Syllable Structure in Spoken Arabic:
a comparative investigation

Rym Hamdi[*], *Salem Ghazali*[**] *& Melissa Barkat-Defradas*[***]

[*] Laboratoire Dynamique Du Langage - UMR CNRS 5596 (Lyon)
[**] I.S.L.T. Université 7 novembre à Carthage (Tunisia)
[***] Laboratoire ICAR-Praxiling - UMR CNRS 5191 (Montpellier)
Rim.Hamdi@univ-lyon2.fr, s.ghazali@irsit.rnrt.tn, melissa.barkat@univ-lyon2.fr

Abstract

The aim of this study is to demonstrate that rhythm variation across Arabic dialects is to a great extent correlated with the different types of syllabic structure observed in these dialects, especially with regard to the relative complexity of onsets and codas. The main focus is on the relationship between syllabic structures on the one hand, and rhythm classes based on segmental duration on the other. Rhythmic variations in Arabic dialects based on proportions of vocalic and consonantal intervals were described in [1] following experimental procedures put forth by [2]. The present investigation consists in computing the frequency of occurrence of the different types of syllables in Moroccan, Tunisian and Lebanese Arabic representing different areas along the dialect continuum. The experimental data is based on the production of 10 minutes of spontaneous speech by Moroccan, Tunisian and Lebanese subjects. The results show that the syllabic patterns observed could be useful in characterizing the different Arabic dialects and may also constitute a basis for discriminating between them. The occurrence of the various types of syllables is significantly different from one dialect to another; the percentage of simple syllables and long vowels (CV, CVC, CVV, CVVC…) is higher in Lebanese Arabic. Moroccan Arabic, on the other hand, shows a tendency towards a high percentage of complex syllables (CCVC, CCVCC,..etc) and short vowels.

1. Introduction

This study is part of our attempts to identify prosodic cues that may play a determining role in discriminating between Arabic dialects. We believe that in addition to the observed phonetic, phonological, morphological and lexical differences, prosodic elements such as intonation, rhythm and syllable structure may constitute reliable cues in distinguishing between the various Arabic dialects.
In previous studies [1] [3], we have highlighted rhythmic variations across these dialects using a phonological approach to rhythmic interpretation proposed by [4] [5]. This approach, which was further elaborated by [2] and later by [6], is based on the quantification of relevant phonological properties. The implementation of this model, which considers rhythm as the sum of these properties, relies on the measurements of syllable complexity and vowel reduction. In applying this modal of rhythmic analysis to Arabic, it became apparent that rhythm varies from one dialect to another and that it does so in a gradual way, thus suggesting the presence of a rhythmic continuum rather than discrete absolute categories [1]. One of the results was that the standard deviation of consonantal intervals (ΔC) is higher in the North African dialects than in those of the Middle East, which suggests the presence of a more complex syllabic structure in the Western dialects.

These values of ΔC are also found to be higher compared to other stress-timed languages such as English [7].
In this paper, we will provide additional data from a detailed investigation of the frequencies of different types of syllables in three Arabic dialects in an attempt to establish a direct relationship between syllable structure type and rhythm tendency.

Furthermore, most studies on Arabic dialectology have included differences in syllable structure, especially between the Western and the Eastern dialect, among the feature that distinguish between these dialects. However, no detailed investigation, to our knowledge, has compared yet these different structures or assessed their role as possible identification cues.

2. Rhythm and syllable

It is widely accepted that the syllable plays a major role in the rhythmic organization of speech. Cyclic movements of the mandible corresponding to syllables have been described as accounting for the acquisition of language by infants as well as for the evolution of human language (Frame then content theory) [8] [9]. In current phonological theory, syllables and related constituents (feet and moras) are basic units in accounting for rhythmic structure in languages.

For both phonetic and phonological theories, two types of dominant syllables emerge: CV and CVC, with CV being the most frequent and the most universal type [9]; perhaps the one from which all other syllables are derived [10]. These two basic syllable forms seem to yield more frequent occurrences of open syllables (CCV, CCCV…) if CV is the dominant syllable in a particular language, and more occurrences of closed syllables (CCVC, CVCC, CCVCC…) when CVC is the most typical one [11].

Lexical databases where words have been segmented into syllables have been used to shed light on the internal structure of syllables as well as on the way syllables combine to form words. Using data from ULSID (UCLA Lexical and Syllabic Inventory Database), [11] set up a typology taking into consideration: the number of syllables in a lexical unit, preferred syllable types, general trends in syllabic patterns and relationships between the different constituents of a syllable. In this paper, we will attempt to follow a similar methodology hoping to contribute to a better understanding of the way syllable structures vary across Arabic dialects, and the degree to which such variations are correlated with different rhythmic patterns.

3. Syllable structure in Modern Standard Arabic (MSA)

In MSA, consonant clusters are not permitted in syllable initial position. Thus, in a word such as [χalfa] "behind" the intervocalic consonants must belong to two different syllables CVC$CV. There are three underlying syllables in MSA: CV, CVC and CVV and two syllables CVVC and CVCC that appear only in surface phonetic forms such as at pause or following other phonological processes. On a basis of data of 100 000 syllables, the frequency of occurrence of CV, CVC and CVV are 49.7%, 23.9% and 17% respectively. If we consider conjunctive *hamzas* as full-fledged consonants then CVVC and CVCC syllables together barely account for 1% of the occurrences of all the syllables [12].

As will be shown below, North African dialects have undergone, in addition to vowel reduction, a process of short vowel deletion in open syllables. This process has resulted in various consonant clusters in both onset and coda positions.

4. Syllabic Structure in the dialects

4.1. Methodology

In this section we will examine the occurrences of different syllabic structures in three Arabic dialects: Moroccan Arabic representing the Western variety, Lebanese Arabic for the Eastern variety and Tunisian Arabic representing the intermediate region. Grouping the dialects into these three sub-classes was motivated, among other things, by our findings pertaining to rhythm classes. The subjects who produced the data are from the cities of Casablanca for Morocco, Tunis for Tunisia and Beirut for Lebanon.

The results are obtained from speech samples representing about 8 to 10 minutes of spontaneous speech at normal speaking rate from each dialect. The speech data comes from the *Araber* corpus [13]. Segmentation was carried out manually following the way spoken utterances were realized and not according to the underlying representations of the separate lexical items. As [14] stated, '*in spontaneous speech, the phonetic realization often differs markedly from the canonical, phonological representation. Entire phone elements are frequently dropped or transformed into other phonetic segments. These patter of deletions and substitutions appear rather complex and somewhat arbitrary when analysed at the level of the phonetic or phonological segment. However, this variation becomes systematic when placed within the framework of the syllable'*. Furthermore, we used the same type of phonetic segmentation in our investigation of rhythmic patterns, which will allow for a reasonable comparison between the results from the two investigations.

4.2. Types of syllable structures:

Syllable Type	Moroccan	Tunisian	Lebanese
cv	32,82	35,40	43,23
cvc	22,74	30,57	33,61
cvv	4,52	4,65	5,23
cvvc	1,94	8,03	5,32
v	3,75	4,29	3,55
vc	0,90	1,64	0,28
cccv	2,45	0	0
cccvc	1,55	0,20	0
cccvcc	0,26	0	0
cccvv	0,52	0	0
ccv	10,98	4,01	2,99
ccvc	10,85	4,38	3,73
ccvcc	1,03	0,64	0
ccvv	0,90	1,73	0,28
ccvvc	1,68	2,65	0,84
cvcc	3,10	1,82	0,93

Table 1 : Frequency of occurrences of syllable types in the three Arabic dialects (expressed in %).

Table 1 shows that, according to this sample, the three dialects differ with respect to type of syllables they permit as well as to their relative frequencies. While Moroccan Arabic exhibits 16 types of syllables, Tunisian Arabic has 13 and Lebanese only 11. In terms of syllable complexity (table 2), some syllables in the Moroccan dialect may include up to 3 consonants in onset position and 2 in the coda. Table 2 also shows that closed syllable types outnumber open ones (9 and 7 respectively).

Empty onset	Simplex onset	Complex onset	
v	cvcc	ccv	cccv
vc	cv	ccvc	cccvc
	cvc	ccvcc	cccvcc
	cvv	ccvv	cccvv
	cvvc	ccvvc	

Table 2 : Types of the syllable structures classified in term of complexity.

Note also that CV and CVC are the two dominant types, since together they account for 55% of all the syllables in Moroccan, 65% in Tunisian and 76% in Lebanese. Within this group, CV is the most frequent syllable in all three dialects as it is the case in the majority of languages. However, CV syllables are much more frequent in the Lebanese dialect than in the North African varieties (figure 1).
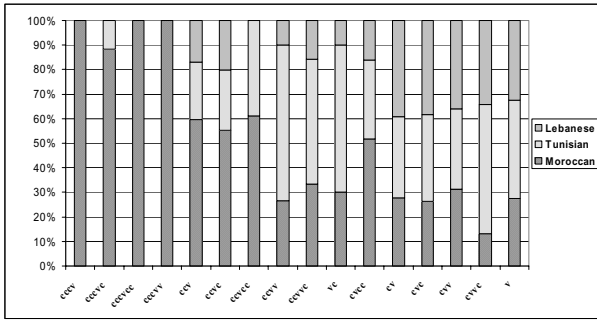
Figure 1 : Distribution of the syllable types in the 3 dialects.

The frequency of occurrence of these types of syllables confirms what was suggested by [15] on the basis of samples from 24 languages. According to [15], the presence in a language of a complex syllable whether closed or open, entails the presence of the less complex one in that language, with complexity being estimated in terms of the number of segments in the onset or the coda. This implicational tendency is confirmed by [11] who suggests that if a language has syllables with complex onsets comprising n consonants then that language also has syllables with onsets containing n-i consonants (i=1 for n-1) as illustrated below

$$CCCV \Rightarrow CCV \Rightarrow CV$$
$$CCCVC \Rightarrow CCVC \Rightarrow CVC$$
or
$$CVCCCC \Rightarrow CVCCC \Rightarrow CVCC \Rightarrow CVC$$
$$VCCC \Rightarrow VCC \Rightarrow VC \Rightarrow V$$

However, results from our sample do not support to the observation that a language will have more occurrences of closed or open syllables depending on whether the dominant syllable is CVC or CV. In our data, although there is a preference for other closed syllables in Tunisian Arabic, there is not much difference in terms of frequency between closed and open syllables in the two other dialects (figure 2).
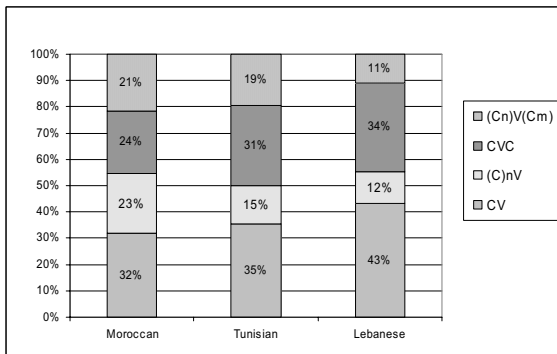

Figure 2: Frequency of occurrence of open and closed syllables in relation to the dominant CV and CVC syllables.

4.3. Syllabic complexity:

The general observation that one can make in the light of this investigation is that consonant clusters are more frequent in the North African dialects, especially in Moroccan Arabic. The latter exhibits syllables that comprise clusters of 3 consonants in syllable-initial position and have the highest frequency of occurrence of all types of complex syllables (figure 3). These clusters are mainly present in onset position, which confirms earlier findings stating that clusters occur more frequently at onset than coda positions [16] [17]. The results also go along the lines of "the maximal onset principle" in phonological theory. The prevalence of clusters in syllable-initial position may in part be explained by the fact that a substantial number of consonants can appear in onset position (from 87 to 100 %) but can not occur at the end of the syllable (from 21 to 82 %), as there are more constraints on coda positions [11].
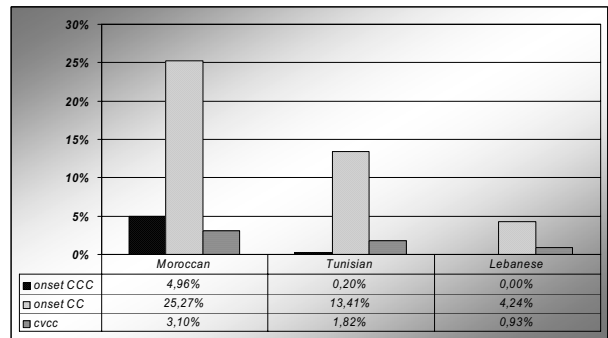

Figure 3 : Proportions of different types of complex syllables in the three dialects.

As mentioned above, Moroccan Arabic is well known for the process of vowel deletion in open syllable resulting in complex onsets and codas. Regarding this dialect, [18] notes: *«This structure of Moroccan Arabic is famous for his consonant overload, similar in that with the Berber structure. It is because of this aspect that Eastern Arabic finds difficulty of understanding Moroccan Arabic».*
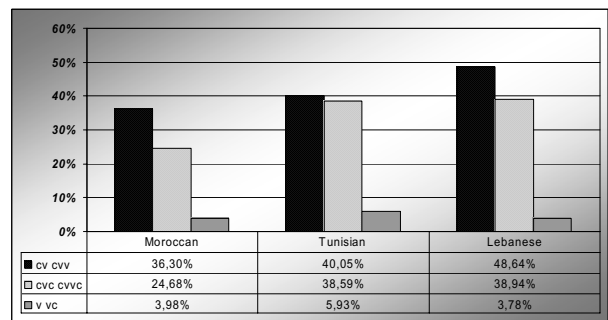

Figure 4 : Proportions of different types of simple syllables in the three dialects.

In Tunisian Arabic however, syllable complexity is somewhat intermediate between the Moroccan and the Lebanese dialects. This dialect represents the Western edge of the central sub-group which includes Egyptian Arabic on its Eastern end. Short vowels in open syllables also drop in most Tunisian dialects including the variety used in this investigation. Vowel reduction, however, is less important in this dialect than in the Moroccan variety which explains the

more frequent occurrences of syllables containing long vowels (CVVC, CVV, CCVV and CCVVC). Note also that syllables beginning with CC cluster are more frequent in the Tunisian than the Lebanese variety which supports the intermediate status of Tunisian Arabic (figures 3 and 4).

4.4. Syllable structure and the acoustic correlates of rhythm

In this section we will compare our observations regarding the structure of syllables in these three dialects to the acoustic correlates of rhythm discussed in previous investigations. Figure 5 illustrates the correlation between ΔC and the proportion of vocalic intervals (%V) in 7 Arabic dialects.
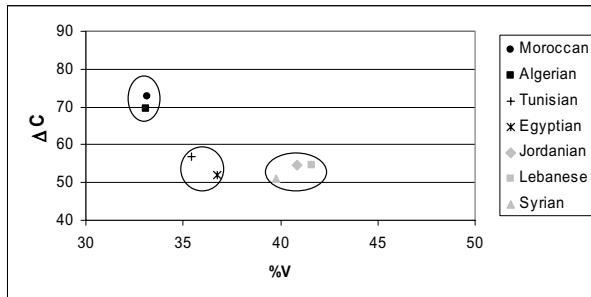
Figure 5 : Distribution of languages & dialects along the %V (x axis) and ΔC dimensions (y axis)

ΔC and %V were shown to be directly associated with vowel reduction and syllabic complexity [2]. Since complex syllables include more consonants than simple ones, the more complex syllables a language has the higher the average proportion of consonants and the lower the %V. At the same time, languages with complex syllables also have simple syllables thus; the large variety of syllable types also increases ΔC.

The distribution of syllable types in the three dialects corresponds to the values of ΔC and %V illustrated in figure 5. Note the negative correlation (r=-0.75) between these two values indicating that Moroccan Arabic which has the largest number of syllable types (16) and the highest number of complex syllables (33% against 15% and 9% for Tunisian and Lebanese Arabic respectively) also shows the highest ΔC values and the opposite is true for %V. Note also the intermediate values both in terms of ΔC and %V and in number of syllable types and their complexity in Tunisian Arabic. These values give more support to our earlier remarks about the division of Arabic dialects in sub-classes [1].

5. Conclusion

In conclusion, our results show that syllable structures can be used as a cue to distinguish between the Eastern and Western varieties of Arabic. These structures are also found to correspond to rhythmic sub-classes established on the basis of measurements from the acoustic signals of ΔC and %V parameters. The fact that geographically intermediate dialects such as Tunisian Arabic also exhibit intermediate values for the parameters investigated provides further support for the suggestion that Arabic dialects form a continuum at least with regard to sound patterns.

6. References

[1] Ghazali, S., Hamdi, R. and Barkat, M., "Rhythm Variation in Arabic Dialects". *In Proceeding of Speech Prosody,* Aix en Provence, 2002.

[2] Ramus, F. and Melher, J.: "Language identification with suprasegmental cues: A study based on speech resynthesis", *J. Acoust. Soc. Amer. Vol 105, 1999, p 512-521.*

[3] Ghazali, S., Hamdi, R. and Knis, K., "Intonational and Rhythmic patterns across the Arabic dialect continuum", to appear *19th Annual Symposium on Arabic Linguistics,* university of Illinois at Urbana-Champaign, 2005.

[4] Dauer, R .M. "Phonetic and phonological components of language rhythm", *Proceedings of the XIth ICPhS,* Tallinn, Estonia, Vol.5, 1987, pp. 447-450.

[5] Roach, P., "On the distinction between 'stress-timed' and 'syllable-timed' languages", *In Linguistic controversies, D. Crystal Ed, 1982, pp. 73-79.* London: Edward Arnold.

[6] Grabe, E. and Low, E.L. "Durational Variability in Speech and the Rhythm Class Hypothesis", *in Laboratory Phonology 7,* Mouton, 2002.

[7] Hamdi, R. Barkat-Defradas, M. and Ben Hamed., M., "Discrimination des langues et des dialectes arabes par le rythme", *JEP*, 2004, Fès. Morocco.

[8] MacNeilage, P.F. and Davis, B., "On the Origin of Internal Structure of Word Forms", *Sciences, Vol. 288, 2000, pp.527-531*

[9] MacNeilage P.F., "The Frame/Content Theory of Evolution of Speech Production", *Behavioral and Brain Sciences, Vol. 21, 1998, pp. 499-511.*

[10] Clements, G.N. and Keyser, S.J., *CV Phonology. A generative theory of the syllable,* Cambridge, Massachusetts, MIT Press 1983.

[11] Rousset, I, and Vallée, N., "Vers une organisation syllabique des lexiques : Tendances, dépendances et cooccurrences segmentales"; *JEP*, 79, Nancy, 2002.

[12] Braham, A and Boukaddida, F, unpublished, IRSIT, Tunisia.

[13] Barkat, M.; Hamdi, R.; Pellegrino, F., "De la caractérisation linguistique à l'identification automatique des dialectes arabes", *MIDL. Workshop, 2004*, Paris.

[14] Greenberg S. "A syllable-centric framework for the evolution of spoken language", *Behavioural and Brain Sciences, Vol. 21, 1998, pp. 518.*

[15] Blevins J., "The syllable in Phonological Theory", *Handbook of Phonological Theory, Goldsmith J.A., lackwell, 1995, pp. 206-235.*

[16] Zerling, J.P., "Structure syllabique et morphologique des mots à caractère onomatopéique et répétitif en français. Étude phonétique", *Travaux de l'Institut de Phonétique de Strasbourg, TIPS, 30, pp. 115-162, 2000.*

[17] Molinu. L and Romano A, "la syllabe dans un parler roman de l'Italie du Sud", *Journée d'études linguistiques 'SyllabeS', AAI*, pp. 148-153, 1999, Nantes.

[18] Youssi A., "La triglossie dans la typologie linguistique", *La linguistique*, 19 (2), 1983, 71-83, Paris.