EDITAREA TEXTELOR DIALECTALE FOLOSIND TRANSCRIEREA FONETICĂ SPECIFICĂ LIMBII ROMÂNE[1]

SILVIU-IOAN BEJINARIU[*], RAMONA LUCA[**], VASILE APOPEI[**], FLORIN IFTENE[*]

1. Introducere

Această lucrare este un omagiu dedicat Domnului Profesor Stelian-Traian Dumistrăcel și în același timp un rezultat al colaborării cu colectivul din care a făcut parte. Sunt prezentate cele mai importante provocări întâlnite şi soluțiile adoptate pe parcursul unei cooperări interdisciplinare inițiate în urmă cu mai mult de 20 de ani între Institutul de Informatică Teoretică și Institutul de Filologie Română „Alexandru Philippide" din cadrul Academiei Române, Filiala Iași. Dacă facem referire doar la domeniul geografiei lingvistice, colaborarea s-a materializat prin proiectarea și implementarea sistemului informatic care a permis publicarea a trei volume din *Noul Atlas lingvistic român, pe regiuni. Moldova și Bucovina* (NALR–Mold. Bucov.). Una dintre cele mai dificile şi incitante probleme a fost şi rămâne cea a editării transcrierilor fonetice specifice limbii române pentru care a fost proiectat un sistem de sinteză grafică în timp real a imaginii grafemelor. În această lucrare propunem un sistem inovativ pentru editarea transcrierilor fonetice în cel mai folosit editor de text. A fost implementată o aplicație pentru generarea automată a grafemelor pentru toate cele 112 791 de combinații de simboluri și fenomene fonetice în zona de coduri private a fonturilor TrueType și de asemenea un AddIn ce permite utilizarea acestora pentru scrierea fonetică în editorul Microsoft Word.

[1] This work presented in this paper was partially supported by the research grant from the Romanian National Authority for Scientific Research and Innovation CNCS = UEFISCDI, project no. PN-III-P4-ID-PCE-2020-0451.

[*] Institutul de Informatică Teoretică al Academiei Române – Filiala Iași/ Institutul de Lingvistică al Academiei Române „Iorgu Iordan – Alexandru Rosetti", București, România (silviu.bejinariu@iit.academiaromana-is.ro / florin.iftene@iit.academiaromana-is.ro).

[**] Institutul de Informatică Teoretică al Academiei Române – Filiala Iași, România (ramona.luca@iit.academiaromana-is.ro / vasile.apopei@iit.academiaromana-is.ro).

ALIL, t. LXII, 2022, Bucureşti, p. 37–53

Colaborarea interdisciplinară între cercetătorii de la Institutul de Informatică Teoretică și Institutul de Filologie Română „Alexandru Philippide” a început la inițiativa domnului profesor Ioan Florea, care a cerut sprijin în implementarea unui sistem de editare a transcrierilor fonetice pentru atlasele lingvistice. Pentru început au fost sistematizate și clasificate simbolurile utilizate în transcrierea fonetică. Deși numărul total de simboluri teoretic posibile, având în vedere aplicarea simultană de până la 5 fenomene fonetice, era suficient de mare, într-o primă etapă a fost încercată soluția de a proiecta manual câte un font pentru fiecare combinație de fenomene fonetice și chiar au fost realizate fonturile pentru simboluri cu unul și două fenomene fonetice. Având în vedere dificultatea procesului de editare din cauza numărului mare de fonturi utilizate, a fost proiectat un nou sistem de transcrieri fonetice bazat pe sinteza grafică în timp real a imaginii simbolurilor. Astfel, utilizând un singur font specializat ce conține doar glyph-urile de bază, imaginea unui grafem cu orice combinație de fenomene fonetice este sintetizată din imaginile componentelor sale (fenomene fonetice asociate). Din punctul de vedere al editării propriu-zise, este necesară introducerea grafemului de bază cu ajutorul tastaturii, după care fenomenele fonetice sunt selectate folosind comenzile disponibile pe bara de instrumente a aplicației. Folosind această modalitate de realizare a transcrierilor fonetice implementată în aplicația *ALR–MB* (Bejinariu *et al.* 2021b), colectivul de dialectologie al Institutului de Filologie Română „Alexandru Philippide” condus de domnul profesor Stelian Dumistrăcel a publicat trei volume ale atlasului lingvistic regional al Moldovei şi Bucovinei în anii 2007, 2014 și 2022 (vezi NALR–Mold. Bucov. V). Rețeaua conține 210 puncte, iar anchetele s-au desfășurat între 1960 și 1980 (Dumistrăcel 1976). În atlas sunt incluse planșe cu hărți analitice, hărți sintetice interpretative, precum și planșe cu material necartografiat (Olariu, Olariu 2010).

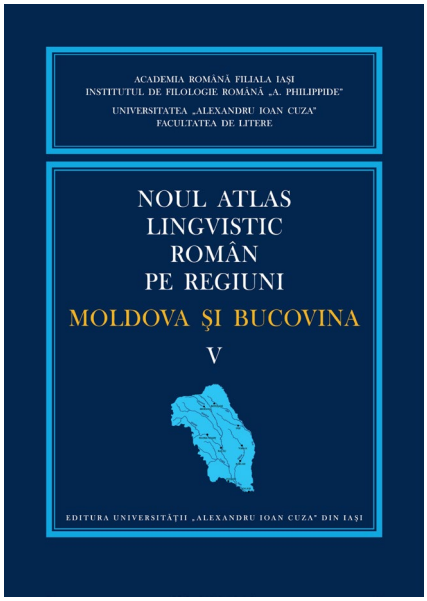

Figura 1. Noul Atlas lingvistic român pe regiuni. Moldova şi Bucovina, vol. V

Același sistem pentru editarea transcrierilor fonetice a fost implementat și în aplicația *AlrMaps*, utilizată de colegii de la Institutul de Lingvistică „Iorgu Iordan – Alexandru Rosetti” al Academiei Române (ILIR) pentru realizarea hărților lingvistice din volumul al doilea al lucrării *Atlasul lingvistic al dialectului aromân*, publicat în 2020 (ALAR II).

Aplicațiile *ALR-MB* și *AlrMaps* sunt specializate pentru introducerea informațiilor dialectale în bazele de date și redactarea asistată de calculator a planșelor atlaselor lingvistice. Cele două aplicații sunt însoțite și de editoare de text (*EditTD* și respectiv *AlrLibrary*), specializate în editarea textelor dialectale, care folosesc același sistem de realizare pentru grafemele utilizate în transcrierea fonetică. Deși sunt disponibile funcții de editare care răspund cerințelor de bază ale unui editor de texte, *EditTD* și *AlrLibrary* nu se pot compara cu un editor profesional. Aplicația *AlrLibrary* include și o componentă multimedia care permite parcurgerea textului cu redarea sincronizată a înregistrării audio în cazul în care aceasta este disponibilă.

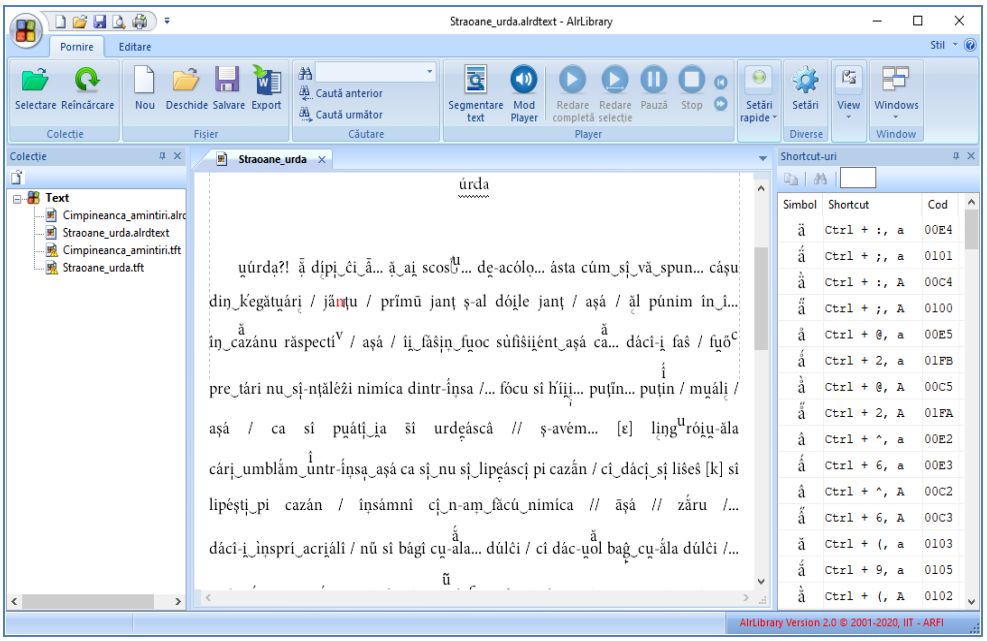

Figura 2. Interfața aplicației AlrLibrary pentru editarea textelor dialectale

Totuși, această metodă de editare în transcriere fonetică poate fi realizată doar în aplicațiile menționate, nu și în editoarele de text uzuale. Iar una dintre cerințele permanente, pe parcursul colaborării de peste 20 de ani cu cercetătorii dialectologi, a fost posibilitatea de a folosi transcrierile fonetice în Microsoft Word. Din acest motiv ne-am propus construirea de fonturi TrueType și implementarea în Word a unei interfețe prietenoase cu utilizatorul care să permită accesul facil la toate simbolurile posibile fără a fi necesară memorarea unui număr imens de shortcut-uri. În secțiunile următoare sunt descrise unele aspecte legate de implementarea componentei *Alr_Word_AddIn* pentru editarea textelor în transcriere fonetică în Microsoft Word.

2. Fonturi pentru editarea în Word a textelor dialectale

2.1. Simboluri de bază și fenomene fonetice

În transcrierea fonetică a limbii române sunt utilizate simboluri de bază ce corespund sunetelor primare având corespondent în setul de caractere obișnuit și simboluri ce corespund sunetelor primare afectate de unul sau mai multe fenomene fonetice. Lista acestor simboluri a fost sistematizată la începutul anilor 2000 de echipa interdisciplinară de cercetători de la Institutul de Filologie Română „Alexandru Philippide" și Institutul de Informatică Teoretică (Bejinariu *et al.* 2009).

Vocale

În tabelul de mai jos (Fig. 3) sunt prezentate cele 19 vocale primare utilizate în transcrierea fonetică a limbii române, cu observația că variantele de vocale din ultimele două coloane nu au fost implementate în versiunea inițială a sistemului. Pentru fiecare dintre aceste vocale există și câte trei variante accentuate (a - á, à, a̋), prin urmare numărul total de simboluri reprezentând vocale primare este de $19 * 4 = 76$.

Simple	Diacritice						
a	ä	ă	â	å		ȧ	
e	ë	ĕ					
i			î		î		ɪ
o	ö	ŏ					
u	ü		û				

Figura 3. Vocale primare folosite în transcrierea fonetică

Consoane

În Fig. 4 sunt prezentate simbolurile corespunzătoare consoanelor primare utilizate în transcrierea fonetică a limbii române.

b c ć ɕ ĉ č̆ č d ḑ đ δ f g ĝ ğ ğ̆ ɡ̂ γ h ħ ɧ χ j k l ɬ
m m̥ n n̥ ŋ ṇ p q r ɾ̆ ɾ̆ ρ s ŝ ś ş t ţ ϑ v w x y z ź ẑ

Figura 4. Consoane primare folosite în transcrierea fonetică

Fenomene fonetice

Fenomenele fonetice ce pot fi aplicate vocalelor sunt în număr de 12 și au fost clasificate în 5 grupe (Fig. 5). Ele pot fi aplicate simultan, dar nu mai mult de unul din fiecare grupă. În cazul vocalelor nu a fost posibilă definirea de reguli de combinare a fenomenelor și, prin urmare, folosind aplicațiile de editare, pot fi

realizate orice combinații, chiar dacă multe dintre ele nu sunt posibile în practică. Menționăm de asemenea că simbolul pentru sunet semivocalic este tratat în aplicație asemănător cu fenomenele fonetice. Astfel, în implementarea propriu-zisă sunt 13 fenomene fonetice împărțite în șase grupe.

În ceea ce privește consoanele, fenomenele fonetice care pot fi aplicate sunt în număr de nouă, împărțite tot în cinci grupe (Fig. 6). În acest caz nu pot fi aplicate mai mult de două fenomene simultan, care nu pot fi din aceeași grupă. Pentru consoane au fost stabilite unele reguli de combinare a fenomenelor fonetice, de aceea în procesul de editare unele combinații nu vor fi acceptate.

Grupe	Fenomen	Exemplu
Durată	Scurtime	ĕ é è ȅ
	Semilungime	e̋ é è ȅ
	Lungime	ē é è ȅ
Nazalizare	Seminazalizare	ẽ é è ȅ
	Nazalizare	ẽ é è ȅ
Ocluzie glotală	Coup de glotte	ʼe ʼé ʼè ʼȅ
Deschidere	Închidere	ę é è ȅ
	Semideschidere	ę́ é è ȅ
	Deschidere	ę́ é è ȅ
	Deschidere mare	ę́ é è ȅ
Afonizare	Semiafonizare	ę́ é è ȅ
	Afonizare	ę́ é è ȅ

Figura 5. Fenomene fonetice aplicate vocalelor, cu exemple (Bejinariu *et al*. 2009)

Grupe	Fenomen
Durată	Semilungime
	Lungime
Palatalizare	Semipalatalizare
	Palatalizare
	Palatalizare mare
Explozie	Explozie
Caracter silabic	Caracter silabic
Afonizare	Semiafonizare
	Afonizare

Figura 6. Fenomene fonetice aplicate consoanelor (Bejinariu *et al.* 2009)

Din punctul de vedere al simbolurilor grafice ce conțin vocale cu mai multe fenomene, acestea sunt aplicate în ordinea grupelor din care fac parte, iar eventualele accente sunt aplicate la sfârșit, așa cum este exemplificat în Fig. 7.

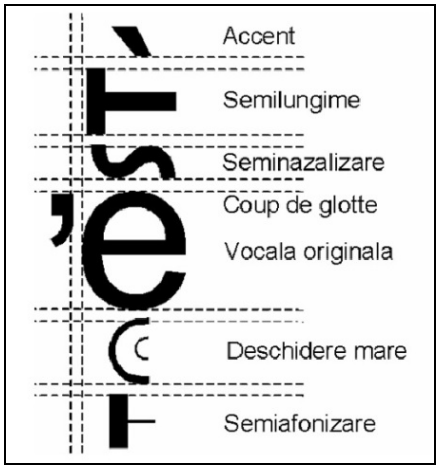

Figura 7. Exemplu de simbol căruia i-au fost aplicate fenomene fonetice (Bejinariu *et al.* 2009)

Este evident faptul că numărul de simboluri posibile este foarte mare (> 100 000), aproape imposibil de proiectat manual folosind un editor standard de fonturi. Folosind experiența în domeniul procesării de imagini din Institutul de Informatică Teoretică, a fost realizat un sistem de sinteză grafică în timp real a imaginii simbolurilor. Astfel, pentru realizarea transcrierilor fonetice este folosit un singur font TrueType pentru literele mici și un alt font pentru literele mari. Inițial au fost utilizate fonturile *ALR_Baza* și *ALR_Baza_Caps*, derivate din Arial, iar apoi au fost proiectate fonturile *ALR_MinionPro* și *ALR_MinionPro_Caps*, derivate din fontul Minion Pro. Sistemul a fost utilizat pentru publicarea ultimelor trei volume ale NALR–Mold. Bucov., precum și publicarea unui volum al *Atlasului lingvistic al dialectului aromân*. Totuși, acest sistem poate fi utilizat doar în aplicațiile specializate și din acest motiv propunem în continuare un sistem inovativ pentru editarea transcrierilor fonetice folosind editoare standard de text.

2.2. Codificarea Unicode pe 32 de biţi în fonturile TrueType

Codificarea Unicode pe 16 biţi utilizată anterior permite definirea a maximum $2^{16} - 1 = 65\,535$ glyph-uri. Totuși, în codificarea Unicode, o parte dintre caractere sunt standardizate prin convenția *Unicode Consortium* și nu este recomandabilă modificarea lor deoarece în anumite situații sistemul de operare poate modifica automat glyphul utilizat în afișare. În acest caz vor fi afișate caractere incorecte și, prin urmare, numărul de poziții disponibile este mult redus. Folosind codificarea UTF-32 (*32 bit Unicode Transformation Format*), numărul de poziții disponibile poate fi extins. Chiar dacă reprezentarea unui caracter este codificată pe lungimea fixă de 32 de biţi, o parte dintre aceștia au valoarea fixă 0 și nu pot fi utilizați. Astfel, doar 21 dintre cei 32 de biţi sunt disponibili pentru reprezentarea unui caracter (Unicode 2011).

Pozițiile care pot fi personalizate de utilizator sunt organizate în așa-numitele *Private Use Area* (*PUA*) – intervale de coduri care nu sunt asignate prin convenția *Unicode Consortium*, după cum urmează (limitele intervalelor sunt exprimate în hexazecimal):

– Private Use Area; Interval: U+0xE000 – U+0xF8FF; 6400 coduri,

– Supplementary Private Use Area A; Interval: U+0xF0000 – U+0xFFFFD; 65 534 coduri;

– Supplementary Private Use Area B; Interval: U+0x100000 – U+0x10FFFD; 65 534 coduri.

Aceste zone private vor fi utilizate pentru definirea celor 112791 glyph-uri pentru toate simbolurile specifice transcrierii fonetice din limba română și toate

variantele lor obținute prin aplicarea tuturor combinațiilor posibile de fenomene fonetice.

În mod evident, numărul de poziții disponibile în PUA este suficient pentru toate combinațiile de simboluri și fenomene. Dar, în funcție de numărul de fenomene fonetice aplicate (fie deasupra simbolului de bază, fie sub acesta), atributele Ascent şi Descent ale fonturilor TrueType au valori diferite. Dacă toate simbolurile ar fi în același fișier-font, cele două atribute vor avea valorile maxime, ceea ce se va traduce printr-o distanță mult prea mare între rândurile de text, chiar dacă numărul de fenomene aplicate este redus. Din acest motiv a fost aleasă soluția de a genera fonturi diferite în funcție de numărul de fenomene aplicate:

– *ALR_MP_0.ttf* – fontul de bază fără nici un fenomen aplicat. Acesta conține glyph-urile pentru simbolurile de bază precum și cele ale fenomenelor fonetice ce urmează a fi aplicate.

– *ALR_MP_1.ttf – ALR_MP_6.ttf* – fonturile pentru simbolurile cărora le-au fost aplicate de la 1 până la 6 fenomene fonetice. Numărul de 6 fenomene provine din cele 5 fenomene propriu-zise în cazul vocalelor plus semnul corespunzător sunetului semivocalic.

2.3. Proiectarea fontului de bază *ALR_MP_0.ttf*

Toate glyph-urile simbolurilor de bază sunt descrise pe pozițiile 0x0100 – 0x01BA (Fig. 8) din fontul *ALR_MP_0.ttf*. Pentru păstrarea compatibilității dintre fonturi și pentru a reduce numărul de schimbări de font în cursul editării, simbolurile de bază se află pe aceleași poziții și în celelalte 6 fonturi TrueType.

Aceste glyph-uri sunt editate folosind editorul de fonturi *FontForge*, un software gratuit de tip open-source. O parte dintre ele provin din fontul Mignon Pro, disponibil în sistemul de operare Windows, iar altele au fost editate manual cu aplicația *FontForge*. Între cele 175 de glyph-uri se află simbolurile de bază, dar și fenomenele fonetice, accentele și celelalte simboluri speciale provenite din alfabetul grecesc care sunt utilizate în transcrieri (codurile sunt exprimate în hexazecimal):

– Vocale, litere mici:	0x0100 – 0x0112
– Consoane, litere mici:	0x0113 – 0x0146
– Vocale, litere mari:	0x0148 – 0x015A
– Consoane, litere mari:	0x015B – 0x018E
– Alte simboluri:	0x0190 – 0x0198
– Fenomene pentru vocale:	0x01A0 – 0x01AC
– Accente:	0x01AE – 0x01b0
– Fenomene pentru consoane:	0x01B2 – 0x01BA

Figura 8. Glyph-urile simbolurilor de bază (captură din aplicația *FontForge*)

Crearea glyph-urilor reprezentând grafeme cu fenomene fonetice

Pentru a genera glyph-urile reprezentând simbolurile de bază cărora le-au fost aplicate fenomene fonetice s-a profitat de faptul că editorul de fonturi *FontForge* pune la dispoziție o interfață cu limbajul Python (FontForge 2022). A fost implementată o aplicație în limbaj Python care generează în mod automat toate glyph-urile reprezentând simbolurile compuse.

Generarea combinațiilor de fenomene fonetice

Având în vedere faptul că vorbim despre un număr foarte mare de simboluri, peste 112 000, chiar fiind împărțite în 6 fișiere font, este evident că nu poate fi utilizată o listă de corespondențe între coduri și fenomenele aplicate simbolului respectiv. De fapt, corespondența este stabilită folosind câteva formule de calcul. Pentru a simplifica aceste formule, poziționarea glyph-urilor în *Private Use Area* este definită după cum urmează mai jos, folosind scrierea hexazecimală. Deoarece numărul de simboluri este diferit în funcție de numărul de fenomene fonetice

aplicate, este indicată doar poziția de început a fiecărui grup în toate cele 6 fișiere font (*ALR_MP_1.ttf – ALR_MP_6.ttf*):

– Vocale, litere mici:	0x000F0000
– Consoane, litere mici:	0x000F2000
– Vocale, litere mici, cu accent grav	0x000F4000
– Vocale, litere mici, cu accent ascuțit:	0x000F8000
– Vocale, litere mici, cu accent dublu ascuțit:	0x000FC000
– Vocale, litere mari:	0x00100000
– Consoane, litere mari:	0x00102000
– Vocale, litere mari, cu accent grav:	0x00104000
– Vocale, litere mari, cu accent ascuțit:	0x00108000
– Vocale, litere mari, cu accent dublu ascuțit:	0x0010C000

După cum a fost deja menționat, a fost urmărită posibilitatea de a identifica rapid codul caracterului pentru orice combinație de fenomene fonetice, dar și invers, fenomenele fonetice corespunzătoare unui anumit cod de caracter. Astfel, valoarea Code / 0x4000 indică tipul caracterului (vocală sau consoană), dacă este literă mică sau mare, precum și tipul accentului (dacă există), în timp ce valoarea Code % 0x4000 identifică în mod unic caracterul și fenomenele fonetice aplicate.

Fenomenele fonetice pot fi aplicate în număr de cel mult 5+1 în cazul vocalelor, respectiv cel mult două în cazul consoanelor, dar nu mai mult de unul din fiecare grupă. Pentru a verifica aceste condiționări, au fost realizate următoarele:

a. A fost definit câte un set de măști pentru vocale, respectiv consoane, care modelează grupele de fenomene. Acestea sunt descrise în continuare folosind sintaxa limbajului Python. Măștile sunt reprezentate ca valori binare, fiecare poziție reprezentând un fenomen fonetic şi fiecare mască conținând toate fenomenele din aceeași grupă.

```
masks_vowels_phenomena = [
     0b0000000000000111,         # durată
     0b0000000000011000,         # nazalizare
     0b0000000000100000,         # ocluzie glotală
     0b0000001111000000,         # deschidere
     0b0000110000000000,         # afonizare
     0b0001000000000000          # semivocala
     ]

masks_consonants_phenomena = [
     0b0000000000000011,         # durată
     0b0000000001111100,         # palatalizare, explozie şi silabic
     0b0000000110000000          # afonizare
     ]
```

În cazul consoanelor, grupele *palatalizare*, *explozie* și *caracter silabic* sunt unificate într-o singură mască, deoarece fenomenele respective se exclud reciproc și, prin urmare, nu pot fi aplicate simultan.

b. Generarea combinațiilor de fenomene este realizată prin parcurgerea tuturor valorilor numerice între 1 și 2^{13}-1, în cazul vocalelor, respectiv 2^{9}-1, în cazul consoanelor. Folosind operații pe biți, sunt reținute doar valorile pentru care numărul de biți '1' este egal cu numărul de fenomene ce se dorește a fi aplicat. Apoi, este verificată consistența din punctul de vedere al grupelor de fenomene: se face AND pe biți cu fiecare dintre măștile corespunzătoare din listă, păstrând doar acele valori pentru care rezultatul operației are cel mult un bit '1' (echivalent cu faptul că se aplică cel mult un fenomen din fiecare grupă).

c. Pentru fiecare dintre valorile care au trecut testul (b) este generat glyph-ul corespunzător și adăugat în următoarea poziție liberă din font.

Generarea automată a glyph-urilor

Glyph-urile corespunzătoare simbolurilor compuse sunt realizate prin gruparea glyph-urilor componente în pozițiile și cu alinierile necesare, în funcție de tipul și modul de poziționare al fenomenelor respective. În aplicarea fenomenelor pentru simbolul de bază trebuie respectate câteva restricții:

– Ordinea de aplicare depinde de grupa din care face parte fenomenul;

– Poziționarea este diferită (sus, jos, stânga, dreapta), în funcție de grupa din care face parte fenomenul;

– Pentru vocale, accentele sunt plasate cel mai sus, deasupra tuturor fenomenelor propriu-zise. Acesta este și motivul pentru care accentele sunt tratate în implementare ca fenomene, fiind ultimele în ordinea de aplicare.

În Figurile 9 și 10 sunt exemplificate rezultatele generării automate a fonturilor pentru transcrierea fonetică a limbii române.

Figura 9. Rezultatul aplicării automate a fenomenelor pentru un grup de caractere (captură din aplicația *FontForge*)

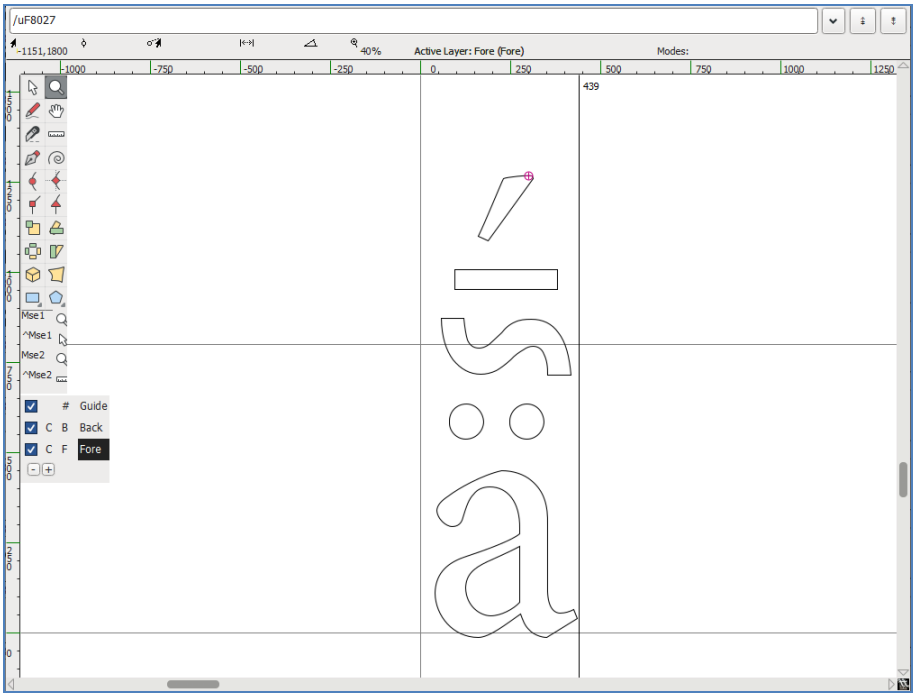

Figura 10. Exemplu cu rezultatul aplicării automate a fenomenelor pentru un singur caracter (captură din aplicația *FontForge*)

Add-in pentru editarea textelor dialectale în Microsoft Word

Editarea textelor dialectale presupune selecția caracterelor dintr-un set de peste 112 000 glyph-uri, codul acestora fiind rezultatul aplicării unei formule de calcul, precum și poziționări particularizate ale acestora, ceea ce este practic imposibil folosind doar tastatura. Prin urmare, a fost implementată o componentă *ALR_Word_AddIn* pentru Microsoft Word, care, după instalare, adaugă un panou în bara de instrumente a aplicației în care sunt puse la dispoziție toate comenzile de editare necesare (Fig. 11). Astfel, este posibilă păstrarea compatibilității cu alte aplicații, precum și utilizarea formatului .docx standard. Un alt avantaj față de versiunile anterioare ale editoarelor de texte dialectale este acela că selecția caracterelor de bază este realizată exclusiv folosind comenzile disponibile pe bara de instrumente și nu necesită memorarea niciunui shortcut sau combinații de taste. O versiune preliminară a acestei componente este prezentată și în Bejinariu *et al.* 2021a.

Figura 11. Bara de instrumente *ALR_Word_AddIn*

Componenta *ALR_Word_AddIn* a fost implementată în limbajul C# folosind mediul de dezvoltare Visual Studio 2022 și este compatibilă cu versiunile 2010 sau mai noi ale editorului de texte Microsoft Word și cu sistemele de operare Windows 7 sau mai noi.

Interfața pentru editare

Pentru editarea de texte în transcriere fonetică trebuie instalată componenta *ALR_Word_AddIn* care adaugă panoul *ALR* în bara de instrumente. Deoarece, în cursul editării, toate interacțiunile utilizatorului cu sistemul sunt interceptate și prelucrate, pentru a nu face greoaie editarea textelor normale, este necesară activarea interfeței, folosind prima comandă de pe bara de instrumente. Aceasta conține 8 panouri pe care sunt grupate comenzile de editare: *Master*, *Formatare*, *Variante simbol*, *Accente*, *Fenomene vocale*, *Fenomene consoane*, *Speciale* și *Diverse*.

Panoul *Master*

În acest panou sunt incluse comenzi generale de editare care sunt descrise mai jos:

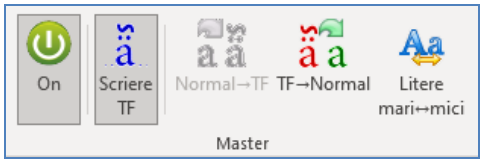

Figura 12. Panoul Master din bara de instrumente *ALR_Word_AddIn*

On	Buton On/Off care permite activarea/ dezactivarea completă a componentei pentru editarea transcrierilor fonetice.
Scriere TF	Activează/ dezactivează scrierea în transcriere fonetică. Dacă este dezactivat, editarea se face folosind fonturile normale.
Normal ⇨ TF	Conversia textului curent selectat din text normal în text de tip transcriere fonetică.
TF ⇨ Normal	Conversia textului curent selectat din text de tip transcriere fonetică în text normal.
Litere mari ⇔ mici	Conversia textului selectat din litere mari în litere mici și invers.

Comenzile de conversie menționate mai sus funcționează doar în cazul în care întreaga selecție poate fi supusă conversiei. Cu alte cuvinte, conversia din text normal în transcriere fonetică funcționează doar dacă toate caracterele din textul selectat sunt de tip text normal.

Panoul *Formatare*

În acest panou sunt incluse comenzi generale de editare. Iconițele utilizate sunt sugestive deoarece sunt asemănătoare celor asociate comenzilor uzuale din Word. Primul grup este aplicabil și secvențelor de mai multe caractere selectate și conține comenzile de formatare la nivel de caracter: Bold, Italic, Underline1şi Undeline2. Al doilea grup de comenzi este specific formatării la nivel de paragraf: aliniere stânga, centrat, aliniere dreapta, justify. Al treilea grup de comenzi este utilizat pentru modificarea poziţiei pe verticală a unui singur caracter: poziționare „în umăr”, respectiv suprapunere.

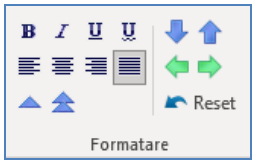

Figura 13. Panoul Formatare din bara de instrumente *ALR_Word_AddIn*

Comenzile din partea dreaptă sunt utilizate pentru deplasarea mai fină a simbolurilor poziţionate „în umăr” sau suprapuse, permiţând deplasarea acestora la nivel de pixel în cele 4 direcţii. Comenzile au fost utilizate în faza de implementare şi testare dar ele pot fi utile şi în utilizarea curentă. Ultima comandă permite resetarea poziţiei caracterelor.

Panoul *Variante simbol*

Variante simbol este un panou dinamic în care sunt afişate variantele posibile ale simbolului curent. În faza de editare trebuie folosită tastatura internațională, nefiind necesară introducerea diacriticelor specifice limbii române. Dacă este activă scrierea în transcriere fonetică, apăsarea unei taste este interceptată și interpretată, iar în panoul *Variante simbol* sunt afișate toate variantele simbolului corespunzător tastei apăsate. Selecția uneia dintre aceste variante produce înlocuirea simbolului curent cu cel corespunzător variantei selectate. Conținutul acestui panou este modificat dinamic, în funcție de caracterul curent selectat.

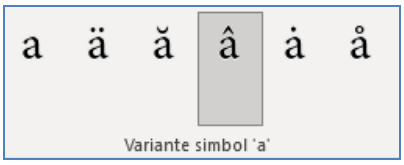

Figura 14. Panoul Variante simbol din bara de instrumente *ALR_Word_AddIn*

Panourile *Accente*, *Fenomene fonetice* şi *Speciale*

În aceste panouri sunt disponibile comenzile pentru aplicarea accentelor (grav, ascuţit şi dublu ascuţit) şi a fenomenelor (13 fenomene fonetice pentru

vocale şi respectiv 9 fenomene ce pot fi aplicate consoanelor). De asemenea, sunt disponibile comenzi pentru introducerea de caractere speciale (litere din alfabetul grec, legato etc.). Grupul de comenzi pentru aplicarea de fenomene fonetice este similar ca funcționalitate cu cel utilizat în variantele anterioare ale editoarelor de texte dialectale. Și în acest caz, selecția unui fenomen sau accent produce înlocuirea caracterului curent selectat cu un nou caracter ce include fenomenul respectiv în glyph. Este evident că, fiind vorba de un număr diferit de fenomene aplicate, este schimbat și fontul selectat pentru caracterul respectiv.

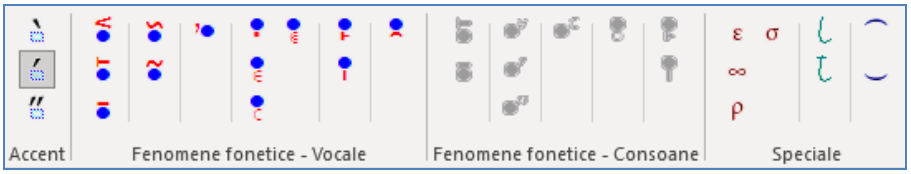

Figura 15. Panourile pentru aplicarea fenomenelor fonetice din bara de instrumente *ALR_Word_AddIn*

Panoul *Diverse*

În panoul *Diverse* (Fig. 16) sunt incluse două comenzi: *Settings* (care permite modificarea setărilor curente ale componentei) şi *Despre Alr* (care produce afişarea ferestrei de dialog cu informaţii despre aceasta).

Figura 16. Panoul Diverse din bara de instrumente *ALR_Word_AddIn*

Activarea componentei *ALR_Word_Addin*

Componenta pentru editarea transcrierilor fonetice este gratuită pentru utilizarea în activități de cercetare și învățământ; totuși, pentru a fi utilizată fără nicio restricție, este necesară activarea acesteia. Activarea este realizată cu ajutorul unei chei de activare obținute după ce sunt transmise informațiile de contact ale utilizatorului (nume, afiliere, adresă de email). Aceste informații sunt necesare pentru a avea o evaluare a numărului de utilizatori. dar și pentru transmiterea către aceștia de informații importante sau pentru trimiterea noilor versiuni ale componentei. În cazul utilizării componentei fără activare, funcțiile principale sunt disponibile, cu observația că din când în când este afișată o fereastră de avertisment în care este cerută activarea.

În figurile 17 și 18 sunt prezentate fereastra de editare a textelor dialectale în Word, respectiv un text complet editat.

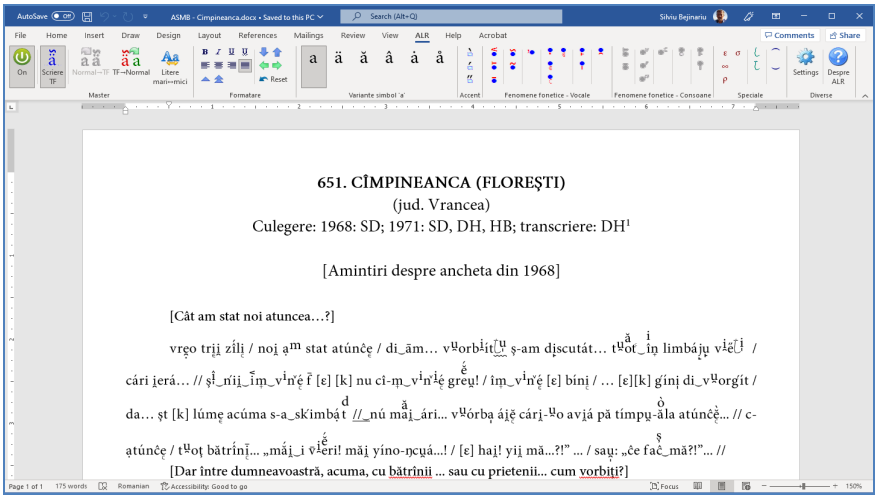

Figura 17. Fereastra principală *Microsoft Word cu ALR_Word_AddIn* activat în bara de instrumente

3. Concluzii

Componenta *ALR_Word_AddIn* reprezintă rezultatul colaborării dintre cercetătorii de la Institutul de Informatică Teoretică, Institutul de Lingvistică „Iorgu Iordan – Alexandru Rosetti” și Institutul de Filologie Română „Alexandru Philippide”. Componenta este funcțională și este utilizată de lingviști pentru editarea textelor dialectale. Prin colaborarea interdisciplinară între colective pot fi identificate posibilități de îmbunătățire și adăugare de noi facilități sistemului de editare propus.

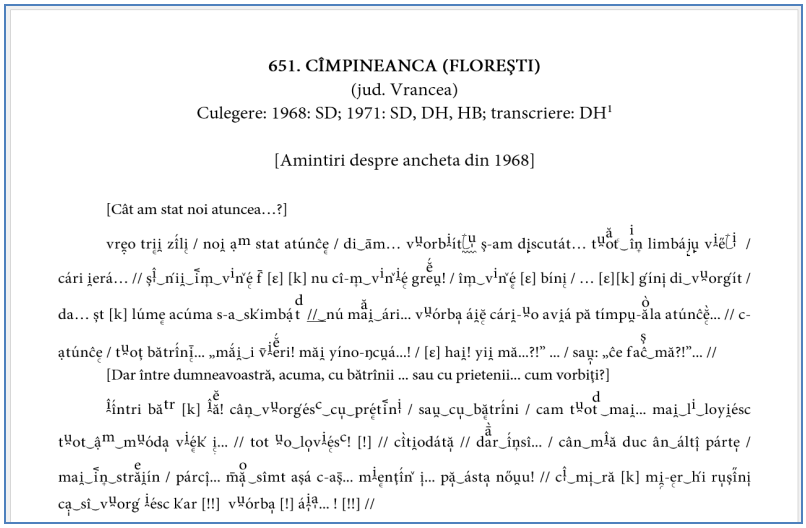

Figura 18. Exemplu de text dialectal editat în *Microsoft Word*

BIBLIOGRAFIE

ALAR II = Nicolae Saramandu, Manuela Nevaci, *Atlasul lingvistic al dialectului aromân*, vol. II, Bucureşti, Editura Academiei Române, 2020.

Bejinariu *et al.* 2009 = Silviu-Ioan Bejinariu, Vasile Apopei, Stelian Dumistrăcel, Horia-Nicolai Teodorescu, *Overview of the Integrated system for dialectal text editing and Romanian Linguistic Atlas publishing – 2009*, în *Proceedings of the 13-th International Conference Inventica 2009*, Iaşi, Editura Performantica, p. 564 – 572.

Bejinariu *et al.* 2021a = Silviu-Ioan Bejinariu, Vasile Apopei, Florin Iftene, *Geolingvistica românească în era digitală*, în M. Nevaci, I. Floarea, I.-M. Farcaș (editori), ***Ex Oriente lux***. *In honorem Nicolae Saramandu*, Alessandria, Edizioni dell'Orso («La colonna infinita» 14), p. 163–176.

Bejinariu *et al.* 2021b = S.-I. Bejinariu, F. Iftene, M. Nevaci, C.I. Floarea, *Preservation of Romanian Linguistic Heritage. Framework for Dialectal Data Management*, în *Proceedings of 16th Edition of The International Conference on Linguistic Resources and Tools for Natural Language Processing – ConsILR-2021*, 13–14 December 2021, Iași.

Dumistrăcel 1976 = Stelian Dumistrăcel, ***Noul Atlas lingvistic român, pe regiuni. Moldova și Bucovina.*** *Probleme ale elaborării*, în „Limba română", anul XXV, nr. 5, p. 547–558.

FontForge 2022 = FontForge Documentation, https://fontforge.org/docs/index.html, ultima accesare la 25.08.2022.

NALR–Mold. Bucov. V = *Noul Atlas lingvistic român, pe regiuni. Moldova şi Bucovina*, material cules de Vasile Arvinte, Stelian Dumistrăcel, Ion Florea, Ion Nuță, Adrian Turculeț și redactat de Stelian Dumistrăcel, Luminița Botoșineanu, Florin-Teodor Olariu, Veronica Olariu, Alexandru-Laurențiu Cohal, Iași, Editura Universității „Alexandru Ioan Cuza", 2022.

Olariu, Olariu 2010 = Florin-Teodor Olariu, Veronica Olariu, *O sută de ani de cartografie lingvistică românească – un bilanţ deschis*, în „Philologica Jassyensia", anul VI, nr. 1 (11), p. 89–118.

Unicode 2011 = Unicode Consortium, *The Unicode Standard Version 6.0 – Core Specification, Ch. 16: Special Areas and Format Characters*, https://www.unicode.org/versions/ Unicode6.0.0/ ch16.pdf, ultima accesare la 25.08.2022.

EDITING DIALECTAL TEXTS USING THE PHONETIC TRANSCRIPTION SPECIFIC TO THE ROMANIAN LANGUAGE

ABSTRACT

This work is a tribute dedicated to Professor Stelian-Traian Dumistrăcel and at the same time a result of the collaboration with the team he was part of. It deals with the most important problems encountered and the solutions adopted during an interdisciplinary cooperation initiated more than 20 years ago between the Institute of Computer Science and the "Alexandru Philippide" Institute of

Romanian Philology of the Romanian Academy Iasi Branch. If we refer only to the field of linguistic geography, the aforementioned collaboration materialized through the design and implementation of the information system that allowed the publication of three volumes of the *New Romanian Linguistic Atlas by Regions – Moldavia and Bukovina*. One of the most difficult and challenging problems was that of editing the phonetic transcriptions specific to the Romanian language, for which a system of real-time graphic synthesis of the image of graphemes needed to be designed. In this paper we propose an innovative system for editing phonetic transcriptions, a system that can be used by any common text editor. An application was implemented for the automatic generation of graphemes for all the 112 791 combinations of symbols and phonetic phenomena in the private code area of TrueType fonts, and also an AddIn that allows their use for phonetic writing in the Microsoft Word editor.

Keywords: *Linguistic atlas*, *Dialectal text*, *Phonetic transcription*, *Text editor*, *TrueType font*.