# Ştiinţele cognitive

Raluca Budiu — mailto:raluca.budiu@cs.cmu.edu
http://www.cs.cmu.edu/~ralucav
Mihai Budiu — mailto:mihaib+@cs.cmu.edu
http://www.cs.cmu.edu/~mihaib/

ianuarie 2000

**Subiect:** relaţia dintre ştiinţele cognitive şi calculatoare

**Cunoştinţe necesare:** idei vagi despre inteligenţa artificială

**Cuvinte cheie:** metaforă, sens literal, model cognitiv, inteligenţa artificială

## Cuprins

# 1 Ştiinţe cognitive — Definiţie şi privire de ansamblu

Acest articol îşi propune să dea o idee despre ce sunt ştiinţele cognitive. Pentru că o discuţie exhaustivă despre ce sunt ele nu îşi are locul aici (şi de altfel depăşeşte capacităţile autorilor), vom încerca să prezentăm un mic studiu de caz, despre abordarea unei probleme din înţelegerea limbajului natural, şi anume înţelegerea metaforelor, presărat pe ici pe colo cu mici consideraţii generale despre metoda şi specificul acestei discipline.

"Ştiinţe cognitive" este traducerea românească a termenului englezesc "cognitive science". Întîmplator, pluralul din numele românesc (construit prin analogie cu "ştiinţe sociale" sau "ştiinţe economice") se potriveşte mai bine — ştiinţele cognitive sunt rezultatul încercării de a pune la un loc munca de cercetare din mai multe domenii distincte: psihologie, inteligenţă artificială, lingvistică, filozofie, ştiinţe neuronale ("neuroscience" — a nu se confunda cu neurologia), educaţie şi antropologie[1]. Istoria disciplinei începe din 1976, o dată cu fondarea jurnalului "Cognitive Science". Trebuie spus că ştiinţele cognitive se suprapun în măsură destul de mare cu psihologia cognitivă; în acest articol o să studiem mai ales intersecţia dintre cele două, şi mai puţin diferenţele. Să spunem totuşi că una din diferenţe este accentul diferit pe care cele două îl pun pe experiment — în timp ce psihologia cognitivă este fundamental înrădăcinată în colectarea de date experimentale de la subiecţi umani, ştiinţele cognitive se folosesc mult de simulări pe calculator.

## 1.1 Metoda

Dar destul cu precizările de terminologie! Cu ce se ocupă deci ştiinţele cognitive? După cum puteţi ghici, cu studiul comportamentului şi inteligenţei umane, mai exact cu construirea unor teorii care să explice diverse aspecte ale acestora. În mod ideal, aceste modele sunt concretizate într-o simulare care se comportă, în condiţii similare, asemănător cu oamenii. Ştiinţele cognitive folosesc ca unealtă fundamentală metoda procesării informaţiei, împrumutată din informatică. De exemplu, teoriile sunt adesea exprimate în termeni de scheme logice (flowcharts). Ele nu încearcă sa explice în detaliu cum se întîmplă procesele în creier, unde sunt (dacă sunt) localizate[2].

Deseori, informaţia este manipulată la nivel pur simbolic: aşa cum atunci cînd scriem un program într-un limbaj de nivel înalt, ca C sau Pascal, nu ne punem problema să descriem exact reprezentarea la nivel de bit a datelor, ci folosim tipuri de date prefabricate, nici ştiinţele cognitive nu îşi propun să descrie cum este reprezentată informaţia la nivel de neuroni. Mai mult, aşa cum este greu de presupus că analizînd fiecare bit manipulat de un program vom înţelege comportamentul programului, metoda procesării informaţiei presupune că este greu de înţeles comportamentul minţii omeneşti în lipsa unui nivel de abstracţie simbolic. Această analogie cu calculatorul a bîntuit multă vreme domeniul (hrănită fiind şi de idei din inteligenţa artificială care s-au dovedit fructuoase pentru ştiinţele cognitive), şi a dat naştere teoriilor bazate pe sisteme de producţie (adică pe condiţionări de tipul "dacă...atunci..."). Deşi folositoare, analogia cu calculatorul a fost deseori privită cu oarecare nesiguranţă, creierul omenesc fiind un lucru atît de diferit fizic de un calculator.

[1] Aceste discipline intră în subtitlul jurnalului *Cognitive Science.*, însă nu toate sunt egal reprezentate; principalele contribuţii în domeniu vin din psihologie şi inteligenţă artificială

[2] Nu este nici o ironie la mijloc — teoria "frenologică" a localizării funcţiilor cognitive în diverse arii ale creierului primeşte din ce în ce mai multe şi mai serioase lovituri.

### 1.2 Conecţionism

Aceasta neîncredere este manifestată cu preponderenţă în curentul ştiinţelor cognitive numit *conecţionism*, care susţine că procesele cognitive trebuie să fie explicate în termeni de interacţiuni între elemente neuronale interconectate între ele. Un pas important în evoluţia conecţionismului a fost marcat de apariţia în 1986 celor două volume "Procesarea distribuită paralelă: explorări în microstructura cogniţiei" editate de James McClelland şi David Rummelhart. Teza cărţii era că teoriile simbolice ale cogniţiei sunt caracterizări aproximative ale ei, şi că mintea omenească poate fi înţeleasă numai cu ajutorul unor modele neurale detaliate.

În ultima vreme, cercetătorii din ştiinţele cognitive tind spre o cale de mijloc: în loc să vadă metoda procesării informaţiei şi conecţionismul ca pe doua soluţii antagoniste, angajate într-o competiţie, le privesc ca fiind complementare. Teoriile abstracte pot cîştiga din încercarea de a explica cum simbolurile pe care le manipulează pot fi implementate neural, iar cele conecţioniste trebuie să încerce să avanseze spre modelarea unor aspecte cognitive mai complexe, identificate de teoriile abstracte.

În cele ce urmează, vrem să exemplificăm metodele cercetării în domeniu cu descrierea evoluţiei unor teorii despre înţelegerea metaforelor. Vom evidenţia un lucru poate mai puţin evident celor cu formaţie de informaticieni, şi anume rolul proeminent pe care colectarea datelor experimentale îl joacă în teoriile despre cogniţie. Este şi firesc să fie aşa, dacă scopul este să înţelegi mintea omenească. Vom încerca apoi să precizăm ce aduce în plus un model computaţional faţă de o teorie verbalizată (adică descrisă doar în proză), care sunt condiţiile pe care trebuie să le îndeplinească, şi cum se poate ataca sarcina creării unui astfel de model; apoi, în acest context, vom discuta un pic noţiunea de teorie unificată a cogniţiei.

## 2 Despre metafore — modele şi date experimentale

Deşi ne-am obişnuit sa asociem noţiunea de "metaforă" de orele de limba română şi de poezie, metaforele sunt un fenomen lingvistic ubicuu. Putem lua drept exemplu următoarele expresii comune, unele deja intrate în limbă, precum "piciorul mesei", "mărul lui Adam", "a rupe un contract", "a i se sfîşia inima", "a lupta cu inflaţia", "limbă de lemn", etc.

Pe lîngă aceste metafore încetăţenite, vorbitorii unei limbi de multe ori inventează altele noi în conversaţii obişnuite. Sintagme precum "păunul a ieşit din nou la plimbare" referindu-se la o persoană ţanţoşă şi ferchezuită sau "slujba meu e o puşcărie" sau "porcul n-a vrut sa-mi dea banii înapoi", pot fi uşor înţelese de către oricine, deşi sunt construcţii metaforice.

Metafora se referă la explicarea unui domeniu sau concept prin evocarea altuia. De exemplu, cînd spunem "timpul este bani", folosim o metaforă în care explicăm un anumit aspect al "timpului" folosindu-ne de un alt concept ("bani"). În această metaforă, "timpul" este *referentul* (adică cel despre care se vorbeşte, "ţinta" metaforei), iar "banii" sunt *vehiculul*, prin care referentul este explicat.

### 2.1 Ipoteza lui Searle

Întrebarea la care vrem să răspundem: cum fac oamenii sa priceapă că e vorba de metafore în cazurile de mai sus? Cum de nu interpretează aceste propoziţii în sens propriu, cuvînt cu cuvînt (sensul propriu al propoziţiei este numit şi *sens literal*)? De ce nu îşi imaginează un păun pe stradă?

O teorie foarte simplă şi foarte celebră, aparţinînd lui Searle, zice că creierul procesează o propoziţie cu metafore în 3 paşi:

1. Construieşte o interpretare *literală* a propoziţiei;
2. Dacă interpretarea literală are sens, stop: se declară mulţumit;
3. Dacă interpretarea literală **nu** are sens, sau este neadevărată, creierul caută o interpretare metaforică.

După această teorie, cînd auzim "porcul n-a vrut să-mi dea banii", ne gîndim mai întîi la un porc în carne şi oase, dar, pentru că ştim că porcii nici nu deţin şi nici nu împrumută bani, ne dăm seama că e vorba de o metaforă.

Teoria este foarte simplă şi plauzibilă, şi, din fericire, uşor de verificat. Dacă teoria ar fi adevărată, atunci oamenii ar trebui sa aibă nevoie de mai mult timp pentru a înţelege o propoziţie metaforică, decît una literală. De ce? Foarte simplu: pentru că pentru o propoziţie literală, din algoritmul de mai sus nu trebuie executaţi decît primii doi paşi, pe cînd, dacă propoziţia este metaforică, cel care o înţelege va trebuie să execute şi pasul 3, necesitînd mai mult timp.

## 2.2 Infirmarea teoriei

Ca urmare, psihologii au proiectat mai multe experimente care să verifice dacă înţelegerea propoziţiilor metaforice ia mai mult timp. O aproximare bună a timpului de înţelegere este timpul de citire a unei propoziţii, care poate fi uşor cronometrat cu ajutorul unui calculator.

Un prim experiment care a arătat că oamenii înţeleg metaforele la fel de repede ca şi propoziţiile literale aparţine lui Ortony, Schallert, Reynolds & Antos (1978). Ei au arătat subiecţilor propoziţii precum "găinile cotcodăceau zgomotos", precedate fie de un pasaj despre nişte păsări de curte la o fermă, fie de un pasaj despre o întîlnire a unei organizaţii de femei. Contrar predicţiilor lui Searle, timpul de înţelegere a propoziţiilor în context metaforic nu a fost diferit de timpul de înţelegere a aceloraşi propoziţii în context literal. Acest experiment a fost considerat drept o bilă neagră pentru teoria lui Searle.

## 2.3 Confirmarea teoriei

Cu aceasta însă fenomenul cognitiv al metaforelor nu a fost dat uitării. În 1990, Gibbs a făcut un experiment, în care subiecţii au văzut un pasaj despre un meci de box, apoi au citit fie o propoziţie literală ("atletul nu a venit la meci"), fie una metaforică ("batoza nu a venit la meci") — unul din boxeri era descris de pasaj ca fiind grăsun. Metaforele ca cele din a doua propoziţie sunt de un gen special ("batoza"): ele se numesc *anaforice*, pentru că referentul metaforei (boxerul) nu este prezent în propoziţia curentă ci apare doar în pasajul anterior.

Măsurătorile subiecţilor umani au arătat că, de data asta, înţelegerea propoziţiei metaforice lua mai mult timp decît înţelegerea celei literale, sugerînd ca teoria lui Searle ar putea fi adevărată.

Cum se împacă acest experiment cu cel al lui Ortony? Principala diferenţă intre cele două este că, deşi amîndouă studiile folosesc metafore anaforice, propoziţiile anaforice din Ortony et al. au de fapt un *şi* sens literal. Cu alte cuvinte, o propoziţie ca "găinile cotcodăceau zgomotos", luată izolat, are un sens perfect limpede, literal. Numai *contextul* în care apare ne poate convinge dacă ea este metaforică sau nu. Textul cu batoza nu poate fi însă interpretat *decît* metaforic.

## 2.4 Concluzii contradictorii

Un alt studiu s-a ocupat de metafore de tipul "Unii A sunt B" — de exemplu: "unele slujbe sunt puşcării" sau "unele drumuri sunt şerpi", etc. Acest studiu a urmat o cale diferită pentru a arăta că teoria lui Searle ar putea fi greşită. Autorii — Glucksberg, Glidea & Bookin (1982) — au pus subiecţii să judece două feluri de propoziţii: (1) propoziţii ca cele de mai sus şi (2) propoziţii fără sens ("unele drumuri sunt puşcării", "unele slujbe sunt şerpi", "unele mere sunt pere", etc.). Subiecţii trebuiau să se pronunţe dacă propoziţiile sunt adevărate sau false **doar în sens literal**. Autorii au cronometrat timpul în care astfel de judecăţi au fost făcute.

Rezultatul a arătat că propoziţiile fără sens au fost clasificate mai repede decît cele cu sens metaforic. Din moment ce nici propoziţiile fără sens, nici cele metaforice nu aveau sens literal şi singura diferenţă între ele era existenţa (sau nu) a unui sens metaforic, autorii au argumentat că oamenii nu pot evita să înţeleagă sensul metaforic al unei propoziţii. Motivul pentru care propoziţiile metaforice iau mai mult timp este că interpretarea metaforică este instantaneu şi obligatoriu calculată. Indirect, ei au dedus astfel că sensul literal nu este accesat înaintea celui metaforic (dacă ar fi fost, atunci n-ar fi avut sens să se continue cu calcularea sensului metaforic).

## 2.5 Despre metoda experimentală

Sper că nu sunteţi plictisiţi de această înşiruire de experimente şi ipoteze contradictorii. Deşi o ştiinţă aparent atît de empirică precum psihologia ar putea să dea în afară impresia că teoriile ei sunt arbitrare şi lipsite de orice certitudine, lucrurile nu stau de loc aşa.

Experimentele concepute de cercetători pentru a verifica o teorie sunt adesea extrem de ingenioase, iar concluziile care pot fi extrase se bazează pe raţionamente riguroase, şi deloc banale.

Putem fi dezamăgiţi de faptul că aparent tot ce putem face este să *infirmăm* teorii false, dar niciodată să demonstrăm adevărul unei teorii. Evidenţa faptelor experimentale pentru o teorie nu garantează adevărul ei. Dar, aşa cum observă filozoful Karl Popper, acestea sunt de fapt trăsăturile generale ale tuturor ştiinţelor naturii, inclusiv fizica: chiar teorii care sunt suportate de extrem de multe fapte experimentale, şi care dau predicţii spectaculoase pentru fenomene neobservate sau ne-explicate (cum a fost de pildă teoria gravitaţiei a lui Newton), pot fi, şi vor fi, eventual, infirmate şi înlocuite de o teorie care aproximează mai bine adevărul (aşa cum a făcut teoria lui Einstein despre gravitaţie).

Chiar şi în teoria informaţiei, se defineşte cantitatea de informaţie prin ceea ce ştim că *nu este adevărat*, şi nu prin ceea ce ştim că poate fi adevărat. Aşa că acest joc minuţios, de a construi castele de cărţi de joc prin teorii aparent efemere, de fapt este o succesiune de paşi (mai mari sau mai mici) pentru o mai bună cunoaştere a realităţii.

## 2.6 O altă teorie de procesare a metaforelor

Revenind la oile noastre, în răspuns la experimentul lui Glucksberg et al., un teoretician (Dascal) a arătat că de fapt este foarte posibil ca autorii experimentului sa fi tras o concluzie greşită. Una din regulile comunicării între oameni este că interlocutorul nu spune lucruri fără sens. Ca atare, după Dascal, este posibil ca subiecţii să fi urmat această regulă cînd au citit propoziţiile metaforice de tipul "unele slujbe sunt puşcării". Cu alte cuvinte, în virtutea regulilor comunicării, subiecţii au parcurs toţi cei trei paşi din modelul lui Searle, şi au găsit un sens metaforic, însă, amintindu-şi în final de cerinţa de a judeca propoziţiile numai în mod literal, au avut nevoie de un timp suplimentar pentru a realiza că propoziţiile metaforice sunt false din punct de vedere literal.

Pe de altă parte, pentru propoziţiile fără sens, subiecţii au parcurs toţi cei trei paşi din algoritmul lui Searle, însă, pentru că nu au găsit o interpretare metaforică pe care să trebuiască sa o rejecteze, au fost mai rapizi. Cu alte cuvinte, Dascal a argumentat că este posibil ca subiecţii din experimentul lui Glucksberg sa fi acţionat după următorul algoritm:

1. Se construieşte o interpretare literală a propoziţiei;
2. Dacă interpretarea literală are sens, se verifică dacă ea este literal adevărată:

   Dacă da, răspunsul la test este "adevărat"; stop;

   Dacă nu, continuă;
3. Dacă interpretarea literală nu are sens, sau dacă este neadevărată, se caută o interpretare metaforică;
4. Dacă nu se găseşte nici o interpretare metaforică, răspunsul este "fals";
5. Dacă se găseşte o interpretare metaforică, este interpretarea metaforică adevărată şi în sens literal?
6. Dacă nu este, răspunsul este "fals";
7. Dacă este, răspunsul este "adevărat".

Acest algoritm poate părea ineficient — adică, dacă ştii că o interpretare este literală, de ce sa mai testezi dacă este literal adevărată? De fapt, ideea este că interpretarea unei propoziţii se caută automat şi "subconştient" — pentru mintea noastră nu există noţiunea de "interpretare literală" sau "metaforică" (de multe ori nici nu ne dăm seama că o propoziţie este metaforică, ca în exemplele date la începutul acestui text). Ca atare, atunci cînd s-a calculat o interpretare, trebuie sa se testeze dacă ea răspunde cerinţelor situaţiei, în cazul nostru dacă interpretarea este literal adevărată.

Aţi înţeles deci de ce, după Dascal, este firesc ca propoziţiile metaforice să ia mai mult timp pentru a fi rejectate decît cele fără sens: pentru că cele fără sens, neavînd nici o interpretare, pot fi clasificate imediat ca fiind false, în timp ce pentru cele metaforice, subiecţii trebuie să execute paşii suplimentari 4 şi 5, necesari rejectării interpretării.

## 2.7 O teorie infirmată

Polemica nu se opreşte însă aici. În răspuns la teoria lui Dascal, Keysar (1989) a proiectat un experiment ingenios; el a manipulat valoarea de adevăr literală şi metaforică a propoziţiilor expuse subiecţilor.

De exemplu, Keysar a arătat participanţilor la experimentul său un pasaj despre un magician la circ, Ion, care era foarte priceput în a face rost de bani. În cazul ăsta, propoziţia "Ion este un magician" este adevărată în ambele sensuri, atît metaforic cît şi literal. Dacă însă pasajul era modificat astfel încît Ion era total nepriceput să strîngă bani, propoziţia "Ion este un magician" devenea adevărată literal, însă falsă din punct de vedere metaforic. Dacă însă pasajul era despre un cercetător care era foarte priceput la a face rost de bani, propoziţia "Ion este un magician" era adevărată numai din punct de vedere metaforic. Într-un context în care Ion era un inginer nepriceput în a face rost de bani, propoziţia cu pricina nu era adevărată nici în sens literal, nici metaforic.

Predicţia teoriei lui Dascal este că propoziţiile adevărate literal vor fi judecate mai repede decît propoziţiile adevărate metaforic sau complet false. Rezultatul experimentului lui Keysar a infirmat însă această predicţie — în mod surprinzător, propoziţiile cu valori de adevăr congruente (adevărate sau false atît din punct de vedere literal cît şi din punct de vedere metaforic) au fost judecate mai repede decît cele cu valori de adevăr incongruente (adevărate dintr-un punct de vedere şi false din celălalt). În felul acesta, Keysar pare să demonstreze că sensul metaforic este accesat în paralel cu cel literal, sau oricum, nu după cel literal.

După cum vedeţi, psihologia cognitivă, deşi se foloseşte de concepte precum scheme logice sau algoritmi ca să descrie diverse teorii, pare destul de bine echipată ca sa răspundă întrebărilor proprii. Chiar şi cu aşa mijloace "sărace" precum experimentarea şi logica elementară, este capabilă sa ajungă la nişte concluzii. Atunci care este rolul calculatorului în toată povestea asta — ce aduce el nou?

## 3 Modele computaţionale

Problema modelelor discutate anterior este că, deşi ele pot face predicţii generale, o mulţime de detalii nu sunt luate în considerare. De exemplu, oamenii au nevoie de cîteva secunde ca să înţeleagă o propoziţie metaforică. La nivel teoretic se poate cu uşurinţă imagina o teorie foarte complicată, însă nimic nu dovedeşte că acea teorie poate să "fie executată" suficient de repede de un mecanism (creier, model cognitiv) real.

De asemenea, în teorie este uşor de folosit într-un algoritm o operaţie de genul "caută o interpretare metaforică", dar de fapt ce înseamnă acest proces de căutare este lăsat în umbră. Cum se reprezintă textul care precede propoziţia în memorie, cum se reprezintă cunoştinţele generale în memorie, care din ele sunt accesate şi în ce ordine, cum este determinat referentul pentru o metaforă anaforică, toate acestea sunt întrebări rămase deschise. Ceea ce calculatorul aduce în acest domeniu este rigoarea — un model computaţional cît de simplu obligă la afirmarea explicită a asumpţiilor (şi simplificărilor) pe care se bazează modelul şi la explicarea tuturor detaliilor care nu au fost incluse în acele asumpţii. De asemenea, în măsura în care poate produce predicţii concrete (procentaj de răspunsuri corecte şi timpi de procesare), un model computaţional este verificabil.

Dar ce înţelegem prin "model computaţional"? Pur şi simplu un set de structuri de date (numite uneori şi "reprezentări") şi un algoritm care operează cu ele. De multe ori, astfel de modele computaţionale se implementează în cîte un "limbaj" specializat, adică se exprimă în termenii unei teorii cognitive mai generale. Astfel de teorii sunt fie bazate pe reţele neuronale, fie pe sisteme de producţie. Diferenţa fundamentală între cele două constă în nivelul de abstracţie la care încearcă să rezolve problema. Sistemele de producţie manipulează simboluri şi nu încearcă neapărat să explice *cum* se implementează aceste simboluri la nivel neuronal; ele declară mecanismele cognitive un fel de “software” care este independent de “hardware”-ul pe care se execută; pe de altă parte, preoponenţii reţelelor neuronale pornesc de la un nivel mai scăzut şi susţin că nu se poate atinge o înţelegere deplină a cogniţiei făcînd abstracţie de specificul “hardware”-ului pe care ea funcţionează – creierul omenesc.

Să vedem cam ce ar însemna un model computaţional pentru înţelegerea metaforelor. În primul rînd se pune întrebarea cum se reprezintă sensul unui cuvînt. Aici avem mai multe opţiuni — putem sa îl reprezentăm fie localizat, adică concentrat într-o singură locaţie în memorie (numită uneori “chunk” în literatura de specialitate), fie distribuit, adică format din mai multe astfel de “chunk-uri”, corespunzînd fiecare cîte unei trăsături a cuvîntului; între chunk-uri sunt legături care reprezintă relaţiile dintre ele. Aceste două tipuri de reprezentare sunt uşor accesibile unui programator — cea localizată corespunde unei singure structuri (ca un `struct` în limbajul C), iar cea distribuită corespunde unui arbore, a cărui rădăcină este sensul cuvîntului şi ale cărui frunze sunt trăsăturile lui individuale. Muchiile pot fi structuri care să conţină cîte un pointer la cele două vîrfuri (vedeţi figura 1[3]).

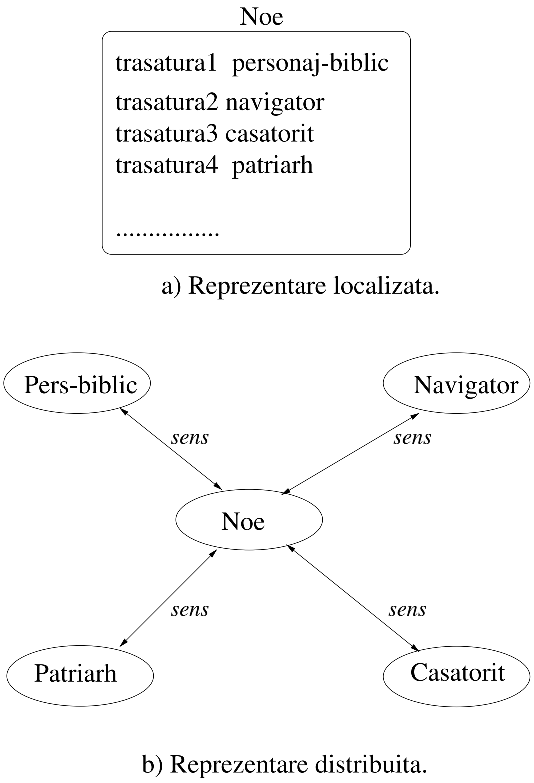


Figura 1: Reprezentarea sensului cuvintelor se poate face în două feluri: fie localizat, în care un sensul unui cuvînt este o colecţie de trăsături, fie distribuit, în care un sensul este reprezentat printr-o relaţiile cu trăsăturile sale. In figură vedeţi reprezentarea localizată (figura a) sau distribuită (figura b) a sensului cuvîntului “Noe”.

[3]Figura nu prezintă arborele desenat in stilul tradiţional; rădacina arborelui trebuie considerată a fi “Noe”

Cum reprezentăm o propoziţie? De exemplu, ca pe un arbore cu muchii etichetate, ca în figura 2: propoziţia este rădăcina arborelui, iar frunzele sunt cuvintele (mai exact sensurile cuvintelor din propoziţie). Muchiile sunt reprezentate ca mai înainte, cu diferenţa că structura corespunzătoare conţine un nou cîmp cu rolul tematic pe care sensul îl are în propoziţie (exemplu: agent, pacient, etc.).

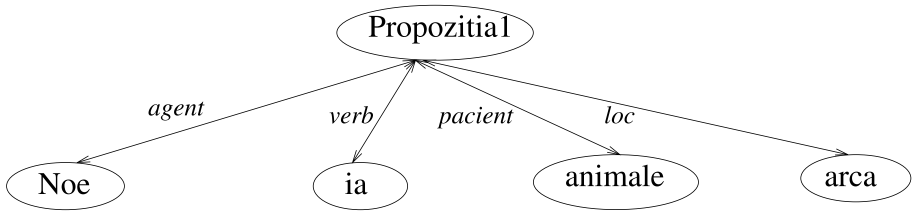


Figura 2: O propoziţie poate fi reprezentată ca un arbore în care frunzele sunt părţile componente ale propoziţiei, iar muchiile indică rolul tematic al fiecărui cuvînt. Fiecare frunză este un sens al unui cuvînt, care poate fi reprezentat în oricare din formele din figura 1.

Modelul trebuie să aibă şi o reprezentare a cunoştinţelor generale despre mediul înconjurător; acestea pot fi reprezentate de pildă ca o colecţie de propoziţii independente, existente în memoria de termen lung.

### 3.1 Modele cognitive şi Inteligenţa Artificială

Cu o astfel de reprezentare, pentru a implementa teoria lui Searle (sau o alta alternativă), trebuie explicat ce înseamnă "a căuta o interpretare pentru o propoziţie". Mai întîi, putem presupune că o interpretare a unei propoziţii este o experienţă anterioară "analoagă" acelei propoziţii şi existentă în mulţimea de cunoştinţe generale. Cu alte cuvinte, înţelegem ceva în funcţie de experienţe similare pe care le-am avut. Acum, cu aceste structuri de date pe care tocmai le-am definit, o interpretare pentru o propoziţie este un arbore din baza de cunoştinţe generale care să se "suprapună" peste arborele propoziţiei pentru care căutăm o interpretare. Suprapunerea poate fi mai mult sau mai puţin exactă. De exemplu, putem presupune că pentru o interpretare literală, suprapunerea trebuie să fie mai riguroasă decît pentru o interpretare metaforică. De aici putem construi cu uşurinţa un model (adică un program) care să urmeze teoria lui Searle.

Poate aţi recunoscut în aceste reprezentări noţiunea de reţea semantica din inteligenţa artificială. Într-adevăr, multe din încercările iniţiale din inteligenţa artificială au fost preluate de către ştiinţele cognitive. Diferenţa fundamentală este că ştiinţele cognitive îşi propun pe de o parte să construiască modele (programe) care să facă ce face omul, aşa cum face omul. În plus, nefiind constrînse să construiască sisteme cu aplicaţii industriale (deocamdată), ştiinţele cognitive îşi pot permite luxul unor asumpţii simplificatoare şi pot sparge astfel o problemă complicată în mai multe bucăţele. De exemplu, un program pentru înţelegerea metaforelor, ca să fie pe deplin funcţional, ar trebui să aibă în interior un algoritm de analiză sintactică, care să indice rolul sintactic al fiecărui cuvînt. Noi însă putem porni de la premiza că rolul sintactic este dinainte calculat[4].

Pe de altă parte, modelele cognitive au constrîngeri suplimentare: ele sunt de pildă constrînse să dea răspunsuri într-un timp comparabil cu cel pe care oamenii îl iau ca să îndeplinească aceeaşi

[4]Aceasta este o premiza în spiritul controversatei idei a modularităţii limbajului.

sarcină.

Deci nu orice program este un model cognitiv. Pe lîngă constrîngeri de timp, un model computaţional este constrîns de tot felul de date empirice din domenii înrudite cu cel pe care îşi propune sa-l modeleze. De exemplu, se ştie că oamenii nu pot să îşi amintească informaţie instantaneu — deci programul trebuie să se respecte anumite reguli despre extragerea unei informaţii din memorie. De asemenea, pentru că oamenii procesează într-un timp totuşi scurt, ideea unei căutări exhaustive în memorie nu este o alternativă viabilă. Se ştie că memoria umană funcţionează asociativ — contextul în care suntem adesea influenţează ceea ce ne amintim, iar procesul de extragere a informaţiei din memorie este fundamental.

Alan Newell a fost primul care a propus ideea unei teorii unificate a cogniţiei umane. Aceasta ar fi un cadru foarte general construit pe baza unor asumpţii fundamentale despre natura memoriei, despre selecţia unei alternative atunci cînd mai multe sunt posibile, despre învăţare etc. Pe baza acestor asumpţii se produc modele care adresează date empirice din domenii diferite — de exemplu memoria pentru liste de cuvinte[5] sau pentru propoziţii care au sens, alegerea unei strategii optime, descompunerea unei probleme în subprobleme etc. Fiecare din modelele reuşite (în sensul că reproduc fenomenele pe care încearcă să le modeleze), produse fiind sub asumpţiile teoriei, dau validitate acestor asumpţii. Deseori, unele din aceste modele duc la reconsiderarea asumpţiilor fundamentale – cercetarea este un proces de "încercare şi eroare" ("trial and error"). Însă de multe ori fenomene dintr-un domeniu, prin faptul că chestionează asumpţii fundamentale ale teoriei, aruncă lumină asupra unor aspecte dintr-un domeniu paralel.

## 3.2 ACT-R

Există mai multe teorii unificate ale cogniţiei. Universitatea Carnegie Mellon, unde Alan Newell a fost profesor, este legată de mai multe dintre ele (Soar, ACT-R, CAPS etc.). Toate se bazează pe sisteme de producţie.

Pentru exemplificare, vom vorbi pe scurt despre unul dintre ele: ACT-R (`http://sands.psy.cmu.edu/ACT/act-home.html`). Acesta este un limbaj de programare, dar ale cărui instrucţiuni şi mod de evaluare sunt concepute conform cu o anumită teorie despre modul de funcţionare al cogniţiei umane.

Există două tipuri fundamentale de concepte cu care ACT-R operează: memoria declarativă (compusă din "chunk"-uri) care conţine informaţii "statice", fapte asemănătoare cu cele pe care le învăţam în orele de geografie din clasele primare ("Chişinău este capitala Moldovei") şi memoria procedurală, populată de reguli atomice, indivizibile (de exemplu, toate regulile pe care le urmăm cînd conducem maşina). Distincţia între procedural şi declarativ nu este întîmplătoare — există motive să credem că aceste două tipuri de memorie sunt diferite — de exemplu, există amnezici care nu sunt capabili să memoreze fapte noi sau să îşi amintească informaţii din memoria pe termen lung, dar cu toate astea, pot conduce o maşină şi pot chiar învăţa să execute operaţii procedurale complicate (cum ar fi o problemă precum turnurile din Hanoi).

Un program ACT-R menţine o stivă de "scopuri" şi încearcă întotdeauna să execute scopul care este în vîrful stivei; face aceasta examinînd regulile (sau producţiile) din memoria procedurală şi selectînd spre aplicare "cea mai bună" regulă care poate servi la îndeplinirea scopului curent. Cum se defineşte "cea mai bună" este o întreagă poveste — de exemplu, cea mai bună poate fi cea

[5] E vorba despre modul în care oamenii reţin liste lungi de cuvinte fără legătură între ele.

care ia cel mai puţin timp pentru a fi executată sau cea care s-a dovedit corectă în experienţele anterioare[6]. Aşa cum vă puteţi imagina, nici criteriul de selecţie nu este ales la voia întîmplării — nenumărate experimente arată că oamenii şi animalele învaţă din eşecuri sau succese (de exemplu, alegerea locului de păşune de către animale sălbatice depinde de experienţa pe care ele au avut-o în trecut cu locul respectiv).

## 4 Rezumat

Sperăm că am reuşit să ilustrăm în oarecare măsură scopul şi mijloacele ştiinţelor cognitive. Interesul ştiinţelor cognitive este de a înţelege cum gîndesc oamenii; în acest scop, cercetătorii formulează teorii despre cogniţie, care sunt adesea descrise sub formă de algoritmi. Aceşti algoritmi sunt implementaţi folosind interpretoare specializate. “Instrucţiunile” de bază modelează mecanismele elementare ale cogniţiei umane; algoritmii sunt apoi folosiţi pentru a prezice rezultatele experimentale, confirmînd (sau infirmînd) teoria care a stat la baza modelului. Rezultatele experimentale sunt obţinute din studierea unor subiecţi umani.

Cu paşi mărunţi, ştiinţele cognitive ne apropie de înţelegerea acestui fascinant univers care este creierul uman.

## 5 Bibliografie selectivă


Anderson, J.R. and Lebiere, C., 1998. “The atomic components of thought”, Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey.

Gibbs, R., 1990. “Comprehending Figurative referential descriptions”, Journal of Experimental Psychology: Learning, Memory and Cognition, 16, pp. 56–66.

Searle, J.R., 1979. “Metaphor.” În Ortony, A. (editor) “Metaphor and Thought”, Cambridge University Press.

Ortony, A. and Schallert, D. and Reynolds, R. and Antos, S., 1978. “Interpreting metaphors and idioms: Some effects on comprehension”, Journal of Verbal Learning and Verbal Behavior, 17, pp. 465-477.

Dascal, M., 1987. “Defending Literal Meaning”, Cognitive Science, 11, pp. 259-281.

Keysar, B., 1989. “On the functional equivalence of literal and metaphorical inte rpretations in discourse.” Journal of Memory and Language, 28, pp. 375-385.

Anderson, J.R., 1995. “Cognitive psychology and its implications.” W.H. Freeman & Co, New York.


## 6 Despre autori


**Raluca Budiu** este doctorand în cadrul departamentului de calculatoare al universităţii Carnegie Mellon. Domeniul ei de interes este modelarea cognitivă a fenomenelor lingvistice.


[6]De fapt, soluţia exactă este un compromis între cost şi succesul anterior.

**Mihai Budiu** este un corespondent fidel al revistei.