

Компютърните технологии и текстологията

Мирослав Янакиев

Софийски университет "Св. Климент Охридски"

ABSTRACT: Computer technology today reveals a potential for specialists in applied linguistics that would have been difficult to imagine some twenty years ago. On the other hand, some of the innovative approaches made possible by advancements in computer technology may prove unfamiliar or difficult to understand to colleagues who are not familiar with these new tools, and with the new possibilities they create. For example, computer-supported textology can facilitate searching and indexing of material of arbitrary length from very large texts, and can do so within an arbitrary context.

A textologist with proper computer support will no longer require the voluminous *Indices Verborum* that were one of the mainstays of our profession in earlier generations. Neither will there be any need to combine lexical forms into lexemes in order to enter them in a common lemmatized form into an *Index Verborum*.

Whatever the current limitations of computer-aided textology, it clearly has a future in providing, through optical scanning and character recognition, for the transformation of any incunabulum or manuscript into a sequence of coded characters suitable for computer processing. And yet the textologist working with a computer should be aware that the machine is indifferent to the transcription or transliteration used — whether it is in Latin script, the Cyrillic script, the Latin phonetic script, the Latin phonemic one, or some other. Regardless of the specific transcription or transliteration employed, the textual information will be stored in the computer memory as a sequence of zeros and ones, and will be available in its internal, machine form for the type of processing described above.

Развитието на компютърната техника разкрива пред човек, който желае да се занимава с научно изследване на текстове, такива перспективи, каквито той до преди двайсетина година трудно можеше да си представи. Развитието на компютърната техника обаче и изисква от филолога текстолог да се опита да развие мисленето си в насока, която едва ли ще бъде приемлива за традиционно мислещите свои колеги.

Каква е тази насока?

За снабдената с компютри текстология е възможно да се осигури за разумно кратко време намирането и изброяването както на съвсем кратки, така и на много

дълги поредици от знакове в много обемни текстове при това в произволно голям контекст.

Поради това снабденият с компютър текстолог ще е разумно да привикне с мисълта, че няма да са му нужни неекологично похабяващите тонове хартия Indices Verborum, както и че няма да е нужно да се труди да обединява в "лексеми" словоформите, за да ги представя в Index Verborum обединени.

Ще се опитам с пример да поясня твърдението си.

Наша колежка от Свободния факултет на Софийския университет "Св. Климент Охридски" в момента подготвя за издаване като книга ръкописен сборник слова на Агапий, монах, живял в Рилския манастир през миналия век. Ръкописът съдържа - приемаме - 200 страници, което означава 32 хиляди текстови думи (монолекси) или 143 хиляди букви.

Да си представим, че в изданието бъде включен Index Verborum. Това ще означава че към изданието да се добавят най-малко още стотина страници, ако искаме в Index'а да включим адресите на всяка словоформа от текста, т.е. обемът на книгата ще трябва да нарасне поне с една четвърт.

А да си представим сега, че ръкописът е записан на дискета. Даже записването да е направено с най-разточителната програма за записване на текстове Word for Windows 6 и без така нареченото компресиране, целият ръкопис ще се съхранява на една четвърт от мястото, което ни представя една обикновена дискета DSHD. Даже при днешните значително завишавани цени на дискетите в цената (себестойността) на един екземпляр от изданието ще се добавят десетина лева.

Ако текстът е записан на дискета, не е нужно да му се съставя Index Verborum. Нещо повече - текстообработващите програми като споменатия Word for Windows 6 предоставят удобства, с каквито индексите не могат да ни обслужват.

Трябва ни например в споменатия ръкопис на Агапий да потърсим даваджи. Извиквате чрез "менюто" Edit командата Find или направо натискате едновременно клавишите Ctrl и F, появява се на екрана специално меню, което ви подканва да наберете търсената дума, набирате я, натискате клавиша Enter миг и на екрана се появява даваджи заедно с контекста си. Искате да потърсите дава. Постъпвате по аналогичен начин и за мигове компютърът ви дава последователно 3 дава с контекстите им: в оутвърждавали, дават и даваджи.

Чрез друга команда - Replace - можете да установите за броени секунди колко пъти в текста се среща съюзът да, буквата д. Чрез лесно реализуеми програмки (макроси) можете да намирате и да изброявате дисперсни буквосъчетания като и . . . е.

Би ли станал нужен специален Index Verborum, ако изданието е на текст, преведен от друг език и включва информация за чуждосичните съответствия в съответствие с текстографическите традиции?

Ще си позволя, като говоря за тези традиции, да напомня за една публикация в *Советское славяноведение*, М., 1970, 6, с. 58-64: М. Янакиев, Н.В. Котова. О некоторых принципах эдичионной палеославистики. Тези принципи, за които говорим в тази статия, остават актуални, а, уви, не се спазват в изданията на древни текстове и до сега. Но едно положение е загубило актуалността си. За него искам да

разкажа, защото е свързано с шеметното развитие на компютърната техника. И пак ще си послужи с пример.

Ето как изглежда първият ред на л. 17 (21) на готвеното от Н. Котова и мене издание на Григоровичевия паримейник:

Притчи Сол., гл 1, 23-30

се прѣдлагаа нлгъ моего дѣха ∴ нн наδυα же вѣγ

idou proihsomai umin emēs pnoēs didaxō de umās

При подготовката на изданието трябваше да се пишат фишове за всяка дума, съдържащи всеки думата в минимален контекст. Всичко над 100 хиляди. И целият този труд днес се оказва ненужен. С проста клавишна манипулация, след като сме набрали в два различни, но свързани файла преводния и гръцкия текст, компютърът е в състояние да ни подаде всяка монолекса или полилекса, всяка морфема (буква или буквосъчетание) от текста на паримейника или от гръцкия файл заедно с всякакъв контекст и другоезичното му съответствие.

И всичко това, така да се каже "безплатно", без всякакъв отделен Index Verborum, без всякакъв отделен морфемарий.

С помощта на така наречените сканери, на компютърна дискета може да се пренесе всяка страница от ръкопис, и то с всичките ѝ цветове отсенки, с всичките ѝ дефекти, а после чрез съответните компютърни програми да се повиши четливостта на поизбледнели или зацапани знакове и на компютърния екран да видим в "освежен" вид древното пѣстроцветие на оригинала. И всичкото това даже и като "под лупа". И всичкото това не само на екранчето на компютъра, но напоследък

(ако компютърът Ви е пуснатият в продажба от IBM от ThinkPad 755CV) с помощта на обикновен шрайбпроектор даже на голям стънен екран.

А сега - едно "лирическо отстъпление".

Отново за цената на тези възможности. Вярно е, че за да се съхранява на дискета едно сканирано цветно изображение е нужно много място и една дискета не стига, ако искаме да направим един голям ръкопис достъпен за компютърния екран и страница по страница ("фототипно"), и в удобна за обмисляне транслитерация с текстологически коментари. Но появата на така наречените CD ROM устройства отпрати в миналото "дискетната теснотия" и едновременно направи единицата компютърен запис десетки пъти по-евтина. Сега няма съмнение, че издаването "в цвят" на ръкописи вече е не само неекологично, то е и неикономично.

Наистина остава като че ли нерешена главната проблема, проблемата за цената на компютрите, без които нито дискетите, нито компакт-дискете могат да бъдат използвани. Няма нужда да говоря за "медицинската" страна на използването на компютрите — не подлежи на съмнение предстоящата "смърт" на електроннолъчевите (телевизионните) компютърни екрани. Бъдещето е на LCD (дисплеите с течни кристали), които не вредят нито с "рентгеново лъчение", нито с "трепкане". Само че и те още повече оскъпяват компютрите.

И тъкмо по тази проблема ще си позволя да споделя някои съображения, които изискват разискване не в средите на икономисти, а в средата на милионите, даже милиардите бъдещи собственици на компютри, част от които са и намиращите се тук сега. Тези съображения непременно ще бъдат оспорвани от производителите на компютри. Даже ми става страшно, като си помисля за огромната мощ на производителите на компютри. Но ще се осмеля да споделя с вас своите мисли, които ще се сторят "опасни" за няколко стотин хиляди

специалисти, ангажирани в производството на компютри в света, но мисля, които са напълно добронамерени за стотици милиони потребители на компютърна техника, а в последна сметка са напълно добронамерени и за ангажираните в производството на компютри.

В една книжка - "Електрониката в помощ на учителя филолог", издадена през 1988 г. аз се опитах да предложа изискванията, които трябва да задоволява един компютър, за да стане пълен заместник на учебната литература и другите учебни помагала, нужни на един ученик в училище и в къщи. Тогава сметката ми беше, че такъв "универсален учебен компютър" не струва повече от стотина долара.

Ако добавим и още най-много около двеста долара за компакт-дисково устройство, трезво погледнато, един портативен компютър с цветен LCD и с встроена CD ROM може да се продава на цена много под триста долара, ако човечеството реши всеки ученик да бъде снабден със свой индивидуален "универсален учебен компютър", защото огромното производство ще намали многократно себестойността му, а оттук и продажната му цена.

Но за да стане това, производителите на компютри ще трябва да се откажат от игрите с "непрекъснатото усъвършенстване", които се играят със скритата цел да се повишават цените на компютрите, като се обявяват "неусъвършенстваните" компютри за "остарели морално", без да са остарели физически.

Преди много години Пентагонът в САЩ обяви конкурс за създаване на нов компютърен HIGH-LEVEL-Language. Родил се езикът ADA. Езикът ADA остана съвсем малко популярен в средите на програмистите. Но аналогичен подход към неминуемата всеобща компютризация на човечеството е напълно приемлив, ако световна организация, например ООН, обяви конкурс за конструиране на нужния

днес за младите, а утре за всички хора на земята “универсален учебен компютър”, с който човечеството безплатно да снабдява всяко дете, когато то започва да се учи.

Приключих с “лирическото отклонение”.

Защо ми беше нужно то? Защото съм сигурен, че световен конкурс за конструиране на “универсален учебен компютър” ще бъде обявен и текстологията ще трябва да даде своя принос за постигането на добър резултат от конкурса.

Какъв според мене може да бъде този принос?

Текстолозите (филолозите, лингвистите), снабдени с компютри, ще трябва да осигурят компютърни технологии, които да направят максимално плавен, не нараняващ психиката на възрастните преход към една за целия свят азбука.

За снабдената с компютри текстология е възможно (но в момента все още не съвсем!) да осигури за разумно кратко време чрез сканиране превръщането на всеки стар печатен и ръкописен текст в удобна за компютърна обработка поредица от знакове. Поради това снабденият с компютър текстолог ще е разумно да привикне с мисълта, че за компютъра е безразлично “в каква транскрипция (транслитерация)” — латинска, кирилска, латинска фонетична, латинска фонематична, . . . ще се пази в компютърната памет текстовата информация.

Един пример. Приведеният ред от Григоровичевия паримейник:

се предлагалъ нмѣ моего дѣха ∴ ни наѣдъ же вѣи

може с просто превключване на така наречените True-Type-фонтове и да се латинизира:

se pryadlagayen imw moego doyha ∴ ni naoychen zhe vwi,

и да се елинизира:

σε προδλαγαиεν ιμω μοεγο δουχα vi ναουτσιεν ζιε βωι

Гръцкият текст:

ιδου προήσομαι υμῖν ἐμῆς πνοῆς διδάξω δὲ ὑμᾶς

също с просто превключване може да се кирилизира:

igoŷ proŷsomaŷ xŷomŷn emŷs pnoŷs rŷsin digŷkco gŷ xŷomŷs

и да се латинизира:

idoŷ proeŷsomaŷ hyumŷn emeks pnoeŷs reŷsin didŷxo dŷ hyumŷs

И може би тъкмо лекотата, с която компютърът транслитерира, ще отстрани постепенно от съзнанието на хората преградите, които им пречат да минат към една азбука, както вече са почти узрели да минат към една валута.