

Distributional gaps in Slavic initial clusters are accidental

It is generally held that the distribution of consonants in word-initial clusters is the result of grammatical activity, i.e. that gaps are systematic. For example, typical Indo-European languages such as French, English, German etc. show only obstruent-sonorant sequences here (s+C clusters lain aside). These #TR-only languages (where T represents obstruents, R sonorants) are opposed, on the other end of the initial complexity scale, to modern occidental Arabic varieties (Moroccan, Algerian, Tunesian) and Berber, where any combination of two consonants (hence #TR, #RT, #TT, #RR) is found (anything-goes languages).

I have examined the Slavic situation in regard of initial clusters in detail. Slavic languages are known for their permissive initial clusters, but examples are often eclectic (except in the literature on Polish). In order to allow for an assessment of the overall picture, I have compiled a list that ambitions to document all initial sonorant-obstruent clusters (i.e. the most offending ones in regard of sonority sequencing) in 13 Slavic languages (Cz, Slovak, Po, Kaš, Up Sor, Low Sor, Bu, Mac, S-Cr, Sloven, Ru, Ukr, Bieloru). Leaving aside those languages that do not show any #RT clusters (Bu, Mac, Sloven, Bieloru), where sources were too scarce (Up Sor, Low Sor, Kaš) or the initial sonorant of #RT clusters always syllabic (hence enjoying vocalic value, S-Cr, e.g. *rvati* "to fight"), we are left with 5 languages where the distribution of initial #RT clusters may be studied: Cz, Slovak, Po, Ru and Ukr.

Polish has a long tradition in the study of "offending" initial clusters (e.g. Kuryłowicz 1952, Rubach & Booij 1990, Cyran & Gussmann 1999). The strategy has always been the same as in the familiar #TR-only languages: the initial pattern observed is thought to be enforced by the grammar of the language. Hence the existing as well as the non-existing clusters must somehow form a natural class whose guiding principle needs to be discovered. Theory, then, must be able to accommodate the critical opposition. All attempts at characterizing either the existing or the non-existing clusters in terms of natural classes have failed: whatever segmental or syllabic property is chosen, there are always items that should not be there, and others that should exist but do not. In short, the picture appears to be entirely anarchic, disobeying any possible organizing principle. The same holds true for the other four languages, which actually make the overall picture much worse since the distribution of their initial clusters is also anarchic, but anarchic in four different ways.

Independently of this issue, the theory that I am working in (CVCV, Lowenstamm 1999, Scheer 2004) rebels against the typological situation that would result if initial clusters in the five languages quoted were instances of natural classes. That is, there would be languages where no #RTs are tolerated at all, others where **all** logically possible #RTs are well-formed and **do** exist, finally an in-between group, among which the five Slavic languages in question, where some #RTs exist, but others do not (e.g. 18% of logically possible #RTs exist in Cz, 25% in Po). Indeed, according to the theory of the initial CV (Lowenstamm 1999), the choice of how grammar restricts initial clusters is only **binary**: the beginning of the word may or may not be marked by the presence of an empty CV unit. The presence of an empty Nucleus before the initial consonant enforces the #TR-only pattern, while its absence produces anything-goes languages. Hence there is no third option, and a prediction is made to the effect that a grammar that allows for just one single #RT cluster will also allow for all others - this is a grammar of the Arabic type.

The 5 Slavic idioms must thus be anything-goes languages - in other words, the missing #RTs are possible initial clusters: they are accidental, not systematic gaps. This prediction is confirmed by the diachronic study of the inventory of #RTs: all modern #RTs in all Slavic languages come from a Common Slavic sequence #RyerT. Of course, there were no distributional restrictions between C1 and C2 in a CS #C1yerC2 sequence, and hence the co-occurrence of both consonants is pure lexical accident. The modern words, then, are simply their CS ancestors without the yers. Modern grammar allows for any #RT - the gaps are only due to the accidental absence of a CS yer-separated ancestor.

I show that this result, which explains the anarchic distributional situation in the modern languages and conforms to the prediction of the theory, is confirmed by lexical creation, i.e. loanwords or acronyms with non-native #RTs that sneak into the languages at hand without any problem (e.g. *Mtacminda* "mountain in Tbilisi" or *Mcyri* "poem by Lermontov" into Ru).

Literature

- Cyran, Eugeniusz & Edmund Gussmann 1999. Consonant clusters and governing relations: Polish initial consonant sequences. *The syllable, Views and Facts*, edited by Harry van der Hulst & Nancy Ritter, 219-248. Berlin, New York: de Gruyter.
- Kuryłowicz, Jerzy 1952. Uwagi o polskich grupach spółgłoskowych [Remarks on Polish consonantal groups]. *Biuletyn Polskiego Towarzystwa Językoznawczego* **11**, 54-69.
- Lowenstamm, Jean 1996. CV as the only syllable type. *Current trends in Phonology. Models and Methods*, edited by Jacques Durand & Bernard Laks, 419-441. Salford, Manchester: ESRI.
- Lowenstamm, Jean 1999. The beginning of the word. *Phonologica 1996*, edited by John Rennison & Klaus Kühnhammer, 153-166. La Hague: Holland Academic Graphics.
- Rubach, Jerzy & Geert Booij 1990. Edge of constituent effects in Polish. *Natural Language and Linguistic Theory* **8**, 427-463.
- Scheer, Tobias 2004. *A Lateral Theory of Phonology. Vol.1: What is CVCV, and why should it be?* Berlin: Mouton de Gruyter.