

Prefață

Teoria limbajelor formale și a automatelor a devenit un domeniu deja clasic și poate fi studiată sub o formă bine structurată, așa cum sunt de exemplu algebra liniară sau geometria plană. Inițiată ca un aparat pentru modelarea limbajului natural, ea capătă rapid un statut separat odată cu lucrările lui Noam Chomsky (anii '50) și apoi Arto Salomaa (anii '60). De remarcat aportul școlii românești de limbaje formale, începând cu Academicianul Solomon Marcus și continuând cu nume bine cunoscute în domeniu (Gheorghe Păun, Dan Simovici, Alexandru Mateescu, Victor Mitrana, Lucian Ilie și alții).

Importanța limbajelor formale constă în principal în domeniile largi de aplicabilitate. Plecând de la prelucrarea limbajului natural (de la care a preluat terminologia) și continuând cu limbajele de programare (structura unui compilator este aproape în totalitate definită pe baza gramaticilor și a automatelor), modele economice, biologice (calcul bazate pe lanțuri ADN sau pe membrane celulare), lista domeniilor de aplicabilitate oferă permanent noi posibilități de investigare bazate pe structuri generative sau de recunoaștere. Analiza complexității unui model de calcul este imposibilă fără o referire la mașinile Turing – modelul teoretic standard al unui computer.

Acestea sunt – pe scurt – motivele existenței încă de la începutul anilor '70 a unui curs de Limbaje formale și automate la secțiile de informatică. Cartea de față este un suport de curs pentru studenții Facultății de Matematică și Informatică din Universitatea București și preia în mare parte forma unui curs mai vechi de Bazele Informaticii, ținut de autor în perioada 1984 – 1990.

Structura lui urmărește în principal clasificarea făcută de Chomsky limbajelor formale, prezentând pentru fiecare din cele 4 clase de limbaje clasice (limbaje regulate, independente de context, dependente de context, limbaje recursiv enumerabile) atât mecanisme generative (gramatici) cât și de recunoaștere (automate) specifice fiecărei clase.

Astfel, după un prim capitol dedicat automatelor finite, cu rol de a obișnui pe studenți cu modalitatea de abordare a textelor formale, în Capitolul 2 sunt studiate limbajele regulate împreună cu principalele teoreme de reprezentare (lema de pompare și Myhill-Nerode).

Capitolul 3 introduce noțiunea de sistem de rescriere, pe baza căruia se definesc

metodele generative și de recunoaștere folosite la definirea unui limbaj. Clasificarea Chomsky stabilește ierarhia standard a limbajelor ca punct de referință pentru studiile ulterioare.

Capitolul 4 este dedicat principalelor operații pe limbaje. Se studiază proprietăți de închidere la operațiile de reuniune, concatenare, stelare (operațiile regulate pe limbaje), homomorfisme, substituții, intersecție, oglindire și amestecare.

Capitolul 5 introduce câteva mecanisme de recunoaștere noi, legate strâns de automatele finite: gsm-uri, translatorii finiți, automate Moore și Mealy. După ce se arată că automatele finite sunt mecanismele care recunosc limbajele generate de gramaticile regulate, ideea se extinde asupra lui \mathcal{L}_2 – unde se arată că automatele stivă sunt mecanismele de recunoaștere duale gramaticilor independente de context.

Datorită importanței deosebite pe care o are această clasă de gramatici, ei îi este dedicat Capitolul 6 – cel mai extins capitol. Sunt introduse noțiuni noi: arbore de derivare, gramatici proprii, forme normale (Chomsky și Greibach), teoreme de caracterizare.

În Capitolul 7 este făcut un studiu asupra limbajelor dependente de context și a automatelor liniar mărginite. De remarcat că spre deosebire de alte lucrări care tratează această clasă de acceptori ca un caz particular de mașină Turing, am considerat că *LBA*-urile sunt destul de importante pentru a fi construite și analizate separat.

Ultimul capitol se referă la limbajele recursiv numărabile, acceptate de mașinile Turing sau – dual – generate de gramaticile de tip 0.

De menționat că prin acest curs nu se acoperă în totalitate studiul limbajelor formale și al automatelor. Astfel, nu se menționează nimic despre structuri bine dezvoltate cum sunt gramaticile matriciale, indexate, contextuale (cu o mențiune specială pentru gramaticile contextuale Marcus), sisteme de gramatici, sisteme Lindenmayer, *AFL*-uri, sisteme Post, translatorii, modele biologice (automate Watson-Creek, *P* - sisteme¹) sau despre noțiuni cum sunt vectorii Parikh, operații de ștergere și adăugare, studiul cuvintelor infinite etc, despre care s-ar mai putea elabora alte câteva cursuri.

Prin această carte am intenționat o abordare a elementelor de bază referitoare la cele două domenii. Aceste noțiuni fundamentale sunt necesare atât pentru înțelegerea mecanismelor de lucru folosite în alte discipline care au fost sau vor fi studiate ulterior (arhitectura sistemelor de calcul, elemente de complexitate a calculului, construirea de compilatoare, prelucrare de limbaj natural) precum și o deschidere a interesului de a se alătura colectivelor de cercetare din acest domeniu, colective existente în cadrul Facultăților de Matematică și Informatică din principalele centre universitare românești.

Autorul

¹Litera '*P*' vine de la numele lui Gheorghe Păun.