



ACADEMIA ROMÂNĂ
Institutul de Cercetări pentru Inteligență Artificială

NATURALEȚEA LIMBAJULUI ARTIFICIAL – STUDIU DESPRE
TRADUCEREA AUTOMATĂ A GV

NADIA LUIZA DINCĂ

Conducător: prof. dr. **Dan TUFIȘ**,
Membru Corespondent al Academiei Române

Rezumat

Funcția specifică a limbii este de a servi societatea ca mijloc de comunicare, ca mijloc pentru schimbul de idei și informații între oameni. Dacă din punct de vedere tehnologic pot exista popoare primitive, nu aceeași realitate este valabilă la nivelul limbajului: antropologii au fost în repetate rânduri uimiți de complexitatea și gradul de abstractizare al limbilor societăților primitive ca tehnologie. De asemenea, limbajul vorbitorilor needucați este sistematic și condus de reguli, chiar dacă regulile aparțin altor dialecte decât cel standard.

Fiecare individ uman posedă un algoritm cu abilitatea de a genera un număr infinit de propoziții, fiecare dintre ele corespunzătoare unei idei distincte. De exemplu, informația pe care o deținem despre gramatica limbii române sau a limbii engleze conține reguli de tipul “O propoziție poate fi compusă dintr-un grup nominal (subiectul) și un grup verbal (predicatul)”; “Un grup verbal poate fi realizat dintr-un verb și un grup nominal (obiectul)”.

Gramatica se bazează pe reguli ce alcătuiesc subsisteme. Cel mai important este *sintaxa*, al cărei obiect de studiu îl reprezintă ansamblul regulilor privitoare la îmbinarea cuvintelor și relațiile dintre ele în propoziții și fraze.

Al doilea subsistem al gramaticii este morfologia, în care cuvintele sau părți din cuvinte (afixe) sunt asamblate pentru a produce cuvinte complexe.

Din punct de vedere sintactic, orice traducere se confruntă cu trei tipuri de probleme:

- i. definirea și recunoașterea tiparelor sintactice din limba sursă și a relațiilor care le compun într-un întreg acceptabil formal;
- ii. definirea și recunoașterea structurilor sintactice și a relațiilor echivalente din limba țintă;
- iii. corelarea celor două grupe de fenomene (procedurile de transfer dinspre limba sursă spre limba țintă).

În lucrarea **Naturațea limbajului artificial - studiu despre traducerea automată a GV**, analizez cele trei grupe de probleme, focalizându-mă asupra structurii sintactice a grupului verbal din limbile română (limba sursă) și engleză (limba țintă).

Etapile de lucru în realizarea acestei lucrări au cuprins citirea de documentație pentru o înțelegere mai bună a domeniului Prelucrării Automate a Limbajului Natural, definirea grupului verbal și alegerea unei modalități de reprezentare sintactică a constituenților grupului verbal, definirea și crearea tehnologiei pentru descrierea modulelor ce cuprind studiul de traducere automată a grupului.

Din punct de vedere teoretic, lucrarea își propune să prezinte, pentru limba română, o abordare personală a morfologiei paradigmatică, ținând cont de două tipuri de informații: restricția de nominativ-acuzativ plural, nearticulat, considerată identificator de paradigmă, și operațiile de *alternanță*, *adunare sau scădere de șiruri*. Metoda propusă a fost pusă în aplicare în cadrul modulului de deflexionare nominală și verbală, unul dintre modulele ce compun studiul de traducere automată.

La nivelul dezambiguizării am abordat interactiv modulul de preprocesare a textelor TTL, creat și dezvoltat la Institutul de Cercetare pentru Inteligența Artificială, astfel încât pentru lexemele considerate ambigue prin numărul de etichete morfosintactice este apelat programul TTL care reține pentru fiecare ocurență lema și eticheta morfosintactică specifice.

Flexiunea nominală și verbală pentru limba engleză a fost realizată prin intermediul regulilor flexionare și pe baza opoziției dintre forme regulate vs. forme neregulate. Formele neregulate sunt înregistrate în baza de date a aplicației, nefiind necesară o procedură de flexionare, ci doar una declarativă, strict de memorare. În schimb, formele regulate sunt clasificate pe baza comportamentului morfosintactic, iar regulile au în vedere șabloane de flexiune tratate procedural.

Elaborarea secțiunii de sintaxă a grupului verbal pentru limbile română și engleză are la bază o etapă de cercetare în care am definit constituenții grupului verbal, caracterul obligatoriu sau facultativ de ocurență în cadrul grupul verbal și diferite tipare verbale.

Definirea determinanților verbali a fost urmată de alegerea unui instrument pentru reprezentarea grafică a arborilor sintactici, selecția oprindu-se asupra programului Linguistic Tree Constructor, preluat din catalogul de programe al site-ului www.sil.org. Implementarea ulterioară în mediul de lucru Delphi a regulilor de identificare a grupului verbal, create în etapa de cercetare și realizarea aplicației AGV1.0 (Analizor de Grup Verbal 1.0) au constituit etapa finală a studiului de traducere automată a grupului verbal.

Doresc sa multumesc I.C.I.A. pentru acceptul de a folosi în lucrarea de față modulul de preprocesare a textelor TTL, precum și celor care au contribuit la realizarea acestui program performant și eficient.

Lucrarea **Naturațea limbajului artificial - studiu despre traducerea automată a GV** reprezintă o experiență care are nevoie de un punct terminus pentru a trece la un alt nivel, din perspectiva căruia să fie reorganizate conceptele și modalitățile de lucru. O direcție viitoare de lucru poate fi aceea de a demonstra că limbajul natural este un sistem adaptiv complex, păstrând modul de realizare a grupului verbal, dar receptându-l dintr-o altă dimensiune.