

METODOLOGIE PENTRU CONSTITUIREA ȘI ANALIZA UNUI CORPUS ADNOTAT DE SEMNALE VOCALE – CAZUL *SROL*

HORIA-NICOLAI TEODORESCU¹

¹*Academia Română și Universitatea “Gheorghe Asachi” din Iași, Iași – România*

Rezumat

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

1. *Introducere*

Complexitatea și amploarea proceselor limbilor vorbite, procese influențate de un foarte mare număr de factori, face ca analizele fonetice – fie ele de natură articulator fonetică, acustic-fonetice, perceptiv fonetică, sau fonologică – să fie adesea bazate pe seturi de date a căror reprezentativitate este incertă sau insuficient caracterizată. Față de uzanțele din domeniul foneticii și dialectologiei „clasice”, din secolele trecute, standardele științifice actuale impun câteva condiții obligatorii: (i) reprezentativitate statistică; (ii) reproductibilitate; (iii) satisfacerea condițiilor pentru analiza varianței. Conform acestor condiții, foarte puține dintre studiile din secolele trecute, care la momentul realizării au fost studii de mare profunzime – rămân azi valide științific dincolo de nivelul de observații semi-empirice. Ele își pierd valoarea de stabilire de fapte fonetice, păstrându-și doar valoare orientativă, estimativă și 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 cerințele actuale. În această lucrare, pornim de la condițiile precizate mai sus și derivăm un set de imperative metodologice pentru constituirea unui corpus vocal adnotat, cerințe pe care le-am aplicat la constituirea corpusului *SROL*¹. Condițiile le încadrăm în două categorii: cele care privesc constituirea primară a corpusului și validarea lui (Secțiunea 2 a lucrării) și cele care privesc analiza faptelor lingvistice (Secțiunea 3).

2. *Metodologie de constituire a corpusului*

2.1. *Criterii fundamentale*

Criteriul reprezentativității statistice impune: (1). Pentru un model de limbă vorbită, un număr de vorbitori suficient de mare, suficient de reprezentativ ca proporții (procentaje) pentru populația care vorbește limba respectivă (unde, prin limbă, înțelegem

¹ Corpusul “Sunetele Limbii Române –*SROL*” a fost realizat la inițiativa 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 Trandabăt, Alina Untu, A. Verbuță, Oana Voroneanu, M. Zbancioc (și D. Scheianu, Univ. Pitești), cu cooperarea unui număr de 53 de persoane care au contribuit cu înregistrări.

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 pronunții, pentru stabilirea reprezentativității intra-vorbitor și a eliminării unor eventuale pronunții accidental deficitare. (3). Un număr de cuvinte reprezentativ pentru limbă, în sensul că setul include o proporție semnificativă de silabe frecvente ale limbii, astfel încât să fie satisfăcute condițiile pentru analiza varianței la nivelul influenței contextului asupra pronunției vocalelor și consoanelor. (4). Un set de pronunții care să reflecte o proporție neneglijabilă a proceselor limbii, printre altele tonalitatea, încărcătura emoțională și prozodia generală, chiar dacă fenomenul fonetic analizat este punctual, astfel încât să poată fi determinate influențele acelor procese. (5). O caracterizare în detaliu a vorbitorilor, astfel încât să se respecte condiția de verificabilitate (reproductibilitate a analizei) și să se poată ulterior face analize de varianță (de determinare a influențelor diversilor factori asupra limbii vorbite – v. mai jos).

Criteriul reproductibilității impune precizarea cu acuratețe și complet a metodologiei de culegere de date, a parametrilor instrumentarului folosit, a etapelor de prelucrare primară a înregistrărilor, a algoritmilor folosiți în prelucrare, precum și precizarea oricăror alte informații necesare pentru reproducerea ulterioară a analizei sau constituirii unui corpus echivalent. De asemenea, criteriul reproductibilității impune „publicarea” (facerea publică) a tuturor datelor, începând cu înregistrările și datele subiecților și terminând cu protocoalele utilizate și instrumentele proprii dezvoltate pentru prelucrarea și analiza datelor. Un aspect esențial al reproductibilității 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 și comunicări precizată – ne-exhaustiv – de (Turculeț, 2002, p. 76): „monolog, dialog, povestire, interviu, știri, expunere, dispută; discuție în familie, discuție amicală, discuție particulară, discuție publică particulară (în pauze în instituție, la piață, pe stradă etc. – loc?), discuție oficială, discurs public, pledoarie juridică” etc. (citare prelucrată)). Este indubitabil că, până la crearea de instrumente de adnotare automate, cu performanțe 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 „încrucișată” asupra modificării triunghiului vocalelor în silabele accentuate în discursurile publice științifice față de pledoariile juridice și asupra diferențelor ce apar în funcție de sexul vorbitorului în variațiile dintre discurs științific și pledoarie”.

Criteriul completitudinii (de satisfacere a condițiilor pentru analiza varianței) impune cunoașterea în primul rând a tuturor factorilor despre vorbitor – factori familiari – precum limba maternă a mamei, locul unde a copilărit – educaționali (locul unde a urmat școala primară, nivelul maxim de educație atins, profesia – care, ultima, influențează familiaritatea cu vocabularul – gradul de educare / cultură a limbii folosite), factori sociali, medicali (foarte puține corpusuri dau informații asupra factorilor medicali, care pot influența pronunția, dar și folosirea mai largă a limbii, de exemplu ambitusul

vocal sau expresivitatea emotivă) etc. În al doilea rând, analiza varianței presupune ca în corpus să se regăsească toți factorii importanți de variabilitate a pronunției, pentru a se putea face o analiză a cauzelor ce produc o anumită pronunție² și a distinge între diversele influențe. Deci, orice corpus trebuie să includă informații complete despre trei categorii: (A) despre vorbitori; (B) despre contextul și tipul vorbirii (monolog, privat, înregistrare de laborator etc.), precum și data realizării înregistrării; (C) despre (C1) tehnica înregistrărilor, (C2) criteriile de validare și acceptare a înregistrărilor; (C4) modul de pre-prelucrare a semnalelor, (C5) modul de segmentare și (C6) adnotare, (C7) modul de extragere a caracteristicilor și (C8) de validare și eliminare a valorilor „anormale” ale caracteristicilor; (C9) modul de prelucrare statistică a datelor, inclusiv instrumentele utilizate³; (D) Analiza lingvistică a cuvintelor, propozițiilor, frazelor înregistrate. Pentru fiecare factor implicat, sunt necesare minim 5, preferabil 10 înregistrări pentru fiecare sex (fiecare cu minim trei repetări).

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

Din câte știm, SRoL este până în prezent singurul corpus care satisface toate aceste criterii pentru limba română – și unul dintre puținele pe plan internațional.

2.2. *Vorbitorii și fișa vorbitorului*

Pentru fiecare vorbitor, s-a întocmit o fișă a vorbitorului, fișă care conține toate datele necesare pentru caracterizarea socio-educatională și medicală a vorbitorului. Aceste fișe sunt publice, ca și 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 și deci validate sau invalidate de către alte echipe de cercetare, conform cerințelor 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 și utilizat în conjuncție cu alte corpusuri de voce pentru cercetări mai ample.

Fișa include date despre vârstă, sex, zonă geografică în care a copilărit și s-a format ca vorbitor, zona în care a făcut studiile primare, care se știe că fixează în mare măsură

² Subliniem printr-un exemplu relevanța acestui deziderat. Să presupunem că se dorește analiza diferențelor dintre pronunțiile, pentru o vocală dată, în hiatus, ca vocală glisată și respectiv ca diftong. Dacă numărul de vorbitori din corpus este redus, de exemplu 6, iar ei au un fundal educațional-familial foarte diferit (dialectal, social), cu implicații în diftongizarea vocalei respective, fundal nedocumentat, în plus dacă printre ei unul are afecțiuni neurologice (nedocumentate) care reduc viteza de reacție (de ex., viteza impulsului electric pe nervi mai mică decât normal), cu implicații în producerea glisandoului vocalic, rezultatele statistice vor indica, indiferent de faptul lingvistic, real sau nu, al diftongizării glisantelor în limba dată, o „tendință de diftongizare”. Într-o asemenea analiză ipotetică, criteriul de satisfacere a condițiilor pentru analiza varianței nu este satisfăcut, iar cauzele personale ale tendinței de diftongizare sunt „văzute” ca tendință a limbii.

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

variantele dialectologice a limbii vorbite, date despre studiile universitare și locul absolvirii, eventualele studii post-universitare, obiceiuri de viață care pot influența vocea (de ex., fumatul), date biometrice și 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. Mulți vorbitori au vocea „educată”, fie prin profesia didactică, fie prin obișnuința unor prezentări publice.

Studiul s-a făcut cu respectarea în totalitate a păstrării privațiunii subiecților; numele subiecților este cunoscut doar realizatorilor sitului și nu este menționat în fișa publică a vorbitorului, unde identificarea se face printr-un cod numeric. Toți vorbitorii au fost informați asupra obiectivelor generale ale cercetării și condițiilor de difuzare a datelor primare (înregistrările vocale); toți vorbitorii și-au dat consimțământul scris pentru înregistrări. Cercetarea a fost avizată sub raport etic de Consiliul Facultății ETTI, Universitatea Tehnică „Gheorghe Asachi” din Iași.

2.3. Statistica regională și socio-educatională a vorbitorilor

Setul de 53 de vorbitori include cca. 65% bărbați și cca 35% vorbitori feminini. Grupa de vârstă reprezentată este majoritar 20-35 de ani (peste 50%) și doar cațiva vorbitori sunt în grupa de vârstă 45-60 ani. Cu rare excepții, toți vorbitorii sunt educați și au copilărit în zona de NE a Moldovei, din București, în județele Iași, Vaslui, Bacău, Botoșani, Suceava. Trei vorbitori sunt din Transilvania, unul din zona Argeș, doi din Vâlcea, doi din Muntenia - București, iar unul din Maramureș. Ca urmare, SRoL permite comparații între vorbitorii (de limba cultă) din NE Moldovei cu cei din alte regiuni ale României. Profilul socio-educational al vorbitorilor este, pentru marea lor majoritate, acela al tânărului născut în anii 1980-1990 și educat până la nivel de facultate, mulți având un masterat în informatică, lingvistică, sau în domenii conexe.

2.4. Protocolul de înregistrare

Înregistrările au fost efectuate de mai mulți membri ai echipei SRoL, folosind programul GoldWave™ 5.0, cu următorii parametri: frecvență de eșantionare de 22050 Hz; rezoluție de 16 și 24 biți (utilitarul Praat™ prelucrează numai fișierele pe 16 biți), monofonie. Programul GoldWave este un program comercial, suficient de performant la nivelul anului scrierii lucrării, care asigură o înregistrare de calitate, cu parametrii doriți, precum și o prelucrare primară a datelor la nivel acustic.

2.5. Metodologia de validare a fișierelor de semnal vocal și de segmentare precisă

Toate fișierele au fost analizate manual, prin ascultare, de echipa SRoL. Cu această ocazie, au fost făcute prelucrări primare, anume au fost eliminate secțiunile zgomotoase sau pauzele prea mari; eventual, întreaga înregistrare a fost rejectată dacă nu satisfăcea cerințele pentru o înregistrare de calitate, de ex. dacă prezintă „limitări” (trunchieri în amplitudine) de semnal, sau dacă avea zgomot intens, ușor 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 și cu instrumentul informatic utilizat. Metoda poate fi considerată a fi o tehnică specifică foneticii perceptive, cu limitele ei – în principal percepția dependentă de contextul lingvistic în care ascultătorul evaluator s-a dezvoltat (limba natală, educația, cultura lingvistică). A doua metodă⁴, dezvoltată semi-empiric de autor de-a lungul anilor este subordonată în egală măsură foneticii acustice și teoriei semnalelor, deși este aplicată manual și subiectiv.

Segmentarea fișierelor de voce a fost realizată manual, spre deosebire de alte studii similare (vezi de ex. (Gendrot & Adda-Decker, 2005)), pentru a se asigura o acuratețe și un grad de certitudine cât mai mare și o compatibilitate mare între criteriile perceptive și 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 granițele detectate auditiv între fonemele adiacente. Atunci când ascultarea nu permitea o precizare suficient de netă în timp a granițelor, s-a analizat suplimentar forma de undă și s-a decis plasarea graniței 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 și zone de tranziție 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 propoziție – cuvânt – silabă – fonem – segment central de fonem. Au fost distinse la același nivel cu fonemele trei tipuri de pauze: pauze între propoziții, pauze între cuvinte, pauze în interiorul cuvântului⁵. Frazele au fost adnotate de echipa SRoL folosind utilitarul PraatTM, 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 și pentru plozive, respectiv până la ordinul zecilor de ms, pentru vocale și fricative prelungite, accentuate. În starea emoțională „tristete”, duratele monoftongilor sunt sensibil mai mari (cu până la 50%). Variabilitatea duratelor, între foneme și la același fonem funcție de stare, de locul în cuvânt (silaba accentuată fiind tipic mai lungă) etc. face ca „populația de ferestre” de analiză să aibă dispersii mari, ca număr de ferestre per fonem.

2.7. Descriere generală a conținutului corpusului

Sumar, corpusul conține înregistrări de laborator, controlate, cu zgomot redus, de voci în general culte dar nu cultivate, fără voci profesionale din categoriile artiști, reporteri,

⁴ Metoda constă în urmărirea vizuală a formei de undă și separarea segmentelor vocalice pe baza apariției unei deosebiri substanțiale în formele de undă și în spectrul acelor segmente, respectiv pe menținerea în cadrul aceluiași segment a zonelor în care forma de undă și spectrul își 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) și distanțe definite pe spațiul 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 %.

avocați, dar cu câteva voci de persoane din învățământ, cu vorbitori fără patologie cronică sau acută, cu vârsta dominantă 20-35 de ani. Propozițiile sunt pronunțate cu ton neutru sau cu ton emotiv, iar emoțiile au fost simulate și auto-stimulate. Corpusul conține vocale susținute, izolat pronunțate, diftongi, triftongi, hiatusuri, consoane, cuvinte izolate și propoziții scurte sau de lungime medie. Specific corpusului este includerea pe de o parte a unor structuri gramaticale particulare, precum apozitia și subiectul dublu, iar pe de altă parte a propozițiilor pronunțate 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 interpretate în suspensie (...). În acest fel, prin alegerea cu grijă a propozițiilor, s-a asigurat nu doar libertate de interpretare 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 actual⁶.

2.8. Tipuri de contexte pentru vocale

Pentru a satisface criteriul completitudinii (de satisfacere a condițiilor pentru analiza varianței) 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 (fundamentală) unei

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.

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 care apare vocala) semnificative statistic pentru limbă trebuie să fie suficient de mare pentru o inferență statistică. În prezent, corpusul acoperă cca. 35% din limba română, la nivelul silabic⁷ (probabilitatea de regăsire a unei silabe din limba română în corpus este 35%, altfel spus, suma probabilităților silabelor din corpus în l.r. este aproape de 0.35).

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

Statisticile efectuate „încrucișat”, „în diagonală” (pentru selecții de contexte în care apar vocalele și pentru grupuri de vorbitori) se realizează în aplicațiile dezvoltate de noi și de colectivul SRoL prin selectarea acelor instanțieri de vocale care corespund criteriului respectiv. Asupra lor revenim în altă lucrare transmisă la CONSILR2010.

2.9. Tipuri de propoziții, intonații și încărcături emoționale

Statistica a fost realizată pe un set de propoziții prefixate, selectate, după cum s-a mai spus, pentru a satisface riguros criteriul de „compatibilitate multi-emoțională” (Teodorescu 2002 a). Anume, criteriul impune compatibilitatea propoziției respective cu interpretări emoționale pentru toate cele trei emoții selectate drept relevante pentru corpus (bucurie, furie, tristețe), alături de tonul neutru⁸. Alegerea stărilor emoționale a fost realizată după teste preliminare (împreună cu dr. M.S. Feraru) asupra unui set de șapte emoții; selecția s-a făcut astfel încât separarea între emoții să fie suficient de netă (ambiguitate redusă la recunoaștere, deci recognoscibilitate rezonabil de ușoară de către ascultători neutri) și în același timp, încărcarea emoțională a propoziției să fie suficient de facil de realizat pentru vorbitori. Relevanța pentru corpus a tipurilor de emoție este determinată de antagonismele bucurie – tristețe și bucurie – furie, de caracterul intens al acestor emoții, precum și de ușurința cu care pot fi exprimate suficient de distinct una de alta. Propozițiile incluse în această statistică sunt toate de natură afirmativă; faptul că nici una nu include negații este probabil limitarea principală a SRoL în prezent.

Gradul de încărcare emoțională pentru fraze a fost determinat cu ajutorul unei aplicații *online* realizată de colectiv (Pistol & Teodorescu, 2010), aplicație cu ajutorul căreia au

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

⁸ Propoziții și interpretările posibile (// /): Aseară (. / ! / ...), Vine mama (. / ! / ...), Ai venit iar la mine (! / ...), Cine a făcut asta (? / ? ! / ... /), Omul meu îl lucră (. / ! / ...), Oricum, îți poți câștiga locul dorit (. / ! / ...). Fraze cu subiect dublu sau apoziție: Vine ea mama!, "A trecut el așa un răstimp", (M. Sadoveanu), O ști el careva cum să rezolve asta, Mama vine și ea mai târziu, Mama știe ea ce face, Chiar știe el ce face?

fost colectate opiniile asupra recognoscibilității emoției în pronunția respectivă de la un număr mare de evaluatori. Conform scării folosite de (Beller, Obin, Rodet, 2008), „gradul de activare” (a emoției), pentru pronunțiile de propoziții din corpusul SRoL este între inactivare și activare mare, pe o scară cu patru trepte (inactivare, activare mică, medie și mare). Ambele tipuri de activare a emoției, activare extrovertită și introvertită sunt prezente în pronunțiile de propoziții din corpus. Toate emoțiile au fost de tip simulat („acted emotions”), dar „trăit”, în sensul că li s-a recomandat subiecților să se gândească la o situație reală care ar elicită emoția respectivă și deci pronunția emoțională dorită⁹.

3. Metodologia de analiză

3.1. Metodologia de prelucrare și analiză formantică (acustică)

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

Pentru prelucrări spectrale (detecția formanților și a frecvenței fundamentale), s-a selectat un instrument informatic dintre cele mai puternice la ora actuală și în același timp gratuit, utilitarul Praat™, conceput și pus la dispoziție comunității internaționale de către P. Boersma și D. Weenink (Boersma, 2002). Analiza minuțioasă a influenței deplasării ferestrei¹⁰ asupra rezultatelor (Teodorescu, Feraru) a indicat că alegerea este convenabilă și nu produce erori la variații mici ale pasului de deplasare sau ale dimensiunii ferestrei. Utilitarul Praat furnizează valorile frecvențelor medii ale formanților F_0 [Hz], F_1 , F_2 , F_3 , F_4 [Hz], pentru zonele delimitate la segmentare, precum și valori instantanee ale formanților pentru fiecare „fereastră” de analiză.

3.2. Eliminarea erorilor

Utilitarul Praat™, deși printre cele mai performante în prezent, produce cca. 10% erori evidente, din numărul total de determinări de valori medii, la valorile primilor doi formanți și aproape același procent de erori la determinările de frecvență fundamentală. Deoarece valorile „evident eronate” la care ne referim sunt adesea mult mai mari decât cele credibile, aceste erori afectează semnificativ, chiar la frecvențe de apariție 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 fișiere, nu se pune problema determinării manuale a valorilor corecte. Soluția la care

⁹ Gradul de activare al emoțiilor 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 puțin expresivă, iar ultima slab expresivă (ne-expresivă). Facem distincția între *expresivitățile pentru fiecare emoție în parte*, deoarece unele persoane pot activa și exprima ușor o emoție, de exemplu bucuria, dar nu și altele, de ex. „tristețea”. O persoană cu scoruri printre primii 25% dintre vorbitori (scorurile sunt calculate astfel: 4 pct. = f. expresiv, 0 = puțin expresiv; scor personal între 0 și 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 și cu suprapuneri ale ferestrei aluneceătoare (spurapuneri între ferestre succesive) de ordinul a 40%, ceea ce corespunde la deplasări de 100 ms ale ferestrei.

am recurs este de a elimina valorile „evident” eronate precum și 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 și a unor valori reale, dar anormale, nespecifice populației globale. Aceasta este corect din punctul de vedere al cercetării dacă suntem interesați de „procesele medii”, de caracterizarea globală și nu de elemente specifice, rare de pronunție, 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 instrumentului sau altor factori (înregistrări deficitare, pronunții deficitare). Aplicația 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 fișierul cu date primare sunt eliminate valorile evident eronate. După această operație de curățare, se determină statistica primară – valoarea medie și varianța pentru fiecare formant – și 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 populație. Statistica formantului fonemului se calculează cu valorile rămase după acest șir de corecții. Metoda este prezentată pe larg în alte lucrări.

3.3. Metoda de validare a expresiei emoționale

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

4. Discuție și concluzii

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

Înceiem cu un exemplu de aplicație deja inițiată de noi. Să considerăm problema determinării „triunghiului vocalelor” ($F_{1(k)}^v, F_{2(k)}^v$) pentru limba română¹¹. La nivelul actual, această determinare presupune: determinarea „norilor” vocalelor, a elipselor de încredere, global pe întreaga populație (masculin și feminin), separat pe sexe, separat pentru vocale izolate, vocale în cuvinte izolate, în propoziții, influența emoțiilor asupra deplasării triunghiului formanților (deplasare cunoscută în literatură, pentru alte limbi), compararea triunghiului pentru vocale susținute cu cel obținut 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 aceeași familie, eventual analiza pe dialecte, profesii etc. La acestea se adaugă analiza modificării funcțiilor densitate de probabilitate și a parametrilor lor (varianță, asimetrie, aplatizare) pentru cazurile enunțate mai sus. Într-un sens, acesta este și (parțial) programul echipei SRoL pentru viitorul apropiat.

Mulțumiri. Autorul multumește tuturor colegilor care au lucrat la corpusul SRoL pentru nenumărate și îndelungi discuții. Activitatea la SRoL a fost parțial sprijinită de către Academia Română.

Referințe bibliografice

- Beller, G., Obin N., & Rodet X., 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. *Glott 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. Științifică & Encicl., București, 1982
- SRoL, Voiced Sounds of the Romanian Language Project, (autori: Teodorescu, H.-N., Trandabăț, D., Feraru, M., Ganea, R., Verbuță, A., Zbancioc, M., Hnatiuc, M., Voroneanu, O., Pistol, L., Untu, A., Păvăloi, I.), www.etc.tuiasi.ro/sibm/romanian_spoken_language/ro/voci_emoționale.htm.
- Teodorescu, H.-N., Buchholtzer, L., Poșa, C. (1986). *Comunicarea orală om-mașină*. Cap. 2 – Vocea și vorbire, și Cap. 4 – Sinteza vorbirii, Editura Tehnică, Seria “Tehnica la zi”, București, 1986.
- Teodorescu, H.-N., Feraru, M., Trandabăț (2006), D., *Situl ‘Limba Română Vorbită’*. *Lucrările atelierului Resurse lingvistice și instrumente pentru prelucrarea limbii române*. Editori: Corina Forăscu, Dan Tufiș, Dan Cristea Iași, 3 nov. 2006, Editura Universității A.I.I. Cuza Iași, 3-7.
- Teodorescu, H.-N. (2010 a), *Noisy Speech Files Perceptual Mining for Speech and Noise Patterns*, PROMISE, 29 martie 2010, Iași
- Teodorescu, H.-N. (2010 b), *AI Tools for Speech Analysis Applied to the Romanian Language*. (Plenary paper), ECC 2010, 20-22 aprilie 2010, București
- Turculeț, A. (2002), *Tipuri de texte orale*. In: Klaus Bochmann, Vasile Dumbravă (Eds.), *Limba română vorbită în Moldova istorică*, Volume 1. Leipzig: Leipziger Universitätsverlag, 2002, pp. 53-78.

¹¹ Din nefericire, l. română este una dintre puținele limbi europene pentru care încă nu există decât studii semi-empirice (Rosetti, Lăzăroiu, 1982), (Teodorescu et al. 1986), invalide la nivelul cerințelor prezente, privind triunghiul formanților. V. si (Teodorescu, Feraru, Trandabăț, 2005).