

Why Moroccan Arabic tolerates anything word-initially, but Slavic does not

Current generalisations on syllabic structure make crucial reference to the word-initial site in order to determine which sequences are well-formed branching Onsets (=those occurring initially) and which ones are to be regarded as Coda-Onset clusters. This kind of approach has been developed on the basis of, and is designed for languages of the Indo-European (IE) kind where only #TR clusters occur word-initially (T=any obstruent, R=any sonorant). It has nothing to say about languages that do not obey this pattern. Clements (1990) quotes a number of them, and it appears that two genetically defined groups are prominent, i.e. modern occidental Afro-Asiatic languages (Berber, Moroccan, Algerian, Tunesian Arabic) on the one hand, and Slavic on the other. The goal of this talk is to 1) reconcile these "misbehaving" languages with syllabic theory and 2) show why any combination of two consonants is allowed for initially in the Afro-Asiatic type, but only certain #RT-sequences are tolerated in Slavic.

First, I present the analysis of the well behaving IE languages under the assumption that syllable structure boils down to a strict alternation of non-branching Onsets and non-branching Nuclei (Lowenstamm 1996), and that the phonological identity of the left margin of the word is an empty [CV]-unit (Lowenstamm 1999). This initial CV is present in a language like German. [#RTV] = /CV₀-RøTV₁/ sequences are ruled out in this kind of language because the only proper governor available, V₁, cannot satisfy the ECP of both empty Nuclei V₀ and ø at the same time. By contrast in a [#TRV] = /CV₀-TøRV₁/ cluster, R may establish a consonantal governing domain over T and thereby satisfy the ECP of ø (Scheer 1999). V₁ governs properly V₀ to the effect that the structure is well formed.

In "misbehaving" languages, [#RTV] clusters may be well-formed only if the empty Nucleus enclosed within RøT is properly governed (as before, RT may not form a consonantal governing domain). This means, in turn, that there is no initial CV unit in these languages. Note that the fact of being initial does not automatically mean that the CV-unit is distributed, just as a given morphological border may be relevant to the phonology in some languages, but not in others. In absence of the initial CV, the theory thus predicts that any combination of initial consonants is allowed for. This is indeed the situation met by a language like Moroccan Arabic. If all "misbehaving" languages were like Moroccan Arabic, the entire typology of the world's languages, as far as initial clusters are concerned, could thus be captured by setting one simple parameter: presence (==> only #TR) vs. absence (==> #TR and #RT) of the initial CV. The third logical possibility, i.e. only #RT, is ruled out since the only way to take care of the empty Nucleus located within #RøTV is Proper Government coming from V, which would legitimate the ø in #TøRV as well. And indeed, a language where #RT clusters occur, while #TR sequences are excluded, is not attested.

But what about Slavic? In contrast to Moroccan Arabic, only a small subset of logically possible initial combinations of two consonants exist. For instance, #rd and #lp do occur (Czech *rdít se* "go red", *lpít* "stick"), but no word begins with #rp, #lt in any Slavic language. All attempts to characterise the existing or the non-existing set of initial clusters as a natural class have failed (e.g. Cyran&Gussmann 1999). No phonological principle appears to govern their distribution. Recall that as soon as a language tolerates a single instance of #RT, theory predicts that the grammar does impose no restriction on initial clusters.

The answer to this paradox, I submit, comes from diachronic analysis. I show that in both

types of languages, Arabic and Slavic, ALL present-day initial #RT clusters come from former #RVT sequences. This is obvious in Arabic since Classical Arabic did not admit any initial cluster at all. In the evolution to Moroccan Arabic for example, short vowels became schwa, and long vowels were shortened. As schwas are subject to phonotactic alternation with zero, a verbal form such as *katab-a* comes out as [ktIb] (where "I"=schwa). Hence, whatever the first vowel of the word in Classical Arabic, it is reduced to zero in the modern Moroccan variety whenever it was short, thereby establishing the contact between the initial and the second consonant. As the first vowel was always short in verbal forms of Classical Arabic and no restrictions applied to the distribution of the first two consonants, any initial combination is attested in Moroccan Arabic.

The Slavic situation is the same to the extent that all modern #RT clusters come from former #RVT sequences. I present a list that ambitions to collect the exhaustive record of all roots that are instantiated as #RT in any Slavic language (total number: 45) in order to show that all of them go back to #RyerT in Common Slavic. However, Common Slavic had 11 different vowels (counting short and long versions of the same vowel for two), but only two of them, i.e. the two yers, did reduce to zero. This situation is in sharp contrast with the one encountered in Moroccan Arabic where any initial vowel was lost.

In sum, I propose that the distribution of initial #RT clusters is always driven by the lexicon, never by any phonological principle. If all initial vowels fall out as in Moroccan Arabic, anything is possible word-initially. If only two out of 11 vowels are reduced, only the arbitrary subset of lexical items in which they occupied the first vocalic position may come out as #RT. In any case, the gaps in the distribution of #RT, I argue, are not systematic but accidental.

This way, the "misbehaving" languages have no longer to be regarded as phonological aliens. Their difference with respect to "well behaving" languages may be described as a simple parameter setting (presence vs. absence of the initial CV). Moreover, the contrastive behaviour of two major genetically defined groups of "misbehaving" languages, Arabic and Slavic, is explained.

References

- Clements, George N. 1990. The role of the sonority cycle in core syllabification. Papers in Laboratory Phonology I, édité par J.Kingston & M.Beckmann, 283-333. Cambridge: Cambridge University Press.
- Cyran, Eugeniusz & Edmund Gussmann 1999. Consonant clusters and governing relations: Polish initial consonant sequences. The syllable, Views and Facts, édité par Harry van der Hulst & Nancy Ritter, 219-248. Berlin, New York: de Gruyter.
- Lowenstamm, Jean 1996. CV as the only syllable type. Current trends in Phonology. Models and Methods, édité par Jacques Durand & Bernard Laks, 419-441. Salford, Manchester: ESRI.
- Lowenstamm, Jean 1999. The beginning of the word. Syllables?!, édité par John Rennison & Klaus Kühnhammer, 153-166. La Hague: Holland Academic Graphics.
- Scheer, Tobias 1999. A theory of consonantal interaction. *Folia Linguistica* 32, 201-237.