Introducere

Credem că limba română este răsplătită de eforturile de analiză, documentare, păstrare şi publicare ale institutelor de lingvistică şi universităților în mai bine de 100 de ani de cercetare (pentru a remarca numai perioada inaugurată de Haşdeu prin activitatea la dicționarul tezaur). În aceşti ani s-au elaborat şi tipărit dicționare, s-au emis şi dezbătut teorii, s-au constituit puncte de vedere oficiale şi personale şi a fost suficient timp chiar şi pentru contestarea unora dintre ele şi perpetuarea unor dispute.

Între timp, limba română nu a stat nici ea pe loc, iar mijloacele de a studia limba s-au schimbat de asemenea. Dacă, pentru a-i studia evoluția sau pentru a găsi filoane lingvistice încă nedescoperite, atenția cercetătorului rămâne captată în continuare de aspecte de fonologie, sintaxă, semantică, lexicologie, terminologie etc., randamentul şi precizia observațiilor lui creşte dacă face apel la metode de investigare informatică a limbii. De câtva timp accesul la o carte se poate face şi altfel decât ținând-o în mână şi deschizând-o. Şi nu mai e nevoie ca ea să existe în biblioteca de lângă noi ca s-o putem citi. Dintr-o dată a devenit posibil să ne uităm la o carte şi altfel decât parcurgând-o în secvența ei liniară. Rafturile cu fişe de ocurențe ale lexicografilor, care luau ani pentru a fi completate, sunt acum generate automat prin metode de indexare de către programe şi regăsirea unui context se face cât ai clipi...

Dar domeniile lingvisticii computaționale şi ale tehnologiilor limbajului uman au repercusiuni şi de altă natură decât ca metode de cercetare asupra unei limbi. Aplicații de prelucrare a limbajului natural care să deschidă un nou tip de acces la informații pot fi acum concepute. Textul, chiar şi în format electronic, începe să fie privit şi altfel decât ca un şir de caractere sau de cuvinte. Au început să apară metode de a pătrunde în structura lui sintactică şi semantică încât structura şi înțelesul textului să poată fi relevate maşinii şi ea să poată opera cu ele aşa cum operează cu numere, de când a fost ea inventată. Începem să ştim cum să facem maşinile noastre să execute un alt tip de „calcul", mai apropiat de modul nostru de gândire, şi care-şi găseşte originea în text...

Limba română trebuie să ajungă la nivelul de tehnologizare de care se pot mândri astăzi alte limbi intens studiate. Rostul acestei cărți, pe care o dorim prima dintr-o serie, trebuie ataşat acestei ambiții. Ea este scrisă de lingvişti şi informaticieni români care, spre marea noastră bucurie, încep să se înțeleagă din ce în ce mai bine. Este exact ceea ce a urmărit acel grup de constituire a Comisiei de Informatizare pentru Limba Română, când, în martie 2001, s-a reunit pentru prima dată în sediul de pe Calea Victoriei al Academiei Române. Ulterior, această întâlnire a devenit o tradiție prin organizarea anual în Bucureşti, Iaşi şi Chişinău a unor sesiuni de lucru ale unui grup lărgit, care, din acest motiv s-a numit Consorțiu. De doi ani am dorit să invităm la aceste întâlniri şi cercetători aflați la mai mare distanță de noi. Ca urmare, ultimele două întâlniri au căpătat caracterul de ateliere de lucru şi au fost organizate în regim de teleconferință. Am putut asculta astfel glasuri de români care lucrează în universități din America, Germania, Italia, Franța şi Anglia, după cum şi ei ne-au putut urmări pe noi.

Întâlnirea din 3 noiembrie 2006 a Atelierului, a fost găzduită de Biblioteca Facultății de Informatică a Universității „Al.I.Cuza" din Iaşi şi a beneficiat de implicarea MEC în finanțare. Această generoasă contribuție bănească ne-a permis să-i îmbunătățim organizarea, dar mai ales, să tipărim această carte. Îi suntem recunoscători pentru acest ajutor, cu precădere d-nei Veronica Bubulete. Mulțumim totodată participanților la atelier, aflați în sală sau conectați prin Internet, cât şi colectivului de recenzori care ne-au ajutat să îmbunătățim calitatea lucrărilor.

Editorii Iaşi, decembrie 2006

Cuprins