DOCUMENT RESUME

ED 102 825 FL 006 314

AUTHOR	Sherman, Don; Vihman, Marilyn
TITLE	The Language Universals Phonological Archiving Project: 1971 - 1972. Working Papers on Language Universals, No. 9.
INSTITUTION	Stanford Univ., Calif. Committee on Linguistics.
PUB DATE	Nov 72
NOTE	18p.
EDRS PRICE	MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS	*Computational Linguistics; Computer Programs; Information Processing; Information Retrieval; Information Storage; *Language Research; *Language Universals; Phonetics; *Phonology; *Program Descriptions
IDENTIFIERS	Eskimo; MARC; Modern Irish

ABSTRACT

This project was designed to develop machine-searchable files of linguistic data to be interrogated by researchers looking for patterns, examples, and other kinds of evidence bearing on language universals. This preliminary report of the project focuses on a detailed description of the computer program system used, MARC (Machine-Readable Catalog), and on the development of an archive record for storing and retrieving phonetic and phonological data. In the final section of the report some specific examples of the archive material on modern Irish and the North Greenlandic dialect of Eskimo are presented. (PMP)

Working Papers on Language Universals
No. 9, November 1972
pp. 163-179

SURVEY PAPERS

The Language Universals Phonological Archiving Project: 1971 - 1972

Don Sherman
Marilyn Vihman
Stanford University

1. Introduction

The basic goal of the project is to develop machine-searchable files of linguistic data to be interrogated by researchers looking for patterns, examples, and other kinds of evidence bearing on language universals. The project is therefore an investigation of two basic questions: first, what constitute adequate descriptive categories for linguistic phenomena; and second, what are the appropriate media and formats for storing, controlling, and accessing descriptive linguistic data?

The first question is particularly important in that it involves trying to construct the major dimensions in terms of which linguistic phenomena can be described, with the constraint that such descriptions be cross-linguistically comparable. The most serious problem is not to decide which theory or explanation of the facts is correct, but to try to record at least the observational data without undue bias toward possible explanations.

In other words, the project aims at recording "statements regarding individual languages which rest in some direct way on a body of observations."[1] Greenberg also specifies that observationally adequate descriptions should meet two criteria: "(1) particularity, i.e. absence of generalization; (2) the use of terms based directly on physically observable characteristics..." (loc. cit.)

Greenberg's criteria can be taken as the design guidelines for an archive useful for language universals research. Within these guidelines, the analysis of data into descriptive or explanatory patterns is the free

[1] Joseph Greenberg, On the 'language of observation' in linguistics, Working Papers on Language Universals 4, 1970, p.G3.

prerogative of the archive users, regardless of their particular theoretical position. An archive, after all, is much like a library: its job is to organize and store data in as simple and direct a manner as possible. The pleasures of browsing, searching, and discovering are properly reserved for the interests and inspirations of the patron.

The technical problems of how to organize and store data, however, are very much at the center of building an archive. We attempted to deal with these problems of media and formats in terms of three major desiderata: (1) diversity of languages and topics; (2) variety of source material; (3) generalized computer system. The first requirement was to find a means for systematically storing and retrieving, on a language-by-language basis, the data which is useful for universals research. Different topics will be included in the archive, and the data retrieval mechanisms must be capable of searching both language and topic categories.

The second technical goal of the archive is to be able to accept descriptive data from different kinds of source grammars. Typically, grammars show wide differences in style, theoretical outlook, completeness, reliability, etc. The archive must be able to accept and integrate data derived from many different kinds and levels of sources, while imposing some minimum standards of uniformity and interpretation on the data. The third desideratum is to use a computer system as the basic organizing and recording medium. The computer resources which will be of greatest utility are facilities for defining sophisticated record formats and data structures without requiring that archive users become active programming experts.

As the first steps toward putting our principles into practice we concentrated this past year on provisionally selecting a computer system, and on developing an archive record for storing and retrieving phonetic and phonological data. In our selection of a computer system we opted for a combination of a well-organized scheme of data element definitions plus a set of existing file creation and data retrieval programs. The system may require some modification and re-emphasis before it will give optimum performance on linguistic data; however, the record definition conventions show a surprising conformity to the feature and segment oriented descriptions which were appropriate for phonetic and phonological data.

The archive record for phonetics and phonology was designed to be able to represent the 'observational' data available to linguists in published descriptions of languages. The data extracted from each language's grammar is structured into two separate sections: an inventory of phonetic segments and a set of phonological rules. The definitions of the data elements of both parts of the record are designed so that the data can be examined from many different points of view and so that data retrieval can have many options for organizing or combining different access points.

In the remaining sections of this report, we will discuss features both of the computer system and of the archive record for phonetic and phonological data. In section V we also present some brief examples of encoded data.

2. The MARC computer system

In order to provide a starting context for the initial archiving work we chose to work with an existing program system which had been developed by libraries and library schools to store, process, and retrieve machine records of bibliographic data. This system, called MARC (Machine-Readable Catalog) and developed initially by the Library of Congress,[2] was selected as an experimental medium because it offered most of the features we were looking for in terms of computer storage and retrieval of language universals data.

[2]See U.S. Library of Congress, Information Systems Office, MARC Manuals Used by the Library of Congress. Chicago, 1969.

The heart of the MARC system is a generalized data structure which is acceptable to all the operating programs within the system. This data structure consists of various machine-programming conventions such as: character set, configuration of fixed length and variable length record components, and most importantly, conventions for naming data elements and sub-elements. The lowest level of data element is a <u>subfield</u>, which consists of an identifying code plus a text, i.e., a string of characters which may be codes or narrative prose or numeric material. The next level of MARC data elements is the <u>field</u>, which consists of one or more subfields. The field also has an identifying label (the field tag) consisting of five digits. The highest level of structure in the MARC system is the <u>record</u>, which consists of a defined set of fields and subfields relevant for some specific topic. MARC records may be defined to describe very specific items, such as lexical entries in a dictionary, or much more general topics, such as the language universals definitions which represent the phonetic inventory and phonological rules of an entire language or dialect.

The MARC record structure differs from other systems in that it uses explicit tags and codes to identify data fields and subfields. Subfield identifier codes are embedded in the running text, while field level tags are all consolidated in a single record directory. This consolidated directory not only lists all the field tags in a record, but also provides an index or pointer (similar to a page number in a table of contents) to the initial character location of the data string. A schematic version of this arrangement is as follows:

$$\underbrace{\text{Tag}_1,\ \text{Tag}_2,\ \text{Tag}_3 \ldots \text{Tag}_n}_{\text{DIRECTORY}} : \underbrace{\text{Field}_1,\ \text{Field}_2,\ \text{Field}_3 \ldots \text{Field}_n}_{\text{DATA}}$$

The MARC record structure, as pictured above, contains two basic areas: directory and data. The directory serves as a table of contents to the major fields in the record, and for each field identifies the descriptive category as well as the location and length of the data field. The data section of the record contains the texts of the fields plus a second level of subfield indentifiers which are embedded in the running text. The field tags in the directory can be scanned and searched rapidly, but require an overhead in extra storage space. Subfield identifiers conserve storage space but require a complete text search before a specific subfield can be retrieved.

Because the MARC data structure is so broad, it is possible to define many different kinds of record formats within this structure, each with its own scheme of data elements and data element names. The actual operating programs within the MARC system are all parameter controlled and can process any MARC-structure record, whether its content is the bibliographic description of a book or the phonological description of a language. The MARC system programs include a broad range of data processing functions, including record generation and correction, file sorting and display, and most importantly, record retrieval based on complex search requests which may include several search terms linked by the Boolean operators AND, OR, or NOT.[3]

3. Archive definition of phonetic segments

In the initial language universals archive record for phonetic and phonological data it seemed logical to let a single record represent (the dialect of) a language. The phonetic data fields of the record will represent the complete phonetic inventory of a single language, and a large file of such records will eventually comprise a crosslinguistic archive of different phonetic inventories which can be studied and analysed as a self-consistent body of data.

---

[3]See Aiyer, Arjun. The CIMARON System: Modular Programs for the Organization and Search of Large Files. Berkeley, California, 1971.

In order to organize the scheme of phonetic data elements to reflect a traditional taxonomy, we defined field tags which would associate phonetic segments according to their major class (consonant or vowel) and their place and/or manner of articulation. The choice of a predominantly articulatory framework for the field tag scheme for this portion of the record was motivated by a desire to use an easily determined set of "physically observable characteristics" (see section I). This scheme also seemed to us least likely to lead to ambiguous interpretation either by coders or by users. There is in principle no reason, however, why it could not be supplemented by an acoustic or perceptually based categorization.

The first three digits of the field tag are thus used as a primary matrix within which to locate various kinds of phonetic segments. The following scheme of values is used.

TABLE 1 -- FIELD TAG SCHEME FOR PHONE SEGMENTS

1st Digit:	Major Phone Segment Type	
	2--Consonant	3--Vowel
2nd Digit:	Place of Articulation or	Tongue Depth
	0--Unspecified	0--Unspecified
	1--Labial	1--Front
	3--Dental	3--Central
	5--Palatal	5--Back
	7--Velar	
	9--Glottal	
3rd Digit:	Manner of Articulation or	Tongue Height
	0--Unspecified	0--Unspecified
	1--Stop	1--Low
	2--Double Articulation	
	3--Fricative	3--Mid
	5--Nasal	5--High
	7--Lateral	
	9--Vibrant	9--Glide

For example, p (bilabial stop) will be tagged as '211', and i (high front vowel) will have the tag '315'. This scheme is of course only a partial classification, and assigns the same tag to both voiced and voiceless consonar p and b), rounded and unrounded vowels, etc. The full and unique specification of each of the segments will be represented at the subfield level; the tagging scheme is purely an aid to simple retrieval and emphasizes only some of the classifying features of each data field.

The fourth and fifth digit of the tag (known as Indicators) are used to carry an assigned phoneme control number, so that various kinds of sub-phonemic phenomena (e.g. allophonic, dialectal) can be linked together as variants of a single basic phonemic entity, while the basic unit stored remains the phone. Categorization of the individual phone-segments into "phonemes" allows us to preserve information available in some grammars and to provide a functional overview of the phonetic patterning in each language. At the same time, storage in terms of phones, rather than phonemes, and further analysis of phones into features, means that the phonological rules need make no reference to the categorization in terms of phonemes.

The major task of characterizing a phonetic segment as part of the phonetic inventory of a language cannot be accomplished by a brief tag code such as that outlined above. The bulk of the specification must be located at a level which allows more freedom both in representation and in the definition of descriptive categories. In MARC this is the subfield level, and in our record we found it very natural to equate subfields with phonetic features. We wished to avoid the controversy over binary versus multi-valued features; since our coding conventions could accomodate either or both, we have chosen to use each feature-type where it seems to fit the data most naturally (e.g., multi-valued for place of articulation, vowel height, etc., vs. binary for lip-rounding, duration, etc.). Binary features are expressed as a feature-value followed by a parenthetic suffix containing "plus" or "minus": e.g. for the feature duration, the binary values are long (+), long (-). For multi-valued features, each feature-value has a distinct expression: e.g. bilabial, dental, etc., as values of the feature place of articulation. No suffix is used in this case.

For the phone segment fields we currently have defined fourteen subfields which may be used to characterize and specify the features of the phonetic segment. Each subfield is introduced by a three-character sequence: "$" plus 'letter' plus 'blank or number'. The first component ('$') is simply a graphemic representation of the MARC-structure conventions for a subfield delimiter; it signifies that the next character is a subfield identifier code. The second component, the identifier code itself, consists of an upper or lower case letter which serves as the subfield label; we use lower-case codes to identify 'observationally derived' subfields, and the corresponding upper-case letter may be used freely for comments, explanations, disclaimers, source document quotations, notes, etc.

The third component of this introductory sequence is usually blank. A number value is used, however, whenever it is appropriate to analyse the features of a segment into sequential parts, e.g. partial voicing of obstruents, pre-nasalization, pre- or post-palatalization, etc. This technique is a special adaptation of the MARC structure to linguistic purposes, and seems to be a reasonable way of describing phonetic events (as they are currently understood) in terms of applying features in temporal sequence to portions of segmental entities.

The full set of subfields used for characterizing both consonantal and vocalic phonetic segments are defined and explained below.

FIELD 2XY CONSONANT PHONE SEGMENTS

(For values of X and Y, see Table 1)

INDICATORS: The two indicator positions are used to represent an arbitrarily assigned Phoneme Control Number. A tag phone and all its allophonic or other variants will all have the same Phoneme Control Number.

SUBFIELDS:

$a--IPA Symbol. This will be a computer-code version of the IPA symbol for this segment, including a full range of diacritics. The details of this scheme will be given in a later report.

$b--Segment Class. The values for this subfield for the 2XY fields are 'obstruent', 'sonorant', and 'syllabic'.

$c--Place of Articulation. The values for this subfield are conventional names for points of articulation, e.g. 'bilabial', 'labiodental', etc. See Ladefoged, Preliminaries to linguistic phonetics, 1971: 92.

$f--Aspiration. This is a binary-valued subfield whose representations, where specified, are 'aspir (+)' and 'aspir (-)'. Heavy vs light aspiration contrasts will be coded as comments in the $F subfield.

$g--Glottal Mode. This is a multi-valued subfield structured along the lines suggested by Ladefoged (1971: 21), with three values corresponding to three points along the continuum of glottal adduction: 'voiceless', 'voice', 'glottal stop'. Other values, such as 'creaky voice', 'breathy voice', etc. may be used wherever there is data to support these descriptions.

$h--Tenseness. This is a binary-valued subfield whose representations, where specified, are 'tense (+)' and 'tense (-)'.

$j--Prosodic Features. The values of this field have not yet been determined.

$l--Length. This is a binary-valued subfield whose representations, where specified, are 'long (+)' and 'long (+)' and 'long (-)'.

$m--Manner of Articulation. The conventional names for consonantal manners of articulation are used here, such as stop, fricative, lateral, vibrant, nasal, etc. Affricates are represented as a temporal sequence of '$m1 stop' and '$m2 fricative'. This is used as the model for other doubly-articulated consonant segments.

$n--Nasality. This is a binary valued subfield which describes velic opening or closure. The values are 'nasal (+)' and 'nasal (-)'.

$r--Lip-Rounding. This is a binary-valued subfield whose representations, where specified, are 'round (+)' and 'round (-)'.

$s--Segment Status. This field is used to express the phonemic status of segments within the language system being archived. The list below gives the values to be used.

'tag (-)' --no subphonemic variation.
'tag (+)' --this is the tag or basic phone of a phoneme which includes other allophones or free variants.
'allo' --secondary allophone of a phoneme.
'free' --phone segment in free variation, with no other status in the language system.
'loan' --phone segment which appears in loan words only.
'out' --phone segment which occurs only outside normal language system, e.g. in exclamations.
'unspec'--phone segment with major unspecified features.

$x--Source Reference. Citation of relevant pages/paragraphs from source grammar. (The full bibliographic citation for source reference will be found in a separate field (Field 010), whose definition is not given here.

$z--Phonological Rule Reference. This field will carry the numbers of the phonological rules in which this segment may be involved as input, output, or conditioning environment.

FIELD 3XY VOWEL PHONE SEGMENTS

(For values of X and Y, see Table 1)

Note: The Indicator and subfield definitions for the 3XY fields are almost entirely identical with the definitions of the 2XY fields. The exceptions are given below. Vowel feature subfields '$d' and '$e' can be used with consonant phones to express the parameter of tongue height, especially if it is distinctive, as in the cases of palatalization or velarization. These two phenomena would be expressed as '$d front $e high' and '$d back $e high' respectively.

SUBFIELDS (WHICH DIFFER FROM 2XY DEFINITIONS):

$b--Segment Class. The values for this subfield are 'vowel' and 'glide'.

$d--Tongue Height. For vowels, seven potentially contrasting levels of tongue height are defined as follows: high, lower-high, higher-mid, mid, lower-mid, higher-low, low.

$e--Vowel Depth. The values for this subfield are intended to cover the full spectrum of tongue positions used in vowel articulations. These values are: front, front-central, central, back-central, back.

To summarize thus far: each consonant phone segment is located in a place/manner matrix by the second and third digits of the field tag code; each vowel phone segment is located in terms of tongue height and depth. The status of the segment is given in the $s subfield. For example, 'free' identifies those segments which are in unconditioned, free variation with each other, e.g., NORTH GREENLANDIC m and ŋ. Where segment status is unspecified we will identify those segment fields where there are major gaps in our information about important features. An example would be nasalized vowels in NORTH GREENLANDIC where the conditions of nasalization are explicit, but there is no data about the qualities (height or depth) of the resulting segments. In this case we will have a single segment represent all nasalized vowels and code 'unknown' in the various feature subfields (height, depth) as well as coding 'unspecified' in the status subfield.

The subfields we have chosen to include in our current definition represent those features which we tentatively believe to be both codeable (in the sense of being available in the descriptive phonological literature for a large number of languages), and also of interest from an analytic language-universals point of view. This set of features is by no means closed or frozen, but can be changed or expanded by: (a) adding new features or conventions; or (b) adding new values to existing feature subfields. Neither of these forms of change or expansion is especially costly or complicated; the only caution is to maintain consistency with existing data, and to avoid extensive re-coding.

4. Phonological rules

The second major component of the language universals archive record under discussion deals with the formulation of phonological rules in a way compatible with computer storage retrieval and analysis. Within the phonological rule section of the archive record our goal has again been to allow maximum flexibility for access as well as consistency of presentation. There are considerable difficulties involved in encoding information regarding phonological or "morphophonemic" systems described by linguists belonging to quite different traditions. An archive containing phonological rules from a wide range of languages and organized according to a single system, however, would provide an important base for discussion of a great many topics currently of interest, ranging from obvious applications -- such as testing the notion "plausible rule" (cf. Chomsky and Halle, Sound Pattern of English, 1968, p. 428) -- to many other issues, such as the question of rule-ordering, "conspiracies" or the functional interconnection of phonological rules, etc. In formatting phonological rules into highly atomized field components we hope to make the data accessible from enough different points of view, for retrieval purposes, to provide information (once a large data base has been created) on many different questions or theoretical issues.

Briefly, the major requirement here is to define a format for computer-stored rule formulations which will facilitate retrieval, not rule-checking or segment generation. The definition is therefore based on trying to provide the means for answering questions such as:

--In what language does (segment X or feature Y) change, or result from change, or constitute the condition for change?

--In what languages does (segment X or feature Y) remain stable or unaffected by (synchronic, diachronic, allophonic, dialect) rules?

We understand the general linguistic form of a phonological rule to be: A → E/C, which we interpret as: there is an input (A) which becomes an output (B) conditional upon the presence of an environment (C). The environment portion may take the form: /L ___ R; that is, it may consist of an optional preceding left environment and/or an optional succeeding right environment. Thus, from our point of view there are four rule components: Input, Output, Left Environment and

Right Environment. It is this four-way division which we attempt to reflect in our tag-code scheme for phonological rules:

1st Digit: 5 (Field Block Reserved for Rules)

2nd Digit: Rule Component

1 -- Input

2 -- Output

3 -- Left Environment

4 -- Right Environment

3rd Digit: Type of Expression (Disjunction, etc.)

0 -- No Disjunctive Expression

1 -- Unrelated Disjunctive Expression

2 -- Related Disjunctive Expression

Indicators: Used for Rule Number

Under this scheme a single rule is broken up into several fields, with different fields representing different rule components. In addition, a separate field (tag 500) is reserved for describing the general characteristics of the rule as whole. Dividing the rule in this way allows for independent access to each of the rule components, and, equally important, permits the use of the same paradigm of subfields for all rule components. However, since a single rule is divided among several separate fields, it becomes necessary to devise a means for linking together all the parts; the fourth and fifth digits of the tag are used for this purpose and carry an arbitrary two-digit rule number.

The paradigm of subfields to be used in the phonological rule fields (except for field 500) follows exactly the structure proposed for the consonant and vowel phone segments. In this context, however, we will be as concise as possible in using subfields; that is, we will enter the subfield structure at the highest level of generality. For example, if the segments participating in the rule can be uniquely characterized by a single common feature value, e.g. stop, than a listing of all the stop phones need not be given. Similarly, if a feature plays no critical or "selecting" role in a rule, the entire subfield representing that feature (e.g. voice, place of articulation) can be omitted. If the rule expression is based on individual phones, however, there will of course be no choice but to detail all the relevant features -- type, place, manner, etc. Negative feature values can also be used to uniquely characterize the feature-set involved in a rule component. For binary features, the negative is expressed simply by reversing the sign of the suffix: The opposite of long (+) is long (-). For multivalued subfields a minus-sign is used as a prefix to negate the value, e.g. -stop. The meaning of this expression is: all values in the manner of articulation feature except stop. Sometimes it is necessary to construct a complement set by excluding more than one value; e.g. "-obstruent" would be expressed as -stop and -fricative. It should be noted that the $a (IPA symbol) subfield can be thought of as a list of all the segments of a language; thus '$a -p' is a legitimate expression and means every phonetic segment in the language except p.

5. Sample fields from Language Universals Phonetics and Phonology Archive

To conclude this preliminary report, we wish to present some specific examples of the archive material we have been describing. Normally it would be preferable to give an entire record as an example. However, it happens that the particular linguistic topic for which we have developed our first archive record format is very extensive, since it includes the complete phonetic inventory and (some of) the phonological rules for a language. The practical consequence of this is, first, that our archive will grow somewhat more slowly than we might wish; and, second, that the presentation of examples cannot include complete records but must be limited to selected fields.

Accordingly, in this section we present an example of two contrasting consonant fields, taken from our material on MODERN IRISH; and a single example of a phonological rule, taken from the NORTH GREENLANDIC dialect of ESKIMO. In all cases we also give a copy of the source grammar statements which served as a basis for the archive data. The sources are: The Irish of Erris, Co. Mayo (1968) by Eamonn Mhac An Fhailig; and A Phonetical Study of the Eskimo Language (1904) by William Thalbitzer.

LANGUAGE UNIVERSALS DATA ARCHIVE:

SAMPLE RECORD FIELD -- WESTERN IRISH CONSONANT SEGMENTS

"116. b' denotes a voiced bilabial plosive consonant. The lips are drawn inwards to the teeth. There is simultaneous raising of the of the front of the tongue towards the hard palate. b' occurs in all positions, initially, medially, and finally in words.

117. p' corresponds to b'. It differs in being voiceless, in having greater force of exhalation, and in being aspirated."

Eamonn Fhailigh, The Irish of Erris, Co. Mayo, 1968

DATA		MEANING	STRUCTURE
211		Bilabial Stop	Field Tag
01		Segment Number	Indicator
$a	b-palatalized	IPA Symbol	Subfield
$b	obstruent	Segment Number	Subfield
$c	bilabial	Place of Articulation	Subfield
$d	high	Tongue Height	Subfield
$D	"simultaneous rasing of tongue towards hard plate"	Comment on Tongue Height	Subfield
$e	front	Vowel Depth	Subfield
$g	voice	Glottal Mode	Subfield
$m	stop	Manner of Articulation	Subfield

DATA		MEANING	STRUCTURE
	$M less force of exhalation than p-palatalized	Comment on Manner	Subfield
	$n nasal (-)	Nasality	Subfield
	$r round (-)	Lip Rounding	Subfield
	$R "lips are drawn inward"	Comment on Tongue Height	Subfield
	$s tag (-)	Segment Status	Subfield
	$x Fhailigh, par 116	Source Reference	Subfield
	$F	End-of-Field	Terminator
211		Bilabial Stop	Field Tag
02		Segment Number	Indicators
	$a p-palatalized	IPA Symbol	Subfield
	$b obstruent	Segment Class	Subfield
	$c bilabial	Place of Articulation	Subfield
	$d high	Tongue Height	Subfield
	$D "simultaneous raising of front of tongue towards hard plate"	Comment on Tongue Height	Subfield
	$e front	Vowel Depth	Subfield
	$f aspir (+)	Comment on Lip Rounding	Subfield
	$F "greater force of exhalation than b-palatalized"	Comment on Aspiration	Subfield
	$g voiceless	Glottal Mode	Subfield
	$m stop	Manner of Articulation	Subfield
	$n nasal	Nasality	Subfield
	$r round (-)	Lip Rounding	Subfield
	$R "lips are drawn inward"	Comment on Lip Rounding	Subfield
	$s tag (-)	Segment Status	Subfield
	$x Fhailigh, par 117	Source Reference	Subfield
	F	End-of-Field	Terminator

LANGUAGE UNIVERSALS DATA ARCHIVE:
SAMPLE RECORD FIELD -- NORTH GREENLANDIC PHONOLOGICAL RULE NO. 14

"When i or u (high vowels) is followed by an aspirated fricative (ɬ, s, Φ) the whole surface of the tongue is raised tolerably high during the articulation of both the vowel and the consonant." p. 146

Thalbitzer, William. A Phonetical Study of the Eskimo Language. 1904

DATA	MEANING	STRUCTURE
510	Rule Input	Field Tag
14	Rule Number	Indicator
$b obstruent	Segment Class	Subfield
$c -uvular	Place of Articulation	Subfield
$g voiceless	Glottal Mode	Subfield
$m fricative	Manner of Articulation	Subfield
F	End-of-Field	Terminator
520	Rule Output	Field Tag
14	Rule Number	Indicator
$b obstruent	Segment Class	Subfield
$c -uvular	Place of Articulation	Subfield
$d high	Tongue Height	Subfield
$e front	Vowel Depth	Subfield
$g voiceless	Glottal Mode	Subfield
$m fricative	Manner of Articulation	Subfield
F	End-of-Field	Terminator
530	Rule Left Environment	Field Tag
14	Rule Number	Indicator
$b vowel	Segment Class	Subfield
$d -low	Tongue Height	Subfield
F	End-of-Field	Subfield

Note: The following is the analyst's representation of this rule:

1. Voiceless non-uvular fricative is articulated with a high tongue position after a non-low vowel (i and u plus allophones).
2. $$\underset{\begin{bmatrix}\text{-voice}\\ \text{+fric.}\\ \text{-uvular}\end{bmatrix}}{C} \longrightarrow \underset{\begin{bmatrix}\text{+high}\\ \text{+back}\end{bmatrix}}{C} \;/\; \underset{[\text{-low}]}{V} \;\underline{\quad\quad}$$