

Highly complex syllable structure

A typological and diachronic study

Shelece Easterday

Studies in Laboratory Phonology





Studies in Laboratory Phonology

Chief Editor: Martine Grice

Editors: Doris Mücke, Taehong Cho

In this series:

1. Cangemi, Francesco. Prosodic detail in Neapolitan Italian.
2. Drager, Katie. Linguistic variation, identity construction, and cognition.
3. Roettger, Timo B. Tonal placement in Tashlhiyt: How an intonation system accommodates to adverse phonological environments.
4. Mücke, Doris. Dynamische Modellierung von Artikulation und prosodischer Struktur: Eine Einführung in die Artikulatorische Phonologie.
5. Bergmann, Pia. Morphologisch komplexe Wörter im Deutschen: Prosodische Struktur und phonetische Realisierung.
6. Feldhausen, Ingo & Fliessbach, Jan & Maria del Mar Vanrell. Methods in prosody: A Romance language perspective.
7. Tilsen, Sam. Syntax with oscillators and energy levels.

ISSN: 2363-5576

Highly complex syllable structure

A typological and diachronic study

Shelece Easterday


Easterday, Shelece. 2019. *Highly complex syllable structure: A typological and diachronic study* (Studies in Laboratory Phonology). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/000>

© 2019, Shelece Easterday

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/> 

ISBN: no digital ISBN

no print ISBNs!

ISSN: 2363-5576

no DOI

ID not assigned!

Cover and concept of design: Ulrike Harbort

Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono

Typesetting software: Xe_{La}TeX

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

langsci-press.org

Storage and cataloguing done by FU Berlin

For Elise, Ada, Astrid, and Maria:

May you always be curious!

Contents

Acknowledgments	iii
Abbreviations	v
1 Language sample	1
1.1 Language sampling	1
1.1.1 Common sources of bias in language sampling	2
1.1.2 Other factors which may influence phonological structure and syllable complexity	4
1.1.3 Specific considerations in the current study	5
1.2 Defining the categories of syllable structure complexity	6
1.3 Constructing the language sample	12
1.4 Language sample for current study	13
1.4.1 Areal features of sample	13
1.4.2 Genealogical features of sample	18
1.4.3 Sociolinguistic features of sample	19
1.5 Data collection	20
References	23
Index	27
Name index	27
Language index	27
Subject index	27

Acknowledgments

This book is a modified version of my Ph.D. dissertation, which was defended in May 2017 at University of New Mexico. While the book features an updated language sample, the findings and interpretations are largely unchanged from the original study. Neither work would have been possible without the mentorship and support of many people along the way.

I am particularly grateful to the members of my dissertation committee. My advisor, Caroline Smith, provided helpful and constructive feedback on all aspects of the dissertation and my progress through the Ph.D. Joan Bybee and Ian Maddieson introduced me to language change and phonological typology, respectively, and I have greatly benefited from the inspiring work and mentorship of both. I thank Ioana Chitoran for many thought-provoking discussions of the findings and the next steps in this line of research. Thanks also to Bill Croft.

While writing the original dissertation I was fortunate to have the financial support of the Joseph H. Greenberg Endowed Fellowship (2011–2015) and the Russell J. and Dorothy S. Bilinski Fellowship (2016–2017).

Since October 2017, I have been employed as a postdoctoral researcher at Laboratoire Dynamique du Langage (UMR5596, CNRS & Université Lyon 2) in Lyon, France. While revising this book, I have been funded by LABEX ASLAN (ANR-10-LABX-0081) of the Université de Lyon within the French program Investissements d’Avenir. DDL has been a warm, welcoming, and intellectually exciting environment in which to work, and for this I am grateful to my supervisor François Pellegrino, laboratory director Antoine Guillaume, and Anetta Kopecka, the head of the Description, Typologie, et Terrain (DTT) research axis. Special thanks also go to my fellow “postdoc ladies” – Natalia Chousou-Polydouri, Marine Vuillermet, Laetitia de Almeida, Elisa Demuru, and Natasha Aralova – for their friendship and moral support during this stage of our lives and careers, and to DDL alumna Kasia Janic for the use of her magical apartment.

Many specialists kindly answered my inquiries about details of the languages and language families discussed in this book, including Doug Marmion (Wutung), Matthew Carter (Bashkir), Netra P. Paudyal (Darai), Jonathan Bobaljik (Itelmen), Andrey Filchenko (Eastern Khanty), Bonny Sands (Hadza), Kirk Miller (Hadza),

Acknowledgments

Logan Sutton (Kiowa-Tanoan), Rosa Vallejos (Cocama-Cocamilla), Sylvia Tufvesson (Semai), Kasia Wojtylak (Murui Huitoto), and Albert Alvarez Gonzalez (Uto-Aztecan). Tim Zingler, Ricardo Napoleão de Souza, Lexie Adams, and Marc Adams helped me with the translation of reference material from the original German, Portuguese, and Spanish.

I am grateful to many dear friends for their support over the years that this work was in progress, including Caron Treloar, Lexie Adams, Erin Debenport, Cora McKenna, Francesca Di Garbo, Giorgio Iemmolo, Logan Sutton, Jason Timm, Ricardo Napoleão de Souza, Fabian Armijo, Nick Mathews, Amanda Delgado Galvan, and Sylvia Tufvesson.

Finally, to my family: thank you for everything.

Abbreviations

1	first person
3	third person
AGT	agentive
CAUS	causative
CL	noun class
CONT.P	past continuous
DAT	dative
DEM	demonstrative
DUB	dubitive
F	feminine
FUT	future
INCL	inclusive
INF	infinitive
INTERROG	interrogative
ITER	iterative
M	masculine
NMLZ	nominalizer
NOM	nominative
PL	plural
POSS	possessive
PS	predicate specifier
PST	past
REALIS	realis
RECP	reciprocal
REFL	reflexive
REL	relativizer
SG	singular
RNS	translated from the original by Ricardo Napoleão de Souza
SME	translated from the original by Shelece Easterday
TZ	translated from the original by Tim Zingler

1 Language sample

This chapter describes the language sample used in the study. In §1.1 I discuss general issues of language sampling and specific considerations for sampling in the current study. In §1.2 I examine a previous typology of syllable structure complexity and propose a definition for a category of Highly Complex syllable structure. In §1.3 I discuss the procedure underlying the construction of the language sample. In §1.4 I present the language sample and describe its areal, genealogical, and sociolinguistic features. In §1.5 I briefly discuss the general method of data collection.

1.1 Language sampling

Crosslinguistic comparison is “the fundamental characteristic of typology” (Croft 2003 : 6). In order to make general statements about some linguistic property such as syllable complexity, it is necessary to examine the properties of and variation within that feature in a wide variety of languages. Today, linguists have access to a greater array of grammatical descriptions, corpora, and audiovisual materials than ever before. However, for many languages, reference materials are either not available or not descriptive enough for inclusion in most typological studies. Therefore researchers must rely on samples much smaller than the set of the 7,097 languages known to be living today (Simons & Fennig 2018). Because typology is a data-driven science, the issue of sampling - that is, determining which languages will serve as data sources for addressing the question(s) at hand - is critical in any study. The relative merits of different sampling techniques and methods of controlling for various types of bias have often been the subject of debate in the field (see Bakker 2011 for an overview of the relevant literature).

Before introducing the sample used in the current study, I discuss some of the known sources of bias in typological work. I also discuss the potential effect of these factors on investigating issues of syllable structure or phonology more generally.

1.1.1 Common sources of bias in language sampling

The three most commonly discussed sources of bias in language sampling are *genealogical*, *areal*, and *bibliographic* bias (Bakker 2011).

A typological study may suffer from *genealogical bias* if it includes data from related languages. This presents a potential confound in the interpretation of results, because similar patterns in related languages may not be independent of one another, but instead inherited from a common ancestor. Of all the sources of bias in language sampling, genealogical bias is perhaps the most discussed, and the one most explicitly controlled for. Strategies for minimizing this kind of bias include systematic stratification of the language sample itself at a particular time depth or level of genealogical classification (Bell 1978 ; Maddieson 1984 ; Dryer 1989), or postponement of sampling to post-hoc analysis, when the independence of the feature(s) under study can be determined at each taxonomic level within a language family (Bickel 2008). In practice, though, most typological studies are based on small convenience samples which are heavily skewed towards large, well-known, and well-described language families. Language isolates, which account for up to one-third of known language families (Campbell 2016), and smaller, lesser-known language families often go altogether unrepresented.

Another bias which is often considered in constructing language samples is *areal bias*, in which languages spoken in the same geographical and/or cultural area may have influenced one another through prolonged contact. The literature has long noted the existence of linguistic areas, in which languages from more than one family share sets of traits in common with each other but not with other related languages spoken outside the area (Aikhenvald & Dixon 2001 ; Chirikba 2008). Attempts to minimize such bias include consideration of cultural areas in addition to genealogical affiliations in constructing a language sample, with the ideal sample containing no two languages from the same family or area (Perkins 1985). But while traditional linguistic areas are relatively small and geographically delimited (e.g., the Balkan Sprachbund), studies of individual features have revealed even larger areas of linguistic convergence; e.g., North America has a higher prevalence of head-marking agreement strategies as compared to the rest of the world (Dryer 1989). To minimize the effects of areal bias, Dryer proposes a method by which a language sample may be divided into five large continental areas (later refined to six, Dryer 1992), which can be shown to be independent of one another along at least some typological measures.

As many as two-thirds of spoken languages do not have reference materials which are thorough enough to be consulted in any but the most basic of typolog-

ical surveys (Bakker 2011 : 106). As a result, virtually every language sample suffers from severe *bibliographic bias*. The best documented languages in the world tend to correspond to those which, for whatever historical reason, have the greatest political, social, and economic power and wide geographic spread (e.g., English, Mandarin Chinese, Spanish, Arabic). Thus we find that bibliographic bias is genealogically and areally skewed, with small, less powerful language families and remote regions particularly underrepresented. The least documented language families in the world, for example, tend to be found in lowland New Guinea and parts of the Amazon region (Hammarström 2010). Similar to, or perhaps a subset of, bibliographic bias is what Moreno Cabrera describes as the *written language bias*. Most of the world’s languages do not have a written or standard form, and references for such languages often describe a specific dialect or idelect. When languages do have a written or standard form, reference materials often describe that form. A result of this is that typological studies often compare data from “highly heterogeneous sources”: standardized, highly formal registers for written languages, and unstandardized, informal registers of idelects for unwritten languages (2008: 118).

All three of the sources of bias described above — genealogical, areal, and bibliographic bias — may complicate typological studies of syllable structure. As discussed in the introduction, the prevalence of particular consonant cluster patterns in Indo-European languages may skew general perceptions regarding the universality of these patterns. There are also clear areal asymmetries in the global distribution of syllable structure complexity. Languages with Simple syllable structure tend to be spoken near the equator (Maddieson 2013). Complex syllable structure has been described as a prominent areal feature of specific regions such as the Caucasus (Chirikba 2008) and the Pacific Northwest (Thompson & Kinkade 1990), though there is evidence that maximal syllable structure is more strongly associated with genealogical affiliation even within these linguistic areas (Napoleão de Souza 2017). Insofar as the global distribution of syllable structure complexity is genealogically and areally skewed, the issue of bibliographic bias is relevant. For example, underdocumented areas such as the New Guinea are known to have a higher than average proportion of languages with canonical (C)V syllable structure (Maddieson 2013).

1.1.2 Other factors which may influence phonological structure and syllable complexity

1.1.2.1 Population

Speaker population may have an affect on language structure. It has been proposed that rare or linguistically marked structures, such as Object-initial basic word orders, are more likely to arise and persist in small speech communities (Nettle 1999a). Lingua franca which have wide areal spreads and large speaker populations with many second-language speakers may be more vulnerable to simplificatory pressures than languages whose use is limited to small, close-knit communities (Nettle 1999b; Lupyán & Dale 2010). These proposals suggest that simpler syllable structures may be found in languages with large speaker populations or in situations of heavy language contact. Recent research on emergent phonology in creoles does not necessarily support the latter claim: despite many claims to the contrary, a broad range of complex syllable patterns may occur in these languages (Schramm 2014). Concerning other phonological properties, a positive correlation between phoneme inventory size and speaker population has been noted (Hay & Bauer 2007). Proposed motivations for this pattern, including founder effects analogous to those found in genetics (Atkinson 2011), are controversial (Bybee 2011; Maddieson et al. 2011; Hunley et al. 2012, *inter alia*).

1.1.2.2 Language endangerment

Today, language diversity is in precipitous decline due to rapid social, economic, and environmental changes which have global reach. It has been projected that 84% or more of today's languages may be lost within the next century (Nettle 1999b : 113-114). There is evidence that language vitality status may influence linguistic structure itself. Languages which are obsolescing (dying) — currently about 13% of all living languages (Simons & Fennig 2018) — are known to undergo special kinds of structural change (see Romaine 2010 for a review of research on the structural effects of language obsolescence). In the phonological systems of obsolescing languages, it is common for distinctive contrasts to be leveled and for the regularity of phonological processes to break down. Attested effects of obsolescence on syllable structure include reduction and simplification of syllable margins (Cook 1989, for obsolescing dialects of Chipewyan and Sarcee) and the deletion of entire unstressed syllables in certain word positions (Mithun 1989, for Oklahoma Cayuga).

1.1.2.3 Ecological factors

There is a growing body of work investigating the effect of ecological factors on language structure. Recent studies have explored the hypothesis that the sound systems of human languages may be adapted to features of the natural environment such as elevation, ambient temperature, density of vegetation, and ambient dessication (Everett 2013 ; Maddieson & Coupé 2015 ; Everett et al. 2016 ; Everett 2017). Syllable structure has been a subject of particular focus in this research paradigm. A positive correlation between warm climates and the frequency of CV shapes in languages has been proposed to reflect the communication needs of an outdoor lifestyle, in which more sonorous elements of the speech signal may overcome environmental noise and large distances between speakers (Munroe et al. 1996 ; Fought et al. 2004). Similarly, a negative correlation between density of vegetation and complex syllable structures may reflect the poor transmission qualities of higher frequency sounds (certain consonants and consonant clusters) in densely forested environments (Maddieson & Coupé 2015).

1.1.3 Specific considerations in the current study

Bakker (2011 : 106) states that “the construction of a sample should follow the precise formulation of the research questions one would like to answer on the basis of it”. Recall the two broad research questions motivating this study, presented in (1)-(2):

- (1) *Do languages with highly complex syllable structure share other phonetic and phonological characteristics such that this group can be classified as a linguistic type?*
- (2) *How does highly complex syllable structure develop over time?*

These research questions, and their more specific formulations laid out in Chapter ??, necessitate a carefully constructed language sample. The first seeks to determine whether there are structural features correlated with highly complex syllable structure such that this phenomenon can be considered characteristic of a language type. For the purposes of gaining a thorough understanding of a language type, genealogical and areal diversity are important considerations in the construction of the sample. However, because this study does not aim to quantify the range or overall distribution of highly complex syllable structure, it is not necessary that the sample be genealogically *balanced* in a systematic way. If quantitative analysis reveals crosslinguistically robust patterns of features correlated with syllable structure complexity, these can inform specific predictions

about the evolution of highly complex syllable structure, addressing the second research question. Thus it may actually prove useful to have pairs of related languages with different syllable structure complexity in the sample. The effectiveness of the predictions can then be tested and qualitatively evaluated within these pairs, in which the structures are known to derive from the same origin at some point in the past.

In order to test for correlations between syllable structure complexity and other structural features, typological bias (Comrie 1989 : 12) must be built into the sample. That is, the languages of the sample must be deliberately chosen on the basis of their syllable patterns, so that features of languages with different syllable structure complexity may be compared against one another in a principled way to determine whether the hypothesized correlations exist and are crosslinguistically robust. An important question, then, is how syllable structure complexity is defined in the current study.

1.2 Defining the categories of syllable structure complexity

Definitions of syllable structure complexity typically consider the size and phonological structure of the onset and coda.¹ A classification which considers the size, shape, and differential behavior of onsets and codas, as well as dominant crosslinguistic patterns is given in Maddieson (2013). While broad, this classification captures prevalent crosslinguistic trends in syllable structure and has proven useful in establishing correlations between syllable structure complexity and other features of language structure, such as consonant phoneme inventory size. In a 486-language survey, languages are classified into three categories according to the size and shape of their largest observed onsets and codas:

Simple: languages in which the onset is maximally one C, and codas do not occur.

Moderately Complex: languages in which the onset is maximally two Cs, the second of which is a liquid or a glide; and/or the coda consists of maximally one C.

Complex: languages in which the maximal onset is two Cs (the second of which is something other than a liquid or a glide) or larger than two Cs; and/or the maximal coda consists of two or more Cs.

¹The size and structure of the nucleus may also be a consideration; cf. Maddieson et al. (2013).

1.2 Defining the categories of syllable structure complexity

The distribution of the 486 languages in Maddieson’s survey according to these three categories can be found in Table 1.1.

Table 1.1: Distribution of syllable structure complexity in languages of Maddieson (2013).

Syllable structure complexity	N languages	Percentage
<i>Simple</i>	61	12.6
<i>Moderately Complex</i>	274	56.4
<i>Complex</i>	151	31.1

In contrast to the tightly-defined Simple and Moderately Complex categories, the Complex category in Maddieson (2013) is diverse and open-ended. Languages in this category range from those whose most complex syllable is only a slight expansion on the Moderately Complex types – such as a CVCC shape in which the first consonant of the coda is limited to a liquid or glide – to those having far more complex structures involving up to eight consonants in a tautosyllabic cluster. In order to better understand the internal structure of the Complex category and the distribution of languages within it, I analyzed the size and shape of the maximal syllable structure of these 151 languages. A list of the languages in the sample, along with the references consulted, can be found in Appendix C. The distribution of these languages can be found in Table 1.2, according to the size of their maximal onsets and codas.²

The visual distribution of the languages in Table 1.2 is striking: most languages cluster toward the upper lefthand corner of the table. Over half of the languages in the Complex category have syllable-marginal clusters of two consonants or fewer; that is, onsets and codas no more complex than a sequence of two obstruents. Though onsets of up to seven consonants and codas of up to eight consonants occur, it is very rare for a language to have more than four consonants at either margin (7/147, 4.8%).

Another interesting property of the distribution in Table 1.2 is that as cluster size increases, the crosslinguistic size asymmetry in onsets and codas appears to level out and then reverse. As discussed in Chapter ??, it is more common for a language to have complex onsets than complex codas. When smaller cluster sizes are considered, this pattern is clear even within the Complex category: lan-

²Note that there are only 147 languages represented in Table 1.2. Four of the languages in the Complex category in Maddieson (2013) – Canela-Krahô, Ik, Indonesian, and Yagaria – have been reclassified as having Moderately Complex syllable structure in the current analysis.

1 Language sample

Table 1.2: Distribution of languages in Complex syllable structure category in Maddieson (2013), according to the size of the most complex onset (columns) and coda (rows) structures observed in each language.

Coda	Onset						
	C	CC	CCC	CCCC	CCCCC	CCCCCC	CCCCCCC
(none)		2	1	1	-	-	-
C		23	9	1	-	-	-
CC	35	21	13	3	-	-	-
CCC	3	5	8	3	-	-	-
CCCC	4	2	4	2	-	1	-
CCCCC	1	-	-	1	-	-	1
CCCCCC	-	1	-	-	1	-	-
CCCCCCC	-	-	-	-	-	-	-
CCCCCCCC	-	-	-	1	-	-	-

guages which have three or more consonants in the onset are more common (50 languages) than languages which have these patterns in the coda (38 languages). However, when larger clusters are considered, we find that there is a reversal of the pattern: it is *less* common for languages to have onsets of four or more consonants (15 languages) than to have codas of the same size (19 languages). Examining the specific patterns in more depth than what is presented in Table 1.2, it seems that this reversal occurs when clusters of three *obstruents* are taken as the cutoff point: 24 languages have these, or more complex clusters, as an onset, and 26 languages have these, or more complex clusters, as a coda.

2.2.1 Long sequences of obstruents: tautosyllabic clusters or syllabic consonants?

The above point relates to an interesting feature of attested large onset and coda patterns, also mentioned in Chapter ??, which is that many of these structures do not exhibit the sonority-related sequencing restrictions and contours that are so common of languages with smaller onsets and codas. It is not unusual to observe tautosyllabic clusters consisting entirely of obstruents in languages with large onsets or codas (3)-(4):

- (3) **Cocopa** (*Cochimi-Yuman*; USA, Mexico)
pʂtʃʔaːw
‘I hang up several (things)’

(Crawford Jr 1966 : 36)

- (4) **Itelmen** (*Chukotko-Kamchatkan*; Russia)

qsatxtʃ

‘Follow!’

(Georg & Volodin 1999 : 44)

Thus a common characteristic of the more extreme cases of Complex syllable structure is the presence of long sequences of obstruents. This phenomenon can also be found in Tashlhiyt, a language which is classified as having Complex syllable structure, but which descriptions list as having only simple (Ridouane 2008) or maximally biconsonantal (PuechLouali1999) onsets and codas. Because the language also has syllabic obstruents, it is possible to find words without vowels which consist entirely of voiceless obstruents (5a-b):

- (5) **Tashlhiyt** (*Afro-Asiatic*; Morocco)

a. *tʃsχt*

‘you cancelled’

b. *tʃtχtst*

‘you rolled it (F)’

(Ridouane 2002 : 95)

There is evidence in the literature that word-initial *onset clusters* and word-initial *consonant sequences with syllabic obstruents* exhibit different gestural timing patterns between the rightmost consonant and vowel. Goldstein et al. (2007) instrumentally investigated the predictions of a coupled oscillator model of syllable structure for these two different kinds of word-initial sequences. In Georgian, the timing lag between gestures associated with the rightmost consonant and the vowel decreased with larger word-initial consonant sequences, in line with the model’s predictions for onset clusters. Thus word-initial consonant sequences, such as that in *tʃs’k’riala* ‘shiny clean,’ can be interpreted as syllable onsets in Georgian. In Tashlhiyt, no such pattern was found with increasingly larger word-initial consonant sequences; that is, the instrumental evidence suggested that only the consonant immediately preceding the vowel was coupled to it. Thus the initial sequence in *tsmun* ‘3FS-CAUS-accompany’ was interpreted as being syllabified as [ts.mun]. These results, expanded and confirmed by HermesEtAl2011, lend support to the analysis of Tashlhiyt as having simple onsets and nucleus patterns which include syllabic obstruents, rather than large onset clusters.

While the results summarized above suggest that onset clusters and word-initial consonant sequences with syllabic consonants are categorically different structures in terms of articulation, the situation is actually more complex than that. The Georgian data in Goldstein et al. (2007) was from two speakers, but only one speaker showed the expected effect. The other speaker had timing patterns which more closely resembled the Tashlhiyt pattern. An examination of this speaker's productions revealed the regular presence of an 'epenthetic' vowel between [k'] and [ɾ] in *k'riali* 'glitter' and *ṡs'k'riala* 'shiny clean.' When epenthesis occurred, the decreased timing lag between the rightmost consonant and vowel was not observed; that is, the word-initial sequence of consonants was split by a syllable boundary. The authors note that epenthesis occurred in the speech of both Georgian speakers, but not always in the same forms. However, when epenthesis did occur, it was in very specific phonological environments: when the place of articulation for C₁ was more posterior than that of C₂. Setting aside the issue of vowel epenthesis for now,³ an interesting finding from Goldstein et al. (2007) is that both timing patterns may occur for the same word-initial sequence in the same language.

A related observation regarding large tautosyllabic clusters and syllabic obstruents is that some languages are analyzed as having both in their canonical syllable structure. For example, Crawford Jr (1966) analyzes Cocopa as having large onset clusters, such as [pʃtʃʔ] in (3) above. In addition to these, he proposes that some unstressed syllables can be entirely consonantal, consisting of "an onset only or of an onset and a coda with a predictable 'murmur' vowel following the onset as phonetic peak" (1966: 34). For example, *pʃtʃxmuka' p* 'he embraced her' is syllabified as [p¹.ʃtʃ^a.mu.ka' p] (1966: 43; quality of 'murmur' vowel determined by environment). Patterns of the latter sort occur for specific consonantal combinations in Cocopa, much like the Georgian patterns described above.

There are also a number of languages which are analyzed by one author as having large tautosyllabic clusters, and by another as having simpler syllable margins but also syllabic obstruents. Hoard (1978) gives several examples of languages from the Pacific Northwest which had previously been analyzed as having large onsets and/or codas but might be better thought of as having smaller syllable margins in addition to syllabic stops and affricates (e.g. Quileute, Puget Salish, and Nez Perce). Hoard bases his analyses on impressionistic transcriptions, considering syllables to reflect "the number of audible pulses" in the form which are delineated by "relative separation or detachment" from other segments or

³The issue of vowel epenthesis is an important one which has obvious implications for the analysis of syllable structure; this will be discussed further in §??.

groups of segments (1978: 59-60). Similar disagreements of analysis can be found for Piro (Arawakan; Matteson 1965 ; Lin 1997 ; Hanson 2010). Perhaps these languages are like Georgian and Cocopa above, in which different timing patterns may be produced by different speakers or for different consonant combinations. Without experimental data it is difficult to say whether one or both analyses are appropriate. What is clear, however, is that it is not uncommon for the same language, when exhibiting long sequences of consonants (especially obstruents) at word margins, to be alternately described as having large tautosyllabic clusters and/or syllabic obstruents. For this reason I argue that it is appropriate to group languages of both types together in the current study.

The above approach also has some support in instrumental findings in the literature regarding the articulatory properties of consonantal nuclei. Pouplier & Beňuš (2011) found that for syllabic liquids in Slovak, the kinematic properties of the consonantal gestures did not undergo consistent changes in nucleus position. The authors conclude that in this respect, syllables with consonantal nuclei behave like consonant clusters. According to other instrumental measures, syllabic liquids exhibit timing and gestural coordination relationships with adjacent consonants which are reminiscent of those found in onset or coda clusters, but still distinct from those found in both non-syllabic consonant clusters and vocalic nuclei. Similarly, Fougeron & Ridouane (2008) found that /k/ in Tashlhiyt does not undergo consistent acoustic or articulatory changes when syllabic. However, it does exhibit consistent and stable patterns of temporal alignment and coordination relationships with flanking consonants which are, as noted in the discussion of Goldstein et al. (2007) above, similar to those expected for onset-nucleus-coda patterns. Thus the instrumental evidence point to syllabic consonants as phenomena which exhibit some of the coordination properties of vocalic nuclei while maintaining articulatory properties of non-syllabic consonants.

2.2.2 Highly Complex syllable structure: a definition

Motivated by all of the observations above — the distribution of languages in Table 1.2, three-obstruent clusters as the locus of the weakening of the onset/coda asymmetry, and the similar patterns observed in languages with large tautosyllabic obstruent clusters and those with syllabic obstruents — I define Highly Complex syllable structure as follows:

Highly Complex: languages in which the maximal onset or coda consists of three obstruents, or four or more Cs of any kind; and/or in which syllabic obstruents occur, resulting in word-marginal sequences of three or more obstruents.

Table 1.3 shows how the 486 languages in Maddieson (2013) are distributed with the addition of the Highly Complex category as defined above. The defini-

1 Language sample

tion of the Complex category is adjusted accordingly, and the Simple and Moderately Complex categories are defined as in the previous work.

Table 1.3: A reanalysis of the data in Maddieson (2013), with the added category of Highly Complex as defined in the current work.

Syllable structure complexity	<i>N</i> languages	Percentage
<i>Simple</i>	61	12.6
<i>Moderately Complex</i>	278	57.2
<i>Complex</i>	110	22.6
<i>Highly Complex</i>	37	7.6

While the structural divisions between the syllable structure complexity categories are not evenly distributed, they capture predominant crosslinguistic patterns and serve the specific aims of the current study. These categories provide a four-point scale by which syllable structure complexity can be correlated with other structural features. The Highly Complex category is also defined in such a way as to include the extreme end of the syllable structure complexity cline, but not so narrowly as to introduce extreme genealogical and areal bias into this group. This will allow for meaningful examination, both quantitative and qualitative, of the characteristics of this group of languages.

1.3 Constructing the language sample

Three criteria shaped the design of the language sample for this study: (i) the size of the sample must be large enough for meaningful quantitative analysis, but small enough for in-depth qualitative analysis of the languages with highly complex syllable structure; (ii) the proportional representation of the four syllable structure complexity categories must be similar enough to allow for meaningful comparisons between the groups, and (iii) as per the discussion of bias in typological studies in §1.1, the sample should be both genealogically and areally diverse.

In addressing (i), a sample of approximately 100 languages is appropriate. While this sample size is moderate for a phonological typology survey, it is comparable to those used in studies such as Bateman (2007) and Bybee & Easterday (2019), both of which have quantitative and qualitative components. Given the relative crosslinguistic rarity of languages in the Simple and Highly Complex categories (Table 1.3) and the skewed geographical and genealogical distribution of syllable

structure complexity, meeting criteria (ii) and (iii) is a more difficult endeavor. In constructing the sample, an attempt was made to strike a reasonable balance between the ideal sample composition and these practical considerations. The resulting sample is described in the following section.

1.4 Language sample for current study

The language sample includes 24 languages in the Simple category, 26 languages in the Moderately Complex category, 25 languages in the Complex category, and 25 languages in the Highly Complex category. Tables ??–1.7 list the languages of the sample by syllable structure complexity, geographical macro-area, and genealogical affiliation. A more detailed list which includes ISO 639-3 codes, speaker populations, and language endangerment and development status, can be found in Appendix A.

1.4.1 Areal features of sample

The 100 languages are roughly evenly distributed among the six geographical macro-areas as defined by Dryer (1989 ; 1992).⁴ Africa, Australia & New Guinea, North America, and South America are represented by 17 languages each. Two regions, Eurasia and Southeast Asia & Oceania, are represented by 16 languages each. See Figure 1.1.1 for a plotted map of the languages of the sample.

[MAP]

Geographic distribution of languages in the sample, with colors denoting syllable structure complexity. S=Simple, MC=Moderately Complex, C=Complex, HC=Highly Complex.

Figure 1.1: .1.

⁴The geographical macro-areas are specifically defined as follows (Dryer 1989 : 268; Dryer 1992 : 83, 133–135). **Africa**: continent of Africa, including Semitic languages of southwest Asia. **Australia & New Guinea**: Australian continent and Melanesia, excluding Austronesian languages of Melanesia. **Eurasia**: Eurasian landmass, excluding Semitic and languages from families of southeast Asia as defined below, and including the Munda languages of Austroasiatic. **North America**: North American continent, including languages of Mexico, Mayan and Aztec languages in Central America, and some branches of Chibchan-Paezan. **South America**: South American continent, including languages of Central America except Mayan and Aztec languages, and some Chibchan-Paezan branches. **Southeast Asia & Oceania**: Southeast Asian region, including all Sino-Tibetan, Tai-Kadai, Hmong-Mien, and Austroasiatic languages excluding Munda, and Oceania region (Austronesian languages).

1 Language sample

Table 1.4: Languages in **Simple** syllable structure category, by macro-area and genealogical affiliation.

Region	Language	Top-level family	Subfamily
AFRICA			
	Hadza	(isolate)	
	Southern Grebo	Atlantic-Congo	<i>Volta-Congo</i>
	Yoruba	Atlantic-Congo	<i>Volta-Congo</i>
	Ma'di	Central Sudanic	<i>Moru-Madi</i>
	Southern Bobo Madaré	Mande	<i>Western Mande</i>
AUSTRALIA & NEW GUINEA			
	Savosavo	(isolate)	
	Grass Koiari	Koiarian	<i>Koiaric</i>
	Rotokas	North Bougainville	<i>Rotokas-Askopan</i>
	East Kewa	Nuclear Trans New Guinea	<i>Enga-Kewa-Huli</i>
NORTH AMERICA			
	Towa	Kiowa-Tanoan	
	Pinotepa Mixtec	Otomanguean	<i>Eastern Otomanguean</i>
	Ute	Uto-Aztecan	<i>Northern Uto-Aztecan</i>
SOUTH AMERICA			
	Urarina	(isolate)	
	Warao	(isolate)	
	Apurinã	Arawakan	<i>Southern Maipuran</i>
	Murui Huitoto	Huitotoan	<i>Nuclear Witotoan</i>
	Cavineña	Pano-Tacanan	<i>Tacanan</i>
	Cubeo	Tucanoan	<i>Eastern Tucanoan</i>
SOUTHEAST ASIA & OCEANIA			
	Rukai (Budai dialect)	Austronesian	
	Maori	Austronesian	<i>Malayo-Polynesian</i>
	Tukang Besi North	Austronesian	<i>Malayo-Polynesian</i>
	Saaroa	Austronesian	<i>Tsouic</i>
	Sichuan Yi	Sino-Tibetan	<i>Burmo-Qiangic</i>
	Sumi Naga	Sino-Tibetan	<i>Kuki-Chin-Naga</i>

1.4 Language sample for current study

Table 1.5: Languages in **Moderately Complex** syllable structure category, by macro-area and genealogical affiliation.

Region	Language	Top-level family	Subfamily
AFRICA			
	Kambaata	Afro-Asiatic	<i>Cushitic</i>
	Ewe	Atlantic-Congo	<i>Volta-Congo</i>
	Fur	Furan	
	Kanuri	Saharan	<i>Western Saharan</i>
AUSTRALIA & NEW GUINEA			
	Maybrat	Maybrat-Karon	
	Kamasau	Nuclear Torricelli	<i>Marienberg</i>
	Selepet	Nuclear Trans New Guinea	<i>Finisterre-Huon</i>
	Alyawarra	Pama-Nyungan	<i>Arandic-Thura-Yura</i>
EURASIA			
	Kharia	Austroasiatic	<i>Munda</i>
	Telugu	Dravidian	<i>South Dravidian</i>
	Darai	Indo-European	<i>Indo-Iranian</i>
	Tu	Mongolic	<i>Southern Periphery</i>
			<i>Mongolic</i>
	Eastern Khanty	Uralic	<i>Khantyic</i>
NORTH AMERICA			
	Karok	(isolate)	
	North Slavey (Hare dialect)	Athabaskan-Eyak-Tlingit	<i>Athabaskan-Eyak</i>
	Kalaallisut	Eskimo-Aleut	<i>Eskimo</i>
	Choctaw	Muskogean	<i>Western Muskogean</i>
SOUTH AMERICA			
	Carib	Cariban	<i>Guianan</i>
	Imbabura Highland Quichua	Quechuan	<i>Quechua II</i>
	Cocama-Cocamilla	Tupian	<i>Maweti-Guarani</i>
SOUTHEAST ASIA & OCEANIA			
	Pacoh	Austroasiatic	<i>Katuic</i>
	Paiwan	Austronesian	
	Kim Mun (Vietnam dialect)	Hmong-Mien	<i>Mienic</i>
	Atong	Sino-Tibetan	<i>Brahmaputran</i>
	Cantonese	Sino-Tibetan	<i>Sinitic</i>
	Lao	Tai-Kadai	<i>Kam-Tai</i>

1 Language sample

Table 1.6: Languages in **Complex** syllable structure category, by macro-area and genealogical affiliation.

REGION	Language	Top-level family	Subfamily
AFRICA			
	Mpade (Makari dialect)	Afro-Asiatic	<i>Chadic</i>
	Jola-Fonyi	Atlantic-Congo	<i>North-Central Atlantic</i>
	Lunda	Atlantic-Congo	<i>Volta-Congo</i>
	Dizin (Central dialect)	Dizoid	
	Gaam	Eastern Jebel	
AUSTRALIA & NEW GUINEA			
	Mangarrayi	Mangarrayi-Maran	
	Nimboran	Nimboranic	
	Oksapmin	Nuclear Trans New Guinea	<i>Asman-Awyu-Ok</i>
	Bardi	Nyulnyulan	<i>Western Nyulnyulan</i>
	Ngarinyin	Worrorran	
EURASIA			
	Basque (Central dialect)	(isolate)	
	Burushaski	(isolate)	
	Nivkh (West Sakhalin dialect)	(isolate)	
	Bashkir	Turkic	<i>Common Turkic</i>
	Ket	Yeniseian	<i>Northern Yeniseian</i>
NORTH AMERICA			
	Pech	Chibchan	
	Tzeltal (Aguacatenango dialect)	Mayan	<i>Core Mayan</i>
	Lakota	Siouan	<i>Core Siouan</i>
SOUTH AMERICA			
	Kadiwéu	Guaicuruan	
	Mamaindê	Nambiquaran	<i>Nambikwara Complex</i>
	Apinayé	Nuclear-Macro-Je	<i>Je</i>
	Chipaya	Uru-Chipaya	
SOUTHEAST ASIA & OCEANIA			
16	Koho (Sre dialect)	Austroasiatic	<i>Bahnaric</i>
	Lelepa	Austronesian	<i>Malayo-Polynesian</i>
	Lepcha	Sino-Tibetan	<i>Himalayish</i>

1.4 Language sample for current study

Table 1.7: Languages in Highly Complex syllable structure category, by macro-area and genealogical affiliation.

REGION	Language	Top-level family	Subfamily
AFRICA			
	Tashlhiyt	Afro-Asiatic	<i>Berber</i>
	Doyayo	Atlantic-Congo	<i>Volta-Congo</i>
	Bench	Ta-Ne-Omotic	
AUSTRALIA & NEW GUINEA			
	Menya	Angan	<i>Nuclear Angan</i>
	Kunjen (Oygangand dialect)	Pama-Nyungan	<i>Paman</i>
	Alamblak	Sepik	<i>Sepik Hill</i>
	Wutung	Sko	<i>Nuclear Skou-Serra-Piore</i>
EURASIA			
	Kabardian	Abkhaz-Adyge	<i>Circassian</i>
	Itelmen	Chukotko-Kamchatkan	
	Albanian (Tosk dialect)	Indo-European	<i>Albanian</i>
	Polish	Indo-European	<i>Balto-Slavic</i>
	Georgian	Kartvelian	<i>Georgian-Zan</i>
	Lezgian	Nakh-Daghestanian	<i>Daghestanian</i>
NORTH AMERICA			
	Passamaquoddy-Maliseet	Algic	<i>Algonquian</i>
	Cocopa	Cochimi-Yuman	<i>Yuman</i>
	Mohawk	Iroquoian	<i>Northern Iroquoian</i>
	Yakima Sahaptin	Sahaptian	<i>Sahaptin</i>
	Thompson	Salishan	<i>Interior Salish</i>
	Tohono O'odham	Uto-Aztecan	<i>Southern Uto-Aztecan</i>
	Nuu-chah-nulth	Wakashan	<i>Southern Wakashan</i>
SOUTH AMERICA			
	Camsá	(isolate)	
	Yine	Arawakan	<i>Southern Maipuran</i>
	Tehuelche	Chonan	<i>Continental Chonan</i>
	Qawasqar	Kawesqar	<i>North Central Alacalufan</i>
SOUTHEAST ASIA & OCEANIA			
	Semai	Austroasiatic	<i>Aslian</i>

1 *Language sample*

There are some asymmetries in the areal representation of syllable structure complexity in the sample. In Eurasia, the Simple category is entirely unrepresented. In Southeast Asia & Oceania, the Highly Complex category is represented by only one language. In North America, the Highly Complex category is relatively overrepresented, accounting for seven of the languages in that region.

1.4.2 Genealogical features of sample

The 100 languages of the sample belong to 74 different language families.⁵ 64 of the language families are represented by one language each; this figure includes nine language isolates and many small and mid-sized families. Another three language families — Arawakan, Pama-Nyungan, and Uto-Aztecan — are represented in the sample by two languages each. All of the language families represented by more than two languages in the sample — Afro-Asiatic, Atlantic-Congo, Austroasiatic, Austronesian, Indo-European, Nuclear Trans New Guinea, and Sino-Tibetan — are within the top ten in the world by size in number of languages (Hammarström et al. 2018). Every attempt was made to maximize the diversity of subfamilies within the top-level families represented by more than one language in the sample. Subfamilies represented by more than one language are Volta-Congo (Atlantic-Congo family, five languages), Malayo-Polynesian (Austronesian family, three languages) and Southern Maipuran (Arawakan family, two languages). There are no pidgins or creoles in the sample.⁶

Most often, a language family is represented by only one language within a syllable structure complexity category. The Simple category is somewhat less genealogically diverse than the others, with only 19 families represented by the 24 languages.

Recall from §1.1.3 that an important feature of the language sample design is the inclusion of pairs of related languages with maximally different syllable structure complexities. This allows for hypotheses about the diachronic development of highly complex syllable structure to be tested at the local level. The sample includes five relevant pairs from five macro-areas: Ute (Simple) and Tohono O’odham (Highly Complex), both Uto-Aztecan from North America; Apuriná (Simple) and Yine (Highly Complex), both Arawakan from South America;

⁵For genealogical affiliations, I use the classifications in Glottolog 3.3 (Hammarström et al. 2018).

⁶There is debate over whether Cocama-Cocamilla may be appropriately classified as a creole. While the language is typically classified as Tupi-Guaraní, it is quite divergent from other languages of the family in aspects of its phonology and morphosyntax, perhaps owing to factors of language contact. See Vallejos Yopán (2010) for a critical discussion of this topic.

Yoruba (Simple) and Lunda (Complex), both Atlantic-Congo from Africa; Darai (Moderately Complex) and Albanian (Highly Complex), both Indo-European from Eurasia; and Maori (Simple) and Lelepa (Complex), both Austronesian from South-east Asia & Oceania. Some pairs of languages are more closely related than others.⁷

1.4.3 Sociolinguistic features of sample

Speaker population data and language vitality status classifications for the languages of the sample can be found in Appendix A.

The L1 speaker populations for the languages of the sample vary widely, ranging from five for Tehuelche (Chonan, Patagonia) and Yakima Sahaptin (Sahaptian, Pacific Northwest) to 74,244,300 for Telugu (Dravidian, Southern India). In general, the languages in the Simple and Moderately Complex categories have larger speaker populations than those in the Complex and Highly Complex categories, though there are many exceptions to this trend. Of the 26 languages with fewer than 1,000 speakers, 18 of them have Complex or Highly Complex syllable structure, and half of the ten languages with fewer than 100 speakers have Highly Complex syllable structure. The prevalence of very smaller speaker populations among languages in these categories is no doubt related to the high concentration of these languages in geographical areas with high rates of language endangerment.

The Expanded Graded Intergenerational Disruption Scale (EGIDS) is an assessment of language vitality developed by Lewis & Simons (2010), following work by Fishman (1991), UNESCO (Brenzinger et al. 2003), and others. It considers many different factors of language use, including rates and means of intergenerational transmission, domains of language use, and official recognition of the language. Using EGIDS as a starting point, Ethnologue (Simons & Fennig 2018) has developed a coarse-grained estimate of the relative development versus endangerment of languages. By this measure, languages are classified into categories according to their development/vitality status: Institutional, Developing, Vigorous, In Trouble, and Dying.⁸ Languages with robust vitality are more com-

⁷The region of Australia & New Guinea is not represented in these pair comparisons. The only family represented by more than one language in this region is Nuclear Trans New Guinea; however some of its subfamily classifications are controversial or not well demonstrated (Pawley 2005).

⁸The language development categories are defined as follows (Simons & Fennig 2018). **Institutional**: language has wide use in the home and community and official status at educational, provincial, national, and/or international levels. **Developing**: language is used in the home,

1 Language sample

mon in the Simple and Moderately Complex portions of the sample. The Complex and Highly Complex categories have the highest proportion of languages classified as Dying (6/25 languages in each category). It should be noted that in North America, which has perhaps the highest proportion of languages with Highly Complex patterns of all the macro-areas, rates of language endangerment and language loss are extreme. Ethnologue 21 classifies 237/254 (93%) of the living languages spoken north of the US-Mexico border in North America as In Trouble or Dying (Simons & Fennig 2018).

The sociolinguistic features of the language sample lead to an interesting observation: highly complex syllable structure is a rare language feature by many different measures, including non-structural ones. As described in Chapter ??, highly complex phonotactic patterns are often treated as anomalies and theoretical outliers, especially when occurring in underdescribed non-European languages. A very small proportion of the world's languages have structures of this kind. Many of the languages with these structures are found in parts of the world where language endangerment rates are particularly high, and have correspondingly small speaker populations. Thus highly complex syllable structure is a marginalized pattern in theoretical, descriptive, historical, and social terms. This is all the more reason to dedicate a typological study to this linguistic feature.

1.5 Data collection

The data used for this study was collected from published reference grammars, phonetic and phonological studies, and other relevant language descriptions. In a few cases, an expert researcher on the given language was additionally consulted. Data was collected with the guidance of coding sheets whose questions were designed to address the research questions and hypotheses of each chapter. As the methodology behind the coding of the data differs for each part of the study, it will be discussed separately within each chapter.

Because the language references consulted were written in different time periods and reflect a variety of descriptive practices, they use diverse transcription

community, and sometimes broader contexts, and in initial stages of developing a system of writing and standardization. **Vigorous:** language is used in the home and community by speakers of all generations, but has not yet developed a system of graphization or standardization. **In Trouble:** language is currently in the process of losing intergenerational transmission, with the community shifting to other languages for daily use, but there are still speakers of child-bearing age. **Dying:** language has lost intergenerational transmission entirely, all fluent speakers are above child-bearing age.

standards. Using the phonetic descriptions of sounds in the language references, all phoneme inventories and phonetic and phonemic transcriptions have been transcribed using International Phonetic Alphabet (IPA) symbols. Where there is some ambiguity in interpreting the phonetic description provided by the source consulted, this has been noted in Appendix B. In the examples given in this chapter and throughout the book, non-IPA symbols have been replaced by their IPA equivalents.

References

- Aikhenvald, Alexandra & R. M. W. Dixon. 2001. Introduction. In Alexandra Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 1–26. Oxford: Oxford University Press.
- Atkinson, Quentin D. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332(6027). 346–349.
- Bakker, Dik. 2011. Language sampling. In Jae Jung Song (ed.), *Handbook of linguistic typology*, 100–127. Oxford: Oxford University Press.
- Bateman, Nicoleta. 2007. *A crosslinguistic investigation of palatalization*. University of California, San Diego dissertation.
- Bell, Alan. 1978. Language samples. In Joseph H. Greenberg (ed.), *Universals of human language: Method and theory*, vol. 1, 123–156. Stanford: Stanford University Press.
- Bickel, Balthasar. 2008. A refined sampling procedure for genealogical control. *Sprachtypologie und Universalienforschung* 61. 221–233.
- Brenzinger, M., A. Yamamoto, N. Aikawa, D. Koundioubu, A. Minasyan, A. Dwyer, C. Grinevald, M. Krauss, O. Miyaoka, O. Sakiyama, R. Smeets & O. Zepeda. 2003. *Language vitality and endangerment*. Paris: UNESCO Ad Hoc Expert Group Meeting on Endangered Languages. <http://www.unesco.org/culture/en/endangeredlanguages>, accessed 2016-9-12. (Available online at Accessed on .
- Bybee, Joan. 2011. How plausible is the hypothesis that population size and dispersal are related to phoneme inventory size? Introducing and commenting on a debate. *Linguistic Typology* 15(2). 147–153.
- Bybee, Joan & Shelece Easterday. 2019. *Consonant strengthening: A crosslinguistic survey and articulatory proposal*. To appear in *Linguistic Typology* 23.
- Campbell, Lyle. 2016. Language isolates and their history, or, what's weird, anyway? In Jeremy Steffman Rolle Nicholas & John Sylak-Glassman (eds.), *Proceedings of the 36th annual meeting of the Berkeley linguistics society, february 6-7, 2010*, 16–31. Berkeley, CA: Berkeley Linguistics Society.

References

- Chirikba, Viacheslav A. 2008. The problem of the caucasian sprachbund. In Pieter Muyskens (ed.), *From linguistic areas to areal linguistics*, 25–93. Amsterdam: John Benjamins.
- Comrie, Bernard. 1989. *Language universals and linguistic typology*. Chicago: University of Chicago Press.
- Cook, Eung-Do. 1989. Is phonology going haywire in dying languages? Phonological variations in Chipewyan and sarcee. *Language in Society* 18(2). 235–255.
- Crawford Jr, James Mack. 1966. *The cocopa language*. University of California at Berkeley dissertation.
- Croft, William. 2003. *Typology and universals*. 2nd edition. Cambridge: Cambridge University Press.
- Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.
- Everett, Caleb. 2013. *Evidence for direct geographic influences on linguistic sounds: The case of ejectives*. PLOS ONE 8: e65275.
- Everett, Caleb. 2017. *Languages in drier climates use fewer vowels*. *Frontiers in Psychology* 8: 1285.
- Everett, Caleb, Damián E. Blasi & Sean G. Roberts. 2016. Language evolution and climate: The case of desiccation and tone. *Journal of Language Evolution* 1(1). 33–46.
- Fishman, Joshua A. 1991. *Reversing language shift*. Clevedon: Multilingual Matters.
- Fougeron, Cécile & Rachid Ridouane. 2008. On the phonetic implementation of syllabic consonants and vowel-less syllables in tashlhiyt. *Estudios de Fonética Experimental* 18. 139–175.
- Fought, John G., Robert L. Munroe, Carmen R. Fought & Erin M. Good. 2004. Sonority and climate in a world sample of languages: Findings and prospects. *Cross-cultural research* 38(1). 27–51.
- Georg, Stefan & Alexander P. Volodin. 1999. *Die itelmenische Sprache: Grammatik und texte*. Weisbaden: Harrassowitz Verlag.
- Goldstein, Louis, Ioana Chitoran & Elisabeth Selkirk. 2007. Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the XVIth International Congress of Phonetic Sciences*, 6–10 August 2007, Saarbrücken, Germany. 241–244.

- Hammarström, Harald. 2010. The status of the least documented language families in the world. *Language Documentation and Conservation* 4. 177–212.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath. 2018. *Glottolog* 3.3. Jena: Max Planck Institute for the Science of Human History. (Available online at <http://glottolog.org>, Accessed on .)
- Hanson, Rebecca. 2010. *A grammar of Yine (piro)*. La Trobe University, Melbourne dissertation.
- Hay, Jennifer & Laurie Bauer. 2007. Phoneme inventory size and population size. *Language* 83(2). 388–400.
- Hoard, James E. 1978. Syllabification in Northwest Indian languages, with remarks on the nature of syllabic stops and affricates. In Alan Bell & Joan Bybee Hooper (eds.), *Syllables and segments*, 59–72. New York: North-Holland Publishing Company.
- Hunley, Keith, Claire Bowern & Meghan Healy. 2012. Rejection of a serial founder effects model of genetic and linguistic coevolution. *Proceedings of the Royal Society B: Biological Sciences* 279(1736). 2281–8.
- Lewis, M. Paul & Gary F. Simons. 2010. Assessing endangerment: Expanding Fishman’s GIDS. *Revue Roumaine de Linguistique* 55(2). 103–120.
- Lin, Yen-Hwei. 1997. Syllabic and moraic structures in piro. *Phonology* 5. 403–436.
- Lupyan, Gary & Rick Dale. 2010. *Language structure is partly determined by social structure*. PLOS ONE 5(1): e8559.
- Maddieson, Ian. 1984. *Patterns of sounds*. With a chapter contributed by Sandra Ferrari Disner. Cambridge: Cambridge University Press.
- Maddieson, Ian. 2013. Syllable structure. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Maddieson, Ian, Tanmoy Bhattacharya, D. Eric Smith & William Croft. 2011. Geographical distribution of phonological complexity. *Linguistic Typology* 15(2). 267–279.
- Maddieson, Ian & Christophe Coupé. 2015. *Human spoken language diversity and the acoustic adaptation hypothesis*. Proceedings of Meetings on Acoustics 25(1).
- Maddieson, Ian, Sébastien Flavien, Egidio Marsico, Christophe Coupé & François Pellegrino. 2013. LAPSyD: Lyon-Albuquerque Phonological Systems Database. In Proceedings of the 14th Interspeech Conference, Lyon, France, 25–29 August 2013.
- Matteson, Esther. 1965. *The piro (arawakan) language*. Berkeley & Los Angeles: University of California Press.

References

- Mithun, Marianne. 1989. The incipient obsolescence of polysynthesis: Cayuga in Ontario and Oklahoma. In Nancy C. Dorian (ed.), *Investigating obsolescence: Studies in language contraction and death (studies in the social and cultural foundations of language 7)*, 243–257. Cambridge: Cambridge University Press.
- Munroe, Robert L., Ruth H. Munroe & Stephen Winters. 1996. Cross-cultural correlates of the consonant-vowel (CV) syllable. *Cross-Cultural Research* 30. 60–83.
- Napoleão de Souza, Ricardo. 2017. *A comparison of maximal syllable structure in four linguistic areas*. Paper presented at 91st Annual Meeting of the Linguistic Society of America; Austin, Texas, 5–8 January 2017.
- Nettle, Daniel. 1999a. *Is the rate of linguistic change constant?* *Lingua* 108. 119–136.
- Nettle, Daniel. 1999b. *Linguistic diversity*. New York: Oxford University Press.
- Poupplier, Marianne & Štefan Beňuš. 2011. On the phonetic status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology* 2. 243–273.
- Ridouane, Rachid. 2002. Words without vowels: Phonetic and phonological evidence from tashlhiyt berber. *ZAS Papers in Linguistics* 28. 93–110.
- Ridouane, Rachid. 2008. Syllables without vowels: Phonetic and phonological evidence from tashlhiyt berber. *Phonology* 25. 321–359.
- Romaine, Suzanne. 2010. Contact and language death. In Raymond Hickey (ed.), *The handbook of language contact*, 320–339. Oxford: Wiley-Blackwell.
- Schramm, Mareile. 2014. *The emergence of creole syllable structure: A cross-linguistic study*. Berlin/Munich/Boston: De Gruyter Mouton.
- Simons, Gary F. & Charles D. Fennig (eds.). 2018. *Ethnologue: Languages of the world, twenty-first edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Thompson, Laurence C. & M. Dale Kinkade. 1990. Languages. In Wayne Suttles (ed.), *Handbook of North American Indians* (, Northwest coast), vol. 7, 30–51. Washington: Smithsonian Institution.
- Vallejos Yopán, Rosa. 2010. *A grammar of kokama-kokamilla* (ph. D. dissertation, University of Oregon.)

Did you like this book?

This book was brought to you for free

Please help us in providing free access to linguistic research worldwide. Visit <http://www.langsci-press.org/donate> to provide financial support or register as a community proofreader or typesetter at <http://www.langsci-press.org/register>.



Highly complex syllable structure

Set blurb on back with \BackBody{my blurb}

