

# Universals, Their Violation and the Notion of Phonologically Peculiar Languages

**Vladimir Pericliev**

*Bulgarian Academy of Sciences*

## Abstract

A language can be said to be “peculiar” if it violates a universal pattern that admits only very few exceptions. In the paper, we propose a typology of phonological peculiarities concerning the content of segment inventories, and deal in detail with one of these types. The type involves the illegitimate absence of a segment that an implicational universal predicts should actually be present in the segment inventory of a language. A computer program retrieved 33 phonological peculiarities of this type in the *UCLA Phonological Segment Inventory Database* (UPSID), comprising 451 languages. It turned out that 391 of these languages had no peculiarity of the type studied, 43 had one, and 17 more than one peculiarity. Some observations are made regarding the last category of 17 “strongly peculiar” languages. In particular, it is shown that despite their strong phonological idiosyncrasy, the peculiar languages have only a limited variability in that they cannot violate more than six universals or have more than three distinct segments lacking.

Keywords: phonological universals, types of phonological peculiarities, gaps in segment inventories

## 1. Introduction

A language can be said to be “peculiar” (or “deviant”) if there is a universally valid norm for the class of all languages and this language belongs to a very small subclass of languages that violate the norm. The violated norm (or universal) can be any sort of linguistic object (entity, relation, rule, etc.), and depending on the domain to which the norm belongs, linguistic peculiarities can be phonological, morphological, syntactic or semantic.

A language like !Xu, for instance, is phonologically peculiar in that it has 141 segments, unlike other languages, that normally have inventories of  $27 \pm 7$  segments (Maddieson 1984: 7). In contrast, Hawaiian and Rotokas are phonologically peculiar in having unusually small segment inventories (of 13 and 11 segments, respectively). Hungarian is morphologically peculiar in that it has exceptionally many cases (more than 20, cf. e.g., Plank 1986). Some Mayan languages have the peculiar Verb-Object-Subject word order, while some Carib languages have the peculiar Object-Verb-Subject order (Derbyshire & Pullum 1981, Keenan 1978), the usual orders being stated in Greenberg’s first universal “In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object” (Greenberg 1966: 77). Eskimo (Pullum 1991) is semantically peculiar in that it has about a dozen different lexemes for snow, with no general term for the concept, while no other language is known to make so many sense differentiations. Bulgarian is semantically peculiar in that it makes unusually many, viz. four, relative-age distinctions for the kin term “husband’s sister”, even though the language has a general term for the term (Gerov 1897).

Similar examples can be easily multiplied, for linguistics has always been interested in finding universally valid patterns and their eventual exceptions or violations. Some recent manifestations of this research interest are e.g. the extension of the popular electronic *Universals Archive* at <http://ling.uni-konstanz.de/pages/proj/sprachbau.htm> (Plank & Filimonova 2000) with a collection of rare linguistic phenomena or the inclusion of entries on spectacular linguistic phenomena in the *World Atlas of Linguistic Structures* (Dryer et al. forthcoming).

In this paper, we shall be concerned with phonological peculiarities, and more specifically with such related to the content of segment inventories. In section 2, we outline four types of phonological peculiarities related to the content of segment inventories and discuss some problems of their study. In section 3, one of these types of peculiarities, viz. the type resulting from the absence of a segment in a language that should be present according to an implicational universal, is described in detail. Thirty-three such peculiarities are computationally retrieved from the *UCLA Phonological Segment Inventory Database* (UPSID). Section 4 makes some observations based on these results. In particular, it is shown that, out of the 451 languages in the UPSID database, forty-three (“mildly peculiar” languages) exhibit only one peculiarity (or violate just one universal), and seventeen (“strongly peculiar” languages) exhibit more than one peculiarity (violate more than one universal). It is suggested that the degree of peculiarity of a language be measured by the number of universals it violates, and it is observed that no language can violate more than six universals (of the type studied). Also, it is observed that no language can illegitimately lack more than three distinct segments. Finally, section 5 summarises the contributions of the paper.

## 2. Four Types of Phonological Peculiarities and How to Study them

A phonological peculiarity is any rare phonological phenomenon that is a violation of some well-established phonological norm (or universal). In this paper, we shall deal with phonological peculiarities related to the content of “segment inventories” (i.e., the collection of sounds used by a language). In particular, we shall describe in detail phonological peculiarities of what will be called peculiarities of Type III below.

The peculiarities concerning the content of segment inventories can be viewed from two distinct perspectives.

The first perspective focuses on whether this peculiarity involves an illegitimate *presence* or, alternatively, an illegitimate *absence* of a segment in a language (in the first case, the norm/universal prescribes that this segment should be absent from the language, while in the second, that the segment should be present in the language).

The second perspective focuses on whether the illegitimate presence or absence of a segment in a language is *conditional* (i.e., triggered by the presence/absence of another segment in the inventory of the same language) or *unconditional* (i.e., *not* triggered by the presence/absence of another segment in the inventory of the same language). Put differently, conditional phonological peculiarities are those that violate a “conditional universal” (=implicational universal), while unconditional phonological peculiarities are those that violate a “non-conditional universal” (=unrestricted universal).

Table 1 summarizes the resulting four logically possible types of phonological peculiarities (the variables A and B in the table stand for any individual segment or a sequence of segments).

Based on Table 1, we elaborate on the four types of phonological peculiarities in more detail.

As for the first type of peculiarity, *Type I: Illegitimate conditional presence of segment*, a language possesses this type of peculiarity if there is a firmly established norm (universal) prescribing that no language should have two specific segments A and B in its segment inventory, but this language in fact has both segment A and segment B.

As an example of phonological peculiarity Type I, consider:

Table 1. Types of phonological peculiarities pertaining to the content of segment inventories

| Type No. | Type Name                                      | Peculiarity Specifics              | Universal Violated |
|----------|--|------------------------------------|--------------------|
| I        | Illegitimate conditional presence of segment   | Language has both segments A and B | A implies not B    |
| II       | Illegitimate unconditional presence of segment | Language has segment A             | not A              |
| III      | Illegitimate conditional absence of segment    | Language has segment A but not B   | A implies B        |
| IV       | Illegitimate unconditional absence of segment  | Language does not have segment A   | A                  |

“A language has both a voiced bilabial fricative [β] and a voiced labiodental fricative [v].”

The peculiarity violates the implicational universal  $\square[\beta] \square \rightarrow \text{not } [v]$ <sup>1</sup> (cf. Maddieson 1984: 14), or what is logically equivalent, the proposition that a language cannot contain both [β] and [v]. By way of an explanation for a universal of this type, it has been suggested that languages generally do not tolerate “closely related” sounds, and in this particular case, the voiced bilabial fricative [β] and the voiced labiodental fricative [v] are such sounds. Given some norm (or universal), it is always interesting to see what the exceptions to this norm are. In this case, the languages in UPSID-451 exhibiting the exceptional phenomenon are Ewe (Niger-Kordofanian language family/Kwa branch), Kohumono (Niger-Kordofanian language family/Cross-River branch), Ogbia (Niger-Kordofanian language family/Cross-River branch), and Dahalo (Afro-Asiatic language family/Cushitic branch). Remarkably, all peculiar languages belong to different branches of the same language family, Niger-Kordofanian, suggesting that this peculiarity might be viewed as a “family peculiarity”.

The second peculiarity, *Type II: Illegitimate unconditional presence of segment*, involves languages that illegitimately possess a segment in violation of a firmly established unrestricted universal to the effect that no language should possess this segment.

As an example of phonological peculiarity Type II, consider: “A language has a voiceless vowel.”

This peculiarity violates the unrestricted (=unconditional) universal that languages do not normally have voiceless vowels. This is a common sense universal since it is all too well known that vowels are as a rule pronounced with voice. On inspection of the UPSID-451 database, it is seen that the languages exhibiting the unusual phenomenon are Dafla, a Sino-Tibetan language of the Himalaic branch, Ik, a Nilo-Saharan language of the E. Sudanic branch, and Sandawe, a Khoisan language of the Sandawe branch of the family.

By way of another example of the same type of peculiarity, consider: “A language has a fricative vowel.”

This peculiarity violates the unrestricted universal that languages do not have fricative vowels. We found that the languages in UPSID-451 exhibiting the rare phenomenon are Bai and Naxi, both belonging to the Sino-Tibetan language family, but the first language of the Sinitic branch, while the second from the Lolo-Burmese branch. Yet another language in the database studied is exceptional in this sense, viz. Ewondo, a Niger-Kordofanian language from the Bantoid branch.

The third type of phonological peculiarity, *Type III: Illegitimate conditional absence of segment*, involves the violation of a sound implicational universal, Segment A  $\square \rightarrow \square$  Segment B, prescribing that if a language possesses Segment A, then it should also possess Segment B.

As an example of this type of peculiarity, consider: “A language has a voiced palato-alveolar sibilant fricative [ʒ] but lacks the voiceless palato-alveolar sibilant fricative [ç].”

<sup>1</sup> The arrow “ $\rightarrow$ ” between two segments A  $\square \rightarrow \square$  B stands for logical implication, holding between these segments. Segment A is usually referred to as the “antecedent” of the implication, and segment B as the “consequent” of the implication. The implication A  $\square \rightarrow \square$  B is true in a language in all cases unless the language has A (the antecedent), but lacks B (the consequent).

This peculiarity violates the implicational universal that if a language has [ʒ] then it should also have [ʃ]. The peculiarity will be found in languages with [ʒ], but lacking [ʃ]. The languages in UPSID-451 exhibiting the unusual phenomenon are Mbum (Niger-Kordofanian language family, Adamawa branch), Atayal (Austro-Tai language family, Atayalic branch), and Apinaye (South American/Macro-Ge).

The fourth type of phonological peculiarity, *Type IV: Illegitimate unconditional absence of segment*, involves the violation a well-established unrestricted (unconditional) universal prescribing that all languages should possess some segment.

As an example of this type of peculiarity, consider: “A language lacks a bilabial plosive.”

This peculiarity violates the unrestricted universal that all languages have a bilabial plosive. The languages in UPSID-451 exhibiting the spectacular phenomenon are Aleut (Eskimo-Aleut/Aleut), Eyak (Na-Dene/Eyak), Hupa (Na-Dene/Athabaskan), Cherokee (Northern American/Keresiouan), Wichita (Northern American/Keresiouan). It is a striking circumstance that all these are North American languages. On what is perhaps the prevalent view today, assuming that the language families Eskimo-Aleut, Na-Dene and Keresiouan are genetically unrelated, this fact would most probably suggest that it might well be an “areal peculiarity”, possessed by languages in one geographical area, North America. The proponents of “macro-families” in contrast, assuming that these (together with some other Eurasian languages) are themselves genetically related in their distant past, would probably view this peculiarity as a “family peculiarity” and thus as further evidence for their genetic relationship. In any case, this peculiarity will be of substantial theoretical interest.

By way of another example of the same type of peculiarity, consider: “A language lacks a front vowel.”

This peculiarity violates the unrestricted universal that all languages have a front vowel. The languages in UPSID-451 exhibiting this exceptional phenomenon are Qawaskar (or Alakaluf) (South American family, Andean branch) and Yessan-Mayo (Papuan family, Sepik-Ramu branch).

How can we go about describing the phonological peculiarities of these four types?

The notion of “(phonological) peculiarity” is closely related to language universals, and in particular to the notion “near universals”, conceived as firmly established patterns allowing of only a very small number of exceptions. Language universals are usually divided into “absolute universals”, valid unexceptionally for all languages, and “statistical (or near) universals”, that can have only a limited number of exceptions (for a general discussion of universals, cf. Comrie 1981, Croft 1990; for a recent interesting methodological discussion, cf. Odden 2003). If absolute universals split the universe of all languages into “possible” and “impossible” types, near universals can be said to split languages into “majority” and “minority” types. If absolute universals completely denounce language variability of certain types (viz. the types that violate the universals), near universals actually *assert* language variability, by admitting all types (including the minority types that violate the universals). Thus, the task of discovering phonological peculiarities reduces to the task of discovering the violations of near universals, and the task of discovering phonologically peculiar languages reduces to the task of discovering the languages that are the exceptions to these near universals.

As with any grammatical description, we can impose some requirements as to what is to be considered an “adequate description” of phonological peculiarities. Such a description should include all the phonological peculiarities of the type studied (which is what might be called the *completeness condition*) and only these peculiarities (which is what might be called the *soundness condition*).

The *soundness condition* in essence boils down to the question of whether or not the near universals the found peculiarities are violations of are relatively uncontroversial universals. Indeed, to be in a position to label a phenomenon “peculiar” (“deviant”, “rare”, etc.) one needs to have at one’s disposal some firmly established norm, or standard, to serve as a yardstick. A reliable universal is a statement that is valid for all (or most) languages such that it is due to some universal processes and not due to

such reasons as: (i) borrowing of some languages from others, (ii) genetic inheritance of some languages from others, or (iii) chance (cf. Comrie 1981: 194).

The *completeness condition* amounts to listing exhaustively, without any omissions, of all phonological peculiarities of the type studied.

An adequate description of phonological peculiarities thus needs to satisfy both the soundness and the completeness conditions.

The example “A language has a voiced palato-alveolar sibilant fricative [ʒ] but lacks the voiceless palato-alveolar sibilant fricative [ʃ].” above (belonging to Type III phonological peculiarities, which will be the subject of our study) for instance fulfils the soundness condition, though this may not be immediately obvious from its presentation above. In actual fact, the phenomenon described in this example is truly peculiar in that, first, it holds only in 3 languages out of 61 relevant ones (i.e., in 5% of the languages in USPID-451, and by extrapolation, in 5% of all known languages), and second, the universal violated is quite a reliable one (cf. the sixteenth phonological peculiarity in Section 3, which corresponds to our example). Indeed, the universal is not due to borrowing or genetic inheritance being supported in 58 (=61 – 3) languages coming from 3 large geographical areas, viz. EURASIA, AFRICA, and the AMERICAS. Besides, as we have computed, the universal pattern is statistically highly significant ( $p < 0.001$ ) and therefore most unlikely to have occurred by chance.

In contrast, the suggestion made in a familiar paper by Gamkrelidze (1978) that languages with the voiced velar fricative phoneme [ɣ] but lacking the voiceless velar fricative [x] are “rare exceptions” is not a really sound phonological peculiarity. Thus, our inspection of USPID-451 shows that although the implicational pattern violated is a reliable one (it is non-chance and is supported in 14 language families from 3 geographical areas), the phenomenon is by far not exceptional being valid in 27 out of 55 languages (or in 49% of all languages).

Similarly, if one were to posit as peculiar languages that have in their segment inventories the high back unrounded vowel [ɯ], but lack the voiced bilabial nasal [m], we should be very suspicious regarding such a proposal. Thus, irrespectively of the fact that such languages are exceptional (2 languages in UPSID-451 out of 41, or in 5%), and the violated implication is typologically well supported (in 10 language families belonging to all 4 large geographical areas), the pattern is most likely the result of pure chance (we estimated that  $p > 0.9$ ). This will come as no surprise as one recognizes the fact that the absence of the consonant [m] in a language is in itself rare (in only 26 out of 451 languages) and therefore the languages having the additional property of also possessing the vowel [ɯ] will, of necessity, be still rarer; this implicates that no correlation exists between [ɯ] and [m].

As regards the completeness condition, which an adequate description of a type of phonological peculiarity should fulfil, we may say that previous research on language peculiarities neither presented exhaustive descriptions nor indeed considered the imposing of such an adequacy condition. Our description in the next section will conform to the condition.

### **3. Phonological Peculiarities Involving an Illegitimate Conditional Absence of a Segment**

Below we describe the phonological peculiarities of Type III, viz. such involving an illegitimate conditional absence of a segment.

We base our investigation on the most detailed collection of segment inventories of the world languages, the *UCLA Phonological Segment Inventory Database* (UPSID-451), compiled by Maddieson and colleagues at UCLA (Maddieson & Precoda 1991, Maddieson 1991, cf. also Maddieson 1984), which contains the segment inventories of 451 languages.

UPSID-451 is compiled on a genetic principle, classifying the languages into 18 major genetic groupings (=language families). These language families fall into 4 large geographical areas: EURASIA, AFRICA, the AMERICAS, and AUSTRALASIA, as shown below.

- |   |  |
|---|--|
| <p>I. EURASIA</p> <ol style="list-style-type: none"> <li>1. Indo-European</li> <li>2. Ural-Altaic</li> <li>3. Austro-Asiatic</li> <li>4. Austro-Tai</li> <li>5. Sino-Tibetan</li> <li>6. Caucasian</li> <li>7. Other Eurasian minor families</li> <li>8. Dravidian</li> </ol> | <p>III. AMERICAS</p> <ol style="list-style-type: none"> <li>13. Na-Dene</li> <li>14. North Amerind</li> <li>15. South Amerind</li> <li>16. Eskimo-Aleut</li> </ol> |
| <p>II. AFRICA</p> <ol style="list-style-type: none"> <li>9. Niger-Kordofanian</li> <li>10. Nilo-Saharan</li> <li>11. Afro-Asiatic</li> <li>12. Khoisan</li> </ol>   | <p>IV. AUSTRALASIA</p> <ol style="list-style-type: none"> <li>17. Australian</li> <li>18. Papuan</li> </ol>  |

UPSID-451 can be assessed as a very representative sample. The primary goal of the database is to provide a sample from which statistically valid statements concerning frequency and co-occurrence can be drawn. The database is thus a reliable empirical source for a computational investigation of the sort undertaken in this article.

We have developed sophisticated computational machinery for the discovery and verbalization in English of language universals/peculiarities. The program was previously applied to the discovery and verbalization of word order universals (Pericliev 1999, 2000, 2002a). The computational details (described in Pericliev 2002b and 2003) are outside the interest of a typologist and will therefore be ignored here. We will only outline some of the program's features that are directly relevant for this study. The computational system can find, by conducting an exhaustive search of the examined database, all the phonological peculiarities (and near universals), thus fulfilling the completeness condition. To ensure the soundness of discovered peculiarities, the system can find phonological peculiarities (and near universals) with a user-specified percentage of validity such that are supported by a user-specified number of languages belonging to a specified number of language families and geographical areas. Additionally, the system can estimate the statistical significance of the peculiarities (or near universals) found using the standard chi-square test (e.g., McClave & Dietrich 1988), thus minimizing the possibility for chance peculiarities to smuggle in the description.

Below we list the phonological peculiarities of the investigated Type III (illegitimate conditional absence of segment) that the system found exploring UPSID-451. We ran the program requiring that only peculiarities be found that are valid in at most 5% of the languages (although 5% is, of course, a somewhat randomly chosen cut-off point). Also, these peculiarities are such that violate statistically highly significant universals ( $p < 0.001$ ) (in order to minimize the smuggling of purely chance patterns) and are further supported in at least 2 distinct language families from 2 distinct large geographical areas (in order to minimize the smuggling of genetic/areal patterns).

The format of the description is as follows. Each phonological peculiarity listed states a segment that is present and a segment that is "illegitimately" absent. The peculiar languages in UPSID-451 (relative to this phenomenon) are then given with their genetic affiliation alongside with some further comment.

The first phonological peculiarity we address is observed when a language has a voiced dental plosive [d] but lacks a voiced bilabial plosive [b].

Languages tend to exhibit correlation between segments of the same manner such that differ only as regards their place of articulation. One (of many) examples is the universal [d] □ → □ [b], in which the two segments have an identical feature structure, except that the place of the first is dental, and the place of the second bilabial. The universal violated by this phonological peculiarity is massively supported in 78 languages from 10 out of all 18 language families recognized in UPSID, the languages belonging to all 4 geographical areas, EURASIA, AUSTRALASIA, AFRICA, and the AMERICAS. The languages exhibiting the phonological peculiarity in question are Irish, an Indo-European language, and Mixe, a North Amerind language. Irish doesn't have the voiced bilabial plosive [b], but as some kind of "compensation" has the palatalized voiced bilabial plosive [bʲ] and the labialized-velarized voiced bilabial plosive [bʷ]. Mixe, which is very different in segment inventory from Irish, does not have a voiced bilabial plosive at all, but only the voiceless bilabial plosive [p].

The second phonological peculiarity we mention refers to a language which has a voiced alveolar plosive [d] but lacks a voiced bilabial plosive [b].

Like the previous case, this phonological peculiarity violates a universal, viz. [d] □ → □ [b], holding between segments differing only in their place of articulation (alveolar vs. bilabial). Likewise, the universal is strongly supported by 116 languages from 15 language families situated in all 4 geographical areas. The languages exhibiting the peculiarity are Irish (Indo-European), Gadsup (Papuan), Wapishana (South Amerind), and Cherokee (North Amerind). We saw above that Irish seems to compensate its lack of [b] by having some of its variants. Gadsup doesn't have a voiced bilabial plosive at all (but only the voiceless bilabial plosive [p]). Wapishana has only its laryngealized version, [ɓ], and Cherokee has no bilabial at all (a fact we mentioned in Section 2 as an example of phonological peculiarity Type IV).

A third phonological peculiarity is observed when a language has a voiced dental/alveolar plosive [ " d " ] but lacks a voiced bilabial plosive [b].

This phonological peculiarity, again, violates a universal, viz. [ " d " ] implies [b], holding between segments of the same manner, differing only in their place of articulation (dental/alveolar vs. bilabial). It should be mentioned at this point that the feature "dental/alveolar", as applied to consonants, is used as a cover term by the compilers of UPSID-451, to denote unspecified segments for which it is not clear exactly whether they are true dentals or true alveolars (the reason for this uncertainty occurring in the database being in the fact that many of the linguistic sources used in the compilation do not specify this information). Put in the context of the previous two peculiarities, the third phonological peculiarity therefore, though not strictly speaking carrying novel information, can be viewed as providing an undifferentiated support to these peculiarities and we prefer to list it separately. The violated universal, [ " d " ] □ → □ [b], is massively supported in 89 languages out of 91 relevant languages. The supporting languages come from 16 language families from all geographical areas. The deviant languages exhibiting the peculiar phenomenon are Sentani, a Papuan language and Eyak, a Na-Dene language. If Sentani has the voiceless bilabial plosive [p], Eyak is most exceptional in not having any bilabial plosive at all.

A fourth phonological peculiarity is to be found in a language which has a voiceless dental plosive [t̪] but lacks a voiceless velar plosive [k].

This one, and the following two peculiarities, are similar to the preceding three peculiarities with the exception that the illegitimately absent segment is not the bilabial plosive [b], but the velar plosive [k]. The violated universal in this particular case is [t̪] □ → □ [k]. It is quite well supported in 104 languages. These come from 16 language families situated in all 4 geographical areas. It is interesting to note that both languages that exhibit the peculiar phenomenon, viz. Bella Coola and Tzeltal, are North Amerind languages. It thus looks like a "family peculiarity". Both languages have some variant of the

voiceless velar plosive [k]. Bella Coola has the labialized velar plosive [k<sup>w</sup>], and Tzeltal the aspirated velar plosive [k<sup>h</sup>].

A fifth phonological peculiarity is observed when a language has a voiceless alveolar plosive [t] but lacks a voiceless velar plosive [k].

The violated universal, [t] → [k], is massively supported in 174 languages, coming from 15 languages from all 4 geographical areas. Three of the deviant languages, viz. Kewa, Vanimo, and Usan, are Papuan; two exceptional languages are Niger-Kordofanian, viz. Beembe and Kohumono; one language, Hupa, belongs to the Na-Dene language family, and Kwaio belongs to the Austro-Tai family. Kewa and Usan lack [k], but both have its voiced counterpart [g]. Vanimo, the other deviant Papuan language, doesn't have a velar plosive at all. Beembe has its aspirated counterpart [k<sup>h</sup>], and Kohumono, the other Niger-Kordofanian language, both [k<sup>h</sup>] and its labialized version [k<sup>wh</sup>]. Hupa does not have a velar plosive at all, and neither has Kwaio.

A sixth phonological peculiarity amounts to the following: a language has a voiceless dental/alveolar plosive [t̪] but lacks a voiceless velar plosive [k].

This peculiarity reports a similar state of affairs as the previous two (see also the considerations concerning the feature “dental/alveolar” in the third phonological peculiarity). The violated universal, [t̪] → [k], is corroborated in 150 out of 152 relevant languages (or in 99% of the cases). This supportive evidence comes from languages from 16 language families from all geographical areas. The languages exhibiting the deviant phenomenon are Klao (Niger-Kordofanian) and Zuni (North Amerind). Klao does not have a velar plosive, and Zuni has the aspirated [k<sup>h</sup>] and the aspirated labialized [k<sup>wh</sup>].<sup>2</sup>

The seventh phonological peculiarity can be found in a language that has a voiceless aspirated dental/alveolar plosive [t̪<sup>h</sup>] but lacks a voiceless aspirated velar plosive [k<sup>h</sup>].

This peculiarity also reports a violation of a universal holding between segments of the same manner such that differ only insofar as their place of articulation is concerned (in this particular case, a dental/alveolar segment implies its velar counterpart). The violated universal, [t̪<sup>h</sup>] implies [k<sup>h</sup>], is supported in 47 languages, belonging to 11 language families from all 4 geographical areas. The exceptional languages exhibiting the phenomenon both come from the Americas and are Shiriana (South Amerind) and Picuris (North Amerind). Both languages lacking [k<sup>h</sup>] have its plain counterpart [k].

The eighth instance of a phonological peculiarity to be addressed is observed when a language has a prenasalized voiced velar plosive [ᵑg] but lacks a prenasalized voiced bilabial plosive [ᵑb].

This is an analogous peculiarity, infringing a norm holding between similar segments pronounced at different places. The norm ([ᵑg] → [ᵑb]) is supported in 44 out of 45 relevant languages from 11 language families from all 4 areas. The only deviant language exhibiting the phenomenon is Mazatec, a North Amerind language. Mazatec has several other prenasalized segments, but not such at the bilabial place of articulation. All other languages from the Oto-Manguean branch (in UPSID-451) to which Mazatec belongs actually lack the prenasalized voiced velar plosive [ᵑg], and are in this sense irrelevant, with the exception of one language, Amuzgo, which, unlike its relative Mazatec, is non-deviant and conforms to the universal, having both the prenasalized voiced velar plosive [ᵑg] and the prenasalized voiced bilabial plosive [ᵑb].

The ninth example of phonological peculiarity occurs in a language which has a voiced retroflex plosive [d.] but lacks a voiced bilabial plosive [b].

This is yet another peculiarity violating a universal between segments identical in feature structure except for the feature of place. The violated universal, [d.] → [b], is supported in 26 out of 27 relevant languages. The latter belong to 10 language families from all 4 areas. The only exceptional lan-

<sup>2</sup> When this article was ready to go to print, it was suggested to me that Klao and Zuni actually have the voiceless velar plosive [k]. I was not in a position to check this suggestion properly in other specialised literature, so stick to the data in UPSID-451, stating that these languages lack [k].



guage exhibiting the peculiar phenomenon is Diyari, an Australian language. Diyari not only lacks the voiced bilabial plosive [b], but is deviant in that it does not have a bilabial plosive at all. Diyari belongs to the Pama-Nyungan branch of Australian, with the following other languages listed in UPSID-451: Wik-Munkan, W. Desert, Arrernte, Gugu-Yalandyi, Kala Lagaw Ya, Bandjalang, Yolngu, and Yidiny. It is interesting to note that neither of these languages actually has the voiced retroflex plosive [d.], as Diyari has, and therefore neither is either an example or an exception to the universal [d.] □ → □ [b].

The tenth phonological peculiarity is observed when a language has a voiced retroflex plosive [d.] but lacks a voiced velar plosive [g].

This peculiarity is similar to the previous ones as regards the nature of the violated universal, and similar to the immediately preceding one in having the same language, Diyari (Australian), manifesting the peculiar phenomenon. Diyari doesn't have the voiced velar plosive [g], but has its voiceless counterpart [k]. The violated universal, [d.] □ → □ [g], is supported in 26 languages from 10 language families from all geographical areas. As was mentioned in the previous peculiarity, the other languages from the Pama-Nyungan branch of Australian, listed in UPSID-451, will be irrelevant to this universal/peculiarity since neither of them actually has the voiced retroflex plosive [d.]

The eleventh phonological peculiarity we deal with is when a language has a voiced palatal plosive [ɟ] but lacks a voiced velar plosive [g].

This is yet another peculiarity involving violation of a universal in which a segment implies its equivalent segment pronounced at a different place. Here, a voiced palatal plosive [ɟ] implies a voiced velar plosive [g]. This universal pattern is corroborated in 42 out of 43 relevant languages. The support comes from 12 language families from all 4 areas. The exceptional language exhibiting the phenomenon is Kwakw'ala, a North Amerind language. Kwakw'ala does not have the plain voiced velar plosive [g], but only its labialized counterpart [g<sup>w</sup>]. Kwakw'ala belongs to the Almosan branch of North Amerind. UPSID-451 has data on the following other languages belonging to this branch: Wiyot, Ojibwa, Quileute, Tseshaht (Nootka), Lushootseed, Upper Chehalis, Shuswap, and Bella Coola. Kwakw'ala looks different from all these relatives in that neither of these languages has the voiced palatal plosive [ɟ].

The twelfth phonological peculiarity can be found in a language that has a voiced velar plosive [g] but lacks a voiced bilabial plosive [b].

This peculiarity is of the same general kind, as those discussed above. The violated universal, [g] □ → □ [b], is massively supported in 244 languages. The evidence comes from 17 (out of a total of 18 language families distinguished in UPSID), belonging to all geographical areas. There are nine exceptional languages conforming to this peculiarity. The languages in which the peculiarity shows up are: Kewa and Rotokas (both Papuan), Eyak (Na-Dene), Irish (Indo-European), Brao (Austro-Asiatic), Wapishana (South Amerind), Cherokee, Mazahua, and Mixe (all three North Amerind). Instead of the expected voiced bilabial plosive [b], Kewa has its prenasalized counterpart [m̥b], Rotokas, Brao, Mazahua, and Mixe have its voiceless fellow [p], Eyak and Cherokee, as already mentioned, have no bilabial at all, and Irish has two variants: the palatalized voiced bilabial plosive [b<sup>ɰ</sup>] and the labialized-velarized voiced bilabial plosive [b<sup>w</sup>]. Wapishana has only its laryngealized version, [b̥].

The thirteenth phonological peculiarity involves a language which has a voiced labial-velar plosive [ɡ̠b] but lacks a voiced velar plosive [g].

This is yet another peculiarity infringing a norm holding between two segments sharing all their feature-value structures, but the feature-values for place, in this case “labial-velar” vs. “velar”. The violated universal, [ɡ̠b] □ → □ [g], is supported in 37 out of 39 relevant languages. The evidence comes from 5 language families from 3 areas, EURASIA, AUSTRALASIA, and AFRICA. This exceptional phenomenon shows up in Klao and Temne, both Niger-Kordofanian languages. This peculiarity thus

seems to be a family peculiarity. Indeed, neither of these languages has a velar plosive at all (but for Klao, cf. footnote 2).

The fourteenth phonological peculiarity concerns languages that have a prenasalized voiced alveolar plosive [ʎ] but lack a voiceless alveolar plosive [t].

This peculiarity is of a different kind than those discussed so far in that the violated universal, viz. [ʎ] □ → □ [t], does not hold between segments differing only in their place of articulation. Here, both segments are pronounced at the same place, but have different feature-values for voicing, and besides, the antecedent segment [ʎ] has the additional feature “prenasalized”. The universal is supported in 25 out of 26 languages, the languages being members of 8 families from all 4 geographical areas. The only deviant language is Yessan-Mayo, a Papuan language. In place of the expected voiceless alveolar plosive [t], Yessan-Mayo has its aspirated equivalent [tʰ]. Yessan-Mayo belongs to the Sepik-Ramu branch of Papuan. UPSID-451 contains data on the following other languages of this branch: Iwam, Kwoma, and Alamlak. Iwam and Alamlak do not have the prenasalized voiced alveolar plosive [ʎ], while Kwoma does have it, but unlike Yessan-Mayo, conforms to the norm and possesses also the voiceless alveolar plosive [t].

The fifteenth phonological peculiarity is observed when a language has a voiceless bilabial plosive [p] but lacks a voiceless velar plosive [k].

This peculiarity involves a violation of a universal between identical segments but for their place of articulation. The violated universal, [p] implies [k], is massively supported in 368 languages from all 18 language families recognized in UPSID. There are nine deviant languages. These languages are Usan and Vanimo (from the Papuan language family), Beembe and Klao (from the Niger-Kordofanian language family), and Bella Coola, Tzeltal, and Zuni (from the North Amerind language family). As regards the languages Klao and Zuni, see footnote 2. Usan and Vanimo lack not only the expected voiceless velar plosive [k], but also its voiced counterpart [g]. So do Beembe and Klao (Bembe has the aspirated variant [kʰ]). Bella Coola has the labialized counterpart [kʷ], Tzeltal [g] and [kʰ], and Zuni has the aspirated [kʰ] and the aspirated labialized [kʷʰ].

The sixteenth phonological peculiarity has to do with languages that have a voiced palato-alveolar sibilant fricative [ʒ] but lack a voiceless palato-alveolar sibilant fricative [ʃ].

Generally, the presence of a voiced fricative in the segment inventory of a language implies also the presence of its voiceless fricative, but this norm can have a fairly large number of exceptions for some specific segments (Maddieson 1984: 47). This peculiarity concerns the violation of a quite well supported universal. The violated universal, [ʒ] □ → □ [ʃ], holds in 58 out of 61 relevant languages; the languages are members 13 language families from EURASIA, AFRICA, and the AMERICAS. The languages exhibiting the peculiar phenomenon are Mbum (Niger-Kordofanian), Atayal (Austro-Tai), and Apinaye (South Amerind). The expected voiceless palato-alveolar sibilant fricative [ʃ] is absent in Mbum, but it has the voiceless alveolar sibilant fricative [s], while Atayal and Apinaye both have the voiceless dental/alveolar sibilant fricative [s].

The seventeenth phonological peculiarity concerns languages that have a voiced velar nasal [ŋ] but lack a voiced bilabial nasal [m].

In nasal stops, similarly to oral stops, there is a strong correlation between identical segments that differ only as regards their place of articulation. This peculiarity is an example of a violation of such a universal, viz. [ŋ] □ → □ [m], in which the segments differ only in that the first is velar, and the second bilabial. The infringed universal is massively supported in 235 out of 237 relevant languages (17 language families from all 4 large geographical areas). The languages exhibiting the rare phenomenon are Irish (Indo-European), and Mixtec (North Amerind). Irish, perfectly analogously to the way it handles its lack of the plain plosive [b] (cf. the first phonological peculiarity), “compensates” its lack of the voiced bilabial nasal [m] with the presence of its palatalized counterpart [mʲ] and its labialized and

velarized counterpart [m<sup>w</sup>]. Mixtec, in contrast, is additionally exceptional in that it does not have a bilabial nasal at all.

The eighteenth phonological peculiarity is found in such languages that have a voiced alveolar nasal [n] but lack a voiced bilabial nasal [m].

This peculiarity is of the same general kind as the preceding one. Again, the violated norm ([n] □ → □ [m]) is massively supported in 200 out of 202 relevant languages. They are distributed in 17 language families from all 4 areas. The exceptional phenomenon shows up in Konkani (Indo-European) and South-Nambiquara (South Amerind). Konkani has no other nasal than [n], and South-Nambiquara has only the laryngealized voiced dental/alveolar [ɲ], and not any bilabial nasal.

The nineteenth phonological peculiarity occurs in languages that have a nasalized lower-mid front unrounded vowel [ẽ̃] but lack a lower-mid front unrounded vowel [ɛ].

This peculiarity violates the very general tendency that the presence of segments with secondary articulations in the segment inventory of a language imply the presence of their simpler equivalents without these secondary articulations. This tendency is frequently referred to as “markedness”, the marked member being the more complex segment and the unmarked member the simpler segment (cf. Greenberg 1966). The violated universal, [ẽ̃] □ → □ [ɛ], is supported in 34 languages, members of 7 language families from all 4 geographical areas. There is only one language exhibiting this rare phenomenon, viz. Seneca, a North Amerind language. The segment [ẽ̃] is the only lower-mid segment in the language.

The twentieth phonological peculiarity we address amounts to the following circumstances: a language has a nasalized low central unrounded vowel [ã̃] but lacks a low central unrounded vowel [a].

This phonological peculiarity is similar to the previous one in that it violates a universal holding between two segments, one of which is the nasalized counterpart of the other segment. The universal, [ã̃] □ → □ [a], is supported in 82 out of 83 relevant languages. The corroborating languages are members of 10 language families from all 4 geographical areas. The sole language exhibiting the rare phenomenon is Kashmiri, which belongs to the Indo-European language family. It is interesting to note that Kashmiri has about half a dozen cases of other nasalized long/short vowels, in all of which the language has the pair nasalized vs. non-nasalized equivalent, but not in the particular case under study.

The twenty-first phonological peculiarity is observed when a language has a long higher-mid front unrounded vowel [e:] but lacks a long higher-mid back rounded vowel [o:].

This is yet another peculiarity holding between vowels. However, not between a nasalized and a non-nasalized segment, but between a front unrounded and a back rounded vowel. This is a common pattern. The particular violated universal, [e:] □ → □ [o:], an instance of this pattern, is supported in 20 out of 21 relevant languages, which belong to 9 language families situated in all 4 geographical areas. The only language exhibiting the spectacular phenomenon is Norwegian, a member of the Indo-European language family.

The twenty-second phonological peculiarity we list occurs when a language has a high back rounded vowel [u] but lacks a high front unrounded vowel [i].

This peculiarity, again, involves vowels. It violates the universal, [u] implies [i], holding between a back rounded and a front unrounded vowel. This norm is massively supported in 360 languages, which are members of all 18 language families in UPSID. This universal is an instance of a common trend in languages to have a front unrounded vowel equivalent to correspond to a back rounded vowel. However, there are exceptions to this general trend, and there are 9 exceptions to the particular universal at issue. The languages exhibiting the peculiar phenomenon are Even, Selkup and Yukaghir (all three being Ural-Altaic languages), Basque (isolate), Chukchi (Chukchi-Kamchatkan), Nivkh (Paleo-Siberian), Pohnpeian (Austro-Tai), Hixkaryana (South Amerind), and Angas (Afro-Asiatic).

The twenty-third phonological peculiarity we deal with occurs in a language that has a voiced dental/alveolar plosive [d] but lacks a voiced dental/alveolar nasal [n].

This peculiarity holds between segments of different manners, however sharing their place of articulation and the feature-value for voicing (viz. voiced). The violated universal, [ "d" ] □ → □ [ "n" ], is supported in 87 out of 91 relevant languages. The languages are members of 16 language families from all 4 geographical areas. The particular languages manifesting the rare phenomenon are Yakut (Ural-Altai family), Maba (Nilo-Saharan family), Klao (Niger-Kordofanian family), Lithuanian (Indo-European family).

The twenty-fourth phonological peculiarity we mention refers to a language that has a voiced alveolar implosive [ɗ] but lacks a voiced alveolar nasal [n].

This phonological peculiarity, again, holds between segments of different manners, but such sharing the feature-values for voicing and place of articulation. The infringed norm is [ɗ] □ → □ [n]. The norm is corroborated in 22 out of 23 languages, which are members of 7 language families from the areas EURASIA, AFRICA and the AMERICAS. The only language exhibiting the rare phenomenon is Vietnamese (Austro-Asiatic). Vietnamese lacks the expected voiced alveolar nasal [n], but in fact has the voiced dental nasal [ɳ].

The twenty-fifth phonological peculiarity can be found in such languages which have a labialized voiceless velar ejective-stop [kʷ] but lack a glottal plosive [ʔ].

This phonological peculiarity involves the violation of a universal holding between two segments of totally different feature structure (unlike all peculiarities/universals so far, in which there was a significant feature-value overlap). The violated universal, [kʷ] □ → □ [ʔ], is supported in 22 out of 23 relevant languages. The corroborating languages belong to 6 language families from EURASIA, AFRICA, and the AMERICAS. The only language exhibiting the rare phenomenon is Amharic, an Afro-Asiatic language.

The twenty-sixth phonological peculiarity to mention in this article is observed when a language has a voiced dental/alveolar sibilant fricative [ "z" ] but lacks a voiced dental/alveolar nasal [ "n" ].

This peculiarity involves the violation of a norm, holding between a fricative and a nasal sharing their feature-values for voicing and place of articulation. The violated universal, [ "z" ] □ → □ [ "n" ], is supported in 42 out of 43 languages. The languages are members of 13 language families from all 4 geographical areas. The only language in which the rare phenomenon shows up is Hindi-Urdu, an Indo-European language. Hindi-Urdu however has the voiced dental nasal [ɳ], and since [ "n" ] stands for unspecified voiced dental or alveolar nasal, it is not really a clear exception, and all that can be said with some certainty is that the language does not have an unspecified [ "n" ].

The twenty-seventh phonological peculiarity to consider here is when a language has a voiceless aspirated dental/alveolar sibilant affricate [ "tsʰ" ] but lacks a voiceless aspirated velar plosive [kʰ].

This is a peculiarity infringing a norm holding between segments of different manners and places of articulation, but sharing the feature-values for voicing and aspiration. The violated norm, [ "tsʰ" ] □ → □ [kʰ], is supported in 24 out of 25 relevant languages, belonging to 8 language families from EURASIA and the AMERICAS. The language exhibiting the phenomenon is Chipewyan from the Na-Dene family. Chipewyan has the plain [k] and the labialized variant [kʷ], but not really the expected aspirated [kʰ].

The twenty-eighth phonological peculiarity we consider is when a language has a voiced palato-alveolar sibilant affricate [dʒ] but it lacks a voiced bilabial plosive [b].

This phonological peculiarity, like the preceding one, violates a universal holding between an affricate and a plosive. These segments have the same feature-value for voicing, viz. "voiced". The violated universal, [dʒ] □ → □ [b], is massively supported in 108 languages, members of 17 language families situated in all 4 geographical areas. There are 5 languages exhibiting the rare phenomenon. They are Zulu (Niger-Kordofanian), Eyak (Na-Dene), Irish (Indo-European), Ache (South Amerind), and Achumawi (North Amerind). As already mentioned, Eyak has no bilabial plosive at all, and Irish has

two variants of the segment, rather than the expected voiced bilabial plosive [b]. Zulu, Ache, and Achumawi have only the voiceless bilabial plosive [p].

The twenty-ninth phonological peculiarity occurs in languages which have a voiced alveolar sibilant affricate [dz] but still lack a voiced bilabial plosive [b].

This is a similar peculiarity to the one above. The violated universal, [dz] implies [b], is supported in 23 out of 24 relevant languages. The corroboration comes from languages belonging to 5 language families from AUSTRALASIA, AFRICA, and the AMERICAS. There is just one exception to this universal. The language exhibiting the phenomenon is Cherokee (North Amerind), which, as we have already mentioned, has no bilabial plosive at all.

The thirtieth phonological peculiarity to mention is when a language has a voiced alveolar sibilant affricate [dz] but the language lacks a voiced velar plosive [g].

This is yet another peculiarity involving the violation of a universal holding between an affricate and a plosive. The two segments share the feature-value “voiced”, but have different places of articulation. The violated universal, [dz] → [g], is supported in 23 out of 24 languages from 17 language families from all 4 geographical areas. The only language in which this rare phenomenon shows up is Kwakw’ala, a North Amerind language. Instead of the expected voiced velar plosive [g], Kwakw’ala has its labialized variant [g<sup>w</sup>].

The thirty-first phonological peculiarity occurs in languages that have a voiceless palato-alveolar sibilant affricate [tʃ] but still lack a voiceless velar plosive [k].

Yet another peculiarity involving the violation of a universal holding between an affricate and a plosive. The two segments share the feature-value “voiced”, but have different places of articulation, palato-alveolar vs. velar. The violated universal, [tʃ] → [k], is massively supported in 182 languages, belonging to 16 language families from all 4 geographical areas. The languages exhibiting the peculiar phenomenon are Alambak (Papuan), Kohumono (Niger-Kordofanian), Hupa (Na-Dene), Farsi and Irish (Indo-European), and Somali (Afro-Asiatic). Hupa has no velar plosive at all. Alambak, Farsi, and Somali have only the aspirated version [k<sup>h</sup>] of the expected voiceless velar plosive [k]. Kohumono has both the aspirated [k<sup>h</sup>] and the labialized aspirated [k<sup>wh</sup>], and Irish has [k<sup>h</sup>] and the palatalized variant [k<sup>h</sup>].

The thirty-second phonological peculiarity to consider is when a language has a voiceless palato-alveolar sibilant ejective-affricate [tʃʼ] but lacks a voiceless velar ejective-stop [kʼ].

Another peculiarity holding between segments of different manners of articulation. The violated universal, [tʃʼ] → [kʼ], is corroborated in 43 languages from 8 language families from EURASIA, AFRICA, and the AMERICAS. The only language conforming to this peculiarity is Hupa (Na-Dene). Hupa has a voiceless uvular ejective-stop [qʼ] rather than the expected voiceless velar ejective-stop [kʼ].

And the final, thirty-third, peculiarity which we will consider in this article is observed when a language has a voiced dental/alveolar r-sound [ʀ] but lacks a voiced labial-velar approximant [w].

This is yet another peculiarity infringing a norm holding between segments of different manners of articulation. The violated universal, [ʀ] → [w], is supported in 35 out of 36 relevant languages. The supporting languages are members of 12 language families from all 4 geographical areas. The only language exhibiting the peculiar phenomenon is Wappo (North Amerind). Wappo has other approximants but not the expected labial-velar one [w].

#### 4. Peculiar Languages and Degrees of Peculiarity

In Section 3, we listed 33 phonological peculiarities involving a conditional absence in the segment inventory of a language. Our computer system listed all and only the peculiarities of the type studied that hold in UPSID-451.

We can measure the “degree of phonological peculiarity” of a language by counting the number of peculiarities it has, or what is the same, counting the number of universals it violates. The more the peculiarities of a language (or the more universals it violates), the more phonologically peculiar it is.

Table 2 shows the number of languages having none, one, two or more phonological peculiarities.

Table 2. Number of languages having none, one or more phonological peculiarities

| No. of peculiarities in a language | No. of languages with this No. of peculiarities |
|------------------------------------|---|
| 0                                  | 391   |
| 1                                  | 43  |
| 2                                  | 12  |
| 3                                  | 3   |
| 4                                  | 1   |
| 5                                  | 0   |
| 6                                  | 1   |

It is seen in Table 2 that 391 out of the 451 or 87% of the languages investigated do not manifest any of the 33 peculiarities listed in the preceding section. 43 languages manifest only a single peculiarity and can therefore be said to be only “mildly phonologically peculiar” (relative to the peculiarity type studied). The remaining 17 languages, manifesting two or more deviations from the norm, can be said to be “strongly phonologically peculiar”.

It is instructive to look more closely at the strongly phonologically peculiar languages and the peculiarities they have.

The following observations can be made:

First, no language was found that has more than six peculiarities, or violates more than six universals (Irish, the most strongly peculiar language in our data set, has six deviations). Generally, languages that have more than two peculiarities are rare (only five languages in our data set, viz. Eyak, Hupa, Cherokee, Klao, and Irish). This suggests some limit to language variation (relative to the examined peculiarity type): languages can indeed be idiosyncratic, but only to a certain limit, not exceeding six deviations.

Table 3. Seventeen strongly phonologically peculiar languages

| No. of Peculiarities | Language  | Peculiarities                                    |
|----------------------|---|--|
| 2 Peculiarities      | <i>Kewa</i> (Papuan/Trans-New-Guinea)           | has [g] but lacks [b]<br>has [t] but lacks [k]   |
| „                    | <i>Usan</i> (Papuan/Trans-New-Guinea)           | has [p] but lacks [k]<br>has [t] but lacks [k]   |
| „                    | <i>Vanimu</i> (Papuan/Sko)                      | has [p] but lacks [k]<br>has [t] but lacks [k]   |
| „                    | <i>Beembe</i> (Niger-Kordofanian/Bantoid)       | has [p] but lacks [k]<br>has [t] but lacks [k]   |
| “                    | <i>Kohumono</i> (Niger-Kordofanian/Cross-River) | has [t] but lacks [k]<br>has [tʃ] but lacks [k]  |
| “                    | <i>Diyari</i> (Australian/Pama-Nyungan)         | has [d.] but lacks [b]<br>has [d.] but lacks [g] |

|                 |  |   |
|-----------------|--|---|
| “               | <i>Wapishana</i> (South American/M. Arawakan)  | has [d] but lacks [b]<br>has [g] but lacks [b]  |
| “               | <i>Bella Coola</i> (North American/Almosan)    | has [p] but lacks [k]<br>has [t̥] but lacks [k]   |
| “               | <i>Kwakw’ala</i> (North American/Almosan)      | has [j] but lacks [g]<br>has [dz] but lacks [g]   |
| “               | <i>Mixe</i> (North American/Penutian)          | has [d̥] but lacks [b]<br>has [g] but lacks [b]   |
| “               | <i>Tzeltal</i> (North American/Penutian)       | has [p] but lacks [k]<br>has [t̥] but lacks [k]   |
| “               | <i>Zuni</i> (North American/Penutian)          | has [̥t̥] but lacks [k]<br>has [p] but lacks [k]  |
| 3 peculiarities | <i>Eyak</i> (Na-Dene/Eyak)                     | has [̥d̥] but lacks [b]<br>has [g] but lacks [b]<br>has [dʒ] but lacks [b]  |
| “               | <i>Hupa</i> (Na-Dene/Athabaskan)               | has [t] but lacks [k]<br>has [tʃ] but lacks [k]<br>has [tʃʰ] but lacks [kʰ]   |
| “               | <i>Cherokee</i> (Northern American/Keresiouan) | has [d] but lacks [b]<br>has [g] but lacks [b]<br>has [dz] but lacks [b]  |
| 4 peculiarities | <i>Klao</i> (Niger-Kordofanian/Krue)           | has [̥t̥] but lacks [k]<br>has [p] but lacks [k]<br>has [gb̥] but lacks [g]<br>has [̥d̥] but lacks [̥n̥]  |
| 6 peculiarities | <i>Irish</i> (Indo-European/Celtic)            | has [d] but lacks [b]<br>has [d̥] but lacks [b]<br>has [g] but lacks [b]<br>has [dʒ] but lacks [b]<br>has [tʃ] but lacks [k]<br>has [ŋ] but lacks [m] |

Secondly, regarding the content of phonological peculiarities, it is seen that all strongly peculiar languages turn out to have a “gap” in their (oral or nasal) stop systems. Thus, they illegitimately lack only one of the following segments: [b], [k], [g], [kʰ], [̥n̥], and [m]. The missing segments belong to the bilabial, dental/alveolar and velar series.

It is interesting to note that many of the strongly peculiar languages lack only one segment, even though they have several peculiarities, or violate several diverse normative rules or universals. Examples are Usan, Vanimó, Beembe, Kohumono, Wapishana, Bella Coola, Kwakw’ala, Mixe, Tzeltal, Zuni, Eyak, and Cherokee. E.g. say Eyak has just one gap in its segment inventory, viz. it lacks a [b], but violates three universals with the antecedents [̥d̥], [g], and [dʒ], any one of which implies the consequent [b].

The languages that lack most segments (=have most gaps) are: Klao (lacking three segments, [k], [g], and [̥n̥]), and Irish (also lacking three segments, [b], [k], and [m]). Thus, complementing the conclusion in point (1) above, we can say that the limit to language idiosyncrasy (relative to the examined peculiarity type) is violation of no more than six universals and no more than three gaps in segment inventory.

And, thirdly, regarding the language family/areal distribution of the strongly phonologically peculiar languages, it is interesting to note that they seem to be clustered into the Papuan family (3 languages), the Niger-Kordofanian family (3 languages), and the Amerindian languages (9 languages).

The observed circumstances do not seem random, and therefore call for an explanation. Indeed, why cannot languages violate more than six universals and cannot have more than three segment gaps? Or why are the missing segments only bilabial, dental/alveolar or velar stops to the exclusion of all other segments? Or why should the strongly peculiar languages be genetically and areally so unevenly distributed?

## 5. Conclusion

The contributions of the paper can be summarized as follows. We propose a typology of phonological peculiarities concerning the content of segment inventories of languages. One type of phonological peculiarity, amounting to an illegitimate lack of a segment predicted by the presence of another segment, is described. Our computerized discovery of 33 peculiarities encompasses all and only the phonological peculiarities of the type studied and thus our description satisfies the adequacy conditions for completeness and soundness. The degree of phonological peculiarity of a language can be measured by the number of peculiarities it has, or equivalently, by the number of universals it violates. It is observed that 87% of the languages in UPSID-451 do not violate any universal, while the remaining languages can appropriately be split into “mildly phonologically peculiar” (a majority violating just one universal), and “strongly phonologically peculiar” (a minority violating more than one universal). The strongly peculiar languages found are observed to be both subject to only a limited variability (they cannot violate more than six universals or have more than three segment gaps) and be quite unevenly distributed from a genetic and areal perspective. The latter circumstances are interesting and need an explanation.

We will conclude noting that among other consequences, our results have an important consequence regarding historical reconstruction of languages. Universals have been traditionally invoked in language reconstruction (at least as far back as the work of Roman Jakobson). The rationale behind this is the assumption that extinct languages must not have been essentially different from extant languages, so that patterns or universals that have been established on the basis of extant languages must have held true also for the unknown proto-languages. Put differently, language reconstruction on a typological basis assumes that reconstructed languages “generally should not violate universals”. Here, we make this assumption more specific, suggesting an upper bound on the number of universals violations and segment gaps.

## References

- Comrie, B. 1981. *Language Universals and Linguistic Typology*. Chicago: Chicago University Press.
- Croft, W. 1990. *Typology and Universals*. Cambridge: Cambridge University Press.
- Derbyshire, D. & G. Pullum. 1981. Object-initial Languages. *International Journal of American Linguistics* 47, 192-214.
- Dryer, M., M. Haspelmath, D. Gil & B. Comrie (eds.). (forthcoming). *World Atlas of Language Structure*. Oxford: Oxford University Press.
- Gamkrelidze, T. V. 1978. On the Correlation of Stops and Fricatives in a Phonological System. In J. Greenberg (ed.), *Universals of Human Language* 2: 9-46. Stanford: Stanford University Press.
- Gerov, N. 1897. *Rečnik na Bălgarskii Ezik*. (A Dictionary of Bulgarian Language). Plovdiv: Druzestvenna pečatnitsa “Săglasie”.



- Greenberg, J. 1966. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In J. Greenberg (ed.), *Universals of Language* 73-113. Cambridge, MA: MIT Press.
- \_\_\_\_\_. 1966. *Language Universals, with Special Reference to Feature Hierarchies*. The Hague: Mouton.
- Keenan, E. 1978. The Syntax of Subject-final Languages. In W. Lehmann (ed.), *Syntactic Typology* 267-327. Austin: University of Texas Press.
- Maddieson, I. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.
- \_\_\_\_\_. 1991. Testing the Universality of Phonological Generalizations with a Phonetically Specified Segment Database: Results and limitations. *Phonetica* 48, 193-206.
- \_\_\_\_\_. 1999. In Search of Universals. In *Proceedings of the 14th International Congress of Phonetic Sciences* 2521-28. San Francisco.
- Maddieson, I. & K. Precoda 1991. Updating UPSID. *UCLA Working Papers in Phonetics* 74, 104-114.
- McClave, J. & F. Dietrich 1988. *Statistics*. San Francisco & London: Dellen Publishing Company & Collier MacMillan Publishing Company, 4th edition.
- Odden, D. 2003. Languages and Universals. *Journal of Universal Language* 4, 33-74.
- Pericliev, V. 1999. Further Implicational Universals in Greenberg's Data (a Computer-generated Article). *Săpostavitelno ezikoznanie (Contrastive Linguistics)* 24, 39-51.
- \_\_\_\_\_. 2000. More Statistical Implicational Universals in Greenberg's Data (another Computer-generated Article). *Săpostavitelno ezikoznanie (Contrastive Linguistics)* 25, 115-125.
- \_\_\_\_\_. 2002a. Economy in Formulating Typological Generalizations. *Linguistic Typology* 6, 49-68.
- \_\_\_\_\_. 2002b. A Linguistic Discovery System that Verbalises its Discoveries. In *COLING, Proceedings of the 19th International Conference on Computational Linguistics* 1258-62, August 24 -September 1. Taipei, Taiwan.
- \_\_\_\_\_. 2003. An Appraisal of UNIVAUTO---the First Discovery Program to Generate a Scientific Article. In G. Grieser, Y. Tanaka, & A. Yamamoto (eds.), *Proceedings of the 6th International Conference on Discovery Science (DS 2003)*, Sapporo, Japan, 429-436. Lecture Notes in Artificial Intelligence vol. 2843. Berlin, Heidelberg, New York: Springer-Verlag.
- Plank, F. 1986. Paradigm Size, Morphological Typology, and Universal Economy. *Folia Linguistica* 20, 29-48.
- Plank, F. & E. Filimonova. 2000. The Universals Archive: A Brief Introduction for Prospective Users. *Sprachtypologie und Universalienforschung* 53, 109-123.
- Pullum, G. K. 1991. *The Great Eskimo Vocabulary Hoax*. Chicago: The University of Chicago Press.