

ИСПОЛЬЗУЮТ ЛИ (ПОЛНОСТЬЮ) СЛАВИСТЫ ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ?

Габор Л. Балаж

(Balázs L. Gábor, Szegedi Tudományegyetem, Szláv Filológiai Tanszék
H–6722 Szeged, Egyetem u. 2.)

На двойной вопрос заглавия одновременно ответить невозможно, поэтому в дальнейшем мы попытаемся изложить свои мысли об этой проблеме по очереди. На вопрос: **используют ли** слависты компьютер, можно с радостью и большей уверенностью ответить: да, все больше, все чаще и все многостороннее. Это безусловно факт, несмотря на то, что полный ответ существенно сложнее, ведь подразумевается и то, что есть и такие коллеги, которые или вообще не пользуются компьютером, или прибегают к нему лишь изредка, для решения отдельных конкретных задач. Только одной возможной причиной этому может быть то, что не у всех имеется компьютер дома, а на местах работы количество вычислительных машин, приходящих на человека, меньше одного, хотя данное обстоятельство само по себе оказывает значительное влияние на формирование настоящего положения. К тому же, надо принять во внимание и трудности приобретения программных обеспечений, юридические вопросы пользования им, а также иногда и недостаток или недоступность специальных целевых программ. Нам кажется, однако, что кроме материальных факторов существуют и иные, не менее значительные аспекты, которые часто не принимаются к сведению, в то время как в известном смысле от этого зависит – по меньшей мере отчасти – будущее исследований и обучения. Дело в том, что даже имеющимися компьютерами не пользуется каждый, кто имеет на это возможность, потому что или он чуждается техники, или не верит в то, что ему вообще удастся усвоить, как пользоваться той или иной программой. Многие до сих пор думают, что для работы с компьютером нужны математические знания, осведомленность в специальных языках программирования, и, пожалуй, в электротехнике. Те, кто хотя бы однажды попробовали, как можно работать с помощью вычислительной машины, легко могли убедиться в том, что это вовсе не так. Настороженность такого типа не является новым явлением в обучении и исследовательской работе, ведь было время, когда эпидиаскопы, проекторы, магнитофоны, а позже видеоаппаратура и лазерные приборы разделили работников исследовательских институтов на две партии. Есть специалисты, которые до сих пор не пользуются этими наглядными пособиями,

вопреки тому, что уже никто не может сомневаться в том, что применяя их можно повышать уровень отдельных лекций. Естественно не случайно, что все эти замечания относятся прежде всего к представителям гуманитарных наук, ведь, например, в кругу работников естественных и медицинских наук картина намного благоприятнее. Вместо чрезмерных обобщений или упрощений, однако, стоит конкретнее и более оптимистически рассмотреть положение пользования вычислительными машинами у славистов. Мы нарочно не будем стремиться к полноте раскрытия данного вопроса, нашей целью является, в первую очередь, привлечение внимания и побуждение к нему, ведь тема лежит на улице...

Что касается тривиальных, рутинных программных обеспечений, к счастью, пользование вычислительными машинами становится все благополучнее как с количественной, так и с качественной точки зрения. Вместо точных статистик здесь мы намереваемся ссылаться лишь на самые общие явления. Например на то, что уже очень многие признали, что применение какой-либо редакторской программы намного удобнее и эффективнее даже самой лучшей пишущей машины, потому что таким образом растрчиваем меньше бумаги, из-за незначительных поправок не приходится заново напечатывать весь текст, на одной дискетке помещается целая книга. Более того, всю рукопись можно послать с помощью электронной почты в любой край света или по нескольким адресам, если это нужно, и таким образом, что адресат через пару минут уже может прочитать ее. За относительно недолгое время можем составить глоссарии, указатели, содержание к существующему документу. И это, конечно, еще не все. Те, кто пользуются какой-нибудь редакторской программой, знают, что кроме перечисленных возможностей в ходе работы можно пользоваться десятками или сотнями удобных функций и практичных услуг. Подобное положение имеется и в случае программ для обработки базы данных или сложных таблиц. В то время как обработка данного текста раньше была немыслима без карточек и картотеки, сегодня можно не только избежать этой операции, но даже желательно заменить ее компьютерной обработкой. Правда, и таким образом приходится как-нибудь ввести данные в машину, но после этого предоставляется возможным систематизирование, добавление, отпечатание, пересылка и т.д. материала, причем на основе любых точек зрения, после любых поправок, и сколько раз угодно, практически без объемных ограничений. К тому же, даже внос данных можно облегчить путем применения в наши дни уже недорогих сканнеров и т.н. программ оптического распознавания текста. А что касается возможностей поиска в документах, закрепленных электронным, т.е. в конечном счете бинарным способом, они уже почти сами по себе представляют достаточную причину на пользование вычислительной машиной, просто потому, что предоставляют удобные, не-

трудные, полностью надежные (практически без ошибок) операции, независимо от того, какое программное обеспечение имеется в виду.

Из ежедневных рутинных действий, совершаемых с помощью компьютерной сети следует припомнить скорее всего электронную почту (*e-mail*) и пользование Интернетом. Первое, к счастью, уже применяется и такими коллегами, которые раньше вообще не пользовались компьютером. Нет надобности особо утверждать преимущества данной услуги, ведь общеизвестно, что *e-mail* является быстрой, удобной и даже бесплатной возможностью. Стоит подчеркнуть, однако, что с помощью электронной почты кроме простой переписки – как мы отметили выше – можно послать адресату (или одновременно нескольким адресатам) большие и маленькие файлы данных, программы, даже звуковые или видеозаписи. Для осуществления последних задач, конечно, имеются и другие способы (FTP, Internet), но безусловно самым распространенным (и самым простым) остается *e-mail*. Об Интернете можно слышать самые разные ошибочные мнения, поэтому важно осознать, что, несмотря на безусловно встречающиеся отрицательные явления, мировая сеть представляет собой, в первую очередь, неисчерпаемый источник, международный форум, который доступен всем, и хотя существуют и платные услуги, он обеспечивает в сущности неограниченное количество знаний, данных, программ, адресов, телефонных номеров и т.д., бесплатно и быстро, для каждого. В наши дни, когда приобретение все более дорогой специальной литературы и межбиблиотечные заказы часто затруднительны или требуют слишком много времени, данное обстоятельство само по себе заслуживает внимания. Такие возможности, как например видеоконференция, известны пока лишь немногим, но это вовсе не означает, что они не могут стать обыкновенным коммуникационным средством в течение нескольких лет, даже для славистов. К такому положению, с другой стороны, надо подготовиться вовремя, ведь развитие технического фона – очевидно – поразительно ускоренное.

На второй вопрос, заданный в заглавии, т.е. **используют ли полностью** слависты компьютеры, нам хочется дать немного более подробный и конкретный ответ. Кроме выше перечисленных наспех рутинных, практически независимых от языкового посредства и темы областей применения, вычислительные машины – если мы снабжаем их надлежащими программами – способны решать значительно более сложные и специализированные задачи или ряд комплексных задач. Именно из-за специфичности задач, однако, в дальнейшем мы попытаемся сосредоточиться на одной – хотя довольно сложной – области исследования: на палеославистике. В ходе обзора, однако, будут затронуты и общие, толкуемые и в более широком ракурсе аспекты.

На основе доступных мне данных,¹ в рамках палеославистики впервые стали пользоваться компьютером в августе 1980 г. в Голландии, в Ниймегенском университете. Группа ученых, управляемая профессорами Груйсом и Костером, к которой позже присоединился и В. Федер, на основе голландской версии программы SDDR, написанной для операционной системы CP/M, разработала систему, подходящую для описания и каталогизации рукописей, обширная документация которой закончилась в 1983 г. В 1986 г. болгарский профессор Денев приготовил доступную для более широкой исследовательской среды разновидность этой программы для MS-DOS, для компьютеров типа IBM ПК, дополнив ее кириллическим, свободно изменяемым шрифтом. В 1987 г. журнал *Полата књигописьная* опубликовала материал запланированной в Ниймегене,² но в конце концов неосуществленной конференции, посвященной международной базе данных средневековых рукописей. В том же году, по инициативе CIBAL, появилось описание другой системы для базы данных, которое вышло в свет также в журнале *Полата књигописьная*. По разным причинам ни одна из упомянутых программ не распространилась среди славистов. К попыткам прошлых двух десятилетий относится инициатива Института русского языка *Машинный фонд русского языка*, первичной целью которой было создание электронного собрания русских текстов (прежде всего с лексикографической целью), но из-за местной платформы и эта программа не могла стать популярной. В наши дни существует немало коммерческих продуктов, предлагаемых специально для палеославистов (TSAR, IS T и т.п.), но, с одной стороны, они являются довольно дорогими, а, с другой стороны, их нельзя назвать удобными для использования. К тому же, в них имеются определенные недостатки. Из более надежных инициатив – опять не претендуя на полноту – мы хотим обратить внимание на три.

Первая инициатива – финская попытка описать морфологию старославянского языка на двух уровнях. Ее разработал Линдстедт, адаптируя модель Коскенниemi для старославянского языка.³ Главными чертами модели являются следующие: 1) она описывает морфологическую систему старославянского языка с помощью правил, действующих на двух уровнях, определяющих соотношение между лексическим и поверхностным представлением данной словоформы, 2) правила не однонаправленные, то есть

¹ *The First International Conference on Computer Processing of Medieval Slavic Manuscripts, 24–29 July 1995, Blagoevgrad, Bulgaria.* <http://www.ceu.hu/medstud/ralph/report.htm>.

² *Полата књигописьная*, December 1987: 17–18.

³ J. Lindstedt, *A Two-Level Description of Old Church Slavic Morphology: Scando-Slavica* 30, 1984: 165–189.

модель одинаково подходит как для образования (порождения), так и для распознавания (анализа), 3) для иллюстрации действия системы Линдстедт написал компьютерную программу, которая на ограниченном наборе лексем, но убедительно показывает функционирование данной модели.

Модель, выбранная нами в качестве второго примера – т.н. программа OCSGen, которая является старославянской имплементацией системы BETA, разработанной в Стокгольмском университете.⁴ Она симулирует исторические фонологические процессы с индоевропейской эпохи до старославянской, в форме хронологически систематизированных правил (rewriting rules). Система при входе принимает индоевропейские словоформы, а при выходе обеспечивает старославянские рефлексы. Эта модель предназначена для облегчения сравнительно-исторических исследований тем, что она способна порождать все фонологически возможные старославянские словоформы, естественно, за существенно более короткое время, чем это можно было бы делать без вычислительной машины.

Третье программное обеспечение является примером того, как можно использовать компьютер для текстологической обработки старых славянских текстов. М. Камулья из Пизы сделала сопоставительный анализ семи славянских псалтырей с помощью программы DBT.⁵ В ходе работы, сделанной действительно эффективной именно при помощи компьютера, проливался свет на многочисленные лексические варианты, которые различаются от текста выбранной образцом Синайской псалтыри. Рассмотрение собранных пар слов по автору подтвердило такое мнение, согласно которому славянскую традицию псалтыри следует считать единой. В связи с программой DBT, разработанной в Институте компьютерной лингвистики в Пизе, Камулья замечает, что она делает возможным невероятно быстрое и точное совершение «операций» на (в данном случае) старых славянских текстах, в реальном времени и интерактивно.

Из вышеупомянутых осязаемо, что в славистических исследованиях пользование вычислительными машинами стало все популярнее и распространнее за прошлые два десятилетия. Нельзя не признаться, однако, что время от времени возникают и определенные трудности. Самой серьезной проблемой является отсутствие унифицированности почти во всех областях. Что касается аппаратных средств, известно, что наряду с наиболее распространенными в Венгрии ПК типа IBM, многие пользуются машинами

⁴ L. Borin, A Computer Model of Sound Change: an Example from Old Church Slavic. *Literary and Linguistic Computing* 3, No. 2, 1988: 105–108.

⁵ M. Camuglia, The Psalter, its tradition and the computer: a new method of textual analysis. *Paleobulgarica/Старобългаристика* XX, No. 1, 1996: 3–13.

иной архитектуры, прежде всего Apple Macintosh и Sun. Часто бывает, что обмен данными между этими компьютерами затруднителен, что дальше осложняется разнообразием операционных систем (MS-DOS, Windows, Mac OS, OS/2, UNIX, LINUX). Нередко даже внутри одной платформы проблематична конверсия данных, переносимость документов, шрифтов, программ. В качестве характерного примера можно сослаться на следующее обстоятельство. Хотя и чрезвычайно распространенная редакторская программа Microsoft Word, и подобно популярная WordPerfect имеют версии для операционной системы Windows, все равно нельзя беспроблематично прочитать текст, написанный с помощью одной программы, на компьютере, где доступна только другая программа, несмотря на то, что теоретически обе поддерживают конверсию туда и обратно. Это особенно так, если в данном тексте одинаково встречаются, например, и кириллические, и латинские буквы. До сих пор исследования немало затруднил тот факт, что нет настоящей координации между славистами, с одной стороны, и специалистами, занимающимися греческой, латинской и еврейской палеографией и кодексами, с другой. Именно поэтому следует считать важной вехой возникновение проекта под названием "Компьютерная обработка старых славянских рукописей",⁶ представленного следующими исследователями: Д. Бирнбаум (Питтсбургский университет), А. Бояджиев (Софийский университет), М. Добрева (Институт математики БАН), А. Милтенова (Институт литературы БАН). Проект основывается на следующих принципах: 1) унифицирование файлового формата документов, 2) многократная применимость (данные независимы от переработки), 3) переносимость электронных текстов (независимость от местных платформ), 4) необходимость сохранения в электронной форме языковых памятников, 5) стремление к структурированному распределению данных в соответствии со славянскими традициями орфографии, текстологии, палеографии и т.д.

В рамках данного проекта осуществилась в июле 1995 г., в Благоевграде (Болгария) конференция, превзошедшая все ожидания, под названием "Компьютерная обработка средневековых славянских рукописей". На пятидневное мероприятие приехали свыше семидесяти участников из одиннадцати стран. Тридцать четыре доклада были распределены в десяти секциях, а именно: I. Состояние исследования, II. Проблемы кодирования, III. Обработка текстов: проблемы, методы и средства, IV. Сравнение сводов, V. Шрифты и графическая обработка, VI. Компьютерная обработка дан-

⁶ Л. Тасева, М. Йовчева, Компьютерна обработка на средновековни славянски ръкописи. *Paleobulgaria/Старобългаристика* XX, No. 1, 1996: 121–127. См. еще: *The First International Conference on Computer Processing of Medieval Slavic Manuscripts, 24–29 July 1995, Blagoevgrad, Bulgaria*. <http://www.ceu.hu/medstud/ralph/report.htm>.

ных из средневековых славянских рукописей путем использования языка SGML, VII. Применение систем для базы данных, VIII. Частные исследования: типы рукописей, IX. Multimedia, X. Программа KLEIO. На конференции участники пришли к общему согласию в том, что т.н. язык SGML (Стандартизированный генерализированный Markup язык, англ. Standard Generalized Markup Language), особенно его применение TEI (см. подробнее об этом ниже), предоставляет адекватные рамки для удовлетворения нужд исследователей, работающих в разных областях, в то время как он обеспечивает стандартизованную модель для членения документов и для кодирования знаков препинания и символов. Несколько специалистов, принимающих участие в значительных проектах кодирования, согласилось о том, что они объединенными усилиями будут стремиться к приспособлению директив TEI к особым требованиям славянских филологических исследований (TEI = Text Encoding Initiative, т.е. Инициатива для кодирования текстов). Другим конкретным результатом конференции было немедленное создание доступного каждому архива FTP для ранних кириллических текстов и связанных с ними электронных файлов (<ftp://ftp.pitt.edu/dept/slavic>), и создание информационной страницы в мировой сети (<http://www.ceu.hu/medstud/obsht.htm>). На этой странице можно найти, между прочим, проекты в связи с кириллическими и глаголическими рукописями, данные бесплатно доступных в электронной форме памятников, библиографические подробности опубликованных в прошлые годы изданий, касающихся данной проблематики, новейшие микрофильмы Хиландарской библиотеки в Огайо, немало полезных адресов по Интернету, местонахождение основных документов и т.д. Можно только пожалеть, что на этой значительной и результативной международной конференции – насколько нам известно – не принимал участие ни один венгерский славист. Несколько смягчает положение, однако, что в представленный здесь проект можно включиться в любое время.

Ввиду важности двух вышеупомянутых понятий (язык SGML и система TEI), нам хотелось бы остановиться на них кратко.

SGML является международным стандартом, определяющим независимые от средств и систем методы для представления текстов в электронной форме. Другими словами, SGML – это метаязык, формально описывающий какой-либо язык, в данном случае т.н. "этикетировочный" язык (markup language). Этикетирование означает снабжение текста такой дополнительной информацией (замечания, комментарии), которая содержит в себе эксплицитные указания в связи с интерпретацией текста. Таким образом, этикетировочный язык является множеством правил маркировки для кодирования текстов, определяющих, что считается допустимым, какая маркировка нужна, и как следует ее разграничивать от самого текста. Здесь важно подчеркнуть независимость от средств и систем, ведь это значит, что

совершенно все равно, на каком компьютере, под какой операционной системой и с помощью какой редакторской программы вводится текст для обработки. Благодаря стандартизированному формату, текст, написанный на языке SGML, можно изобразить на мониторе в готовом для отпечатки виде. Одни редакторские программы, как например последние версии Corel WordPerfect для Windows, содержат удобные возможности изображения и манипулирования текстов в формате SGML. (Microsoft Word, к сожалению, пока не поддерживает стандарт SGML, но ввод таких текстов возможен.)

Данный стандарт был принят и используется вышеупомянутой инициативой для кодирования текстов TEI, разработка которой началась в 1987 г., также в форме международного проекта. Ее возникновению содействовала возникшая у филологов потребность, согласно которой для сложных текстовых структур необходима единая система кодирования. Таким образом предоставится возможным минимизировать многообразие существующей практики кодирования, упростить машинную обработку и стимулировать обмен электронными документами. Скоро выяснилось, что одна достаточно гибкая схема может обеспечить общее решение для проблем кодирования. Для такой цели систему TEI стали распространять таким образом, чтобы она была пригодной для удовлетворения многообразным требованиям практически любой области науки или любого применения. Следовательно, TEI стала единственной систематизированной попыткой разработать всеобщую модель кодирования текстов, подходящей как для обработки и анализа любого текста, написанного на любом языке, так и для обслуживания широкого круга существующих и будущих применений. Подробная документация системы общедоступна, между прочим, по Интернету, в форме так называемых директив (TEI Guidelines). Эти директивы, без ограничений в форме или содержании, действительны для любого текста, написанного на естественном языке, независимо от того, когда написали данный текст, или к какому литературному жанру или типу он принадлежит. Сверх того, эти же директивы относятся и к связным текстам, и к словарным материалам, и к лингвистическим корпусам. Стоит обратить внимание и на то, что правила и рекомендации, сформулированные директивами, соответствуют качественному стандарту ISO 8879, определяющему язык SGML, и стандарту ISO 646, определяющему семибитовый набор знаков для взаимного обмена кодированных текстов. Самыми важными целевыми установками директив TEI являются следующие: обеспечение стандартного формата для обмена данными, указания для кодирования текстов в таком формате, поддержка для кодирования всякого рода свойств изучаемого исследователями всякого рода текста, независимость от программных обеспечений. Кроме этого, важной целью является и стремление к сохранению как можно большей взаимозаменяемости со существующими

стандартами, и поддержка для конверсии уже существующих электронных текстов в новый формат. Представление системы конкретных этикеток для кодирования текстов, а также и описание синтаксиса, определяющего употребление таких этикеток, намного вышло бы за рамки настоящей работы, поэтому здесь не будем заниматься этим. Основная идея, которую мы попытались очертить выше, однако, сама по себе пригодна для возбуждения интереса. Для тех, кто интересуется подробностями, свободно доступна целая система TEI, включительно и описание метаязыка SGML, как об этом уже говорилось.

Мы подготовили этот беглый обзор в надежде, что более точное понимание настоящего положения и возможностей придаст новый размах компьютерной обработке текстов среди славистов Венгрии. Ведь уже с полным правом можно установить, что слависты все чаще и все удачнее используют вычислительные машины, хотя возможности практически неограничены, следовательно, обязанностей в данной области еще много. Машины, конечно, никогда не заменят деятельность креативного человеческого ума, но дело не в этом. Пользоваться компьютером стоит для того, чтобы облегчить механические, продолжительные, иногда монотонные рабочие процессы и одновременно сократить количество ошибок. А результат: единообразная обработка, точность, экономность времени и энергии, стандартизированные методы и форматы (т.е. переносимость), взаимозаменяемость, копируемость, сохраняемость, простая пересылаемость. С другой стороны, нам нечего терять, ведь не нужно отказываться от традиционных, проверенных исследовательских принципов, значит не надо приносить никаких жертв. Несколько лет тому назад – особенно в Средней и Восточной Европе – пользование компьютером считалось редкой привилегией. Сегодня, однако, не можем позволить себе пренебрегать доступными нам средствами информатики или оставлять их невключенными на долгое время. В завершение хочется всем нам пожелать пользоваться вычислительными машинами как можно больше, лучше и результативнее, во славу науки и в назидание самим себе.