

THE ROMANIAN MULTIMEDIA PROSODIC ATLAS (AMPROM)

Anca-Diana BIBIRI & Oana PANAITE & Prof. Adrian TURCULEȚ, Social-Human Interdisciplinary Research Department & Alexandru Ioan Cuza, University of Iași

The Multimedia Atlas of Romania (AMPROM) is a last generation atlas which combines principles of geolinguistics with techniques of instrumental phonetics and those of informatics, followed the scientific-methodological approach used by AMPER (*L'Atlas Multimédia Prosodique de l'Espace Roman*). AMPROM is conceived as an interactive database bringing together data collection and acoustic analysis concerning prosodic features of linguistic varieties specific to the Romanian language. The objective of the atlas is the documentation of geopro-sodic variation of Romanian language both at the level of dialects and idioms and at the level of colloquial literary language, which will enable the achievement of an intonative typology of the varieties of Romanian language.

1 Introduction

Romanian multimedia prosodic atlas - AMPROM - is the first prosodic atlas which aims to present the main prosodic patterns (intonation patterns) of the Romanian language varieties identified both at the level of the diatopic variants of the standard language and at the level of the dialect variants. AMPROM will be a completion of the Romanian linguistic atlases – The Romanian Linguistic Atlas (ALR) and NALR (The New Romanian Linguistic Atlas) by region (NALR and ALRR) – which have not recorded the intonation. This paper aims a broad presentation of the AMPROM project with current state of research, results and prospects for the Romanian language.

Starting point and pattern for AMPROM is *L'Atlas Multimédia Prosodique de l'Espace Roman* (AMPER), both atlases are in the stage of ongoing projects. AMPROM applies, with some own development, the design, methodology and tools developed within the Romanic project. A number of points in the AMPROM network will be found in AMPER (the part of the Romanic atlas dedicated to the Romanian language: AMPER-ROM[ÂNIA]).

AMPROM is designed in a binary structure, in terms of networks of points: the first deals with Romanian dialects and dialectal varieties (about 50 points of survey in rural areas, where the subjects have elementary education), while the second deals with the standard varieties of the diatopic language (about 20 points of survey made in cultural centers, aimed at subjects with higher education).

2 AMPROM methodology

As a computerized atlas, electronically (by the means of recording, storing, processing and audio-graphical representation of the data), AMPROM is a database, the multimedia atlas contains recorded utterances and sequences (with the informants' own speech), orthographically and phonetically transcribed, acoustically processed in text files (1, 2, 3. txt for the three repetitions of a sequence and 0. txt for the average of three repetitions), including the main physical features of each vowel from the utterance: the duration, maximum intensity (sound energy) and fundamental tone (F0) – the last one is measured in three points of vowel duration (at the beginning, in the middle and at the end). Based on these texts, graphics of intensity and duration are generated (in the form of histograms), melodic profiles of each statement, as well as average profiles (results based

on the three repetitions of a sentence). For the average profiles, tone .wav files are created through synthesis, containing ringtones of the content of segmentation; these synthetic tones can be used in tests of perception and recognition of intonation patterns, to assess and auto-evaluate.

3 Corpora

During the prosodic dialect investigations for AMPROM two questionnaires are used. The first questionnaire consists of a series of statements used for AMPER-ROM. The sets of statements that make up the questionnaire - established by morpho-syntactic and phonetic criteria - are formed by: *declarative* sentences (affirmative and negative) and total *interogative* sentences (affirmative and negative), having the syntactic structure of SVO (subject - verb - object) where S and O receive, in turns, adjective and / or prepositional determinants; the nouns and adjectives that are used in the utterances are trisyllabic oxitones, paroxitones and proparoxitones. Since in the Romanian language the negation receives usually the stress of the phrase, the negative declarative and interrogative-negative were also introduced in the questionnaire.

AMPER-ROM questionnaire (sequence):

twk Nevasta vede un căpitan/ The wife sees a captain

kwt Un căpitan vede nevasta/ A captain sees the wife

dwk Nevasta tinerea vede un căpitan/ The young wife sees a captain

gwt Un căpitan elegant vede nevasta/ An elegant captain sees the wife

swk Nevasta frumoasă vede un căpitan/ The beautiful wife sees a captain

pwt Pasărea vede nevasta/ The bird sees the wife

zwk Nevasta harnică vede un căpitan/ The hardworking wife sees a captain

bwt Pasărea papagal vede nevasta/ The parrot bird sees the wife

twg Nevasta vede un căpitan elegant/ The wife sees an elegant captain

fwg Pasărea frumoasă vede nevasta/ The beautiful bird sees the wife

Syntactic and phonetic restraints to which the 'fix' minimal corpus has to respond that is set for AMPER will be found, for the same reasons of contrastive analysis conditions, also in the questionnaire designed for AMPROM. However, to capture a larger number of Romanian intonation patterns in their territorial distribution, a second questionnaire includes other statements, simpler (without many formal constraints) to facilitate the contact with the subjects and to prepare them for the fixed questionnaire, the AMPER-ROM. That includes about 100 sentences and has two variants: low version (compulsory) and extended version (optional), the latter is done only in some points of inquiry being applied once, with the best informants.

There are presented the types of syntactic structures that make up the AMPRom questionnaire:

- VO structures (with inclusive subject): (1) 1a: L-ai văzut pe Ion?/ Have you seen John? (2) 2a: L-ai văzut pe Vasile?/ Have you seen the Basil? (3) 3a: Ai văzut fetele?/ Have you seen the girls?
- Structures pursuing the ratio of the word order and prosody: (1) 1b: Pe Ion l-ai văzut?/ John was that you have seen? (2) 2b Pe Vasile l-ai văzut?/ Basil was that you have seen? (3) 3b: Fetele le-ai văzut?/ Girls were that you have seen?
- VS/SV Structures: (25) 25a Vine Ion./ There comes John 25b Ion vine./John is coming (28) 28a Cine vine?/ Who is coming? 28b Ion vine./John is coming.
- Structures with double negation elements both in the question and the answer: (26): Nu vine nime(ni) la noi?/ There comes nobody(none)to us? (30): N-a venit nime(ni) la noi./Nobody(none)came to us.
- Structures in which modulators are used (adverbs of manner and semi-adverbs – sure, precisely, certainly, immediately, surely, maybe, whether, really or even modal verbs – I think, it might): 20b: Chiar vine Ion?/ Really, is John coming? 21a: Sigur/Precis (că) vine/Sure/precisely he is coming. (23) 23b S-ar putea să nu vină./It might be that he is not coming. 23c. Cred că vine./ I think he is coming.
- Structures containing different types of questions: partial, alternative, confirmation: (56) 56a: Cât e ceasul?/What time is it? 56b: Cât e ora? / What time is it? (41) Vii ori nu vii?/ Are you coming or not? (55) 55b: Pleci mâine la Iași, nu-i așa?/ You are going tomorrow to Iași, aren't you?
- Structures containing vocative addressing and calling: (40): *Ion (Ioane), dă-mi un măr (te rog)! / Ion (John), give me an apple (please)! (35): 35a: Ana! / Ann!, 35b: Maria! / Mary!, 35c: Ileana! / Helen!*
- Structures that require intonation suspension (to express the continuity): (49) – *Apucă-te/Ia și-nvață, că de nu... / Start/Let's learn, or else...*
- Structures that are intended to stress the prosodic expression that emphasizes the subordination relationship: (72) 72a: *Când am ajuns la piață, ploua cu găleata / When I arrived at the market, it was pouring; 72b: Ploua cu găleata, când am ajuns la piață. / It was pouring rain when we arrived at the market.*
- Exclamatory structures: (84): Ce batic frumos ai! / That's a beautiful scarf! (85): Ce miroase a pâine caldă! / That's a good smell of hot bread!
- Structures on intercalation prosody: (74) 74a: Tata mi-a zis: Du-te repede și cheam-o pe soră-ta! / My father said, 'Go quickly and call your sister'! 74b: Du-te repede și cheam-o pe soră-ta! mi-a zis tata. / Go quickly and call your sister! my father said. 74c: Du-te repede – mi-a zis tata – și cheam-o pe soră-ta! / 'Go quickly' my father said. 'and call your sister!'
- Structures containing enumerations: (66): Am fost la piață/târg și am cumpărat: roșii, ceapă, morcov și ardei. / I was at the market / fair and bought tomatoes, onions, carrots and peppers. (84): Luni, marți, miercuri...(și) duminică. / Monday, Tuesday, Wednesday ... (and) Sunday.

- Structures containing a sequence of short sentences: (79): De dimineață m-am trezit, am pregătit micul dejun și apoi am plecat la serviciu./ This morning I woke up, I made breakfast and then went to work.
- Sentences with the same structure (V) for the affirmative, interrogative and imperative mood: (80): *Așteaptă/Wait..* (81): *Așteaptă?/Wait?* (82): *Așteaptă!/Wait!/ Așteaptă-mă!/Wait for me!*
- Structures with a focus on constituents of the statement: (4): 4a Pe **Vasile** l-ai văzut ?/Was **Basil** that you saw? 4b L-ai văzut pe **Vasile**?/ Did you see **Basil**?; (58): Bei **vin**?/Are you drinking **wine**?
- Structures with a successive focus on constituents of the statement (64): *Mănânci pește?/ Are you eating fish?* 65a: *Mănânci pește?/ Are you eating fish?*
- Affective structures: (56f): E/îi amiază? / Is it/ It's noon? It's already noon? (59): **Bei vin?/Are you drinking wine?**

The extended form of the questionnaire contains other type of syntactic structures:

- Structures pursuing the prosody of idioms and phrases: (89 a, b, c...): da de unde/what, no way; nu mai spune/ yah, do not say; ce folos;/ so, what; nici vorba/pomeneala/no way/not at all; cum/unde să facă ea așa ceva/what/how did she do that; da mai știi?/that could be?, ei și?/so, what?, iote/iete/there/;; măi/uite/inga/ni la el!/Hey/look/you/you there.
- Structures containing greetings and politeness: (91): Bună ziua!/ Good afternoon!; (97): Poftim/There you go!/ Na!/Here! – Mulțumesc/mulțam!/ Thank you/Thanks – Poftim, pentru puțin, cu plăcere, să crești mare (la copii)./ There, Don't mention it, you are welcome, May you grow strong! (for children).
- Structures that use adverbs and adverbial phrases to strengthen the assertion and negation: (104): Da,/Yes Sigur,/Sure Firește/Surely, Negreșit!/No doubt! (105): Nu,/No Nicidecum,/No way Niciodată,/Never Nici în ruptul capului!/On no account!
- Imprecations: (107 a, b, c...): Arde-l-ar focu să-l ardă!/ May he burn in hell! Lua-l-ar naiba/dracu să-l ia!/ The hell/the devil with him! Fir-ar/fi-o-ar a dracului!/ Damn it/Damn with it! – Du-te dracului/la dracu/la satana!/ Go to the devil/to Satan!

In some sentences with neutral intonation required by the AMPER-ROM, questionnaire, during the prosodic dialects surveys we asked for limited focus (especially by contrast) of some constituents of the statement. Through the two questionnaires, AMPROM exceeds the fixed corpus of AMPER-ROM restrictions, both in terms of morpho-syntactic structures that are investigated and addressing other aspects of intonation than the neutral one. Statements are recorded at least three times and obtained through indirect questions (for attaching the involved words) and by verbal and non-verbal implications (facial expressions, gestures) to the context and / or forming some speech situations during the continuous dialogue between the investigator and informant, and, in some ultimate circumstances the investigator is saying the statement (with a monotonous intona-

tion, mechanical that does not suggest the actual intonation), the informant statements should not read the statements to avoid the specific "reading" intonation."

The investigation usually begins with discussions between the investigator and locals (free corpus), while the investigator has the opportunity to observe the intonation patterns of local speakers and to choose the best informants to achieve the prosodic dialectal investigations. Then it follows the AMPRom questionnaire, considering that it contains statements similar to usual speech, achieving a favorable atmosphere for the investigation. Then the questionnaire AMPER-ROM is accomplished, demanding that in the end is recorded the focused sentences which the accent placed on constituents from different sentences. The two surveys are repeated three times in different sessions on the same day or / and in the following days.

4 AMPRom network of points

The Romanian language field – as it appears in ALR (*Romanian Linguistic Atlas*) – contains the following dialects: D a c o - r o m a n i a n , A r o m a n i a n (M a c e d o r o m a n i a n), M e g l e n o r o m a n i a n and I s t r o r o m a n i a n with their subdialects and patois; that is the research area for both atlases, which have additionally included the diatopic varieties of standard language/culture. If in the project of AMPER 10 points of survey are included (with codes 90-99) for Romania and The Republic of Moldova (most of them being rural, but there are also cities where we want to record for the diatopic variants of the literary language), the network of points for the AMPRom starts from the historical provinces and it covers the whole territory of the Dacoromanian dialect: Moldovia, Greater Wallachia, Transylvania, Maramureș (part of the historic Partium region), Crișana (part of the historic Partium region), Banat (part of the historic Partium region), Lesser Wallachia, Dobruja, Bessarabia, but also the territories of the other three dialects. For the coding areas and survey points, the Romanian language was divided into 21 areas. Daco-Romanian dialect includes the provinces of Romania, labeled as follows: A = Moldova, B = south of Bucovinei (Romania), D = Transylvania, E = Maramureș, F = Crișana, G = Banatul, H = Greater Wallachia, J = Oltenia, K = Dobruja, to which there have been added the territories outside the Romanian borders where Romanian speakers are found in compact area: C = Republic of Moldova (Bessarabia), L = Chernivtsi and Transcarpathian, N = region of Ukraine, Odessa Ukraine = M, Vojvodina Timoc Valley (Serbia, Bulgaria) = N, Hungary = O. Aromanian dialect areas are: P = Greece, R = Albania, Macedonia = S Bulgaria = T Megleno dialect is found in: Republic of Macedonia = U, Greece = V; Istro-Romanian dialect is in Croatia = Z. Regarding Aromanian and Meglenoromanian dialects the record/ surveys are made/ will be conducted both in the Balkan Peninsula, and with the Aromanians and Meglenoromanians in Dobruja, where speakers of these dialects were settled in the third decade of the twentieth century.

The density of survey points for AMPRom will depend on their representativeness. We appreciate that it may be sufficient dozens (approximately 70) of points of prosodic dialectal investigation for the intonation patterns.

5 Informants

To achieve the speech prosody documentation, for both the local and standard language there may not be used the same informants; in the first case there will be surveyed rural subjects (selected from ALR and NALR network) and in the second case, the survey will

take place in the cities, in the most important cultural centers for that province. Among the classical criteria required in a dialectical survey that the informants must meet include: communicative availability, average age, spontaneity, minimal school education, good diction and voice, regular voice (not to lose his voice during the pronunciation of the utterance).

In rural areas two informants are used (with the code 1 (odd) for women and 2 (even) for male) indigenous, representative for the local speech, with elementary education (up to high school), middle-aged - 30-50 years (if necessary psychophysical conditions are met, they may be older) who speak natural under the conditions of the investigation. In urban areas the surveys twofold: besides informants 1 and 2 (belonging to lower social class / low and / or middle / middle, with influences of the local dialect), there are used informants 5 (female) and 6 (man) with higher education (belonging to the upper social class / high), speaking cultural language, but which are normal people (i.e. not the "professional speakers", more precisely those who work in the media, teachers, especially those teaching Romanian language and other languages). If there are more than four informants interviewed according to their socio-cultural status, they will receive tokens 3, 4 ... and 7, 8.

6 Acoustic processing of the recorded data

Acoustic analysis tools that are used in processing the prosodic dialectal material that was recorded during investigations are PRAAT / SCRIPT PRAAT for AMPER (Antonio Romano, Albert Rilliard), Matlab, AMPER 2006, Computer interface of prosodic curve. Statements are recorded in digital format (files with .wav extension - Waveform Audio File Format) and acoustic analysis using software tools. The sequence analysis goes through several stages: changing the sampling frequency sound wave of 48 kHz to 16 kHz (GoldWave) delineation and labeling according to the statements used in the questionnaire: numbering / coding of the statements is made for AMPER-ROM, keeping for AMPRom the first four indices of code. The numbering of the points includes number 9 – the code for the Romanian language, one letter - which indicates the area of the survey point (of the 21 mentioned above) and another figure - from 0-9 – from the number of (up to) 70 points of investigation. Encoding a sentence consists of 6 symbols: for the investigated point, the sign of the informant (by gender and training), encoding the statement - with a symbol consisting of numbers and letters (because of the various structure there is not permitted the coding type AMPER) - and its registration number (1, 2, 3 ...). For example, 9A011A1 is the first record of the first utterance of AMPRom questionnaire, formulated by the first female informant in the locality of Iași, Moldavia, in the Romanian language.

Assisted by the software - PRAAT – there follows the segmentation and labeling vowel elements (in the case of diphthongs the two vowel marks go together), based on oscillograms, spectrograms and by hearing, thus there occur, for each analyzed utterance, texts in which are found physical correlates of vowels: duration, intensity and fundamental frequency (F₀ – for the three points of the vowel). Based on these texts, there are obtained, using Matlab routines, average values (o.txt), duration graphs, intensity and individual melodic outlines etc.

7 Achievements and prospects.

The investigations for AMPRom began with a CNCSIS (Project Director Adrian Turculeț) grant was obtained and conducted during 2007-2008. During the preparation of surveys for AMPRom, for the Seminar of Phonetics and Dialectology, involving Bachelor and Master students of the Faculty of Letters, there were conducted sample surveys in some localities of Moldavia, Bucovina, Wallachia, Transylvania, Maramures and The Republic of Moldova. A first database was established for the geoprosodic research of the literary language variation and its regional variants. Data and their processing have been exploited in various communications and papers. We are currently in the stage of the processing of acoustic data collected from surveys conducted in dialectal areas of Moldavia, Bucovina, Maramures, Crisana, a stage in which there attend a team of PhD students and researchers from the *Department of Interdisciplinary Research in Humanities* at the "Alexandru Ioan Cuza" University of Iasi, the last two areas corresponding to AMPRom. *Maramureș* and AMPRom. *Crișana*, ongoing projects financed from the European Social Fund through the Sectoral Operational Program for Human Resources Development in the Project *Innovative Development and Research Impact through Postdoctoral Programs POSDRU/89/1.5/S/49944*); some of the points of the surveyed are part of the international project AMPER-ROM.

AMPRom website was made in spring 2011 and it contains the first database available for the specialists, as well as to anyone interested in issues of prosody. The site has been loaded with processed data analyzed from 5 cities in the area A, represented by Moldavia on the right bank of Prut river: 9A0 – Iași, 9A1 - Dolhasca, 9A4 - Pufești, 9A5 - Muntenii de Sus and 9A9- Liesti.

For the database – BD in Grenoble were sent eight points of the survey for the Romanian language (AMPER-ROM), which are also included on the DVD accompanying the paper *Intonations romanes*, done in 2011 in Grenoble. In addition, there were still processed 11 cities and 12 villages of Dacoromanian dialect and in the locality Gevgelja in the (Republic of Macedonia) (9Uo) for the Meglenoromanian dialect..

8 Conclusions

Acoustic-auditory research on Dacoromanians idioms done so far have shown that areas intonation do not coincide with dialectal areas that can be drawn using segmental data that are present in the Romanian linguistic atlases (RLA, NRLA). AMPRom aims the research of the dacoromanian dialects from a prosodic perspective, to demonstrate the overlap and differences between prosodic and dialectal areas. The Romanian Multimedia Prosodic Atlas (AMPRom) brings new data to the research of diatopice variation, but also to some diastratum aspects and diaphasic prosody for the Romance languages.

The Romanian language geography was incorporated into the Roman geolinguistic by traditional accomplishments found in language atlases: *Linguistischer Atlas für dacorumänischen Sprachgebietes* of Gustav Weigand, *Atlasul lingvistic român* (ALR), *Noul Atlas lingvistic român, pe regiuni* (NALR/ALRR) and subsequently through the atlases of the second generation, interpretative: *Atlasul limbilor romanice* (ALiR) and *Atlas linguarum Europae* (ALE). Through the accomplishment of proposed project there can be possible to integrate the Romanian language in the first prosodic atlas (AMPER) and to accomplish the first multimedia romanian prosodic atlas (AMPRom).

9 Bibliography

- Intonations romanes*, 2011. *Intonations romanes*, coordonné par Paolo Mairano. În „Géolinguistique“, Hors-série n° 4, Ellug, Université de Grenoble.
- Projet AMPER: Atlas Multimédia Prosodique de l'Éspace Roman*, 2005. În „Géolinguistique“, hors série 3 (ed. J.P. Lai), Ellug.
- Michel Contini. 2e Séminaire international du projet AMPER, 2005. În *Projet AMPER*, pp. I- XI.
- Michel Contini. *Le projet AMPER: passé, present, et avenir*, 2007. În L. de Castro Moutinho, R. L. Coimbra (ed.), *I Jornadas Científicas AMPER-POR. Actas* (Aveiro, 29-30 octobre 2007), pp. 9- 19.
- Laurenția Dascălu-Jinga, 2001, *Melodia vorbirii în limba română*, Editura Univers Enciclopedic, București.
- Adrian Turculeț, L. Botoșineanu, Ana-Maria Minuț, *Atlasul multimedia prosodic roman (AMPRom). Chestionarul și rețeaua de puncte*, 2006. În „Limba și literatura română. Regional-Național-Universal“ (Simpozion internațional Iași-Chișinău, 24-27 noiembrie 2005), Iași, Casa Editorială „Demiurg“, p. 283-293.
- Adrian Turculeț, Ana-Maria Minuț, *De la AMPER la AMPRom*, 2007. În Luminița Hoarță Cărăușu (coord.) *Rezultate și perspective actuale ale lingvisticii românești și străine*, Editura Universității „Al. I. Cuza“, Iași, pp. 349-361.
- Adrian Turculeț, 2007, *Un nou atlas lingvistic romanic: AMPER*,. În „Studii și cercetări lingvistice“, ianuarie-iunie, 2007, București, pp. 203-214.
- Adrian Turculeț (ed.), 2008, *La variation diatopique de l'intonation dans le domain roumain et roman*, Editura Universității „Al. I. Cuza“, Iași.
- Adrian Turculeț, Oana Beldianu, Anca-Diana Bibiri, 2012, *Pentru un Atlas multimedia prozodic român (AMPRom)*. În *Colocviul Internațional „Filologia modernă: Realizări și perspective în context european. Abordări interdisciplinare în cercetarea lingvistică și literară“*, Academia de Științe a Moldovei, Institutul de Filologie, Chișinău, pp. 397-402.

<http://amprom.uaic.ro/>

<http://w3.u-grenoble3.fr/dialecto/AMPER/amper.htm>

Acknowledgements: The paper was done with support from the European Social Fund through the Sectoral Operational Program for Human Resources Development in the Project Innovative Development and Research Impact through Postdoctoral Programs POSDRU/89/1.5/S/49944.

Keywords: AMPRom, acoustic analysis, prosodic features, intonative typology of the varieties of Romanian language.