

ASPECTE INTENSIONALE ÎN LIMBAJUL NATURAL: PROCESAREA

¹Sergiu CREȚU, ²Anatol POPESCU

¹Academia de Studii Economice din Moldova

²Universitatea Tehnică a Moldovei

Abstract. This study aims to develop techniques for determining the meaning (Sinn) of a sentence written or spoken in natural language, based on meaning (Meaning, Bedeutung) of lexical units fixed by a vocabulary (elaborated ad-hoc) in conjunction with syntactic rules for combining values in lexical syntactic units.

Index Terms – extensional context, intensional context, interpretation, meaning, sentence.

I. INTRODUCERE

Scopul prezentului studiu rezidă în:

1. Studiarea tehnicilor de procesare a componentei intensionale a textelor în limbaj natural (LN);
2. Elaborarea unor medii de stocare și procesare a informației conținute în textele în LN sub formă intensională;
3. Proiectarea tehnicilor de utilizare a informației stocate în vederea asigurării contextului sintactico-semantic adecvat, necesar interpretării informației extrase;

Funcțiile principale ale unei astfel de interfață, de fapt, sunt:

1. Depistarea și identificarea unităților morfo-sintactice în textul LN analizat.
2. Sintetizarea contextului sintactic (gramatica) adecvat;
3. Interpretarea sintactico – semantică a textului LN în contextul sintetizat;
4. Generarea unor reacții la interpretare: glose, exegeze, expuneri succinte, acțiuni concrete.

II. PRECIZĂRI ALE CONCEPTELOR

Limbajul natural reprezintă un sistem de utilizare a unor semne verbale (cuvintele). Această definiție succintă are un caracter semiotic. Există păreri (Ch. Morris), că semiotica poate servi drept metalimbaj pentru definirea lingvisticii. Au fost facute chiar și încercări în acest sens [1]. De la Ch. Pierce, F. de Saussure, Ch. Morris și alții semnele, ca obiecte semiotice, trebuie să includă: 1. un semnificat; 2. un semnificant; 3. un interpretant. Semnificatul este obiectul la care trimite semnul (renvoi, R. Jakobson), semnificantul este imaginea semnului (grafică, acustică etc.), prin care semnul se face recunoscut, iar interpretantul este un semn sau mai multe semne corelate cu semnul definit. Ultima dimensiune a

semnului- interpretantul- ține de componenta sa intensională.

Observație. Limba, limbajul și vorbirea sunt entități diferite. O limbă poate avea mai multe limbaje, dată fiind pragmatica sa (modul de utilizare). Prin extensie limbajul poate fi și artificial formal, logic, matematic etc. Sunt mai multe teorii despre apariția limbii. Vom reieși în prezentul studiu din teoria [1], că limba este o adaptare a ființelor vii fiind un mijloc de modelare a lumii reale. Lumea reală este interpretată prin semne. Vorbirea (componenta comunicativă) este de dată recentă în comparație cu apariția limbii – 300 mii ani în urmă. Există interpretarea lumii reale (orientată în mediu) fără vorbire.

Concluzie. Fiecare semn al limbii are două componente (aspecte) determinante: 1. aspectul extensional. 2. aspectul intensional. Aspectul extensional ține de referirea (trimiterea la, renvoi) unui lucru din lumea reală, iar aspectul intensional ține de multiplicarea interpretărilor semnului. Am evitat până acum folosirea denumirii „cuvânt” pentru semnul verbal, cu toate că a fost menționată mai sus. Acest fapt este justificat, deoarece nu există o corespondență biunivocă între cuvintele unei limbi și semnele verbale. De exemplu, frazele clasice „Luceafărul de dimineață”, „Luceafărul de seară” și „Venus” referă același corp ceresc. Ar fi instructiv să numărăm semnele verbale incluse în frazele prezentate mai jos.

Vom face o analiză succintă:

1. Fraza „Luceafărul de dimineață” este primul semn. Constă din: semnificant - cuvintele: „Luceafărul”, „de”, „dimineață”, semnificat – corpul ceresc Venus, interpretant - Luceafărul de dimineață.
2. Fraza „Luceafărul de seară” este al doilea semn. Constă din: semnificant - cuvintele:

„Luceafărul”, „de”, „seară”, semnificat – corpul ceresc Venus, interpretant - Luceafărul de seară.

3. Fraza „Venus” este al treilea semn. Constă din: semnificat – cuvântul „Venus”, semnificat – corpul ceresc Venus, interpretant – Venus.

Concluzie. Semnificantul unui semn poate include mai multe cuvinte. Mai multe semne pot referi același lucru din lumea reală. Interpretantul poate include mai multe semne. Ultima afirmație trebuie explicată. Cele trei semne evocate mai sus nu sunt echivalente în sensul aspectelor activate la observator. Luceafărul de dimineață nu este totuna cu sensul Luceafărul de seară. Mulțimea de sensuri evocate într-o relație dintre semnificat și semnificat formează aspectul intensional.

Exemple. Fie date propozițiile:

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

O analiză superficială a fragmentului de text prezentat mai sus ar trebui să dea un răspuns concludent la următoarele întrebări:

1. Sunt adevărate faptele expuse în propozițiile de mai sus?
2. Cine este președintele R. Moldova? Nicolae Timofti?
3. Putem, eventual, folosi părerea lui Ion pentru a răspunde la primele două întrebări?

Vom preciza acum contextele pentru propozițiile enunțate mai sus.

Prima propoziție poate fi interpretată atât în context extensiv cât și în context intensional.

Contextul va fi extensiv, dacă există un semnificat pentru semnificantul „Președintele R. Moldova”. Acest fapt depinde, în general, de niște factori extralingvistici, regrupați cu denumirea de „lume posibilă” (possible world, w). În unele lumi posibile propoziția va fi tratată extensiv, iar în altele – intensional, adică când nu va exista o ființă reală pe post de președinte.

A doua propoziție este în context extensiv. Semnificantul „Nicolae Timofti” este un termen singular, adică în toate lumile posibile există o persoană cu acest nume.

A treia propoziție este în context intensional evident. Din lipsă de spațiu nu vom mai aprofunda analiza

Concluzie. Interpretarea textelor în LN rezidă în elaborarea:

1. Modelului categorial de reprezentare a obiectelor lingvistice;
2. Modelului denotativ de atribuire de sens obiectelor lingvistice;
3. Modelului de reprezentare a rezultatelor interpretării.

În continuare vom prezenta succint primele două modele.

III. REPRESENTAREA SINTAXEI TEXTULUI.

Modelul de reprezentare a sintaxei textului este elaborat pentru un fragment al limbii române și va fi bazat

pe așa numita gramatică categorială [2]. Ea operează cu unități morfo-sintactice care se numesc categorii. Gramatica utilizată în acest studiu va include următoarele categorii:

1. N – nume proprii (termen singular).
2. CN – substantive comune (nume predicative).
3. Vi – verbe intransitive.
4. Vt – verbe tranzitive.
5. Vf – frază verbală.
6. S – propoziție (sentence)

Din lista categoriile gramaticii utilizate sunt evidențiate categoriile de bază B – mulțimea denumirilor categoriilor de bază. În cazul dat $B = \{B_N, B_{CN}, B_{Vi}, B_{Vt}\}$. Conform înțelegerii elementele B reprezintă niște etichete de mulțimi. De exemplu, B_N este eticheta pentru mulțimea numelor proprii utilizate. Categoriile de bază au o semnificație lexicală, iar valorile categoriilor VP și S trebuie sintetizate conform regulilor de interpretare. Unele entități sintactice nu au o expresie categorială (sub formă de categorii), de exemplu, conjuncțiile etc. Ele se numesc sincategorematice și trebuie incluse în regulile sintactice de generare a frazelor lingvistice. Frazele lingvistice se construiesc conform regulilor sintactice. Din mulțimea tuturor frazelor posibile (mulțimea de șiruri de caractere) pot avea încărcătură semantică doar frazele generate în concordanță cu regulile sintactice declarate. Regulile sintactice sunt niște producții de forma:

$\langle \text{producție} \rangle ::= \langle \text{condiție} \rangle$, atunci $\langle \text{concluzie} \rangle$
Elementul $\langle \text{condiție} \rangle$ este o expresie logică.

Exemple. Fie dată propoziția: „**Ion arborează drapelul**”. În acest caz categoriile vor fi:

$B_N = \{\text{Ion}\}$, $B_{CN} = \{\text{drapel}\}$, $B_{Vt} = \{\text{arbora}\}$.

Regula (sintactică) corespunzătoare de generare (producția) este:

$S1. \text{arbora} \in Vt \text{ and drapel} \in CN$, atunci $\text{arbora drapel}' \in VP$, sau în caz general:

$S1. \alpha \in Vt \wedge \beta \in CN$, atunci $\alpha\beta' \in VP$

Fiecare regulă sintactică este precedată de o etichetă de forma: S_n , unde n este un număr întreg. Acest lucru este necesar pentru identificarea regulii, când va fi declarată semantica și pragmatica. VP este categoria frazelor verbale. Desigur, regula definește numai partea verbului din propoziție. Pentru a defini propoziția mai declarăm o regulă:

Abordarea prezentată ne permite să punem în evidență existența unor relații între categoriile gramaticii. Acest lucru poate fi formalizat prin introducerea a doi operatori: $\wedge(A,B)$ – categoria A se află la dreapta categoriei B și $\backslash(A,B)$ – categoria A se află la stânga categoriei B. De exemplu, categoria Vt este echivalentă cu expresia: $\backslash(N, \wedge(CN, S))$. Expresia prezentată poate fi redată prin cuvintele: „... la stânga categoriei Vt se află categoria N, iar la dreapta – categoria CN și în acest caz elementele frazei lingvistice formează categoria S ...”. Amintim, categoria S este propoziția descrisă. Imediat concludem, că atât categoria Vt, cât și categoria VP pot fi excluse din definiția gramaticii. Urmează o nouă definiție pentru gramatica categorială

Exemplu. Fie dată propoziția: „**Ion arborează abil drapelul**”. Atunci gramatica categorială clasică care poate analiza (sau genera) această frază lingvistică este:

1. Categoriile de bază sunt: $\mathbf{B}=\{N, CN, S\}$, unde N desemnează numele proprii (termii singulari), CN sunt numele comune, S reprezintă propoziția și ține locul de axiomă.
2. Vocabularul $\mathbf{V}=\{\text{Ion, drapel, abil, arbora}\}$.
3. Gramatica categorială clasică (relația finită)
 $\mathbf{G}=\{\langle \text{Ion}, N \rangle, \langle \text{drapel}, CN \rangle, \langle \text{arbora}, \backslash(N, /(\text{CN}, S)) \rangle, \langle \text{abil}, \backslash(\backslash(N, /(\text{CN}, S)), \backslash(N, /(\text{CN}, S))) \rangle\}$.

Gramatica categorială prezentată aici face analiza propoziției enunțate. Pentru a extinde fragmentul analizat vor trebui modificate componentele B , V și G ale gramaticii.

IV. MODELUL SEMANTIC DENOTATIV

Gramatica G din paragraful precedent va fi explicată cu ajutorul interpretării gramaticii într-un Univers tipat de indivizi (individuals). În acest Univers (mulțime) sunt definite două tipuri de indivizi: e (entity) și t (true, false). Tipul e desemnează niște entități cu caracter constant, natura cărora este irelevantă. Tipul t desemnează două constante (indivizibile) identificate de cuvintele „adevăr” (true, 1) și „fals” (false, 0). Natura acestor constante este și ea irelevantă.

Definiția 4.1. Mulțimea *Type* de tipuri este mulțimea minimală care include următoarele elemente:

1. $e \in \text{Type}$.
2. $t \in \text{Type}$.
3. Dacă $a \in \text{Type}$ și $b \in \text{Type}$, $\langle a, b \rangle \in \text{Type}$.

În definiția 4.1 $\langle a, b \rangle$ desemnează o funcție cu domeniul de definiție D_a (mulțime de tipul a) și domeniul de variație D_b (mulțime de tipul b). De exemplu, $\langle e, t \rangle$ desemnează o mulțime de indivizi sau, altfel, o funcție caracteristică care regrupează indivizii cu o anumită însușire, iar $\langle \langle e, t \rangle, t \rangle$ este un predicat de gradul 2 etc.

Cum trebuie de procedat pentru a interpreta (a determina înțelesul) o propoziție a unui limbaj natural? În primul rând, poate fi vorba numai de o funcție, care ar pune în corespondență o constantă din mulțimea de constante de tipul t dintr-un anume Univers de indivizi ales pentru interpretare. Conform lui A. Tarski: „,p’ este adevărată atunci și numai atunci când p este adevărată”, unde p desemnează propoziția, iar ghilimelele singulare evocă aspectul lingvistic al propoziției. Ca rezultat, drept obiect de studiu va trebui luată propoziția tratată ca obiect logic. În consecință, funcția de interpretare pentru o propoziție în limbaj natural va fi o funcție de traducere din limbajul natural în formulele unui limbaj logic la o primă etapă. La etapa a doua formulele vor fi interpretate pe un Univers tipat de indivizi. Propoziția va fi adevărată sau falsă în dependență de valoarea formulei logice interpretate. Cum ajungem la aceste formule logice? Ne vom baza pe Principiul de Compoziționalitate al lui G. Frege, adică vor trebui mai întâi interpretate părțile componente ale propoziției, apoi, pe această bază se va

obține interpretarea propoziției în întregime. Interpretarea propoziției bazată pe interpretarea părților sale componente presupune o abordare conformă cu ideile expuse de R. Montague în [2]. Conform acestora se vor evidenția mai întâi funcții propoziției ca mai apoi să se caute argumentorii funcțiilor. Ne va interesa structura limbajului logic pe care îl vom utiliza pentru a elabora formulele de interpretare. Acest limbaj este un limbaj formal cu sintaxă și semantică fixă. În această prezentare nu vor fi prezentate regulile care definesc sintaxa și semantică acestui limbaj.

Semantica fragmentului de text analizat va fi obținută în două etape:

1. Pe baza expresiilor categoriale (sintaxa) ale textului analizat vor fi obținute expresiile interpretante corespunzătoare în limbajul logic;
2. Expresiile logice obținute astfel vor fi interpretate în temeiul Universului.

Astfel se obține emergența semnificativității în fragmentul de text analizat. Semantica fragmentului de text numa este ambiguă.

Conform celor expuse mai sus, la o primă etapă, trebuie de definit o corespondență între categoriile sintactice ale gramaticii categoriale utilizată pentru descrierea frazelor lingvistice și tipurile semantice folosite la elaborarea formulelor limbajului logic.

Definiția 4.2.

- I. Pentru categoriile de bază (definiția 4.1) ale gramaticii vom stabili următorul morfism f :
 1. $N \rightarrow e$, numelor proprii li se vor pune în corespondență indivizii Universului tipizat.
 2. $S \rightarrow t$, propozițiilor li se vor pune în corespondență elementul de tipul t (true, false).
 3. $CN \rightarrow \langle e, t \rangle$, numelor comune li se vor pune în corespondență predicatele de gradul 1.

Morfismul f va asigura caracterul injectiv al corespondenței definite.

- II. Pentru celelalte categorii ale mulțimii $\text{Cat}(B)$ va fi asigurată relația:

$$f(\backslash(A, B)) = \langle f(A), f(B) \rangle \text{ și } f(/(A, B)) = \langle f(A), f(B) \rangle, \text{ unde } A \text{ și } B \in \text{Cat}(B).$$

Exemplu. Vom apela la gramatica G din definiția 2.2. Conform morfismului f definit vom obține:

$$\begin{aligned} \langle \text{Ion}, N \rangle &\rightarrow e \rightarrow \text{Ion} \\ \langle \text{drapel}, CN \rangle &\rightarrow \langle e, t \rangle \rightarrow \lambda x[\text{drapel}''(x)] \\ \langle \text{arbora}, \backslash(N, /(\text{CN}, S)) \rangle &\rightarrow \\ &\rightarrow \langle e, \langle \langle e, t \rangle, t \rangle \rangle \rightarrow \exists x[\text{drapel}''(x) \wedge \text{arbora}''(\text{Ion}, x)] \\ \langle \text{abil}, \backslash(\backslash(N, /(\text{CN}, S)), \backslash(N, /(\text{CN}, S))) \rangle &\rightarrow \\ &\rightarrow \langle \langle e, \langle \langle e, t \rangle, t \rangle \rangle, \langle e, \langle \langle e, t \rangle, t \rangle \rangle \rangle \rightarrow \\ &\text{abil}''(\exists x[\text{drapel}''(x) \wedge \text{arbora}''(\text{Ion}, x)]) \end{aligned}$$

Comentariu. Verbul tranzitiv „a arbora” a fost tratat absolut extensional. Cuvintele cu apostrof dublu sunt

traducerile într-un limbaj logic ipotetic ale cuvintelor din limbajul natural.

V. ASPECTE INTENSIONALE ALE LIMBAJULUI NATURAL.

Limbajul natural conține contexte, care nu răspund abordării de mai sus. Propoziția „Alexandru își aliniază falanga” este ambiguă. Ea nu poate avea valoarea „adevăr” sau „fals” fără precizarea contextului (aspectul intensional). Cuantificarea, în acest caz, se va face pe niște obiecte extralingvistice. Ele au obținut diferite denumiri pe parcursul dezvoltării ideilor. Acum, de cele mai multe ori, sunt cunoscute ca lumi posibile (possible worlds). Fiecare lume posibilă gestionează o anumită stare de lucruri a realității. Conceptul lumilor posibile a fost explorat la elaborarea logicilor modale (S. Kripke) și logicilor intensionale (R. Montague). Ca rezultat al remarcilor făcute, interpretarea propozițiilor va avea loc în contextele asigurate de lumile posibile. Vom exploata varianta lumilor posibile adusă de R. Montague în lucrarea clasică [3] cu modificările noastre. În continuare, deoarece în p.4 al studiului a fost prezentată abordarea extensională, se vor prezenta doar adăugirile la limbajul logic necesare pentru a asigura abordarea intensională. Aceste adăugiri vor fi regrupate pe compartimentele limbajului.

Adăugirile sunt următoarele:

- I. La mulțimea Type (definiția 4.1) se adaugă rândul:
 4. Dacă $a \in \text{Type}$, atunci $\langle s, a \rangle \in \text{Type}$, unde s este al treilea tip de obiect adăugat la
- II. La regulile sintactice ale limbajului logic se adaugă rândurile:
 1. Dacă $\alpha \in \text{ME}_a$, atunci $\wedge \alpha \in \text{ME}_{\langle s, a \rangle}$.
 2. Dacă $\alpha \in \text{ME}_{\langle s, a \rangle}$, atunci $\vee \alpha \in \text{ME}_a$.

Comentariu. Elementul s adăugat permite ieșirea din limbaj și operarea cu entități extralingvistice (lumile posibile). Există posibilitatea de a defini funcții de tipul $\langle s, a \rangle$. ME este mulțimea expresiilor cu sens (înțeles). În sfârșit, verbul tranzitiv „a arbora” a propoziției din p.2. va fi exprimat în abordare intensională de formula:

arbora (Ion, $\wedge \lambda A \exists x [\text{drapel}(x) \wedge A(x)]$)

VI. CONCLUZII

Prezentul studiu are drept scop elaborarea sistemelor de interpretare a textelor în limbaj natural. Abordarea propusă permite perfectarea unui sistem informatic, care ar ține cont de complexitatea problemei. Din lipsă de spațiu au rămas în afara prezentării unele rezultate obținute la subiectul abordat. Ar fi de remarcat elaborarea unui limbaj formal pentru definirea regulilor sintactice și semantice ale etapelor de interpretare evocate. Rezultatele se regăsesc în lucrările publicate [4], [5], [6]. Această interpretare este una sintactico-semantică. Pentru o interpretare eficientă se cere elaborarea unei baze de cunoștințe [7], [8], [9], care ar asigura contextul semantico-pragmatic. În prezentarea de față nu a fost

abordată această problematică. Scopul lucrărilor inițiate rezidă în implementari concrete care ar asigura:

1. Elaborarea de interfețe eficiente între sistemele informatice și utilizator.
2. Automatizarea stocării informației conținute în textele în limbaj natural.
3. Identificarea și extragerea informației stocate.

Schema generală a abordării propuse poate fi consultată în [10], [11]

BIBLIOGRAFIE

- [1] T. Sebeok, *Semnele: O introducere în semiotică*. Humitas, București, 2002, p. 237.
- [2] R. Montague, „Universal Grammar”, Reprinted in *Formal Philosophy; Selected Papers of Richard Montague*, 1974, p. 222-246.
- [3] R. Montague, „The Proper Treatment of Quantification in Ordinary English”, Reprinted in *Formal Philosophy; Selected Papers of Richard Montague*, 1974, pp. 247-270.
- [4] S. Crețu, *Interpretarea semantică a textelor în limbaj natural*. EduSoft, Bacău, 2007, p. 208.
- [5] S. Crețu, A. Popescu, „Meaning of the Sentence in the Natural Language: semantic insight”, *Meridian ingineresc*, no. 4, Tehnica UTM, Chișinău, 2013, p. 46-50.
- [6] S. Crețu, „A system for natural language text syntactic – semantic interpretation (SSI)”, *The 2nd supplement of the review Informatica Economică, International Conference Knowledge Management: Projects, Systems and Technologies*, Bucharest, vol. 1, November, 2006, p. 171 – 174.
- [7] S. Crețu, A. Popescu, „Definirea semanticii propoziției în limbaj natural”, *The 7th International Conference, Microelectronics and Computer Science*, september 22-24 2011, UTM, Chișinău, 2011, p. 174-177.
- [8] S. Crețu, A. Popescu, „Aspecte semantice și pragmatice ale sensului propoziției în limbaj natural” *Meridian ingineresc*, no. 3, Tehnică UTM, Chișinău, 2013, p. 18-23.
- [9] S. Crețu, „O metodă de definire a semanticii pentru limbajul natural”, *Competitivitatea și inovarea în economia cunoașterii: Probleme și soluții pentru România și Republica Moldova: Conferința științifică internațională din 26 – 27 septembrie 2008*. Chișinău: ASEM, 2008, vol. II, p. 342 – 345.
- [10] S. Crețu, A. Popescu, „Definirea semanticii propoziției în limbaj natural”, *The 7th International Conference, Microelectronics and Computer Science Conference*, september 22-24 2011, UTM, Chisinau, 2011, p. 174-177.
- [11] S. Crețu, „A system for natural language text syntactic – semantic interpretation (SSI)”, *The 2nd supplement of the review Informatica Economică, International Conference Knowledge Management: Projects, Systems and Technologies*, Bucharest, vol. 1, November, 2006, p. 171 – 174.