

On the syllabic structures of Aromanian

Sergiu Nisioi

Faculty of Mathematics and Computer Science

Center for Computational Linguistics

University of Bucharest

Bucharest, Romania

sergiu.nisioi@gmail.com

Abstract

In this paper we have investigated the syllabic structures found in Aromanian a Romance language spoken in the Balkans across multiple countries with important communities which spread from Greece to Romania. We have created a dictionary of syllabified words and analyzed a few general quantitative and phonological aspects of the dictionary. Furthermore, we have approached the syllabic complexities, the sonority patterns present in the syllable's constituents and the degree in which the Sonority Sequencing Principle (SSP) holds for this language. Based on all the information gathered we have devised an automatic syllabification algorithm which has a 99% accuracy on the words in the dictionary. In this way we hope to extend the existing phonological studies on Eastern Romance and to spread and preserve meta-linguistic information on this endangered language.

1 Introduction

Aromanian, according to linguists (Papahagi, 1974) or (Saramandu, 1984) is part of a larger family of Eastern Romance languages consisting from Daco-Romanian (standard Romanian), Aromanian, Megleno-Romanian and Istro-Romanian. We underline this characteristic because some of the linguistic properties that are present in Aromanian are also present, more or less, in the other three languages largely due to the common historical and linguistic context in which they formed and evolved. Unfortunately, Istro-Romanian and Megleno-Romanian are labeled by the UNESCO Red Book (2010) as “severely endangered” languages with approximately 300 (Filipi, 2002) and respectively 5000 speakers (Atanasov, 2002).

Aromanian is not in a better situation carrying the label of “definitely endangered” with approximately 500,000 speakers (Atanasov, 2002). Currently, the language has no accepted standard, being written in various forms depending on social and political factors of the regions in which it is spoken (Kahl, 2006). This is why we believe that a study on Aromanian can only be done empirically, on corpus data or dictionaries or by adopting a multi-valued dialectal approach justified by field work.

2 Previous work

Although Capidan (1932) offers an exhaustive study on Aromanian with valuable comments on previous research, very few modern linguistic studies target this language and there are no computational linguistic studies as far as we are aware at this point. Caragiu-Marioțean (1968) offered one of the first modern studies with respect to the phonology and structural morphology of the language; her work represents our linguistic baseline. Since Aromanian is spoken in various regions, there are expected geographical particularities. Caragiu-Marioțeanu (1997) classifies the Aromanian sub-dialects into two types: *type F* - the variants that resemble the Farsherot, predominant in Albania and some parts of Greece and *type A* - all the other variants. She argues that type F sub-dialects are spoken by smaller communities which have been influenced by type A. Type F sub-dialects have certain phonetic features: the closed-central vowel [i] does not exist, the groups of consonants [rl], [rn] are transformed into a sound that could be classified as a velar [r] (Capidan, 1932) and the diphthongs [ɛ̞a] and [o̞a] are transformed into [e] and [o].

Place →	Labial		Coronal			Dorsal	
Manner ↓	Bilabial	Labio-dental	Dental	Alveolar	Postalveolar	Palatal	Velar
Trill				~r			
Lateral approximant				~l		ʎ	
Nasal	~m			~n		~ɲ	
Sibilant fricative				s~z	f~ʒ		
Non-sibilant fricative		f~v	θ~ð			ç~j	x~χ
Stop	p~b		t _̃ ~d _̃			c~ç	k~g
Affricates			t̃s~d̃z		tʃ~tʒ		

Table 1: The consonant inventory of Aromanian as it is described by Caragiu-Marioțeanu (1975). Our dictionary uses the same alphabet to store the syllabified words.

3 Dictionary of syllables

The dictionary used in our study is compiled by Cunia (2010) from the dictionaries of Papahagi (1974) and the one of Dalametra (1906). The main advantage of this resource is its lexical richness, including specific variants of words known by the author. Moreover, for most of the words, the syllabification is reproduced from a different dictionary compiled by Papahagi (1974). The later is considered a valuable and reliable linguistic resource for this language. The final size of the dictionary has approximately 69.000 syllabified words. Among the disadvantages, we could count a significant amount of misspelled words (that we have manually corrected), and words syllabified incorrectly. Also, the dictionary is written with the orthography proposed by Cunia (1997) using an alphabet that has only one diacritic. The purpose of the alphabet is to be more practical in the digital era, departing from other related phonemically spelled languages like Italian or Spanish. One of the main drawbacks of this alphabet is the compression of two different sounds (the mid-central vowel [ə] and the closed-central vowel [i]) into the letter “ă”, leaving the reader to decide which phoneme is actually used. The motivation behind the compression comes from the two vowels being geographical allophones for Aromanian (depending on type F or type A sub-dialect). In this case we should be cautious when analyzing vowels as phonological units. This is not the only case where orthography can influence our study.

Aromanian, together with Romanian (Chitoran, 2001), contrasts diphthongs [ɛa], [ɔa] from semivowel (glide) to vowel sequences [ja] and [wa] (written here as [j̣a] and [ẉa]). This means that the phonological representation of diphthongs is the one proposed by Chitoran (2002): *both ele-*

ments of the diphthong are represented as sharing a syllable nucleus. According to this representation, diphthongs are predicted to function as a single unit and a single segment.

The orthography of our dictionary restricts us to partially operate with these distinctions. Similar to Romanian (Chitoran, 2001), the glide-vowel sequence [ɔa] is less frequent - in our dictionary with less than 100 occurrences - compared to the larger number of occurrences for the diphthong [ɔa]. However, other dictionaries like the one of Caragiu-Marioțeanu (1997), which is complete only to the letter ‘D’, uses a different orthography and the actual contrasts might differ from resource to resource. Since this doesn’t guarantee consistent results we represent internally all the above pairs as single units in the nucleus.

In the current state, the exact phonetic value of the letter “ă” is ambiguous and the disambiguation is a non-trivial task unless parallel resources are available.

To overcome the difficulty of using a specific alphabet, we have decided to convert the entire set of consonants from Latin script into IPA (International Phonetic Alphabet). The consonant inventory of Aromanian is detailed in Table 1, the same alphabet is used internally to store the syllabified words. Our representation does not have the same amount of detail as a phonetic transcription would, instead, it offers a general unified format which could be used by linguists in future studies. Our dictionary also contains the accented vowels and a distinction between [u, e, o, i] and the semivowels [ɥ, ɛ̣, ɔ̣, ɨ̣].

The palatalized pairs of consonants [c]~[ç] and [ç]~[j] are to be found only before [i, ɨ̣, e, ɛ̣] according to Caragiu-Marioțeanu (1968). In all the other places the velar ones will be encountered. In practice, native speakers do not always use the

palatalized before [i, i̯, e, e̯] but we will keep this rule to be consistent across the dictionary.

Romanian		Aromanian	
CV structures	Percentage	CV structures	Percentage
cv	55.04%	cv	24.84%
cvc	15.06%	cvc	6.28%
v	6.91%	cvcv	5.07%
ccv	5.82%	cvv	2.76%
cvv	5.43%	ccv	2.54%
vc	3.40%	v	2.36%
cvcv	2.86%	vc	1.47%
ccvc	1.33%	ccvc	1.29%
vv	0.83%	cvvcv	0.91%
cvcc	0.73%	cvvc	0.37%
cvvc	0.43%	cccv	0.36%
ccvv	0.24%	vv	0.26%
cvccv	0.23%	ccvcv	0.25%
vcv	0.23%	vcv	0.22%
cvvv	0.22%	ccvv	0.20%
cccv	0.22%	cvcv	0.13%
vvc	0.16%	cvvv	0.12%
ccvcc	0.12%	cccvc	0.12%
cvvcv	0.11%	vvcv	0.09%
cccvc	0.11%	vvc	0.09%

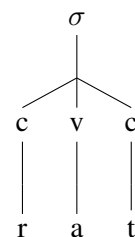
Table 2: The first most common CV structures in Romanian and Aromanian. Semivowels are also denoted with “v”. The syllables in both languages allow complex CV structures. Both languages share similar structures the difference is made in the distribution of each. The most frequent structure is, in both cases, “cv” (consonant vowel).

4 Syllabic structures

4.1 CV structures

Both Aromanian and Romanian share a large degree of common features, but no comparative studies have been made on the CV (here: consonant-vowel) structures available in the two languages. For Romanian, a database of syllables was already provided by (Barbu, 2008) under the title “RoSyllabiDict”. This database contains almost all the morphological forms of Romanian words. Among the existing studies with respect to the CV structures in Romanian we count the one of Dinu and Dinu (2006). Comparing two grammatically similar languages in terms of the distribution of CV units can bring a new perspective on the similarities at the phonological level. In CV-theory (Clements and Keyser, 1983) a syllable σ is represented in a three-tier form.

For example the word “rat”:



The CV structures are a part of the phonological layer in the universal grammar and the most common one, encountered in all the natural languages is the “cv” structure. In Table 2, both Romanian and Aromanian have this structure as the most frequent one. Theoretically, the four primary types of CV-structures are “cv”, “v”, “cvc” and “vc”. The CV-theory (Clements and Keyser, 1983) predicts that if a language has “vc” syllables then all the other three primary structures will be encountered in that language.

In Table 2, it’s not unusual to see CV structures of the form [ccvcv], this is because the standard Romanian orthography, as opposed to English in most of the cases, makes no distinction between the grapheme of a semivowel and the one of a vowel. If the CV structure has the following form [ccvcv] then the second “v” is a glide. In our internal representation of our dictionary two glides can be encountered at the end of syllables: [i̯] the mark for plural in all the Romanian dialects and [u̯] - frequently emphasized in the texts since the first Aromanian writers of the eighteenth century (Papahagi, 1909).

4.2 Menzerath-Altmann law

Menzerath-Altmann law (Altman, 1980) states that the size of a linguistic construct is inversely correlated with the size of its constituents. Which means, in this particular case, that the average size of a word in syllables increases as the average size of the syllable (in phonemes) decreases. Previous studies proved that this law has applicability for more general linguistic constructs, in syntax (Buk, 2007) and even beyond linguistics in genome structures (Baixeries et al., 2013). In the syllable-phoneme context, extensive studies have been made. Fenk et al. (2006) investigated this law on 33 different languages and found an active correlation between the CV complexity of the syllables and the decay of the ratio between number of syllable and number of phonemes per syllable. In this sense, we have investigated the relation between Romanian and Aromanian with re-

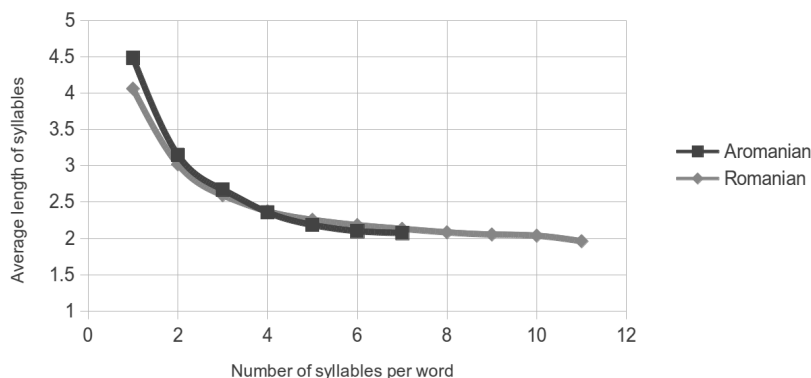


Figure 1: Menzerath-Altmann law. The word/syllable ratio in Romanian and Aromanian

spect to the Menzerath-Altmann law. We used the “RoSyllabiDict” dictionary of Romanian syllables compiled by Barbu (2008). This dictionary is constructed from almost all the morphological word-forms in Romanian, having a considerable size of about 520.000 entries.

The results can be visualized in Figure 1 - Romanian has a smaller average length of syllables, somewhere close to 4 but a significantly larger average length of words - close to 11. While Aromanian has a slightly larger length of syllables, close to 4.5 but the length of the words does not exceed 7 syllables in average. Moreover, Romanian is a highly developed language containing a vast set of neologisms and loans that can affect the word length. While Aromanian is more an archaic language spoken in small communities usually used between family members lacking the lexical richness of a general-use language. On one hand, Aromanian has a smaller average length of words than its developed relative, on the other hand, at the phonological level, even though the decay is similar (overlapping most of the times), Aromanian presents a slightly larger length of syllables (in phonemes). This suggests that Aromanian is slightly more complex in terms of syllable phonotactics than Romanian.

4.3 The structure of the syllable

In order to investigate the complexity and diversity of the syllables in Aromanian, we have chosen to examine the constituents of the syllables in terms of sonority sequences. Phonetically, the spoken chain consists in waves of sonority or sound intensity (Lehmann, 2005). The Sequence Sonority Principle (SSP) (Clements, 1990) regards the syllable as the phonological unit behind the waves of

sonority. A representation of this concept can be visualized in Figure 2.

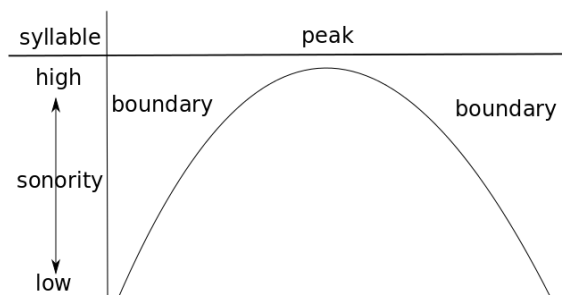


Figure 2: The sonority wave

The sonority of a phoneme can be regarded as “its loudness relative to that of other sounds with the same length, stress, and pitch” (Ladefoged, 1975). The sonority is given by a concept called strength (Escure, 1977), on one side strength can be represented by the sonorance in which the phonemes are ordered by their acoustic energy: Stops → Fricatives → Nasals → Liquids → Glides → Vowels. On the other side, the scale can be represented by the articulatory resistance (Anderson and Ewen, 1987) of the phonemes as in Figure 3.

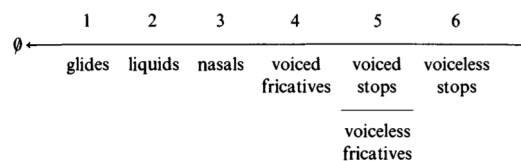
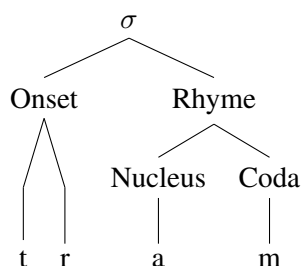


Figure 3: Scale of articulatory resistance

The Sonority Sequence Principle states that the syllable’s peak is a group of segments of high

sonority while the syllable's boundaries consists of phonemes of low sonority. In almost every natural language there are exceptions to this principle and investigating it can be valuable in speech recognition and in automatic syllabification based on phonotactic rules (Frampton, 2011). If the exceptions are accounted then the number of sonority peaks in a word is correlated with the number of syllables. In the same manner, the number of syllable boundaries is correlated with the number of low sonority phonemes.

As previously mentioned, the "cv" structure is universal in every language, thus the syllable may have two basic constituents (Fudge 1969, 1987): an onset (governed by the consonant) and a rhyme (governed by the vowel). The rhyme is further divided into a nucleus (forming the syllable's sonority peak) and a coda (consonants of descending sonority), the following schema exemplifying the word "tram" is relevant to the definition:



A constituent (onset, nucleus, coda) is branching if multiple phonemes are to be found in its structure and non branching if it is constructed from a single unit. Onsets and codas in Aromanian can be empty (syllable made of nucleus only - "v"), branching (two or more consonants "ccv" for onsets and "vcc" rarely for coda) or non-branching ("cvc" - the most frequent construct, single consonant only). In Table 2 the CV structures already suggested this fact. Compared to codas, the onsets in these languages tend to be more complex branching in up to three consonants.

4.4 Sonority and the coda

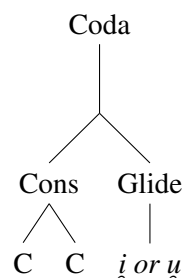
The first observation arises with respect to the Aromanian coda and the fact that it can end in glides [i̯] and [u̯]. This creates a peak of sonority inside the structure of the coda, thus a sonority reversal right at the end of the syllable. In this situation the SSP is broken since the sonority is not decreasing towards the end boundary of the syllable. These types of codas appear in final syllables, the semivowel [i̯] being morphologically

determined while the semivowel [u̯] is a particular feature of Aromanian. Table 3 contains the percentages regarding the sonority of the codas. Because of the final glides, a large number of codas (41%) will break the SSP by having a sonority reversal (a sequence of phonemes in ascending sonority inside the coda). Mixed codas have a sequence of phonemes that is neither ascending neither descending.

Coda sonority	Percentage
Ascending	41.21%
Descending	58.72%
Mixed	0.06%

Table 3: Percentages of the syllable coda. The majority labeled with 'Ascending' are sonority reversals constructed from Cons + Glide.

Given that the presence of these glides in word-final syllables is frequent in all the Romanian dialects, we decided to adopt the following structure of the coda for word-final syllables:



Using this design we have also investigated the sonority of the coda in its "Cons" structure which is limited in Aromanian to not having more than two consonants.

It was noticeable to observed that the SSP is always obeyed by the "Cons" substructure of the coda.

4.5 Sonority and the onset

All the results so far, indicate that the coda is not a very complex structure in this language. In fact, Caragiu-Marioțeanu (1968) stated that Aromanian previously had open syllables. A fact attested in the early works of the eighteenth century writers (Papahagi, 1909) describing the syllables as being opened. The linguistic study of Davis and Baertsch (2011) offers a model in which the structure of the onset is related to that of the coda through a shared component called margin. The model predicts that a complex onset in a language

requires the presence of a coda (Kaye and Lowenstamm, 1981). For Aromanian, in particular, the formation of the coda could be a result of the increased complexity of the onsets.

We have investigated the different patterns of sonorities found in onsets, Table 4 contains the percentages of each of these patterns. Aromanian has four types of sonority sequences in the onset:

- “Constant” \longrightarrow - one or more consonants (sonority plateau) with equal sonorities
- “Ascending” \nearrow - a sequence of phonemes with ascending sonorities
- “Descending” \searrow - a sequence of phonemes with descending sonorities
- “Nadir” $\searrow \nearrow$ - a sequence in which the sonority descends and then rises towards the nucleus (e.g. the onset “mbr” or other Nasal+Stop+Liquid)

Onset sonority	Percentage
Constant	89.10%
Ascending	5.63%
Descending	4.11%
Nadir	1.13%

Table 4: Sonority patterns found in syllable onsets.

The “Constant” and “Ascending” sequences of sonorities in onsets obviously obey the SSP and they count as the majority in the language. The ascending onsets can take the following forms: $[bl]$, $[br]$, $[dr]$, $[\ðr]$, $[fl]$, $[f\lambda]$, $[fr]$, $[gl]$, $[g\lambda]$, $[gn]$, $[gr]$, $[\gamma l]$, $[\gamma n]$, $[\gamma r]$, $[kl]$, $[k\lambda]$, $[kr]$, $[ks]$, $[kf]$, $[pç]$, $[pl]$, $[p\lambda]$, $[pr]$, $[ps]$, $[sl]$, $[sm]$, $[tl]$, $[tr]$, $[tsr]$, $[vl]$, $[vr]$, $[xl]$, $[x\lambda]$, $[xr]$, $[zl]$, $[zm]$, $[zn]$, $[\theta r]$ from which the onsets ending in nasal consonants (i.e. $[n]$, $[m]$ or $[p]$) are found only in word initial syllables.

About 5% of the onsets can be classified as exceptions from the SSP and the majority of them are to be found in word-initial syllables. In word-medial syllables we could count only rare examples of fricative + stop clusters. The descending consonant clusters in the onsets can be constructed by the patterns in Table 5.

The most interesting phonotactic constraint to mention is related to “Nadir” onsets - all of them appear only in word initial syllables. Intuitively, we may consider these onsets as being constructed

$[f] \sim [v]$	+ $[t] \sim [dz]$
$[m]$	+ bilabial, fricative
$[n]$	+ most of the less sonorous consonants
$[s] \sim [z]$	+ $[p] \sim [b]$, $[t] \sim [d]$, $[k] \sim [g]$, $[\underline{t}f] \sim [j]$
$[j]$	+ $[k]$, $[p]$, $[t]$, $[\underline{t}s]$

Table 5: Descending consonant clusters in the onset. The marker \sim underlines the voiceless/voiced feature of the sounds. The phonemes tend to cluster together depending on the voice (e.g. $[z]$ being a voiced consonant is more likely to be encountered near other voiced consonants - $[b]$, $[d]$, $[g]$ or $[j]$).

from two types of clusters: on one hand “Descending” + “Constant” onset clusters and on the other hand “Constant” + “Ascending” onset clusters.

Quantitatively, the two approaches are equivalent and have the following form:

1. “Descending” onset cluster + liquid ($[l]$, $[\lambda]$ or $[r]$)
2. $[m]$, $[n]$, $[s]$, $[f]$, $[z]$ + “Ascending” onset cluster

From a linguistic perspective the onsets that do not respect the SSP can be analyzed using the concept of semisyllable. Cho and King (2008) proposed this model of a syllable by imposing certain restrictions:

- no nucleus
- no codas
- no stress/accent/tone
- prosodically invisible
- well-formed onset clusters (observing SSP)
- restricted to morpheme peripheral positions

The concept has been applied on Georgian, Polish and Bella Coola - languages with highly complex clusters of consonants (Cho and King, 2008) and even on French (Féry, 2003) to split complex codas. In Aromanian, for both “Nadir” and the “Descending” sequences, the semisyllable are word-initial. These semisyllables contain only the onset from one of the following phonemes: f , v , m , n , s , j or z .

5 Phonetic syllabification algorithm

Studying the phonotactics of a language can be valuable for rule based automatic syllabification (Clements, 1990). Previous studies on phonologically complex languages like Imdlawn Tashlhiyt Berber (Frampton, 2011) validated the universality of this approach. The work of Iacoponi and Savy (2011) addresses the same problem on Italian, their rule-based phonetic algorithm reaching a precision of over 98%. Not any phonetic algorithm can be generalized or applied to different languages but in all the cases the same pattern is preserved: the syllable boundary is defined by a point of low sonority in a sequence of phonemes, see Figure 2.

In our particular case, based on the previous analysis of the sonority patterns, we have devised the following seven rules for establishing a syllable boundary:

1. diphthongs [ɛ̃a] and [ɔ̃a] are treated as single units (Chitoran, 2001)
2. maximal onset principle: $\searrow |c+ \nearrow$ - if the phoneme c is a sonority minimum then the syllable boundary is placed before the minimum and c is added to the next onset (Kahn, 1976)
3. word-medial c +nasal split: $\searrow +c| \nearrow$ - if the consonant c is a sonority minimum and it is followed by a nasal consonant (i.e. [n], [m] or [ɲ]) then the syllable boundary is placed after the minimum and c is added to the current syllable's coda - based on the results in Section 4.5
4. special [s] + stop cluster - the fricative consonant [s] will be treated as having the same sonority as any other stop consonants ([p], [b] etc.)
5. split plateau: $\longrightarrow | \longrightarrow$ - if a sonority plateau is found then put a syllable boundary in between
6. initial semisyllables consisting from consonants will be glued to the immediately next syllable
7. word-final coda can end in one or two glides - as described in Section 4.4

The second and the third rules are the ones referring to the actual minimum points of sonority within a word. The key is whether we want to cut the syllable before or after the minimum point and this fact is determined by phonotactic constraints. The initial semisyllables, although they respect the SSP, are merged within the next syllable and the word final codas may end in glides.

We have compared the output of this algorithm with the actual data already in our dictionary. Almost one percent of the words were incorrectly syllabified because they were exceptions to the above rules and the overall precision was 99%. This algorithm can easily be extended to other Eastern Romance languages by verifying the seven rules provided.

6 Conclusions and future work

In this paper we have offered a quantitative approach to the syllable structures and substructures found in Aromanian. In addition, we propose a dictionary resource to inspire future studies and to help preserve a “definitely endangered” Romance language. It is not an easy task to execute a study on a language that lacks an institutionalized standard. Our approach is empirical, corpus based and the quality of the results is strictly dependent on the quality of the corpus. This is why we have focused on investigating general phonological properties of the language in the limits afforded by the corpus at our disposal. Comparisons with existing studies on Romanian reaffirm the tight relation between the two languages and offer confidence that the results obtained are in accordance with existing facts about them. Moreover, based on phonotactic investigations, we have constructed an algorithm for automatic syllabification in Aromanian that has a 99% accuracy.

Future studies involve developing this resource to disambiguate the letter “ã” and adding more detail in the phonetic representation by recording native speakers. The current phonological study will help us to further develop rule-based resources on morphology considering existing theoretical studies (Caragiu-Marioțeanu, 1968) and may help us to better understand the evolution of the Eastern Romance languages and the relations between them. Last but not least, we hope that developing these linguistic and computational resources will encourage a widespread use of Aromanian.

References

- Altmann, G. 1980 *Prolegomena to Menzerath's law* In *Glottometrika* 2, pp. 1-10
- Anderson, John M. and Colin J. Ewen 1987 *Principles of Dependency Phonology* In Cambridge University Press
- Atanasov, Petar 2002 *Meglenoromänisch. Lexikon der Sprachen des europäischen Ostens*, Wieser Enzyklopädie des europäischen Ostens 10
- Baixeries, Jaume, Hernandez-Fernandez, Antoni, Forns, Nuria, Ferrer-i-Cancho, Ramon 2013 *The parameters of Menzerath-Altman law in genomes* In *Journal of Quantitative Linguistics* 20 (2), pp. 94-104.
- Barbu, Ana-Maria 2008 *Romanian Lexical Data Bases: Inflected and Syllabic Forms Dictionaries* In Proceedings of the Sixth International Conference on Language Resources and Evaluation, pp. 28-30
- Buk, Solomiya and Rovenchak, Andrij 2007 *Menzerath-Altman Law for Syntactic Structures in Ukrainian* In *CoRR Journal* 2, pp. 1-10
- Boiagi, M. C. 1813. *Grammatiki Romaniki itoi Make-dones Vlachikiki*, Wien
- Capidan, Theodor 1932. *Aromânii. Dialectul aromân. Studiu lingvistic* Monitorul Oficial și Imprimeriile Statului, Imprimeria Națională, București
- Caragiu-Marioțeanu, Matilda 1968 *Alterntion. Fonomorfologie aromână. Studiu de dialectologie structurală* Editura Academiei Române, Bucharest
- Caragiu-Marioțeanu, Matilda 1975 *Compendiu de dialectologie română* București, Editura științifică și enciclopedică
- Caragiu-Marioțeanu, Matilda 1997 *Dicționar aromân (macedo-vlah) DIARO. A-D. Comparativ (român literar-aromân)* Editura Enciclopedică, Bucharest
- Chitoran, Ioana 2001 *The Phonology of Romanian: A Constraint-Based Approach*, Mouton de Gruyter, Berlin, New York
- Chitoran, Ioana 2002 *A perception-production study of Romanian diphthongs and glide-vowel sequences* *Journal of the International Phonetic Association*, 32, pp. 203-222
- Cho, Young-mee Yu and King, Tracy Holloway 2008 *Semisyllables and Universal Syllabification* In *The Syllable in Optimality Theory*, Cambridge University Press, pp. 183-212
- Clements, George N. and Keyser, Samuel Jay 1983 *CV Phonology: A Generative Theory of the syllable* MIT Press. Cambridge
- Clements, George N. 1990 *The role of the sonority cycle in core syllabification* In J. Kingston and M. E. Beckman (eds.) *Papers in Laboratory Phonology I*, pp. 283-333
- Cunia, Tiberius 1997. *On the Standardization of the Aromanian System of Writing*, The Bituli-Macedonia Symposium of August 1997
- Cunia, Tiberius 2010 *Dicționar a limbăljei armânească*, Editura Cartea Aromână
- Dalamestra, I. 1906 *Dicționar macedo-român*, Editura Academiei Române, Bucharest
- Davis, Stuart and Baertsch, Karen 2011 *On the relationship between codas and onset clusters* In *Handbook of the syllable*, Leiden, pp. 71-98 Netherlands: Brill.
- Dinu, Anca and Dinu, Liviu P. 2006 *On the data base of Romanian syllables and some of its quantitative and cryptographic aspects* In Proceedings of the Fifth International Conference on Language Resources and Evaluation, pp. 1795-1798
- Escure, G. J. 1977 *Hierarchies and phonological weakening* *Lingua* 43 (1), pp. 55-64
- Fenk, A., Fenk-Oczlon, G. and Fenk, L 2006 *Syllable complexity as a function of word complexity*, In The VIII-th International Conference "Cognitive Modeling in Linguistics" Vol. 1, pp. 324-333
- Féry, Caroline 2003 *Markedness, Faithfulness, Vowel Quality and Syllable Structure in French* *Syllable complexity as a function of word complexity*, *Journal of French Language Studies*, Volume 13, Issue 2, pp. 247 - 280
- Filipi, Goran 2002 *Istrorumänisch. Lexikon der Sprachen des europäischen Ostens*, Wieser Enzyklopädie des europäischen Ostens 10
- Fudge, E.C. 1969 *Syllables* In *Journal of Linguistics* 5, pp. 253-286.
- Fudge, E.C. 1987 *Branching Structures within the Syllable* In *Journal of Linguistics* 23:359-377
- Frampton, John 2011 *GDE syllabification A generalization of Dell and Elmedlaoui's syllabification algorithm* *The Linguistic Review*, 28, (3), pp. 241-279
- Iacoponi, L. and Savy, R. 2011 *Sylli: Automatic Phonological Syllabification for Italian* In proceeding of: INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association, pp. 27-31
- Kahl, Thede 2006. *Istoria aromânilor*, Editura Tritonic, București
- Kahn, Daniel 1976. *Syllable-based generalizations in English phonology*, Doctoral dissertation, MIT

- Kaye, Jonathan and Lowenstamm, Jean 1981 *Theory of markedness in generative grammar* Pisa, Italy: Scuola normale superiore di Pisa.
- Lehmann, Christian 2005 *Latin syllable structure in typological perspective* Journal of Latin Linguistics, 9 (1), pp. 127-148
- Ladefoged, Peter 1975. *A course in phonetics* Harcourt Brace Jovanovich : New York
- Moseley, Christopher (ed.) 2010. *UNESCO Atlas of the Worlds Languages in Danger* 3rd edn. Paris, UNESCO Publishing
- Papahagi, Pericle (ed.) 1909. *Scriitori aromâni in secolul al XVIII-lea (CAVALIOTI, UCUTA, DANIL)* Editura Academiei Române, București
- Papahagi, Tache 1974. *Dictionarul dialectului aromân. General și etimologic* Editura Academiei Române, București
- Saramandu, Nicolae 1984. *Româna, in Tratat de dialectologie românească* Scrisul românesc, Craiova