Journal of Universal Language 5
March 2004, 119-148

Word Final Coda Typology

Mark VanDam
Indiana University

Abstract

Language families were chosen based on the classification of Ruhlen (1987) in order to determine the number and kind of word final coda types languages tend to permit. It was found that languages tend to prefer fewer word final segments. Indeed, the extreme case is that which allows no word final consonants at all, a very common description of languages of the world. Next, of those languages that do allow word final coda consonants, the phonological kind of constituency was investigated. It was discovered that languages tend to simultaneously prefer a manner hierarchy (nasal > liquid > obstruent > glide) and a place hierarchy (alveolar > velar > retroflex, tap). The languages in the sample are shown to bear out the predictions made by the quantity and quality tendencies. Finally, the results predict the universality of word final coda constituents preferred by the languages of the world.

Keywords: phonology, place of articulation, manner of articulation, universals, coda segment, typology

1. Introduction

Languages differ in whether or not they permit word final consonants and, if so, the number and kind of consonants they permit. This paper investigates the range of possible and permissible word final coda types cross-linguistically, seeking to determine coda presence and constituency. It is demonstrated that languages tend to disprefer word final coda constituents, but of those languages that do permit word final codas, two proposed coda-constituent hierarchies are simultaneously respected, the manner of articulation hierarchy and the place of articulation hierarchy.

Word final codas, although a small part of the entire linguistic inventory of any language, are of interest because of their special status within the structure of language. As Kager (1999) points out, "*no languages* are known in which syllables *must have codas,*" although many languages do allow codas. Languages which disallow codas are of minimal interest to this study, but of the languages that do allow codas, there are several reasons to prefer word final codas (compared to medial codas). First, not all languages treat medial and final coda consonants equally. No language allows only word internal codas and disallows word final codas, but it is not always the case that coda sequences found medially are equally likely to be word final. For example, Kollimalai Tamil allows complex codas in the initial or medial syllable, but requires word final syllables to have a simple coda (Rajaram 1972). Second, syllabic structure of words is often difficult or impossible to determine word internally, especially in consideration of issues such as resyllabification, amibisyllabicity, or (de)linking and the various instantiations these might have in different theoretical frameworks such as Optimality Theory or Lexical Phonology. On the other hand, coda determination is virtually indisputable word finally. Third, although word final consonant clusters are not necessarily predictable from medial clusters, "strict restrictions on consonant clustering are often relaxed at the

margins of the word" (Kenstowicz 1994). If restrictions are relaxed, we expect to find the most permissive coda consonant clusters at the edges of the word. Since the larger typological question is *what is possible or potential human language* we can answer this question at the limits of possibility or potentiality with word final coda types. Fourth, internal morphological boundaries can affect possible production or perception of consonants or consonant clusters. For example, contrasting the English words *mistake* 'to misunderstand' and *miss-take* 'bad take' (on a film set, for example) shows that although linguistic-phonetic input is identical, morphological input can affect consonant clusters within words. Since word final consonant clusters are always at a morpheme boundary on their right edges, the variability of right-edge morpheme contact is eliminated in the present study. It should be noted that although right-edge morphological variation is controlled for, this does not imply that all morphological boundaries are controlled for. For example, the English word *mistake* above could appear in the plural form *mistakes*, thus adding morphology and phonology to the word final coda. Finally, many sources containing phonetic or phonological information of syllabic structure restrict themselves to word final coda sequences with little or no detailed attention to medial clusters. Because of this scant attention, it is more practical in a typological survey to focus on source material that is overtly concerned with word final consonants.

2. Language Sampling

Under the assumption that phonological variation is more readily influenced by contact (areal, genetic, or otherwise) than other parts of language, such as morphological type or the nature of function words, a diversified variety sample is best suited for this study. Using Ruhlen (1987) as a basis for classifying language families into a

large and diverse sample, the current study samples one language from each of Ruhlen's families, with a modification of two classifications Ruhlen suggests. The first adjustment to Ruhlen's classification is to conflate all isolates into a single category and select only one from the larger category rather than identify each individual isolate as a family by itself. Second, Ruhlen's classification of Pidgins and Creoles is not included in the current study for two reasons: redundancy and unpredictability. According to Rickford, "a pidgin usually involves mixture or compromise between native languages of its users ... [including] avoidance of consonant clusters and other marked phonological segments" and "creoles are simpler than older languages because of their pidgin ancestor" (Rickford 2003). Because of this mixed parentage, languages included in the Pidgins and Creoles family have partially genetically-related languages represented in other language families separately identified in Ruhlen's system. If some or all of the elements of a language are represented elsewhere, the inclusion of this family would be at least partially redundant. Next, Ruhlen's Pidgins and Creoles classification is excluded from the present work because the "phonemic system is loose, so that great variation in actual pronunciation—even to the point of several phonemically distinct shapes for many morphemes—can occur without necessarily impairing understanding" (Hockett 1958), as well as its speakers described as using "a very basic and crude way of speaking, stitched together from bits and pieces drawn from several different languages, with a tiny vocabulary, a variable phonology, and nothing much in the way of a grammar" (Trask 1996). This partial redundancy and unpredictability inherent in the nature of pidgins and creoles makes them undesirable for the current study.

With the preceding two adjustments made to Ruhlen's definition of language families, there are 18 families, hence 18 languages, included in the current study. A language from each family was selected on the basis of available descriptive phonetic or phonological

literature.[1] For each language included here, published phonological or phonetic studies were examined to determine coda presence and constituency. From these primary sources, the possible types of word final codas were recorded in terms of number and kind.

2.1. Phonological Data

The present study deals exclusively with word final codas and makes no claim about other phonological aspects of language. Thus, although some aspects of onsets are related to codas, the presence of "constraints holding among prevocalic C-sequences are not simply the mirror image of those which constrain postvocalic C-sequences" (Blevins 1995).

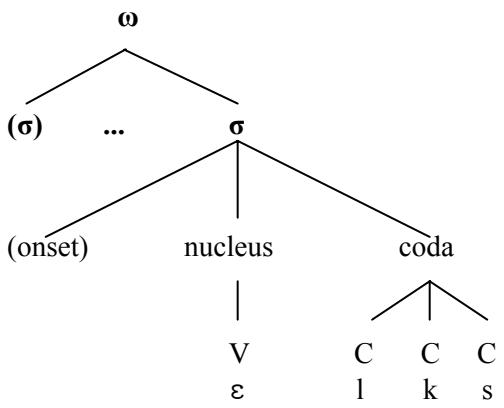
Although the available literature is not equally suited for gathering information, the following assumptions are made about the collected data. First, only vowels are nucleic constituents and only consonants are coda constituents. Second, because a coda is part of syllabic structure, all languages are assumed to have syllabic structure minimally consisting of a nucleus. In this respect, every coda will be preceded by its syllabic nucleus (vocalic peak) and followed by a right edge of a word. I assume a rather simple approach which does not provide for extra-syllabic or unsyllabified elements (Fudge 1969, Goldsmith 1990, VanDam 2003) or final, non-coda consonants as onsets of syllables without a pronounced nucleus (McCarthy & Prince 1990, Pigott 1999). Third, contact between words (a word final coda adjacent to the onset or nucleus of a following word) is not

[1] Language sources include: Tamazight-Berber (Jilali 1976), Biloxi (Einaudi 1976), Waljbiri (Capell 1962), Mokilese (Harrison 1976), Georgian (Job 1977), Chukchee (Kaempfe & Volodin 1995), Tamil (Karunakaran 1971, Rajaram 1972), Inupiaq (Kaplan 1981), Armenian (Job 1977), Kobon (Davies 1980), Nama-Hottentot (Hagman 1977), Navaho (Austin & McDonough 2000, Hoijer 1945), Xhosa (Beach 1938), Kanuri (Hutchison 1981), Chinese (Ning 1993), Hungarian (Vago 1980), Basque (Hualde 1991), Bella-Coola (Bird 2001), English (Kenstowicz 1994).

considered. Although in running speech this is a significant factor (affecting resyllabification, spreading of features, or other factors), the final codas considered in the present work are assumed to be in isolation or careful speech. As a consequence of these assumptions, environmental influence is controlled: the preceding environment will always be vocalic and the following environment will always be a word boundary.

To summarize the preceding discussion, the following diagram in (1) shows the environment of word final coda consonants. This top-down schema represents the phonological word **ω** at the top of the tree. Data not relevant to this study appear in parenthesis. The sample word in phonetic transcription at the bottom of the tree is the English word *elks*. In this case, the cluster /lks/ is the word final coda consisting of a tri-consonantal cluster.

(1)



3. Method: Number

Since no language prohibits open syllables, and no language requires closed syllables, languages can be classified into two general

types: those which allow codas and those which do not. The languages which do not allow codas are not considered in this study. Those languages that do allow (word final) codas will be classified according to the maximum number of coda consonants each language permits.

The maximum number of permissible coda consonants, henceforth MAX CODA, will be expressed with a single whole number representing only the maximum number of possible segments. Since no known human language requires a coda, the minimum permissible coda will always be zero.[2] It has also been observed that "if clusters of *n* Cs are possible syllable-finally, then clusters of *n*-1 Cs are also possible finally" (Blevins 1995). Whatever value of Max Coda a language permits, it also permits all lesser values, too.

Finally, the languages of the world can be categorized by their Max Coda values. From the current pilot study of 18 languages, no language was found to allow more than six consonants within a single coda, but there are no representative languages with maximally four or five word final consonant clusters in this study, although those languages presumably do exist. For example, some dialects of English allow a Max Coda value of four in words such as *sixths* /sIksəs/ and *strengths* /strɛŋkəs/, or even a Max Coda value of five in words such as *warmths* /wærmpəs/. It does seem likely, however, that there is some upper limit to Max Coda, but only attested maximums are examined in this study. If languages are found which allow more than six coda consonants (or languages change to allow more or fewer), revised Max Coda values can easily be incorporated into the data.

Table 1 is a sample table showing the Max Coda distribution in schematic form in the first column, the Max Coda numerical value

[2] Because it is certainly conceivable that a language could exist requiring a word final coda, a minimum should be theoretically allowed for, but, until at least one example is described, only the maximum is relevant to this discussion.

in the second column, and number and percent of languages in the current sample classified in each Max Coda value in the third and fourth columns.

Table 1. Max Coda values

schema	Max Coda	number of languages in sample	percent of languages in sample
V	0	*t*	*a*
VC	1	*u*	*b*
VCC	2	*v*	*c*
VCCC	3	*w*	*d*
VCCCC	4	*x*	*e*
VCCCCC	5	*y*	*f*
VCCCCCC	6	*z*	*g*

3.1. Method: Quality

Categorization of languages by Max Coda shows number constituency, but does not make reference to the quality of constituents within each particular Max Coda. Breaking down each Max Coda category into constituent segments can better give us an understanding of possible and potential human languages.

If a specific Max Coda value is extracted from Table 1 and separated into a language specific analysis, the constituents of word final coda sequences can be determined.[3] Languages which permit only a single word final coda (Max Coda-1 languages) are shown by per-

[3] Many authors of source material use non-conventional transcription systems, use the traditional system in a non-traditional way, do not describe their diacritics, use non-phonetic (English) orthography, or have no discernable system. Using the best (and often most current) sources I was able to establish a reasonably well supported/well documented transcription method within this paper based on the IPA. All transcriptions cited in the text are converted to IPA standard transcription if they are not already.

mitted segments on a language-specific basis. From these data, generalizations are observed and extended with respect to languages with greater Max Coda values.

Languages with Max Coda values greater than 1 are also examined based on language specific constituency. For example, Table 2 is a representation of two hypothetical Max Coda-2 languages divided into possible coda types for each permissible coda length. Each column represents data for individual languages. Rows indicate the name of the language, the consonantal phoneme inventory, and the actually-occurring word final codas when each number of consonants is permitted. Languages are arranged by Max Coda values so that all languages appearing in a table share the maximum number of coda consonants.

Table 2. Max Coda-2 languages (hypothetical)

Language (family)	LngA (FamY)	LngB (FamX)
C-inventory	n, ŋ, m, l, r, p, b, t, d, k, g, q, x, ʡ	n, ŋ, m, ɹ, l, p, b, t, d, g, ɤ, w, h, ʘ, ǀ, ǃ, ǂ, ǁ
C coda	n, ŋ, m, l, r, p, t, k, q, x, ʡ	n, ŋ, m, ɹ, l, d, g, ɤ, w
CC coda	nt, nk, ŋt	ɹm, ɹn, ɹl

There are two important observations about quality tables. First, notice that the languages in Table 2 permit some, but not all, logically possible complex sequences. For both languages there are logically possible combinations not appearing. In fact, all logically possible combinations *never* actually appear. For example, Biloxi, a Max Coda-2 language, permits the phonemes /x/ and /k/ word finally, either as the first or second member of a cluster or as singletons, but does not allow word final geminate consonants, so the sequences /xx/ and /kk/ are not permitted. Since Biloxi is not an isolated case, each exceptional datum is explained with the best available resources.

Second, there are more segments allowed to occur as single-member codas than segments allowed to occur in multiple-member codas (single-member codas are shown in the second row of data in Table 2). That is, the quality of single-member codas includes more phonemes than in complex clusters. This observation is interesting because as the number of logically possible segmental constituents in a given language increases, the number of actually-occurring combinations never increases proportionally. Greenberg (1978) observes that, in a language with 22 consonantal phonemes in the inventory, the number of single-member codas is equal to the number of phonemes, 22, and, "the logically possible sequences of length 2 are 22^2 = 484 ... For length 3 the logically possible number of combinations is 22^3 = 10,684" and so on. Using this method, the logically possible number of combinations for Georgian coda sequences, to be discussed below in greater detail, is 113,379,904.[4] It is remarkable that from the immense number of possible combinations, there is only one actually-occurring six-consonant coda sequence in Georgian.

This trend is largely respected by languages in the current sample: as the complexity of codas and number of logically possible coda sequences increase, the actual occurrence of sequences decreases.[5] This trend is shown in Tables 3 and 4 below. Table 3 shows the number of actually-occurring sequences[6] preceding the

[4] By coincidence Georgian also has a 22-phoneme consonantal inventory facilitating comparison to Greenberg's example. Greenberg's formula to calculate logical possibilities is essentially to raise the number of phonemes in the inventory to the power of Max Coda for that language. Thus, the calculation for Georgian is 22^6 = 113,379,904.

[5] Although the number of different coda sequences decreases, this does not imply that the number of occurring words is directly linked with this value. Since the word final coda distinguishes only part of the phonological input, other parts of a word could potentially provide sufficient linguistic input to differentiate words with identical complex coda sequences.

[6] In some cases, the number of actual sequences was not explicit in the literature. In these cases, I have indicated at least the attested minimums that are cited in the relevant sources accompanied by a (+). The (+) is intended to mean that more ex-

number of logically possible sequences. Table 4 shows the percent of actually occurring sequences given the logically possible combinations (the calculations from Table 3 are carried out). Logically possible sequences vary depending on phonemic inventory of respective languages: more phonemes allow more logical possibilities.[7] Languages without values in columns do not allow that number of word final coda consonants.

Table 3. Actual sequences and logically possible sequences

language	C	CC	CCC	CCCC	CCCCC	CCCCCC
Georgian	21 / 22	189 / 484	154 / 10,684	56 / 324,256	18 / 5,153,632	1 / 113,379,904
Hungarian	19 / 20	40+ / 400	1 / 8,000			
Tamazight-Berber	21 / 21	27+ / 441	2 / 9,261			
Biloxi	3+ / 15	2 / 225				
Navaho	9+ / 26	1 / 676				

Table 4. Percent of actual sequences (# actual sequences / # possible sequences)

language	C	CC	CCC	CCCC	CCCCC	CCCCCC
Georgian	96	39	1.4	>0.1	>0.01	>0.001
Hungarian	95	7+	>1			
Tamazight-Berber	100	6+	>1			
Biloxi	20+	>1				
Navaho	35+	>1				

amples of phonemes are assumed to be in the language. However, even with the values presented, the argument is unaffected. Numbers without (+)s are specifically referred to in the relevant literature.

[7] Number of phonemes in each language's inventory is the second number in the first column. As stated above, the number of logical possibilities is calculated as the number of phonemes in the inventory raised to the power of the allowed number of phonemes in the sequence (for that column): $(\#\text{ phonemes})^{(\#\text{ phonemes allowed})}$. Specific sources of this data for languages are as follows: Georgian (Job 1977), Hungarian (Vago 1980), Tamazight-Berber (Jilali 1976), Biloxi (Einaudi 1976), Navaho (Hoijer 1945, Austin & McDonough 2000).

4. Data from the Languages of the World

As stated above, representative languages from families adapted from Ruhlen (1987) were selected. One language from each family was selected on the basis of available published literature describing the language. The following sections present the results from analysis of the data collected on each representative language from each family beginning with the discussion of quantity and its typological implications. Results and implications from the quality analysis of Max Coda-1 languages are shown in Section 4.2. The languages permitting complex word final codas are discussed in light of these implications in Section 4.3.

4.1. Quantity Typology

Table 5 shows quantity sampling (see Section 4.2 for Max Coda assignments of individual languages). Since this study is interested in the maximum number of allowable word final coda consonants (not in particular words, but each language as a representative whole), only the maximum number is considered when counting quantity.

Table 5. Max Coda values

schema	Max Coda	number of languages in sample	percent of languages in sample (n=18)
VC	1	13	72
VCC	2	2	11
VCCC	3	2	11
VCCCCCC	6	1	6

Although the wide range of permissibility within codas attests to the variety of human language, there is clearly a preference to allow

fewer coda consonants than more. The number of languages representative of each Max Coda never increases as the Max Coda value increases. This strongly suggests the propensity to allow as few coda consonants as the language will bear. From the perspective of a quantitative analysis, these distinctions clearly favor fewer word final consonant coda constituents.

4.2. Quality Typology of Max Coda-1 Languages

As shown in Section 4.1, Max Coda values classify languages into groups that allow different maximal quantities in word final consonant clusters. Although Table 5 shows how many languages allow a specific Max Coda value, it does not show what kind of consonants can or cannot occur qualitatively or in what sequence for any given language. The following section is dedicated to languages that allow only a single consonant as a word final coda. Languages allowing more complex word final codas are discussed in Section 4.3.

Max Coda values were determined for each language. Table 6 shows the thirteen Max Coda-1 languages (the first row of language data in Table 5). Each individual language is listed in the left column with its family membership following in parentheses. The four columns on the right (separated by dashed lines) are the only permissible word final coda consonant by manner (N(asal), L(iquid), G(lide), O(bstruent)). Since all languages in Table 6 allow maximally one word-final coda consonant, each segment in the right columns can appear only as a single consonant in a $(C_0)VC$ structure.

The segmental word final coda inventories of the thirteen languages in Table 6 reveal several interesting observations. First, all languages tend to permit the alveolar nasal /n/.[8] Particularly interest-

[8] The only exception, Xhosa, allows only bilabial nasal /m/ in word final codas. But the presence of /m/ in Xhosa is limited to third-person objective concord, and

ing is the case of Waljbiri, which allows only /n/ in word final coda position. This suggests that if a language permits a single word final coda, it will be /n/. If word final coda segments are added incrementally (synchronically or diachronically) the first segment allowed into word final coda positions will likely be /n/.

Table 6. Max Coda-1 languages (where word final coda =C_1)

Language (family)	C_1			
	N	L	G	O
Korean (Altaic)	m, n, ŋ	l		p, t, k,
Waljbiri (Australian)	n			
Mokilese (Austric)	m, n, ŋ, m^w			
Chukchee (Chukchi-Kamchatkan)	n	l	w, j	t, k
Tamil (Elamo-Dravidian)	m, n, ɳ	l, ɭ,ɾ, r,	j	
Inupiaq (Eskimo-Aleut)	m, n, ɲ, ŋ			t, c, k, q
Armenian (Indo-European)	m, n	ʁ, ɹ, r	w	p, t, k, p', t', k', b, d, g, f, v, s, z, ʃ, ʒ, χ, h
Kobon (Indo-Pacific)	m, n, ɲ, ŋ	ɹ, l, ɭ, ʎ		p, b, d, c, g, f, s, x,
Nama-Hottentot (Khosian)	m, n			p, t, s
Xhosa (Niger-Kordofanian)	m			
Kanuri (Nilo-Saharan)	m, n	l, ɹ		
Chinese (Sino-Tibetan)	n, ŋ	ɹ		
Basque (*Isolate*)	n, ɲ	l, r, ɹ, ʎ		t, k, s, z, c, x, ʃ, s̻, ʃ̻

because of this very limited semantic/syntactic role, Xhosa is considered somewhat exceptional. However, even though Xhosa is somewhat exceptional, its lone permissible coda element is not entirely arbitrary because it is nasal.

Further analysis of Table 4 reveals other interesting patterns beyond the initial presence of /n/. If /n/ is permitted, then the language will next permit /m/, /ɲ/, or /ŋ/. From the data in Table 6, it appears that languages do not permit /m/, /ɲ/, or /ŋ/ before first permitting /n/. In the same way, /ɳ/ does not become a possible coda until after /m/, /ɲ/, or /ŋ/ are permitted. The hierarchy in (2) captures these observations.

(2) n > m, ɲ, ŋ > ɳ

The nasal hierarchy in (2) implies that if a language allows codas, it will allow segments to the left before it allows segments to the right; or, if a segment appears in a coda, at least some representative segment from each category to the left will also be represented. This appears to be a tendency of language, or perhaps the languages in the sample, but cannot be considered a strong universal property until many more languages are investigated. If the phonemic inventory of a language does not include a particular segment in the hierarchy, that segment cannot precede a segment lower on the scale in order for the lower ranked segment to be allowed. Also, a language might skip over a segment, but languages tend not to skip a category (contained by >) in favor of a lower ranked category. For example, no language simultaneously allows /n/ and /ɳ/ but disallows /m/, /ɲ/, and /ŋ/.

Those languages that allow nasals and permit other segments also respect a further segmental hierarchy. The next manner allowed after nasal is liquid. As shown in Table 6, eight of the thirteen languages allow nasals and liquids, with two permitting only nasals and liquids and no other manners. When liquids (and nasals) are permitted, the first liquids will be either /l/ or /ɹ/, as seen in Korean, Chukchee, Kanuri, and Chinese. After either /l/ or /ɹ/ is permitted, then /r/, /ɻ/, or /ʁ/ is permitted. Only after these classes are satisfied /ɭ/, or /ɾ/ will follow. The hierarchy in (3) shows which liquid segments appear before others (which are preceded by the nasal hierarchy):

(3) nasal >> l, ɹ > r, ʎ, ʁ > ɻ, ɾ

Languages that permit nasals and liquids tend to include obstruents next. This hierarchy is shown in (4):

(4) nasal >> liquid >> t > k, p > s, z, c, q, ʃ > b, d, g, x, h

Since only three of the thirteen sampled languages permit glides, and none permit only nasals and glides, glides are ranked below obstruents in the hierarchy shown in (5). The ranking of glides, however, is somewhat arbitrary. First, glides are under-represented in the sample data. Second, as Kenstowicz (1994) notes, the "*semivowels* or *glides* [y] and [w] are close kin to the corresponding high vowels [i] and [u]" and are thus interpreted differently by different authors. Because of these observations about glides, absolute placement should be reserved for future research, but for present purposes glides are situated below obstruents.

(5) nasal >> liquid >> obstruent >> w, j

From a broader perspective—one of natural classes rather than individual segments—the range of permissible manners other than nasal is varied: not all logically possible combinations of manners are represented. Assuming nasals occur if codas occur, the three relevant manners other than nasal—**L**iquid, **G**lide, and **O**bstruent—occur in exactly eight different logically possible combinations: N, NL, NG, NO, NLG, NLO, NGO, NLGO. From the previous examination of hierarchies, the allowable manners represented in Table 7 below show what kind of consonants can appear in the coda of any given language in that category.[9]

[9] Of course, the order of the natural classes is irrelevant since each language therein can only have a simple coda.

Table 7 shows the manners in the right columns with the indication 'yes' if the language(s) in that row permit that class of consonants in word final coda position. Since the generalization in (5) predicts that Max Coda-1 languages will at least have a nasal, the nasal class is presumed always to be occupied. If no representative languages appear in this study, question marks appear.

Table 7. Max Coda-1 constituency by natural class of manner

attested languages	C_1 summary	C_1 nasal	liquid	glide	obstruent
Waljbiri, Xhosa, Mokilese	N	yes			
Kanuri, Chinese	NL	yes	yes		
???	NG	yes?		???	
Nama-Hottentot, Inupiaq	NO	yes			yes
Tamil	NLG	yes	yes	yes	
Korean, Basque, Kobon	NLO	yes	yes		yes
???	NGO	yes?		???	???
Armenian, Chukchee	NLGO	yes	yes	yes	yes

The only types of Max Coda-1 languages unattested in this study (Tables 6 and 7) are NG and NGO languages. However, it is not the absence of these language types that is striking, but the presence of examples in every other category. With only thirteen Max Coda-1 languages, six of eight logically possible types are attested. Furthermore, in languages attested, there is a pattern languages tend to observe when allowing coda consonants. Languages tend to allow word final coda consonants according to THE MANNER HIERARCHY in (6).

(6) nasal > (liquid) > (obstruent) > (glide)

This hierarchy implies two things. First, if a language has coda consonants, it will have nasal coda consonants. Second, if a language allows coda consonants other than nasals, it will also permit

at least some segment from the natural class(es) higher (to the left) on the hierarchy.

Reconsider the observation that if a language permits word final codas then it will permit the alveolar nasal /n/. The previous discussion described this tendency in terms of manners of articulation, but it can also be extended to describe the tendency of languages to allow specific featural specifications in word final coda position in terms of place of articulation. The first (nasal) feature languages allow is the alveolar /n/; the first group of liquid features allowed includes the coronal (alveolar) /l/; the first obstruent allowed is the alveolar /t/. The second group of segments allowed includes the segments /ŋ/, /ʁ/, and /k/—all velar. The third group allowed includes the retroflexed and tapped segments /ɳ/, /ɭ/, /ɾ/. From these observations, THE PLACE HIERARCHY in (7) is proposed.

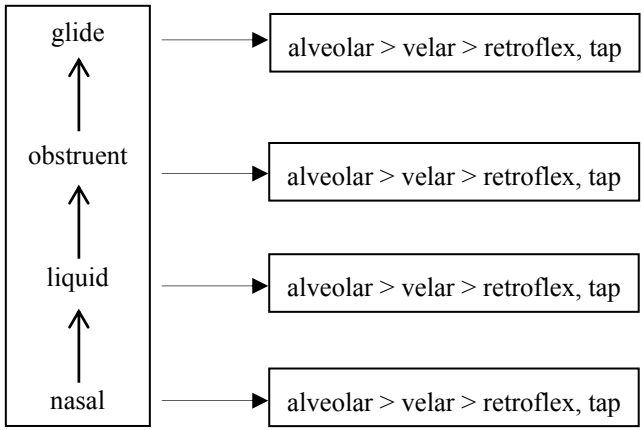
(7) alveolar > velar > retroflex, tap

This hierarchy suggests that the tendency is to permit alveolar coda consonants before velar, and velars before retroflex or tap segments. If a language has a word final coda segment in a category on the right, it will also have at least one representative from each category to its left.

The hierarchies in (6) and (7) are complimentary to one another such that each category in the manner hierarchy in (6) will also respect the place hierarchy in (7). If, for example, a language has only one segment in each category of the manner hierarchy, each segment is predicted to be alveolar; if a language has a velar segment in each category of the manner hierarchy, it will also have an alveolar segment in each category of the manner hierarchy. The display in (8) represents the manner and place hierarchies integrated into a single scheme. The manner hierarchy is shown vertically: segments tend to occur additively in a bottom-to-top fashion. Complimenting the manner hierarchy, the place hierarchy is represented horizontally:

segments tend to occur additively in a left-to-right fashion.

(8) Incorporated manner and place hierarchies



Finally, Table 8 shows the cumulative application of the manner and place hierarchies. Similar to the display above in (8), Table 8 below shows the manner hierarchy predicts a bottom-to-top occurrence of segments and the place hierarchy predicts a left-to-right occurrence of segments. Table 8 includes representative segments within each manner and place category. Some segments do not fit neatly into one specific category and are listed in parenthesis. Importantly, Table 8 also shows the cumulative effects of the incorporated hierarchies: all segments up to and including the x-axis value (place) and the y-axis value (manner) of the attested segment are expected to also appear.

The shaded region in Table 8 shows the application of the word final coda segments of Chinese within the cumulative hierarchies table (the word final phonemes of Chinese are in bold). As predicted, the permitted word final consonants of Chinese also include the more highly ranked categories with each hierarchy: Chinese permits

the velar /ŋ/ as well as the higher ranked alveolar /n/ within the place hierarchy; similarly, Chinese permits the word final liquid /ɹ/ as well as the higher ranked nasal /n/ within the manner hierarchy.

Table 8. Cumulative natural class and place hierarchies

manner hierarchy	alveolar	velar	retroflex, tap
glide	j	w, j	
obstruent	t, (s, z)	k (p)	ʃ, (c, q)
liquid	l, ɹ	ʎ, ʁ (r)	ɭ, ɾ
nasal	**n**	**ŋ** (m, ɲ)	ɳ
no coda			

place hierarchy

4.3. Quality Typology of Max Coda 2 through 6 Languages

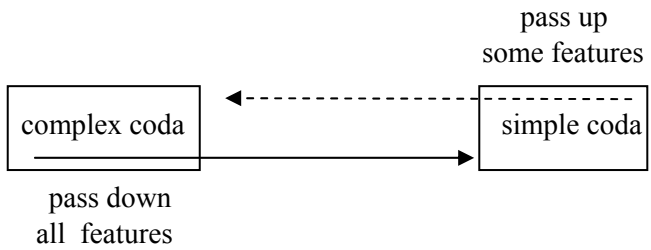
In the following section languages with Max Coda values greater than 1 are described in terms of quality. The quality of word final coda constituents permitted by these languages, regardless of how many consonants are permitted, tend to behave similarly to each other: if segments of quality *abc* are possible word final codas, then segments of quality *a, b, c, ab, ac,* and *bc* also tend to be possible. For example, Biloxi allows the complex codas /xk/ and /kx/ as well as the simple codas /x/ and /k/. When the maximally permitted coda allows combinations within the cumulative hierarchy, all of those combinations will also tend to be allowed, or inherited, by codas with fewer constituents as well.

Although coda sequence combinations tend to be passed down in full as possible sequences, it is not a bi-directional exchange with simpler codas "passing up" features in the hierarchy. As discussed

in Section 3.1, those languages that allow complex coda sequences always allow fewer combinations of sequences as the complexity of the coda (in number or quality) increases, even though the potential number of logically possible sequences increases exponentially (see Tables 3 and 4).

Not only are the quantities of sequences passed down in full, but so also are the qualities. For each quality in a larger word final coda sequence, that quality is also observed in sequences with fewer constituents. Coda complexity is passed down in full to simpler codas, but never passed up in full to more complex codas. (Of course, some segments must be passed up in order for larger codas to exist, but all segments are never passed up.) For example, Navaho allows only /ʔʃ/ as a complex coda, and among the permitted single-member codas are /n/, /l/, /s/, /d/, /ʃ/, and /ʔ/. Not only are both segments from the complex coda also allowed as simple codas, but the single-member coda segments exactly follow the cumulative hierarchy in Table 8 and respect the observation that all qualities are passed down while only some are passed up. This relationship is shown in (9) below. In a given language, all features of complex codas are shared with simpler codas, but simpler codas only share some features with more complex codas.

(9)



Max Coda-2 languages are shown in Table 9. The two languages shown in this table represent a small number of possible Max Coda types compared to Max Coda-1 languages. In Biloxi, only two com-

plex codas are permitted, /xk/ and /kx/; in Navaho, only one complex coda is permitted, /ʔʃ/.

Table 9. Max Coda-2 language coda constituency

Language (family)	Biloxi (Amerind)	Navaho (Na-Dene)
C inventory	n, m, p, b, f, t, d, c, k, s, ʃ, x, j, h, w	n, n', m, m', l, ɬ, b, t, t', k', d, k, k^w, g, s, z, ʃ, , x, x^w, ɣ, $ɣ^w$ ʔ, j, h, h^w
C coda	x, k, s	n, l, d, s, z, ɵ, ʔ, ʃ, h
CC coda	xk, kx[10]	ʔʃ

The languages in Table 9 confirm both the cumulative hierarchy in Table 8 as well as the generalization in (9) that coda features are fully passed down, but only partially passed up. Both languages show a wider range of quality and quantity variation as the complexity of the coda decreases.

Of special interest to the proposed hierarchies is Biloxi. Biloxi seemingly does not allow word final nasal or liquid consonants as single-member codas, yet allows the obstruents /x/, /k/, and /s/. On the surface, this appears to be contradictory to the manner hierarchy which predicts obstruents will occur only after liquid and nasal. However, the inspection of the phoneme inventory reveals the complete absence of liquids,[11] which, if not available, obviously cannot be employed. The availability of nasals is also called into question, because, according to Einaudi (1976), "it is often difficult to tell whether we are dealing with /ã/ or /an/". In either case, whether a nasal vowel or a nasal consonant is present in word final position, it seems to suggest that Biloxi does not disprefer at least some word

[10] Biloxi does not permit word final geminates so the sequences /xx/ and /kk/ are not allowed.

[11] Einaudi notes that /l/ and /r/ do occur, but "*l* occurs only in two modern names [and] *r* occurs in one proper name" (1976). This strongly suggests that liquids are functionally not part of the Biloxi phonemic inventory.

final nasality. The words in (11) show examples of word final nasal vowels.

(11) Biloxi

hã	'and'
dã	'he holds'
natõ	'brain'
õ	'make'
hĩ	'he arrives'
ekedĩ	'that is why'

(Einaudi 1976)

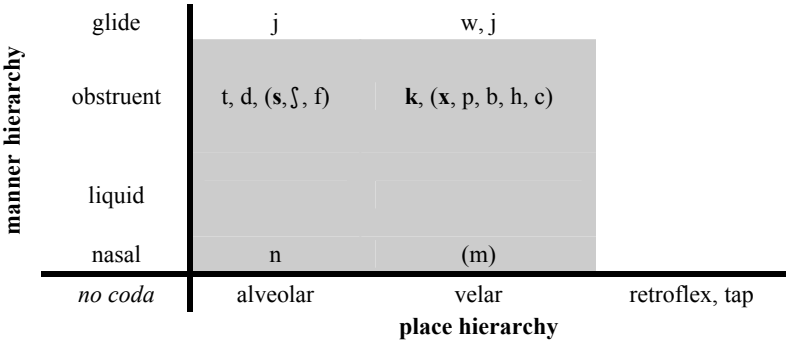
The nasal vowels /ã/, /õ/, and /ĩ/ are phonemes contrasting with the non-nasal phonemes /a/, /o/, and /i/. Although this is not a case of allophonic variation *preferring* word final nasality, the status of nasalized phonemes at least allows for the presence of nasality in word final position.

In light of these observations about Biloxi liquids and nasals, the cumulative hierarchy seems very well satisfied. The cumulative hierarchy table is reproduced in Table 10 with the phonemic inventory of Biloxi entered into the relevant places. The shaded region shows where at least some segments are permitted as word final codas, and bold segments indicate actually-occurring coda segments.

As stated above, Biloxi does not allow nasal consonants as word final codas. However, as the sample data in (11) show, Biloxi does allow the *manner* nasal to occur word finally (realized on vocalic segments), thus allowing the manner hierarchy to be filled, even though there is no consonantal *place* associated with that manner. Since liquids are not available, that manner is irrelevant to Biloxi. The next manner category available, obstruent, is occupied by the three occurring segments /s/, /k/, and /x/. These segments also satisfy the place hierarchy as well as the observation that only some features

are passed up: /s/ occupies the alveolar place, but is not allowed to be passed up in order to become a member of complex coda.

Table 10. Biloxi cumulative natural class and place hierarchies



Max Coda-3 languages are shown in Table 11. Again, variability among word final clusters is limited as the Max Coda value increases. Hungarian permits only one sequence of a tri-consonantal word final coda. Tamazight-Berber, although more possible combinations are permitted, also strictly limits what segments may appear and in what order. Hungarian permits only the sequence /psʃ/ in tri-consonantal word final codas (Vago 1980); if a Tamazight-Berber word ends in a tri-consonantal cluster, it will be either /ʃʃɵ/ or /qqɵ/ (Jilali 1976).

Table 11 confirms the cumulative hierarchy in Table 8 and the observation in (9): both languages show wider varieties of word final codas as the complexity decreases while respecting the manner and place hierarchies.

Table 12 shows the Max Coda-6 language, Georgian. Although a Max Coda value this large is presumably rare among the languages of the world, it nonetheless is represented by the word /bandʒgvl/ 'tuft of hair' (Job 1977).[12] The consonantal inventory of Georgian

[12] Bird (2001) also reports six-consonant clusters in Bella Coola. However, analyses

as well as word final clusters containing fewer consonants are also shown in Table 12.

Table 11. Max Coda-3 languages

Language (family)	Tamazight-Berber (Afro-Asiatic)	Hungarian (Uralic-Yukaghir)
C inventory	n, m, b, d, g, f, ɵ, s, z, ʃ, ʒ, g, ç, ʝ, x, q, h, ɹ, l, w, g^w	p, b, m, f, v, t, d, ʃ, ʒ, n, l, r, s, z, k, g, c, ɟ, ɲ, h
C coda	b, f, m, ɵ, d, t, k, s, z, ʃ, ʒ, n, l, ɹ, ç, ʝ, g^w, q, x, w, h	p, b, m, f, v, t, d, ʃ, ʒ, n, l, r, s, z, k, g, c, ɟ, ɲ
CC coda	mm, nn, bb, ff, tt, dd, ss, zz, ʃʃ, ʒʒ, ç ç, ʝʝ, kk, g^wg^w, qq, mɵ, ɹɵ, ɹb, ɹf, ɹm, ɹd, ɹs, ɹz, ɹʃ, ɹʒ, ɹ ç, ɹʝ, ts, tʃ, t ç, ts, dʒ, dʝ, dz, st, zt, sɵ, ʒɵ, fɵ, ɹɵ, xɵ, mn.....(etc.)	ss, zz, pp, bb, mm, tt, dd, ʒʒ, nn, zz, kk, gg, rr, pt, kt, st, ʃt, nk, ng, mp, mb, ld, lt, rt, lv....(etc.)
CCC coda	ʃʃɵ, qqɵ	psʃ

Table 12. Max Coda-6 language

C inventory (=22)	m, n, l, r, p, b, p', v, t, d, t', s, z, ʃ, ʒ, k, g, k', χ, ʁ, q', h
C coda (=21)	m, n, l, r, p, b, p', v, t, d, t', s, z, ʃ, ʒ, k, g, k', χ, ʁ, q'
CC coda (=189)	bm, ʒm, sp', nn, tv, bl, t'l, zr, ʁr, pt, gs, t's, nt', rʃ, nk, mk',

of languages that include word final clusters this large encounter problems. The Bella Coola cluster reported by Bird, [c'ktsk^{w}c'] 'he arrived', cannot be considered in the current study according to the criteria established in section 2.1 because there is not a preceding vocalic nucleus defining the coda environment. The maximal example noted by Job, /bandʒgvl/ 'tuft of hair', is subject to the criticism that the second and third members form the affricate /dʒ/, considered to be a single segment under some analyses and two adjacent segments under others. Although he attempts to justify this and the other affricates, /ts/, /t's/, /tʃ/, /t'ʃ/, /dz/, and /dʒ/, as two adjacent segments, the distinction is at least controversial.

	ʃχ, t'q', gm, lp, ʃp, p'n, lv, gl, pl, tr, p'r, kt, vs, bz, p't', pʃ, rk, nk', bʁ, ʃq', zm, mp, bn, rn, ʒv, dl, kl, k'r, ʃr, ʃt, ls, vz, rt', kʃ, pk, p'k', zʁ, tm, rp, gn, sn, rv, vl, ʁl, mr, χr, χt, ms, lz, st', q'ʃ, ʃk, rk', lʁ, k'm, lb, dn, pn, sv, zl, q'l, nr, bt, zd, ns, nz, pt', χʃ, dg, sk', nʁ, lm, mb, vn, kn, t'v, tl, ʃl, p'r, vt, k'd, p's, rz, kt', tʃ, zg, ʃk', ʒʁ, rm, rb, zn, ʃn, kv, k'l, χl, ʒr, lt, ld, rs, dz, ʃt', t'ʃ, lg, tχ, rʒ, sm, ʁb, tn, χn, ʁv, ml, br, sr, mt, nd, ps, vt', ʁt, rʒ, ng, lχ, lq', t'm, lp', k'n, gv, q'v, p'l, gr, t'r, nt, rd, ks, k't', vʃ, dʒ, rg, rχ, nq', km, mp', ln, dv, ʃv, rl, dr, pr, rt, ʁd, χs, lt', mʃ, tk, bk', sχ, p'q', ʁm, rp', mn, zv, ʁv, sl, vr, kr, st, bs, ts, mt', nʃ, lk, lk', pʁ, rq'
CCC coda (=154)	rsm, rk'n, t'ʃn, rk'v, rχv, tʁv, ngl, rk'l, rpl, t'sl, q'dr, st'r, mʁr, rds, lt's, nkt', nt'ʃ, ʁʃm, t'k'n, tʁn, sk'v, t'q'v, lʁv, rgl, mp'l, nkl, t'ʃl, ʁvr, kt'r, k'vt, kvs, nt's, rkt', rt'ʃ, tsm, kmn, rʁn, lt'v, vʃv, rʁv, rdl, rp'l, rkl, rʁl, mtr, ʃkr, kvt, vrs, rt's, ltʃ, ʁt'ʃ, t'ʃb, brn, dʒn, rt'v, tʃv, sʁv, k'vl, nʒl, pkl, sʁl, ʃk'r, p'q'r, mrt, lts, tʁz, mtʃ, bdʒ, rdn, rt'n, ngv, st'v, tsv, pʁv, nvl, msl, nχl, dʒl, zmr, tsr, t'ʃd, mts, ndz, ntʃ, ldʒ, tvn, st'n, rgv, tkv, dzv, ʃʁv, rtl, rsl, rχl, mbr, mp'r, t'sr, sʁd, nts, rdz, rtʃ, mdʒ, kvn, tsn, rtv, lkv, t'sv, mbl, ʁtl, nt'l, t'q'l, vdr, lt'r, t'ʃr, dʒd, rts, rst', ʃtʃ, rdʒ, ʁvn, t'sn, nk'v, rlv, t'ʃv, rbl, nk'l, rt'l, dzl, ndr, nt'r, tʁr, ngs, kts, kst', ʁtʃ, ntk, rtk, dʒk', t'sq', tʃk, ntʁ, t'ʃq', dzg, rtʁ, dʒg, tʃʁ, psk', tsʁ, nt'k', ndχ, t'sk', dʒχ, t'ʃk, rt'q'
CCCC coda (=56)	rt'sb, rtsn, rt'q'v, tʃʁv, rdχl, nt'sl, ntck, ntsv, dzgn, ndʒn, ntʃv, tsʁv, nʒχl, rt'sl, rdzg, rtsv, dʒgn, dzgv, rtʃv, rdʒv, rt'q'l, nt'ʃl, rdʒg, ndzχ, sk'vn, dʒgv, ntsv, ndgl, ntʃl, rtʁl, rt'sk', rdzχ, t'sk'n, t'sk'v, rtsv, rʁvl, rtʃl, tsʁl, rt'sk', rdʒχ, rgmn, tʃkv, ndzv, ntkl, rtsl, ndʒl, ntʃʁ, nt'ʃq', tʃkn, tskv, mt'sv, rtkl, rdzl, zχvr, rtʃʁ, rt'ʃq'
CCCCC coda (=18)	dzχvn, rtsʁv, rt'sq'l, rtʃkn, ndzgl, nt'ʃq'l, grdzn, ndʒgl, ntʃʁl, rtsʁn, nt'sk'l, rtʃʁl, rt'sq'v, rt'sk'l, rtsʁl, ntsʁv, rdʒχl, t'svrt
CCCCCC coda (=1)	ndʒgvl

The five languages represented in Tables 9, 11, and 12 confirm the hierarchies in Table 8 as well as the generalization that coda features are passed down in full, but only partially passed up to larger coda sequences.

5. Summary and Conclusion

The current study demonstrates several interesting tendencies and predictions about human language structure. Two broad issues were presented concerning word final codas: quantity and quality.

The number of word final coda segments varies significantly across the languages of the world. Although representative examples of 18 different languages reveal four different maximal word final coda structures, fewer coda consonants are clearly favored over more.

The quality of segments each language permits within its coda also reveals the structure of human language. The most common type of coda-permitting languages are those which permit only simple codas. Those languages permitting only simple codas clearly reveal the tendency for languages to permit segments within two newly proposed hierarchies—the manner hierarchy and the place hierarchy. The combination of these hierarchies, along with the observation that codas tend to fully pass down features, yields empirically accurate predictions for languages which permit complex codas.

Although these observations are not without exception, there is clearly the tendency within natural human language to respect these constraints on permissible structures through the proposed hierarchies. Overwhelmingly, examples are shown in this study that support the analysis.

Languages yet to be incorporated are expected follow the observations made here as well as other observations of language structure. Future research will be directed in several directions, but in all

cases should include many more languages and appropriate statistical techniques suited to handle observations from large numbers of languages. First, those languages which do not allow coda consonants must be considered in some systematic way. Of course, these languages will not have input for quality analyses. The observation that languages tend to allow fewer coda consonants predicts a large number of these languages. Second, the presence of glides, underrepresented in the manner hierarchy, should be examined in greater detail. Third, diachronic aspects of language change and variation will be incorporated in order to test the reliability of proposed hierarchies. Diachronic analyses are expected to confirm the presence and quality of coda constituents within the proposed hierarchies. Fourth, the present study can be incorporated into and tested in studies of artificial languages such as Unish and Esperanto. Relevant artificial languages could be tested to determine their fitness within the proposed hierarchies, or, if possible, if a diachronic analysis supported the claims and proposed hierarchies. In addition to specific areas of interest with respect to the current findings, there are other linguistic properties of interested to the current classification. For example, such linguistics items as prosody, salience, syntactic constituency, perceptual quality (such as Steriade's P-Map), and others might prove complimentary to the current study.

References

Austin, M. & J. McDonough. 2000. Navajo Wordlist for a Phonemic Inventory. Available at URL <http://ling.rochester.edu/ftp/pub/mcdonough/westernnavlist.pdf>.

Beach, D. 1938. *The Phonetics of the Hottentot Language*. Cambridge: W. Heffner.

Bird, S. 2001. Language Diversity. Available at URL <http://www.ldc.upenn.edu/sb/courses/cogsci/diversity.htm>.

Blevins, J. 1995. The Syllable in Phonological Theory. In J. Goldsmith (ed.), *The Handbook of Phonology* 206-244. Cambridge, MA: Blackwell.

Capell, A. 1962. *Some Linguistic Types in Australia*. Sydney: The University of Sydney.

Davies, H. 1980. *Kobon Phonology*: *Pacific Linguistics, Series B-No. 68*. Canberra: The Australian National University.

Einaudi, P. 1976. *A Grammar of Biloxi*. New York & London: Garland Publishing, Inc.

Fudge, E. 1969. Syllables. *Journal of Linguistics* 5, 253-286.

Goldsmith, J. 1990. *Autosegmental and Lexical Phonology*. Oxford: Blackwell.

Greenberg, J. 1978. Some Generalizations Concerning Initial and Final Consonant Clusters. In C. Ferguson, J. Greenberg, & E. Moravcsik (eds.) *Universals of Human Language 2: Phonology* 243-280. Stanford, CA: Stanford University Press.

Hagman, R. 1977. *Nama Hottentot Grammar*. Bloomington, IN: Indiana University Press.

Harrison, S. 1976. *Mokilese Reference Grammar*. Honolulu, HI: The University Press of Hawaii.

Hockett, C. 1958. *A Course in Modern Linguistics*. New York: MacMillan.

Hoijer, H. 1945. *Navaho Phonology*. Albuquerque, NM: The University of New Mexico Press.

Hualde, J. 1991. *Basque Phonology*. London & New York: Routledge.

Hutchison, J. 1981. *A Reference Grammar of the Kanuri Language*. Madison, WI: University of Wisconsin-Madison African Studies Program.

Jilali, S. 1976. *A Phonological Study of Tamazight Berber: Dialect of the Ayt Ndhir*. Unpublished PhD Dissertation, Los Angeles, CA: University of California.

Job, D. 1977. *Probleme eines Typologischen Vergleichs Iberokaukasischer und Indogermanishcer Phonemsysteme in Kaukasus*. Frankfurt am Main: Peter Lang.

Kaempfe, H-R. & A. Volodin. 1995. *Abriss der Tschuktschischen Grammatik*. Wiesbaden: Harrassowitz.

Kager, R. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kaplan, L. 1981. *Phonological Issues in North Alaskan Inupiaq 282*. Fairbanks, AK: University of Alaska.

Karunakaran, K. 1971. *The Kollimalai Tamil Dialect*. Annamalainagar: Azhahu Printers, Chidambaram.

Kenstowicz, M. 1994. *Phonology in Generative Grammar*. Cambridge, MA: Blackwell.

McCarthy, J. & A. Prince 1990. Prosodic Morphology and Templatic Morphology. In M. Eid & J. McCarthy (eds.), *Perspectives on Arabic Linguistics II: Papers from the second annual symposium on Arabic linguistics*. *Current Issues in Linguistic Theory* 72, 1-54. Amsterdam & Philadelphia: John Benjamins.

Ning, C. 1993. *Communicating in Chinese*. New Haven, CT: Far Eastern Publications.

Pigott, G. 1999. At the Right Edge of Words. *The Linguistic Review* 16, 143-196.

Rajaram, S. 1972. *Tamil Phonetic Reader*. Mysore-6: Central Institute of Indian Languages.

Rickford, J. 2003. Pidgins and Creoles. In W. Frawley (ed.), *International Encyclopedia of Linguistics, Second Edition* 3: 340-344. Oxford: Oxford University Press.

Ruhlen, M. 1987. *A Guide to the World's Languages 1: Classification*. London: Edward Arnold.

Trask, R. 1996. *Historical Linguistics*. London: Arnold.

Vago, R. 1980. *The Sound Pattern of Hungarian*. Washington, D.C.: Georgetown University Press.

VanDam, M. 2003. On the Phonological Structure of /i/-suffixed English Nicknames. *Indiana University Linguistics Club Working Papers Online* 3.1, 1-30.