Similar Place Avoidance in Slavic and Other Languages

Aleš Bičan

Abstract: The paper discusses a constraint on the distribution of homorganic CVC sequences known as Similar Place Avoidance (SPA). Though proposed as a statistical universal, it has been little considered in Slavic and other Indo-European languages. We evaluate the CVC distribution in 100 recorded and reconstructed varieties, of which 18 are Slavic, 44 are non-Slavic Indo-European, and 38 are non-Indo-European. The SPA principle has been formulated as pertaining to CVC sequences of two consonants sharing the same place, but it has also been suggested that coronals are dependent on sonorancy agreement for the constraint to take effect. This dependency is indeed observable but concerns dento-alveolars only, not coronals as a whole class. SPA weakly restricts combinations of dento-alveolar sonorants with palatal sonorants. Combinations of different-place coronal obstruents are disfavored, but this is instead due to sibilancy avoidance (a restriction of the co-occurrence of two sibilants in a CVC sequence, previously unreported). Finally, combinations of palatals (including post-alveolars) are less often subject to an SPA effect, and the Slavic languages virtually lack this kind of restriction.

Keywords: Obligatory Contour Principle, phonotactics, palatal, sonorant, coronal, sibilant, labial-coronal effect

1. Introduction*

One of the best-known phonotactic constraints operative in spoken languages is the avoidance of repetitions of similar consonants separated by a nuclear segment or appearing in other kinds of adjacency. This avoidance is at the core of several principles known as Obligatory Contour Principle (McCarthy 1986), Similarity Avoidance (Frisch et al. 2004), Identity Avoidance (Tang 2000), Repetition Avoidance (Walter 2007), and Similar Place Avoidance (Pozdniakov and Segerer 2007). The constraint primarily pertains to the place of articulation. However, other consonantal properties, such as the manner of articulation

* The paper has profited enormously from two anonymous inspirational and insightful reviews, especially from the second 22-page review. Needless to say, all remaining errors are our own.

or the state of the glottis, may contribute to the same-place consonant co-occurrence restriction, or the non-place features may have their own avoidance restrictions (MacEastern 1999; Pozdniakov 2010; see also Gallagher 2015 and references therein). There is also some evidence that vowel co-occurrence is restricted in a similar manner (Walter 2010; Doucette et al. 2024).

Since Cantineau's (1946) and Greenberg's (1950) pioneering studies of Classical Arabic, where the dislike for similar consonant repetition was first noted, the constraint has been observed in nearly every language studied. To date, the largest and most comprehensive treatment of the phenomenon is Pozdniakov and Segerer's seminal work (2007; henceforth P&S). Analyzing CVC sequences in large lexical databases for 31 languages from several families, they show that although these languages allow repetitions of consonants belonging to four basic place classes of labials, dentals/alveolars, palatals, and velars, the frequency of these homorganic combinations is much smaller than one would expect from the occurrence of the particular consonants. Their data also revealed that a comparable, though less strong, kind of avoidance may apply to consonants belonging to two place superclasses sharing some similarity, one grouping together labials and velars ("peripherals" from the perspective of the oral cavity), the other uniting dentals/alveolars and palatals ("medials"). These statistical cross-linguistic tendencies have become known as Similar Place Avoidance (SPA).[1]

Previously and subsequently, other studies have confirmed the cross-linguistic avoidance of homorganic CVC combinations of labials and velars, but the situation with "medial" or coronal consonants is more complicated. Researchers disagree on whether the avoidance pertains to coronals as a whole class (Cooper 2009; Mayer et al. 2010) or to dento-alveolars and palatals separately (P&S; Nikitina 2022), or whether palatals should be included in the class of coronals at all (Rousset 2004; Kinney 2005). The question of whether palatals pattern with dento-alveolars or with velars (the latter claimed by Padgett 1992 for Russian) is also unclear. In any manner of classification, however, the avoidance effect for "medial" place classes has turned out to be weaker than for the peripheral classes, or the effect manifests itself only in conjunction with non-place features, such as sonorancy, stricture/continuancy, or voicing (Pierrehumbert 1993; McCarthy 1994; Coetzee and Pater 2008).

Despite the uncertainty concerning the most appropriate formulation of the SPA principle, the fact that languages or their users tend to disfavor repetitions of similar consonants is undeniable. Besides the 31 languages in P&S's study, most of which are from Africa, the effect has been documented or noted

[1] As already pointed out, some writers subsume this phenomenon under other names and/or principles such as Obligatory Contour Principle (OCP, or OCP-Place when it concerns the place of articulation). We will speak about SPA irrespective of how the cited works label the phenomenon.

to exist in at least 60 other languages from Afro-Asiatic (including Semitic), Austronesian, Indo-European (IE), and other families, including creoles (see the references throughout this paper). The effect has further been confirmed for as many as 3,200 languages in Mayer et al. (2010), but this otherwise formidable study must be treated with some caution because the authors examined the limited vocabulary of the Swadesh list. SPA has also been considered in studies on the evolution of speech from babbling through first words to adult languages, especially in relation to the preference for Labial-Vowel-Coronal sequences (the Labial-Coronal effect; MacNeilage and Davis 1999, 2000; MacNeilage et al. 2000; Rousset 2004; Kinney 2005; Vallée et al. 2009; Carrissimo-Bertola 2010).

Support for the SPA principle is also found in areas other than statistical analyses of morpheme or word lists. Experiments have demonstrated that nonwords violating SPA are judged less well-formed by speakers of Hebrew (Berent and Shimron 1997), Arabic (Frisch and Zawaydeh 2001), and English (Davis 1991; Frisch 1997; Coetzee 2005). Other experiments indicate that SPA-violating sequences help Dutch speakers identify a word boundary or spot a word's presence in artificial utterances (Boll-Avetisyan and Kager 2014). There is, furthermore, evidence that loanwords resist adaptation to the recipient language when violating SPA (Frisch et al. 2004 on Maltese). Further support comes from Hulden (2017), who reports that an algorithm created for automatically detecting and separating vowels and consonants in a running text has proven to be more successful when SPA distributional restrictions have been implemented into the algorithm. Finally, some studies mention that SPA also limits consonant combinability in English consonant clusters (Pierrehumbert 1994). Explanations of why SPA is such a robust distributional restriction are usually grounded in language perception and processing, and confusability avoidance (Pierrehumbert 1993; MacNeilage et al. 2000; Frisch et al. 2004; Frisch 2004; Graff 2012; Boll-Avetisyan and Kager 2014; cf. also Wright 2004 and Zukoff 2017).

The great attention SPA has received suggests that the validity of this effect is firmly established, and any further confirmation is merely a matter of empirical support. However, as already mentioned, several unresolved theoretical issues remain concerning how the place classes subject to SPA restrictions are defined, and to what extent non-place features as well as segment distance, prosody, or morphology influence the effect. Since space prevents us from addressing all these issues, we will only pay attention to the role of sonorancy in SPA restrictions and to the way the class of coronals is delimited and subdivided. Also, we will look into the nature of palatals, which happen to be resistant to SPA in many languages. These problems will be subsumed under the main goal of our paper, which is to provide a comprehensive survey of the SPA effect using large lexical datasets for Slavic, IE, and other languages.

Many, though not all, previous SPA studies are potentially deficient in drawing generalizations about entire language families or subbranches without having investigated a sufficiently large sample. The sampling may lead to incorrect conclusions if, for instance, Russian and English are selected as representatives of Slavic and Germanic languages, respectively. Therefore, we will conduct an analysis of one language class (Slavic) supplemented by a similarly extensive treatment of its superordinate family (IE languages). The 38 non-IE languages widen the perspective but are not in the main focus. We have chosen Slavic to compensate for the neglect this family has received in SPA studies. To the best of our knowledge, only Russian has been analyzed from the perspective of SPA (Padgett 1992, 1995; Pozdniakov 2007, 2010, 2011). Croatian has been considered from a related perspective of segment repetition in general (MacKay 1970).

The rest of the paper is organized as follows. In the next section, we introduce our Slavic and non-Slavic material (§2.1, §2.2), explaining how it is phonologically classified (§2.3) and statistically evaluated (§2.4). Section 3 discusses the strength and diffusion of SPA in Slavic; the next section does the same for non-Slavic IE (§4.1) and non-IE languages (§4.2). Since many languages fail to avoid palatal homorganic CVC sequences, we will present a typology of palatal systems to find correlations between the nature of SPA and the nature of these systems. Section 5 is dedicated to the problem of coronals. One of the major findings of this paper is that two separate classes of coronals should be distinguished for the SPA effect to be adequately accounted for, namely dento-alveolars on the one hand and post-alveolars together with (alveolo-) palatals on the other. The previously described distributional restrictions on coronals are shown to be properties of dento-alveolars rather than coronals as a whole (§5.1). We also discuss the role of sonorancy agreement in SPA restrictions, showing that it is only dento-alveolars that strongly depend on this agreement, whereas palatals as well as labials and velars do not (§5.2). However, there is one cross-linguistically observable restriction on the combinability of coronal obstruents pertaining to sibilants, and this sibilancy avoidance is another major and previously unreported finding discussed in §5.3. Finally, §5.4 briefly addresses the problem of the labial-coronal effect and other combinations that are *favored* cross-linguistically instead of being avoided. In the final section of this paper, we evaluate various models of SPA, proposing our formulation, and briefly discuss the etymological nature of native Slavic roots with homorganic CVC sequences violating SPA.

2. Data and Methodology

It is generally agreed that SPA is a constraint on morpheme (root) structure (Booij 2011), and that SPA violations across morpheme boundaries are often more readily tolerated (P&S). However, while word lists are easily obtainable for many languages, morpheme lists are either too small or simply lacking. Our analysis relies on word lists but considers only those C_1VC_2 sequences in which the V element is the first syllabic nucleus of a word, namely $\#C_1VC_2$, $\#CC_1VC_2$, $\#CCC_1VC_2$, and more complex patterns, irrespective of whether C_2 belongs to the first or second syllable (# is a word boundary). Henceforth, we will just speak about CVC or C_1VC_2 sequences. The choice of the first syllable has been motivated by the fact that these syllables are less likely to contain a morpheme boundary, at least in Slavic and other IE languages, which prefer suffixation over prefixation. In this section, we describe our Slavic (§2.1), non-Slavic Indo-European, and non-IE language data (§2.2), the way phonemes of these languages are classified (§2.3), and the way CVC distribution is statistically evaluated with Yule's Q (§2.4).

2.1. The Slavic Material

Our analysis of the SPA effect rests on the Slavic vocabulary included in the indices of the etymological dictionary of the Old Church Slavonic (OCS) language, *Etymologický slovník jazyka staroslověnského* (Havlová et al. 1989–2022; henceforth *ESJS*), and on the OCS headwords in the dictionary. The dictionary painstakingly lists all known cognates of OCS words in every Slavic language, thanks to which the indices are, with some reservations to be mentioned presently, collections of etymologically interrelated vocabularies. *ESJS* also offers reconstructions of Proto-Slavic roots underlying the OCS lexemes, but the indices provide many more Proto-Slavic reconstructions than OCS headwords. This is because the dictionary evaluates all major etymological explanations, some of which are highly conjectural. Since we want to cover the Proto-Slavic stage, we have decided to rely instead on the list of Proto-Slavic reconstructions (headwords) from Derksen's (2008) dictionary.

Though largely etymologically related, the *ESJS* vocabularies are not equivalent. First, not all lexemes of all Slavic languages with a reconstructable Proto-Slavic etymon have an attested OCS cognate. Furthermore, the indices contain some other Slavic words mentioned in support of the etymological explanations of the OCS words, which are not necessarily cognates or whose relation to a particular OCS lexeme is disputed. Moreover, *ESJS* explains all OCS vocabulary, including loanwords. Their equivalents are selectively mentioned for the other Slavic languages and hence recorded in the indices. Finally, the wordlists consist of basic (lemma) forms, though occasionally orthographic and other variants are also included. Derivatives are not mentioned in the dictionary and its indices unless they are etymologically interesting.

Simply put, the *ESJS* indices provide a sample of Slavic cognate vocabulary that is fairly exhaustive regarding the languages covered and reasonably large for each of the languages considered. Besides Old Church

Slavonic and Proto-Slavic, the languages are Bulgarian, Macedonian, Serbo-Croatian, and Slovenian for the South branch; Slovak, Old Czech, Modern Czech, Upper Sorbian, Lower Sorbian, Polabian, Pomeranian, and Polish for the West branch; and Belarusian, Ukrainian, Old Russian, and Modern Russian for the East branch. The Pomeranian data omit Kashubian. Slovincian is also excluded because its vocabulary listed in *ESJS* is too small. For the same reason, we leave out all words from older or immediate stages of the Slavic languages except those of Old Russian and Old Czech, which are sufficiently represented in the dictionary. Polabian has the smallest lexical dataset (745 word-initial CVC sequences), and Modern Russian has the largest one (4,259 sequences). On average, there are 2,195 word-initial CVC sequences for every variety.

All the wordlists were transformed into phonological transcription by simple grapheme-to-phoneme substitutions. Though this must have produced some errors, the procedure can be regarded as essentially reliable thanks to the phonological spelling of the languages. Additionally, we will refer to other lexical databases of some Slavic languages that have been processed similarly or have a phonetic/phonological transcription.

2.2. The Non-Slavic IE and Non-IE Material

To enrich our analysis, we have collected phonologically or phonetically transcribed lexical material for 44 non-Slavic IE languages and 38 non-IE languages from various lexical sources.[2] Preference has been given to databases that include at least 1,000 lexical items and provide phonetic or phonological transcription for these items. When such a source is unavailable or unknown to us, we have chosen languages where the phonological form could be reliably inferred from the orthography (e.g., Lithuanian). The databases differ in size (ranging from 859 CVC sequences for Chukchi to 282,512 sequences for Norwegian; 26,960 sequences on average), contents (lemmas only or inflected forms included; loanwords included or excluded), detail (broad or narrow transcription), perspective (phonetic or phonological), and last but not least, reliability (probability of transcription errors). Thus, the data must be treated with some caution. The non-Slavic IE sample also includes a list of Proto-Indo-European reconstructions and some other reconstructed varieties.

While we have aimed for exhaustivity in the case of the IE languages, the selection of the non-IE languages has been motivated by data availability, their size, and the desire to cover as many families as possible. Consequently, the non-IE sample is not representative because some families are instantiated by

---

[2] The reader will find the list of sources for all the lexical data as well as the statistical evaluation of the data at https://ojs.ung.si/index.php/JSL/article/view/469/345. The datasets themselves are not provided because they are either publicly available or protected by copyright.

more than one language (especially the Austroasiatic family), while other families contain only one (e.g., the Dravidian family) or no representative (e.g., the Tupian family). The primary purpose of the non-IE languages is to illustrate the nature of SPA outside the Indo-European family.

2.3. Phonological Classification

We will describe the distribution of the non-syllabic consonants in CVC sequences/combinations (henceforth just sequences/combinations unless specified otherwise). The nuclear V segment includes syllabic consonants. Long and nasalized vowels are counted as singular Vs. Diphthongs are generally interpreted as single Vs (always in the Slavic languages), but they are sometimes treated as two-phoneme sequences depending on the particular database used or the nature of the diphthongs in a given language (e.g., in Danish; cf. Basbøll 2005). Stress, tone, and other prosodic features are ignored for the sake of simplicity, though they may affect the strength of SPA in some languages (see Frisch 1997 and Dmitrieva 2008 on English, and van Goch 2010 on Dutch).

The non-syllabic consonants (i.e., the Cs; henceforth just consonants) are divided into four major place classes for which we will use the labels labials, dentals, palatals, and dorsals, symbolized with letters P, T, Č, and K, respectively (we distinguish Č for palatals from C for consonants in general). Following P&S, we unite dentals and palatals on the one hand and labials and dorsals on the other. P&S call these superclasses medials and peripherals, respectively, but we prefer the more traditional labels coronals and non-coronals (see §5 for a further discussion of coronals).

The labial class consists of the consonants in which at least one lip is the primary moving articulator, including the labio-velar approximant /w/, which agrees with the classification of P&S and Mayer et al. (2010) as well as with the findings of Kumagai (2020). The dorsal (short for dorso-guttural) class gathers the consonants articulated in the back of the oral cavity or in the throat: velars, uvulars, pharyngeals, epiglottals, and glottals. In some languages (such as the Semitic ones; see, e.g., McCarthy 1994), the throat consonants may behave as a separate class in the consonantal distribution. Since the Slavic languages virtually lack these segments, we refrain from positing a separate place class here. The velar nasal [ŋ] is classified as an allophone of the dental /n/ before velars in the Slavic languages. In the non-Slavic ones, it is usually treated as a separate dorsal segment (depending on the language or the database analyzed).

The palatal class comprises consonants produced with a raised tongue middle in a post-alveolar or palatal articulatory region (Ladefoged and Maddieson 1996: 14–15), namely, post-alveolars like /ʃ/, alveolo-palatals like /ɕ/, and true palatals like /c/, /ɟ/, /ɲ/, /j/ and like /ç/, /ʝ/ (Hall 1997 separates the latter two from the former four; cf. also Rocasens 1990). We discuss palatal systems in more detail in §4.1. Finally, the dental (short for dento-alveolar) class contains (inter)dentals, (pre-)alveolars, and retroflexes.[3] Retracted sibilants are also classed here. The classification of retroflexes requires a few notes. Although retroflexion is often understood as a separate place of articulation (which is reflected in the IPA chart), some linguists view it as an articulatory gesture (Hamann 2002). The exact target of this gesture ("place of articulation") varies across languages and speakers but tends to be in the post-alveolar region. It would thus make sense to classify retroflexes as palatals. However, unlike post-alveolars, alveolo-palatals, and true palatals, retroflexes do not have a raised (domed) tongue middle. Moreover, retroflexes may be interpreted as allophones of dentals/alveolars in some languages, such as Norwegian, Swedish, or Kukatja. We, therefore, place retroflexes into the dental class.

It has been argued that Polish, Russian, Lower Sorbian, and Serbo-Croatian also contain retroflex consonants (Zygis 2003), namely those affricates and fricatives that are usually transcribed in Slavic literature with the symbols /č/, /š/, /ž/ or /t͡ʃ/, /ʃ/, /ʒ/ and that are generally described as post-alveolar sibilants. Articulatorily and perceptually, these segments differ slightly from "true" retroflexes, for example, in Indo-Aryan languages. We will demonstrate in §5.3 that their distribution parallels the distribution of other palatals in Polish. Consequently, the Slavic retroflex/post-alveolar sibilants are included in the palatal class.

Another problem concerns the classification of palatalized consonants (such as /sʲ/ in Russian) and other consonants with secondary articulation. Since secondary articulation is often contextually determined (allophonic), the standard treatment is to group them with their non-palatalized counterparts (see Padgett 1992 and Pozdniakov 2007 on Russian and Pierrehumbert 1993 on Arabic). The same approach will be followed here, though the problem needs to be considered in more detail.

In the rest of the paper, the terms labial, dental, palatal, dorsal, coronal, and non-coronal stand for the phonological classes just defined. They

---

[3] In contrast to P&S, we place /s/ as well as /z/, /t͡s/, and /d͡z/ among dentals, not palatals (P&S: 312, fn 4; see also Pozdniakov 2007). P&S support this choice by stating that the /s/ fricative often appears in the /ʃ/ slot in African languages, further noting that "[the reclassification of /s/ as a dental] would have slightly changed the figures, but not the tendencies [for avoidance]". However, elsewhere (p. 341) they admit that the classification of /s/ is the most problematic one, which might have been a source of some discrepancy. Note also that Mayer et al. (2010) classify /ʃ/ and /ʒ/ as dentals but /t͡ʃ/ and /d͡ʒ/ as palatals without any explanation. For P&S and us, all these four consonants are palatals. Another case of disagreement concerns liquids; Mayer et al. (2010) treat /r/ and /l/ as palatals, whereas P&S treat them as dentals. The latter is what we also do (together with most other studies).

are written with small capitals, while regular letters indicate that a more traditional phonetic meaning of these terms is implied or that we refer to other phonological conceptions (the latter especially in the case of coronals). Thus, when we state that some language's PALATAL class contains true palatal obstruents like /c/, /ɟ/ and post-alveolar sibilants like /ʃ/, /ʒ/, we mean that the consonants produced at the two articulatory targets (the hard palate and the back of the alveolar ridge) are grouped into one SPA identity class.

Following Yip (1989), we understand an SPA identity place class as a well-defined set of consonants for which the following hypotheses are assumed.

(1) Hypotheses concerning SPA identity classes

a. CVC sequences consisting of consonants from the same identity class (henceforth: homorganic CVC sequences) are statistically underrepresented (avoided).

b. CVC sequences consisting of consonants from different identity classes (henceforth: heterorganic CVC sequences) are not statistically underrepresented; they are either overrepresented (favored) or show no statistically significant distributional anomalies.

The corroboration and/or rejection of the hypotheses in (1) depends on how identity classes are defined, to which problem we will return in §5. Another question is how statistical underrepresentation is determined.

2.4. Statistical Evaluation

Linguists have put forth several ways to determine whether the distribution of some elements differs from some norm, and whether this difference is not just a random distortion in the data population. One standard measure is a comparison of the observed frequency (O) and the expected frequency (E). The former is the actual frequency of some combination of elements, whereas the latter is a theoretical frequency the combination would have had if it were derived from the frequencies of the elements in the combination. An easy comparison of O and E is obtained by dividing these two frequencies, and this simple method has been used in most SPA studies, at least since Pierrehumbert (1993), though it was already employed in other analyses (e.g., Janson 1986). An O/E ratio greater than one means overrepresentation (a combination occurs more frequently than expected); a ratio of less than one implies underrepresentation (it occurs less frequently than expected).

Since it is unlikely that the observed frequency of a combination would equal its expected frequency in real data, the O/E ratio always indicates that a combination occurs more or less often than expected. We must have criteria for deciding what discrepancy from the equation is statistically significant and linguistically interesting. This can be determined by statistical measures such as the χ^2 test, often used in this connection, or we can set up some arbitrary and reasonable limits. The latter was done by P&S, who regarded the situation when O was different from E by at least 15% as interesting; an O/E ratio below this threshold was interpreted as falling within the range of natural distortions in the data, that is, as no evidence for overrepresentation or underrepresentation.

Although the O/E ratio is a simple and intuitively interpretable test, it has been subject to criticism. For example, Wilson and Obdeyn (2009) demonstrated that the measure was mathematically flawed because it confounded co-occurrence restrictions with positional probabilities. They argued for an alternative loglinear model, but it was proven to suffer from the same weaknesses (Stanton and Stanton 2022). Other, linguistically more profound arguments against the O/E ratio are laid out in Grotberg's (2022) dissertation, which discusses several statistical measures used for the analysis and evaluation of SPA.

Most previous studies have shown that CORONAL (DENTAL) homorganic CVC sequences are less often underrepresented than LABIAL or DORSAL ones, or that their underrepresentation effect is less strong than that for LABIALS and DORSALS. At the same time, it is well known that CORONALS are present in nearly every phoneme inventory (Maddieson 1984). Our material confirms this omnipresence and further testifies that the average frequency of nearly all combinations containing a DENTAL is greater than the average frequency of any other combination type. Now, Grotberg (2022: 54ff.) demonstrates that the O/E test is sensitive to this skewed distribution of CORONALS as opposed to the NON-CORONALS so that it underestimates the amount and force of avoidance among CORONALS (see the next paragraph). To ward off this bias, Grotberg proposes another statistic known as Yule's Q or Yule's coefficient of association, which is common in psychological or sociological studies, and which does not suffer from the aforementioned dependence.

Yule's Q belongs to the odds ratio family that quantifies the strength of association between two events, which for us is the occurrence of a consonant C_1 with a consonant C_2 in a C_1VC_2 sequence. Both tests are calculated from 2×2 contingency tables such as Table 1 (on the following page), with four values: A, B, C, and D (the shaded area). Should we want to calculate Yule's Q for combinations of two consonants C_1 and C_2 (e.g., LABIALS co-occurring with other LABIALS in a CVC sequence), then value A is the observed frequency of these combinations. Value B is the observed frequency of C_1 occurring with all other consonants but C_2 (LABIALS with NON-LABIALS). Value C is the frequency of the reverse combinations, that is, of all consonants other than C_1 occurring with C_2 (NON-LABIALS with LABIALS). Finally, value D corresponds to the frequency of all remaining combinations (NON-LABIALS with NON-LABIALS). The sum of A, B, C, and D equals the frequency of all CVC combinations (Total). The sums of the particular rows and columns are marginal totals (e.g., the total frequency of LABIALS). Since DENTALS are so frequent cross-linguistically, the marginal totals for any combination involving a DENTAL will be greater than the other marginal totals. The O/E ratio and the χ^2 test are sensitive to this discrepancy, but the odds ratio and Yule's Q are not.

Table 1. A 2 × 2 contingency table for the calculation of Yule's Q of a C_1VC_2 combination

	C_2	**$\neg C_2$**	**Marginal totals**
C_1	A	B	A + B
$\neg C_1$	C	D	C + D
Marginal totals	A + C	B + D	Total (A + B + C + D)

The odds ratio (OR) is the ratio between two odds, A to B and C to D, that is, the ratio of the frequency of C_1 occurring with C_2 to the frequency of C_1 occurring with consonants other than C_2, and of the frequency of consonants other than C_1 occurring with C_2. Two equivalent formulas for calculating OR are given under (2a). Yule's Q is a transformation of OR that can be obtained either from OR or directly from values A, B, C, and D (2b).

(2) a. Odds Ratio b. Yule's Q

$$OR = \frac{\frac{A}{B}}{\frac{C}{D}} = \frac{AD}{BC} \qquad Q = \frac{OR-1}{OR+1} = \frac{AD-BC}{AD+BC}$$

The range of possible results is zero to infinity for OR and −1 to +1 for Yule's Q. It is this finite and symmetrical range that makes Yule's Q a more helpful tool for expressing the strength of association between two consonants (O/E also ranges from zero to infinity). When there is no association, the difference between the frequency of C_1VC_2 (LABIALS with LABIALS) multiplied by the frequency of $\neg C_1V\neg C_2$ (NON-LABIALS with NON-LABIALS) equals the frequency of $C_1V\neg C_2$ (LABIALS with NON-LABIALS) multiplied by the frequency of $\neg C_1VC_2$ (NON-LABIALS with LABIALS). It means that C_1 is equally associated with C_2 and all other consonants. In actual data, this zero Yule's Q is probably only a theoretical possibility; there will always be some deviation from zero in either direction.

Positive Yule's Q expresses the strength of association between C_1 and C_2, whereas negative Q expresses the strength of disassociation. In the first case,

C_1 is more likely to occur with C_2 than other consonants; in the latter, it is less likely. If C_1 occurs with C_2 (i.e., A is non-zero) but does not combine with other consonants (i.e., either B or C or both are zero), and at the same time there exists at least one combination of $\neg C_1$ and $\neg C_2$ (D is non-zero), Yule's Q equals 1. This is the complete association of C_1 with C_2. On the other hand, if C_1 does not combine with C_2 (A is zero) but combines with at least one other consonant (B or C or both are non-zero), and at the same time there are combinations of $\neg C_1$ and $\neg C_2$ (D is non-zero), Yule's Q is −1. It is the complete disassociation of C_1 with C_2 or the absence of C_1VC_2 combinations.[4]

Again, in real frequencies of CVC sequences, it is rare to obtain the values of +1 and −1, though both cases appear in our data, the latter being more frequent. For example, Otomi has one combination of a PALATAL obstruent with another PALATAL obstruent but no combination of a PALATAL obstruent with any other consonant; hence Q = 1. Afrikaans has no combination of two PALATALS, but PALATALS combine with other consonants; hence Q = −1. However, in most other cases, Q is a non-zero value between these two extremes. Table 2 on the following page gives an example of a positive Q calculated for CVC combinations of DENTALS with LABIALS in Polish. Table 3 on the following page is an example of a negative Q for combinations of DENTALS with DENTALS in the same language. The former value speaks for the association between DENTALS and LABIALS, and the latter speaks for the lack of association between DENTALS and DENTALS.

Since it is practically impossible to determine Yule's Q for all possible events (for all CVC sequences in all words), we do not know the true value of the strength of association between two variables. We can only determine an estimate of Yule's Q based on a particular sample. These estimates will vary from sample to sample, but the Q values themselves do not tell us how good they are. One way to evaluate the reliability of an estimate is to determine its confidence interval (CI). A 95% CI means that, were we to calculate Yule's Q for a theoretically infinite number of other samples, 95% of those samples would be expected to have a value in the reported range. It is thus highly probable that the true value of Yule's Q lies within the range reported as the 95% confidence interval.[5]

[4] If both A and B or both A and C are zero, the Yule's Q formula fails because we divide by zero. The standard procedure is to add 0.5 or even 0.1 to each value (Hollander et al. 2014: 512). Although this practice produces a meaningful Q value, it shadows the real linguistic significance of the underlying CVC distribution. A zero value means that either C_1 does not combine with any consonant or that C_2 does not combine with any other consonant, or both. These situations cannot be evidence for SPA because languages of this kind restrict the occurrence of both homorganic and heterorganic CVC sequences involving a given consonant.

[5] Yule's Q is essentially a normal distribution (Grotberg 2022: 94, fn 4). The lower and upper limits of CI are calculated from the lower and upper limits for Odds Ratio (OR),

Table 2. Calculation of Yule's Q and its confidence interval for combinations of DENTALS with DORSALS in Polish

		C_2	
		DORSAL	NON-DORSAL
C_1	DENTAL	83	566
	NON-DENTAL	191	1660

$$Q = \frac{83\times1660 - 566\times191}{83\times1660 + 566\times191} = 0.12$$

95% CI = [–0.02, 0.25]

Table 3. Calculation of Yule's Q and its confidence interval for combinations of DENTALS with DENTALS in Polish

		C_2	
		DENTAL	NON-DENTAL
C_1	DENTAL	362	533
	NON-DENTAL	759	846

$$Q = \frac{362\times846 - 533\times759}{362\times846 + 533\times759} = -0.14$$

95% CI = [–0.22, –0.06]

θL and θU, respectively (Hollander et al. 2014: 521):

$$Q_{Lower} = \frac{\theta L - 1}{\theta L + 1}, \text{ where } \theta_L = e^{\ln(OR) - Z \times SEln(OR)}$$

$$Q_{Upper} = \frac{\theta U - 1}{\theta U + 1}, \text{ where } \theta_U = e^{\ln(OR) + Z \times SEln(OR)}$$

The Z value defines the appropriate confidence limits, such as 2.58 for the 99% CI or 1.96 for the 95% CI. The constant *e* is Euler's number. The $SE_{ln(OR)}$ value is the standard error calculated from the following formula in which A, B, C, and D are the frequency values in a contingency table (see Table 1 on p. 185):

$$SE_{ln(OR)} = \sqrt{\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}}$$

Thus, the 95% CI upper and lower values for Table 2 are calculated as follows:

$$SE_{ln(OR)} = \sqrt{\frac{1}{83} + \frac{1}{566} + \frac{1}{191} + \frac{1}{1660}} = 0.14 \qquad OR = \frac{AD}{BC} = \frac{83\times1660}{566\times191} = 1.27$$

$$\theta_L = e^{\ln(OR) - Z \times SEln(OR)} = e^{\ln(1.27) - 1.96 \times 0.14} = 0.97, \text{ hence } Q_{Lower} = \frac{\theta L - 1}{\theta L + 1} = \frac{0.97-1}{0.97+1} = -0.02$$

$$\theta_U = e^{\ln(OR) + Z \times SEln(OR)} = e^{\ln(1.27) + 1.96 \times 0.14} = 1.68, \text{ hence } Q_{Upper} = \frac{\theta U - 1}{\theta U + 1} = \frac{1.68-1}{1.68+1} = 0.25$$

The wider the CI, the greater the chance that it overlaps zero and includes both positive and negative values, which means that neither genuine overrepresentation (expressed by positive values) nor genuine underrepresentation (negative values) can be excluded. Returning to tables 2 and 3, we see that both Q values for Polish (0.12 and −0.14, respectively) are almost equally distant from zero (the strength of association and disassociation is very similar for both combination types), but we also witness that the CI for the Q value for the combinations of DENTALS with DORSALS includes zero, while the CI for the combinations of two DENTALS does not. Consequently, we lack reliable evidence for an association between DENTALS and DORSALS in this language. On the other hand, we can claim with some confidence that combinations of DENTALS with other DENTALS are avoided there.

The CI width is not only dependent on our confidence level but also on the size of our sample and the distribution of certain CVC combinations. Small samples will produce wider CIs than larger samples at the same confidence level. Similarly, CVC sequences containing very infrequent consonants will likely produce greater CIs. Considering our Slavic data, we have decided to work with 95% CIs. In Slovenian, for example, wider intervals make the Yule's Q value for DORSAL homorganic sequences insignificant, but the value is significant for a more extensive alternative dataset with the same CI.

Let us return to Yule's Q proper. The more it deviates from zero, the greater the (dis)association there is between C_1 and C_2. Following Rosenthal (1996: 51), Grotberg (2022: 92) distinguishes several bands of the strength of effect size. The values of $|Q| < 0.2$ correspond to a very weak (dis)association, and those of $0.2 < |Q| < 0.43$ to a weak (dis)association. We may conflate all stronger effects into one strong (dis)association category. Although Yule's Q for homorganic CVC sequences exceeds $|0.43|$ in many languages of our sample, this value is rarely reached for non-homorganic sequences. In the tables to come, we parenthesize the Q coefficients whose 95% CIs do not make them reliable estimates. These values are included in the counts of average Yule's Q for language classes and subclasses.

When there is a significant disassociation, that is, negative Yule's Q with a CI not containing zero, we will speak for convenience and stylistic diversity about a CVC sequence being disfavored, dispreferred, underrepresented, or avoided. On the other hand, a CVC sequence will be described as being favored, preferred, overrepresented, or not avoided when its Yule's Q is positive and the CI does not contain zero (significant association).

Sometimes a combination of two consonants turns out not to be attested at all in our data even though some languages in question combine these phonemes with other consonants. Afrikaans was mentioned above as an example of such a language, one lacking homorganic PALATAL sequences. Although Yule's Q for these cases is −1, which equals the Odds Ratio of 0, its confidence interval cannot be calculated by the standard procedure outlined in footnote 5 because the logarithm of 0 is undefined. The measure mentioned in footnote 4 (adding 0.1 to values A, B, C, and D) makes the CI calculable, but in most cases the interval is too wide to include zero. It is either due to the small size of some of our data or due to the nature of the Yule's Q statistic. Unless additional data are examined, we cannot confidently conclude that such combinations are completely avoided. We have therefore decided not to use languages that lack a certain CVC sequence (we will call them –1 languages) in the calculations of average Yule's Qs. Yet we mention them, as they are obviously of linguistic interest.

In what follows and in the online supplementary materials, Yule's Q values and CIs are calculated from the frequencies of the CVC sequences in which V is the first word nucleus as recorded in the lexical databases (dictionaries) used. A CVC sequence may thus occur in several words containing the same morpheme.

3. Pure SPA in Slavic Languages

Our default hypothesis is that CVC sequences containing consonants from the same identity class are avoided in most or all Slavic languages. Evidence for this avoidance will be negative Yule's Q whose CI does not include zero. We assume four identity classes, LABIALS, DENTALS, PALATALS, and DORSALS, as defined in §2.3. This is what P&S call Pure SPA, in which only place features are involved. The role of non-place features will be considered in §5. A language will be said to exhibit a pure SPA effect if at least three of the four identity classes are avoided (i.e., the majority).

Table 4 (on p. 191) provides Q coefficients for the four identity classes in the 18 Slavic languages.[6] Since the lexemes for which we quantify the CVC distribution are essentially cognates, the differences between the Q values can be understood as an approximation of the development of SPA from Proto-Slavic to its individual offspring.[7] Recall that the statistics have been calculated only for the CVC sequences where V is the first word nucleus. The default hypothesis is corroborated for LABIALS, DENTALS, and DORSALS, but not PALATALS (see the parenthesized values, which are statistically non-significant). Unless the latter class is wrongly delimited, which possibility we address in §5, the Slavic languages are characterized by the absence of PALATAL SPA.

[6] For brevity, we will use the following abbreviations in this section: PSl = Proto-Slavic, OCS = Old Church Slavonic, Bul = Bulgarian, Mac = Macedonian, SCr = Serbo-Croatian, Sln = Slovenian, Slk = Slovak, OCz = Old Czech, MCz = Modern Czech, USo = Upper Sorbian, LSo = Lower Sorbian, Plb = Polabian, Pom = Pomeranian, Pol = Polish, Bel = Belarusian, Ukr = Ukrainian, ORu = Old Russian, MRu = Modern Russian.

[7] See also Pystynen (2014) and Cathcart (2023) on diachronic views on SPA.

Labial avoidance is encountered in all Slavic varieties (as well as most other languages; see below). On average, it is stronger in the South and East varieties (average Yule's Q is −0.60 in both branches) than in the West branch (−0.39). PSl also displays a strong avoidance effect here (−0.70). The decreased strength within the West languages may be a function of the development of the original PSl palatalized labials, which have been dissolved into two-phoneme sequences such as /pj/ in Czech or Polish. Dental avoidance is also observable in all varieties, and if we compare the PSl state with the descendant languages, we could say that the development has been toward a decrease in the strength of this type of SPA. The weakening is again most obvious in the West languages; it may relate to the development of palatals out of dentals/alveolars.

The dorsal homorganic sequences are also widely disfavored, but the Yule's Q test does not provide reliable evidence for an avoidance effect in ORu, OCz, Plb, and USo. For the first three varieties, this result looks suspicious and must be a consequence of the small size of our data. It seems unlikely that ORu and OCz would fail to avoid K_K when these sequences are disfavored both in the previous and subsequent developmental stages (in PSl, and MRu and MCz, respectively). Our datasets for ORu and OCz are among the smallest compared to the other Slavic languages (see Table 4 on the following page). Likewise, the small data size is probably why Plb does not avoid K_K. On the other hand, the absence of dorsal SPA in USo appears to be a genuine feature of this language independent of the data size. USo is the only Slavic language having dorsal sonorants, namely the uvular rhotic /ʁ/ (and its palatalized variant) corresponding to the dental/alveolar rhotic /r/ in the other languages (Howson 2017). When USo /ʁ/ is reclassified as a dental, the Yule's Q statistic produces a high negative significant value for K_K (−0.56), and the value for T_T increases, too (−0.28 instead of −0.16). Thus, though the USo rhotics may be classified as dorsals based on their articulation, phonotactically they do not behave like these consonants, which may be due to them being dentals in origin.

The attitude of the Slavic languages toward palatal homorganic sequences requires more attention. Only three languages significantly disfavor these sequences (Bul, Sln, and SCr); two significantly favor them (LSo and Pol). We lack reliable statistical evidence for this avoidance (or preference) in the remaining 13 languages, though. This diversity may be explained by both the common development within particular Slavic subbranches and the nature of the palatal systems, though the problem is more complex, as is suggested by our alternative data discussed further below.

Within Slavic, we observe five kinds of palatal systems, the breakdown of which is reproduced in Table 5 (on the following page): types P2, P3, P4, P6, and P9 (see the next section and Table 9 on p. 199 for the other types). Each type is characterized by the presence of the consonant classes we place into

Table 4. Yule's Q values for homorganic CVC sequences in the Slavic languages

	Bul	**Sln**	**SCr**	**PSl**	**Mac**	**OCS**	**Bel**	**Ukr**	**MRu**
P_P	−0.69	−0.62	−0.52	−0.70	−0.62	−0.57	−0.65	−0.59	−0.56
T_T	−0.44	−0.25	−0.34	−0.47	−0.32	−0.34	−0.43	−0.42	−0.35
Č_Č	−0.33	−0.27	−0.22	(−0.42)	(−0.26)	(−0.10)	(−0.25)	(−0.17)	(0.04)
K_K	−0.41	−0.22	−0.35	−0.33	−0.31	−0.49	−0.50	−0.53	−0.49
N	2,809	2,899	3,831	1,259	1,584	2,073	2,490	2,814	4,259

	Slk	**MCz**	**Pom**	**LSo**	**Pol**	**ORu**	**OCz**	**Plb**	**USo**
P_P	−0.59	−0.42	−0.29	−0.22	−0.16	−0.61	−0.61	−0.51	−0.29
T_T	−0.30	−0.25	−0.14	−0.19	−0.14	−0.40	−0.20	−0.30	−0.16
Č_Č	(−0.08)	(−0.01)	(−0.09)	0.17	0.11	(−0.10)	(−0.05)	(−0.27)	(0.01)
K_K	−0.47	−0.38	−0.56	−0.40	−0.56	(−0.18)	(−0.40)	(0.12)	(−0.07)
N	2,354	3,189	1075	1,701	2,500	1,292	1,101	745	1,679

(P_P = labial sequences, T_T = dental sequences, Č_Č = palatal sequences, K_K = dorsal sequences. Statistically non-significant values are parenthesized. Shading highlights significant positive values. The row N gives the total number of CVC sequences in a particular language.)

Table 5. A partial typology of palatal systems, applied to the Slavic languages

Type	**/j/**	**/ʃ/**	**/ɲ/**	**/c/**	**/ç/**	**/ɕ/**	**O**	**Languages**
P2	•	•						PSl, **Bul**, **Sln**, Bel, Ukr, ORu, MRu, USo, Plb
P3	•	•	•					Pom
P4	•	•	•	•				Mac, OCS, OCz, MCz, Slk
P6	•	•	•			•		**SCr**, *Pol*
P9	•	•				•		*LSo*

(/j/ = palatal approximants, /ʃ/ = post-alveolar sibilants, /ɲ/ = palatal sonorants (nasals and liquids), /c/ = palatal plosives, /ç/ = palatal fricatives, /ɕ/ = alveolo-palatal sibilants, O = other segments (such as clicks). The languages in bold avoid palatal sequences; those in italics favor them; the rest show no distributional anomaly in either direction.)

the palatal identity class, namely the palatal approximant /j/, post-alveolar sibilants (/ʃ/, which includes both affricates and fricatives), palatal sonorants other than /j/ (/ɲ/, typically the nasal but also /ʎ/ in Slk or the /ř/ phoneme in MCz),[8] true palatal plosives (/c/), true palatal fricatives (/ç/, of the kind found in German but not attested in Slavic), alveolo-palatal sibilants (/ɕ/, e.g., in Pol, again including both affricates and sibilants), and some other segments (O, not found in Slavic).

Type P2 is the most common palatal system, limited to post-alveolar sibilants and /j/. The Č_Č avoidance has been evidenced in two of the nine languages belonging here (Bul and Sln). The other seven instances are all four East languages, PSl, USo, and Plb, each of them lacking palatal avoidance. This absence is shared with type P4, which differs from P2 in containing true palatal plosives (and sonorants). P4 consists of two South Slavic languages, Mac and OCS, and three West languages, OCz, MCz, and Slk (the Czech-Slovak subgroup).

All these facts support our assumption that the presence of palatal SPA is both related to the nature of palatal systems and to the common development of Slavic subbranches. First, palatal avoidance occurs in the South Slavic varieties that lack true palatal plosives; the non-South languages without true palatal plosives fail to disfavor Č_Č.[9] Second, palatal avoidance is also absent in languages with true palatal plosives irrespective of their affiliation. Finally, it is also absent in Pom, which has true palatal sonorants but not plosives. However, it is not obvious whether the lack of palatal SPA is really due to the sonorants because the non-Slavic languages belonging in P3 tend not to underrepresent Č_Č, while those belonging in P4 do (see Table 9 on p. 199).

Let us move to the remaining three varieties, SCr, Pol, and LSo. The typology in Table 5 does not explain the nature of palatal SPA in these languages. SCr and Pol belong in P6, and LSo belongs in P9, but SCr avoids Č_Č, whereas Pol and LSo favor these sequences. All three of these languages have two sibilant series within the palatal class, the post-alveolars and alveolo-palatals.

8 In a more minute typology, the palatal nasals and liquids could be separated, but since the purpose of this typology is to find out correlations between SPA and palatal systems, we have grouped them together. We have not found any such correlation in OCz, MCz, Slk, OCS, and SCr, which all possess palatal liquids. Neither has any such correlation been detected in 17 other non-Slavic languages that possess these sonorants. The interpretations of the phoneme systems are in accord with the illustrations of the IPA published in the *Journal of the International Phonetic Association*, which are available for most Slavic languages.

9 PSl transcends the classification into West, East, and South branches but belongs in P2 and does not avoid Č_Č. It is worth noting that Yule's Q for Č_Č in PSl reaches the greatest negative value among the Slavic languages (−0.42), and its confidence interval, which is rather large due to the small size of the PSl data, mostly occupies the negative area ([−0.72, 0.01]). This suggests avoidance.

In SCr, however, alveolo-palatals are limited to affricates, whereas the other two languages have both affricates and fricatives. Thus, it seems that the preference of Č_Č encountered in Pol and LSo is a West Slavic innovation related to the development of a rich system of alveolo-palatal sibilants in addition to post-alveolar sibilants.

The unique nature of Č_Č in Pol can further be illustrated with the help of a larger set of lexical data taken from WikiPron dictionaries extracted from Wiktionary (107,924 first-syllable CVC sequences).[10] Table 6 on the following page provides Q values for the combinations of the three sibilant series Pol possesses, the dento-alveolars /t͡s/, /d͡z/, /s/, /z/; the post-alveolars/retroflexes /t͡ʂ/, /d͡ʐ/, /ʂ/, /ʐ/; and the alveolo-palatals /t͡ɕ/, /d͡ʑ/, /ɕ/, /ʑ/. All combinations of the dento-alveolar sibilants with any other sibilants in any order are avoided (including repetitions of dento-alveolar sibilants). In contrast, three of the four combinations of the post-alveolar and alveolo-palatal sibilants (the shaded cells) are favored; it is only sequences of an alveolo-palatal plus a post-alveolar that are avoided.[11] These facts suggest that the post-alveolar and alveolo-palatal sibilants phonotactically belong to one class distinct from that of the dento-alveolar sibilants. This conclusion supports our decision to place Pol and LSo post-alveolar sibilants into the PALATAL class, even though some linguists (Hamann 2002; Zygis 2003) have interpreted these sibilants as retroflexes (which are placed in the DENTAL class in the other languages of our sample).

Having access to some other lexical databases, we can examine the nature of SPA even further. Our primary datasets of the words drawn from the *ESJS* indices are essentially cognates, native to the Slavic family, and basic, uninflected, and underived forms. However, the Slavic languages contain more vocabulary, including inflected and derived forms and many words imported from other languages. All these items may influence the SPA effect and its strength. Table 7 on the following page reproduces the Q values for alternative lexical databases for nine Slavic languages. The databases are not necessarily related to each other or unbiased. The MCz data contain both native and borrowed words extracted from several dictionaries; the MRu and Sln data are derived from headwords of two etymological dictionaries; the material for the remaining languages is again taken from WikiPron dictionaries, which may contain a rather large number of loanwords.

10 See the online appendix for the sources of these additional databases, https://ojs.ung.si/index.php/JSL/article/view/469/345.

11 The negative Q value for the CVC sequences of alveolo-palatals followed by post-alveolars may be due to a more general avoidance of the CVC sequences that begin with an alveolo-palatal. First, notice that the combinations of alveolo-palatals with alveolars and post-alveolars reach the greatest negative Q values. Second, the C_1VC_2 sequences in which C_1 is an alveolo-palatal are less common than the ones in which C_2 is such a consonant.

Table 6. Yule's Q values for combinations of sibilants in the Polish alternative data (107,924 CVC sequences)

	Dento-alveol.	Post-alveolar	Alveolo-pal.
Dento-alveolar	−0.20	−0.31	−0.29
Post-alveolar	−0.23	0.12	0.14
Alveolo-palatal	−0.53	−0.34	0.20

(Shading marks homorganic combinations of the sibilants belonging to the palatal class.)

Table 7. Yule's Q values for alternative data of selected Slavic languages

	Bul	Sln	SCr	Mac	Bel	Ukr	MRu	MCz	Pol
P_P	−0.49	−0.56	−0.59	−0.53	−0.29	−0.57	−0.66	−0.47	−0.22
T_T	−0.40	−0.31	−0.33	−0.29	−0.34	−0.36	−0.33	−0.28	−0.17
Č_Č	−0.49	(−0.06)	(0.07)	(−0.03)	−0.35	−0.09	−0.11	+0.09	0.09
K_K	−0.65	−0.45	−0.44	−0.30	−0.38	−0.45	−0.43	−0.49	−0.37
N	30,527	8,460	19,643	49,091	29,748	23,441	16,360	62,107	104,924

(The shaded values are markedly different from those reproduced in Table 4 on p. 191.)

The individual Q values for P_P, T_T, and K_K reproduced in Table 7 differ from those in Table 4 on p. 191 (= our primary data), but they are all significantly negative. Hence, the avoidance of these sequences is a stable feature of the Slavic languages, independent of the type of lexical data. However, what interests us most is Č_Č, and here we observe substantial differences (shaded in the table). We have previously seen that no other Slavic languages but Bul, Sln, and SCr avoid palatal homorganic sequences. The alternative data produce a slightly different picture. Sln and SCr do not disfavor these sequences, in which they agree with Mac. On the other hand, all three East Slavic languages show a palatal avoidance effect, though its strength is relatively weak. The *ESJS* data provide no evidence for this avoidance here. To complicate the situation even further, we have calculated Yule's Q for 3,750 native roots in MRu, using the quantitative data from Pozdniakov (2007), and these data do not indicate that Č_Č is avoided. Thus, the MRu situation recorded in Table 7 may be due to the presence of loanwords. Finally, the MCz data manifest a slight but statistically insignificant degree of *preference* for palatal homorganic sequences. When we calculated Yule's Q for 7,609 native roots in MCz from a different source (see the online appendix), these data provided no evidence for preference or dispreference of Č_Č. All these alternative datasets suggest that loanwords, inflected words, and derivates influence the nature of palatal avoidance.

Despite the variability, one obvious conclusion is inferable from the facts presented in this section: the Slavic languages are not characterized by any strong or stable tendency to avoid homorganic combinations of palatals.

4. Pure SPA in Non-Slavic Languages

To see whether the state observed for Slavic is specific to this language group, we have evaluated lexicons for other languages, of which 44 belong to the IE family, and 38 are recruited from several non-IE families. We will not provide individual Yule's Q values for these languages; they can be found in the online appendix (see fn 2). The IE languages will be discussed in more detail because they are closely related to the Slavic languages, and because no other family is represented by so many members in our data. We believe the more comprehensive and exhaustive the analyzed language data are, the more useful information they can provide. We reject the idea that one language represents the whole family and that limited language representatives lead to meaningful, cross-linguistically valid generalizations ("universals"). In short, the non-IE languages only supplement our analysis, providing tentative evidence about the situation outside the IE world.

4.1. Non-Slavic IE Languages

Table 8 on the following page reproduces average Q values for Proto-Indo-European (PIE) and all IE branches, some of which are represented by one language only.[12] The table includes the Slavic branch too. The parenthesized numbers express how many languages have produced a reliable Q value; only these languages are included in the calculation of the averages. The Hellenic languages (Ancient and Modern Greek) do not have any palatals, at least in our data. Modern Greek palatal consonants are allophones of velars and partly

[12] Our analysis confirms, in essence though not in detail, the findings of previous SPA studies. See P&S, Cooper (2009), and Sandell (2015) on PIE; Berkley (2000) and Grotberg (2022) on Latin; Grotberg (2022) on Medieval Castilian; Plénat (1996), Berkley (2000), Rousset (2004), and Carrissimo-Bertola (2010) on French; MacNeilage et al. (2000) on Spanish; Frisch et al. (2004) on Italian; Carrissimo-Bertola (2010) on Portuguese; Berkley (1994, 2000), Frisch (1997), MacNeilage et al. (2000), and Dmitrieva (2008) on English; van Goch (2010) and Boll-Avetisyan and Kager (2014) on Dutch; Khan (2007) on East Bengali.

Table 8. Average Yule's Q values for homorganic CVC sequences in particular IE branches

	PIE (n = 1)	**Hellenic (n = 2)**	**Italic + Romance (n = 8)**	**Anatolian (n = 1)**	**Celtic (n = 4)**
P_P	−0.58 (1)	−0.71 (2)	−0.65 (8)	−0.63 (1)	−0.60 (4)
T_T	−0.51 (1)	−0.42 (2)	−0.38 (8)	−0.44 (1)	−0.36 (4)
Č_Č	0 (0) [1]	×	−0.25 (3) [1]	N.S.	0 [1]
K_K	−0.32 (1)	−0.46 (2)	−0.49 (8)	−0.63 (1)	−0.42 (3)

	Baltic (n = 2)	**Germanic (n = 15)**	**Albanian (n = 1)**	**Indo-Aryan (n = 8)**	**Armenian (n = 1)**
P_P	−0.57 (2)	−0.56 (15)	−0.56 (1)	−0.53 (8)	−0.52 (1)
T_T	−0.43 (2)	−0.35 (15)	−0.27 (1)	−0.31 (7)	−0.19 (1)
Č_Č	−0.25 (2)	−0.50 (10) [4]	−0.13 (1)	−0.30 (6) [1]	N.S. (×)
K_K	−0.54 (2)	−0.33 (14)	−0.14 (1)	−0.52 (8)	−0.52 (1)

	Tocharian (n = 1)	**Slavic (n = 18)**	**All non-Slavic (n = 44)**	**All IE (n = 62)**
P_P	−0.32 (1)	−0.51 (18)	−0.58 (44)	−0.56 (62)
T_T	−0.20 (1)	−0.30 (18)	−0.36 (43)	−0.34 (61)
Č_Č	N.S. (×)	−0.27 (3)	−0.37 (22) [8]	−0.36 (25) [8]
K_K	−0.58 (1)	−0.43 (14)	−0.43 (42)	−0.40 (56)

(The numbers in round brackets, or parenthesis, indicate how many languages with significant *negative* Q values have been used for the calculation. The numbers in square brackets express how many languages lack certain combinations altogether (the −1 languages); they have not been included in the average calculation. Square-bracketed values are not given when they equal 0. N.S. = not significant.)

of alveolars (Arvaniti 2007). Although Table 8 naturally invites us to compare the strength of the SPA effect across various IE branches and its development from PIE, we should keep in mind that the comparison is only approximate.[13] First, we compare average Yule's Q values calculated from samples of various sizes and contents. The Q values tend to be very similar to each other, and

[13] We would like to thank one of the reviewers for pointing this out.

their confidence intervals, moreover, largely overlap. Second, since Yule's Q is more or less normally distributed, we can say that one Q is larger or smaller than another, but the absolute difference between these two values does not really tell us how different the values are (Grotberg 2022: 93ff.).

Looking at the last three columns in Table 8, we witness that the non-Slavic IE languages are similar to the Slavic ones both in the extent and strength of the avoidance of P_P, T_T, and K_K as well as in the fact that palatals are again exceptional, though less obviously. All non-Slavic IE languages avoid P_P just like the Slavic ones do, and this type of avoidance is the strongest in nearly every branch. The exceptions are Anatolian (i.e., Hittite), Armenian, and Tocharian (B), in which dorsals equal or exceed the avoidance strength of labials. The average Q value for P_P is a little greater in the non-Slavic languages, which is due to the consistently higher values across the non-Slavic branches and due to the comparatively weaker avoidance of P_P in the West Slavic languages. Curiously, the average Q value for the non-Slavic languages agrees with the value of PIE (see the first column). It is also worth pointing out that four IE branches have developed a stronger avoidance effect for P_P than PIE. They are placed in the upper part of Table 8 (Hellenic, Italic + Romance, Anatolian, and Celtic).[14] All the remaining branches, including the Slavic one, are characterized by a decrease in the strength of labial avoidance.

As mentioned, K_K sequences show the strongest avoidance in Hittite, Armenian, and Tocharian B. On average, however, the strength of dorsal avoidance is virtually identical in the Slavic and non-Slavic languages. If we accept PIE and its Q values as an approximation of the original state of the IE languages, we notice that all the IE branches have evolved into having stronger avoidance of the dorsal homorganic sequences than PIE originally had. This may result from the development of the three PIE velar series (plain, palatalized, and labialized). Dorsal avoidance is not evidenced for Breton and Plautdietsch, though. While in Breton the lack of evidence may be a function of the small lexical sample (n = 888), the size of the Plautdietsch data is high enough to be reliable (n = 13,510). In fact, the absence of dorsal avoidance in this language may not be surprising at all because Germanic languages generally show weaker avoidance here compared to the other languages.

The dental CVC sequences are underrepresented in nearly all non-Slavic IE languages except for Pashto. However, this may only be a consequence of the small size of the Pashto data, which may also be the reason why this language fails to avoid palatal sequences (see below). At the same time, T_T sequences tend to be less strongly avoided than the non-coronal sequences, which we have already observed in the Slavic languages, and which is also true for the non-IE ones (see the next section). And if the K_K avoidance strength appears to have been increased in all IE branches except for the Germanic one, the

[14] Individually, only the Portuguese Q value for P_P does not exceed that for PIE.

dental avoidance strength has instead been relaxed in every IE branch and, in fact, in nearly every IE language. The languages in which Yule's Q is greater than that for PIE (−0.51) are Latin (−0.61) and Modern Greek (−0.55).[15] This change is likely a consequence of the rise of palatals other than /j/ (including true palatals and post-alveolars) in the particular IE varieties. Other factors must have played a role, too, because some IE languages have a rather limited palatal inventory.

This brings us to the last place class, palatals, whose homorganic sequences are again avoided in fewer languages than sequences of the other places. However, the portion of the avoidant languages is considerably larger than in the Slavic group. The avoidance is observed in 30 (68.18%) non-Slavic IE languages, of which eight lack Č_Č sequences altogether (Afrikaans, Gothic, Old Frisian, Proto-Germanic, Ossetian, Proto-Celtic, Latin, and PIE). In the Slavic family, there are only three such languages (16.67%).

The presence or absence of palatal avoidance thus apparently depends both on the languages' genealogy and the nature of their palatal inventory. In Table 9 on the following page, we present the full typology of the palatal systems across all three language groups, which was already partly reproduced in Table 5 on p. 191 for Slavic. The columns with IPA symbols stand for the consonant classes we group into the palatal SPA class.[16] The last four columns provide the ratio between the languages that significantly avoid Č_Č and all languages of a particular type and a particular language group.

---

15 We have also calculated the Q value for T_T from the quantitative data for PIE used in P&S, which relied on a different material. The value is −0.49, which is still relatively high. Yule's Q for P_P is −0.66, for K_K, it is −0.46, and the value for Č_Č is not statistically significant. Note also that P&S are silent about their exact classification of the PIE consonants, but it is likely that they treat palatovelars as palatals, which is what Sandell (2015: 11, fn 16) does as well. Classifying our data in the same way, we get similar results to those of P&S. In contrast, Cooper (2009: 60) views palatovelars and labiovelars as dorsals with secondary articulations, which has also been suggested by one of the reviewers (see also Weiss 2016). We have followed this classification here, but it is worth noting that when the palatovelars are excluded from the dorsal class, Yule's Q for this class increases (−0.43 for our data).

16 The sibilant classes /ʃ/ and /ɕ/ include both affricates and fricatives (post-alveolar and alveolo-palatal ones, respectively; see also Kokkelmans 2021 on sibilants in general). On the other hand, the class of true palatal fricatives /ç/ excludes affricates because they are either non-occurrent in our data or interpreted as variants of palatal plosives, with which they vary in some languages (e.g., in Albanian, Kolgjini 2004; or Hungarian, Siptár and Törkenczy 2000). Note also that the typology reflects the phonological databases we use and the phonological classification they employ. Thus, for example, Finnish is classified as a language with post-alveolar sibilants (contrary to Kokkelmans 2021) because our Finnish database contains a number of loanwords with these segments.

Table 9. Full typology of PALATAL systems

Type	/j/	/ʃ/	/ɲ/	/c/	/ç/	/ɕ/	O	Slavic	Other IE	Non-IE	All
P0								×	×/2	×	×/2
P1	•							×	6/7	0/1	6/8
P2	•	•						2/9	14/17	3/5	19/31
P3	•	•	•					0/1	3/7	2/7	5/15
P4	•	•	•	•				0/5	2/3	3/4	5/12
P5	•	•	•	•	•			×	0/3	1/1	1/4
P6	•	•	•			•		1/2	×	1/1	2/3
P7	•	•	•				•	×	×	0/1	0/1
P8	•	•		•				×	×	1/1	1/1
P9	•	•				•		0/1	1/1	1/3	2/5
P10	•		•	•	•			×	1/1	1/1	2/2
P11	•		•	•				×	×	4/7	4/7
P12	•		•	•		•		×	×	3/3	3/3
P13	•		•			•		×	1/1	1/1	2/2
P14		•		•				×	1/1	×	1/1
P15		•						×	×	0/1	0/1
P16						•		×	1/1	0/1	1/2
Total								3/18	30/44	21/38	54/100

(/j/ = palatal approximants, /ʃ/ = post-alveolar sibilants, /ɲ/ = palatal sonorants (nasals and liquids), /c/ = palatal plosives, /ç/ = palatal fricatives, /ɕ/ = alveolo-palatal sibilants, O = other segments (such as clicks). The last four columns provide the ratio between the languages that avoid Č_Č and all languages of a particular type. The statistics include languages that lack PALATAL sequences.)

For the sake of completeness, the typology includes type P0 for two languages without any PALATALS (Ancient and Modern Greek).

The simplest PALATAL inventory among the IE languages is the one that just contains /j/ (P1), although this system is not found in the Slavic group. Likewise, its representation is too small in the non-IE language group to draw any conclusion (only Greenlandic belongs here, lacking PALATAL SPA). All seven non-Slavic IE languages with this system avoid Č_Č. Proto-Celtic is of this kind, thereby differing from the three non-reconstructed Celtic languages in our sample (Welsh, Irish, and Breton) that have different systems and that lack palatal avoidance.

Another simple type is P2, with just /j/ and post-alveolar sibilants. It is by far the most frequent palatal system cross-linguistically. It is represented by 17 non-Slavic IE varieties, of which as many as 14 avoid Č_Č. This tendency conspicuously contrasts with the situation in the Slavic group, where type P2 is also the most common system but where only Bulgarian and Slovenian show a palatal SPA effect. The ratio for the non-IE languages is also suggestive of the tendency, but our sample is again too small to draw any conclusion.

Of the non-Slavic IE languages belonging in P2, it is only Armenian, Welsh, and Pashto for which we lack clear evidence for palatal avoidance. Since Armenian is the only representative of its branch, it is hard to judge to what extent the absence is an anomaly. In contrast, the Welsh case may be a feature characterizing the whole Celtic branch (see above). However, the absence of palatal avoidance in Pashto is very likely due to the limited data, which is also a reasonable explanation for the lack of dental avoidance in this language, too (see above). Moreover, the Č_Č sequences are underrepresented in the other three Iranian languages, all belonging in P2 as well (Ossetian, Kurmaji, and Persian).

Let us move to P3 (/j/, post-alveolar sibilants, and other palatal sonorants, usually the nasal /ɲ/). It is another very common system encountered among the non-Slavic languages (both IE and non-IE), but not in the Slavic group, where only Pomeranian belongs here, lacking palatal SPA. The ratios in Table 9 indicate that P3 languages tend *not to avoid* Č_Č, which appears to relate to the presence of palatal sonorants other than /j/. Recall that P2 languages, without these sonorants, instead tend *to avoid* palatal sequences. This conclusion is further supported by the non-IE group, where five out of seven P3 languages lack palatal avoidance and where three out of five P2 languages have this avoidance.

The non-Slavic IE languages belonging in P3 and not avoiding Č_Č are three Romance varieties: Italian, Catalan, and French (the last one significantly favors these sequences; see below). In contrast, another two Romance languages belonging in P3 have palatal avoidance, namely Spanish and Portuguese; Yule's Q for the former is relatively high, −0.43, while the latter attests only very weak avoidance (−0.15; so does Romanian, −0.17).[17] Finally, type P3 also includes Breton (no palatal avoidance; see above on the Celtic languages) and Hindi (avoidance). The Hindi case should perhaps be reconsidered be-

[17] The contents of the Spanish palatal system may be a matter of discussion. First, the phonemic status of [j] is uncertain because some analysts interpret it as a non-syllabic allophone of /i/ (Hualde 2005). Secondly, the obstruent /ʝ/ has merged in Spanish dialects with the sonorant /ʎ/, which is reflected in the transcription of the lexical database that is the source of our data (i.e., Spanish is interpreted as a language lacking palatal plosives). This might have influenced the nature of palatal SPA.

cause the palatal nasal occurs here (a) in clusters before another palatal consonant, in which case it might be allophonic, and (b) in several loanwords, in which case the occurrence is marginal (Ohala 1999).

A fairly common system among the Slavic languages, though not so frequent elsewhere, is type P4, differing from P3 by containing true palatal plosives. All five Slavic languages of this kind lack palatal avoidance, which contrasts again with the other language groups. The Č_Č sequences are avoided in two out of three non-Slavic IE languages (Albanian and Latvian, but not Tocharian B) and in three out of four non-IE languages (Nancowry, Khasi, and Hungarian, but not Basque). This again supports our conclusion that the increased absence of palatal avoidance is a peculiarity of the Slavic family.

The last type instantiated by more than one IE language but absent in Slavic is P5, in which we find both true palatal plosives and fricatives. All three non-Slavic IE languages belonging here do not underrepresent Č_Č (Galician, Irish, and Plautdietsch). However, Turkish, as the only non-IE representative, avoids these sequences. There are two other languages that contain true palatal fricatives but lack post-alveolar sibilants, and both avoid palatal sequences, too (Icelandic and Nyah Kur). More languages are needed to evaluate the possibility that the nature of palatal avoidance is dependent on the presence of true palatal fricatives like /ç/. Note that Hall (1997) presents arguments for not regarding these fricatives as coronals (we have classified all palatal consonants as coronals).

There is not a lot to say about the remaining types. Many are attested only in the non-IE group, and those found among the two IE groups are largely represented by one language only. The only tentative generalization we can make here is that the Č_Č sequences are inclined to be avoided in languages without post-alveolar sibilants (types P10–13 including the only discussed type P1). To conclude our discussion of the non-Slavic IE languages, let us mention that there is one language that favors Č_Č, namely French. We have related the preference of palatal homorganic sequences to the presence of two palatal sibilant series (post-alveolars and alveolo-palatals) in Polish and Lower Sorbian. However, the same explanation does not hold for French, which has a fairly standard palatal system (post-alveolar sibilants, /j/, and /ɲ/). The non-Slavic languages, whether from the IE family or not, cannot be helpful in corroborating our hypothesis that the preference in Polish and Lower Sorbian is due to the presence of the two sibilant series. Four non-Slavic languages have two sibilant series within the palatals. Norwegian and Tatar avoid Č_Č but limit alveolo-palatals to fricatives, whereas Polish and Lower Sorbian also have alveolo-palatal affricates. Alveolo-palatal affricates and fricatives are present in Lezgian and Chukchi, but their palatal Q values are not significant.

4.2. Non-IE Languages

Our sample comprises 38 non-IE languages selected from several macro-families, most of which are represented by one variety only. An exception is the Austroasiatic family, with nine members from the Mon-Khmer group, which has been motivated by the availability of phonologically transcribed lexical data. Since P&S provided statistical information on the distribution of all CVC sequence types, we take advantage of their data for the 28 non-IE languages in their sample, using Yule's Q to gain a broader perspective.[18] Table 10 on the following page reproduces average Yule's Q values for both non-IE language samples, comparing them with the values for the Slavic and other IE languages repeated from the previous sections.

The non-IE languages confirm the trends already spotted in the other groups. Labials are most strongly and widely avoided, though there are languages without such avoidance. Dorsals show the second strongest avoidance effect; several languages are without such avoidance within each group. Dentals are also very widely avoided but less strongly. Finally, palatals are avoided by markedly fewer languages than the other places of articulation, and the non-IE languages are similar in avoidance strength to the Slavic languages, while the non-Slavic IE languages are not. There are also several other differences in detail to mention.

All IE languages have been found to have a labial avoidance effect. In contrast, we lack reliable evidence for this avoidance in one non-IE language in our sample (Otomi) and two in P&S's sample (Bullom and Sua). What is more, two non-IE languages in our sample significantly *prefer* P_P (Chinese and Mazahua). The weaker degree of labial avoidance seen in P&S's bundle is probably a trait of African languages, which make up most of the sample. In particular, Niger-Congo languages tend to have a weaker avoidance effect (including Xitsonga in our sample). On the other hand, the decreased negative Q value for labials in our non-IE sample can be related to the nine Mon-Khmer languages, of which eight have labial Q values below the average.

[18] Another re-evaluation using Yule's Q was already done by Grotberg (2022), though she did not provide actual Q values. From P&S's sample of non-IE languages, we have left out Basque because it is already part of our material, though the source of the data is different. The re-calculated statistic using Yule's Q is also available at https://ojs.ung.si/index.php/JSL/article/view/469/344. We should also note that there is some uncertainty about the statistics of P&S's languages. Having recalculated their numerical values from the paper's appendix, we have found out that there are probably some errors either in the values or in the tables reproduced in the paper proper. For example, in their Table 15, Mpongwe is marked as not significantly avoiding P_P, even though the O/E deviation value for this pattern is −25.8, which they otherwise regard as significant avoidance. Also, the sum of the occurrences of particular combinations for Mongolian and Kiga-Nkore does not equal the sum provided in the appendix.

Table 10. Average Yule's Q values for homorganic CVC sequences and the numbers of languages avoiding these sequences

	Slavic (n = 18)	**Non-Slavic IE (n = 44)**	**Non-IE ours (n = 38)**	**Non-IE P&S (n = 28)**
P_P	−0.51 (18)	−0.58 (44)	−0.49 (35)	−0.41 (26)
T_T	−0.30 (18)	−0.36 (43)	−0.26 (38)	−0.30 (25)
Č_Č	−0.27 (3)	−0.37 (22) [8]	−0.30 (21)	−0.28 (13)
K_K	−0.43 (14)	−0.43 (42)	−0.38 (35)	−0.45 (25)

(See Table 8, p. 196, regarding the round and square brackets.)

The already-mentioned Chinese and Mazahua are exceptional within the whole language bundle. They are the only varieties (including P&S's sample) that prefer LABIAL as well as DORSAL homorganic sequences. Admittedly, the association effect is very weak, though still significant (0.06 and 0.14 for LABIALS and DORSALS in Chinese, 0.08 and 0.16 for LABIALS and DORSALS in Mazahua, respectively). It is yet to be determined to what extent this is due to the nature of our lexical data and/or the statistical measure used, but we do not rule out the possibility that there are really languages preferring two NON-CORONAL sequences. The preference for LABIALS has been previously reported for Yup'ik (Rousset 2004: 196), though it concerned sequences of bilabials only (we are unaware of any mention of languages preferring DORSALS).

There is little to say about DENTALS (but see the next section). The strength and extent of DENTAL avoidance in the non-IE languages are comparable to the situation in the other families. This avoidance is observed in all non-IE languages of our sample but is, curiously, absent in three of the P&S sample, namely Kiga-Nkore, Quechua, and Sue. There is only one language in our sample for which we lack evidence for DORSAL avoidance, Otomi (see below). In P&S's sample there are three languages without this avoidance (Fula, Manjaku, and Sua); since all are from the Atlantic family, the absence may be an areal/genetic feature.

The avoidance of PALATAL homorganic sequences is again somewhat anomalous because many non-IE languages fail to avoid them, just as the IE languages do. In the previous subsection, when discussing the PALATAL system typology, we argued that the presence or absence of this avoidance might relate to the contents of the PALATAL inventory. However, since most types we introduced are generally represented by eight or fewer non-IE languages, we need more comprehensive data to draw reasonable conclusions. Therefore, our discussion will be limited here to a couple of general observations.

To begin with, there is agreement among all four bundles of languages on the fact that Č_Č sequences are avoided by the smallest number of languages in contrast to the other places. The Slavic family clearly differs from the rest because the portion of the avoidant languages is the smallest here (only three languages avoid Č_Č, i.e., 16.67%). The portion is larger in both non-IE bundles (55.26% for our own sample, 46.43% for P&S's one), but by far the greatest number of the languages avoiding Č_Č is found in the non-Slavic IE group (61.36%, including the −1 languages that lack these sequences altogether).

Another language-specific feature appears to be the absence of palatal avoidance in all nine non-Atlantic Niger-Congo languages included in our and P&S's sample (the Atlantic languages proper do not have this feature). Three of the non-Atlantic languages even prefer Č_Č (Proto-Bantu, Bemba, and Kiga-Nkore). The preference is also observable in languages from other families, namely, Korean, Tagalog, Maliseet-Passamaquoddy, Cebuano, and Semai (all included in our own sample). Above we argued that the preference of Č_Č in Polish and Lower Sorbian may relate to the presence of alveolo-palatal sibilants, but this cannot explain the same kind of preference within the non-IE languages mentioned because only Korean has alveolo-palatal sibilants.

One final remark on the non-IE bundle: We find two languages here for which Yule's Q statistic provides no evidence for the SPA effect. One is Otomi (our sample), which significantly avoids dentals only; the other is Sua (P&S's sample), where no place is significantly underrepresented. The latter case is suspicious because the data P&S used (495 CVC sequences) are too small to warrant reliability. Our Otomi dataset is larger (2,242 sequences) but may be non-representative or selective. It is taken from Lexibank (List et al. 2022), which is itself derived from a word list included in the World Loanword Database (Haspelmath and Tadmor 2009; note, however, that only 11% of these words are loanwords).

In the IE bundle, most languages avoid three or four places, and only seven avoid two out of the four places. Thus, at most we could claim that there is no conclusive evidence for avoidance or non-avoidance in some IE languages. On the other hand, the possibility of there being a non-IE language without SPA remains open. Mayer et al. (2010) mention that such languages do exist, especially around the equator, without providing any examples. However, the size of the lexical samples upon which they based their observations were too small to be reliable (no more than 246 CVC sequences for any of 3,200 languages). Judging from the data kindly shared by Thomas Mayer (personal communication, December 2024), some examples are Pirahã, Amuesha, Yavitero, Kalina, and Sambe. We have evaluated one larger lexical sample for

Kalina (615 CVC sequences), and Yule's Q has indicated that this language has an SPA effect for all places but PALATALS.

5. Coronals

We have demonstrated that SPA is observable across various language groups, but two crucial reservations apply. The avoidance effect is smaller for DENTALS than for LABIALS and DORSALS. Second, PALATALS are avoided only in 53.9% of languages (including P&S's 28 ones, or 56% in our 100-language sample), primarily due to the Slavic languages lacking this avoidance. Moreover, homorganic sequences of PALATALS happen to be *favored* in 10 languages (7 in our sample). These unusual properties invite us to review how the CORONAL place classes are delimited and how SPA is formulated.[19] As explained in §2.3, we use the term CORONAL as a convenient cover label for DENTALS and PALATALS but assume that SPA applies to each of them. However, some studies hold that the SPA principle concerns the whole class of CORONALS (to be discussed in §5.1). Moreover, since the whole class of CORONALS happens to manifest a weaker avoidance effect, some other studies argue that CORONAL avoidance is sensitive to manner features and that the class should be divided into obstruents and sonorants for SPA to obtain (§5.2). Finally, the separation of DENTALS and PALATALS may be essentially correct, but the membership of some consonants in these classes should be reconsidered (§5.3). Our discussion will conclude with remarks on the so-called labial-coronal effect and the CVC sequences not directly affected by the SPA principle.

5.1. Coronal Avoidance

In line with P&S and others, we assume that SPA applies to four basic place classes: LABIALS, DENTALS, PALATALS, and DORSALS. However, some researchers assume fewer classes, namely those that correspond to the three major articulators: labials, coronals, and dorsals. The most straightforward reduction is the one that merges DENTALS and PALATALS into CORONALS.[20] Such a union

---

[19] It is also possible that the difference between DENTALS and PALATALS on the one hand and LABIALS and DORSALS on the other is a consequence of the statistical measure used. However, we have specifically chosen Yule's Q statistic because Grotberg (2022) has convincingly demonstrated that other statistical tools previously used by researchers, like O/E or χ^2, are biased. They consistently underestimate the avoidance of CORONALS (DENTALS) because they are sensitive to the high frequency these consonants have in most languages.

[20] Remember that small capitals indicate that CORONALS is a label for the union of the classes of DENTALS and PALATALS. Other usages of the term *coronal* are discussed in the following subsections. They are written in lowercase letters. The union of dentals and

Table 11. Average Yule's Q values for homorganic CORONAL CVC sequences and the numbers of languages avoiding these sequences

	Slavic (n = 18)	Non-Slavic IE (n = 44)	Non-IE ours (n = 38)	Non-IE P&S (n = 28)
Cor_Cor	−0.29 (18)	−0.39 (43)	−0.35 (36)	−0.46 (27)
T_T	−0.30 (18)	−0.36 (43)	−0.26 (38)	−0.30 (25)
Č_Č	−0.27 (3)	−0.37 (22) [8]	−0.30 (21)	−0.28 (13)

(See Table 8 on the round and square brackets.)

was explicitly argued for by Mayer et al. (2010) for various languages on the grounds that these consonants behave similarly in the SPA statistic (see also Teil-Dautrey 2008 for Proto-Bantu or Grotberg 2022 for Latin and Medieval Castilian and a re-analysis of P&S's data using the single class of CORONALS).

Applying this merger to our sample, we discover that all but two languages avoid homorganic sequences of CORONALS (see Table 11 above). The exceptions are Pashto and Chinese; another is Kamilaroi in P&S's data. The avoidance of CORONALS shows a stronger avoidance effect than the avoidance of DENTALS in both non-IE language bundles; the effect also slightly weakens in the non-Slavic IE languages. In the Slavic languages, the avoidance effect is virtually the same, though it drops a little (average Yule's Q = −0.29 for CORONALS and −0.30 for DENTALS).

Although the general increase in the avoidance strength speaks for the union of DENTALS and PALATALS, there is one serious problem with this union: If homorganic sequences of CORONALS are subject to SPA, we expect that *any* combination of two CORONALS is also avoided, that is, SPA should be observable for repetitions of DENTALS and PALATALS as well as combinations of DENTALS with PALATALS in both orders (T_Č and Č_T). This question is part of a broader problem concerning the distribution of CORONALS and NON-CORONALS, namely the effect called "extended SPA".

P&S noticed that their sample languages showed a tendency to avoid not only CVC combinations of the four primary place classes but also of two superclasses, that is, the combinations T_Č and Č_T involving different CORONALS and P_K and K_P involving different NON-CORONALS. To the best of our knowledge, this extended SPA effect has not been reported before P&S or afterward. However, the underrepresentation of said combinations is noticeable in some other studies (e.g., Mayer et al. 2010), though not explicitly named.

palatals is not the only possible one; palatals could also be interpreted as dorsals. Due to the lack of space, we will not consider this option, but see §5.4 and fn 2.

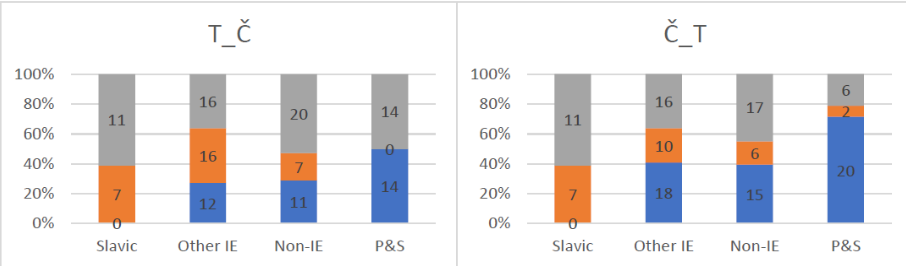

Figure 1. Graphs showing the proportion of the languages avoiding and not avoiding combinations T_Č and Č_T in our and P&S's language sample (Blue = avoidant languages, orange = favoring languages, gray = no significant inclination in either direction or the absence of such combinations. The statistics include –1 languages.)

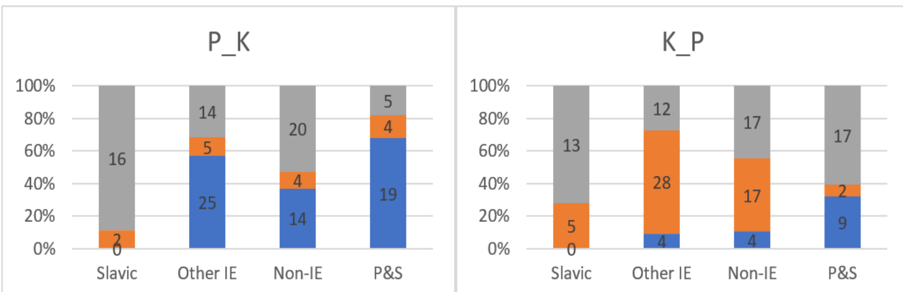

Figure 2. The same graphs for P_K and K_P (see Figure 1)

The attitude of our own languages to these combinations is visualized in Figure 1 (above), also including a re-analysis of the 28 non-IE languages from P&S. The graphs offer a comparison of the proportion of the varieties that significantly disfavor (blue) or significantly favor (orange) the combinations T_Č and Č_T, or in which there is no significant disproportion in either direction (gray; including two cases of the languages without palatals, Ancient and Modern Greek). Figure 2 gives the same for P_K and K_P. For simplicity, we do not report average Yule's Q values, though it is worth noting that all the combinations tend to be avoided with slightly more strength than they are preferred—with the exception of Slavic (e.g., T_Č has the average value of −0.31 for the avoidant non-IE languages in our sample and 0.17 for the favoring languages of the same group; but it is −0.15 against 0.22 for P_K in the same group).

It is immediately clear that no Slavic language avoids T_Č and Č_T; instead, a large portion of these languages *favors* them. In the case of T_Č, this includes all East Slavic languages together with Bulgarian, Serbo-Croatian, and Lower Sorbian; in the case of Č_T, the favoring languages are all South Slavic varieties together with Old Russian, Modern Czech, Slovak, and Proto-Slavic. The preference is apparently connected with their shared historical development. And it also seems to be an areal or genetic trait. Considering the situation in the other groups, we witness a noticeable inclination to avoid rather than prefer combinations of two different coronals. The proportion of the languages disfavoring T_Č and Č_T is largest in P&S's sample, of which only two languages happen to favor these sequences (Joola Kwaatay and Manjaku). Most of the avoidant languages are from Africa. Note also that the sequence Č_T tends to be avoided by more languages than its reverse, which has already been pointed out by P&S.

These observations go against the predictions of coronal SPA that unites dentals and palatals into one class. The proportion of the languages that avoid combinations of two different coronals is too small to warrant the conclusion that *any* combination of coronals is statistically underrepresented. The fact that coronal avoidance has been so widely reported must be due to the properties of dentals rather than coronals as a whole class. CVC sequences of dentals are almost uniformly avoided, but combinations of dentals with palatals happen to be avoided in a considerably smaller portion of languages. Many languages disprefer combinations of two palatals, but far from all. Moreover, some languages, such as the Slavic ones, tend to favor the mixed combinations of coronals.

Before moving to other topics, let us briefly look at Figure 2, which compares the proportions of the avoidant and favoring languages for the other combination types falling under the extended SPA principle, those involving two different non-coronals. The Slavic languages fail to provide evidence for this type of avoidance. Once again, no Slavic variety disfavors any of these combinations. Instead, P_K is favored by Slovak and Old Czech but not by Modern Czech (which is curious). K_P is favored by all South Slavic languages but Bulgarian, and by Upper Sorbian. And once again it is P&S's non-IE languages in which these two sequences are most widely avoided, especially P_K. However, comparing the proportion of avoidant and favoring languages, we can hardly claim that the non-Slavic IE languages and the non-IE languages in our own sample tend to underrepresent K_P. There are more languages favoring these sequences than languages avoiding them. Therefore, only P_K sequences are demonstrably avoided across language families, except Slavic. We do not have any explanation for this discrepancy.

In short, our data do not support the existence of the extended SPA principle in Slavic, and in general, there are more languages without this effect than those that impose restrictions on the combinability of CORONALS with NON-CORONALS.

5.2. Place-*cum*-Sonorancy Avoidance

Although many linguists have demonstrated the existence of coronal avoidance in several languages, we have repeatedly pointed out that the reader should be cautious about the meaning of the term *coronal*. It is often used as an equivalent to DENTALS either because (alveolo-)palatals or post-alveolars are not classified as DENTALS (see the following subsection) or because languages have very few (alveolo-)palatals and post-alveolars (e.g., Arabic, McCarthy 1994) or none (e.g., Maori, Rácz et al. 2016). Still, even these languages are characterized by a weaker coronal SPA effect, which is the same as saying that these languages have a weaker DENTAL SPA effect. And as demonstrated in the previous subsection, the effect is weaker even if the class of coronals includes alveolo-palatals, palatals, and post-alveolars. In short, no matter how coronals are defined, their avoidance tends not to be as strong as in other places.

It has been argued that the weakness or even the lack of a coronal SPA effect is a consequence of the large number of coronals present in phoneme inventories compared to the other place classes (Pierrehumbert 1993; Frisch et al. 2004). More distinctive features are therefore needed to capture differences between coronals. If similarity is measured by distinctive features, coronals are individually more dissimilar. And since SPA is essentially the avoidance of *similar* segments, additional, non-place features are expected to influence the coronal SPA effect.

Researchers have indeed reported that non-manner features influence SPA not only in the case of coronals but also other places. Padgett (1992) called them *subsidiary features*, claiming that they are, in a feature-geometric sense, dependent on and attached to the place node, which are the features [sonorant] and [continuant] in his conception. Yet phonological properties other than place have been shown to play a role in SPA (see, e.g., Berkley 1994, Dmitrieva 2008, and van Goch 2010, Rose and King 2007 on voicing, vowel type, stress, and distance). So, perhaps subsidiary features should simply be understood as SPA-influencing and -increasing properties (Wilson and Obdeyn 2009). The weight of such properties in SPA restrictions may be language-specific, though (see Coetzee and Pater 2008).

Padgett (1992) only formalized what had been known about the coronal SPA effect. Initially, the division of coronals by the feature [sonorant] was put forth for the first language in which SPA was described, Classical Arabic (McCarthy 1986, 1994).[21] This language restricts the occurrence of coronals

[21] As already mentioned, McCarthy (1994) and others (e.g., Padgett 1992 for Russian; see also the other references in this subsection) further demonstrate how the feature

within consonantal roots (already observed by Cantineau 1946 and Greenberg 1950). In the first two consonant positions, coronal obstruents freely combine with coronal sonorants, whereas repetitions of coronal obstruents and coronal sonorants are prohibited. The same division of coronals was confirmed in SPA studies of other Semitic languages (Akkadian by Reiner 1966; Tigrinya by Buckley 1997; Hebrew by Bachra 2001; Chaha and Amharic by Rose and King 2007) and other languages like Qafar (Hayward and Hayward 1989), Imdlawn Tashlhiyt Berber (Elmedlaoui 1995), Yamato Japanese (Kawahara et al. 2006), Shona and Wargamay (both in Wilson and Obdeyn 2009), Ngbaka (Danis 2017), Russian (Padgett 1992, 1995), English (Berkley 1994, 2000), and Dutch (van Goch 2010).

Our 100-language sample supports the observation that the avoidance of CORONALS depends on sonorancy, which we view as a difference between obstruents and sonorants. Eighty-six languages avoid repetitions of CORONAL obstruents (including all Slavic languages), and 92 avoid repetitions of CORONAL sonorants (including all Slavic languages except for Upper Sorbian). In contrast, combinations of CORONAL obstruents with CORONAL sonorants are avoided only in 29 languages (of which none is Slavic), and combinations of CORONAL sonorants with CORONAL obstruents are avoided in 15 languages (none is Indo-European!). These heterorganic CVC sequences are instead favored by 36 languages (obstruent + sonorant) and 66 languages (sonorant + obstruent). In both cases, this is true for virtually all Slavic languages (one positive Q value is not statistically significant).[22]

However, once again, we should be mindful of what these observations tell us. In the previous paragraph, we deliberately wrote "CORONALS" rather than "coronals" to indicate that we refer to a consonant class that may or may not include PALATALS (depending on whether a language has them), not just DENTALS. Thus, if combinations of CORONAL obstruents are indeed avoided, we expect that the avoidance concerns not just repetitions of DENTAL and PALATAL obstruents, and of DENTAL and PALATAL sonorants (on which see Table 13 on p. 213) but also combinations of the obstruents and sonorants without agreement in place. Our data provide weak support for these predictions.

Table 12 on the following page quantifies the extent and degree to which our languages avoid combinations of the CORONALS not agreeing in place

[continuant] affects SPA. We will not consider this problem here for lack of space. Suffice it to say that the influence this feature might have among coronals largely coincides with what we will call sibilancy avoidance in the following subsection.

[22] The sonorancy SPA effect for the 28 non-IE languages of P&S cannot be ascertained because we do not have information on the distribution of obstruents and sonorants. The authors only mention that "[their] measurements show that the dominant effect concerns PLACE OF ARTICULATION [emphasis original], not manner, nasality, or state of the glottis" (p. 342).

Table 12. Average Yule's Q values for CVC combinations of DENTALS and PALATALS with or without sonorancy agreement for a subsample of 86 languages

		Dental		Palatal	
		Obstruents	**Sonorants**	**Obstruents**	**Sonorants**
Dental	**Obstruents**			−0.35 (43) [2]	−0.29 (17) [1]
	Sonorants			−0.45 (3) [4]	−0.39 (16) [1]
Palatal	**Obstruents**	−0.31 (41) [4]	−0.15 (12) [3]		
	Sonorants	−0.34 (18) [1]	−0.37 (31) [0]		

(Shading marks sonorancy agreement. See Table 8, p. 196, on the round and square brackets and Table 13, p. 213, for the empty cells.)

and/or sonorancy. The calculations are limited to a subsample of 86 languages that possess at least one PALATAL obstruent and one PALATAL sonorant (each language has at least one DENTAL obstruent and sonorant). All Slavic varieties are included. The average Yule's Q is again calculated only for the languages that significantly avoid a given combination.

The table reveals several things. To begin with, the sonorant-first combinations are avoided more strongly than the combinations beginning with an obstruent. The sequences agreeing in manner/sonorancy are more widely disfavored than those without the agreement (see the shaded cells). The obstruent combinations are avoided in approximately 50% of the languages, whereas the sonorant combinations are in less than 35%. Finally, the sequences without agreement in place and manner are avoided in the smallest number of languages, never making a portion larger than 20% of the sample. In short, Table 12 supports our conclusion arrived at in the previous subsection: CVC combinations of DENTALS with PALATALS tend not to be avoided.

Even though there is little evidence for sonorancy having a substantial role in CORONAL avoidance, we must still account for the widely reported fact that nearly all the languages analyzed do show a place-*cum*-sonorancy avoidance effect for coronals (mind the typographical difference!). We offer an answer to this question in the next subsection, but first we would like to address another problem: Will SPA account better for the distribution of NON-CORONAL consonants when they are divided into subclasses by sonorancy?

Berkley (2000: 29) paid attention to this question in English, showing that all mutual combinations of labial obstruents and sonorants, as well as all combinations of velar obstruents and sonorants, are almost equally underrepresented irrespective of sonorancy agreement. Van Goch (2010: 79–82) reached the same conclusion for Dutch, and Padgett (1992, 1995) reported the same for labials in Russian. In contrast, Frisch et al. (2004: 192–93) state that there is some evidence for the subclassification of labials by manner in Classical Arabic, but the effect is not as strong as for coronals. Similarly, Coetzee and Pater (2008) indicate that sonorancy agreement influences, to some degree, the strength of SPA in Muna for all places.

Our material does not prove that SPA would be considerably influenced by sonorancy agreement. Table 13 on the following page compares the strength and extent of SPA as a function of sonorancy agreement for the four basic places. The values for labials and dentals are derived from the whole 100-language sample. Those for palatals and dorsals rely on fewer languages (86 and 33, respectively) because not all languages have obstruents and, especially, sonorants for these places.[23]

There is a striking difference between the dentals and the remaining places. While almost the same number of varieties avoid all combination types with very similar strength in the case of labials and palatals, the avoidance of dentals is more common and stronger when there is sonorancy agreement (contrast the shaded and unshaded cells). This fact immediately explains why many previous studies reported sonorancy-dependent avoidance of coronals: As pointed out above, coronals are usually equivalent to dentals either because the class of coronals excludes (alveolo-)palatals or post-alveolars or because the languages analyzed have a very small inventory of (alveolo-) palatals and post-alveolars.[24]

The way sonorancy influences the distribution of dorsals is less clear because our observations can be based on just 33 languages. The other languages either lack such segments or disallow CVC combinations of a dorsal sonorant with any other consonant. Yet the statistics of Table 13 do not indicate that dorsal SPA would be influenced by sonorancy agreement. It is only the presence of a dorsal sonorant (especially as the C_1 consonant) that lowers the number of languages in which sequences of two dorsals are avoided. On the other hand, sequences of two dorsal obstruents are avoided in the majority of these 33 languages. The combinations with at least one dorsal sonorant

[23] Since some of our source databases use phonetic transcription, some segments, like the velar nasal [ŋ], may be allophonic variants rather than separate phonemes. Note also that the dorsal subsample includes only those languages that have dorsal sonorants that combine with other consonants in the first word syllable. This is to say that our 100-language sample contains other languages with dorsal obstruents and sonorants, but in many the sonorants have very restricted combinability.

[24] Palatals once again stand out because their mutual combinations are absent in many languages (see the numbers in square brackets), even though the languages have both palatal obstruents and sonorants.

Table 13. Average Yule's Q values for homorganic CVC sequences with or without sonorancy agreement

		Labial Obst.	Labial Son.			Dental Obst.	Dental Son.
Labial	Obst.	−0.47 (76) [1]	−0.51 (80) [2]	Dental	Obst.	−0.27 (79)	−0.15 (28)
	Son.	−0.64 (75) [9]	−0.55 (76) [3]		Son.	−0.22 (11)	−0.51 (93)

		Palatal Obst.	Palatal Son.			Dorsal Obst.	Dorsal Son.
Palatal	Obst.	−0.41 (34) [7]	−0.41 (25) [13]	Dorsal	Obst.	−0.37 (30)	−0.35 (20)
	Son.	−0.49 (16) [23]	−0.55 (23) [19]		Son.	−0.43 (1) [7]	−0.66 (7) [14]

(The values for LABIALS and DENTALS are based on the total sample of 100 languages; the values for PALATALS on a subsample of 86 languages (the same as in Table 12); the values for DORSALS on a subsample of 33 languages. Shading marks sonorancy agreement. See Table 8 on the round and square brackets.)

are furthermore significantly preferred in 12 languages, of which six possess a velar (or uvular) rhotic. This includes Upper Sorbian, the only Slavic variety in this subsample. None of the languages that limit DORSAL sonorants to nasals prefer combinations containing a DORSAL sonorant. Hence, the preference is at least partially explainable by the presence of DORSAL trills.

5.3. Sibilancy Avoidance

Another explanation for the unusual distribution of homorganic sequences of DENTALS and PALATALS could reside in the contents of these identity classes. Recall that we define DENTALS as a class consisting of (inter)dental, (pre-)alveolar, and retroflex consonants, whereas the class of PALATALS comprises post-alveolars, alveolo-palatals, and true palatals. Although we call them collectively CORONALS, not all consonant classes mentioned have been characterized as coronals (e.g., by sharing the articulation of the tongue blade or by the feature [+coronal]). Since alveolo-palatals of the kind found in Polish (e.g., /ɕ/) are poorly represented in our language sample, we will not consider how their reclassification impacts the SPA effect. Likewise, we will not go into the problem of palatal fricatives such as /ç/ and /ʝ/ in German and elsewhere, which Hall (1997: 15ff.) interpreted as non-coronals. They, too, are poorly represented in our data. Instead, we can look into another question suggested by one of the reviewers, namely the status of post-alveolars (palato-alveolars), which always happen to be sibilants.

First of all, it turns out that removing post-alveolars from the class of palatals does not improve the SPA statistic for this identity class. There are 14 languages in our sample for which this can be reliably tested. They limit their palatal obstruent inventory to true palatal plosives and post-alveolar sibilants.[25] One half of these languages avoid the CVC sequences that contain either true palatal plosives or post-alveolar sibilants; no language favors these sequences. Sequences just containing true palatal plosives are avoided in five languages. To be more precise, five languages lack such combinations, which may be a lacuna in our lexical data rather than total avoidance. Three languages significantly favor repetitions of palatal plosives (Modern Czech, Latvian, and Hungarian). In short, if the class of palatals includes post-alveolar sibilants, more languages avoid palatal CVC sequences than when the sibilants are not part of the class.

We are better positioned to assess how the avoidance effect is changed when post-alveolar sibilants are classified as dentals. Our sample contains 74 languages with post-alveolar sibilants. When these segments are treated as dentals, the repetitions of dental obstruents happen to be underrepresented in more languages than when post-alveolars are members of the class (63 instead of 60). Furthermore, the avoidance strength measured by averaged Yule's Q increases from −0.26 to −0.28. Though slight, the increase can be sufficient proof that post-alveolar sibilants pattern with dento-alveolars rather than palatals. However, the increase is actually a manifestation of another type of avoidance not mentioned in the literature: the avoidance of repetitions of sibilants in CVC sequences.

To demonstrate this restriction, let us consider a smaller selection of the 47 languages that have both (pre-)alveolar and post-alveolar sibilants and no other coronal affricates and fricatives to affect the statistics; all these languages have dental/alveolar plosives too. Thus, the 47-language subsample excludes English, which contains the non-sibilant fricatives /θ/ and /ð/, or Polish, in which we find alveolo-palatal sibilants. Table 14 on the following page re-

---

[25] Four other languages have true palatal plosives and post-alveolar sibilants, but they also have true palatal fricatives. Since some linguists consider these fricatives to be non-coronals (Hall 1997), we exclude these four languages from the statistics. Also, we do not consider how palatal obstruents, however they are delimited, combine with palatal sonorants, as sonorancy agreement may play a role. To put it simply, the 14 languages we describe here are those belonging in types P4, P8, and P14 (see Table 9, p. 199).

Table 14. Average Yule's Q for combinations of dento-alveolar plosives with dento-alveolar sibilants and post-alveolar sibilants

	Dento-alv. plosives	**Dento-alv. sibilants**	**Post-alv. sibilants**
Dento-alv. plosives	−0.28 (32)	−0.30 (22) [2]	−0.24 (13) [2]
Dento-alv. sibilants	−0.28 (21) [1]	−0.41 (36) [3]	−0.47 (20) [5]
Post-alv. sibilants	−0.27 (16) [1]	−0.49 (21) [6]	−0.42 (19) [5]

(The values are calculated for 47 languages. Shading marks sibilancy avoidance. See Table 8 on p. 196 on the round and square brackets.)

produces average Yule's Q values for all mutual CVC combinations of dento-alveolar plosives, dento-alveolar sibilants, and post-alveolar sibilants.

The table demonstrates two things. First, the combinations involving no more than one sibilant (the unshaded cells) are underrepresented in fewer languages and less strongly than the combinations involving two sibilants (the shaded cells). Two-sibilant combinations are even absent in many languages (the number of languages is given in the square brackets). Second, the combinations of plosives with sibilants (affricates and/or fricatives) are disfavored in more languages and in a stronger manner when they agree in the place of articulation than when there is no such agreement, that is, when both are dento-alveolars rather than a combination of a dento-alveolar and a post-alveolar. To put it otherwise, if we set aside combinations of sibilants, the avoidance of homorganic sequences of DENTAL obstruents is indeed less strong than the avoidance of LABIALS and DORSALS (refer to Table 10, p. 203). The inclusion of post-alveolars in the class of DENTALS will increase the DENTAL avoidance strength, but only because many languages disfavor repetitions of sibilants, obviously due to their acoustic and perceptual similarity (cf. Kokkelmans 2021). We find it more appropriate to assume the existence of sibilancy avoidance as a special kind of similarity avoidance rather than to unite post-alveolar sibilants with dento-alveolars instead of palatals.

There are two other, interrelated pieces of evidence for treating post-alveolars as belonging to the distributional class of PALATALS, at least in Slavic languages. First, almost all Slavic languages significantly and strongly avoid CV sequences of PALATALS and back vowels; the only exception is Polabian, in which Yule's Q for such sequences is not significant. This avoidance is observable both for combinations with true palatals (= plosives and nasals in

Table 15. Yule's Q for CV combinations of PALATALS with back vowels in Slavic languages

	Bel	Bul	LSo	Mac	MCz	MRu	OCS	OCz	ORu
Plo	×	×	×	absent	−0.87	×	absent	1.00	×
Aff	−0.58	−0.51	−0.73	−0.59	−0.70	−0.74	−0.82	absent	absent
Fri	(−0.17)	−0.53	−0.49	−0.54	−0.57	−0.58	−0.63	−0.72	−0.46
Nas	×	×	×	absent	−0.77	×	absent	1.00	×
Son	−0.52	(0.02)	−0.42	−0.49	−0.93	−0.53	−0.35	−0.67	−0.60
	Plb	**Pol**	**Pom**	**PSl**	**SCr**	**Slk**	**Sln**	**Ukr**	**USo**
Plo	×	×	×	×	×	absent	×	×	×
Aff	×	absent	(−0.14)	−0.88	−0.49	−0.47	−0.82	−0.57	−0.80
Fri	(0.23)	−0.57	−0.61	−0.75	−0.59	−0.55	−0.51	−0.31	−0.64
Nas	×	absent	−0.86	absent	(−0.60)	absent	×	×	×
Son	(0.24)	−0.79	−0.61	0.71	−0.29	(−0.15)	(0.06)	−0.49	−0.64

(Parenthesized values are not significant. Based on the same data as Table 4. "Absent" means a language lacks a certain combination of PALATALS even though it has these consonants.)

Table 16. Average Yule's Q values for CVC sequences unaffected by pure and extended SPA

	Slavic (n = 18)	Non-Slavic IE (n = 44)	Non-IE ours (n = 38)	Non-IE P&S P&S (n = 28)
P_T	0.29 (17)	0.35 (44)	0.24 (32)	0.27 (21)
T_P	0.35 (18)	0.32 (43)	0.25 (31)	0.26 (25)
T_K	0.19 (12)	0.29 (41)	0.22 (34)	0.29 (24)
K_T	0.19 (6)	0.20 (27)	0.23 (28)	0.30 (22)
P_Č	× (0)	0.17 (18)	0.23 (17)	0.28 (19)
Č_P	× (0)	0.22 (17)	0.22 (15)	0.23 (18)
Č_K	× (0)	0.32 (17)	0.28 (15)	0.26 (16)
K_Č	× (0)	0.16 (5)	0.24 (9)	0.17 (11)

(The parenthesized numbers indicate how many languages with significant *positive* Q values have been used for the calculation.)

Table 15 on the previous page) and post-alveolar sibilants (= affricates and fricatives in Table 15).[26]

The second piece of evidence comes from Modern Czech. This language (and probably other Slavic languages, too) not only disfavors combinations of palatals with back vowels but severely limits their distribution, which concerns both the true palatals /c/, /ɟ/, /ɲ/, and /j/ as well as the post-alveolars /t͡ʃ/, /d͡ʒ/, /ʃ/, /ʒ/, and /r̝/ (ř). Both types of combinations are mostly found across morpheme boundaries, in loanwords, or expressive and onomatopoeic vocabulary (Mathesius 1931). In other words, these CV combinations are virtually absent within native-origin morphemes; the few exceptions there are could be etymologically explained as being old and obscured instances of the three contexts mentioned (Bičan 2018 and Bičan, forthcoming).

5.4. Labial-Coronal Effect

The four basic place classes of labials, dentals, palatals, and dorsals produce 16 possible CVC combinations. The SPA principle, as formulated by P&S, divides them into three groups. One consists of homorganic combinations directly affected by pure SPA, the main research focus of this paper. The second group consists of coronal combinations of dentals with palatals and of non-coronal combinations of labials with dorsals. We saw in §5.1 that some languages disfavored these sequences, which phenomenon P&S had called extended SPA. The final group includes all the remaining combinations. It follows from the nature of SPA that if some sequences are disfavored/underrepresented, others must be favored/overrepresented. It is the third group of combinations that strongly inclines to this preference.

Table 16, shown on the previous page, provides average *positive* Yule's Q values for the eight kinds of combinations of non-coronals and coronals, which belong to the third group and which P&S found to be preferred cross-linguistically. This preference is confirmed by our sample languages and the recalculation of P&S's data, though again less obviously in Slavic. The eight sequence types break down into three bundles, which division is most obvious among the Slavic languages.

The first consists of combinations of labials with dentals (P_T and T_P). The preference for these sequences is cross-linguistically the strongest and most diffused one. Of the IE languages, it is only Pomeranian and Pashto for which this avoidance is not statistically significant, which may be due to the small size of our lexical data. The preference was already well spotted by P&S, but they did not explain it. It is a manifestation of a so-called labial-coronal

[26] The statistics are a little distorted for Polish, Pomeranian, and Serbo-Croatian because these languages also have alveolo-palatal affricates and fricatives, which have been included in the classes Aff and Fri in the table.

or fronting effect. Numerous cross-linguistic studies have demonstrated that sequences of a (bi)labial and a coronal are preferred both in infant and adult languages (see overviews in MacNeilage and Davis 2000 or Carrissimo-Bertola 2010). Since Table 16 on p. 216 indicates that the effect pertains little to palatals, the tendency should be dubbed a labial-dental effect instead. In fact, palatals were not included in the coronal class in the studies mentioned.

On the other hand, we do not notice any straightforward tendency to favor or disfavor the remaining combinations. The second bundle of sequences includes combinations of dentals with dorsals (T_K, K_T). They are favored by fewer languages and less strongly, especially among the IE languages. In the case of K_T, the drop in favor may be connected to there being relatively many languages that significantly disfavor these sequences. Since none of these 12 languages are from Slavic, the explanation actually fails. Neither can we explain the attitude toward the T_K sequences (they are disfavored by three languages, again none being Slavic). The last bundle of sequences consists of those involving a palatal combined with a non-coronal. None of these combinations are favored in Slavic languages, and the non-Slavic languages also tend to favor them to a lesser degree. Instead, these sequences happen to be disfavored in not a small number of languages. It may partly be diachronic hold-over, especially in the case of Č_K and K_Č in Slavic. It is well known that Slavic palatalizations produced postalveolar sibilants from velars; a similar mutation of velars into palatals is attested elsewhere (Bateman 2007). In other words, many palatal segments were originally velars, so the avoidance of Č_K and K_Č might have originally been the avoidance of K_K.[27]

6. Conclusion

Similar Place Avoidance (SPA) or its close variant, Obligatory Contour Principle (OCP or OCP-Place), is, besides the Sonority Sequencing Principle, perhaps the best-known and studied phonotactic constraint shared by many natural languages. Yet the studies do not agree on the formulation of the SPA principle, which slightly diminishes its power and casts doubt on its universality. We calculated the SPA statistics for the lexicons of 100 language varieties, of which 18 are Slavic, 48 non-Slavic Indo-European, and 38 non-IE languages. These languages avoid CVC combinations of consonants at the same place of articulation, but the situation is more complex than the existing SPA models suggest.

SPA models set up several consonant classes (identity classes), assuming that the CVC combinations of same-class consonants are significantly

[27] It is then not surprising that Padgett (1992) and Pozdniakov (2007, 2010) found that combinations of dorsals with /t͡ʃ/, /ʃ/, and /ʒ/ are underrepresented in Russian. Our data are less suggestive of this kind of avoidance.

underrepresented, whereas different-class combinations are not. The difference between the models concerns whether coronal consonants are divided or grouped by place features only or whether non-place features such as sonorancy interfere. Repetitions of labials and dorsals are, with very few exceptions, avoided in all languages. The subdivision of labials by sonorancy does not improve the model because all four logically possible combinations of labial obstruents and sonorants are avoided in approximately the same number of languages. The same could be said about dorsals, though the rarity of dorsal sonorants distorts the picture. However, subdivision by sonorancy reveals that repetitions of sonorants agreeing in place are widely and strongly avoided cross-linguistically, which is true not only for dorsals but also for dentals and palatals, and as already mentioned, for labials too.

Table 17 on the following page visualizes the three most common models used in typological studies of SPA or recurring in the literature on SPA or OCP (model A in P&S; B in Mayer et al. 2010; C in several studies on Semitic languages). They disagree on the number and definition of the consonant classes for which consonant co-occurrence restrictions hold. Analyses of singular languages may work with more classes (e.g., velars and gutturals separated in Arabic) or fewer classes (e.g., Greek lacks palatals). The first, non-parenthesized values are averaged Yule's Q coefficients, measuring the association strength between two consonants in a CVC sequence or the degree to which the sequences are avoided. The second, parenthesized values are percentages of the languages from three groups (Slavic, non-Slavic IE, and non-IE) that significantly avoid a given sequence.

None of the three models adequately describes the distribution of coronal CVC sequences in Slavic and other languages (unless model C is equivalent to model D due to the lack of palatals; see below). Model B unites dentals and palatals into one class of coronals but leads to wrong predictions about the distribution of the united class. It is apparent from model A that palatals are markedly less affected by similarity avoidance than dentals. Homorganic sequences of palatals even tend to be favored in some languages (including Polish and Lower Sorbian), which is not evidenced for dentals. Furthermore, CVC combinations of coronals in which one consonant is dental and the second palatal (in either order) are not avoided in any Slavic language, and the same is true for around 60% of the other languages. Model B is thus not superior to model A.

The third model, C, the most common alternative to A and the original form of the OCP-Place principle, divides coronals into obstruents and sonorants. Repetitions of coronal obstruents and coronal sonorants are underrepresented in all Slavic languages and most other languages (several non-Slavic IE and non-IE languages do not avoid sequences of coronal obstruents). At the same time, combinations of coronal obstruents with coronal sonorants tend not to be avoided—in fact, they are favored across all language groups.

Still, this classification must be rejected for two reasons. The first follows from the same but elaborated argument against model B: Combinations of CORONAL obstruents agreeing in sonorancy but not in place are avoided in no more than 50% of Slavic languages; even elsewhere the avoidance is not much diffused. Still, we should note that most languages disfavor combinations of DENTAL obstruents with PALATAL obstruents (in both orders). It seems to be a consequence of sibilancy avoidance, that is, the dispreference of CVC sequences in which the consonants are dental, (pre-)alveolar, post-alveolar, and possibly also alveolo-palatal sibilants (but on Polish, see Table 6 on p. 194). Though not reported previously, a strong sibilancy avoidance effect is observable in many languages (but less obviously in the Slavic group).

Table 17. Three rejected SPA models

A	LABIALS	DENTALS	PALATALS	DORSALS
	S: −0.51 (100%)	S: −0.30 (100%)	S: −0.27 (16.67%)	S: −0.43 (77.78%)
	I: −0.58 (100%)	I: −0.36 (97.73%)	I: −0.37 (50.00%*)	I: −0.43 (95.45%)
	N: −0.49 (92.10%)	N: −0.26 (100%)	N: −0.30 (55.26%)	N: −0.38 (92.11%)

B	LABIALS	CORONALS	DORSALS
	S: −0.51 (100%)	S: −0.29 (100%)	S: −0.43 (77.78%)
	I: −0.58 (100%)	I: −0.39 (97.73%)	I: −0.43 (95.45%)
	N: −0.49 (92.10%)	N: −0.35 (97.37%)	N: −0.38 (92.11%)

C	LABIALS	CORONAL OBSTRUENTS	DORSALS
	S: −0.51 (100%)	S: −0.31 (100%)	S: −0.43 (77.78%)
	I: −0.58 (100%)	I: −0.29 (90.91%)	I: −0.43 (95.45%)
	N: −0.49 (92.10%)	N: −0.23 (73.68%)	N: −0.38 (92.11%)
		CORONAL SONORANTS	
		S: −0.56 (94.44%)	
		I: −0.51 (100%)	
		N: −0.35 (81.58%)	

(The first numerical values are the average Yule's Q values for 18 Slavic ("S"), 44 non-Slavic IE ("I"), and 38 non-IE ("N") languages. Parenthesized values are percentages of the languages significantly avoiding a given homorganic sequence. The asterisked percentage includes those languages that lack PALATAL sequences altogether.)

Table 18. The proposed SPA model

D	labials	dental obst.	palatals	dorsals
	S: −0.51 (100%)	S: −0.32 (94.44%)	S: −0.27 (16.67%)	S: −0.43 (77.78%)
	I: −0.58 (100%)	I: −0.26 (90.91%)	I: −0.37 (50.00%*)	I: −0.43 (95.45%)
	N: −0.49 (92.10%)	N: −0.26 (57.89%)	N: −0.30 (55.26%)	N: −0.38 (92.11%)
		dental son.		
		S: −0.63 (94.44%)		
		I: −0.56 (100%)		
		N: −0.44 (84.21%)		

(The numerical values are average Yule's Q values for 18 Slavic ("S"), 44 non-Slavic IE ("I"), and 38 non-IE languages ("N"). Parenthesized values are percentages of the languages significantly avoiding a given homorganic sequence. The asterisked percentage includes those languages that lack palatal sequences altogether.)

The second reason for rejecting model C concerns the predominance of dentals within the class of coronals. The observation that the sonorancy agreement determines the strength and diffusion of SPA is valid, but the agreement affects only dentals, not palatals. Irrespective of sonorancy agreement, combinations of palatals are avoided by almost the same number of languages (again, less obviously in Slavic, but these languages do not restrict homorganic CVC sequences of palatals in general). However, since most coronals are, in fact, dentals (including alveolars), the sonorancy agreement effect manifests itself even if palatals are included in statistical calculations.

The objections against models B and C and the combinability of dentals and palatals lead us to the fourth model of SPA reproduced in Table 18, shown above. Here, it is only dentals that are divided by sonorancy. The Slavic languages fit this model perfectly as they avoid both dental subclasses, but even the non-Slavic IE languages show a high degree of dispreference, particularly for repetitions of dental sonorants.

Model D is a combination of model A, originally introduced by P&S for African languages, and model C, derived from the phonotactics of Semitic languages. However, its validity for non-IE languages in general is only tentative. If there is a clear correlation between avoidance and preference within the IE languages, the contrast is less sharp in the non-IE language group. In the IE languages the sonorancy-agreeing combinations tend to be avoided and not favored, whereas the sonorancy-disagreeing combinations tend not to be avoided; instead, they tend to be favored. In contrast, there are many non-IE languages that avoid the sonoracy-disagreeing combinations and few languages that prefer them. For example, combinations of DENTAL obstruents with DENTAL sonorants are underrepresented in 15 non-IE languages (39% of the 38 languages!), but they are overrepresented only in five languages. Thus, other factors may influence the distribution of DENTAL CVC sequences, and it is quite likely that the form and strength of the avoidance affect is to a large extent language-specific.

Be that as it may, it remains an open question why DENTAL SPA is so prominently affected by sonorancy agreement, at least in some languages. It has been suggested that this has to do with the size of coronal (i.e., DENTAL) inventories. More segments within this class require more distinctive features to distinguish them from each other and other segments. Consequently, "similarity for coronal pairs like /s, n/ is less than the similarity of /f, m/ due to the larger space of contrasts in the coronals[; as] a result, the co-occurrence restriction for /f, m/ is stronger than the co-occurrence restriction for /s, n/" (Frisch et al. 2004: 199; see also Pierrehumbert 1993: 375). However, while we agree that answers to the mystery of DENTALS should be sought in the richness and commonality of these segments, we reject the explanation by distinctive features. It can only be valid if we agree that it is distinctive features (in general or in certain conceptions) and/or phonological contrasts that truly express similarities between segments rather than their articulatory and acoustic properties (cf. Ohala 1990). Moreover, if SPA in general is to be explained by language perception and processing (as it usually is, see Introduction), then we should be able to demonstrate that, say, /f/ and /m/ are indeed perceived as more similar than /s/ and /n/.

Other mysteries lie behind the behavior of PALATALS. First, homorganic CVC sequences of these consonants are avoided only in three out of 18 Slavic varieties. Although these sequences are not universally disfavored even in other language groups, the portion of the avoidant languages is still larger than in Slavic—58% of the other languages (including those in P&S's sample) show a PALATAL avoidance effect or lack PALATAL sequences altogether. The contrast may partly be due to there being no Slavic language without post-alveolar sibilants (a possible exception may be Polabian once we leave out loanwords). The languages lacking these sibilants tend to avoid PALATAL sequences (22 non-Slavic languages are like this, 18 of which have an avoidance effect). However, there are striking differences among the languages that limit their PALATALS to post-alveolar sibilants and /j/, and the languages that have post-alveolar sibilants, /j/, other sonorants such as /ɲ/, and true palatal plosives such as /c/. Both types are among the most frequent PALATAL systems cross-linguistically; in Slavic they are the two most frequent systems. The vast majority of the non-Slavic IE and non-IE languages of both kinds have an avoidance effect for the palatal sequences, whereas the effect has been found only for two Slavic languages out of fourteen.

The second mystery concerns the very lack of palatal avoidance. Why do many languages fail to impose distributional restrictions on the repetitions of palatals when most of these languages avoid homorganic sequences of labials, dorsals, and dentals (dental obstruents and dental sonorants)?[28] It is possible that the lack and/or weakness of palatal SPA relates to the origin of palatal consonants. They have often arisen from the palatalization of dorsals or dentals/alveolars before front vowels (Bateman 2007). Also, the approximant /j/ is a favored epenthetic and prothetic segment (Blevins 2008), or it might result from the dissolution of palatalized consonants (cf. Proto-Slavic **běžati* 'to run' with non-homorganic /běž/ > Modern Czech /bjɛʒɛt/ where /bj/ is from palatalized /bʲ/). Hence, homorganic palatal sequences were originally non-homorganic, and the lack of SPA may reflect this original state. It is then no surprise that there are other restrictions on the distribution of palatals that do not concern their co-occurrence with other consonants, but rather with vowels. One such restriction is the avoidance of immediate combinations of palatals with back vowels (see §5.3).

There is one final issue we would like to address briefly here. Although languages rather strongly and widely avoid CVC sequences with consonants from the same place of articulation, this avoidance is not a strict prohibition. Homorganic CVC sequences exist in all languages so far considered. The obvious question is what SPA-violating words are like. The most often cited examples are CVC sequences with identical consonants. In some languages, these are claimed to be the only violations possible (e.g., in Chol, as shown by Gallagher and Coon 2009; see also Rose 2000), for which reason some researchers ascribed a special status to repetitions of identical consonants within the SPA restrictions (Frisch 2004; Frisch et al. 2004). Yet, the morphological, lexicological, and etymological nature of SPA violations remain little explored. Possible SPA-violating lexical items and morphosyntactic contexts are summarized in P&S; they mention ideophones and other sound-symbolic vocabulary, child-language words and expressive lexical items in general, loanwords, morphological reduplications, and other morpheme-boundary sequences. Pozdniakov (2007) subsequently provided a detailed but partial lexicological analysis of SPA-violating CVC sequences in Modern Russian.

Our analysis of Slavic sheds some light on the nature of SPA-violating items and partly confirms the previous observations. To begin with, the question of loanwords cannot be adequately resolved within IE languages because these languages have mostly borrowed words from each other, that is, from languages restricted by SPA. Secondly, the problem of morpheme boundaries

[28] Another mystery is why the sequences of dental obstruents are not avoided in so many non-IE languages (see Table 18, p. 221).

is more complex than suggested. The situation in Modern Czech indicates that the same kind of SPA restrictions apply for roots and across suffixal boundaries, but no SPA effect is observable at prefixal boundaries. Interestingly, a labial and dorsal SPA effect is consistently present across word boundaries in the poetry of some writers, but neither effect has been observed at word boundaries in prose.

We have also looked at the origin, meaning, and function of the 1,256 native roots with a reliable etymology that contained a CVC sequence of labials, dental obstruents, dental sonorants, palatals, and dorsals. The roots had been excerpted from Modern Czech, Old Church Slavonic, and Slovenian (including many verbal roots from Czech dialects). As shown below in Table 19, around 60% of the labial and dorsal homorganic sequences occur in phonologically motivated (sound-symbolic) roots, which includes onomatopoeia, phonaesthematic or expressive formations, and child-language formations. The share of these roots is much smaller for dentals and palatals. The homorganic CVC sequences of these consonants are mostly secondary developments rather than the original building blocks of the roots. They arose from ancient morphological processes, of which reduplication and suffixation are most common (= "different domain" in Table 19) or are products of sound changes like assimilation or lenition (= "phonological changes"), or both. Labial and dorsal sequences also arose from these processes, but to a much smaller degree. Only dental sequences occur in roots inherited from Proto-Slavic or Proto-Indo-European in the unchanged phonological and morphological form (= "unmotivated original").

This etymological analysis of Slavic confirms that homorganic CVC sequences appear in phonologically motivated vocabulary or are products of sound changes and morphological processes but also reveals that there is a

Table 19. Origin of native Slavic roots with homorganic CVC sequences (n = 1,256)

	N	Unmotivated non-original			Unmotivated orig.	Motivated
		Dif. domain	Phon. changes	Both		
labials	188	20.7%	9.6%	5.3%	0.0%	64.4%
dorsals	115	16.5%	7.8%	3.5%	0.0%	72.2%
dent. obs.	244	34.0%	25.4%	5.3%	4.5%	30.7%
dent. son.	290	40.3%	5.5%	24.8%	1.0%	28.3%
palatals	419	8.1%	42.7%	17.7%	0.0%	31.5%

sharp and statistically significant difference between homorganic sequences of non-coronals (labials and palatals) and coronals (dentals and palatals), which has been advocated throughout this paper. The former are much more inclined to be conveyors of sound symbolism than the latter. The former are also more strongly avoided cross-linguistically. The least associated with sound symbolism are homorganic sequences of palatals, which are also least often avoided cross-linguistically.

Sound symbolism could then be a place to look for an explanation of the differences in SPA strength. Non-coronals may be better predisposed for mimicking the extra-linguistic world than coronals; labials perhaps thanks to their visual saliency (cf. Kumagai 2020 and Kilpatrick et al. 2023), dorsals thanks to their acoustics (cf. Barrera-Pardo 2013 and Elsen 2017). It has been repeatedly pointed out that sound-symbolic vocabulary often features phonological abnormalities (e.g., Kořínek 1934 or Haiman 2018: 114–16). Sound symbolism and phonological peculiarity may go hand in hand, enhancing each other. If so, CVC sequences of labials and dorsals may be more strongly avoided than those of dentals so that their sound-symbolic capacity is more prominent and apparent among otherwise entirely arbitrary signs. However, it is also possible to offer an alternative explanation for the results in Table 19: non-coronals are less likely to be results of sound changes than coronals, and non-coronals are also less likely to figure in word-formative processes of Proto-Slavic and Proto-Indo-European roots. In short, we witness the effectiveness of language here in that some patterns are avoided only to be utilized more prominently elsewhere.

References

Arvaniti, Amalia. (2007) "Greek phonetics: The state of the art". *Journal of Greek linguistics* 8: 97–208.

Bachra, Bernard N. (2001) *The phonological structure of the verbal roots in Arabic and Hebrew*. Leiden/Boston/Köln: Brill.

Barrera-Pardo, Darío. (2013) "The sound symbolism of velar sounds in English popular fiction". Eva Parra-Membrives and Albrecht Classen, eds. *Literatur am Rand: Perspectiven der Trivialliteratur vom Mittelalter bis 21. Jahrhundert*. Tübingen: Narr Verlag, 47–62.

Basbøll, Hans. (2005) *The phonology of Danish*. Oxford: Oxford University Press.

Bateman, Nicoleta. (2007) *A crosslinguistic investigation of palatalization*. PhD dissertation, University of California.

Berent, Iris, and Joseph Shimron. (1997) "The representation of Hebrew words: Evidence from the Obligatory Contour Principle". *Cognition* 64: 39–72.

Berkley, Deborah M. (1994) "Variability in Obligatory Contour Principle effects". Katherine Beals, Jeannette Denton, Robert Knippen, Lynette Melnar, Hisami Suzuki, and Erica Zeinfeld, eds. *The parasession on variation in linguistic theory*. Vol. 2 of *Papers from the 30th regional meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 1–12.

———. (2000) *Gradient Obligatory Contour Principle*. PhD dissertation, Northwestern University.

Bičan, Aleš. (2018) "Etymologie a fonologie: případ Mathesiova fonotaktického pravidla" [Etymology and phonology: The case of Mathesius's phonotactic rule]. Petr Malčík, ed. *Vesper Slavicus: Sborník k nedožitým devadesátinám prof. Radoslava Večerky*. Praha: Nakladatelství Lidové noviny, 13–31.

———. (forthcoming) "Consonant-vowel combinations in Slavic and other languages".

Blevins, Juliette. (2008) "Consonant epenthesis: Natural and unnatural histories". Jeff Good, ed. *Linguistic universals and language change*. Oxford: Oxford University Press, 79–107.

Boll-Avetisyan, Natalie, and Réne Kager. (2014) "OCP-PLACE in speech segmentation". *Language and speech* 57(3): 294–421.

Booij, Geert. (2011) "Morpheme structure constraints". Mark Van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, eds. *Phonological interfaces*. Vol. 4 of *The Blackwell companion to phonology*. Oxford: Blackwell, 2049–69.

Buckley, Eugene. (1997) "Tigrinya root consonants and the OCP". *Penn working papers in linguistics* 4(3): 19–51.

Cantineau, Jean. (1946) "Esquisse d'une phonologie de l'Arabe Classique". *Bulletin de la Société Linguistique* 43: 93–140.

Carrissimo-Bertola, Manon. (2010) *Structures syllabiques des unités lexicales: "The fronting effect"*. MA thesis, Université Stendhal.

Cathcart, Chundra A. (2023) "The evolution of similarity avoidance: A phylogenetic approach to phonotactic change". Noah Elkins, Bruce Hayes, Jinyoung Jo, and Jian-Leat Siah, eds. *Proceedings of the 2022 Annual meeting on phonology*, 1–12. Available at: https://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/5434/5206. Last accessed 5 February 2024.

Coetzee, Andries W. (2005) "The Obligatory Contour Principle in the perception of English". Sonia Frota, Marina Vigário, and Maria João Freitas, eds. *Prosodies*. Berlin/New York: De Gruyter Mouton, 223–45.

Coetzee, Andries W., and Joe Pater. (2008) "Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic". *Natural language and linguistic theory* 26: 289–337.

Cooper, Adam I. (2009) "Similarity avoidance in the Proto-Indo-European root". *Penn working papers in linguistics* 15(1): 55–64.

Danis, Nicholas S. (2017) *Complex place and place identity*. PhD dissertation, State University of New Jersey.

Davis, Stuart. (1991) "Coronals and the phonotactics of nonadjacent consonants in English". Carole Paradis and Jean-François Prunet, eds. *Phonetics and phonology 2: The special status of coronals*. New York: Academic Press, 49–60.

Derksen, Rick. (2008) *Etymological dictionary of the Slavic inherited lexicon*. Leiden/Boston: Brill.

Dmitrieva, Olga. (2008) "The gradient phonotactics of English CVC syllables". Unpublished manuscript. Available at: https://web.ics.purdue.edu/~odmitrie/papers/Dmitrieva2007_OnEnglishOCP.pdf. Last accessed 5 February 2024.

Doucette, Amanda, Timothy J. O'Donnell, Morgan Sonderegger, and Heather Goad. (2024) "Investigating the universality of consonant and vowel co-occurrence restrictions". *Glossa* 9(1): 1–39.

Elmedlaoui, Mohamed. (1995) "Géométrie des restrictions de cooccurrence de traits en sémitique et en berbère: synchronie et diachronie". *Canadian journal of linguistics/Revue canadienne de linguistique* 40(1): 39–76.

Elsen, Hilke. (2017) "The two meanings of sound symbolism". *Open linguistics* 3: 491–99.

Frisch, Stefan A. (1997) *Similarity and frequency in phonology*. PhD dissertation, Northwestern University.

———. (2004) "Language processing and segmental OCP effects". Bruce Hayes, Robert Kirchner, and Donca Steriade, eds. *Phonetically based phonology*. Cambridge, UK: Cambridge University Press, 346–71.

Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. (2004) "Similarity avoidance and the OCP". *Natural language and linguistic theory* 22: 179–228.

Frisch, Stefan A., and Bushra A. Zawaydeh. (2001) "The psychological reality of OCP-place in Arabic". *Language* 77(1): 91–106.

Gallagher, Gillian. (2015) "Natural classes in cooccurrence constraints". *Lingua* 166: 80–98.

Gallagher, Gillian, and Jessica Coon. (2009) "Distinguishing total and partial identity: Evidence from Chol". *Natural language and linguistic theory* 27: 545–82.

Graff, Peter N. H. M. (2012) *Communicative efficiency in the lexicon*. PhD dissertation, Massachusetts Institute of Technology.

Greenberg, Joseph A. (1950) "The patterning of root morphemes in Semitic". *Word* 6: 162–81.

Grotberg, April L. (2022) *Quantifying phonological feature co-occurrence*. PhD dissertation, Purdue University.

Haiman, John. (2018) *Ideophones and the evolution of language*. Cambridge, UK: Cambridge University Press.

Hall, T. Alan. (1997) *The phonology of coronals*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hamann, Silke. (2002) "Postalveolar fricatives in Slavic languages as retroflexes". Sergio Baauw, Mike Huiskes, and Maaike Schoorlemme, eds. *OTS yearbook 2002*. Utrecht: Utrecht Institute of Linguistics, 105–27.

Haspelmath, Martin, and Uri Tadmor, eds. (2009) *World loanword database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: http://wold.clld.org. Last accessed 1 November 2024.

Havlová, Eva, Adolf Erhart, and Ilona Janyšková, eds. (1989–2022) *Etymologický slovník jazyka staroslověnského* [Etymological dictionary of the Old Church Slavonic language], vols. 1–21. Prague: Karolinum; Brno: Tribun.

Hayward, Katrina M., and Richart J. Hayward. (1989) "'Guttural': Arguments for a new distinctive feature". *Transactions of the Philological Society* 87(2): 179–93.

Hollander, Myles, Douglas A. Wolfe, and Eric Chicken. (2014) *Nonparametric statistical method*. 3rd edition. New Jersey: Wiley.

Howson, Phil. (2017) "Upper Sorbian". *Journal of the International Phonetic Association* 47(3): 359–67.

Hualde, José Ignacio. (2005) "Quasi-phonemic contrasts in Spanish". Benjamin Schmeiser, Vineeta Chand, Ann Kelleher, and Angelo Rodriguez, eds. *WCCFL 23: Proceedings of the 23rd West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Press, 374–98.

Hulden, Mans. (2017) "A phoneme clustering algorithm based on the Obligatory Contour Principle". Roger Levy and Lucia Specia, eds. *Proceedings of the 21st conference on computational natural language learning (CoNLL 2017)*. Vancouver: Association for Computational Linguistics, 290–300.

Janson, Tore. (1986) "Cross-linguistic trends in the frequency of CV sequences". *Phonology yearbook* 3: 179–95.

Kawahara, Shigeto, Hajime Ono, and Kiyoshi Sudo. (2006) "Consonant co-occurrence restrictions in Yamato Japanese". *Japanese/Korean linguistics* 14: 27–38.

Khan, Sameer ud Dowla. (2007) "Similarity avoidance in East Bengali fixed-segment reduplication". Erin Bainbridge and Brian Agbayani, eds. *Proceedings of the 34th Western Conference on Linguistics*. Fresno: Department of Linguistics at California State University, 257–71.

Kilpatrick, Alexander, Alexandra Ćwiek, Eleanor Lewis, and Shigeto Kawahara. (2023) "A cross-linguistic, sound-symbolic relationship between labial consonant, voiced plosive, and Pokémon friendship". *Frontiers in psychology* 14: 1–11.

Kinney, Ashley L. (2005) *Serial organization of speech sounds in Creole languages*. PhD dissertation, University of Texas.

Kokkelmans, Joachim. (2021) *The phonetics and phonology of sibilants*. PhD dissertation, University of Verona.

Kolgjini, Julie M. (2004) *Palatalization in Albanian: An acoustic investigation of stops and affricates*. PhD dissertation, University of Texas at Arlington.

Kořínek, Josef M. (1934) *Studie z oblasti onomatopoje* [Studies from the area of onomatopoeia]. Prague: Nákladem Filosofické fakulty University Karlovy.

Kumagai, Gakuji. (2020) "The pluripotentiality of bilabial consonants: The images of softness and cuteness in Japanese and English". *Open linguistics* 6: 693–707.

Ladefoged, Peter, and Ian Maddieson. (1996) *The sounds of the world's languages*. Oxford: Blackwell.

List, Johann-Mattis, Robert Forkel, Simon J. Greenhill, Christoph Rzymski, Johannes Englisch, and Russell D. Gray. (2022) "Lexibank, a public repository of standardized wordlists with computed phonological and lexical features". *Scientific data* 9(316): 1–16.

MacEastern, Margaret. (1999) *Laryngeal co-occurrence restrictions*. New York: Garland.

MacKay, Donald. (1970) "Phoneme repetition in the structure of languages". *Language and speech* 13(3): 199–213.

MacNeilage, Peter F., and Barbara L. Davis. (1999) "Consonant (vowel) consonant sequences in early words". *Proceedings of the 14th International Congress of Phonetic Sciences*. Vol. 3. San Francisco: University of California, Berkeley, 2489–92.

———. (2000) "Evolution of speech: The relation between ontogeny and phylogeny". Chris Knight, Michael Studdert-Kennedy, and James R. Hurford, eds. *The evolutionary emergence of language*. Cambridge, UK: Cambridge University Press, 146–60.

MacNeilage, Peter F., Barbara L. Davis, Ashley Kinney, and Christine Matyear. (2000) "The motor core of speech: A comparison of serial organization patterns in infants and languages". *Child development* 71: 153–63.

Maddieson, Ian. (1984) *Patterns of sounds*. Cambridge, UK: Cambridge University Press.

Mathesius, Vilém. (1931) "O výrazové platnosti některých českých skupin hláskových" [On the expressive validity of some Czech sound groups]. *Naše řeč* 15: 38–40.

Mayer, Thomas, Christian Rohrdantz, Frans Plank, Peter Bak, Miriam Butt, and Daniel A. Keim. (2010) "Consonant co-occurrence in stems across languages: Automatic analysis and visualization of a phonotactic constraint". Fei Xia and Lori Levin, eds. *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*. Uppsala: Association for Computational Linguistics, 70–78.

McCarthy, John. (1986) "OCP effect: Gemination and antigemination". *Linguistic inquiry* 17(2): 207–63.

———. (1994) "The phonetics and phonology of Semitic pharyngeals". Patricia A. Keating, ed. *Phonological structure and phonetic form*. Cambridge, UK: Cambridge University Press, 191–223. [Papers in Laboratory Phonology III.]

Nikitina, Tatiana. (2022) "At the intersection of synchrony and diachrony: A phonotactic analysis of the lexicon of Wan". *Languages in Africa* 3(2): 196–211.

Ohala, John J. (1990) "There is no interface between phonology and phonetics". *Journal of phonetics* 18: 153–71.

Ohala, Manjari. (1999) "Hindi". International Phonetic Association, ed. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge, UK: Cambridge University Press, 100–103.

Padgett, Jaye. (1992) "OCP subsidiary features". *NELS* 22(1): 335–46.

———. (1995) *Stricture in Feature Geometry*. Stanford: CSLI Publications.

Pierrehumbert, Janet. (1993) "Dissimilarity in the Arabic verbal roots". *NELS* 23: 367–81.

———. (1994) "Syllable structure and word structure: A study of triconsonantal clusters in English". Patricia A. Keating, ed. *Phonological structure and phonetic form*. Cambridge, UK: Cambridge University Press, 168–88. [Papers in Laboratory Phonology III.]

Plénat, Marc. (1996) "De l'interaction des contraints: Une étude de case". Jacques Durand and Bernard Laks, eds. *Current trends in phonology: Models and methods*. Salford: ESRI, 585–617.

Pozdniakov, Konstantin. (2007) "O *'ploxix' i 'xorošix' slovax* v russkom jazyke" [On 'bad' and 'good' words in the Russian language]. N. Vakhtin and G. Levinton, eds. *Hommage Albert Baiburin*. St. Petersburg: Université Européenne, 78–119. Available at: http://pozdnia-kov.free.fr/publications/2007_Baiburin.pdf. Last accessed 22 February 2024.

———. (2010) "Ograničenija distantnoj sočetaemosti soglasnyx v russkom jazyke" [Limitations of the distant combinability of the consonants in CVC sequences in Russian]. V. Z. Dem'jankov and V. J. Porhomovskij, eds. *V prostranstve jazyka i kul'tury: zvuk, znak, smysl. Sbornik statej v čest' 70-letija V. A. Vinogradova*. Moscow: Jazyki slavjanskix kul'tur, 113–26. Available at: http://pozdniakov.free.fr/publications/2010_Vinogradov.pdf. Last accessed 22 February 2024.

———. (2011) "Fonotaktika russkogo kornja i ee potencial'nye resursy" [Phonotactics of the Russian root and its potential resources]. *Hommage à Vadim Kasevich*. St. Petersburg University, 119–74. Available at: http://pozdniakov.free.fr/publications/2011_Phonotactics_Russian.pdf. Last accessed 22 February 2024.

Pozdniakov, Konstantin, and Guillaume Segerer. (2007) "Similar place avoidance: A statistical universal". *Linguistic typology* 11: 307–48.

Pystynen, Juho. (2014) "Similar place avoidance in language history". *Freelance reconstruction*. Blog post. Available at: https://protouralic.wordpress.com/2014/10/11/similar-place-avoidance-in-language-history/. Last accessed 5 February 2024.

Rácz, Péter, Jennifer Hay, Jeremy Needle, Jeanette King, and Janet Pierrehumbert. (2016) "Gradient Māori phonotactics". *Te Reo* 59: 3–21.

Reiner, Erica. (1966) *A linguistic analysis of Akkadian*. The Hague: Mouton and Co.

Rocasens, Daniel. (1990) "The articulatory characteristics of palatal consonants". *Journal of phonetics* 18: 267–80.

Rose, Sharon. (2000) "Rethinking geminates, long-distance geminates, and the OCP". *Linguistic inquiry* 31(1): 85–122.

Rose, Sharon, and Lisa King. (2007) "Speech error elicitation and co-occurrence restrictions in two Ethiopian Semitic languages". *Languages and speech* 50(4): 451–504.

Rosenthal, James A. (1996) "Qualitative descriptors of strength of association and effect size". *Journal of social service research* 21(4): 37–59.

Rousset, Isabelle. (2004) *Structures syllabiques et lexicales des langues du monde: Données, typologies, tendances universelles et contraintes substantielles*. PhD dissertation, Université Stendhal.

Sandell, Ryan. (2015) "Obligatory Contour Principle effects in Indo-European phonology: Statistical evidence and the morphology-phonology interface". Stephanie W. Jamison, H. Craig Melchert, Brent Harmon Vine, and Angelo Mercado, eds. *Proceedings of the 26th Annual UCLA Indo-European Conference*. Bremen: Hempen, 141–60.

Siptár, Péter, and Miklós Törkenczy. (2000) *The phonology of Hungarian*. Oxford: Oxford University Press.

Stanton, Juliet, and John F. Stanton. (2022) "In defense of O/E". Unpublished manuscript. Available at: https://ling.auf.net/lingbuzz/006391. Last accessed 5 February 2024.

Tang, Sze-Wing. (2000) "Identity avoidance and constraint interaction: The case of Cantonese". *Linguistics* 38(1): 33–61.

Teil-Dautrey, Gisèle. (2008) "Et si le proto-bantu était aussi une langue … avec ses contraintes et ses déséquilibres". *Diachronica* 25(1): 54–110.

Valleé, Nathalie, Solange Rossato, and Isabelle Rousset. (2009) "Favoured syllabic patterns in the world's languages and sensorimotor constraints". François Pellegrino, Egidio Marsico, Ioana Chitoran, and Christophe Coupé, eds. *Approaches to phonological complexity*. Berlin/New York: Mouton de Gruyter, 111–39.

van Goch, Wil. (2010) *The Obligatory Contour Principle: Consonant co-occurrence restrictions in Dutch*. PhD dissertation, Utrecht University.

Walter, Mary A. (2007) *Repetition avoidance in human language*. PhD dissertation, Massachusetts Institute of Technology.

———. (2010) "Harmony versus the OCP: Vowel and consonant cooccurrence in the lexicon". *Laboratory phonology* 1: 395–413.

Weiss, Michael. (2016) "The Proto-Indo-European laryngeals and the name of Cilicia in the Iron Age". Andrew Miles Byrd, Jessica DeLisi, and Mark Wenthe, eds. *Tavet Tat Satyam: Studies in honor of Jared S. Klein on the occasion of his seventieth birthday*. Ann Arbor, MI: Beech Stave, 331–40.

Wilson, Colin, and Marieke Obdeyn. (2009) "Simplifying subsidiary theory: Statistical evidence from Arabic, Muna, Shona, and Wargamay". Unpublished manuscript. Available at: https://linguistics.ucla.edu/people/hayes/219/papers/WilsonObdeyn2009.pdf. Last accessed 5 February 2024.

Wright, Richard. (2004) "A review of perceptual cues and cue robustness". Bruce Hayes, Robert Kirchner, and Donca Steriade, eds. *Phonetically based phonology*. Cambridge, UK: Cambridge University Press, 34–56.

Yip, Moira. (1989) "Feature geometry and cooccurrence restrictions". *Phonology* 6: 349–74.

Zukoff, Sam. (2017) *Indo-European Reduplication: Synchrony, Diachrony, Theory*. PhD dissertation, Massachusetts Institute of Technology.

Zygis, Marzena. (2003) "Phonetic and phonological aspects of Slavic sibilant fricatives". *ZAS papers in linguistics* 3: 175–213.

Aleš Bičan
Department of Linguistics
and Baltic Studies
Faculty of Arts, Masaryk University
Brno, Czech Republic
and
Czech Language Institute
Czech Academy of Sciences
Brno, Czech Republic
ORCID: 0000-0003-2481-6023
bican@phil.muni.cz