On the phonotactic judgments of Czech native speakers[*]

Péter Rácz

1 INTRODUCTION

This paper aims to examine the claim that Czech has no restrictions on word-initial consonant clusters.[1] Support for this claim is two-fold: firstly, the language has word forms with initial *sonorant + stop* clusters, unattested in a number of other languages, including English, Hungarian, or French. Secondly, speakers have no alleged qualms accepting nonce forms or loans with similar initial clusters either. Consequently, the phonotactics of the language should not impose restrictions on word-initial clusters.

I have two major arguments in support of my suspicion concerning the equal status of words with an initial *sonorant + stop* cluster (henceforth #RT) and words with an initial *stop + sonorant* cluster (henceforth #TR) in Czech. Firstly, the number of #TR forms in the lexicon of Czech seems to be far greater than that of #RT ones. Secondly, as a nonce-word experiment shows, while Czech native speakers do accept the latter to an extent, they show a clear preference for the former.

These details suggest that a view of phonotactics as a set of across-the-board restrictions is slightly too generous: in the case of Czech, such a view either admits #RT into the set of acceptable phonotactic sequences, thereby severely over-generating (as most of these clusters are not attested at all, and some only show up in small numbers), or excludes them, considering a bulk of forms as lexical exceptions. The categorical approach can be replaced with a gradient one, where the well-formedness of forms is determined by comparison to existing forms, giving a scale of phonotactic well-formedness from completely ordinary to rare to impossible.

The structure of the paper is as follows: in Section 2, I review the claims on Czech word-initial phonotactics. Then, in Section 3, I discuss the nonce-word experiment conducted with native speakers. Finally, I provide a general discussion in Section 4.

---

[1]In this paper T stands for any obstruent, R for any sonorant, V for any vowel, # for a word boundary.

2   WORD-INITIAL CC CLUSTERS IN CZECH

Czech seems to allow a large number of word-initial consonant clusters that other languages tend to evade. Here are a few examples from Scheer (2006):

(1)   **stop + fricative:**   *pšota* 'misery'
       stop + stop: *kterak* 'which'
       fricative + stop: *sto* 'hundred'
       affricate + stop: *čtení* 'reading'
       nasal + fricative: *mžít* 'drizzle'
       nasal + stop + liquid: *mdlít* 'faint,weak'
       liquid + fricative: *rzi* 'rust'
       liquid + stop: *lkát* 'to whine'

Of course, the language has *stop + liquid* clusters as well. The above list is curious as most languages, including the majority of Indo-European languages, forbid *sonorant + stop* clusters word-initially. The standard claim is that these clusters violate the principle of *sonority* (Selkirk 1984): sonority, the relative loudness of a segment, must rise towards the syllable peak and fall towards the coda. As stops are less sonorous than sonorants, a #RT word violates this principle.

While sonority seems to evade attempts at a tangible definition (Harris 2005), we can agree that *sonorant + stop* (and *stop + stop*) clusters cross-linguistically imply *stop + sonorant* clusters: if a language has the former, it has the latter, and there are languages with only the latter.

2.1   Syllabicity?

The first thought that springs to one's mind when encountering written forms like *rtut'* 'quicksilver' or *drtit* 'to crush' is that in these forms the sonorant is syllabic, and forms its own syllable peak, saving the sonority patterning of the whole form itself. Trubetzkoy (1939) and Scheer (2001), however, claim that Czech has a predictable distribution of syllabicity: sonorants are syllabic between two obstruents and word-finally, following a consonant, but not word-initially before another consonant.

We can see this for ourselves by looking at the spectrograms of two forms, *vrtit* and *rtut'*, as uttered by a native speaker, in Figure 1. In the case of *vrtit* (upper panel), syllabic *r* is predicted: observe that the largest and longest intensity peak falls on the *r*, and there is robust voicing – in the case of *rtut'* (lower panel), we expect a non-syllabic *r*: clearly, the duration is much shorter than in the above case, there is less voicing, and the main intensity peak falls on the vowel. This is, of course, only an illustration rather than
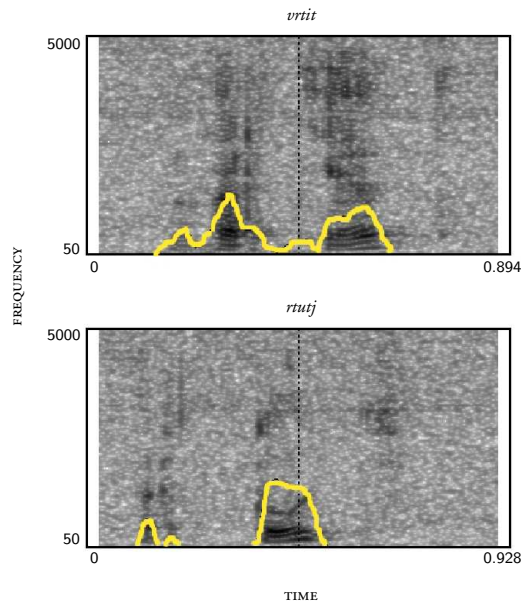
Figure 1: *Syllabic and non-syllabic* r *in Czech*

actual evidence, as the nature of the phonetic correlates of syllabicity is far from clear (Sylvia Blaho p.c.).[2]

Despite the predictable syllabicity we could still claim that the syllabic sonorants differ phonologically from their non-syllabic counterparts if it were not for vowel–zero alternations in the language linking these together in forms of the same lexeme (cf. below).

2.2 Restrictions on word-initial phonotactics

There is agreement in the literature that the presence of non-syllabic #RT clusters in the language is due to a number of diachronic changes which resulted in the loss of certain reduced vowels, the so-called *yers* (Comrie 1987; Sussex & Cubberley 2006). Furthermore, most authors tend to agree that the only synchronic model that can describe the word-initial phonotactics of Czech is one which places no restrictions whatsoever on word-initial clusters (Sawicka 1974; Bethin 1998; Scheer 2004, 2007): anything goes. There are certain flaws in this argument, since, as Scheer (2007: 351) duly notes, *Czech attests 28 combinations of 108 logically possible clusters (6 sonorants, 18 obstruents), which amounts to 26%.*

---

[2]As Martin Krämer points out, we expect a language with trochaic feet to have syllabic sonorants in weak positions, like in English *bottle, button*. Czech has fixed initial stress, which falls on the syllabic sonorant in #TRTV, but on the vowel in #RTV, so a syllabic sonorant would be somewhat surprising there.

Scheer claims that there are no restrictions on word-initial clusters (in Czech and in Slavic languages in general), and the missing combinations are lexical gaps.[3] He goes on to argue that a Slavic language could borrow any kind of #RT word from a phonotactically permissive language like Berber, quoting borrowings of proper names from Caucasian languages to Russian. The fact that 28 combinations are observed in Czech (from the possible 108) results from diachronic changes creating these clusters, but not others.

He quotes further instances of unruly behaviour in word-initial phonotactics: for instance, while #mC clusters are possible in Czech, #nC clusters are absent, despite the fact that *n* is commonly agreed to be a less marked nasal than *m* (a common belief resting on observations such as 'if a language has only one nasal, it is *n*').

2.3   The ratio of #RT and #TR forms

The fact that a large number of combinations do not occur is not the only argument against a categorical *laissez-faire* approach to Czech phonotactics. While I will not embark upon a detailed corpus study, a casual stroll through a dictionary can illustrate that even the 28 attested combinations do not occur in multitudes. Scheer (2006)'s corpus of #RT forms gives four 'archaic' examples, two of which were replaced by forms with no such onset cluster by a native speaker who has been asked to produce these forms (*lsknouti se* 'shine' to *lesknout, mknouti se* 'to get together' to *semknout se*).

The dictionary I consulted (Čermák 1963) lists only one example for most #RT clusters: *rt, lp, lk, lz, rz, rž, md*. For *lž, mž*, it lists three; for *rv*, five. Turning to #TT clusters, only the #sC ones occur in larger numbers, along with *ct, čt* (six, eleven). *tk* appears once, *kt* seven times. In comparison, there are about 80, 50, and 70 entries for *pr, kr, tr*.

Naturally, I do not believe that waving a dictionary or anecdoting about an informant can be substantial arguments against a phonological theory. While a dictionary gives an idea of what the frequent forms are in the language, it does not control for token frequency. More importantly, it only lists stems, which is problematic inasmuch as the intricate vowel–zero alternations of Czech can give a large number of #RT forms, like *jsem* 'be 1SG'. These alternations are even more important as they show that the sonorant in an #RT form and that in a #RVT form (or a #TRT form) is of the same nature. They show different phonetic behaviour perhaps, but they cannot easily be claimed to be different segments.

---

[3]The 28 to 108 ratio is less startling if we take into account the fact that Czech has progressive voicing assimilation, which already excludes a number of possible combinations.

Nonetheless, the claim that there is no restriction on word-initial CC clusters grants identical status to clusters which either rarely show up in the language or do not appear at all, like *kt, pk*, and clusters which are all over the place, such as *kr, tr*.

2.4 Summary

As we have seen, Czech allows a large number of word-initial onset clusters that other languages would not consider well-formed. The claim about this is that Czech has no restrictions on this environment, and the fact that not all possible combinations are attested comes from the whims of language change.

The problem with this argument is that it grants equal status to all kinds of clusters, even though the #TR type is much more numerous than the #RT one (while the existence of the latter is undeniable). A formal approach to this is proposed by Scheer (2007), whose typology predicts two language types: the first, containing languages like English or Italian, (roughly) only allows #TR forms, while the second, containing the Slavic languages, allows any #CC form, including #RT ones as well.

Recall from section 2.2 the proposition that a language like Czech could freely borrow all kinds of #RT forms. This proposition was put under scrutiny by conducting a nonce-word experiment devised to find out whether native speakers have misgivings about some word-initial clusters or indeed accept any #CC cluster word-initially.

3 THE EXPERIMENT

3.1 Methods

I conducted a pilot experiment with seven Czech speakers. The participants were given a list of 40 Czech nonce-words, and had to judge on a scale of 1 to 10 how unfamiliar they sounded: a score of 1 meant that the word looked very Czech for the subject, whereas a score of 10 meant it looked completely alien. The list contained six #TR forms, seventeen #TT forms, and eleven #RT forms, along with four #CCC forms, and two inflected Berber verbs (*txznt* 'you.Sɢ.stored' and *bddl* 'exchange.Iᴍᴘ').

The methodology had its weaknesses. Some forms, for instance, were very similar to existing forms (some ending in *t* could be interpreted as infinitives), and the group of informants was not selected too carefully. Two of the subjects marked the majority of forms with 10, even those with a TR

	Good	Don't Know	Bad
TR	93%	0%	7%
TT	76%	6%	18%
RT	63%	10%	27%
TTT	0%	0%	100%
Berber	0%	0%	100%

Table 1: *Acceptability and phonotactic shape*

onset. These subjects were excluded from the analysis, giving five speakers and five judgment tokens for each type.

3.2 Results

Table 1 shows acceptibility judgments about forms with structurally different word-initial clusters. The form was regarded as 'good' if the subject gave it a mark between 1–4, 'bad' between 6–10, while 5 was considered as 'don't know'.

There is a correlation between the judgments and the structure of the cluster. While tokens with frequent #TR clusters were almost all accepted, the lower the type-frequency of a cluster is (and the less it conforms to generalizations about sonority), the more reluctant the subjects become to accept it: the proportion of 'good' judgments drops from 93 to 63 per cent from #TR to #RT. Furthermore, negative judgments raise from 7 per cent to 27. The #CCC forms, along with the Berber words, were rejected.

It is possible to say, then, that #TR forms still seem more natural to the subjects, which contradicts any claims which equate these with #RT or #TT. On the other hand, looking at judgments based on whether the cluster in question occurs in base forms in the language or not, one does not see vastly broad differences (Table 2).

Clusters supported by existing base forms do score better, but are not accepted categorically; moreover, a large number of factors were ignored: clusters occurring once in a base form (like *rt* as in *rtutj'*) were counted

Occurs	Good	Don't Know	Bad
Yes	70%	8%	22%
No	57%	10%	30%

Table 2: *Acceptability and occurrence*

as existent, while others which could show up as the result of vowel–zero alternations were counted as non-existent.

3.3 Discussion

While the scope of the experiment does not allow it to cover all the nooks and crannies of Czech word-initial phonotactics, it does provide some interesting evidence against categorical approaches to phonotactics, which claim that Czech allows any possible #CC cluster, including #RT forms, just because it seems to allow some. The results show a correlation between acceptability and phonotactic shape, as well as frequency of occurrence: frequent and 'usual' #TR forms were more acceptable to the subjects than the less frequent #RT forms.

4   GENERAL DISCUSSION

Since the #CC clusters attested in Czech evade any attempt to round them up in a natural class, categorical approaches to phonotactic competence can only account for them by predicting that all combinations of word-initial consonants form valid clusters in the language. There are two problems with such a predicton: Firstly, a large number of these combinations do not occur in the language at all (and some of them only occur marginally), and secondly, my nonce-word experiment hints at the possibility that native speakers themselves show preferences for certain combinations over others.

A similar nonce-word experiment in English (Greenberg & Jenkins 1964) showed a gradient scale of phonotactic well-formedness, a scale I believe to be present in Czech. These observations led some authors to suggest that a view of phonotactics which boils down to acceptable versus unacceptable is crudely simplified, and that the role of the general similarity to existing forms of the language is also crucial (Bybee 2001).

We can pursue this approach and claim that, indeed, the only thing that is crucial is similarity to existing forms. No doubt, there is a reason why these forms are usually of the #TR kind: tokens of *stop + liquid* onsets are more perceivable and easier to produce, promoting their long-term survival during language change. At the same time, however, language change can provide unusual word-initial cluster types as well, but we should take guard against hugely extending our grammar to include a small number of such forms.

REFERENCES

Bethin, C. Y. (1998). *Slavic Prosody*. Cambridge University Press, Cambridge.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge University Press, Cambridge.

Čermák, A. (1963). *Czech-English English-Czech Dictionary*. Saphograph, New York.

Comrie, B. (ed.) (1987). *The Major Languages of Eastern Europe*. London: Routledge.

Greenberg, J. & J. J. Jenkins (1964). Studies in the psychological correlates of the sound system of Amerian English. *Word* 20, pp. 157–177.

Harris, J. (2005). The phonology of being understood: Further arguments against sonority. *Lingua* 116, pp. 1483–1494.

Sawicka, I. (1974). *Struktura grup spółgłoskowych w językach słowiańskich*. Ossolineum, Wrocław.

Scheer, T. (2001). Čeština ve waršawě. Handout of a class given on Czech in Warsaw.

Scheer, T. (2004). *A lateral theory of phonology. Vol 1: What is CVCV, and why should it be?* Mouton de Gruyter, Berlin.

Scheer, T. (2006). List of words with initial Sonorant-Obstruent clusters in 13 Slavic languages. Available at http://www.unice.fr/dsl/tobias.htm.

Scheer, T. (2007). On the status of word-initial clusters in Slavic (and elsewhere). Compton, R., M. Goledzinowska & U. Savchenko (eds.), *Annual Workshop on Formal Approaches to Slavic Linguistics. The Toronto Meeting 2006*, Ann Arbor: Michigan Slavic Publications, pp. 346–364.

Selkirk, E. (1984). On the major class features and syllable theory. Aronoff, M. & R. T. Oerhle (eds.), *Language Sound Structure: Studies in Phonology Dedicated to Morris Halle by his Teacher and Students*, MIT Press, Cambridge, MA.

Sussex, R. & P. Cubberley (2006). *The Slavic Languages*. Cambridge University Press, Cambridge.

Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie*. Vandenhoeck & Ruprecht, Göttingen.