REVISTA DE ANALIZĂ NUMERICĂ ŞI TEORIA APROXIMAŢIEI
Volumul 3, Fascicola 1, 1974, pp. 35—51

UN ALGORITM PENTRU GENERAREA CUVINTELOR LIMBII ROMÂNE

de
MINERVA BOCŞA
(Timişoara)

§ 1.

Se ştie că în numeroase limbi naturale cuvintele pot fi clasificate în flexibile şi neflexibile. *Cuvîntul neflexibil* are o singură realizare pe planul expresiei. *Cuvîntul flexibil* poate fi conceput ca o mulţime de elemente — numite forme flexionare ale cuvîntului unitate lexicală — între care există raportul pe plan morfematic şi semantic. În numeroase împrejurări cînd se enumeră cuvintele unei limbi sau porţiuni de limbă, în dicţionare, liste de cuvinte, analize de texte, etc., se au în vedere doar *cuvintele unităţi lexicale*. Se presupune că cel care le utilizează poate reconstitui întreaga mulţime de forme flexionare, adică întreaga *paradigmă* a fiecărui cuvînt.

În prelucrarea informaţiei lingvistice cu ajutorul calculatorului, traducere automată, rezumare automată, stabilirea caracteristicilor cantitative ale limbii ş.a. e nevoie să se realizeze *generarea cuvintelor* şi *analiza* acestora prin program. *Generarea automată a cuvintelor* (sinteza) înseamnă producerea automată a întregii paradigme, pornind de la cuvîntul unitate lexicală, ca dată. *Analiza automată a cuvintelor* înseamnă stabilirea cuvîntului unitate lexicală, adică apartenenţa la o anumită paradigmă, pornind de la un cuvînt oarecare al limbii, ca dată.

Ne-am propus să rezlovăm generarea cuvintelor limbii române cu ajutoul calculatorului IRIS 50. Ne-am mărginit la un număr de cca. 2000 cuvinte regulate, substantive, adjective şi verbe, majoritatea cuprinse în listele de cuvinte din „Morfologia structurală a limbii române" de Valeria Guţu Romalo, [6], avînd în vedere doar *formele sintetice* (compuse dintr-un singur cuvînt). Am ales limbajul ASSIRIS deoarece dispune de o listă largă de instrucţii adecvate problemei. Prezentăm în cele ce urmează algoritmul pentru generarea cuvintelor *flexiunii nominale şi verbale* din limba română.

§ 2.

Cuvîntul flexibil se obţine prin juxtapunerea a două elemente : o parte constantă, numită *radical* (rădăcină, temă) şi o parte variabilă, numită *flectiv* (terminaţie, modificator). Flectivele posibile se grupează în clase,

dînd naştere la *clasele flexionare* ale limbii. Avem substantive de declinarea I-a, a II-a, ... verbe de conjugarea I-a, a II-a ... Astfel substantivul MAMĂ din declinarea I-a are ca radical MAM- iar seria de flective este -Ă, -E, -A, -EI, -ELE, -ELOR. Paradigma substantivului MAMĂ este formată din cuvintele distincte MAM-Ă, MAM-E, MAM-A, MAM-EI, MAM-ELE, MAM-ELOR.

Pentru realizarea cuvintelor ar fi suficient să separăm radicalul şi să ataşăm succesiv seria de flective. Un asemenea algoritm a fost utilizat de ERICA NISTOR DOMOKOŞ [11], în experimentele sale de traducere automată din limba engleză în limba română pe calculatorul MECIPT-1, în 1962 şi anii următori. O variantă a sa mai amplificată a fost la baza sintezei substantivelor ruseşti cu calculatorul DACICC-1, imaginată de Paul Schveiger şi LIVIU NEGRESCU, 1966, [12].

Limba română însă are multe neregularităţi care nu permit generalizarea acestui algoritm. Există cuvinte care nu au un radical constant. Cuvîntul MASĂ are radicalele MAS-(Ă) şi MES-(E), cuvîntul MĂR admite radicalele MĂR-(Ø)*) şi MER-(E), cuvîntul FLOARE are radicalele FLOAR-(E) şi FLOR-(I), cuvîntul CAL are radicalele CAL-(-Ø, -UL, -ULUI) şi CA-(-I, -II, -ILOR). Aceste forme diferite ale radicalului se numesc *alomorfe* iar modificările de tip A-E, Ă-E, A-Ø, L-Ø, etc. se numesc *alternanţe fonetice*.

S-a stabilit că în limba română radicalele substantivale au maximum două alomorfe, cele adjectivale au trei, FRUMOS-Ø, FRUMOŞ-I, FRUMOAS-E) iar verbele ajung la cinci şi şase alomorfe, fără totuşi să prezinte ceea ce se numeşte o variaţie aberantă. (VED-EA, VĂD-Ø, VEZ-I, VAD-Ă, VĂZ-ÎND; RAD-E, RAZ-I, RĂD-EAM, RĂ-SEI, RA-SE, RĂZ-ÎND).

Alternanţele fonetice (vocalice şi consonantice) se produc foarte frecvent şi nu sînt circumscrise decît aproximativ de condiţionări fonetice. (CAS-Ă, apropiat de MAS-Ă şi din aceeaşi declinare are totuşi un singur radical, ŞACAL- din aceeaşi declinare cu CAL-Ø nu-l pierde pe L în cursul paradigmei ş.a.m.d.). Nici segmentarea cuvîntului în alt mod — problemă controversată — nu detaşează net partea fixă din cuvînt de partea variabilă. Singurul mod de descriere utilizabil în problema noastră este *stabilirea unui inventar complet* de astfel de alternanţe şi stabilirea repartiţiei valorilor în paradigmă.

Considerăm că o asemenea descriere se face în cel mai economic mod utilizînd l i t e r e l e v a r i a b i l e, noţiune concepută de GR. C. MOISIL în cercetările sale legate de problemele traducerii automate [9], [10]. Litera variabilă T/Ţ din BĂRBAT-Ø are valoarea *(realizarea)* T în (a) 1-Nom., sing., neart., 2-Gen., sing., neart., 3-Dat., sing., neart., 4-Ac., sing., neart., etc. şi valoarea Ţ în (b) 1-Nom. pl., neart., 2-Gen., pl., art., etc. Astfel întîlnim litera variabilă A/Ă din PACE (PA/ĂCE), care dă PACE, PĂCI; literele variabile Z/J şi A/Ø din VITEAZ (VITEA/ØZ/J) care dă VITEAZ, VITEJI; literele variabile Ş/S, T/C, Ø/A din CREŞTE (CREØ/AŞ/ST/CE) care dă CRESC, CREŞTI, CREASCĂ ş.a.m.d.

---

* Simbolul vid, lanţul vid, numit şi zero.

Regăsim literele variabile în [6] unde au servit la descrierea structurală a substantivului, adjectivului şi verbului.

Sinteza automată a cuvintelor trebuie să prevadă două aspecte:

I. o *codificare* a întregii informaţii necesare flexionării;

II. un *program de decodificare* şi de construire a paradigmei.

§ 3.

Codificarea informaţiei se face ţinînd cont de fazele operaţiei:

A) cuvîntul de pe cartela perforată se introduce în memoria centrală.

B) programul prelucrează cuvîntul dat;

C) se tipăreşte la imprimantă paradigma obţinută.

Se ştie că la calculatorul IRIS 50 cititorul de cartele preia informaţia de intrare de pe cartelele perforate într-un anumit cod (Hollerith) şi o introduce în memoria centrală unde se codifică în EBCDIC.

Codificarea *literelor variabile* trebuie să conţină următoarele elemente: a) *realizările* (valorile efective) ale literei variabile în toate alomorfele radicalului; b) repartizările radicalului în paradigmă. Astfel codificarea trebuie să permită construirea paradigmei după schema (*). Aici $\overline{AL_i}$ cu $i = 1, n$ sînt cele n alomorfe ale radicalului, FL_{ij} cu $j = 1, k_i$ sînt cele k_i flective ce se ataşează radicalului AL_i iar P^g_{ij} sînt poziţiile

$$\text{CUVÎNT} \begin{cases} AL_1 \begin{cases} -\,FL_{11} & P^g_{11} \\ -\,FL_{12} & P^g_{12} \\ \dots\dots\dots \\ -\,FL_{1k_1} & P^g_{1k_1} \end{cases} \\ AL_2 \begin{cases} -\,FL_{21} & P^g_{21} \\ \dots\dots\dots \\ -\,FL_{2k_2} & P^g_{2k_2} \end{cases} \\ \dots\dots\dots \\ AL_n \begin{cases} -\,FL_{n1} & P^g_{n1} \\ \dots\dots\dots \\ -\,FL_{nk_n} & P^g_{nk_n} \end{cases} \end{cases}$$

Schema de formare a paradigmei. (*)

din paradigmă ce se realizează prin lanţul $AL_i\,FL_{ij}$. Numărul g arată gruparea poziţiilor P_{ij} în cazul concret al fiecărui cuvînt în parte.

La *substantive* gruparea elementelor P_{ij} e simplă, reducîndu-se la două (trei) situaţii standard, substantive masculine şi feminine, excepţiile fiind rare. Numărul total de forme flexionare de la substantive este la noi 16:

număr — sing. sau plur. —, *caz* — nom. gen., dat., ac., — și *determinarea* — neart., art. (nu am considerat vocativul). Dar numeroasele omonimii reduc paradigma la 6 (7) cuvinte diferite. Cele două alomorfe posibile se repartizează în grupările (1) și (2) după cum urmează:

(1) Substantive masculine și neutre

AL_1 { CM1: sing., [nom., gen., dat., ac.,] neart. (BĂIAT-Ø)
CM2: sing., [nom., ac.,] artic. (BĂIAT-UL)
CM3: sing., [gen., dat.,] artic. (BĂIAT-ULUI)

. . . . . .

AL_2 { CM4: pl., [nom., gen., dat., ac.,] neart. (BĂIEȚ-I)
CM5: pl., [nom., ac.,] art. (BĂIEȚ-II)
CM6: pl., [gen., dat.,] art. (BĂIEȚ-ILOR)

(2) Substantive feminine

AL_1 { CF1: sing., [nom., ac.,] neart. (FAT-Ă)
CF2: sing., [nom., ac.,] art. (FAT-A)

. . . . . .

AL_2 { CF3: { CF3′: sing., [gen., dat.,] neart.
CF3″: pl., [nom., gen., dat., ac.,] neart. (FET-E)
CF4: sing., [gen., dat.,] art. (FET-EI)
CF5: pl., [nom., ac.,] art. (FET-ELE)
CF6: pl., [gen., dat.,] art. (FET-ELOR).

În cazuri izolate la substantivele feminine distingem CF3′ diferit de CF3″ (cuvintele SARE, CARNE), care fac să avem 7 cuvinte diferite în paradigmă.

Cuvîntul VRABIE), cere însă o altă grupare, diferită de (1) și (2), deoarece CF4 are același alomorf cu CF1 (VRABIEI). (Circulă paralel și VRĂBIEI, întocmai ca și ARAMEI și ARĂMII).

Din analiza celor circa 1.300 *substantive* am desprins în total două grupări obișnuite, (1) și (2) și încă două grupări *aberante* (3) și (4). Aceasta ne-a condus la codificarea și programul expus în [1].

Adjectivele studiate (circa 250) au la noi 32 de forme flexionare, masculine și feminine, respectînd în principiu aceleași omonimii ca substantivele și cu un număr mic de grupări ale elementelor paradigmei, care ne-a condus la codificarea și programul expus în [2].

Verbele studiate (circa 450) prezintă însă o bogată variație a acestor grupări. [6]. De exemplu la verbe cu *două alomorfe* ale radicalului avem grupările:

(1) AL_1 { infinitiv (LUPT-A)
ind. prez. I. sing. (LUPT-Ø)
ind. prez. III. sing. (LUPT-Ă)
(toate celelalte) (LUPT-...)

. . . . . .

AL_2 { ind. prez. II. sing. (LUPȚ-I)

— — — — — — — — — — — — — — — —

(2) AL_1 { infinitiv (APĂR-A)
ind. prez. I. sing. (APĂR-Ø)
ind. prez. III. sing. (APĂR-Ă)
(toate celelalte) (APĂR-...)

. . . . . .

AL_2 { ind. prez. II. sing. (APER-I)
conj. prez. III. sing. (APER-E)
conj. prez. III. pl. (APER-E)

etc.

Pentru verbele cu două alomorfe avem grupările numerotate (1), (2), (11), la cele cu trei alomorfe avem grupările (1), ... (22) ș.a.m.d.

Literele variabile figurează în alomorfe conform unor configurații, care de asemenea diferă mult. Astfel literele E/I și T/Ț din verbul A PREZENT-A și literele A/Ă și C/Ø din verbul A FAC-E (ambele cu trei alomorfe ale radicalului) dau *realizările*:

AL_1 (a)	AL_2 (b)	AL_3 (c)
PREZENT-A	PREZINT-Ø	PREZINȚ-I
FAC-E	FĂC-EAM	FĂØ-Ø

Avem — notînd alomorfele cu (a), (b), (c) — literele

E/I în configurația (a)/(b)(c)
T/Ț în configurația (a)(b)/(c)
A/Ă în configurația (a)/(b)(c)
C/Ø în configurația (a)(b)/(c).

Literele variabile de la trei alomorfe prezintă și alte configurații (a)(c)/(b), (a)/(b)/(c), la patru alomorfe avem configurațiile (a)(b)/(c)(d), etc. *Numerotarea* alomorfelor AL_1, AL_2 ... AL_n, respectiv (a), (b), (c), ..., se face după *temeiuri lingvistice* așa încît nu putem — renumerotîndu-le — să considerăm ca egale configurațiile (a)(b)/(c) și (a)(c)/(b).

În vederea codificării tuturor informațiilor pentru substantive, adjective și verbe, am exploatat posibilitățile calculatorului IRIS 50 care admite ca unitate minimă adresabilă de informație *octetul* (opt cifre binare) format din cvartetul stîng (ponderea tare, poids fort) și cvartetul drept (ponderea slabă, poids faible). Am alcătuit un inventar al *literelor variabile* întîlnite, care se ridică la 82 și am întocmit ceea ce vom numi *alfabetul generalizat al limbii române*. Acest alfabet conține literele:

A, Ă, Â, B, C, D, E, F, G, H, I, Î, J, K, L, M, N, O, P, Q, R, S, Ș, T, Ț, U, V, W, X, Y, Z, blanc*) precum și cele 82 litere variabile din care dăm mai jos cîteva:

Nr.	*Lit.*	*Nr. alom.*	*Config.*	*Cuvînt*
1	A/Ă	2	(a)/(b)	CART-E/CĂRȚ-I
9	Ø/A	2	(a)/(b)	MOTOR-Ø/MOTOAR-E
23	N/Ø	2	(a)/(b)	SPUN-E/SPU-I
25	A/Ă	3	(a)(b)/(c)	BAT-E, BAȚ-I/BĂT-EAM
49	U/O	3	(a)/(b)(c)	RUG-A/ROG-Ø,ROAG-Ă
54	Ă/E/A	3	(a)/(b)/(c)	SPĂL-A/SPEL-I/SPAL-Ă
58	C/P	4	(a)(b)/(c)(d)	COC-Ø, COAC-E/COP-T, COAP-SERĂ
79	D/Z/Ø	5	(a)(b)/(c)/(d)(e)	ROAD-E, ROD-Ø/ROZ-I/RO-SEI, ROA-SE

Alfabetul obținut, inclusiv numărul alomorfelor radicalului și configurațiile, s-a codificat cu ajutorul codului CGCLR (Codul general al cuvintelor limbii române), ce figurează în tabelul 1.

Se observă că literele fixe au ponderea tare E și F, literele cu două alomorfe ale radicalului au ponderea tare C și D literele ce comută cu zero au ponderile tari 0, 1, 2, 3, 4, ș.a.m.d. În general tabelul nr. 1 conține o mare regularitate privind decodificarea, deoarece ponderea tare (coloana) arată numărul alomorfelor și în multe cazuri configurația, iar ponderea slabă (linia) arată valorile reale ale literelor variabile. Au existat însă litere a căror decodificare nu e așa regulată. Ele ocupă zona dublu încadrată de la coloanele 2, 3, 4, 5, 6, 8, C.

Alfabetul nostru poate fi lărgit dacă se vor analiza în acest mod toate cuvintele limbii române, literele variabile noi fiind plasate în spațiile libere din tabelul 1.

În tabelul 2 dăm elementele principale ale codului EBCDIC.

*) Literele Ă, Â, Î, Ș, Ț care nu figurează la calculatorul IRIS-50 sînt codificate cu semnele ∂, %, |, $, &.

§ 4.

In memoria internă a calculatorului cuvîntul va fi reprezentat în CGCLR. Pe cartelă îl vom codifica astfel:

a) Literele fixe le vom reprezenta printr-un caracter corespunzător literei din alfabetul obișnuit;

b) Literele cu variație regulată vor fi codificate prin trei caractere: un *prefix* format de două cifre zecimale, indicînd coloana din tabelul 1, urmat de una din valorile *realizărilor*. De ex. litera nr. 1 (A/Ă) este 13A, litera nr. 9 (Ø/A) este 00A, litera nr. 25 (A/Ă) este 11A, litera 49 (U/O) este 08U ș.a.m.d.

c) Literele cu variație neregulată vor fi codificate cu patru caractere: *prefixul* format de două cifre zecimale ca la b), o cifră zecimală *separator*, apoi o literă, una din valorile *realizărilor*. De ex. litera nr. 32 (Ø/A) este 020A, litera nr. 63 (A/Ø) este 030A, litera nr. 64 (A/Ø) este 031A, litera nr. 58 (C/P) este 062C ș.a.m.d.

Pe fiecare cartelă figurează un singur cuvînt în felul următor:

col. 1—25: cuvîntul cu literele codificate ca la a), b), c).
col. 26—27: LCV, lungimea cuvîntului de pe cartelă, două cif. zec.
col. 28—29: LRD, lungirea radicalului din forma internă, două cif. zec.
col. 30: NAL, numărul alomorfelor radicalului, o cif. zec.
col. 31—32: GRP, grupările alomorfelor radicalului, două cif. zec.
col. 33: PV, partea de vorbire care poate fi numai S (subst.), A (adj.) sau V (verb), o literă.
col. 34: GT, gen pentru substantive, blanc pentru adjective și tip pentru verbe, o literă sau blanc.
col. 35—36: CLF, clasa flexionară, arătînd declinarea sau conjugarea, două cif. zec.
col. 37—80: neutilizat.

De exemplu avem cuvintele ȚARĂ, SPĂLA, codificate pe cartelă astfel:

col.	1,	2,	3,	4,	5,	6,	7,	8,	...	26,	27,	28,	29,	30,	31,	32,	33,	34,	35,	36
	Ț	1	3	A	R	Ă				0	6	0	3	2	0	2	S	F	1	3
	S	P	0	8	4	Ă	L	A		0	8	0	4	3	0	4	V	A	0	1

§ 5.

Algoritmul propus de noi pentru realizarea fazelor A, B, C, are etapele:

I_A — Codifiarea informației de intrare (de pe cartelă) din codul EBCDIC în CGCLR..
II'_B — Formarea alomorfelor radicalului.
II''_B — Atașarea flectivelor
III_C — decodificarea din CGCLR în codul EBCDIC și tipărirea informației de ieșire.

Tabelul 1

Codul CGCLR

Pondere tare / slabă	0 0000	1 0001	2 0010	3 0011	4 0100	5 0101	6 0110	7 0111	8 1000	9 1001	A 1010	B 1011	C 1100	D 1101	E 1110	F 1111
0 0000	$Ø/_{A}^{9}$	$A/_{Ø}^{10}$	$Ø/_{A}^{43}$ ac/b	$Ø/_{A}^{74}$ abc/d	$A/_{Ø}^{77}$ ae/bcd	$A/_{Ă}^{80}$ abe/cdf	$Ă/_{A}^{62}$ ad/bc	$A/_{Ă}^{56}$ ab/cd	$A/_{Ă}^{45}$	$Ă/_{A}^{46}$		$A/_{Ă}^{25}$	$Ă/_{A}^{2}$	$A/_{Ă}^{1}$	A	Â
1 0001			$Ø/_{A}^{32}$ ab/c	$A/_{Ø}^{63}$ ad/bc				$Ă/_{E}^{57}$ ab/cd			$Ă/_{E}^{36}$		$E/_{Ă}^{4}$	$Ă/_{E}^{3}$	Ă	Ŧ
2 0010				$A/_{Ø}^{69}$ a/bcd											E	K
3 0011								$I/_{Î}^{64}$ ac/bd	$Î/_{I}^{47}$					$Î/_{I}^{5}$	Î	M
4 0100	$Ø/_{J}^{11}$														I	Q
5 0101	$Ø/_{U}^{12}$						$O/_{U}^{71}$ abc/d	$U/_{O}^{68}$ a/bcd	$U/_{O}^{49}$						U	R
6 0110															O	V
7 0111							$Ș/_{Ș}^{72}$ abc/d	$Ș/_{S}^{65}$ ac/bd		$Ș/_{S}^{50}$	$S/_{Ș}^{38}$	$S/_{Ș}^{28}$		$S/_{Ș}^{18}$	Ș	W
8 1000															Ș	X
9 1001		$G/_{Ø}^{21}$	$G/_{Ø}^{33}$ ab/c	$G/_{Ø}^{60}$ ab/cd											G	Y
A 1010		$C/_{Ø}^{22}$	$G/_{Ø}^{34}$ ab/c					$T/_{C}^{66}$ ac/bd		$T/_{C}^{51}$	$C/_{T}^{39}$	$C/_{T}^{29}$		$C/_{T}^{14}$	C	P
B 1011							$A/_{Ă}^{61}$ ad/bc	$T/_{Ț}^{73}$ abd/c	$Ă/_{A}^{26}$ ab/c	$Ț/_{T}^{52}$	$T/_{Ț}^{40}$	$T/_{Ț}^{30}$	$Ț/_{T}^{16}$	$T/_{Ț}^{15}$	T	blanc
C 1100			$B/_{Ø}^{35}$ ab/c				$Ă/_{A}^{67}$ a/bcd		$E/_{A}^{27}$ ab/c				$A/_{E}^{6}$		Ț	B
D 1101		$N/_{Ø}^{23}$		$N/_{Ø}^{70}$ a/bcd	$D/_{Z/Ø}^{55}$ a/b/c		$C/_{P}^{58}$ ab/cd	$D/_{Z}^{59}$ ab/cd	$E/_{A}^{37}$ ac/b	$Z/_{D}^{53}$	$D/_{Z}^{41}$	$D/_{Z}^{31}$	$E/_{A}^{7}$	$D/_{Z}^{17}$	D	N
E 1110	$Ø/_{H}^{20}$				$D/_{Z/Ø}^{79}$ ab/c/de	$D/_{Z}^{78}$ abd/ce	$Ă/_{E/A}^{75}$ a/bd/c		$E/_{I}^{48}$ a/bc		$Z/_{J}^{42}$		$Î/_{A}^{8}$	$Z/_{J}^{18}$	Z	H
F 1111		$L/_{Ø}^{24}$	$L/_{Ø}^{44}$ ac/b		$D/_{Z/Ø}^{82}$ ac/bf/de	$E/_{Ă/A}^{81}$ ac/be/d	$Î/_{Ă/A}^{76}$ ab/c/d		$Ă/_{E/A}^{54}$ a/b/c				$G/_{P}^{19}$		J	L
Config.	a/b	a/b							a/bc	a/bc	ac/b	ab/c	a/b	a/b		
	Comutare cu zero															
Nr. alom.	2 alomorfe		3 alom.	4 alom.	Cazuri speciale		4 alomorfe		3 alomorfe				2 alom.		Litere fixe	
Prefix	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15

Tabelul 2.
Codul EBCDIC

Pondere tare / slabă	0 0000	1 0001	2 0010	3 0011	4 0100	5 0101	6 0110	7 0111	8 1000	9 1001	A 1010	B 1011	C 1100	D 1101	E 1110	F 1111
0 0000					blanc	&										0
1 0001													A	J		1
2 0010													B	K	S	2
3 0011													C	L	T	3
4 0100													D	M	U	4
5 0101													E	N	V	5
6 0110													F	O	W	6
7 0111													G	P	X	7
8 1000													H	Q	Y	8
9 1001													I	R	Z	9
A 1010																
B 1011						$										
C 1100							%	@								
D 1101																
E 1110																
F 1111																

Nucleul întregului program îl formează instrucția de t r a d u c e r e și t e s t a r e pe care o utilizăm de la stînga spre dreapta (TRTR, Translate and test). Această instrucție prelucrează un lanț de caractere de lungime dată, executînd asupra lui ambele sau una sau cealaltă din funcțiuni: *traduce*, adică înlocuiește fiecare caracter (octet) cu un alt caracter, conform unei tabele de corespondențe;
testează, adică verifică dacă octetul are sau nu o proprietate.

Proprietatea poate fi foarte variată: octetul dat, întreg sau orice grupare de biți ai săi, selecționați prin cifrele 1 ale unui alt octet, denumit *mască de comparare*, se compară cu un al treilea octet, numit *criteriu de comparare*. De exemplu cifrele din ponderea tare (biții 0, 1, 2, 3) se selecționează prin masca 1111 0000, iar cifrele din ponderea slabă (biții 4, 5, 6, 7) se selecționează prin masca 0000 1111. Comparația se face cu biții corespunzători din octetul „criteriu de comparare" și este de forma: caracterul dat este $<$, $\leqslant$, $=$, $>$, $\geqslant$, $\neq$ față de criteriul de comparare. Execuția instrucției se oprește în interiorul lanțului, cînd s-a aflat proprietatea testată, sau dacă s-a epuizat tot lanțul fără găsirea caracterului căutat.

Etapa I_A *a algoritmului* are schema logică din figura 1.

Instrucția TRTR are aici ambele funcții. Traduce literele cuvîntului de lungime LCV (șirul de lungime L) după o tabelă de decodificare (tabelul 3). În același timp testează prezența cifrei în lanț (adică a literei variabile), verificînd dacă există un caracter cu ponderea tare egală cu F. (vezi tabelul 2). Apoi se înlocuiesc fie următoarele trei, fie patru caractere cu un singur octet, conform cu tabelul 1.

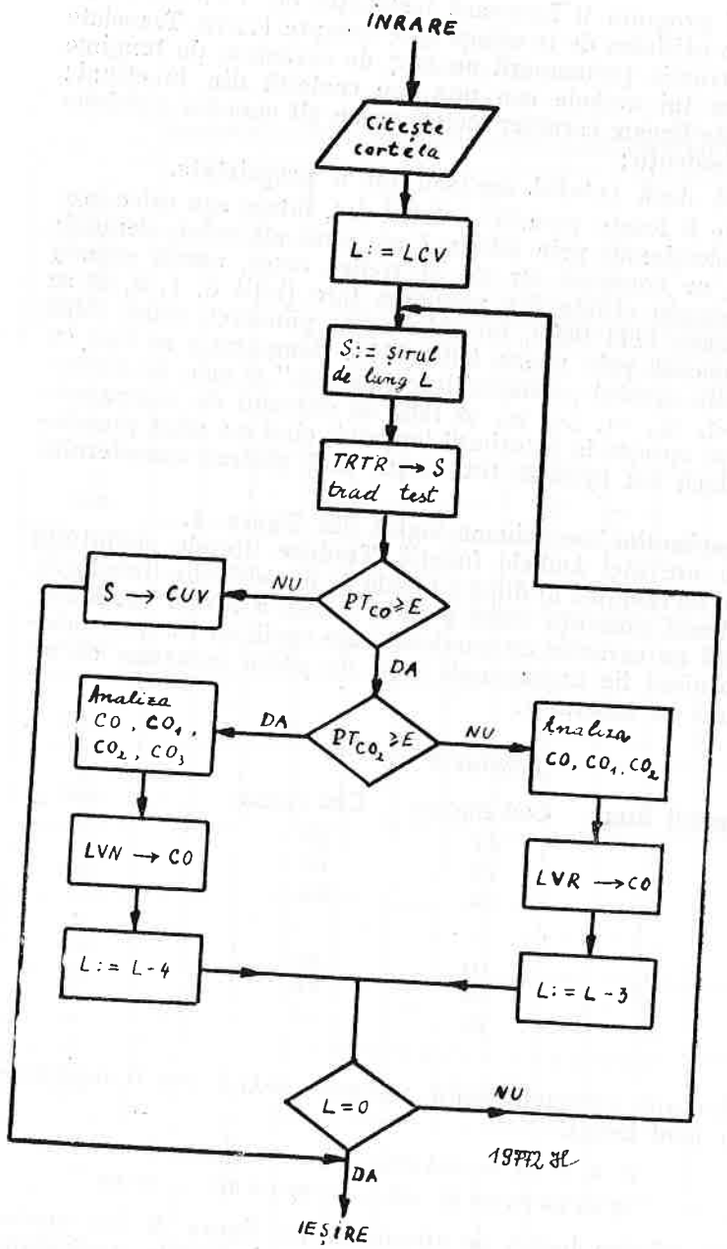
Tabelul 3

Simbol litera	Cod EBCDIC	Cod CGCLR
A	C1	E0
B	C2	FC
C	C3	EA
. . . . .		
M	D4	F3
N	D5	FD
P	D7	FA
. . . . .		

După această etapă cuvintele ȚARĂ, FRUMOS, SPĂLA, vor fi reprezentate în calculator (cod hexa).

Ț A/Ă R Ă; F R U M O Ø/A S/Ș; S P Ă/E/A L A
EC D0 F5 E1; F2 F5 E5 F3 E6 21 A7; E7 FA 8F FF E0

Etapa II'_B are schema logică de ansamblu în figura 2, iar rutina de analiză și depunerea literelor variabile în alomorfele radicalului în figura 3.



LCV = lungimea cuvîntului
Ș = șirul ce se prelucrează prin TRTR
CO = caracter de oprire
CO_1, CO_2 = caractere următoare
PT = ponderea tare
LVR = lit. var. regulată
LVN = lit. var. neregulată
CUV = zona de depunere din memorie

Figura 1. Rutina de codificare din EBCDIC în CGCLR

LRD = lungimea radicalului
NAL = nr. alomorfelor radicalului
PT_{co} = ponderea tare a caracterului de oprire
AL_i = alomorful radicalului de rang i

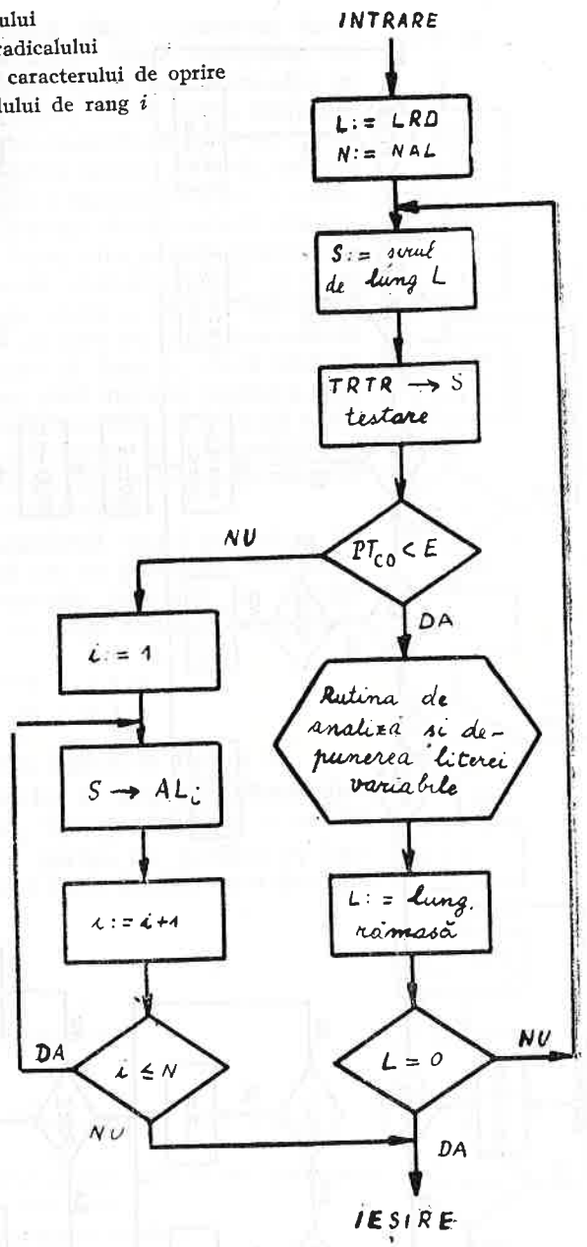


Figura 2. Rutina de formare a alomorfelor radicalului.

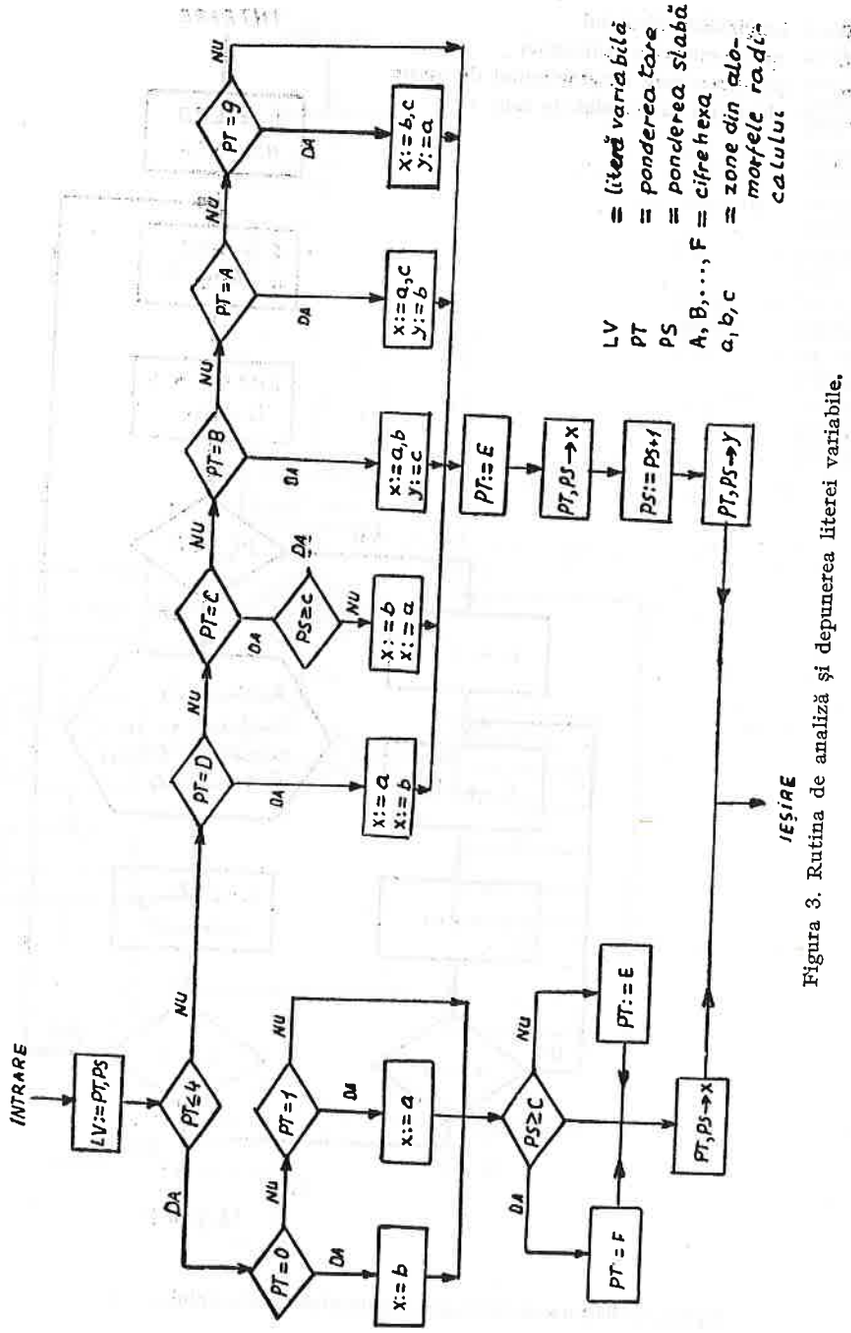


Figura 3. Rutina de analiză și depunerea literei variabile.

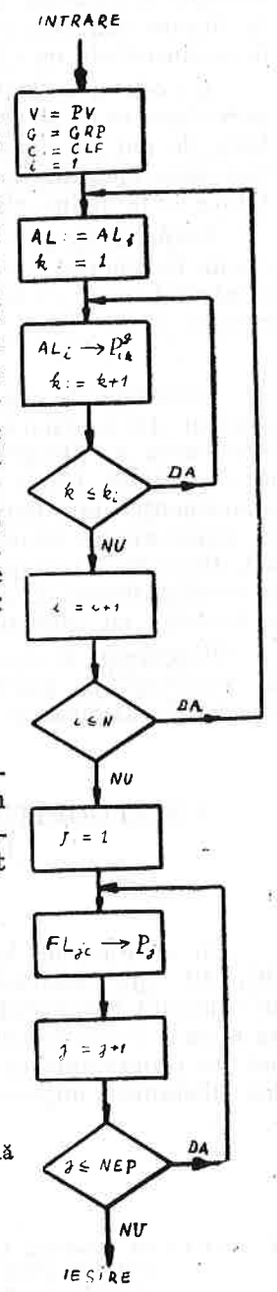
Instrucția TRTR testează (fără traducere) dacă șirul codificat în CGCLR conține litere variabile, cu pondere tare mai mică decît E. În caz afirmativ se analizează caracterul de oprire, adică litera variabilă aflată, dacă e regulată, dacă comută cu zero, etc., testînd pondere tare și ponderea slabă. *Literele regulate ce nu comută cu zero* permit o simplă formare a *realizărilor*. Acestea sînt întotdeauna două, ambele din coloana E și una pe aceeași linie, alta pe linia următoare (tabelul 1). Deci cele două realizări vor fi E,PS și E, PS + 1 (PS este ponderea slabă a literei variabile). *Literele regulate ce comută cu zero* au *realizarea nevidă* pe aceeași linie, în coloana E (sau F, dacă sînt în zona punctată). *Realizarea vidă* impune mutarea spre stînga cu un pas a lanțului următor. Cele două *realizări* ale literei variabile se depun în alomorfele AL_i conform configurației, care de fapt servește la calculul adreselor din interiorul alomorfelor.

În această etapă cuvîntul ȚARĂ (EC DO F5 E1) prezintă caracterul testat DO cu ponderea tare D<E. Fiind un caracter cu variație regulată DO devine întîi E0 iar apoi devine E1. Cele două alomorfe sînt:

EC E0 F5 = Ț A R
EC E1 F5 = Ț Ă R

Etapa II''_B are schema logică în figura 4.

Printr-un ciclu în ciclu se depune fiecare alomorf AL_i în zonele P^g_{ik} din paradigmă — ca în schema (*) — potrivit cu partea de vorbire PV (coloana 33) cu gruparea GRP (col. 31, 32) care de fapt



PV = partea de vorbire
GRP = gruparea alomorfelor radicalului
CLF = Clasa flexionară
AL_i = alomorful de rang i
P^g_{ik} = zona din paradigmă conform grupării GRP asociată radicalului AL_i
FL_{jc} = flectivul din clasa CLF asociat zonei P_j din paradigmă
N = numărul alomorfelor radicalului
NEP = numărul total al elementelor paradigmei

Fig. 4. — Rutina de atașare a flectivelor.

serveşte la calculul adreselor *zonelor* P_{ik} din interiorul paradigmei. Apoi la fiecare din cele N alomorfe ale radicalului AL_i se adaugă flectivele FL_{jc} în conformitate cu clasa flexionară CLF.

La cele N = 2 alomorfe ale radicalului cuvîntului ŢARĂ, care au gruparea GRP = 02 se ataşează seria de 2 + 4 flective din declinarea numerotată de noi 13, care sînt -A -Ă şi —I, —II, -ILE -ILOR. Ele se depun în cele şase zone ale paradigmei cu ajutorul instrucţiei *transfer de caractere* MVSR (Move byte string right).

Etapa III_C prin instrucţia TRTR traduce fără testare din codul CGCLR în EBCDIC lanţurile de caractere ce sînt depuse la paradigmă după o tabelă inversă cu tabela 3 şi apoi tipăreşte toată paradigma cu cele NEP cuvinte aflate.

§ 6.

Cuvintele de pe cartele pot alcătui un *fişier pe disc* ce-l vom denumi Dicţionarul morfologic al limbii române despre care este vorba în [9] şi [10]. Dicţionarul permite *sinteza* formelor flexionare ale cuvintelor. Dotat cu un algoritm de analiză el permite şi recunoaşterea apartenenţei unui cuvînt dat la o unitate lexicală. De exemplu, recunoaşte că POPOARELOR aparţine cuvîntului POPOR şi este CM6, pl., [gen., dat.,] art. Precizăm că proprietăţile codului CGCLR ne permit o asemenea analiză în condiţii destul de simple. Orice altă operaţie de sinteză a frazelor limbii, a textelor cu ajutorul ordinatorului utilizează acest dicţionar.

Programul în ASSIRIS al algoritmului descris mai sus se află în curs de verificare pe calculatorul IRIS 50 de la Centrul Teritorial de Calcul Electronic Timişoara.

UN ALGORITHME POUR LA GÉNÉRATION DES MOTS DE LA LANGUE ROUMAINE

RÉSUMÉ

En utilisant les lettres à valeurs variables, conçues par Gr. C. MOISIL [9], [10], qui decrivent les altérnances phonétiques des radicaux des mots on construit l'alphabet généralisé de la langue roumaine écrite. On a codé les mots à l'aide de cet alphabet en CGCLR (tableau nr. 1) et le décodage par un programme en ASSIRIS permet d'obtenir les formes flexionnelles des substantifs, adjectifs et verbes avec l'ordinateur IRIS 50 (FELIX C-256)

BIBLIOGRAFIE

[1] Minerva Bocşa, *Codage de l'alphabet généralisé du roumain écrit pour l'ordinateur IRIS 50 (FELIX C-256). Le substantif.* — Cahiers de linguistique théorique et appliquée, **2**, 1973.

[2] Minerva Bocşa, *Codage de l'alphabet généralisé du roumain écrit pour l'ordinateur IRIS 50 (FELIX C-256). L'adjectif.* — Cahiers de linguistique théorique et appliquée, **1**, 1974.
[3] *Documentation de l'ordinateur IRIS 50.*
[4] *Documentaţia calculatorului electronic FELIX C-256.*
[5] *Gramatica limbii române,* Editura Academiei Republicii Populare Române (Bucureşti), 1963.
[6] Valeria Guţu Romalo, *Morfologie structurală a limbii române, (Substantiv, adjectiv, verb),* Editura Academiei Republicii Socialiste România (Bucureşti), 1968.
[7] — *Indreptar ortografic, ortoepic şi de punctuaţie,* Editura Academiei Republicii Populare Române (Bucureşti), 1965.
[8] Solomon Marcus, *Limbă şi cod,* Studii şi cercetări lingvistice, **1**, 1966.
[9] Grigore C. Moisil, *Probleme puse de traducerea automată. Conjugarea verbelor în limba română scrisă.* Studii şi cercetări lingvistice, **1**, 1960.
[10] Grigore C. Moisil, *Problèmes posés par la traduction automatique. La déclinaison en roumain écrit.* Cahiers de linguistique théorique et appliquée. **1**, 1962.
[11] Erica Nistor Domokos, *Algoritm de traducere automată din limba engleză în limba română.* Editura Didactică şi Pedagogică (Bucureşti), 1966.
[12] Paul Schveiger, Vera Drondoe, *À propos de la vérification de la sous-routine de déclinaison mécanique des substantifs,* Revue Roumaine de Linguistique, **XII**, *1*, 1967.

Facultatea de Matematică-Mecanică
a Universităţii din Timişoara

Primit la 26.VI.1973.