

# ASPECTE ALE DETERMINĂRII SENSULUI PROPOZIȚIEI ÎN LIMBAJ NATURAL

Sergiu CREȚU  
A.S.E.M.  
[srgcretu@yahoo.com](mailto:srgcretu@yahoo.com)

**Abstract.** This study represents a research concerning the principles of elaboration of an interface in natural language (NL) for informational systems. When we say interpretation of texts in a NL, we mean the process of analyzing and extracting information contained in NL texts, realized according to a precise semantic and pragmatic context.

## 1. Introducere. Cadrul conceptual

Prezentul studiu vizează elaborarea unor tehnici de determinare a sensului (Sinn) unei propoziții scrise sau rostite în limbaj natural, reieșind din semnificația (Meaning, Bedeutung) unităților lexicale fixate de un vocabular (perfectat ad-hoc) în coroborare cu regulile sintactice de combinare a valorilor lexicale în unități sintactice.

Baza acestui mecanism este asigurată de:

- Componenta lexicală reprezentată de un vocabular. Vocabularul definește, în fond, semantica lexicală a textului, adică sensul unităților lexicale ale limbajului.
- Componenta sintactică ce definește sintaxa frazelor (propozițiilor) limbajului.
- Componenta semantică ce definește așa-numita semantică propozițională a textului în limbaj natural: referirea, raportarea unităților sintactice ale propoziției la anumite stări de lucruri (Sachverhalte) ale domeniului (World) de interpretare.
- Componenta pragmatică ce ține de utilizarea limbajului. Sensul unei propoziții depinde de contextul de utilizare a acesteia, adică de locul unde și de timpul când a fost pronunțată, scrisă etc. propoziția.

Pentru a putea contribui la determinarea sensului unei propoziții rostite sau scrise componentele enunțate mai sus sunt pasibile de a fi reunite într-un singur sistem, utilizându-se anumite tehnici integrative elaborate în scopul dat. Aceste tehnici integrative devin posibile în baza Principiului de Compoziționalitate al lui Frege, care presupune, pentru articulația sintaxă – semantică a textului, existența unui homomorfism între unitățile sintactice și cele semantice ale limbajului natural.

Toate acțiunile necesare stabilirii sensului unei propoziții în limbaj natural (consultarea vocabularului etc.) vor fi descrise prin intermediul unei algebre multi-sort, elaborate în acest scop. Teza dată are un caracter fundamental pentru toată expunerea ulterioară. Vom concretiza, pe scurt, conceptele ce țin de această abordare.

Domeniul de aplicabilitate al studiului propus atenției îl constituie sistemele informatice, pentru care este caracteristică reprezentarea informației sub formă de texte în limbaj natural.

În continuare va fi analizată fiecare dintre componentele mecanismului de determinare a sensului unei propoziții în limbaj natural.

## 2. Componenta semantică

Abordarea fenomenului semantic, a înțelesului este atât de complexă, încât s-a încercat înlocuirea semanticii (sinonimă cu înțelesul) prin studiul înțelesului unor entități morfo – sintactice.

Din lipsă de spațiu, nu vom stăruia asupra problemelor apărute în legătura cu o astfel de abordare a semanticii. În consecință, am trecut la o analiză atentă și sistematică a relațiilor semantice propriu-zise: coerența discursului, corectitudinea fenomenelor de referire, sinonimia etc. Or, majoritatea discursurilor analizate conțin incoerențe, generate de erori tehnice sau de erori conceptuale, lipsesc referirile etc. Cu toate acestea, în interpretarea de mai sus, discursurile au un anumit sens, dar sunt incorecte semantic.

Mai mult decât atât, dacă, de exemplu, îi comunicăm vorbitorului că „CATELAL” este substantiv, atunci el poate lesne să reproducă toată morfologia acestui cuvânt și poate, chiar, genera fraze sintactic corecte, dar fără înțeles. Acest fapt relevă necesitatea separării aspectului pur formal al competenței lingvistice a vorbitorului de aspectul denotațional al acesteia, înțeles ca un proces de atribuire a „înțelesului” prin specificarea obiectelor denotate.

Pentru depășirea problemelor evocate mai sus vor fi aduse, aici, trei definiții.

**Definiția 2.1.** Competența lingvistică a vorbitorului de a produce fraze lingvistice fără vreo implicare a sensului o vom numi competență sintactică.

A doua definiție ține de semantică, tratată ca atribuire de sens unităților morfo–sintactice. În domeniul dat există suficientă confuzie și tot atâtea păreri privind depășirea acesteia. Se va adopta următoarea strategie.

**Definiția 2.2.** Competența lingvistică a vorbitorului de a stabili relații semantice (de sens, de înțeles) între unitățile morfo–sintactice: relații de includere (conținere) a sensului (is part of etc.), relații de referire, consistență, coerență se va numi competență semantică.

Rămâne ultimul aspect al competenței lingvistice – aspectul pragmatic.

**Definiția 2.3.** Competența lingvistică a vorbitorului de a utiliza frazele cu caracter lingvistic într-un context sintactico–semantic corect o vom numi competență pragmatică.

Problemele legate de sintaxa limbajului natural au fost abordate într-o serie de lucrări, de exemplu în cele aparținând lui N. Chomsky, W.A. Woods și altor autori, care au dezvoltat tehnici eficiente pentru analiza sintactică a limbajului natural.

Pentru reprezentarea semanticii există mai multe modele populare, de exemplu modelul Grosz Sidner's, modelul centrelor, teoria nervurii și altele, care au permis elaborarea de tehnici sofisticate de determinare a corectitudinii textelor în limbaj natural pe baza relațiilor semantice.

Pragmatica rămâne, deocamdată, neexplorată suficient pentru a propune soluții viabile. Ar fi de remarcat, aici, contribuția lui Ch. Pierce, G. Frege, F. de Saussure, B. Russel, Ch. Morris, R. Carnap, R. Montague etc. Dat fiind spațiul restrâns, nu vom face o analiză comparată a contribuției acestor savanți la evoluția pragmaticei moderne.

Înțelegerea (dar și producerea) unui text în limbaj natural presupune abilitatea vorbitorului de a manifesta cele trei competențe lingvistice menționate mai sus: sintactică, semantică și pragmatică. Această manifestare are caracter integrativ. Este greu de formulat explicit regulile de interacțiune a sintaxei, semanticii și pragmaticei în procesul de înțelegere (producere) a textelor. S-ar putea, însă, presupune caracterul primar al competenței sintactice. Pe baza acesteia se manifestă competența semantică (vorbitorul poate indica sinonimele, antonimele, meronimele, holonimele, este conștient de relațiile de referire etc). În fine, competența pragmatică este rezultatul sintezei competenței sintactice, competenței semantice, dar și a unei competențe pragmatice anterioare actului de înțelegere, bazate pe o experiență pragmatică (stocată într-o bază de cunoștințe, de exemplu sub forma unor sintagme cu un caracter mai mult sau mai puțin lingvistic).

După o analiză atentă se observă că competența sintactică poate fi tratată ca o competență semantică, obiectul semanticii constituindu-l conceptele sintactice și morfologice (subiect, predicat, substantiv, adjectiv).

Competența semantică are caracter static-dinamic, ceea ce înseamnă că unele relații semantice, de exemplu cele de sinonimie, meronimie etc., pot fi stocate într-un vocabular sub forma unor structuri informatice. Pragmatica are, în schimb, caracter dinamic. Nu pot fi, în mod principial, fixate în vocabular sensurile propozițiilor, frazelor, textelor. Acestea urmează a fi construite, generate, sintetizate, ceea ce se va efectua în două moduri: prin definirea explicită a pragmaticei și prin intermediul rețelilor semantice.

Modelul pus la baza definirii competenței semantice a vorbitorului derivă din modelele utilizate pentru descrierea semanticii limbajelor formale. Acest model postulează existența următoarelor entități – mulțimi finite sau infinite (nevide):

- 1)  $O$  – mulțimea obiectelor ce urmează a fi manipulate;
- 2)  $M$  – mulțimea actelor mentale: raționamente, gânduri despre obiectele analizate;
- 3)  $\odot$  – operația de compoziție a actelor mentale.

Nu se va specifica natura actelor mentale. Important este faptul că, fiind aplicat, actul mental produce (forma de reprezentare este deliberativă) un obiect ce aparține mulțimii de obiecte  $O$ . Se postulează ineficiența aplicării actului mental prin admiterea unui obiect  $\omega$  – obiectul vid ce aparține mulțimii  $O$  și luat drept rezultat al aplicării eșuate. Este admisibilă aplicarea față de obiectul vizat a unui lanț de acte mentale. În acest caz, lanțurile de acte

mentale sunt produse prin intermediul operației de compoziție  $\odot$ . Există un act mental ce identifică obiectul, iar rezultatul aplicării acestuia este obiectul însuși. Acest act mental se va nota cu  $I$ .

Mulțimea  $O$  a modelului este structurată și include următoarele mulțimi:

1)  $T = \{ t \mid (\forall m) m \odot t = \omega \}$  – mulțimea obiectelor terminale, adică obiecte al căror sens este fixat într-un vocabular, registru. Orice act mental în această situație este considerat eșuat (obiectul  $\omega$ ). Trebuie să acceptăm interpretarea propusă în vocabular, registru etc.;

2)  $E = \{ e \mid e \neq \omega \wedge (\forall m) m \odot e = \omega \}$  – mulțimea obiectelor terminale elementare. Deci, este eliminat obiectul vid  $\omega$ ;

3)  $C = O \setminus E$  – mulțimea obiectelor compuse, neelementare, non – terminale. Față de aceste obiecte pot fi formulate și aplicate lanțurile actelor mentale înșiruite cu operația  $\odot$ . În prezentul studiu vom arăta geneza acestor obiecte.

Vom cere imperativ îndeplinirea următoarelor relații între entitățile modelului, stipulate mai sus:

$$(R1) \quad m \odot A \in O$$

$$(R2) \quad (\alpha \odot m)A = \alpha(m(A))$$

$$(R3) \quad I \odot A = A$$

$$(R4) \quad (\exists \omega)(\forall m) m \odot \omega = \omega$$

$$(R5) \quad (\forall [(\forall m) m \odot \omega = \omega \rightarrow (\forall A)(\exists \alpha) \alpha \odot A = \omega])$$

$$(R6) \quad (\forall \alpha, e)[\alpha(A) = e \leftrightarrow \alpha(B) = e] \rightarrow A = B$$

**Remarcă.** Relațiile  $R1$ ,  $R2$ ,  $R3$  și  $R4$  sunt clare intuitiv și nu necesită explicații.  $R5$  asigură unicitatea obiectului vid, iar  $R6$  formulează condițiile de egalitate a obiectelor. Aici am folosit următoarea notație: literele latine mici – pentru a desemna actele mentale singulare, literele latine mari – pentru a desemna numele de obiecte. Pe de altă parte, literele grecești sunt utilizate pentru a putea vorbi despre lanțurile de acte mentale obținute cu operația  $\odot$ .

Pentru actele mentale se definește relația de dependență, adică:

**Definiția 2.4.** Două acte mentale  $\alpha_1$  și  $\alpha_2$  sunt dependente, dacă există un al treilea act mental  $\beta$ , astfel încât au loc egalitățile:

$$\alpha_1 = \beta \odot \alpha_2 \text{ sau } \alpha_2 = \beta \odot \alpha_1$$

Relația de dependență permite identificarea prin referire a termenilor singulari dintr-o propoziție.

De exemplu, sunt date propozițiile:

(1) „Manualul de pediatrie este pe masă.”

(2) „Cartea este deschisă la pagina 10.”

Au oare termii „Manualul” și „Cartea” același referent sau este vorba despre doi referenți diferiți? Răspunsul la întrebările de acest tip poate fi asigurat doar în cadrul unui sistem ce ar simula contextul pronunțării propozițiilor respective (Sentence). Aici, prin cuvântul „context” subînțelegem, mai ales, locul și timpul când au fost pronunțate propozițiile analizate. În cadrul modelului propus în studiul de față rezolvarea referințelor se face astfel:

1) Se definește separat repertoriul actelor mentale pentru numele „Manualul” și pentru numele „Cartea”, respectiv.

2) Se generează lanțurile posibile de acte mentale pentru ambele nume.

3) Dacă lanțurile generate nu sunt dependente, atunci numele testate astfel au referenți diferiți, pe care îi și putem identifica în calitate de rezultat al procesului descris.

Chiar și în lipsa detaliilor, este clar că, pentru a avea un răspuns cât de cât satisfăcător la întrebările de mai sus, vom avea nevoie de o informație suplimentară. În paragraful următor vom arăta cum poate fi gestionată această informație suplimentară.

### 3. Componenta pragmatică

În paragraful precedent am făcut referire la informația suplimentară necesară pentru rezolvarea referirilor din propozițiile enunțative. Aceste propoziții sunt pronunțate pentru asertarea unor stări de lucruri din lumea obiectelor (mulțimea  $O$  din modelul p.2). Lanțurile de acte mentale au menirea de a pune în valoare faptele din lumea obiectelor.

Pentru a contura mai eficient problemele care apar în încercarea de a gestiona informația suplimentară, necesară identificării stării de lucruri din lumea obiectelor, vom relua câteva propoziții asertive:

- (1) „Președintele R. Moldova salută drapelul țării.”
- (2) „Nicolae Timofti arborează drapelul.”
- (3) „Ion consideră că fiecare președinte de țară arborează drapelul de stat.”

Vom contura câteva probleme care apar imediat ce se face o analiză (fie și superficială) a fragmentului de text prezentat. Anume:

- (4) Sunt adevărate faptele expuse în propozițiile de mai sus?
- (5) Cine este președintele R. Moldova? Nicolae Timofti?
- (6) Putem, eventual, folosi părerea lui Ion pentru a răspunde la primele două întrebări?

Oferim câteva comentarii. Problema (4) ține, de fapt, de semantică. Conform lui R. Carnap, estimările aserțiunilor cu valorile „adevăr” sau „fals” depind de contextul impus de traducerea în metalimbaj a predicatelor limbajului-obiect. Deci, este vorba despre utilizarea frazelor limbajului.

Pentru problema (5) estimarea cu valoarea „adevăr” depinde de faptul când a fost rostită propoziția. Deci, este o problemă de timp, ce ține de anul, luna și ziua când a avut loc evenimentul rostirii propoziției.

Problema (6), care este, de fapt, o problemă de interpretare a subordonatei din propoziția (3), și-ar găsi o rezolvare dacă s-ar preciza locul (țara), unde președinții arborează personal și de fiecare dată drapelele țărilor respective. În caz contrar, opinia lui Ion ar rămâne doar o opinie.

În concluzie, informația de care avem nevoie pentru a determina sensul propozițiilor rostite sau scrise de către vorbitor ține de determinarea contextului semantic al propozițiilor, de locul unde și de timpul când s-a produs geneza textului. Acesta și este aspectul pragmatic.

Pentru gestionarea aspectului pragmatic al propozițiilor rostite sau scrise, ca o primă aproximare, poate fi propus următorul model inspirat din modelele de gestionare a datelor (abordarea relațională). Modelul este mai puțin realist, dar poate servi la stabilizarea cadrului conceptual.

**Definiția 3.1.** Fie date  $P$  – mulțimea momentelor din spațiul temporal, mulțimea  $O$  – mulțimea obiectelor lumii

și mulțimea  $M$  – mulțimea actelor mentale definite pe mulțimea  $O$ . Prin mulțimea stărilor vom înțelege mulțimea  $S=(P \times M)$  – produsul cartesian.

**Definiția 3.2.** Actul mental  $m_i$  se realizează în starea  $S=(p,m)$ , dacă  $m_i$  aparține lanțurilor de acte mentale rezultative ( $\neq \omega$ ) cu cap de serie  $m$ .

**Definiția 3.3.** Prin mulțimea de stări reale  $R$  se subînțelege mulțimea de valori definită de funcția  $P \rightarrow M$ .

S-ar putea afirma că mulțimea  $R$  este o submulțime a mulțimii  $S$ , unde pentru fiecare moment din spațiul temporal există un singur act mental din  $M$ .

**Definiția 3.4.** Prin starea de lucruri (Sachverhalte) a lumii se va înțelege o submulțime  $H$  a mulțimii  $R$ , pentru care este garantată monotonia.

Pentru a putea demonstra cum funcționează modelul prezentat mai trebuie să definim procesul de atribuire a sensului frazelor lingvistice.

**Definiția 3.5.** Relatia  $V$ , definită pe:

- 1) mulțimea  $L$  a propozițiilor limbajului natural;
- 2) starea de lucruri  $H_1$  în momentul rostirii;
- 3) starea de lucruri  $H_2$  ca rezultat al rostirii,

definește sensul frazelor limbajului  $L$  pentru starea de lucruri  $H$ .

Pentru a demonstra cum funcționează modelul definit mai sus vom prelua un exemplul din p2., anume:

„Manualul este pe masă”

Relatia  $V$  ne permite să scriem următoarea formulă:

$[Manualul](s_1, f, s_2)$

Aici  $s_1$  este starea inițială și momentul rostirii propoziției,  $s_2$  este starea finală, iar  $f$  este o funcție de tranziție între stări. Ea participă la alegerea stării finale conformă cu momentul din spațiul temporal. Astfel, ca rezultat va fi generată starea ce conține actul mental:

$(p_1, \text{denumirea(Manualul)})$

Rezultatul acestei stări va fi constanta textuală „Manual de pediatrie”. Remarcăm ca funcția  $f$  pastrează momentul temporal. Momentul va modificat când vor fi evaluate predicatul și cîrmstantialul – timpul prezent al verbului etc. Din lipsă de spațiu am omis o serie de detalii, uneori importante pentru o mai bună înțelegere a studiului prezent.

Este în lucru o variantă completată și revizuită a lucrării.

### 4. Concluzii și perspective

În final, vom aduce câteva concluzii ce derivă din abordarea prezentată mai sus:

1. Sensurile primare (semnificația), care urmează a fi atribuite fragmentelor textului interpretat, sunt reprezentate sub forma unor obiecte elementare din  $E$ . Formele de reprezentare a sensurilor sunt irelevante. Ele pot fi obiecte lingvistice, un alt text, alte imagini etc. Acest fapt derivă din definiția conceptului de mulțime. Extragerea unui alt sens printr-un act mental asupra obiectului elementar este interzisă, deoarece  $\alpha(e) = \omega$ .
2. Există obiecte compuse (neelementare), care admit mai multe acte mentale (mulțimea  $O \setminus E$ ). Fiecare act mental admisibil este asociat cu obiectul ce denotă aspectul evocat al obiectului vizat –  $m(A)$ . Obiectul compus are o structură imbricată, arborescentă.

3. Obiectul compus, la fel și cel elementar, poate participa ca parte componentă a unui alt obiect și, implicit, contribuie la formarea sensului.
4. Sensul obiectului compus poate fi modificat prin adăugarea și excluderea părților componente.
5. Sensul obiectului compus nu poate fi redus la sensurile părților sale componente, ci este o combinație a sensurilor constituente. Aceste combinații sunt definite dinamic printr-un mecanism, ce are la bază interpretarea prin intermediul lanțurilor de acte mentale. Dimensiunea pragmatică a obiectului se obține prin introducerea stărilor și stărilor de lucruri.

Interpretarea pragmatica a textelor in limbaj natural poate fi eficient implementată, dacă în prealabi s-a lucrat serios la elaborarea unei baze de cunostinte lingvistice și de alta natură suficient de consistentă ca să poată fi folosită drept suport informațional.

În prezenta lucrare nu au fost tratate problemele elaborării unei astfel de baza de cunoștințe. Cerințele de realizare a unei baze de cunoștințe au fost tratate în lucrările [1], [2], [3].

Aceasta bază de cunoștințe ar trebui să formeze contextul semantic - pragmatic al interpretării textului.

Atât mediul semantic (baza de cunoștințe), folosit la interpretare, cât și mecanismul de interpretare propriu – zis trebuie să fie compatibile în sensul formalismelor descriptive. În prezenta abordare această cerință va fi asigurată de limbajul LBC –limbajul de gestionare a bazei de cunostinte. Baza de cunoștințe, utilizată la interpretare, trebuie să aibă dimensiuni considerabile pentru a furniza contextul semantic adecvat. Prin urmare, pentru baza de cunoștințe sunt indicate tehnici extensibile.

Pentru reprezentarea rezultatului interpretării textelor în limbaj natural trebuie formulate două teze, care stipula, că rezultatul interpretării sintactico – semantice a textului este descris de o algebră multi – sort, iar atribuirea de sens construcțiilor sintactice se reduce la homomorfismele de

finite de derivatorii algebrici. În consecință va fi posibilă elaborarea schemei de reprezentare a procesului de interpretare a textelor în limbaj natural.

Pentru menținerea integrității bazei de cunoștințe va trebui elaborat un set de operații și formulată concepția vocabularului virtual – metodă de organizare și de extindere a bazei de cunoștințe.

Va trebui precizată noțiunea de glosă ca rezultat final al interpretării textelor în limbaj natural.

În perspectivă se preconizează întreprinderea următoarelor investigații pentru dezvoltarea abordării propuse:

1. Elaborarea principiilor de implementare a corpus – ului lingvistic adecvat procesului de interpretare sintactico – semantică - pragmatica.

2. Perfectarea unei baze morfologice reprezentative, folosind limbajul LBC și principiile de reprezentare a informației cu caracter lingvistic.

3. Realizarea unui interpretor semantic - pragmatic, care ar automatiza extragerea informației din textul analizat cu interpretarea ei ulterioară în contextul semantic asigurat de baza de cunoștințe existentă.

## Bibliografie

1. Crețu S., Popescu A., Interpretarea sintactico – semantică a textelor în LN, Proceedings of the 5<sup>th</sup> International Conference on „Microelectronics and Computer Science”, v.II, Chisinau, 2007, p. 181-184.

2. Crețu S., Elaborarea unui mecanism formal de interpretare sintactico – semantica a textelor în limbaj natural, in Proceedings of the 30th Annual Congress of the American – Romanian Academy of the Arts and Sciences(ARA), Chișinău, 2005, p. 131 – 133.

3. Crețu S., A system for natural language text syntactic – semantic interpretation (SSI), in the 2<sup>nd</sup> supplement of the review Informatica Economică, International Conference Knowledge Management: Projects, Systems and Technologies, Bucharest, vol. 1, November, 2006, p. 171 – 174.