CONTRIBUTORS

Alan Bell
D. N. S. Bhat
Dwight Bolinger
John Crothers
Charles A. Ferguson
Thomas V. Gamkrelidze
Joseph H. Greenberg
Larry M. Hyman
Ian Maddieson
Merritt Ruhlen
Russell Ultan
Marilyn May Vihman

Universals of Human Language

Edited by Joseph H. Greenberg

Associate Editors:
Charles A. Ferguson & Edith A. Moravcsik

VOLUME 2

Phonology

Stanford University Press, Stanford, California
1978

Some Generalizations Concerning Initial and Final Consonant Clusters

JOSEPH H. GREENBERG

ABSTRACT

Based on a sample of 104 languages, 40 universals regarding initial and final clusters are formulated. These fall into five main groups: 1) the marked status of clusters as such; 2) preferred types of assimilation; 3) preferred types of dissimilation; 4) preferences based on the relation to the peak of the syllable; 5) preferences for certain consonantal types over others not dependent on factors involved in groups 2), 3) and 4).

This article is a somewhat revised version of one originally published in Russian in Voprosy Jazykoznanija (1964) 4. 41-65. An English version appeared subsequently in Linguistics 18. 5-34 (1965). It was subsequently utilized and commented on in Charles E. Cairns, Markedness, neutralization and universal redundancy rules, Language 45.4. 863-85 (1969); Doris L. Pertz, Sensitivity to phonological universals in children and adults, doctoral dissertation, Columbia University (1973), and Doris L. Pertz and Thomas G. Bever, Sensitivity to phonological universals in children and adolescents, Working Papers on Language Universals 13. 69-90 (December 1973).

I am particularly indebted to Doris L. Pertz for critical comments. The only substantial change incorporated into this version is owing to her. Former universal 34, asserting the existence of at least one initial cluster of liquid + nasal implies at least one of liquid + obstruent has been withdrawn through lack of sufficient empirical evidence, and subsequent universals have been renumbered.

The original study was supported by the National Institute of Health (M8 0739-01) and by a grant in aid of my personal research from the Behavioral Sciences division of the Ford Foundation. This assistance is gratefully ackowledged.

In the present study, which is of a preliminary nature, a number of generalizations are proposed regarding initial and final consonant sequences, based on a sample of approximately one hundred languages. All assertions made here are to be understood as not claiming any validity beyond this sample. It is, of course, reasonable to conclude that, although exceptions are to be expected with further investigation, they should be few in number and that, therefore, at least a statistical validity for the statements made here can be claimed.

Most previous study of consonant clusters has been related more or less directly to the question of the possible functional definition of phonemes in terms of their behavior in combinations. Almost the only attempt to generalize about the characteristics of consonant clusters is to be found in an article of Trnka, which was subject to a critique by Trubetskoĭ.[1] The most important suggestion of Trnka in relation to the present paper is that "...phonemes differentiated by a mark of correlation never combine in the same morpheme ..." As indicated by Trubetskoĭ, this is not true as a completely general statement since, for example, nasal and voiced homorganic stop is not only a permitted but even a favored combination. It does, however, hold for some features under certain conditions. The generalizations 11, 12, and 13 below are specific cases of this principle. Trnka also pointed out the incompatibility of combinations of two different kinds of sibilants, reflected in generalization 16 of the present paper. A further point in Trnka's discussion which has proven useful is his appreciation of morpheme boundaries as allowing combinations forbidden internally in the morpheme. This factor is involved in the statement of a number of the universals presented here.

The only other suggestion which has proven useful for this study is that of Hjelmslev in regard to resolvability, that is, the principle that longer consonant sequences in general contain as partial sequences shorter ones which are likewise occurrent.[2] This point is incorporated in generalization 2.

[1] B. Trnka, General Laws of Phoneme Combinations, Travaux du Cercle Linguistique de Prague 4. 75-61 (1931) and N. Trubetskoĭ, Principes de phonologie (Paris, 1949), 264-8. It is indicative of how little has been done in this area that no universals of consonant combinations figure in the general table of universals to be found in B. Uspenskiĭ's review of Universals of language (Cambridge, 1962) in Voprosy Jazykoznanija 5. 121-9 (1963).

[2] L. Hjelmslev, On the principles of phonematics, Proceedings of the Second International Congress of Phonetic Sciences (1936), 49-54.

The language employed in the present study are listed in the appended bibliography with a numerical designation for convenience of reference in the text of the article and with the chief sources employed for each language. Where I consider the material to be significantly defective in completeness or in phonetic information, this has likewise been indicated. It is clear that even incomplete information can yield relevant evidence for some generalizations, while being insufficient for others. Thus, a generalization that a lateral is never followed by a vibrant ("r" sound) can be refuted by a single valid example from a description which is seriously incomplete, while being insufficient to refute a generalization of other types, e.g. that every language which has an initial obstruent followed by a nasal also has some combination of initial obstruent followed by a liquid. In such a case we might find /kn/ among the clusters reported but no examples such as /kr/ because of the insufficiency of the material.

The exclusion of medial clusters was dictated chiefly by practical considerations. The number of such combinations is often very large. Further, many of the sources utilized in the present study contain statements regarding initial and final clusters only. The study of medial clusters also raises some theoretical problems not present in the case of initial and final clusters. For example, in languages with syllabic initial and final single consonants or clusters, except for possible word-sandhi phenomena, the medial clusters produced at word boundaries are in general predictable from initial and final combinations. Such clusters should evidently be distinguished from those which are word-internal and which may or may not be present in languages independently of the question as to whether word-boundary clusters exist. Again, morpheme boundary and morpheme internal clusters should be distinguished among word internal clusters. For these reasons a study of medial clusters would be much more elaborate and difficult to undertake from existing data.

Preliminary to setting forth specific results, it will be necessary to consider a few problems of theoretical method. These are only briefly discussed, insofar as they affect the procedure employed in the present study. The very phrase consonant cluster raises definitional problems which have to be decided in order to compare consistently data from different languages. A first question, then, concerns the definition of consonant. In accordance with the usual notion, consonant is understood here in terms of function in the syllable, i.e. as a margin rather than peak. This will mean, however, that the syllabic and non-syllabic allophones of the same

phoneme, e.g. [u], [w] in some languages, will have only the latter allophones reckoned as consonants. Strictly speaking, if there are both consonant and vowel allophones of the same phoneme, then this distinction is an irrelevant feature for the phoneme concerned at the same time that it is a central question in the present study. It is clear that this same basic consideration arises at other points. Consider, for example, the generalization stated below as number 5. This asserts that the presence of final heterorganic nasal obstruent combinations in any language implies the existence of homorganic combinations. Now many languages have final combinations which are phonetically [ŋk] or [ŋg] where [ŋ] is to be considered as a member of an /n/ phoneme containing [n] and [ŋ] among its variants. A naive reading of a phonemic transcription /nk/ would lead to its classification as a heterorganic combination. Strictly speaking, however, since in this case the difference between dental-alveolar and velar position is a non-distinctive feature, we have no right to classify it as either heterorganic or homorganic on a phonemic basis. Yet a classification of the sounds on a phonetic basis allows us to compare languages and draw general conclusions.[3] Further, it is evidently the same tendency operating in different languages which leads to this particular allophonic distribution of [n] and [ŋ] in one language while in another which contains distinct /n/ and /ŋ/ phonemes, it leads to a preference for sequences like /nd/ and /ŋg/ while it disfavors /ng/ and /ŋd/.

Since allophonic information is, for the reasons just adduced, essential to the present study, statements of phonemic combinations in the literature had to be supplemented by phonetic descriptions in the same or other works. For this reason also, data from languages of the past were not included in the sample of 104, although in certain instances such languages were taken into consideration when the absence of phonetic detail did not seriously affect a particular hypothesis.

The other part of the phrase 'consonant cluster' also raises difficulties. It is well known that for certain classes of sounds, the decision as to whether we have a cluster or succession of phonemes

[3] What is said here is in close agreement with the view of E. Fischer-Jørgensen that "...the tendencies to free combinations or to definite restrictions between different parts of the syllable seem to be more easily formulated when the parts of the syllable are defined on a phonetic basis." (On the definition of phoneme categories on a distributional basis, Acta Linguistica 7. 8-39 (1952)).

as against a single phoneme has not produced a usable unarbitrary criterion which meets with general consent. Thus, in languages with the phonemic contrast of unaspirated and aspirated consonants, the alternative solution as a single phoneme, e.g. /pʻ/ or a cluster /ph/ depends on considerations of symmetry or "pattern" in the sound system in general which often leads to individually different solutions even for the same language. It seems unavoidable, for purposes of valid comparison among languages, that one must make a decision in such matters which, even though it may be arbitrary, will be consistently applied. In general the sequences at issue are well characterized in N. Trubetskoĭ's classic work on phonology as "produced by a single articulatory movement or by means of a progressive dissociation of an articulatory complex."[4] In the former of these cases, that of the affricates, I have considered the articulation to be a cluster of stop + fricative. The latter have all been considered single consonants. These include aspirated, glottalized, labialized, palatalized, velarized, and pharyngealized sounds. The sequence nasal + homorganic voiced stop, treated as a single phoneme in some languages by some analysts, e.g. FIJIAN, is here always treated as a cluster.

A further problem of definition arises regarding the terms 'initial' and 'final.' In principle, initial and final in the utterance is intended. Most studies are in terms of word initial and final which generally comes to the same thing as utterance initial and final. Where there are word sandhi rules, however, only the utterance initial or final forms are considered in the formulation of generalizations. Where word boundaries occur between members of a consonantal sequence which is actually or potentially utterance initial or final, the entire sequence is viewed as a valid cluster. Thus, RUSSIAN v dome, 'in the house,' is considered as having an initial consonant cluster [vd].

A particularly vexing question concerns the treatment of clusters in borrowed words. The line between forms recent enough to be considered borrowings and those which can be considered fully assimilated into the language is difficult to draw. Moreover, some of the studies utilized distinguished borrowed from native clusters, while others did not. As far as possible this distinction was made in compiling the material from the original sources where given, from etymological dictionaries, or from my own knowledge. However, since such data were not obtained from all of the languages,

[4] N. Trubetskoĭ 1949: 58.

it is to be understood that, as a general rule, borrowed are included along with indigenous clusters in the statements of the present article. While the exclusion of combinations in borrowed forms would change the typological assignment of certain languages in the tables below, in no case would such a change have been sufficient to invalidate a hypothesis.

A further question concerns the distinction between clusters which appear only with contained morpheme boundaries and those which do not. Again an attempt was made to record this distinction as far as possible, but in some instances the information at my disposal was not sufficient to resolve this question. In several instances where the relevant evidence was sufficient, the existence of morpheme boundaries figures in the statement of generalizations.

Finally, there is the question as to what particular variety of a language is intended, or even which speech tempo since, in many instances, particular consonant sequences which occur in slower and careful speech are contracted or assimilated in more rapid or more colloquial instances. In general, I have treated so-called "standard" languages simply because they have usually been more carefully described from the phonetic point of view. In general, the source cited in the bibliography will give sufficient indication of the particular variety of a language which is being considered. For HINDI the RANKHANDI dialect was used and for KAREN, SGAW. On the question of speech styles, I have in general utilized the lento forms as against the allegro, but I have recorded data concerning such variation where they were present. It is plausible to consider that allegro forms give important insight into the identification of "difficult" and less-favored sequences and into the direction of historical change. However, their systematic treatment has been left for further investigation.

In general it proved useful to distinguish between initial and final clusters as separate systems with distinct though often similar properties. In the sample of 104 languages there were found to be 90 initial systems and 62 final systems. Although the possibilities of certain connections are not to be excluded, in general initial and final systems seem to function independently and it was not possible to formulate any generalizations connecting them.

The first set of hypotheses concern properties of initial and final systems which correlate with the length of the sequences.

1. For initial and final systems, if x is the number of sequences of length m and y is the number of sequences of length n and $m > n$, and p is the number of consonant phonemes, then $\frac{x}{p^m} \leq \frac{y}{p^n}$.

In other other words the propertion of the logically possible combination utilized decreases or remains the same with increasing length of the sequences. This may be illustrated for ENGLISH initial clusters as follows: the number of consonant phonemes are 22. All of these except /ž/ and /ŋ/ occur as single phonemes. The logically possible sequences of length 2 are $22^2 = 484$. Of these 28 occur. For length 3 the logically possible number of combinations is $22^3 = 10,648$. Of these only 8 occur. No sequences of length greater than 3 are found. Hence,

$$\frac{20}{22}(L=1) > \frac{28}{484}(L=2) > \frac{8}{10,648}(L=3) > 0\,(L=4) = 0\,(L=5) \text{ etc.}$$

It will be noted that the absolute number of combinations of length 2, i.e. 28, is greater than those of length 3, i.e. 8, etc. However, in the limiting case L=1, in this as in many other instances the number of combinations of length 2 is greater than length 1. We can therefore make the following statement regarding absolute length:

2. For initial and final systems, if x is the number of sequences of length m and y is the number of sequences of length n, and $m > n$ and $n \geq 2$, then $x \leq y$.

The statement in the "Memorandum concerning language universals," in Universals of language that "If syllables containing sequences of n consonants in a language are to be found as syllabic types, then sequences of $n-1$ consonants are also to be found in the corresponding position (prevocalic or postvocalic) except that CV $\rightarrow$ V does not hold," can be deduced as a corollary from either 1 or 2 above, insofar as it refers to word initial and final as a special case of syllabic initial and final.[5] Further, 1 and 2 make no assertion concerning L=0 since in this case there is no question of combinations.

In general the validity of 1 and 2, to which no exception was found in the 104 languages of the sample, provides objective evidence of the "difficulty" of clusters. This would seem to correlate with the diachronic tendency towards their simplification, since any simplification automatically reduces the number, both absolutely and

[5] Universals of language, ed. by J. H. Greenberg (Cambridge 1963), p. 263.

proportionally, of sequences of the length subject to reduction and increases the number of shorter sequences.

The next statement refers to the property of resolvability which was first suggested by Hjelmslev. A sequence is here said to be completely resolvable if every continuous subsequence also occurs. For example, if in a language initial fstr occurs then if fs, st, tr, fst and str all occur, it is completely resolvable. If some of these occur but not otherwise, it is partially resolvable, and if none occurs, it is non-resolvable.

3. Every initial or final sequence of length m contains at least one continuous subsequence of length m - 1.

In the overwhelming majority of instances sequences are completely resolvable. In the weaker form asserted here there are still a very small number of unresolvable sequences in the material collected. These were from CHATINO, PAME, and COEUR D'ALENE and totalled 10 in all. Thus this assertion has only statistical validity but far beyond chance within any reasonable confidence limit. That this is not a chance phenomenon in the individual languages can be illustrated from initial clusters in ENGLISH. Here all 8 clusters of length 3 (skw, skr, skl, skj, spl, spj, spr and str) are completely resolvable. Now the only initial clusters of length 3 that can be formed from those of length 2 that conform to the requirement of complete resolvability are the following: spl, spr, spj, str, stw, skl, skr, skj, skw, sfl, sfr, sfj. That 8 clusters chosen at random out of $20^3 = 8,000$ logically possible combinations should all fall within a set of 12 is, of course, highly significant statistically. This generalization could be restated as follows: For every initial and final system and for every length, the number of completely resolvable sequences is greater than the number of those which are not completely resolvable. This statement has no exceptions in the present material.

The reason for the phenomenon of resolvability is, at least partly, that longer sequences are formed from shorter sequences by morphological or syntactic combination. The latter occurs, for example, in initial consonant sequences in the SLAVIC languages where prepositions consisting of a single consonant are found. From this it should follow, as a general result, that the longer a sequence the more likely it is to contain one or more morpheme or word boundary. Unfortunately, it was only possible to classify the sequences in this regard for a few languages. Two classes of sequences were distinguished, those which occurred exclusively with a contained

morpheme or word boundary and those which occurred in at least some cases without such an internal boundary. The following generalization is therefore merely a probable conjecture which was verified in the few cases presented below.

4. In all initial and final systems, if there are sequences of length m and n and m > n, then the proportion of sequences of length m which only occur with internal morpheme boundaries is equal to or greater than the proportion of such sequences which occur of length n. Since every word boundary will also be a morpheme boundary, it is sufficient to state the above generalization in terms of morpheme boundaries.

Among the cases investigated were Cœur d'Alene and Dutch. In Cœur d'Alene initial clusters the ratios were as follows:

$$\frac{0}{42}\,(L=1) < \frac{83}{88}\,(L=2) < \frac{42}{42}\,(L=3) = \frac{2}{2}\,(L=4).$$

The final ratios were:

$$\frac{0}{42}\,(L=1) < \frac{62}{77}\,(L=2) < \frac{11}{12}\,(L=3) < \frac{1}{1}\,(L=4).$$

Dutch initials do not have morpheme boundaries. Hence the results will of course be 0 (L = 1) = 0 (L = 2) = 0 (L = 3). For finals we have:

$$\frac{0}{11}\,(L=1) < \frac{2}{38}\,(L=2) < \frac{20}{32}\,(L=3) < \frac{10}{11}\,(L=4) < \frac{2}{2}\,(L=5).$$

We now turn to hypotheses of a more specific nature. These may be classified from two points of view, logical and material. Logically, we have an implication if, whenever a system has a particular property ϕ it also has some other, ψ. The majority of the hypotheses stated here are of this nature. If properties ϕ and ψ mutually imply each other, they are equivalent. Every language with property ϕ has ψ and vice versa. Another logical type consists of assertions regarding the intersection of all systems. Such properties may be considered basic in that they apply to all systems even those with the fewest combinations. If a particular logically possible type of combination does not occur in any system it belongs to the complement of the class union of all the systems. In this case even the most extensive system does not possess the particular property.

From the material or phonetic point of view hypotheses always have to do with the favoring of certain types of consonants or

consonant sequences over other consonants or consonant sequences. Such hypotheses may be classified as contextual, ordinal and absolute. By contextual will be meant the tendency for one class of consonants to be favored over another in a particular context. Insofar as the context itself shares similarities with the class of consonants in question it is assimilatory, insofar as it differs it is dissimilatory. It is evident that these names which are properly applied to diachronic processes are appropriate here insofar as the operation of these types of change will tend to produce the given contextual properties in sequences described synchronically. By an ordinal hypothesis will be meant one which concerns the favoring of particular classes of consonants not only by context but also by reference to order. Such hypotheses are usually related to the structure of the syllable and to the tendency of certain classes of sounds to be peripheral (e.g. stops) and of others to be central (e.g. liquids, semivowels) in reference to the central vocalic nucleus. Finally, by an absolute hypothesis is meant one which concerns an overall tendency to favor one class of sounds as against another independently of context or orientation in the syllable. Thus, there is a tendency to favor unvoiced over voiced combinations, of unglottalized as against glottalized, and of sibilants as against other spirants. In general we consider hypotheses here in conformity with the foregoing classification into assimilatory, dissimilatory, ordinal and absolute.

The first of our specific hypotheses has to do with the favoring of combinations which are homogeneous in respect to voicing over those which are heterogeneous.

5. There are no initial or final systems in which all obstruent combinations are heterogeneous in regard to voicing.

The force of this statement derives from the fact that there are languages in which all obstruent combinations are homogeneous in regard to voice, e.g. ENGLISH, POLISH.[6] An assertion of similar type cannot be made with regard to obstruent-sonant combinations because the sonant is most usually voiced whether in combination with a voiced or unvoiced obstruent.

[6]An exception outside of the sample is PALAYCHI KAREN all of whose (initial) obstruent clusters are unvoiced + voiced. A revised statement that in all languages with obstruent clusters, there are some with first member unvoiced would, as far as my knowledge goes, be without exception.

Another hypothesis of assimilation concerns the preference of nasals for the following voiced stop to be at the same point of articulation (homorganic combinations).

6. In final systems the existence of at least one sequence consisting of a nasal (voiced or unvoiced) followed by a heterorganic obstruent implies the existence of at least one sequence consisting of a nasal (voiced or unvoiced) followed by a homorganic obstruent.

As with other statements of the form $\phi \supset \psi$, the falsity of the opposite implication $\psi \supset \phi$ is tacitly asserted. If it were true, the universal would be stated in the form of an equivalence. As supporting evidence from the sample of 104 languages, we cite by number those which have both final heterorganic and homorganic sequences of nasal followed by obstruent and those which have homorganic sequences without heterorganic. In the implicitly asserted typology, then, one of the four logically possible types, the class of languages with heterorganic but without homorganic sequences of this kind is null.

Languages with final systems containing both heterorganic and homorganic combinations are, then, as follows: 1, 2, 3, 4, 7, 10, 13, 27, 28, 30, 33, 35, 36, 41, 42, 43, 48, 49, 61, 62, 64, 67, 72, 75, 76, 80, 81, 86, 87, 94, 101, 102, and 103. Languages with final systems containing homorganic without heterorganic combinations are: 5, 11, 15, 16, 23, 34, 44, 45, 46, 51, 66, 69, 70, 77, 82, 89, 93, 96, 97, 98. The data for 57 (KASHMIRI) were not sufficient to decide between these two possibilities. The remaining languages with final systems did not have any combinations of nasals and obstruents.

These results can be shown in the following table:

	Heterorganic	~ Heterorganic
Homorganic	33	20
~ Homorganic	0	8

Table 1

A similar statement for initial systems holds in almost all cases, but a number of SLAVIC languages (e.g. RUSSIAN, POLISH, CZECH) are conspicuous exceptions in that they contain initial heterorganic combinations such as mg without having homorganic sequences. It should be noted that for purposes of the present hypothesis

sequences of nasal followed by obstruent are included even when preceded or followed by one or more consonants. Thus, RUSSIAN /mgla/ counts as an instance of a heterorganic cluster.

We now consider hypotheses of dissimilation. The first group has to do with preference for combinations of stop and spirant as against stop + stop or spirant + spirant. The hypotheses presented here refers to sequences of length two only. In fact sequences of three or more stops or three or more fricatives are excessively rare and implicational universals analogous to the following could doubtless be formulated.

7. In initial systems the presence of at least one combination of stop + stop implies the presence of at least one combination of stop + fricative.

In fact languages with stop + stop almost always have both stop + fricative and fricative + stop combinations. Affricates are counted here as stop + fricative as explained in an earlier section. The following languages which have stop + stop combinations also have both stop + fricative and fricative + stop: 2, 10, 14, 16, 27, 28, 30, 31, 32, 34, 38, 41, 42, 44, 59, 61, 77, 79, 80, 81, 82, 93, and 102. Two languages, 47 and 87, have stop + stop and stop + fricative but do not have fricative + stop. Further details are given in Table 2.

	FS·SF	FS	SF	~(FS∨SF)
SS	23	0	2	0
~SS	35	3	20	7

Table 2

8. In final systems the presence of at least one combination of stop + stop implies the presence of at least one combination of fricative + stop. Here again most languages with stop + stop have both stop + fricative and fricative + stop. Languages with all three combinations are: 2, 3, 4, 7, 16, 27, 30, 33, 35, 36, 38, 41, 42, 43, 44, 48, 61, 62, 64, 67, 72, 77, 80, 81, 82, 86, 87, 94, 96, 101. Only one language, 98, had stop + stop and fricative + stop without stop + fricative. The details are shown in Table 3.

SS	FS·SF	FS	SF	~(FS∨SF)
SS	30	0	3	0
~SS	16	10	4	1

Table 3

The slight preference shown for initial stop + spirant and final spirant + stop in the above two statements, suggests a possible connection with syllabic structure in that the more open fricative is in each case closer to the vocalic center of the syllable.

Similar generalizations concerning initial and final fricative combinations are as follows:

9. In initial systems the existence of at least one fricative + fricative combination implies the presence of at least one stop + fricative combination or at least one fricative + stop combination.

In most cases both stop + fricative and fricative + stop combinations are found in languages with fricative + fricative. These are 1, 3, 10, 13, 14, 27, 28, 30, 33, 35, 36, 38, 39, 41, 42, 44, 48, 52, 61, 62, 63, 64, 75, 77, 80, 81, 82, 86, 93, 103, 104. Only one language has fricative + fricative without fricative + stop, 56. Two languages, 49 and 57, have fricative + fricative without stop + fricative. Further information is given in Table 4.

	FS·SF	FS	SF	~(FS∨SF)
FF	31	2	1	0
~FF	29	18	3	6

Table 4

10. In final systems the existence of at least one fricative + fricative combination implies the presence of at least one stop + fricative or at least one fricative + stop combination.

The languages with fricative + fricative combinations which have both fricative + stop and stop + fricative combinations are 3, 4, 7, 13, 18, 27, 30, 33, 35, 36, 41, 42, 43, 49, 61, 62, 64, 67, 75, 76, 77, 79 and 86. Two languages with final fricative + fricative lack the final stop + fricative combination, 34 and 98, and one language, 1, lacks fricative + stop. These and other results are summarized in Table 5.

	FS·SF	FS	SF	~(FS∨SF)
FF	23	2	1	0
~FF	22	3	9	2

Table 5

The general thesis of Trnka regarding the non-occurrence of sequences which differ in only one feature holds in certain limited cases, chiefly those involving differences in laryngeal adjustment such as voicing, voicelessness and glottalization. In regard to the contrast voiced vs. voiceless, sequences of voiced and unvoiced nasals occur, the unvoiced often being phonemically interpreted as /h/. There are legitimate though rare examples of sequences of otherwise identical voiced and unvoiced fricatives. These are initial /sz/, /fv/ and /ky/ in PALAYCHI KAREN, not in the present sample, and final /vf/ in GILYAK. In regard to stops, initial /td/ was noted in BILAAN and KHASI. In both instances the absence of detail in the accompanying phonetic description in the sources suggest the possibility of a svarabhakti vowel. The only example in final systems is CŒUR D'ALENE /dt/ in which, however, there is a morphological boundary between the consonants. It is reasonable to expect that otherwise non-existent combinations will tend to resist change when a morphological boundary is present through the analogical presence of the other allomorphs which contain the sounds in question. It will be seen later that the fact that in these exceptions in initial systems the unvoiced precedes the voiced but in final the voiced precedes the unvoiced is not accidental. Generalizations partly covering this same ground will be formulated below.

The sequence of otherwise identical voiceless and glottalized consonants is likewise extremely rare. The only example found in the sample was initial /kk̓/ in KUTENAI involving a morpheme boundary. Outside of the sample an instance in EYAK, but without phonetic description, cited by Fang-Kuei Li, namely final /q̓q/ and also involving a morpheme boundary was noted.[7] Finally, not a single instance is found in the sample of a sequence of consonants in which the only difference was between voicing in one and glottalization in the other. Thus not only such sequences as */p̓b/ do not occur, but also */mm̓/ etc. are not found. Indeed, the general tendency for glottalized sounds not to occur in clusters with voiced leads to the formulation of a principle which partly overlaps the preceding, namely, that no sequence of glottalized and voiced obstruents occur. The partial dependence of these two statements is shown by the fact that a sequence */bp̓/ violates both. Their non-identity is shown by the fact that */ll̓/ violates the first but not the second, while */p̓g/ violates the second but not the first. None of the statements in this paragraph are limited to sequences of two consonants only. We have, then, the following generalizations:

[7] IJAL 22 (1956), p. 47.

11. An unvoiced stop in initial systems is never preceded immediately by a stop differing only by voicing, and in final systems is never immediately followed by such a stop.

12. In final systems a voiced stop is only followed immediately by a stop differing only in being unvoiced if there is a morpheme boundary between them.

13. No succession of consonants only differing in that one is unvoiced and the other is glottalized occurs in initial or final systems unless there is a morpheme boundary between them.

14. No combination of a voiced and glottalized obstruent is found in either initial or final systems.

Two further hypotheses concern the well-known tendencies of liquid and sibilants not to occur together. While the final sequence rl is fairly common, initial rl only occurs in MITLA ZAPOTEC with morpheme boundary. No example of lr is found in the present material. In regard to sibilants there is a strong tendency for different type sibilants, most commonly [s] and [š], not to combine. Practically all examples occur with morpheme boundary. The only example without morpheme boundary in clusters with only two members was SHILHA BERBER final /zž/. There was no case of final hushing followed by hissing sibilant, e.g. */šs/ in final two-member clusters. These generalizations can be stated as follows:

15. In initial and final systems a lateral is never followed immediately by an r-type sound. In initial system the sequence r + lateral only occurs with morpheme boundary.

16. In final systems, clusters with two members never consist of a hushing followed by a hissing sibilant.

We next consider hypotheses of the ordinal type as defined earlier. A first group concerns the tendency of liquids to follow obstruents in initial systems and precede them in final systems. It is true that the opposite order is sometimes found. In many such cases, particularly in final position in the word or in contact with voiceless consonants, the liquid itself is voiceless. Even where this is not so, statements of an implicational type can be made.

17. In initial systems the existence of at least one sequence containing a liquid, whether voiced or unvoiced, immediately followed by an obstruent implies the existence of at least

one sequence containing an obstruent immediately followed by a liquid.

The following languages have liquid followed by obstruent and also obstruent followed by liquid: 2, 10, 14, 20, 28, 30, 41, 59, 75, 77, 81, 103. The following have no obstruent liquid combinations of any kind: 6, 21, 29, 31, 34, 47, 54, 58, 66, 71, 73, 79, 90, 99, 100. The remainder, representing the most common type, have obstruent followed by liquid without liquid followed by obstruent. These results are set forth in Table 6.

	OL	~OL
LO	12	0
~LO	65	15

Table 6

A comparable hypothesis is valid for final systems.

18. In final systems, the existence of at least one sequence containing a stop immediately followed by a liquid implies the presence of at least one sequence containing a liquid followed by a stop.

The languages with both obstruent +liquid and liquid +obstruent are: 2, 7, 15, 28, 38, 41, 43, 49, 67, 75, 76, 77, 80, 81, 97, 98 and 103. The largest group of languages has liquid + obstruent but not obstruent +liquid. The remaining languages have no obstruent liquid combinations at all. These are 5, 9, 23, 34, 44, 46, 58, 61, 79, 82, 83, 86, 89, 93, 99, 102 and 104. These data are summarized in Table 7.

	LO	~LO
OL	18	0
~OL	27	17

Table 7

Semivowels, of course, show an equal or even greater tendency than liquids to be adjacent to the peak of the syllable. There are a very few cases in the present sample of sequences in which semivowels are separated from vowels by an intervening liquid or nasal, e.g. initial wr and wl in PASHTO and CHATINO wn and yn. Voiceless semivowels occur with intervening consonants of various types. However, there is no valid example of a voiced semivowel separated

from a syllabic peak by an obstruent. In CZECH at least in literary pronunciation, we have sequences such as /jde/ and /jsou/. However, CZECH j is described phonetically by Trávníček as fricative (třená) and possessing consonantal noise (souhláskový šelest) and by Kučera as "articulated with a moderate degree of lamino-palatal friction."[8] We have therefore the following statement:

19. Voiced semivowels are not followed by obstruents in initial systems or preceded by obstruents in final systems.

Although there are, as we have seen, instances of two voiced sonants adjacent to a vowel, such a sequence is never separated from the vowel by an obstruent. Nor are there sequences of three sonants in either initial or final systems. GILYAK phonemic final sequences such as xlŋ and nmŋ have phonetically short vowels between the consonants.[9] Hence we have this generalization.

20. Two successive voiced sonants are always followed by a vowel in initial systems and preceded by a vowel in final systems.

The strong tendency towards voicing assimilation was noted earlier. In a minority of instances sequences with both voiced and unvoiced members occur. An examination of these cases shows that in almost all such instances of combinations which are heterogeneous in voicing, the voiced single consonant or the voiced sequence of consonants is adjacent to the vowel. An overall count in the languages of the sample of the obstruent combinations with two members classified according to voicing was made. Where one of the consonants was partly voiced, i.e. voiced during part of the duration of the consonant and unvoiced during the remainder, the combination was not counted. Where there was free variation of two sequences which differed in voicing of one of the members, each variant was accorded half value. These cases were not numerous. Although the method is obviously a crude one and no significance should be attached to the exact figures, the results are of interest in the present connection as giving at least an approximate notion of the relative frequency of the different types. A total of 1030 two

[8] F. Trávníček, Mluvnice Spisovné Češtiny (Praha, 1951) I: 19 and H. Kučera, The phonology of Czech ('sGravenhage, 1961), p. 28.

[9] Personal communication, Robert Austerlitz.

member clusters were recorded for initial position and 683 for final. The results are stated in Table 8 in percentages of the totals.

	Initial	Final
Unvoiced + Unvoiced	66.70	78.62
Voiced + Voiced	21.65	12.59
Unvoiced + Voiced	10.68	2.05
Voiced + Unvoiced	0.97	6.73

Table 8

For initial systems it holds in general, not for obstruent sequences only, that an unvoiced consonant or succession of consonants preceding a vowel is not itself preceded by voiced segments. The exceptions were almost all instances of nasal + homorganic unvoiced stop followed by a vowel, e.g. initial /nt/ and were noted for CHATINO, CHRAU and PAME (OTOMI). The other instances of the interruption of voicing in an initial sequence were from KHASI, e.g. bt, dp, and BILAAN, e.g. bt, bs. There is also CZECH js, in which as noted above, CZECH j was interpreted as a voiced fricative. Since for KHASI and BILAAN few phonetic details were forthcoming in the descriptions, it is quite possible that, as in similar instances elsewhere, the initial stop /b/, /d/, etc. are unvoiced lenes. The following statement is therefore made with the awareness that the exceptions just noted may be valid.

21. Except for voiced nasal followed by homorganic unvoiced obstruent, an unvoiced consonant or sequence of unvoiced consonants in initial systems immediately preceding a vowel is not itself preceded by one or more voiced consonants.

In the three languages cited above as having voiced nasal followed by unvoiced homorganic obstruents, the far more common combination of voiced nasal followed by voiced homorganic obstruent is likewise found. We have, therefore, the following implicational statement.

22. In initial systems the presence of at least one sequence of voiced nasal + unvoiced homorganic obstruent implies the presence of at least one sequence of voiced nasal +voiced homorganic obstruent.

The symmetrical thesis for final systems corresponding to that of 21 for initial systems, asserts that an unvoiced consonant or sequence of unvoiced consonants immediately preceded by vowel

is not followed by a voiced consonant or sequence of voiced consonants. These two hypotheses taken together amount to the following. There is a voiced center of the syllable consisting of the vowel and possible successive preceding and following voiced consonants but that voicing is normally confined to this nucleus, i.e. that voicing is not interrupted and resumed within the same syllable. For final systems again there are very few exceptions in the present material and these often of doubtful validity. Exceptions include AMHARIC final kd in a single word, CŒUR D'ALENE final tg^w, HUNGARIAN šd described as rare by Vértes and not found in Hall, and a few instances in ARABIC all involving pharyngeals. Corresponding to the situation in initial systems there are cases of unvoiced obstruent followed by voiced nasal, in this case not necessarily homorganic, e.g. PAME tn, kn. There are no cases of voiceless liquids or nasals being followed by voiced consonants. These results may be summarized in the following statement:

23. In final systems, except for unvoiced obstruents followed by a voiced nasal, an unvoiced consonant or sequence of unvoiced consonants following a vowel is not followed by one or more voiced consonants.

A set of ordinal hypotheses may also be formulated regarding the tendency of liquids to be closer to the syllabic peak than nasals. This can be stated in the common implicational form, with the proviso that unvoiced liquids, as might be expected, do not share with voiced liquids the property of being nearer to the syllabic peak than nasals.

24. In initial systems the existence of at least one sequence consisting of a voiced liquid followed by a nasal implies the existence of at least one combination consisting of a nasal followed by a liquid.

By 21, the second member of these combinations must also be voiced, and by 20 this sequence must be followed by a vowel. Languages containing voiced liquid +nasal as well as nasal +liquid are 14, 30, 41, 59, 75, 77 and 81. Languages with nasal + liquid and without voiced liquid +nasal are 16, 19, 25, 26, 37, 40, 43, 51, 55, 56, 57, 60, 72, 80, 82 and 84. Language 2, 10 and 20 do not have nasal + liquid and do have liquid + nasal but the liquid is voiceless. MITLA ZAPOTEC (103) has /r̃N/ without any nasal + liquid combination. However, the distinction here is between lenis and fortis rather than voice or voiceless and /r/ may be voiceless. Little information on voice and voicelessness is given, hence this is

quite possibly not a real exception. Relevant information is given in Table 9.

	N + L	~N + L
L + N	7	(1)
~ L + N	16	66

Table 9

The corresponding hypothesis for final systems can be put in stronger form. Sequences of liquid + nasal are not only favored over nasal + liquid but in general over all other combinations of nasals and liquids, i.e. nasal + nasal and liquid + liquid.

25. In final systems, the existence of at least one sequence consisting of a nasal followed by a liquid, a nasal followed by a nasal or a liquid followed by a liquid implies the existence of at least one sequence consisting of a liquid followed by a nasal. Languages 4, 7, 13, 16, 27, 33, 36, 38, 41, 42, 43, 49, 67, 72, 76, 77, 80, 86 and 98 have combinations of one or more of the types nasal + liquid, nasal + nasal and liquid + liquid along with liquid + nasal. Languages 3, 15, 18, 30, 35, 44, 57, 64, 65, 75, 81. 82, 94 and 102 have liquid + nasal without having any combinations of the three other types. These data are set forth in Table 10.

	NLVNNVLL	~(NLVNNVLL)
LN	19	14
~ LN	0	29

Table 10

The first set of hypotheses of the absolute type has to do with the favoring of unvoiced obstruents over voiced obstruents. The statistical predominance of unvoiced obstruent + unvoiced obstruent over the other three types of voicing combinations in clusters with two members was strikingly displayed in Table 8. A number of non-statistical generalizations are also possible. Thus for initial systems the following statement can be made.

26. In initial systems the existence of at least one combination consisting of two voiced obstruents implies the existence of at least one combination consisting of two unvoiced obstruents. There is here a single exception, the SGAW

dialect of KAREN (56) which has the voiced obstruent combination /bɣ/ but has no sequences with both members voiced. There are, however, combinations in which the first member is unvoiced and second is /ɣ/, for example /pɣ/ and /sɣ/. The languages with both voiced and unvoiced obstruent clusters are 1, 3, 10, 11, 13, 16, 18, 30, 36, 37, 40, 41, 42, 44, 49, 52, 59, 62, 63, 64, 67, 69, 75, 77, 80, 81, 82, 84, 86, 98 and 103. The remainder have only unvoiced obstruent clusters except 56 (KAREN) mentioned as an exception earlier and the following languages which have no clusters of obstruent + obstruent: 8, 9, 25, 29, 50, 57, 60, 74, 78, 90, 91, 95. These data are summarized in Table 11.

		Unvoiced Obst + Unvoiced Obst.	~Unvoiced Obst. + Unvoiced Obst.
Voiced Obst. + Voiced Obst.		31	1
~ Voiced Obst. + Voiced Obst.		46	12

Table 11

It would be possible to eliminate KAREN as an exception by restating 26 so that the presence of voiced obstruent + voiced obstruent implied the presence of at least one combination with initial unvoiced obstruent.

The even more powerful tendency to the unvoicing of final obstruents is shown statistically in Table 8 as well as in Table 12 below in which the proportion of languages with only unvoiced obstruent combinations is greater than in initial position. Also there are no exceptions in final systems.

27. In final systems, the existence of at least one combination consisting of two voiced obstruents implies the existence of at least one combination consisting of two unvoiced obstruents.

Both voiced and unvoiced obstruent clusters are found in 3, 4, 7, 13, 36, 41, 44, 48, 64, 75, 76, 80, 82, 86, 98 and 103. All of the remaining languages have unvoiced obstruent clusters only, except 23, 28, 65, 88 and 89 which have no obstruent clusters at all. These data are presented in Table 12 below.

	Unvoiced + Unvoiced Obst.	~Unvoiced + Voiced Obst.
Voiced + Voiced Obst.	16	0
~Voiced + Voiced Obst.	41	5

Table 12

The dominance of unvoiced over voiced obstruent is shown not only in combination of obstruent with obstruent but also of obstruent with sonant. For initial systems the single case of OSAGE, reported by Wolff, in which one informant had initial [br-] in free variation with [bər-] and the other [bl-] in free variation with [bəl-] and [bəɬ-] but no other initial obstruent + liquid combination prevents the formulation of a single generalization covering all cases of initial obstruent + liquid. However, the following more limited statements have no exception in the present sample.

28. In initial systems the existence of at least one sequence of voiced obstruent + nasal implies the existence of at least one sequence of unvoiced obstruent + nasal.

29. In initial systems the existence of at least one sequence of voiced obstruent + semivowel implies the existence of at least one sequence of unvoiced obstruent + semivowel.

The distribution of languages in accordance with these two hypotheses is described immediately below in Table 13 and 14.

	Unvoiced + Nasal Obst.	~Unvoiced + Nasal Obst.
Voiced + Nasal Obst.	24	0
~Voiced + Nasal Obst.	27	39

Table 13

	Unvoiced + Semivowel Obst.	~Unvoiced + Semivowel Obst.
Voiced + Semivowel Obst.	39	0
~Voiced + Semivowel Obst.	22	29

Table 14

For final systems a statement of unrestricted generality is possible.

30. In final systems the existence of at least one combination of sonant + voiced obstruent implies the existence of at least one combination of sonant + unvoiced obstruent. Details are given in Table 15.

	Sonant + Unvoiced Obst.	~Sonant + Unvoiced Obst.
Sonant + Voiced Obst.	24	0
~Sonant + Voiced Obst.	33	5

Table 5

Whereas, as has just been seen, voicelessness is dominant over voice in obstruents, in sonants the situation is reversed. Unvoiced nasals, liquids and semivowels, when they appear, are often allophonic variants which are then confined to voiceless environments. Whereas, as has been seen, among obstruent sequences those in which all members are unvoiced are the favorite type, sequences of unvoiced sonants do not occur at all in the sample.

31. In initial and final systems an unvoiced sonant is never immediately preceded or followed by another unvoiced sonant.

Glottalized consonants, which may be in principle classified as a third major group alongside of voiced and unvoiced consonants since they involve a distinctive type of laryngeal adjustment, are subordinate in relation to non-glottalized consonants. This is shown by the fact that in languages with the glottalized and non-glottalized consonants, if there are any clusters, there are always some consisting exclusively of non-glottalized consonants but not necessarily any containing one or more glottalized member. This may be considered the limiting case of the following more general rule.

32. If a language has at least one cluster containing n glottalized consonants, it has at least one cluster with $n-1$ glottalized consonants.

This rule has been verified up to the maximum of three glottalized consonants with occurrence only in 41 (Georgian). Languages with clusters containing two glottalized consonants and which, therefore, by 32, have clusters with single glottalized consonants as well as clusters without any glottalized members are 27 and 102. Language 2, 4, 23, 28, 53, 61, 73, 79, 87 and 92 have clusters with a single glottalized member as well as those consisting exclusively on non-glottalized member. Languages 6, 24, 32, 50, 69, 71, 72, 100 have no glottalized combinations, although they have glottalized phonemes.

By glottalized is meant here the common ejective type. The number of languages with implosives, normally voiced, was too small to draw any safe conclusions. However, it may be plausibly conjectured that sequences of implosives with glottalized ejectives do not occur, nor do sequences of implosives with voiced or unvoiced obstruents nor of different implosives with each other. All sequences noted in the present sample were of implosives followed by voiced liquid in initial systems.

Among basic sound types, a preference for obstruents over nasals is shown, at least in the environment of liquids by the following group of four related hypotheses.

33. In initial systems the existence of at least one cluster consisting of nasal + liquid implies the existence of at least one cluster consisting of obstruent + liquid.

Languages with both nasal + liquid and obstruent + liquid combinations are 14, 16, 19, 25, 26, 30, 37, 40, 41, 43, 51, 55, 56, 57, 59, 60, 72, 75, 77, 80, 81, 82 and 84. The remainder, except for these fifteen which do not have either type of combination (6, 21, 29, 31, 34, 47, 54, 58, 66, 71, 73, 79, 90, 99, 100) have obstruent + liquid while nasal + liquid clusters are lacking. These results are summarized in Table 16.

	Obstruent + Liquid	Obstruent + Liquid
Nasal + Liquid	23	0
Nasal + Liquid	52	15

Table 16

34. In final systems the existence of at least one liquid + nasal cluster implies the existence of at least one liquid + obstruent cluster.

Languages with both liquid + nasal and liquid + obstruent are: 3, 4, 7, 13, 15, 16, 18, 27, 30, 33, 35, 36, 38, 41, 42, 43, 44, 48, 49, 57, 64, 65, 67, 72, 75, 76, 77, 80, 81, 82, 86, 94, 98 and 102. Those with liquid + obstruent and without liquid + nasal are: 1, 2, 10, 28, 51, 62, 69, 87, 88, 96, 97, 101 and 103. The remainder have neither type of combinations. The results are summarized in Table 17.

	Liquid + Obstruent	~Liquid + Obstruent
Liquid + Nasal	34	0
~Liquid + Nasal	13	15

Table 17

36. In final systems the existence of at least one nasal + liquid cluster implies the existence of at least one obstruent + liquid cluster.

Languages with both nasal + liquid and obstruent + liquid clusters are: 2, 7, 28, 38, 41, 43, 49, 76, 80 and 98. Those with obstruent + liquid and without nasal + liquid are: 27, 33, 67, 75, 77, 81 and 103. The remainder do not have either type of cluster. The results are shown in Table 18.

	Obstruent + Liquid	~Obstruent + Liquid
Nasal + Liquid	10	0
~Nasal + Liquid	7	45

Table 18

The dominance of obstruents over nasals is further shown by the following generalization.

36. In final systems the existence of at least one cluster consisting of nasal + nasal implies the existence of at least one cluster consisting of nasal + obstruent.

Nasal geminates are included in this statement. The languages with both nasal + nasal and nasal + obstruent are: 4, 7, 27, 41, 43, 64, 67, 72, 76, 77, 80 and 86. The remainder except for languages 9, 38, 58, 65, 79, 83, 99 and 104 which have neither type, have nasal + obstruent but do not have nasal + nasal. The results may be seen from Table 19.

	Nasal + Obstruent	~Nasal + Obstruent
Nasal + Nasal	12	0
~Nasal + Nasal	42	8

Table 19

A preference for liquids over nasals, at least in one environment, is shown in the following generalization:

37. In initial systems the existence of at least one cluster consisting of obstruent + nasal implies the existence of at least one cluster consisting of obstruent + liquid.

An exception is 31 (SANTEE DAKOTA) which has obstruent + nasal without having obstruent + liquid. However, it has no liquid consonants. Languages with both obstruent + nasal and obstruent + liquid are: 1, 2, 8, 10, 12, 13, 14, 16, 17, 20, 23, 26, 27, 28, 30, 32, 33, 35, 36, 38, 39, 41, 42, 43, 44, 48, 49, 51, 52, 59, 60, 61, 62, 63, 67, 69, 72, 75, 77, 80, 81, 82, 84, 86, 87, 92, 93, 98, 102 and 103. Languages 6, 21, 29, 34, 47, 54, 58, 66, 71, 73, 79, 90, 99 and 100 have neither combinations. The remainder, except for SANTEE have obstruent + liquid while obstruent + nasal combinations are lacking. The results are stated in Table 20.

	Obstruent + Liquid	~Obstruent + Liquid
Obstruent + Nasal	50	1
~Obstruent + Nasal	25	14

Table 20

The final hypotheses of the absolute type have to do with points of articulation. There is some evidence for the dominance of the dental-alveolar region over the labial and the palatal velar. This is most surprisingly shown in what may be called the law of the final dental-alveolar.

38. Every language with final clusters contains at least one cluster with a final obstruent in the dental-alveolar region.

Systems with very few final clusters are, of course, the strongest test of this hypothesis. Examples from ancient languages not found in the sample include CLASSICAL GREEK with the three final clusters ps, ks and ŋks and LATIN which has final clusters all of which end in s or t. Within the sample are BALTI with only

ks, rs, ŋs and ks and MASAI with only rn, rt and rd. From hypotheses 27 and 30 it follows that at least one of the final dental alveolar obstruents is unvoiced.

A corresponding statement for initial systems cannot be confined to obstruents since there are systems with initial nasal + obstruent combination only.

39. Every language with initial clusters contains at least one cluster with an initial consonant in the dental-alveolar region.

A strong confirming example is CHIRICAHUA APACHE which in addition to initial alveolar affricates has only st and sd.

A further piece of evidence is contained in the following hypothesis.

40. A language which has any affricates includes among them at least one in the dental-alveolar or alveopalatal region.

That is, languages like GERMAN with affricate such as pf also have ts (or t̠s̠, or their voiced or glottalized counterparts). Of the languages in the sample 3, 23, 28, 32, 42, 43, 48 and 61 have affricates of the alveolar or alveopalatal type as well as affricates at other points of articulation. Languages 8, 14, 15, 17, 19, 25, 29, 38, 39, 49, 50, 55, 56, 65, 70, 74, 78, 89, 90, 91 and 95 have no affricates. The remainder have dental-alveolar or alveopalatal affricates but no others. These results are tabulated in Table 21.

	Other Affricates	~Other Affricates
Dental and similar Affricates	8	0
~Dental and similar Affricates	75	21

Table 21

We may summarize at this point the chief conclusions concerning initial and final consonant clusters valid for the sample of 104 languages utilized in terms of objective preferences as revealed in the foregoing generalizations.

1. Shorter clusters are preferred over longer ones.
2. Clusters which are analyzable into subclusters which likewise occur are preferred over those which are unanalyzable.

3. In terms of assimilation, homorganic nasals + obstruents are preferred over heterorganic nasals + obstruents and obstruent combinations which are homogeneous in voicing are favored over those which are heterogeneous.
4. In terms of dissimilation, sequences which differ only in glottal adjustment are disfavored as are sequences of different kinds of sibilants or of different kinds of liquids.
5. In relation to the peak of the syllable, combinations are favored in which sonants are closer to the peak than obstruents and in which voiced consonants are closer to the peak than unvoiced.
6. In absolute terms unvoiced obstruents are preferred over voiced, voiced sonants are preferred over unvoiced, nonglottalized consonants are preferred over glottalized, liquids are preferred over nasals, and the dental-alveolar point of articulation is preferred over other positions.

At various points of the exposition allusion was made to the importance of other related topics. These remarks are here summarized and amplified. In general their importance is that they provide independent evidence for the conclusions based on initial and final clusters alone. They also tend to raise further related questions. It is clear that a more complete and valid set of generalizations and their explanation can not be carried out except in such a broader context.

Among the related topics are the study of medial clusters, of phonemic systems, of morphophonemic alternations, of canonical forms of morphemes and of diachronic sound change. The relevance of the first of these, medial clusters, is so obvious that it need only be mentioned as another part of the same general topic of consonant combinations. An example of the connection between the study of phonemic systems and that of clusters is the fact that, parallel to the favoring of obstruents over nasals in combinations is the fact that there are no languages without obstruents while there is a small number of languages without nasals. The study of morphophonemic alternations will show that certain "avoided" combinations are in general the subject of morphophonemic rules which provide replacements. For example, in MAYAN languages, which have the third person prefix with base form s- there is usually a rule by which stems with initial basic -š have a replacement form for the sequence sš. These alternations are, of course, the result of diachronic changes which are documented or inferred.

It is evident that facts of alternation as reflecting historical changes add a certain additional type of information which must be

considered in any explanatory theory and may indeed aid in the solution. Thus, in the above example, a purely synchronic listing of the initial combinations shows only that sš is absent. It is a further fact whether, where sš is expected s appears, š, or as is at least logically conceivable, some third sound or sound combination. From the cure we may perhaps receive some enlightenment regarding the disease.

The study of canonical form of morphemes also provides independent evidence on the same questions. Thus again, in the matter of sibilants, CHIRICAHUA APACHE allows no roots which have both s and š. Such "distance phenomena" probably involve more complex psychological processes and may therefore differ somewhat from those involved in immediate sequences.

Thus, in SEMITIC, as a general rule several consonants with the same point of articulation are not found in the same root. Among the excluded combinations are the sequences of nasal and homorganic stop which are very common as direct sequences.

Finally, it may be pointed out that the generalizations presented here are only a portion of the conclusions that can be drawn either from this material or from other evidence. Thus, possible regularities concerning differential text frequency of different classes of combinations were not considered. Also systemic quantitative hypotheses were not considered as, for example, those regarding the relative size of class membership of phonemes immediately following or preceding particular phonemes or classes of phonemes. It is to be hoped that further investigations will test the hypotheses presented here as well as discover new ones.

LIST OF LANGUAGES

The letters after each entry indicate the following:

I Existence of Initial Clusters
F Existence of Final Clusters
C Information on clusters is seriously incomplete
P Phonetic information is incomplete or inadequate
T One of the sources contains a table or other form of statements concerning consonant clusters

1. AFRIKAANS (I, F, T)
   Meyer de Villiers, Afrikaanse Klankleer (Kapstaad, Amsterdam, 1958).

2. AGUACATEC (MAYAN) (I, T)
International Journal of American Linguistics [IJAL] 22. 72-6 (1956).

3. ALBANIAN (I, F, T)
Structural grammar of Albanian (Bloomington, 1957).

4. AMHARIC (F, P, C)
Armbruster, C.H. Initia Amharica, 2 vols. (Cambridge, 1908).

5. AMUESHA (ARAWAK) (I, F, T)
IJAL 19. 191-4 (1953); Miscellanea Phonetica 3. 15-21 (1958).

6. APACHE (I, T)
Linguistic Structures of Native America by Harry Hoijer and others [LSNA] (New York, 1946) 55-84.

7. ARABIC (EGYPTIAN) (F, T)
Harrell, R. The phonology of colloquial Egyptian Arabic (New York, 1957).

8. ARANTA (P, I, T)
Oceania 12. 255-302 (1942).

9. ARAPAHO (F, T)
IJAL 22. 49-56 (1956).

10. BALTI (TIBETAN) (I, F, P, C)
Read, A.F.C. Balti grammar (London, 1934).

11. BASQUE (I, F, P, C)
Lhande, P. Dictionnaire Basque-Française et Français-Basque (Paris, 1926-); Gavel, H. Eléments de phonétique basque (Bayonne, 1929).

12. BENGALI (I)
Language 36. 22-59 (1960); Wagner, Reinhard, Bengalische Texte in Urschrift und Umschrift (Berlin, 1930).

13. BERBER (SHILHA) (I, F, T)
Applegate, J. An outline of the structure of Shilha (New York, 1958).

14. BILAAN (I, F, T)
Philippine Journal of Science 84. 311-22 (1955).

15. BRETON (I, F, T)
Sommerfelt, A. Le breton parlé à Saint-Pol-Léon (Rennes, 1921).

16. BULGARIAN (I, F, C)
Minkov, M. Bălgaro-anglijski rečnik (Sofia, 1958); Stoikov, S. Uvod v bălgarskata fonetika (Sofia, 1955).

17. CAMBODIAN (I, T)
Bulletin of the School of Oriental and African Studies [BSOAS] 14. 149-74 (1952); Bulletin de la Société Linguistique de Paris [BSLP] 42. 112-31 (1942).

18. CATALAN (I, F, C)
Badia Margarit, A. Gramática histórica catalana (Barcelona, 1951); Falera, P. Diccionari de la llengua catalana (Barcelona, 1954).

19. CHAM (I, P)
Aymonier, E. and Cabaton A. Dictionnaire Čam-Français (Paris, 1906).

20. CHATINO (I, T)
IJAL 20. 23-7 (1954).

21. CHINANTEC (QUIOTEPEC) (I, T)
IJAL 27. 237-50 (1961).

22. CHINANTEC (USILA) (I, T)
IJAL 28. 251-5 (1962).

23. CHONTAL (HOKAN) (I, F, T)
IJAL 16. 35-39 (1950).

24. CHONTAL (MAYAN) (I, T)
IJAL 25. 44-53 (1959).

25. CHRAU (I, C)
BSLP 57. 171-91 (1962).

26. CHUKCHEE (I, P)
Handbook of American Indian Languages [HAIL] II: 639-903 (ed. by Franz Boas).

27. COEUR D'ALENE (I, F, T)
HAIL III: 517-707.

28. COOS (I, F, P, C, T)
HAIL II: 303-429.

29. CUICATEC (I, T)
IJAL 12. 139-146 (1946).

30. CZECH (I, F, C)
Kucera, H. The Phonology of Czech ('s Gravenhage, 1961); Trávniček, F. Mluvnice spisovné čestiny vol. 1 (Praha, 1948).

31. DAKOTA (SANTEE) (I, P, T)
HAIL I: 879-965.

32. DAKOTA (YANKTON) (I, T)
IJAL 21. 56-59.

33. DANISH (I, F, T)
Martinet, A. La phonologie du mot en danois (Paris, 1937).

34. DELAWARE (I, F, T)
LSNA 130-57.

35. DUTCH (I, F, T)
Cohen, A. et al., Fonologie van het Nederlands en het Fries ('s Gravenhage, 1961).

36. ENGLISH (I, F, T)
Bloomfield, L. Language (New York, 1933) 127-138.

37. EWE (I, P, C)
Westermann, D. A study of the Ewe language (London, New York, 1960).

38. FRENCH (I, F, T)
Nyrop, K. Manuel phonétique du français parlé (6th ed., 1951, Copenhagen); Lingua 5. 253-87 (1955-6).

39. FRISIAN (I, T)
Same source as 35.

40. GÃ (I, C, P)
Wilkie, B. Ga grammar, notes and exercises (Oxford, 1930).

41. GEORGIAN (I, F, P)
Norsk Tidsskrift for Sprogvidenskap 18. 5-90; BSOAS 14. 55-72.

42. GERMAN (I, F, T)
Moulton, W.G. The sounds of English and German (Chicago, 1962).

43. GILYAK (I, F, P, T)
Hattori, T. Versuch einer Phonologie des Südostgiljakischen Phonembestands, Journal of Hokkaido Gakugei University 13. 67-96 (1962).

44. GREEK (I, F, T)
Koutsoudas, A. Verb morphology of Modern Greek (Bloomington, 1962).

45. HINDI (I, T)
Language 34. 212-24 (1958).

46. HUAMBISA (F, T)
Lingua Posnaniensis 6. 1-8 (1957).

47. HUICHOL (I, T)
IJAL 11. 31-5 (1945).

48. HUNGARIAN (I, F, T)
Acta Linguistica Academiae Scientarum Hungaricae 3. 125-57; 411-29 (1953); 4. 193-224 (1954); Hall, R. Hungarian grammar (Baltimore, 1944).

49. ICELANDIC (I, F, T)
Kress, B. Die Laute des modernen Islandischen (Leipzig, 1937).

50. IOWAY-OTO (I, C, T)
IJAL 13. 233-250 (1947).

51. IRISH (I, F, T)
Ó Cuív, B. The Irish of West Muskerry, Co. Cork (Dublin, 1944).

52. ITALIAN (I, T)
Hoare, A. An Italian dictionary (2nd ed. Cambridge, 1925); Hall, R. Descriptive Italian grammar (Ithaca, 1948).

53. IXCATEC (I, T)
Miranda, M.T.F. de, Fonémica del Ixcateco (México, 1959).

54. JAPANESE (I, T)
Language 26. 86-125 (1950).

55. JAVANESE (I, T)
Uhlenbeck, De Structuur van het Javaanse Morpheem (Bandoeng, 1949).

56. KAREN (I, P, T)
University of California Publications in Linguistics, vol. 25 (1961).

57. KASHMIRI (I, F, P, C, T)
Acta Orientalia 19. 79-99 (1943).

58. KERESAN (I, F, T)
IJAL 12. 229-36 (1946).

59. KHASI (I, P, T)
Rabel, L. Khasi, a language of Assam (Baton Rouge, 1961).

60. KORYAK (I, P, T)
HAIL II: 639-903.

61. KUTENAI (I, F, T)
IJAL 14. 37-42 (1948).

62. LATVIAN (I, F, C)
Endzelin, J. and Hausenberg, Latviešu valodas vardnica (Chicago, 1956); Endzelin, J. Latviešu valodas gramatika (Riga, 1951).

63. LITHUANIAN (I, F, T)
Otrebski, J. Gramatyka języka litewskiego (Warszawa, 1958).

64. MARATHI (I, F, C)
Kelkar, A.R. The phonology and morphology of Marathi (Cornell dissertation, 1958).

65. MASAI (F, P, C)
Tucker, A.N. and Tompo Ole Mpaayei, J., A Maasai grammar (London, 1955).

66. MAZATEC (I, T)
IJAL 13. 78-91 (1947).

67. NORWEGIAN (I, F, T)
Norsk Tidsskrift for Sprogvidenskap 12. 5-29 (1942).

68. OSAGE (I, T)
IJAL 18. 63-8 (1952).

69. OSSETE (WESTERN) (I, F, C)
BSOAS 13. 36-79 (1951).

70. OSTYAK (F, T)
Steinitz, W. Ostjakische Grammatik und Chrestomathie (2nd ed., Leipzig, 1950).

71. OTOMI (MAZAHUA) (I, T)
IJAL 19. 253-8 (1953).

72. OTOMI (PAME) (I, F, T)
IJAL 22. 242-65 (1956).

73. OTOMI (TEMOAYAN) (I, T)
IJAL 15. 213-22 (1949).

74. PALAUNG (I, T)
BSOAS 23. 544-57 (1960).

75. PASHTO (I, F, T)
Penzl, H. A grammar of Pashto (Washington, D.C., 1955).

76. PERSIAN (F, T)
Nye, G.E. The phonemes of modern Persian (Michigan dissertation, 1955).

77. POLISH
Benni, T. Fonetyka Opisowa, in Benni et al., Gramatyka języka polskiego (Krakow, 1923); Bulas, K. and Whitfield, F.J. The Kosciuszko foundation dictionary, vol. 2 (Mouton, The Hague, 1961).

78. PORTUGUESE (I)
Boletim de Filologia 3.1-30 (Lisboa, 1949).

79. QUILEUTE (I, F, T)
HAIL III. 149-292.

80. RUMANIAN (I, F, T)
Graur, A. and Rosetti, A. Esquisse d'une phonologie du roumain, Bucarest, Universitatea, Facultatea de filosofie şi litere. Bull. Ling. 36. 5-29 (1938).

81. RUSSIAN (I, F, C)
Avanesov, R.I. Fonetika sovremennogo russkogo literaturnogo jazyka (Moscow, 1956).

82. SERBO-CROATIAN (I, F, T)
Language 22. 112-20 (1946).

83. SIERRA POPOLUCA (F, T)
IJAL 13. 13-17 (1947).

84. SINHALESE (I, C)
University of Ceylon Review 18. 163-75 (1960).

85. SPANISH (I, T)
Language 27. 248-53 (1951).

86. SWEDISH (I, F, T)
Studia Linguistica 9. 8-20 (1955).

87. TAKELMA (I, F, C, T)
HAIL II. 7-296.

88. TAOS (F, C)
LSNA 184-221.

89. TEMNE (F, T)
Wilson, W.A.A. An outline of the Temne language (London, 1961).

90. TERENA (I, T)
IJAL 12. 60-3 (1946).

91. THAI (I, T)
Abramson, A.S. The vowels and tones of Standard Thai (Bloomington, 1962).

92. TOJOLABAL (MAYAN) (I, T)
IJAL 12. 34-43 (1946).

93. TOTONAC (I, F. T)
IJAL 12. 34-43 (1946).

94. TURKISH (F)
Sevortijan, E.V. Fonetika tureckogo literaturnogo jazyka (Moscow, 1955); Turkce Sözlük (Ankara, 1955).

95. TWI
Schachter, P. Teaching English pronunciation to the Twi-speaking student (Legon, 1962).

96. UZBEK (F, T)
Bidwell, Charles, A structural analysis of Uzbek (New York, 1955).

97. WAIGALI (I, F, P, C, T)
Norsk Tidsskrift for Sprogvidenskap 17. 146-324 (1954).

98. WELSH (I, F)
Jones, S. A Welsh phonetic reader (London, 1926);
Jones, T.G. and Gwynn, A. Geiradur gymraeg-saesneg a saesneg-gymraeg (Cardiff, 1950).

99. WICHITA (I, F, T)
IJAL 16. 179-84 (1950).

100. WINNEBAGO (I, T)
Susman, A. The accentual system of Winnebago (New York, 1943).

101. YUMA (F, T)
LSNA 249-88.

102. YUROK (I, F, C)
University of California Publications in Linguistics 15 (1951).

103. ZAPOTEC (MITLA) (I, F, T)
Briggs, E. Mitla Zapotec grammar (Mexico, 1961).

104. ZOQUE (I, F, T)
IJAL 17. 105-23 (1951).