San Diego Linguistics Papers **4** *(2013) 53-61*

A Preliminary Statistical Analysis of the Relationship between Syllable Structure Complexity and Sonorant Inventory Size

Bethany Keffala
University of California, San Diego

1. Introduction

The development of online databases of language information, such as WALS (the World Atlas of Language Structures, Dryer & Haspelmath 2011) and UPSID (UCLA Phonological Segment Inventory Database, Maddieson 1984, Maddieson & Precoda 1990), make investigations of cross-linguistic patterns faster and easier, especially on a large scale. For instance, Maddieson (2007) analyzed data from a set of 625 languages (WALS) to challenge the assumption that all languages are equally complex, and the consequent implication that a language with greater complexity in one dimension must also have less complexity in another. If it is the case that total systemic complexity is equal across languages, we might expect that languages with more complex syllable structures should tend to have smaller consonant inventories, while languages that allow only simple syllable structures should have larger consonant inventories; complexity in one area of the phonological system is balanced by simplicity in another. What Maddieson found instead was the opposite pattern. Languages with complex syllables tended to have larger consonant inventories while those with simpler syllable inventories tended also to have smaller consonant inventories.

The aim of the following paper is to further examine the cross-linguistic pattern found in Maddieson (2007) in which complexity of syllable structure positively correlates with size of consonant inventories. Specifically, I will investigate the sizes of obstruent and sonorant inventories across the same syllable type categories used by Maddieson (2007), as well as the size of the total consonant inventory. We should expect the size of the total consonant inventory to increase with an increase in complexity of the syllable type inventory (i.e., we expect replication of Maddieson's results). However, the increase in the total consonant inventory could result from an increase in the sonorant inventory alone, an increase in the obstruent inventory alone, or an increase in both sonorant and obstruent inventories. Since greater syllabic complexity can result not only from singleton codas but onset and coda consonant clusters (Selkirk 1982, Blevins 1995, Maddieson 2007), and because onset and coda clusters tend cross-linguistically to employ sonorants closest to the vowel (Selkirk 1982, Clements 1990, Blevins 1995), I predict that the increase in the size of the consonant inventory as a whole is driven by an increase in the size of the sonorant inventory. Additionally, codas cross-linguistically tend to prefer segments with greater sonority (Selkirk 1982, Clements 1990), further suggesting that larger consonant inventory size in languages with greater syllabic complexity may be attributed to an increase in number of sonorant consonants specifically.

2.0 Syllable Structures and Consonant Inventories

In his (2007) study, Maddieson categorizes languages into three syllable inventory types: simple, moderately complex, and complex. Languages in the 'simple' group allow CV syllables, and may allow onsetless syllables, (C)V. Languages in the 'moderate' group additionally allow simple codas (no more than one consonant in a coda) and/or a restricted set of onset clusters (no more than two consonants, the second of which must be a liquid or glide). Finally, languages in the 'complex' group allow any range of onset or coda consonant clusters in addition to or apart from those allowed in the moderate group.

Maddieson found that languages with simple syllables had statistically significantly fewer consonants (N = 62 languages, M = 17.66 consonants) than languages with either moderate (N = 317 languages, M = 21.30 consonants) or complex (N = 180 languages, M = 25.28 consonants) syllable systems, and that languages with moderate syllable systems had statistically significantly fewer consonants than languages with complex syllable systems. Higher sonority segments are cross-linguistically preferred in the second position of onset clusters, in singleton codas, or the first position of coda clusters (Selkirk 1982, Clements 1990, Blevins 1995). Since increases in syllable complexity involve the addition of one or more of these positions, it stands to reason that the increase in consonant inventory size that is correlated with the increase in syllable structure complexity could come from an increase in the inventory of sonorants specifically.

A potential problem, as Maddieson (2007) notes, is that an increase in the size of sonorant inventories may also be correlated with an increase in the size of consonant inventories in general:

> …the relationships found cannot be obviously accounted for by "natural" or "functional" considerations. For example, the tendency for larger consonant inventory and more complex syllable structure to go together cannot be "explained" by considerations of which types of consonants are more likely, from a cross-linguistic point of view, to occur in clusters. In general, the larger consonant inventories are more likely to contain a higher proportion of consonants that cross-linguistically tend to have their distribution restricted both in clusters and in coda position. (p. 103)

If we find that the size of sonorant inventories increases as the complexity of syllable structures increases, it is possible that this may simply be due to increase in size of consonant inventories overall. Lindblom & Maddieson (1988) argue that languages tend to maintain a relatively constant ratio of obstruents (about 70%) to sonorants (about 30%). However, a recent cross-linguistic study by Hauser (2013) suggests that, while obstruent and sonorant inventory sizes are positively correlated, the relationship is perhaps not as strong as we might expect. Using data from 628 language varieties in P-base (Mielke 2008), Hauser showed that correlation between sonorant and obstruent inventory size (r = 0.23) explained a small to moderate amount of the variance (Cohen 1992). Instead, the size of a given sound class was best predicted by the size of an adjacent sound class along the sonority hierarchy, but only within the sets of obstruents, sonorants, or vowels. In other words, more obstruents beget more obstruents, and more sonorants beget more sonorants, but more obstruents do not necessarily beget more sonorants. This being the case, investigation of a possible relationship between syllable complexity and sonorant inventory size is a reasonable and potentially worthwhile line of inquiry. In the following section, I outline the sample selection procedure.

3.0 Methods and Materials

Using data from WALS and UPSID, I investigated possible relationships between complexity of syllable structure and size of sonorant and obstruent inventories. In the absence of such online databases, collecting and analyzing even moderately sized samples of cross-linguistic data would require a non-trivial amount of time. However, the use of data from these sources is not without issues. In particular, inclusion of a language in the database is not random, with some language families more heavily represented than others (Bickel 2008). Additionally, the method of recording linguistic data requires some level of abstraction, which may in some cases obscure cross-linguistic similarities or differences (Simpson 1999). Rather than using the entirety of the non-random sample, Bickel (2008) suggests sampling that controls for genealogical and areal factors.

In the WALS database online (accessed July, 2013), one can access a list of languages with simple (61 languages), moderately complex (274 languages), and complex (151 languages) syllable structures (categorized according to the method used in Maddieson 2007). Within each syllable type category starting with simple systems, I randomized the list of languages given by WALS, and took the first language as the first data point in the sample. In order to control for genealogical factors, I allowed only one data point from any language Family within a syllable type category sample (e.g. admitting Hawaiian into the sample of simple syllable languages disallowed the admission of any other Austronesian language into the simple syllable sample, though at most one Austronesian language would still be allowed in each of the moderately complex and complex syllable samples). In order to control for areal factors, I allowed at most three data points from the same geographical area (as defined by Ethnologue[1]). Using these strict selection criteria, the largest the sample size for the simple syllable group could be was 8 languages. Therefore, the sample sizes for the moderately complex and complex groups were each also constrained to 8 languages, subject to the same selection criteria, yielding a total sample size of 24 languages.

The following table (1) shows the languages in each sample:

[1] Ethnologue defines five world areas: Africa, Americas, Asia, Europe, Pacific.

(1) Languages included in the study sample

	Simple (N=8)	Moderately Complex (N=8)	Complex (N=8)
Language (Family, Area)	Rotokas (North Bougainville, Pacific)	Bandjalang (Australian, Pacific)	Abipon (Mataco-Guaicuru, Americas)
	Hawaiian (Austronesian, Americas)	Tiruray (Austronesian, Pacific)	Ket (Yeniseian, Asia)
	Ekari (Trans-New Guniea, Pacific)	Hixkaryana (Cariban, Americas)	Sa'ban (Austronesian, Asia)
	Barasano (Tucanoan, Americas)	Temein (Nilo-Saharan, Africa)	Beja (Afro-Asiatic, Africa)
	Warao (isolate, Americas)	Mambila (Niger-Congo, Africa)	German (Indo-European, Europe)
	Lelemi (Niger-Congo, Africa)	Xiamen (Sino-Tibetan, Asia)	Kota (Dravidian, Asia)
	Naxi (Sino-Tibetan, Asia)	Amuesha (Arawaken, Americas)	Amuzgo (Oto-Manguean, Americas)
	Hadza (Khoisan, Africa)	Iraqw (Afro-Asiatic, Africa)	Nez Perce (Penutian, Americas)

For each language in the sample, the number of sonorants, obstruents, and total consonants (sonorants+obstruents) was recorded using the segment inventories available in UPSID (accessed July, 2013). The following section presents statistical analyses of the relationships between syllable structure complexity and sonorant, obstruent and total consonant inventory sizes.

4.0 Results and Discussion

The following tables (2-4) present counts of sonorants and obstruents, sonorant-to-obstruent ratios, and total consonant counts for each language, grouped by syllabic complexity. Averages across languages for each group are found in the last row of each table.

(2) Simple syllable structure

	Language (Family, Area)	Sonorants	Obstruents	Son/ Obs	Total Consonants
Simple Syllable Structure	Rotokas (North Bougainville, Pacific)	1	5	0.2	6
	Hawaiian (Austronesian, Americas)	4	4	1	8
	Ekari (Trans-New Guniea, Pacific)	4	6	0.667	10
	Barasano (Tucanoan, Americas)	3	8	0.375	11
	Warao (isolate, Americas)	5	6	0.833	11
	Lelemi (Niger-Congo, Africa)	7	15	0.467	22
	Naxi (Sino-Tibetan, Asia)	4	34	0.118	38
	Hadza (Khoisan, Africa)	7	51	0.137	58
	AVERAGES:	4.375	16.125	0.475	20.5

Simple syllable languages have an average of 4.375 sonorants and 16.125 obstruents, with an average sonorant to obstruent ratio of 0.475, close to the ratio expected by Lindblom & Maddieson (1988). The average number of consonants total is 20.5, larger than the average of 17.66 found in Maddieson (2007). Upon visual inspection of the data for simple syllable languages, Naxi and Hadza seem to diverge from the rest of the sample, and could be outliers (other languages in the sample have between 6 and 22 consonants and between 4 and 15 obstruents, while Naxi has 38 consonants of which 34 are obstruents, and Hadza has 58 consonants of which 51 are obstruents). However, given that the sample size is relatively small, it is difficult to tell.

Table (3) presents the data from the sample of languages with moderately complex syllable structure (those with singleton codas and/or less marked onset clusters). Languages with moderate syllable structure complexity have an average of 6.875 sonorants and 12.75 obstruents, with an average sonorant to obstruent ratio of 0.714, with a greater proportion of sonorants than predicted by Lindblom & Maddieson (1988). The average number of consonants total is 19.625, smaller than Maddieson's (2007) average of 21.3.

(3) Moderately complex syllable structure

	Language (Family, Area)	Sonorants	Obstruents	Son/ Obs	Total Consonants
Moderately Complex Syllable Structure	Bandjalang (Australian, Pacific)	8	4	2	12
	Tiruray (Austronesian, Pacific)	7	9	0.778	16
	Hixkaryana (Cariban, Americas)	7	11	0.636	18
	Temein (Nilo-Saharan, Africa)	7	11	0.636	18
	Mambila (Niger-Congo, Africa)	7	12	0.583	19
	Xiamen (Sino-Tibetan, Asia)	3	16	0.188	19
	Amuesha (Arawaken, Americas)	8	14	0.571	22
	Iraqw (Afro-Asiatic, Africa)	8	25	0.32	33
	AVERAGES:	6.875	12.75	0.714	19.625

Table (4) presents data from the sample of languages that allow complex syllable structures.

(4) Complex Syllable Structure

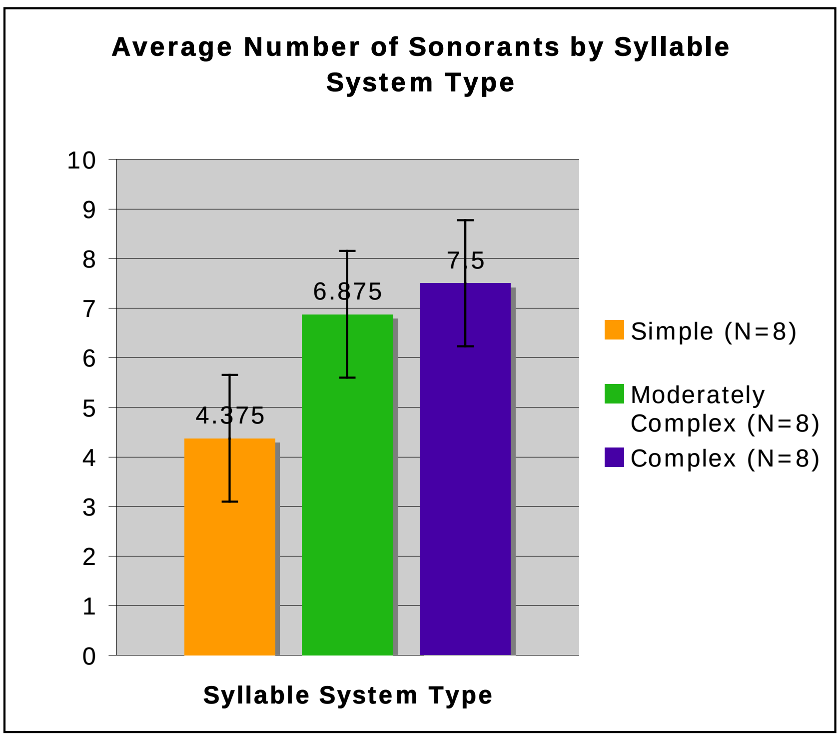
	Language (Family, Area)	Sonorants	Obstruents	Son/ Obs	Total Consonants
Complex Syllable Structure	Abipon (Mataco-Guaicuru, Americas)	7	8	0.667	15
	Ket (Yeniseian, Asia)	6	12	0.5	18
	Sa’ban (Austronesian, Asia)	8	11	0.727	19
	Beja (Afro-Asiatic, Africa)	6	15	0.4	21
	German (Indo-European, Europe)	6	16	0.375	22
	Kota (Dravidian, Asia)	9	14	0.643	23
	Amuzgo (Oto-Manguean, Americas)	8	17	0.471	25
	Nez Perce (Penutian, Americas)	10	15	0.667	25
	AVERAGES:	7.5	13.5	0.582	21

Languages with complex syllable structure have an average of 7.5 sonorants and 13.5 obstruents, with an average sonorant-to-obstruent ratio of 0.582, again with a higher proportion of sonorants than expected by Lindblom and Maddieson (1988). The average number of consonants total is 21, again smaller than Maddieson's average of 25.28 consonants for languages with complex syllable structure.

A one-way ANOVA did not find significant differences in total consonant inventory size between the three syllable groups ($F(2, 21) = 0.029919$, $p = 0.971$). This could be due to the inclusion of Hadza and Naxi in the simple syllable group, however the averages for total consonant inventory size for the moderately complex and complex samples were also smaller than expected. A lack of difference in consonant inventory size between syllable groups is unexpected, given Maddieson's (2007) contrary results. A one-way ANOVA testing for differences in obstruent inventory size was non-significant ($F(2, 21) = 0.2198$, $p = 0.804$), though this could again be due to possible outliers in the simple syllable group. A one-way ANOVA testing for differences in sonorant-to-obstruent ratios was also non-significant ($F(2, 21) = 0.7747$, $p = 0.474$). A one-way ANOVA testing for differences in sonorant inventory size was significant ($F(2, 21) = 7.321$, $p < 0.01$). Post-hoc t-tests revealed significant differences in sonorant inventory size between simple (M=4.38, SD=2.00) and moderately complex (M=6.88, SD=1.64) groups ($t(14)=2.7362$, $p < 0.05$), as well as between simple and complex (M=7.50, SD=1.51) groups ($t(14)=3.5305$, $p < 0.01$). Sonorant inventory sizes between moderately complex and complex groups were not significantly different ($t(14)= 0.7920$, $p=0.4426$).

These results suggest that languages that allow codas or consonant clusters tend to have more sonorant segments than languages that maximally allow (C)V syllables:

(5) Languages with simple syllables have fewer sonorants



This finding fits well with the hypothesis that the existence of positions that prefer higher sonority segments should be correlated with a greater number of sonorants. Such a hypothesis predicts a difference between simple and non-simple syllable system groups, but not necessarily between moderately complex and complex syllable system groups, though it is possible that inclusion of more data could cause such a difference to emerge. Additionally, because there was not a significant difference found for total consonant inventory size but there was a significant difference found for sonorant inventory size, this suggests that sonorant inventory size should be separable from total consonant inventory size. However, it would be interesting to conduct the same investigation separating sonorant and obstruent inventories using Maddieson's larger sample.

5.0 Conclusions

This preliminary study used online language databases to conduct an investigation of cross-linguistic phonological patterns pertaining to syllable complexity and sonorant inventory size; as syllable systems increase in complexity, sonorant inventory sizes also increase. While neither WALS nor UPSID are themselves random samples of the world's languages, neither is the set of accessible linguistic grammars. Such databases can provide useful insights in relatively little time, and can be used while controlling for genealogical and areal factors. Future research should attempt to enlarge the sample sizes to maximize use of the linguistic data available. For example, Bickel's algorithm for genealogically controlled sampling (2008) may allow more languages into the sample while still controlling for confounding genealogical and areal factors. Additionally, future investigations should separate the moderately complex syllable group into languages that do or do not allow complex onsets, since the set of segments allowed in the second position of complex onsets may be more restricted than the set of segments allowed in singleton codas. Finally, relationships between sonorant inventory size and consonant inventory size within syllable structure types should be investigated to ensure that total consonant inventory size does not act as a confounding factor.

6.0 References

Bickel, B. (2008). A refined sampling procedure for genealogical control. *Language Typology and Universals* 61:221-33.

Blevins, J. (1995). The syllable in phonological theory. In: J. Goldsmith (ed.) *The Handbook of Phonological Theory*. Cambridge, MA: Blackwell; 206-44.

Clements, G.N. (1990). The role of the sonority cycle in core syllabification. In: J. Kingston & M. Beckman (eds.) *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech*. New York: Cambridge University Press; 283-333.

Cohen, J. (1992). "A power primer". *Psychological Bulletin* 112 (1): 155–159. doi:10.1037/0033-2909.112.1.155. PMID 19565683.

Dryer, M.S. & M. Haspelmath, (eds.) (2011). *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. Available online at http://wals.info/ Accessed on 2011-12-11.

Hauser, I. (2013). Sonority as a primitive: Evidence from phonological inventories. *WCCFL*. Vol. 31. 2013.

Lewis, M.P. (ed.) (2009). *Ethnologue: Languages of the World, Sixteenth edition*. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/.

Lindblom, B. & I. Maddieson (1988). Phonetic universals in consonant systems. *Language, Speech, and Mind* 62–78.

Maddieson, I. (1984). *Pattern of Sounds*. Cambridge University Press.

Maddieson, I. & K. Precoda (1990). Updating UPSID. *UCLA Working Papers in Phonetics*, 74, 104-111.
Maddieson, I. (2007). Issues of phonological complexity: Statistical analysis of the relationship between syllable structures, segment inventories and tone contrasts. M.J. Solé, P. Beddor and M. Ohala (eds.), *Experimental Approaches to Phonology*. Oxford University Press, Oxford and New York: 93-103.
Maddieson, I. (2011). Consonant Inventories. Dryer, M. S. & M. Haspelmath (eds.) *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library, chapter 1. Available online at http://wals.info/chapter/1 Accessed on 2011-12-11.
Maddieson, I. (2011). Syllable Structure. In: Dryer, M.S. & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library, chapter 12. Available online at http://wals.info/chapter/12 Accessed on 2011-12-11.
Marsico, E., I. Maddieson, C. Coupé & F. Pellegrino (2004). Investigating the "Hidden" structure of phonological systems. *Berkeley Linguistic Society* 30.
Mielke, J. (2008). *The Emergence of Distinctive Features*. Oxford University Press, Oxford.
Selkirk, E. (1982). The syllable. v. d. Hulst, H. & N. Smith (eds.) *The Structure of Phonological Representations (Part II)*. Dordrecht: Foris Publications; 337-83.
Simpson, A. (1999). Fundamental problems in comparative phonetics and phonology: does UPSID help to solve them? Proceedings of ICPhS.

Bethany Keffala
Department of Linguistics
University of California, San Diego
bkeffala@ucsd.edu