

ATELIERUL DE LUCRU "RESURSE LINGVISTICE ȘI INSTRUMENTE PENTRU PRELUCRAREA LIMBII ROMÂNE"

O descriere referitoare la ce înseamnă de fapt "Resurse lingvistice și instrumente pentru prelucrarea limbii române" găsim în volumul care conține lucrările acestui atelier din anul 2006, apărut la editura Universității Alexandru Ioan Cuza:

"Dacă, pentru a studia evoluția limbii române, sau pentru a găsi filoane lingvistice încă nedescoperite, atenția cercetătorului rămâne captată în continuare de aspecte de fonologie, sintaxă, semantică, lexicologie sau terminologie, randamentul și precizia observațiilor lui crește dacă face apel la metode de investigare informatică a limbii. De câțiva timp accesul la o carte se poate face și altfel decât ținând-o în mână și deschizând-o. Și 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 și altfel decât parcurgând-o în secvența ei liniară.

Rafturile cu fișe de ocurențe ale lexicografilor, care luau ani pentru a fi completate, sunt acum generate automat prin metode de indexare de către programe și regăsirea unui context se face cât ai clipi.

Dar domeniile lingvisticii computaționale și ale tehnologiilor limbajului uman au repercusiuni și 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 și în format electronic, începe să fie privit și altfel decât ca un șir de caractere sau de cuvinte. Au început să apară metode de a pătrunde în structura lui sintactică și semantică încât structura și înțelesul textului să poată fi relevate mașinii și ea să poată opera cu ele așa cum operează cu numere, de când a fost ea inventată. Începem să știm cum să facem mașinile noastre să execute un alt tip de „calcul”, mai

apropiat de modul nostru de gândire, și care-și găsește originea în text.

Limba română trebuie să ajungă la nivelul de tehnologizare de care se pot mândri astăzi alte limbi intens studiate."

În vederea găsirii unor resurse lingvistice, se întemeiază, în martie 2001, în cadrul unei întruniri la Academia Română, Comisia de Informatizare pentru Limba Română. Ulterior a devenit o tradiție organizarea anuală în București, Iași și Chișinău a unor sesiuni de lucru ale unui grup lărgit de cercetători care, din acest motiv s-a numit Consorțiu. În ultimii doi ani s-a dorit invitarea la aceste sesiuni de lucru a unor cercetători aflați la mare distanță și din acest motiv, întâlnirile au căpătat caracterul de ateliere de lucru și au fost organizate în regim de teleconferință.

Evenimentul se adresează cadrelor didactice, cercetătorilor, profesioniștilor, studenților, masteranzilor și doctoranzilor interesați de: realizarea de resurse lingvistice ori lexicografice românești în format electronic, realizarea de dicționare, tezaure, corpusuri, adnotate ori nu,

tehnologii informatice aplicabile limbii române, aplicații în care au fost utilizate tehnologii lingvistice pentru limba română, realizări de lingvistică teoretică în studiul limbii române, aplicabile tehnologiilor informatice. Organizatorii tradiționali ai acestui eveniment sunt: Universitatea "Al. I. Cuza" Iași, prin Facultatea de Informatică (FII-UAIC) și Academia Română, prin Institutul de Cercetări în Inteligența Artificială (ICIA-AR) București și Institutul de Informatică Teoretică Iași (IIT-AR). În acest an, atelierul de lucru "Resurse lingvistice și instrumente pentru prelucrarea limbii române" se va desfășura în cadrul Facultății de Informatică (Corpul C al Universității "Al. I. Cuza") cu participare directă și în sistem de video-conferință, în perioada 19 – 21 noiembrie 2008, urmând ca în decembrie 2008 să apară volumul manifestării la Editura Universității "Al. I. Cuza" Iași. În linia evenimentelor din aceasta serie, tematica ce va fi abordată în cadrul atelierului include următoarele subiecte, lista fiind deschisă:

- * realizarea de resurse lingvistice românești, textuale ori vorbite, în forma originală ori adnotate;

- * realizarea de colecții lexicografice românești în format electronic;
- * tehnologii lingvistice aplicate limbii române și/sau altor limbi;
- * aplicații în care au fost utilizate tehnologii lingvistice pentru limba română;
- * realizări de lingvistică teoretică cu aplicații în tehnologia limbii române;
- * crearea, completarea, adnotarea și utilizarea resurselor și a tezaurelor multilingve, în care una din limbi este limba română;
- * alinierea resurselor multilingve de natură diversă, cu accent pe resursele limbii române;
- * aplicații ale resurselor multilingve alinate în traducerea automată, extragerea de informații, dezambiguizarea sensurilor;
- * web-ul semantic;
- * proiecte de cercetare ce implică dezvoltarea de resurse și instrumente dedicate limbii române.