следходните спрямо тях отрязъци. При това се получават абсурдни резултати, напр. следното (във втория пример като лема е „отсечена“ формата до надписаната буква):

родо(м), /дѣлы же] рw(д)мь, /нѣ дѣли UBVC30.txt
родо] родw RM4_5.txt

4.2. Невъзможност на програмите да се справят самостоятелно със сегментирането на текста в рамките дори на една текстова група. Избраните програми не дават приемлив резултат без предварително ръчно сегментиране на текста от изследователя. Откъсът бе разделен на 27 съпоставими сегмента — при преписите от гр. *Ia* има три празни сегмента, на които съответства текст в групата *Iv*, а в гр. *Iv* има един такъв, на който група *Ia* предлага текст. Дори най-простата съпоставителна работа с програмите изисква предварителни текстологически наблюдения, без които е невъзможно правилното въвеждане на данните от преписите по начин, подлежащ на обработване. Нито една от двете програми не разполага със средства за систематизирано обозначаване на мащабни пропуски в различни клонове на традицията — сериозен софтуерен пропуск.

4.3. При работата в рамките на избраната текстова група — невъзможност за произвеждане на критически апарат. *Compare It!* изобщо не произвежда никакъв апарат — различията се визуализират в работния прозорец чрез оцветяване. *Juxta* произвежда апарат в HTML файлове; той има следните *по-важни* слабости:

1) Показва тенденция към замяната на текстови възли с фразова съпоставка, напр.:

б/лгочѣстин'ными оукрашаше(с) /ал'чющее напиташе] б/лгымь оукрашаше се.

алчющеє питаше. Dec94.txt (7)

При това обозначенията / и (7) (номер на ред) сочат, че е нарушена въведената от нас сегментация.

2) Допуска смесване/обединяване на два възела на разночетене, като не ги разграничава и отчита коректно, напр.:

же~ TSL663.txt, MDAI88.txt

же праведными ...] праведными ... TSL666.txt

Така един текстологически факт се среща в две (понякога в повече) различни бележки, фокусирани върху различни възли/групи възли.

3) Пропуска текстови варианти, което е *недопустимо* за колационен софтуер. Напр. за първия сегмент на гр. *1в* дава 16 възела; от тях следните се отнасят до формата цр(с)тво:

1 Въ цр(с)тво и] Въ TSL745.txt, TSL792.txt

1 цр(с)тво и] съ F19–329.txt

1 ии] и. црт(с)ва TSL745.txt

1 ии] и црт(с)вiа TSL792.txt

1 ии трояна ц/ра /иеще идолослужению дръжещоу] и цр(с)тва. трагияна цр(с)а. /еще идолослужению держашю F19–329.txt

За съпоставка ето резултата при „ръчно“ колиране само за този възел:

цр(с)тво NBKM1039] цр(с)тва Dec94 F19–329 TSL745 Rs59; црт(с)вiа TSL792; цр(с)твiе MSPC139

4.4. За *Juxta* е *невъзможно* съставянето на критически апарат за текстове, принадлежащи към различни текстологически групи — генерираният файл с варианти е напълно негоден.

4.5. Анализът на точността при предаване на вариантите показва, че *Juxta* дава отлични резултати при съпоставяне на само два източника; повишаването на броя на източниците намалява коректността на колацията. Следователно *Juxta* дава коректни резултати за цели, за които е допустимо и използването на проста съпоставителна програма като *Compare It!*, което обезсмисля прилагането на специализиран софтуер.

5. Обобщение

В резултат на досегашната работа можем да обобщим, че наличният колационен софтуер не съответства на нуждите на палеослависта. Авторите обаче оценяват експеримента си като неуспешен, но важен от гледна точка на изследователската работа. Разработването на специализирани инструменти за колацията е задача, чието реализиране ще доведе до значително разширяване на изследователските възможности както за езиковеди, така и за литературоведи. Осмислянето на слабостите на съществуващите софтуерни продукти от гледище на предполагае-

мия палеославист — техен ползвател, се оказва важно, защото позволи да се очертаят специфичните изисквания, които трябва да се имат предвид в бъдеще при подготовката на такъв специализиран програмен продукт. Тези изисквания ще бъдат представени в бъдеща публикация.

Литература

- Бобев, Ненова 2012 — *Бобев, А., Л. Ненова*. Някои наблюдения върху Богородичния акатист в южнославянската традиция // *Славистиката в глобалния свят — предизвикателства и перспективи*. (Сборник с доклади от конференцията, проведена на 05.–06.10.2012 г.). Благоевград, 2012, с. 81–89.
- Иванова 2008 — *Иванова, Кл.* *Bibliotheca hagiographica balcano-slavica*. С., АИ „Марин Дринов“, 2008.