

MANNER OF ARTICULATION PATTERNS IN WORD-INITIAL BICONSONANTAL SEQUENCES: THE EFFECT OF MORPHOLOGICAL CONTEXT

Shelece Easterday

Laboratoire Dynamique du Langage UMR5596 (CNRS & Université Lyon 2)
shelece.easterday@cnrs.fr

ABSTRACT

This study investigates the relationship between the phonotactic patterns of word-initial biconsonantal sequences and morphological complexity. In a sample of 32 unrelated languages, the common claim that the phonotactic patterns of heteromorphemic and tautomorphemic consonant clusters are structurally different from one another is tested. Specifically, the manner of articulation patterns of heteromorphemic and tautomorphemic word-initial biconsonantal sequences are analyzed and compared.

While some crosslinguistically frequent shapes are more likely to occur in heteromorphemic contexts, heteromorphemic patterns are in fact more restricted than tautomorphemic patterns in a number of other ways. They are less numerous and diverse, both within and across languages. Additionally, typologically rare and purportedly “dispreferred” word-initial consonant sequences are likely to occur in solely tautomorphemic contexts in the sample. Four general language types are identified according to interactions between the patterns of word-initial CC inventories and morphological complexity.

Keywords: consonant phonotactics, phonological complexity, phonological typology, morphology

1. INTRODUCTION

It has long been known that morphology can play an important role in phonotactic complexity: for example, in English and many other languages, the largest tautosyllabic consonant clusters arise through inflection and other morphological processes. This observation motivates a common assumption that morphologically derived phonotactic patterns are structurally different from those which occur within morphemes. Following from this, many typological studies of the sequencing properties of consonant clusters have excluded morphologically complex patterns altogether [12], [16], [19].

A recent line of research beginning with [8] investigates the patterning of tautomorphemic (morpheme-internal) and heteromorphemic (morphologically complex) clusters within

languages. Fine-grained analyses of complex cluster inventories from this perspective have revealed that tautomorphemic and heteromorphemic clusters tend to correspond to universally “preferred” and “dispreferred” shapes, respectively [9], [20]. However, because the languages examined thus far are limited to a small collection of well-studied and (Indo-)European languages, it is unclear whether these results are representative of global patterns.

2. RESEARCH QUESTIONS

The present study moves toward establishing a crosslinguistic baseline for the interaction between consonant phonotactics and morphological complexity. In particular, it seeks to establish whether (i) there are robust crosslinguistic differences between the most frequent tautomorphemic and heteromorphemic sequences, and (ii) whether there are any sequences, across languages, which occur predominantly in one morphological context or the other. Additionally, (iii) language-internal inventories are examined to determine whether those with solely tautomorphemic, solely heteromorphemic, or both morphological patterns tend to differ from one another structurally, with an aim towards (iv) identifying some general language types on the basis of the patterns observed.

These issues are examined in a genealogically and geographically diversified language sample. The current study is limited to word-initial biconsonantal (CC) sequences, as such structures are crosslinguistically the most frequent type of tautosyllabic cluster [17]. Further, the analysis is limited to an examination of the manner of articulation patterns within these sequences. While the analysis presented here is coarser-grained than those of [9], *inter alia*, such abstractions are necessary in order to establish patterns on a broad crosslinguistic scale.

3. METHOD

3.1. Language sample

The language sample consists of 32 unrelated languages with word-initial CC sequences (Table 1).

Table 1: Language sample. T = tautomorphemic, H = heteromorphemic

[ISO 639-3] Language (<i>Family</i>)		Morph. patterns		Type	[ISO 639-3] Language (<i>Family</i>)		Morph. patterns		Type
		T	H				T	H	
[mpi]	Makary Kotoko (<i>Afro-Asiatic</i>)	✓		II	[niv]	Nivkh (<i>isolate</i>)	✓	✓	IV
[pbv]	Pnar (<i>Austroasiatic</i>)	✓		III	[pbb]	Nasa Yuwe (<i>isolate</i>)	✓	✓	IV
[gni]	Gooniyandi (<i>Bunaban</i>)	✓		II	[trn]	Mojeño Trinitario (<i>Arawakan</i>)	✓	✓	IV
[xub]	Betta Kurumba (<i>Dravidian</i>)	✓		II	[bcr]	Babine-Witsuwit'en (<i>Athabaskan-Eyak-Tlingit</i>)	✓	✓	IV
[ium]	Iu Mien (<i>Hmong-Mien</i>)	✓		II	[erg]	Sie (<i>Austronesian</i>)	✓	✓	IV
[dnn]	Dzùngoo (<i>Mande</i>)	✓		II	[bzd]	Bribri (<i>Chibchan</i>)	✓	✓	IV
[opm]	Oksapmin (<i>Nuclear Trans New Guinea</i>)	✓		III	[bgn]	Western Balochi (<i>Indo-European</i>)	✓	✓	IV
[apn]	Apinayé (<i>Nuclear-Macro-Je</i>)	✓		II	[lzz]	Laz (<i>Kartvelian</i>)	✓	✓	IV
[yak]	Yakama Sahaptin (<i>Sahaptian</i>)	✓		III	[yee]	Yimas (<i>Lower Sepik-Ramu</i>)	✓	✓	IV
[dzg]	Dazaga (<i>Saharan</i>)	✓		II	[mxp]	Ayutla Mixe (<i>Mixe-Zoque</i>)	✓	✓	IV
[jya]	Northern Rgyalrong (<i>Sino-Tibetan</i>)	✓		III	[cta]	Tataltepec Chatino (<i>Otomanguean</i>)	✓	✓	IV
[asb]	Assiniboine (<i>Siouan</i>)	✓		III	[tlp]	Filomeno Mata Totonac (<i>Totonacan</i>)	✓	✓	IV
[wut]	Wutung (<i>Sko</i>)	✓		II	[bjt]	Balanta-Ganja (<i>Atlantic-Congo</i>)		✓	I
[aou]	A'ou (<i>Tai-Kadai</i>)	✓		II	[tzh]	Tzeltal (<i>Mayan</i>)		✓	I
[cod]	Cocama (<i>Tupian</i>)	✓		II	[ood]	Tohono O'odham (<i>Uto-Aztecan</i>)		✓	I
[cap]	Chipaya (<i>Uru-Chipaya</i>)	✓		II	[ket]	Ket (<i>Yeniseian</i>)	✓		I

Languages were selected on the basis of availability of reference material with comprehensive descriptions of phonotactic patterns and their morphological properties, and/or access to language specialists for consultation regarding specific patterns. The language sample is stratified so that half of the languages have word-initial CC patterns in solely tautomorphemic contexts, while the other half have at least some of these patterns in heteromorphemic contexts. Languages with only heteromorphemic word-initial CC patterns are relatively rare in the sample (4 languages), reflecting their generally low crosslinguistic frequency; it is more common for both morphological patterns to be present in CC inventories (12 languages).

The classifications listed under the heading “Type” in Table 1 result from *post hoc* analyses and will be defined in Section 5.

3.2 Data collection and coding

A comprehensive inventory of native word-initial CC sequences was gathered from the reference materials for each language. Each sequence was coded for the manner of articulation of its component consonants according to these categories: **N** = nasal stop; **P** = stop; **A** = affricate; **F** = fricative; **G** = central approximant; **D** = tap/flap; **R** = trill; **L** = lateral approximant. Lateral fricatives were classified as F and lateral affricates were classified as A. Sequences involving glottal stops or fricatives were excluded from the analysis.

The coding reflects the phonetic realization of the CC sequence: the shape of the initial sequence in Balanta-Ganja /b-ɲéfé/ [mɲéfé] ‘CL-belly’ [7] is

coded as NN. Thus the results reported here concern both phonotactic structure and phonetic substance.

Each shape was coded for the morphological contexts it is attested to occur in. For example, the PP shape in Western Balochi [2] is always heteromorphemic, occurring when the Subjunctive prefix attaches to a P-initial stem: /b-guʃ/ [bguʃ] ‘SUBJ-tell.PRES’. For a shape to be counted as occurring in both morphological contexts, it was not necessary that the sequence match exactly: for example, FG in Filomeno Mata Totonac [18] is coded as occurring in both morphological contexts on the basis of patterns /swaká/ [swaká] ‘s/he grinds (coffee)’ and /ʃ-júx-ma/ [ʃjúxma] ‘PAST-fall-PROG’.

The heteromorphemic patterns observed in the language sample include single-consonant prefixes and proclitics in combination with initial stem consonants, or occasionally with another prefix.

4. RESULTS

4.1 Frequent shapes and morphological context

There are 14 word-initial CC shapes which occur in one-third or more of the languages of the sample. These are listed in Table 2 in order of the frequency in which they occur in each morphological context. Frequency is calculated as the number of languages with the shape in the given morphological context, divided by the potential number of languages which could have that pattern: that is, those with the relevant consonant types in their phonetic/phoneme inventories *and* the given morphological context in their word-initial CC inventories.

Table 2. Distributions of most frequent shapes.

Tautomorphic			Heteromorphic		
	freq	(%)		freq	(%)
PL	17/22	(77%)	PF	12/15	(80%)
FL	14/20	(70%)	PG	11/16	(69%)
PG	19/28	(68%)	PP	11/16	(69%)
FG	16/26	(62%)	NF	10/15	(67%)
FP	16/26	(62%)	FA	7/11	(64%)
FN	13/26	(50%)	NP	10/16	(63%)
AG	11/23	(48%)	FP	9/15	(60%)
PF	11/26	(42%)	PL	7/12	(58%)
NG	11/28	(39%)	FG	8/15	(53%)
NF	9/26	(35%)	FN	8/15	(53%)
NP	9/28	(32%)	PN	7/16	(44%)
PN	9/28	(32%)	NG	6/16	(38%)
PP	8/28	(29%)	FL	4/11	(36%)
FA	5/23	(22%)	AG	2/11	(18%)

Most of the frequent shapes corresponding to sonority plateaus or reversals violating the Sonority Sequencing Principle (SSP, [6], [15]) – PP, PF, FA, NP, and NF – do occur with higher frequency in heteromorphic contexts, as predicted by [9], [20], *inter alia*. Within this moderate data set, only the morphological trends for the PF, PP, and FA shapes are statistically significant (in Fisher’s exact test). The FP shape, which is often argued to have special phonological status [19], [22], occurs with similar frequencies in both morphological contexts.

Interestingly, fewer than half of the frequent shapes corresponding to traditional sonority rises – PL, AG, and FL – favor the tautomorphic context, and none of these trends are significant. Other “preferred” phonotactic patterns – PG, FG, FN, NG, and PN – occur with roughly similar frequencies in both morphological contexts.

4.2 Shapes occurring only in tautomorphic contexts

Of the 49 shapes attested in the sample, there are none that always occur in heteromorphic contexts. By contrast, there are 21 shapes that always occur in tautomorphic contexts. These include all G-initial, D-initial, R-initial, and L-initial shapes and 4/8 of the A-initial shapes. This is despite the fact that G, D, L, R, and A consonants are no less frequent in languages with heteromorphic patterns.

Shapes occurring only in tautomorphic contexts in the sample are: AN, LG (4 languages each); GG, RG (3 languages each); AD, AR, AL, GP, RN, RP, LN, LP, LA, LF (2 languages each); GN, GD, RA, RF, RL, DP, DG (1 language each). Some examples of these rare shapes include Yakama Sahaptin /lk^wi/ ‘overwhelm’ [13], Northern Rgyalrong /rzáŋ/ ‘it is long’ [11], Iu Mien /jwən/ ‘to be level, smooth’ [1], and Nasa Yuwe /tsme^hme/ ‘butterfly’ [21].

4.3 Inventory properties by morphological context

Here we examine the patterns of word-initial CC inventories according to the morphological patterns of the languages in which they occur.

Four languages have only heteromorphic patterns. In all of these, C₁ is restricted to F, P, and N, or just one of these, and C₂ is always less restricted than C₁. For example, in Tzeltal, C₁ is always F but C₂ may be P, A, N, G, or L.

Sixteen languages have only tautomorphic patterns. Just over half of these have relatively small inventories in which C₂ is restricted to just one or two manners, usually G, L, R, or D (e.g., A’ou: NL, PL, FL). However, the largest inventory in the sample is also found in this group (Northern Rgyalrong, 27 shapes), and several others have very large and diverse inventories (e.g., Pnar, Yakama Sahaptin).

Twelve languages have both tautomorphic and heteromorphic patterns in their inventories. Within this group, there is little evidence for a dramatic bifurcation of shapes according to morphological context. Most often, when a heteromorphic shape occurs in one of these languages, it is alongside a tautomorphic counterpart: e.g. Laz AP, AF, FF, FP, GP (tautomorphic) and NP, NA, NF, PN, PP, PA, PF, PG, PL (both patterns). Heteromorphic shapes in this group of languages are structurally similar to those described above, with C₁ tending to be N, P, or F. However, the tautomorphic shapes in this group, whether they have heteromorphic counterparts or not, are unlike those described above in that they have diverse C₂ patterns which often include A, F, P, and N in addition to L, R, D, and G.

5. DISCUSSION

Given the long-standing assumptions regarding the structure of morphologically complex phonotactic patterns and the recent empirical findings supporting them, some of the results of this study are unexpected. Although crosslinguistically frequent word-initial CC shapes corresponding to traditional sonority reversals or plateaus do show a greater tendency to occur in heteromorphic contexts, within this diverse sample heteromorphic shapes are in fact more limited than tautomorphic shapes in many other aspects of their patterning.

Languages which have only heteromorphic patterns in their CC inventories are relatively rare. Within languages that have both morphological patterns, most heteromorphic shapes have tautomorphic counterparts, a pattern also noted by [9] for word-initial obstruent clusters in Slovak. There are no shapes which occur only in heteromorphic contexts across the whole sample.

Also surprising, given prior assumptions and findings, is the huge diversity found in the tautomorphemic patterns. Twenty-one shapes, many corresponding to SSP-violating “reversals” and all of them typologically rare, are found only in tautomorphemic contexts. The largest inventory in the sample is also found in a language with solely tautomorphemic patterns.

The results here suggest that there are far more factors at play, both crosslinguistically and diachronically, in the interaction of morphology and phonotactics than abstract universal sequencing preferences, as also noted by the language-specific studies in [3]. Apart from the overall structure of the consonant phonetic/phonemic inventory [8], one must additionally consider crosslinguistic tendencies in the phonological patterns of stems and affixal/clitic material, as well as common vowel reduction patterns. For example, there is weak evidence that relatively “unmarked” consonant types such as plain stops and nasals are favored in inflectional affixes, though this is not universal [4]. It is also sometimes the case that particular consonant types cannot occur stem-initially in a language (for example, D in Yimas [10]). These tendencies are relevant to the finding that heteromorphemic word-initial CCs tend to be more limited in their sequencing properties.

Further, it is known that vowel reduction and deletion, which is often conditioned by stress, word position, and other non-manner consonant properties such as voicing, can dramatically change a language’s phonotactic patterns. In particular, pretonic vowel syncope can create large inventories of word-initial consonant sequences, whether the language has strong prefixation patterns (e.g., Tataltepec Chatino [23]) or not (e.g., Lezgian [5], [10]).

In light of the findings here, general affixation and cliticization properties and synchronic and historical vowel reduction patterns in the language sample as a whole were considered. Four general language types are identified, defined primarily by the interaction between CC shapes and morphological context, but also characterized by these other properties:

- **Type I (4 lgs.):** solely heteromorphemic patterns occur in these inventories, which are only moderately diverse. Shapes show more restrictions on C_1 than C_2 . All such languages have productive prefixation or procliticization processes.
- **Type II (11 lgs.):** solely tautomorphemic patterns occur in these inventories, which are small and non-diverse. C_2 is often limited to just one or two manners of articulation. In the current sample, roughly half of the languages of this type lack prefixation/procliticization or it is not productive.
- **Type III (5 lgs.):** solely tautomorphemic patterns occur in these inventories, which are large and diverse. These languages have relatively few restrictions on C_1 or C_2 and therefore often include shapes which are typologically rare. Interestingly, all such languages in the current sample have productive prefixation or procliticization, but these morphemes are restricted to syllabic structures like CV or CVC.
- **Type IV (12 lgs.):** both morphological patterns occur in these inventories, which tend to be of moderate to large size. While their heteromorphemic patterns resemble those of Type I, their tautomorphemic patterns are closer to those of Type III. All such languages have productive prefixation or procliticization. In the current sample, half of the languages of this type have either ongoing vowel reduction processes affecting the initial syllable or recent historical vowel reduction patterns which have changed the phonotactic patterns of the language.

The distribution of the languages of the sample among these four types can be found in Table 1.

While the above proposal is tentative, being based upon a diverse but relatively small language sample, it is presented as a potentially useful classification for future investigations of this topic. The different bundles of synchronic features possibly point towards different diachronic paths of development for these types. While the prevalence of vowel reduction in Type IV languages suggests that this is a particularly dynamic language type with a clear diachronic origin, the motivating principles underlying the differentiation of Types II and III from each other, and Type I from the others, are somewhat more obscure and worthy of further investigation.

While there are many additional phonetic, phonological, and morphological factors to consider that cannot be included at this stage, what is presented here is a first step toward an empirical crosslinguistic investigation of the interaction of phonotactic patterns and morphological complexity. The value of using a diverse language sample in this endeavor is proven by the several findings here which challenge previous assumptions. In the future, it would be worthwhile to investigate whether these findings hold for larger consonant sequences or those occurring in other (e.g., word-final) environments.

6. ACKNOWLEDGEMENTS

This research was funded by LABEX ASLAN (ANR-10-LABX-0081) of the Université de Lyon within the French program Investissements d’Avenir.

7. REFERENCES

- [1] Arisawa, T. D. 2016. An Iu Mien grammar: a tool for language documentation and revitalization. (Ph.D. dissertation, La Trobe University.)
- [2] Axenov, S. 2006. The Balochi language of Turkmenistan: a corpus-based grammatical description. (Ph.D. dissertation, Uppsala Universitet.)
- [3] Ball, M. J., Müller, N. (eds.). 2016. *Challenging sonority: cross-linguistic evidence*. Equinox Publishing Ltd.
- [4] Bybee, J. 2005. Restrictions on phonemes in affixes: a crosslinguistic test of a popular hypothesis. *Linguistic Typology* 9: 165-22.
- [5] Chitoran, I., Babaliyeva, A. 2007. An acoustic description of high vowel syncope in Lezgian. *International Congress of Phonetic Sciences XVI*: 2153-56.
- [6] Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In Kingston, J., Beckman, M. (eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press. 282-333.
- [7] Creissels, D., Biaye, S. 2016. *Le balant ganja*. Dakar: IFAN Cheikh Anta Diop.
- [8] Dressler, W. U., Dziubalska-Kołaczyk, K. 2006. Proposing morphonotactics. *Rivista di Linguistica* 18(2): 249-266.
- [9] Dressler, W. U., Hliničanová, M., Durčo, M., Mörth, K., Korecky-Kröll, K. 2015. Phonotactic vs. morphonotactic obstruent clusters in Slovak and German. *Rivista di Linguistica* 27(1): 45-60.
- [10] Foley, W. 1991. *The Yimas language of New Guinea*. Stanford: Stanford University Press.
- [11] Gong, X. 2018. Le rgyalrong zbu, une langue tibéto-birmane de chine du sud-ouest: Une étude descriptive, typologique et comparative. (Ph.D. thesis, Institut National des Langues et Civilisations Orientales.)
- [12] Greenberg, J. 1965/1978. Some generalizations concerning initial and final consonant sequences. *Linguistics* 18: 5-34. (Reprinted in 1978 in Joseph H. Greenberg (ed.), *Universals of Human Language*, vol. 2: *Phonology*. Stanford, CA: Stanford University Press.)
- [13] Hargus, S., Beavert, V. 2014. Northwest Sahaptin. *Journal of the International Phonetic Association* 44(3): 319-342.
- [14] Haspelmath, M. 1993. *A grammar of Lezgian*. (Mouton Grammar Library, 9.) Berlin: Mouton de Gruyter.
- [15] Hooper, J. B. 1976. *An introduction to Natural Generative Phonology*. New York: Academic Press.
- [16] Kreitman, R. 2008. The phonetics and phonology of onset clusters: the case of Modern Hebrew. (Ph.D. dissertation, Cornell University.)
- [17] Maddieson, I. 2013. Syllable structure. In Dryer, M. S., Haspelmath, M. (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/12>).
- [18] McFarland, T. A. 2009. The phonology and morphology of Filomeno Mata Totonac. (Ph.D. dissertation, University of California, Berkeley.)
- [19] Morelli, F. 2003. The relative harmony of /s+stop/ onsets: Obstruent clusters and the Sonority Sequencing Principle. In Féry, C., Van de Vijver, R. (eds.), *The syllable in Optimality Theory*. Cambridge: Cambridge University Press. 356-372.
- [20] Orzechowska, P. 2012. Effects of morphology on the syllable structure. In Stolz, T., Nau, N., Stroh, C. (eds.), *Monosyllables: From phonology to typology*. Berlin: Walter de Gruyter. 91-114.
- [21] Rojas Curieux, T. 1998. *La lengua páez : una visión de su gramática*. Santafé de Bogotá, Colombia: Ministerio de Cultura.
- [22] Selkirk, E. O. 1982. The syllable: The structure of phonological representations. In Van der Hulst, H., Smith, N. (eds.), *Advances in Nonlinear Phonology*. Dordrecht: Foris. 337-383.
- [23] Sullivant, J. R. 2015. The phonology and inflectional morphology of Chá?knyá, Tataltepec de Valdés Chatino, a Zapotecan Language. (Ph.D dissertation, University of Texas at Austin.)