METODOLOGIE PENTRU CONSTITUIREA ŞI ANALIZA UNUI CORPUS ADNOTAT DE SEMNALE VOCALE – CAZUL *SRoL*

HORIA-NICOLAI TEODORESCU[1,2]

[1]*Academia Română şi*

[2]*Universitatea Tehnica „Gheorghe Asachi"din Iaşi, Iaşi – România*

hteodor@etti.tuiasi.ro

Rezumat

Descriem o metodologie de constituire şi analiză a unui corpus reprezentativ pentru limba română vorbită. Lucrarea are un caracter în esenţă programatic şi metodologic, dar are şi scopul de a atrage atenţia că un mare număr de corpusuri folosite anterior sau în prezent în inferenţa lingvistică sunt deficitare metodologic după cerinţele actuale şi pot conduce la rezultate invalide.

1. Introducere

Complexitatea şi amploarea proceselor limbilor vorbite, procese influenţate de un foarte mare număr de factori, face ca analizele fonetice – fie ele de natură articulator fonetică, acustic-fonetică, perceptiv fonetică, sau fonologică – să fie adesea bazate pe seturi de date a căror reprezentativitate este incertă sau insuficient caracterizată. Faţă de uzanţele din domeniul foneticii şi dialectologiei „clasice", din secolele trecute, standardele ştiinţifice actuale impun câteva condiţii obligatorii: (i) reprezentativitate statistică; (ii) reproductibilitate; (iii) satisfacerea condiţiilor pentru analiza varianţei. Conform acestor condiţii, foarte puţine dintre studiile din secolele trecute, care la momentul realizării au fost studii de mare profunzime – rămân azi valide ştiintific dincolo de nivelul de observaţii semi-empirice. Ele îşi pierd valoarea de stabilire de fapte fonetice, păstrându-şi doar valoare orientativă, estimativă şi istorică, iar aceasta nu doar din motive de limitări tehnologice (de lipsa de înregistrări de calitate), ci mai ales din motive metodologice în (dez)acord cu cerinţele actuale. În această lucrare, pornim de la condiţiile precizate mai sus şi derivăm un set de imperative metodologice pentru constituirea unui corpus vocal adnotat, cerinţe pe care le-am aplicat la constituirea corpusului SRoL[1]. Condiţiile le încadrăm în două categorii: cele care privesc constituirea primară a corpusului şi validarea lui (Secţiunea 2 a lucrării) şi cele care privesc analiza faptelor lingvistice (Secţiunea 3).

2. Metodologie de constituire a corpusului

2.1. Criterii fundamentale

Criteriul reprezentativităţii statistice impune: (1). Pentru un model de limbă vorbită, un număr de vorbitori suficient de mare, suficient de reprezentativ ca proporţii

[1] Corpusul „Sunetele Limbii Române –SRoL" a fost realizat la iniţiativa noastră în perioada 2004-2010 de un grup care include următoarele persoane (în ordine alfabetică): Monica Feraru, Mihaela Hnatiuc, Raluca Ganea, Ramona Luca, I. Păvăloi, Laura Pistol, H.N. Teodorescu, Diana Trandabăţ, Alina Untu, A. Verbuţă, Oana Voroneanu, M. Zbancioc (şi D. Scheianu, Univ. Piteşti), cu cooperarea unui număr de 53 de persoane care au contribuit cu înregistrări.

(procentaje) pentru populaţia care vorbeşte limba respectivă (unde, prin limbă, înţelegem aici fie întreaga limbă, fie un dialect, fie doar un fenomen lingvistic din limba respectivă), conform criteriilor cunoscute în analiza statistică. (2). Pentru fiecare vorbitor, un număr suficient de repetări ale unei pronunţii, pentru stabilirea reprezentativităţii intra-vorbitor şi a eliminării unor eventuale pronunţii accidental deficitare. (3). Un număr de cuvinte reprezentativ pentru limbă, în sensul că setul include o proporţie semnificativă de silabe frecvente ale limbii, astfel încât să fie satisfăcute condiţiile pentru analiza varianţei la nivelul influenţei contextului asupra pronunţiei vocalelor şi consoanelor. (4). Un set de pronunţii care să reflecte o proporţie neneglijabilă a proceselor limbii, printre altele tonalitatea, încărcătura emoţională şi prozodia generală, chiar dacă fenomenul fonetic analizat este punctual, astfel încât să poată fi determinate influenţele acelor procese. (5). O caracterizare în detaliu a vorbitorilor, astfel încât să se respecte condiţia de verificabilitate (reproductibilitate a analizei) şi să se poată ulterior face analize de varianţă (de determinare a influenţelor diverşilor factori asupra limbii vorbite – v. mai jos).

Criteriul reproductibilităţii impune precizarea cu acurateţe şi complet a metodologiei de culegere de date, a parametrilor instrumentarului folosit, a etapelor de prelucrare primară a înregistrărilor, a algoritmilor folosiţi în prelucrare, precum şi precizarea oricăror alte informaţii necesare pentru reproducerea ulterioară a analizei sau constituirii unui corpus echivalent. De asemenea, criteriul reproductibilităţii impune „publicarea" (facerea publică) a tuturor datelor, începând cu înregistrările şi datele subiecţilor şi terminând cu protocoalele utilizate şi instrumentele proprii dezvoltate pentru prelucrarea şi analiza datelor. Un aspect esenţial al reproductibilităţii îl constituie precizarea tipului de vorbire, conform unui număr de criterii precum: voce educată (cultă) / needucată, voce profesională / neprofesională, monolog / convorbire, voce controlată / necontrolată, caracteristici naturale sau simulate (de ex., încărcătura emotivă), contextul socio-profesional al vorbirii (de ex., conform tipologiei de texte şi comunicări precizată – ne-exhaustiv – de (Turculeţ, 2002, p. 76): „monolog, dialog, povestire, interviu, ştiri, expunere, dispută; discuţie în familie, discuţie amicală, discuţie particulară, discuţie publică particulară (în pauze în instituţie, la piaţă, pe stradă etc. – loc?), discuţie oficială, discurs public, pledoarie juridică" etc. (citare prelucrată). Este indubitabil că, până la crearea de instrumente de adnotare automate, cu performanţe mult mai mari decât cele din prezent, realizarea unui corpus care să cuprindă toate aceste tipuri de voci, cu adnotări corespunzătoare, este foarte dificilă. Ca urmare, este important să fie prezentate clar în corpusuri constrângerile de realizare (deci, tipurile de înregistrări, între altele în raport cu clasele de mai sus, cu toate detaliile posibile). Numai în acest fel se va putea realiza, de ex., un studiu ipotetic dar util, precum „O analiză ‚încrucişată' asupra modificării triunghiului vocalelor în silabele accentuate în discursurile publice ştiintifice faţă de pledoariile juridice şi asupra diferenţelor ce apar în funcţie de sexul vorbitorului în variaţiile dintre discurs ştiintific şi pledoarie".

Criteriul completitudinii (de satisfacere a condiţiilor pentru analiza varianţei) impune cunoaşterea în primul rând a tuturor factorilor despre vorbitor – factori familiari – precum limba maternă a mamei, locul unde a copilărit – educaţionali (locul unde a urmat şcoala primară, nivelul maxim de educaţie atins, profesia – care, ultima, influenţează familiaritatea cu vocabularul – gradul de educare / cultură a limbii folosite), factori sociali, medicali (foarte puţine corpusuri dau informaţii asupra

factorilor medicali, care pot influenţa pronunţia, dar şi folosirea mai largă a limbii, de exemplu ambitusul vocal sau expresivitatea emotivă) etc. În al doilea rând, analiza varianţei presupune ca în corpus să se regăsească toţi factorii importanţi de variabilitate a pronunţiei, pentru a se putea face o analiză a cauzelor ce produc o anumită pronunţie[2] şi a distinge între diversele influenţe. Deci, orice corpus trebuie să includă informaţii complete despre trei categorii: (A) despre vorbitori; (B) despre contextul şi tipul vorbirii (monolog, privat, înregistrare de laborator etc.), precum şi data realizării înregistrării; (C) despre (C1) tehnica înregistrărilor, (C2) criteriile de validare şi acceptare a înregistrărilor; (C4) modul de pre-prelucrare a semnalelor, (C5) modul de segmentare şi (C6) adnotare, (C7) modul de extragere a caracteristicilor şi (C8) de validare şi eliminare a valorilor „anormale" ale caracteristicilor; (C9) modul de prelucrare statistică a datelor, inclusiv instrumentele utilizate[3]; (D) Analiza lingvistică a cuvintelor, propoziţiilor, frazelor înregistrate. Pentru fiecare factor implicat, sunt necesare minim 5, preferabil 10 înregistrări pentru fiecare sex (fiecare cu minim trei repetări).

Criteriul eticii ştiintifice impune păstrarea securizată a datelor personale ale vorbitorilor, informarea şi acordul scris al vorbitorilor pentru a participa la teste şi a face publice înregistrările, precum şi validarea de către un for competent a reflectării cerinţelor eticii în producerea şi analiza corpusului.

Din câte ştim, SRoL este până în prezent singurul corpus care satisface toate aceste criterii pentru limba română – şi unul dintre puţinele pe plan internaţional.

2.2. Vorbitorii şi fişa vorbitorului

Pentru fiecare vorbitor, s-a întocmit o fişă a vorbitorului, fişă care conţine toate datele necesare pentru caracterizarea socio-educaţională şi medicală a vorbitorului. Aceste fişe sunt publice, ca şi restul corpusului de voce, pe situl „Sunetele limbii române", (SRoL). Ca urmare, oricare rezultate publicate de grupul nostru asupra limbii române vorbite, pe baza prelucrării materialului din cadrul SRoL, pot fi verificate de oricine şi deci validate sau invalidate de către alte echipe de cercetare, conform cerinţelor metodice actuale – spre deosebire de marea majoritate a lucrărilor curent publicate, pentru care datele sunt neverificabile (nepublice). În plus, SRoL poate fi complementat de alte grupuri de cercetare cu propriile date şi utilizat în conjuncţie cu alte corpusuri de voce pentru cercetări mai ample.

Fişa include date despre vârstă, sex, zonă geografică în care a copilărit şi s-a format ca vorbitor, zona în care a făcut studiile primare, care se ştie că fixează în mare măsură varianta dialectologică a limbii vorbite, date despre studiile universitare şi locul

[2] Subliniem printr-un exemplu relevanţa acestui deziderat. Să presupunem că se doreşte analiza diferenţelor dintre pronunţiile, pentru o vocală dată, în hiatus, ca vocală glisată şi respectiv ca diftong. Dacă numărul de vorbitori din corpus este redus, de exemplu 6, iar ei au un fundal educaţional-familial foarte diferit (dialectal, social), cu implicaţii în diftongizarea vocalei respective, fundal nedocumentat, în plus dacă printre ei unul are afecţiuni neurologice (nedocumentate) care reduc viteza de reacţie (de ex., viteza impulsului electric pe nervi mai mică decât normal), cu implicaţii în producerea glisandoului vocalic, rezultatele statistice vor indica, indiferent de faptul lingvistic, real sau nu, al diftongizării glisantelor în limba dată, o „tendinţă de diftongizare". Într-o asemenea analiză ipotetică, criteriul de satisfacere a condiţiilor pentru analiza varianţei nu este satisfăcut, iar cauzele personale ale tendinţei de diftongizare sunt „văzute" ca tendinţă a limbii.

[3] Atunci când instrumentele nu sunt publice sau comerciale, de exemplu când sunt dezvoltate de autorii corpusului, instrumentele precum şi algoritmii care stau la baza lor, preferabil şi codul sursă trebuie făcute publice, pentru a îndeplini condiţiile de verificabilitate şi reproductibilitate.

absolvirii, eventualele studii post-universitare, obiceiuri de viaţă care pot influenţa vocea (de ex., fumatul), date biometrice şi patologii cunoscute. Detalii au fost prezentate mai pe larg în alte lucrări. Unul singur dintre vorbitori are patologii cunoscute care să îi afecteze sistemul fonator, respirator, auditiv, sau nervos. Mulţi vorbitori au vocea „educată", fie prin profesia didactică, fie prin obişnuinţa unor prezentări publice.

Studiul s-a făcut cu respectarea în totalitate a păstrării privaţiunii subiecţilor; numele subiecţilor este cunoscut doar realizatorilor sitului şi nu este menţionat în fişa publică a vorbitorului, unde identificarea se face printr-un cod numeric. Toţi vorbitorii au fost informaţi asupra obiectivelor generale ale cercetării şi condiţiilor de difuzare a datelor primare (înregistrările vocale); toţi vorbitorii şi-au dat consimţământul scris pentru înregistrări. Cercetarea a fost avizată sub raport etic de Consiliul Facultăţii ETTI, Universitatea Tehnică „Gheorghe Asachi" din Iaşi.

2.3. Statistica regională şi socio-educaţională a vorbitorilor

Setul de 53 de vorbitori include cca. 65% bărbaţi şi cca 35% vorbitori feminini. Grupa de vârstă reprezentată este majoritar 20-35 de ani (peste 50%) şi doar câţiva vorbitori sunt în grupa de vârstă 45-60 ani. Cu rare excepţii, toţi vorbitorii sunt educaţi şi au copilărit în zona de NE a Moldovei, din Bucureşti, în judeţele Iaşi, Vaslui, Bacău, Botoşani, Suceava. Trei vorbitori sunt din Transilvania, unul din zona Argeş, doi din Vâlcea, doi din Muntenia - Bucureşti, iar unul din Maramureş. Ca urmare, SRoL permite comparaţii între vorbitorii (de limba cultă) din NE Moldovei cu cei din alte regiuni ale României. Profilul socio-educaţional al vorbitorilor este, pentru marea lor majoritate, acela al tânărului născut în anii 1980-1990 şi educat până la nivel de facultate, mulţi având un masterat în informatică, lingvistică, sau în domenii conexe.

2.4. Protocolul de înregistrare

Înregistrările au fost efectuate de mai mulţi membri ai echipei SRoL, folosind programul GoldWave™ 5.0, cu următorii parametri: frecvenţă de eşantionare de 22050 Hz; rezoluţie de 16 şi 24 biţi (utilitarul Praat™ prelucrează numai fişierele pe 16 biţi), monofonie. Programul GoldWave este un program comercial, suficient de performant la nivelul anului scrierii lucrării, care asigură o înregistrare de calitate, cu parametrii doriţi, precum şi o prelucrare primară a datelor la nivel acustic.

2.5. Metodologia de validare a fişierelor de semnal vocal şi de segmentare precisă

Toate fişierele au fost analizate manual, prin ascultare, de echipa SRoL. Cu această ocazie, au fost făcute prelucrări primare, anume au fost eliminate secţiunile zgomotoase sau pauzele prea mari; eventual, întreaga înregistrare a fost rejectată dacă nu satisfăcea cerinţele pentru o înregistrare de calitate, de ex. dacă prezintă „limitări" (trunchieri în amplitudine) de semnal, sau dacă avea zgomot intens, uşor perceptibil auditiv.

Două metode de *segmentare* sunt implicit sau explicit utilizate în SRoL. Prima metodă, numită aici *metoda perceptivă,* constă în ascultarea de către o persoană familiarizată atât cu fonetica generală a limbii respective, cât şi cu instrumentul informatic utilizat. Metoda poate fi considerată a fi o tehnică specifică foneticii perceptive, cu limitele ei – în principal percepţia dependentă de contextul lingvistic în care ascultătorul evaluator s-

a dezvoltat (limba natală, educaţia, cultura lingvistică). A doua metodă[4], dezvoltată semi-empiric de autor de-a lungul anilor este subordonată în egală măsură foneticii acustice şi teoriei semnalelor, deşi este aplicată manual şi subiectiv.

Segmentarea fişierelor de voce a fost realizată manual, spre deosebire de alte studii similare (vezi de ex. (Gendrot & Adda-Decker, 2005)), pentru a se asigura o acurateţe şi un grad de certitudine cât mai mare şi o compatibilitate mare între criteriile perceptive şi cele utilizate în abordarea acustic-instrumentalistă. Segmentarea „de precizie" a fost realizată prin ascultare, conform limitelor percepute între foneme, cu teste repetate pentru a determina cât mai precis graniţele detectate auditiv între fonemele adiacente. Atunci când ascultarea nu permitea o precizare suficient de netă în timp a graniţelor, s-a analizat suplimentar forma de undă şi s-a decis plasarea graniţei acolo unde, vizual, apărea clar trecerea de la o formă de undă a unui sunet la forma de undă a următorului. Atunci când a fost necesar, în special pentru pauze şi zone de tranziţie vocală-consoană, o decizie s-a luat în grup pentru validarea segmentării.

2.6. Metodologia de adnotare

Adnotarea a fost realizată manual, simultan cu etapa de segmentare. Nivelurile de adnotare au fost cele corespunzătoare segmentelor de tip propoziţie – cuvânt – silabă – fonem – segment central de fonem. Au fost distinse la acelaşi nivel cu fonemele trei tipuri de pauze: pauze între propoziţii, pauze între cuvinte, pauze în interiorul cuvântului[5]. Frazele au fost adnotate de echipa SRoL folosind utilitarul Praat™, versiunea 5.1.30 (Boersma, 2002).

La nivel de fonem, segmentele au durate larg variabile, cele mai mici de ordinul câtorva milisecunde (ms), pentru pauze scurte şi pentru plozive, respectiv până la ordinul zecilor de ms, pentru vocale şi fricative prelungite, accentuate. În starea emoţională „tristeţe", duratele monoftongilor sunt sensibil mai mari (cu până la 50%). Variabilitatea duratelor, între foneme şi la acelaşi fonem funcţie de stare, de locul în cuvânt (silaba accentuată fiind tipic mai lungă) etc. face ca „populaţia de ferestre" de analiză să aibă dispersii mari, ca număr de ferestre per fonem.

2.7. Descriere generală a conţinutului corpusului

Sumar, corpusul conţine înregistrări de laborator, controlate, cu zgomot redus, de voci în general culte dar nu cultivate, fără voci profesionale din categoriile artişti, reporteri, avocaţi, dar cu câteva voci de persoane din învăţământ, cu vorbitori fără patologie cronică sau acută, cu vârsta dominantă 20-35 de ani. Propoziţiile sunt pronunţate cu ton neutru sau cu ton emotiv, iar emoţiile au fost simulate şi auto-stimulate. Corpusul conţine vocale susţinute, izolat pronunţate, diftongi, triftongi, hiatusuri, consoane, cuvinte izolate şi propoziţii scurte sau de lungime medie. Specific corpusului este

[4] Metoda constă în urmărirea vizuală a formei de undă şi separarea segmentelor vocalice pe baza apariţiei unei deosebiri substanţiale în formele de undă şi în spectrul acelor segmente, respectiv pe menţinerea în cadrul aceluiaşi segment a zonelor în care forma de undă şi spectrul îşi păstrează un grad mare de similitudine cu zonele anterioare. Această metodă, aplicată vizual de către expertul care realizează segmentarea, a fost dezvoltată într-o metodă obiectivă (instrument informatic) (Teodorescu, 2010 b), folosind patternuri (seturi de trăsături acustice) şi distanţe definite pe spaţiul acestor patternuri.

[5] pauzele intra-cuvânt, nu neapărat între silabe, notate $, pauzele inter-cuvinte: blanc; pauzele de scurtă durată, determinate pe semnal, dar care nu se percep, notate %.

includerea pe de o parte a unor structuri gramaticale particulare, precum apoziția și subiectul dublu, iar pe de altă parte a propozițiilor ponunțate cu încărcătură emoțională. Numărul de pronunții de propoziții cu încărcătură emoțională specificată este, pentru o subclasă (descrisă mai jos) de propoziții, aproximativ egal cu numărul de pronunții cu ton neutru (fără încărcătură emoțională). În acest fel, statistica realizată este echilibrată între tonul neutru (unic în cazurile studiilor anterioare asupra vocalelor limbii române) și pronunțiile emoționale. Ca urmare, statistica prezentată are o mai mare rigoare în privința variabilității pronunțiilor posibile și deci în privința modelului complet al limbii vorbite și, în același timp, capătă o dimensiune suplimentară, a încărcăturii emoționale.

Corpusul conține un set de propoziții cu încărcătură semantică imprecis delimitată și lăsată la latitudinea interpretării vorbitorului. De exemplu, una dintre propoziții poate fi interpretată ca interogativă (în scriere marcată prin ?), interogativ-exclamativă (?!), sau ca rămasă în suspensie (...). Alte propoziții suferă interpretări de tipul simplu afirmativ (.), exclamativ (!), sau pot fi intrepetate în suspensie (...). În acest fel, prin alegerea cu grijă a propozițiilor, s-a asigurat nu doar libertate de intrepretare vorbitorului – și implicit un grad ridicat de naturalețe și expresivitate – dar s-a asigurat și o ponderare implicită rezonabilă între diversele prozodii și încărcături emoționale, aspect original și specific corpusului nostru. Modelul de limbă este mai realist astfel, chiar dacă analiza nu include decât un timp total relativ mic de vorbire.

Studiile viitoare vor putea corecta și crește precizia modelului de limbă vorbită, în principal prin analiza unor înregistrări de durată totală mare, echilibrate conform principiilor formulate în (Teodorescu, 2010a), anume: echilibru între pronunții cu tonalități diferite (neutru afirmativ, neutru exclamativ, neutru interogativ, neutru în suspensie) și cu activări (emoționale) diferite, pentru variate emoții. Acele studii viitoare vor trebui să implice zeci de ore de înregistrări pentru a atinge un grad mai mare de semnificație statistică față de studiul actual[6].

2.8. Tipuri de contexte pentru vocale

Pentru a satisface criteriul completitudinii (de satisfacere a condițiilor pentru analiza variației) este necesar ca în corpus să se afle, pentru fiecare proces studiat și pentru limbă în ansamblu, un număr suficient de cazuri care reflectă condiții diferite. De exemplu, printre condițiile care influențează pronunția vocalelor se numără contextul în care apar vocalele (avem în vedere aici, ca exemplu, determinarea unei statistici pentru triunghiul vocalelor în limbă). Este cunoscut că formantul zero (fundamentala) unei vocale este influențat semnificativ de tonalitatea propoziției, de emoția exprimată, de nivelul de accentuare al cuvântului, de contextul cuvântului și de locul unde se află plasată vocala în cuvânt, influențe reflectate de linia prozodică specifică tonalității și imprimată propoziției ca atare, prin urmare și vocalei. În același timp, este cunoscută influența „contextului imediat", constituit de fonemul (monoftongul) anterior sau ulterior vocalei, de existența unui diftong sau de glisare, asupra formanților, în special asupra lui F_1 și F_2. Pentru fiecare vocală, numărul de ocurențe în diversele contexte (în

[6] Până la apariția unor instrumente a căror precizie de segmentare și adnotare automată (și interpretare) să fie echivalentă cu cea a experților umani (uneori constituiți în echipă, ca în cazul unor analize de finețe făcute de noi), asemenea studii sunt greu de realizat. Într-adevăr, activitatea la corpusul SRoL până în prezent este echivalentă cu peste 2-3 (ani × om), iar un studiu comparabil pentru zeci de ore de înregistrări ar presupune zeci de (ani × om) de activitate.

care apare vocala) semnificative statistic pentru limbă trebuie să fie suficient de mare pentru o inferenţă statistică. În prezent, corpusul acoperă cca. 35% din limba română, la nivelul silabic[7] (probabilitatea de regăsire a unei silabe din limba română în corpus este 35%, altfel spus, suma probabilităţilor silabelor din corpus în l.r. este aproape de 0.35).

Un aspect important al contextului vocalelor este structura prozodică în care sunt pronunţate. În acest sens, distingem contexte prozodice din care vocala face parte, de tip DU (creştere F_0), F_0 constant, respectiv UD (descreştere F_0), eventual primul şi ultimul context cu atributul suplimentar „creştere / descreştere rapidă". Reprezentativitatea SRoL sub acest aspect nu a fost încă determinată.

Statisticile efectuate „încrucişat", „în diagonală" (pentru selecţii de contexte în care apar vocalele şi pentru grupuri de vorbitori) se realizează în aplicaţiile dezvoltate de noi şi de colectivul SRoL prin selectarea acelor instanţieri de vocale care corespund criteriului respectiv. Asupra lor revenim în altă lucrare transmisă la CONSILR2010.

2.9. Tipuri de propoziţii, intonaţii şi încărcături emoţionale

Statistica a fost realizată pe un set de propoziţii prefixate, selectate, după cum s-a mai spus, pentru a satisface riguros criteriul de „compatibilitate multi-emoţională" (Teodorescu 2002 a). Anume, criteriul impune compatibilitatea propoziţiei respective cu interpretări emoţionale pentru toate cele trei emoţii selectate drept relevante pentru corpus (bucurie, furie, tristeţe), alături de tonul neutru[8]. Alegerea stărilor emoţionale a fost realizată după teste preliminare (împreună cu dr. M.S. Feraru) asupra unui set de şapte emoţii; selecţia s-a făcut astfel încât separarea între emoţii să fie suficient de netă (ambiguitate redusă la recunoaştere, deci recognoscibilitate rezonabil de uşoară de către ascultători neutri) şi în acelaşi timp, încărcarea emoţională a propoziţiei să fie suficient de facil de realizat pentru vorbitori. Relevanţa pentru corpus a tipurilor de emoţie este determinată de antagonismele bucurie – tristeţe şi bucurie – furie, de caracterul intens al acestor emoţii, precum şi de uşurinţa cu care pot fi exprimate suficient de distinct una de alta. Propoziţiile incluse în această statistică sunt toate de natură afirmativă; faptul că nici una nu include negaţii este probabil limitarea principală a SRoL în prezent.

Gradul de încărcare emoţională pentru fraze a fost determinat cu ajutorul unei aplicaţii *online* realizată de colectiv (Pistol & Teodorescu, 2010), aplicaţie cu ajutorul căreia au fost colectate opiniile asupra recognoscibilităţii emoţiei în pronunţia respectivă de la un număr mare de evaluatori. Conform scării folosite de (Beller, Obin, Rodet, 2008), „gradul de activare" (a emoţiei), pentru pronunţiile de propoziţii din corpusul SRoL este între inactivare şi activare mare, pe o scară cu patru trepte (inactivare, activare mică, medie şi mare). Ambele tipuri de activare a emoţiei, activare extrovertită şi introvertită sunt prezente în pronunţiile de propoziţii din corpus. Toate emoţiile au fost de tip simulat („acted emotions"), dar „trăit", în sensul că li s-a recomandat subiecţilor să se

[7] Probabilităţile silabelor au fost preluate din modelele de limbă determinate pe corpusuri ample în Adriana Vlad et al., Limba română scrisă ca sursă de informaţie, Ed. Paideia, Buc., 2003

[8] Propoziţii şi interpretările posibile (/ / /): Aseară (. / ! / ...), Vine mama (. / ! / ...), Ai venit iar la mine (! / ...), Cine a făcut asta (? / ?! / ... /), Omul meu îl lucră (. / ! / ...), Oricum, îţi poţi câştiga locul dorit (. / ! / ...). Fraze cu subiect dublu sau apoziţie: Vine ea mama!, „A trecut el aşa un răstimp", (M. Sadoveanu), O şti el careva cum să rezolve asta, Mama vine şi ea mai târziu, Mama ştie ea ce face, Chiar ştie el ce face?

gândească la o situaţie reală care ar elicita emoţia respectivă şi deci pronunţia emoţională dorită[9].

3. Metodologia de analiză

3.1. Metodologia de prelucrare şi analiză formantică (acustică)

După analiza preliminară a calităţii înregistrărilor, acestea au fost filtrate, astfel încât să rămână doar spectrul de frecvenţe de interes în analiza formantică, anume între 70 Hz şi 10 kHz. Filtrarea s-a realizat folosind facilitatea utilitarului GoldWave™ de a se preciza de către utilizator banda de filtrare şi forma caracteristicii filtrului.

Pentru prelucrări spectrale (detecţia formanţilor şi a frecvenţei fundamentale), s-a selectat un instrument informatic dintre cele mai puternice la ora actuală şi în acelasi timp gratuit, utilitarul Praat™, conceput şi pus la dispoziţie comunităţii internaţionale de către P. Boersma şi D. Weenink (Boersma, 2002). Analiza minuţioasă a influenţei deplasării ferestrei[10] asupra rezultatelor (Teodorescu, Feraru) a indicat că alegerea este convenabilă şi nu produce erori la variaţii mici ale pasului de deplasare sau ale dimensiunii ferestrei. Utilitarul Praat furnizează valorile frecvenţelor medii ale formanţilor F_0 [Hz], F_1, F_2, F3, F4 [Hz], pentru zonele delimitate la segmentare, precum şi valori instantanee ale formanţilor pentru fiecare „fereastră" de analiză.

3.2. Eliminarea erorilor

Utilitarul Praat™, deşi printre cele mai performante în prezent, produce cca. 10% erori evidente, din numărul total de determinări de valori medii, la valorile primilor doi formanţi şi aproape acelaşi procent de erori la determinările de frecvenţă fundamentală. Deoarece valorile „evident eronate" la care ne referim sunt adesea mult mai mari decât cele credibile, aceste erori afectează semnificativ, chiar la frecvenţe de apariţie de ordinul 5-10%, rezultatele statisticilor. Ca urmare, este necesară eliminarea erorilor grosiere produse de instrumentul de analiză înainte de realizarea statisticii. Dat fiind numărul mare de fişiere, nu se pune problema determinării manuale a valorilor corecte. Soluţia la care am recurs este de a elimina valorile „evident" eronate precum şi pe cele „aberante" (outliers). Statisticile s-au realizat doar pe valorile „valide". Această eliminare a valorilor „anormale" poate conduce la eliminarea, pe lângă valorile eronate şi a unor valori reale, dar anormale, nespecifice populaţiei globale. Aceasta este corect din punctul de vedere al cercetării dacă suntem interesaţi de „procesele medii", de caracterizarea globală şi nu de elemente specifice, rare de pronunţie, care pot să fie mascate de eliminările operate.

În etapa de prelucrare intermediară a datelor, au fost corectate două tipuri de erori, primul produs de instrumentul de analiză Praat™, iar al doilea este posibil a se datora

[9] Gradul de activare al emoţiilor îl determinăm astfel: cuartila superioară (conform evaluării de către ascultatori) este „foarte expresiv", următoarea cuartilă (50-75%) este „expresivă", a treia cuartilă (25-50%) este puţin expre-sivă, iar ultima slab expresivă (ne-expresivă). Facem distincţia între *expresivităţile pentru fiecare emoţie în parte*, deoarece unele persoane pot activa şi exprima uşor o emoţie, de exemplu bucuria, dar nu şi altele, de ex. „triste-ţea". O persoană cu scoruri printre primii 25% dintre vorbitori (scorurile sunt calculate astfel: 4 pct. = f. expresiv, 0 = puţin expresiv; scor personal între 0 şi 16) o considerăm extrovertită, iar între 25%-75%- persoană medie.

[10] Instrumentul Praat a fost setat să lucreze cu ferestre de analiză de 0,025s şi cu suprapuneri ale ferestrei alunecă-toare (suprapuneri între ferestre succesive) de ordinul a 40%, ceea ce corespunde la deplasări de 100 ms ale ferestrei.

instrumentului sau altor factori (înregistrări deficitare, pronunţii deficitare). Aplicaţia informatică dezvoltată permite eliminarea automată a valorilor eronate date de Praat™: (a) nedefinite pentru F_0, acolo unde există sunet vocalic; (b) valori pentru F_0 acolo unde nu este sunet vocalic; (c) valori anormal de mici sau de mari determinate de Praat pentru F_0, F_1, F_2. Anterior efectuării calculelor statistice pentru fiecare formant în parte, din fişierul cu date primare sunt eliminate valorile evident eronate. După această operaţie de curăţare, se determină statistica primară – valoarea medie şi varianţa pentru fiecare formant – şi se elimină valorile „aberante" (outliers), din afara intervalului $\bar{x}-3\sigma, \bar{x}+3\sigma$. În final, pentru valorile rămase, sunt eliminate „cozile" de 3% din populaţie. Statistica formantului fonemului se calculează cu valorile rămase după acest şir de corecţii. Metoda este prezentată pe larg în alte lucrări.

3.3. Metoda de validare a expresiei emoţionale

Validarea expresiei emoţionale a fost realizată în două etape. În prima fază, cinci evaluatori din cadrul colectivului nostru au efectuat o validare a emoţiei reprezentate (exprimate voluntar) în pronunţia respectivă a propoziţiei. Înregistrările care aveau un nivel de expresivitate inacceptabil de scăzut sau o exprimare confuză a emoţiei au fost eliminate în această etapă dintre înregistrări şi înlocuite eventual cu alte înregistrări. În etapa a doua a validării, am recurs la validarea publică, pe Internet. S-a creat o aplicaţie (Pistol & Teodorescu, 2010) cu ajutorul căreia un utilizator oarecare de Internet, vorbitor de limba română, poate accesa şi asculta oricare dintre propoziţii şi preciza dacă acea propoziţie are încărcătură emoţională, dacă emoţia este evidentă (nu produce confuzie, indecizie asupra tipului de emoţie) şi care este emoţia respectivă.

4. Discuţie şi concluzii

Criteriile expuse au fost aplicate la realizarea corpusului SRoL. Acest corpus nu satisface decât parţial criteriul privind completitudinea necesară pentru analiza varianţei, dar, prin aplicarea coerentă a criteriilor până acum şi în continuare la dezvoltarea corpusului, premisele corectitudinii metodologice ale SRoL sunt asigurate. Nesatisfacerea criteriului completitudinii este relativă şi este datorată numărului încă redus de pronunţii pentru unele tipuri de ocurenţe – de exemplu, pentru vocale glisante şi hiatusuri. Dar în cazul SRoL limitarea nu este datorată, ca în cazul altor corpusuri, necunoaşterii (ne-documentării) unor factori care pot influenţa pronunţia – de la factori specifici vorbitorului la modul de înregistrare sau pre-procesare a fişierelor. La aceste din urmă corpusuri, eroarea este fundamentală şi necorijabilă retroactiv.

Încheiem cu un exemplu de aplicaţie deja iniţiată de noi. Să considerăm problema determinării „triunghiului vocalelor" $(F^v_{1(k)}, F^v_{2(k)})$ pentru limba română[11]. La nivelul actual, această determinare presupune: determinarea „norilor" vocalelor, a elipselor de încredere, global pe întreaga populaţie (masculin şi feminin), separat pe sexe, separat pentru vocale izolate, vocale în cuvinte izolate, în propoziţii, influenţa emoţiilor asupra deplasării triunghiului formanţilor (deplasare cunoscută în literatură, pentru alte limbi),

[11] Din nefericire, l. română este una dintre puţinele limbi europene pentru care încă nu există decât studii semi-empirice (Rosetti, Lăzăroiu, 1982), (Teodorescu et al. 1986), invalide la nivelul cerinţelor prezente, privind triunghiul formanţilor. V. şi (Teodorescu, Feraru, Trandabăţ, 2005).

compararea triunghiului pentru vocale susţinute cu cel obţinut pe întreaga limbă (deci, ca medii ale mediilor pe vocale în diversele lor ipostaze), analiza efectului accentului în cuvânt, analiza efectului contextului (CVC, CV_, _VC, V.V, diftongului, glisării vocalei etc.), compararea cu alte limbi din aceeaşi familie, eventual analiza pe dialecte, profesii etc. La acestea se adaugă analiza modificării funcţiilor densitate de probabilitate şi a parametrilor lor (varianţă, asimetrie, aplatizare) pentru cazurile enunţate mai sus. Într-un sens, acesta este şi (parţial) programul echipei SRoL pentru viitorul apropiat.

Mulţumiri. Autorul multumeşte tuturor colegilor care au lucrat la corpusul SRoL pentru nenumărate şi îndelungi discuţii. Activitatea la SRoL a fost parţial sprijinită de către Academia Română.

Referinţe bibliografice

Beller, G., Obin N., & Rodet X. (2008). Articulation Degree as a Prosodic Dimension of Expressive Speech. Speech Prosody 2008, ISCA, Campinas, Brazil, May 6-9, 2008, 681-684.

Boersma, P.P.G. (2002). Praat, a system for doing phonetics by computer. Glot International, Vol. 5 No. 9/10, p. 341-345.

Gendrot, C. & Adda-Decker, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. INTERSPEECH 2005, Sept. 4-8, Lisbon, Portugal, 2453-2455.

Pistol, L. & Teodorescu, H.N. (2010). A Note on Testing the Recognition of Emotional States and Tones in Speech. Memoirs of the Romanian Academy, 2010 (under press).

Rosetti, Al., Lăzăroiu, A. (1982). Introducere în fonetică. Ed. Ştiinţifică & Encicl., Bucureşti, 1982.

SRoL, Voiced Sounds of the Romanian Language Project, (autori: Teodorescu, H.-N., Trandabăţ, D., Feraru, M., Ganea, R., Verbuţă, A., Zbancioc, M., Hnatiuc, M., Voroneanu, O., Pistol, L., Untu, A., Păvăloi, I.), www.etc.tuiasi.ro/sibm/romanian_spoken_language/ro/voci_emoţionale.htm.

Teodorescu, H.-N., Buchholtzer, L., Poşa, C. (1986). Comunicarea orală om-maşină. Cap. 2 – Vocea şi vorbire, şi Cap. 4 – Sinteza vorbirii, Editura Tehnică, Seria „Tehnica la zi", Bucureşti, 1986.

Teodorescu, H.-N., Feraru, M., Trandabăţ, D. (2006). Situl 'Limba Română Vorbită'. Lucrările atelierului Resurse lingvistice şi instrumente pentru prelucrarea limbii române. Editori: Corina Forăscu, Dan Tufiş, Dan Cristea Iaşi, 3 nov. 2006, Editura Universităţii Al.I. Cuza Iaşi, 3-7.

Teodorescu, H.-N. (2010 a). Noisy Speech Files Perceptual Mining for Speech and Noise Patterns, PROMISE, 29 martie 2010, Iaşi.

Teodorescu, H.-N. (2010 b). AI Tools for Speech Analysis Applied to the Romanian Language. (Plenary paper), ECC 2010, 20-22 aprilie 2010, Bucureşti.

Turculeţ, A. (2002). Tipuri de texte orale. In: Klaus Bochmann, Vasile Dumbravă (Eds.), Limba română vorbită în Moldova istorică, Volume 1. Leipziger Universitätsverlag, 2002, pp. 53-78.