A Preliminary Investigation of Quantitative Patterns in Sonority Sequencing

Stefan A. Frisch

Frisch (2004) claims that phonetically or psycholinguistically-motivated sound patterns should be found typologically across languages (qualitatively), as well as typologically within a language or within the lexicon of a language (quantitatively). This claim emerged from the analysis of Obligatory Contour Principle for Place of Articulation (OCP-Place) consonant patterns in Arabic and a number of unrelated languages as a pattern of similarity avoidance. In this paper, a preliminary study is presented that examines sonority sequencing for consonant clusters within individual languages, and within the lexicons of individual languages for consistent quantitative patterns. Sonority sequencing constraints can be seen as phonetically-motivated in articulation or perception (e.g. Ohala & Kawaski-Fukumori 1997, MacNeilage & Davis 2000). Therefore, sonority sequencing patterns should be present not just in cross-linguistic typology but also in the quantitative patterns within individual languages.

1. Introduction

Linguistically significant phonotactic patterns need not be absolute, but can be statistical. Greenberg (1950) examined consonant combinations in the Arabic lexicon and proposed that cases where consonant pairs were significantly underrepresented compared to expected probability can be used to make linguistically significant generalizations. In these types of studies, the type frequency (frequency in the dictionary) rather than the token frequency (frequency of usage) is generally considered the relevant measure (Bybee 1995) and this approach is supported by children's acquisition of phonological patterns (Munson 2001). More generally, the statistical patterns within the lexicon of phonological forms have been referred to as probabilistic phonotactics (e.g. Vitevitch *et al*. 1999).

1.1. Probabilistic phonotactics

The human cognitive system is known to track frequencies of experiences (in language that would be approximated by corpus token frequencies). These experiences are categorized and classified into abstractions at multiple levels (Munson *et al.* 2011). In the case of phonology, experience with language sound structures is used to define phonological categories (Pierrehumbert 2003). It has been hypothesized that this type of abstraction is also useful in language acquisition to parse words from speech (Saffran *et al.* 1998), assign word internal phonological and morphological structure (Treiman *et al.* 2000, Hay *et al.* 2003), and provide language specific information for bilinguals (Betancourt 2013; Messer *et al.* 2010; Sebastián-Gallés & Bosch 2002)

Beyond phonotactics, frequency has been shown to influence language processing in a variety of ways. Perception and production are influenced by phoneme frequency or word neighborhood density (Luce & Pisoni 1998; Vitevitch *et al.* 1999). Experiment participants appear to have a sense of phonological frequencies that is reflected in metalinguistic judgments of well-formedness of novel nonwords (Bailey & Hahn 2000). Abstract but quantitative patterns are known by language speakers (Frisch & Zawaydeh 2001; Zuraw 2010; Hayes *et al.* 2009) and have been modeled theoretically in statistical detail (Anttila 2007; Coetzee & Pater 2008; Frisch *et al.* 2004, Hayes & Wilson 2008). These patterns can be modeled by relatively unsophisticated connectionist networks as long as the network is forced to make some sort of generalization (Alderete *et al.* 2013).

1.2. Functionally-grounded phonological explanation and quantitative patterns

Linguistic research has often sought functional explanation for the common typological patterns in the world's languages. Early functional explanations were often dismissed due to their unreliable nature within and across languages. For example, since languages have both harmony and dissimilation processes, it would seem implausible that there is a functional constraint either preferring more similar segments or less similar segments. Recent theoretical developments have made these apparent contradictions less relevant. Optimality Theory, for example, contains mechanisms for resolving conflicting constraints and proposes that in the absence of other factors a functionally unmarked form will emerge universally (Prince & Smolensky 2004). In addition, taking a quantitative perspective on these constraints makes it possible to observe functional influences that involve either more subtle distinctions or are not absolute in their effects. For example, in the case of dissimilation versus harmony it would appear that the Obligatory Contour Principle for Place of Articulation (OCP-Place) is pattern of similarity avoidance (e.g. Frisch *et al.* 2004; MacEachern 1999) but that identity or identity along a particular dimension may be a crucial factor (McCarthy 1994; Berent & Shimron 1997).

The probabilistic view has also created the opportunity to observe a functional phonological restriction where there is no categorical pattern. Similarity avoidance has been found statistically in English and many other unrelated languages. Additionally, interactions between forces may provide an explanation for the relative strength of the pattern within a language. In the case of similarity avoidance it may be that repeated similar consonants are especially difficult to process given the templatic morphology of Arabic languages where these constraints were originally observed qualitatively (Berg & Abd-El-Jawad 1996).

Frisch (2004) claims on the basis of the similarity avoidance constraint that patterns that are difficult to process cognitively, in perception, or in production will be less frequent than patterns lacking such difficulty. Traditionally, language universals are observed by examining patterns across languages for commonalities, statistical distributions, or implicational relationships. If we add statistical lexical patterns to our search for universal forces acting on language then we can expect that functional forces acting on the sound patterns of language should be found at any level of phonological analysis. Alternatively, this patterning can be viewed as explaining systematic patterns in the number of apparently accidental gaps in otherwise permissible combinations (Frisch *et al.* 2004).Examining quantitative data provides a strong test for claims of functional patterning in phonology. If there is a claim that a pattern is functionally-motivated, it should be found quantitatively in degrees to the extent to which forms obey or violate the proposed functional restriction. In the present study, typologically known patterns in consonant clusters based on sonority are examined within languages for quantitative patterning.

1.3. Functionality of sonority

Sonority constraints on consonant clusters propose that there are manner of articulation restrictions on consonant clusters. Traditionally, low sonority classes are obstruents (stops and fricatives) and high sonority classes are resonant consonants and vowels. Researchers vary number of sonority sub-classes that are defined. Consonant clusters are generally preferred that maximize sonority differences (and when extended to vowels, sonority classification can potentially explain common patterns in syllable structure, e.g. Basbøll 1994). As with many phonological phenomena, there are two broad approaches to the functionality of sonority constraints, based in production or perception. From the production side, sonority sequencing may be the result of ease of production when chunking speech into units (Lindblom 1983; MacNeilage & Davis 2000; Redford 2000). The cycle of jaw movement from opening to closing to opening again provides a production-based framework for transitioning between low sonority speech sounds (jaw closing or raised for obstruents) to high sonority speech sounds (jaw opening or lowered for vowels). Violations of this sequencing require rapid and precise jaw movements that reverse the overall direction of movement, which would be functionally challenging. Alternatively, violations of this sequencing would be the basis for reorganization of the suprasegmental sound structure. In either case, we might expect violating phonological structures to be dispreferred.

From the perception side, sonority modulation may lead to easier perception of cues to speech sounds and phonological structure. Alternations in amplitude from low to high and/or alternations in high versus low dominant frequency spectra would make speech segments and overall syllabic structure easier to identify (Ohala & Kawasaki-Fukumori 1997). It has also been found that vowel and consonant information are recoverable from either the vowel center or the closure/transition portions of the speech signal, and so the redundant encoding of speech information in both the high and low frequency portions of the spectrum makes speech more robust in a noisy environment.

Given these previous proposals for the functional basis of sonority constraints, specific quantitative hypotheses can be generated. For sonority sequencing, consonant cluster types with larger steps in sonority difference toward the peak of the syllable will be preferred. For sonority modulation, consonant cluster types with larger steps in sonority difference will be preferred, regardless of sequencing. In both cases, a quantitative functional analysis proposes that the degree to which a form is functionally good or bad correlates with the frequency of use of that form in phonotactic patterns within a language. If within language patterns do not vary by frequency as a factor of the degree of sonority difference, then the functional quantitative hypothesis is not supported.

2. Methods

2.1. Quantitative lexical typological study

The present study examined consonant cluster type frequency for 47 languages (see appendix for sources). Consonant cluster type frequency in this study is the frequency of occurring CC clusters within a language compared to the set of possible clusters given the types of C1 and C2 that are used in clusters. Data are aggregated within broad manner of articulation classes in order to compare rates of occurrence of different types of clusters by sonority. The present study examines initial, medial, and/or final clusters within the language, depending on the language's phonology and the available data. Note that the source must specifically list the attested consonant clusters in order to conduct a quantitative analysis. If a source only specifies that types of clusters are found it is unusable for quantitative analysis as the apparently accidental gaps cannot be examined. The present study examined CC clusters as the statistical analysis of two segment constituents is the simplest and most straightforward case to analyze (see Pierrehumbert 1994, Frisch 1996, 2000 for issues related to statistical analysis of larger constituents).

The analysis procedure can be demonstrated with a relatively simple language example, the Papuan language Abun (spoken in the West Papua region of Indonesia, Berry & Berry 1999). Word initially, Abun only allows C-glide clusters. If sonority influences the occurrence of clusters then we would expect a relatively greater occurrence of clusters with a large sonority difference (e.g. stop-glide) compared to clusters with a smaller sonority difference (e.g. nasal-glide). However, the raw count of clusters can be deceiving. Abun has a greater variety of stop consonants than nasals so a straightforward comparison of the number of clusters is a misleading representation of the phonotactic possibilities. Specifically, Abun has 7 stops, 4 prenasalized stops, 3 fricatives, 3 nasals, and 2 glides. Table 1 shows the actual number of C-glide clusters for each of these groups, the possible number of C-glide clusters given the consonant inventory of the language, and the relative type frequency of C-glide clusters in each group computed by dividing the actual number by the possible number (A/P). The Abun pattern follows the predictions of sonority modulation and sonority sequencing if we look at the relative number of cluster types. Prenasalized stops are presented as a separate category with higher sonority than stops, but if they were grouped together with the plain stops it would not change the findings, it would just make the example slightly more trivial.

Table 1. Abun C-glide clusters.

C1	Actual	Possible	A/P
Stop	6	7 x 2 = 14	0.43
Prenasalized stop	3	4 x 2 = 8	0.38
Fricative	1	3 x 2 = 6	0.17
Nasal	1	3 x 2 = 6	0.17

A qualitative analysis of the sonority patterns in Abun has no explanation for the lack of use of some consonant clusters within each sonority level. Such apparently accidental gaps are common, but when examined quantitatively can be seen to be not entirely accidental (Frisch *et al.* 2004). Specifically, there is quantitative variation in cases where otherwise expected consonant clusters are not used. This variation depends on the goodness of the combination from the perspective of sonority, and would be otherwise unexplained if these gaps were considered truly accidental. While in a typical phonological analysis, these gaps are not explained phonologically, they are systematic quantitatively.

As a more complicated example, consider coda CC clusters in English with A/P shown in table 2. In this table, individual consonant counts are not included and the quantitative findings are given in a distilled form with the ratio of use of the possible cluster inventory within each grouping.

Table 2. English CC coda clusters.

C2 / C1	Stop	Fricative	Nasal	Liquid
Stop	0.06	0.10		
Fricative	0.06	0.02		
Nasal	0.22	0.25		
Liquid	0.92	0.69	0.67	0.25

In order to determine the degree to which a language obeys sonority constraints, data like that in table 2 can be examined by comparing the ratio of actual vs. possible clusters in adjacent cells (e.g. stop-stop vs. stop-fricative). These are minimal steps along the sonority scale. In the case of sonority sequencing, for example, we would predict that fricative-stop codas are more frequent than stop-stop codas which are more frequent than stop-fricative codas. Sonority modulation predicts stop-stop codas are less frequent than fricative-stop codas or stop-fricative codas. With bigger steps in the sonority scale, the differences should become even larger. However, in order to minimize the potential influence of a single larger or smaller numerical value (which is a priori equally likely to fit the hypothesis or not) the cells are only compared in adjacent pairs. This is a stronger test of sonority as each step along the scale is required to change quantitatively in a systematic way (increasing for better sonority combinations, decreasing for worse sonority combinations).

The present analysis starts with comparisons based on sonority modulation, which predicts that A/P will increase away from the main diagonal of the table for all cluster positions. This is relatively easy to compute. In the English coda example in table 2, there is one case where the frequency decreases farther from the main diagonal (nasal-fricative at 0.25 greater than nasal-stop at 0.22). There are three cases where clusters with better sonority modulation are unattested adjacent to cells with attested clusters with worse sonority modulation (adjacent to stop-fricative, fricative-fricative, and liquid-liquid). There are also several ties (one of stop-stop and stop-fricative at 0.06 and several where there are no attested clusters). For the analysis, ties where both adjacent cells have no attested clusters will not be counted as data. Ties where adjacent cells both contain clusters will be counted as not violating sonority modulation. In other words, only quantitative reversals in sonority for attested clusters are considered violations of the hypothesized quantitative pattern. In the case of the English coda consonants, there are 13 pairs of adjacent cells where sonority modulation is respected and 4 pairs where it is reversed, so 13/17 = 76.5% of the comparisons respect sonority modulation. In this way, the entire distribution of consonant cluster patterns with respect to sonority can be distilled to a single value representing the degree to which sonority constraints are obeyed quantitatively across the cluster types in the language.

This analysis, while suggestive, is preliminary and should be viewed with caution due to a number of caveats. First, the finding is based on a one-step pairwise comparison measure between adjacent cells, and so does not capture potentially phonologically relevant subpatterns in specific cases. Second, the data points are not independent in two ways. Within each language, a single data cell is involved in multiple comparisons with adjacent cells. Using pairwise comparisons for adjacent steps on the sonority scale minimizes this confound as much as possible. In addition, for most of the languages there is more than one cluster position analyzed (e.g. onsets and medial clusters, which may not be phonologically independent). Finally, it is not entirely clear what a random or unstructured distribution of consonant clusters would be. In other words, what is the baseline against which a 76.5% sonority modulation rate in English coda clusters should be compared? These questions are only partially addressed within the scope of the present paper.

2.1.1. Findings: Cluster type frequency

Analysis just like that for English coda consonant clusters, using a table of A/P values for consonant cluster combinations in broad manner of articulation categories, was conducted for each language in the sample for each cluster position (onset, medial, coda) where data were available. Examining the full set of data across all positions and languages, modulation is obeyed in 72% of adjacent cell comparisons. This is based on 93 cluster data sets from the 47 languages sampled. In other words, there is an overall tendency toward sonority modulation across onset, medial, and coda CC clusters in the relative use of consonant cluster types across a variety of languages. There are quantitative violations of the use of consonant clusters according to sonority modulation within languages, but the overall pattern is robust and consistent with the previously examined case of similarity avoidance (Frisch *et al.* 2004).

With a tendency toward sonority modulation established, the analysis can move forward by considering the influence of sonority sequencing. Sonority modulation is in conflict with sonority sequencing in roughly half of the possible clusters. A sonority fall in an onset is good by sonority modulation but poor by sonority sequencing. Similarly, a sonority rise in a coda is good by sonority modulation but poor by sonority sequencing. Therefore, dividing the data into cases where the two possible functional explanations for sonority either agree or are in conflict can test which of these functional explanations provides a better model of the data. However, in line with other cases of functional explanation for phonological patterns, the interaction between these constraints appears to be cumulative. The interaction between sonority modulation and sonority sequencing across the cluster data sets is shown in table 3.

Table 3. Sonority Modulation and Sonority Sequencing.

		Constraints agree		Constraints disagree	
Position	N	Predicted	SD	Predicted	SD
Initial	38	0.76	0.15	0.69	0.16
Final	18	0.79	0.11	0.58	0.24
Medial*	37	0.71	0.16	0.71	0.18

Table 3 shows that adherence to both sonority modulation and sonority sequencing simultaneously is the most common pattern.

* There is no constraint conflict in medial clusters but the clusters are still split into agree/disagree groups for comparison purposes as though they are onsets (assuming an onset bias, Clements 1990).

When these two constraints are in conflict, the quantitative patterns are less robust. Considering quantitative patterns provides a more detailed understanding of the phonotactics of consonant clusters. What might ordinarily be seen as violations of a strictly defined sonority constraint based on either sequencing or modulation is instead seen as a systematic quantitative patterning that respects, for the most part, both constraints.

There is an additional potential impact on the expected quantitative pattern in consonant clusters described so far. This impact involves the overall use of clusters. If a language has a large number of consonant clusters then they cannot all be optimal with regard to sonority modulation or sonority sequencing. This is similar to the cross-linguistic patterning of vowel segments. Vowels generally distribute themselves evenly around the vowel space (Lindblom 1990), but when a language has a large number of vowels, the phonetic distance between vowels is smaller. This can be seen as a functional constraint against large vowel inventories and quantitatively this appears to be the case across languages. Returning to the case of consonant clusters, having an extremely large number of consonant clusters is likely to create clusters that violate sonority constraints. We can examine the density of clusters in the language (treated as an independent factor) as a functional pressure against obeying sonority.

In addition, considering consonant cluster density also aids in answering the question of the random or unstructured rate of consonant clusters that satisfy sonority constraints. The rate of adherence to sonority modulation patterns for random consonant combinations differs depending on the density of clusters in the language. This was determined by Monte Carlo simulations of an artificial language with densities of 0.1, 0.3, and 0.5 (10%, 30%, and 50%). Consonant cluster pairs were created by randomly distributing the actually occurring consonant clusters among the possible clusters from a relatively generic inventory of obstruents and sonorants. Not surprisingly, with higher cluster density, more violations of sonority are found.

The interaction between cluster density and sonority modulation across the typological data as well as in the Monte Carlo simulations is shown in figure 1. Each data point in the figure is the percent of pairwise comparisons between adjacent sonority cells that respect sonority modulation for a particular language and cluster position in the sample (as exemplified for English coda clusters in table 2). A linear regression shows a decreasing tendency to obey sonority modulation as cluster density increases. However, the mean across languages is well above that predicted by chance in the Monte Carlo simulations for the three different densities (in the vicinity of 50% and decreasing with increasing density). In addition to the density-based pattern in sonority modulation, there is an overall tendency to avoid large numbers of clusters in the languages sampled.

2.1.2. Summary: Cluster type frequency

More "gaps" are found in permissible sonority combinations when the sonority difference is small compared to when the sonority difference is large, showing a tendency to obey a sonority modulation constraint. This constraint is more robust when the sonority difference is also compatible with sonority sequencing and less robust when sonority sequencing is violated, showing that both sonority modulation and sonority sequencing influence the cross-linguistic and within language frequency of cluster types. Functional sonority constraints interact with a third factor, in that it is harder to obey sonority restrictions if the language uses many clusters. The sonority constraints are more robust in languages with sparse cluster use and more likely to be violated with dense cluster use. Overall, then, there

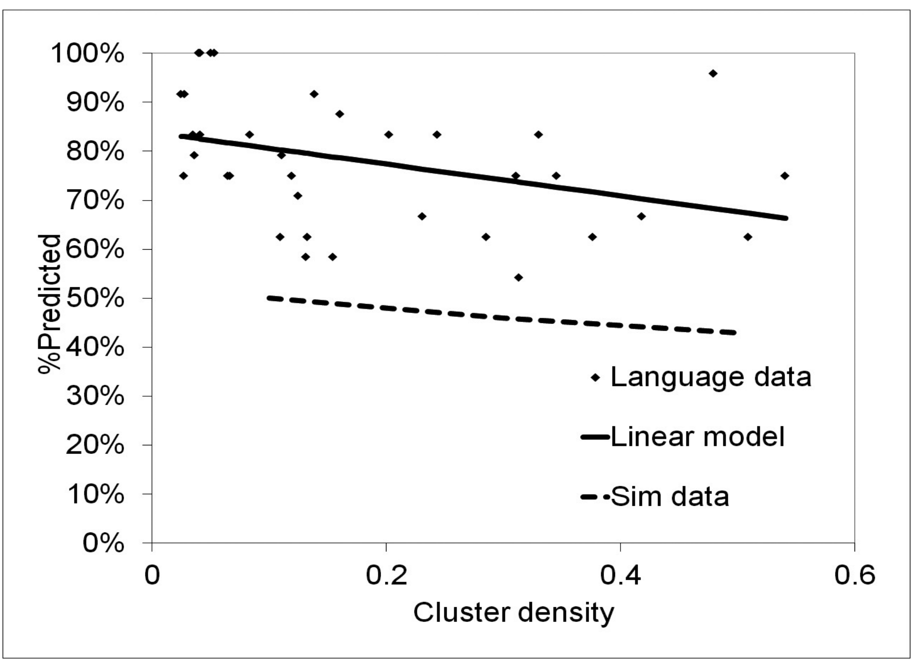

Figure 1. Clusters respecting sonority modulation by cluster density in the language.

are three factors influencing the cross-linguistic distribution of gaps in cluster inventories. Languages avoid large numbers of clusters, avoid clusters that violate sonority modulation, and avoid clusters that violate sonority sequencing. These three constraints interact cumulatively when examined quantitatively.

2.2. Study of cluster lexical frequency

If sonority constraints are functional, grounded in phonetic or psycholingustic processing principles, then we would predict the functional pressures that lead to an avoidance of sonority violations apply as individual words are used. By definition, functional constraints make the individual words more difficult to perceive or produce correctly. Thus, we would expect to find quantitative patterns related to sonority sequencing within the frequency distribution of forms in the lexicon (Frisch 2004). Quantitatively examining sonority patterns in the lexicon requires consideration of the frequency of occurrence of the consonants involved in the cluster. A cluster involving two infrequent consonants would be unlikely to occur regardless of sonority. A gap in clusters involving two frequent consonants is more likely to be systematic. For the present study, the O/E (observed divided by expected) measure of relative co-occurrence is used to account for baseline segment probability (Frisch *et al.* 2004).

An analysis of cluster frequency in the lexicon has been completed for two data sets, Spanish medial clusters and English coda clusters. The case of Spanish medial clusters will be covered in greater detail. Table 4 shows the frequency count of across lexical items for Spanish medial clusters of different sonority classes in the University of South Florida Spanish Frequency Lexicon (Brea-Spahn 2009; Frisch & Brea-Spahn 2010).

Table 4. Frequency count for Spanish medial clusters in the lexicon.

C2 C1	Stop	Fricative	Nasal	Liquid	Glide
Stop	299	385	70	1214	520
Fricative	1522	137	162	515	1852
Nasal	2575	931	43	16	328
Liquid	578	610	308	17	528
Glide	93	92	65	37	0

Given the frequency of different sonority classes of consonant in clusters, the randomly expected frequency of sonority combinations can be computed by multiplying the marginal probabilities for each type of C1 and C2. The observed count divided by the count expected by chance (O/E) provides a relative measure of co-occurrence given baseline frequency. Table 5 shows the O/E for the Spanish medial data. An O/E less than 1 means that fewer combinations than expected are found. An O/E greater than 1 means that more combinations than expected are found. Like with the analysis of consonant cluster types, we would predict relative frequency to increase away from the main diagonal of the table. Comparing adjacent cells as before, the percent of comparisons that respect sonority modulation can be computed to provide a single number that summarizes the degree to which quantitative patterns obey the hypothesized constraints.

Table 5. O/E for Spanish medial clusters in the lexicon.

C2 C1	Stop	Fricative	Nasal	Liquid	Glide
Stop	0.24	0.77	0.47	1.97	0.69
Fricative	0.94	0.21	0.82	0.63	1.87
Nasal	1.74	1.54	0.24	0.02	0.36
Liquid	0.80	1.76	2.94	0.04	1.01
Glide	0.86	2.04	4.87	0.68	0.00

A similar analysis was conducted for English final CC clusters in the Hoosier Mental Lexicon database of English words (Nusbaum *et al.* 1984). As with the analysis of cluster type frequency, comparisons between two empty cells were not used in the computation.

2.2.1. Findings: Cluster lexical frequency

In the Spanish medial cluster data, 68% of comparisons obey sonority modulation for O/E across the Spanish lexicon. In the English final cluster data, 71% of comparisons obey sonority modulation for O/E across the English lexicon. These findings are similar to the findings for cluster type frequency. This pattern is consistent with a soft or gradient constraint against poor sonority modulation clusters acting on the lexicon of the language, influencing the distribution of consonant clusters in lexical items. Finding the general typological pattern for sonority within the quantitative distribution of lexical items within a language provides strong evidence for the functional nature of the constraints involved.

2.3. Limitations of the present study

The current study is typologically limited to a convenience sample of languages. For the most part, this is a subset of languages from Greenberg (1978). A more extensive study to verify the claims of this paper is needed with a larger and typologically balanced sample of languages. In addition, the present study examined only two cases of patterns within a lexicon. To clearly establish the quantitative nature of sonority constraints in consonant clusters, additional lexical sources will have to be investigated.

The present study is also based on a relatively standard division of manner classes for consonants into a sonority hierarchy. However, this division has been claimed to have no logical basis by Basbøll (1994), who advocates a more generic sonority hierarchy with only three distinctions. Given that the quantitative patterns observed here work reasonably well and are relatively robust, some variation in the definition of the sonority hierarchy will not affect the overall result. This is one advantage of taking a quantitative approach. Minor contradictory details have little impact on the overall pattern.

3. Conclusion

The strongest possible test of functional constraints on language processing would find both quantitative and qualitative patterns in language that reflect functional forces. If a typological pattern is functionally-based, then the cross-linguistic pattern should also be visible within individual languages quantitatively (Frisch 2004). In the present study, articulatory and perceptual motivations for the sonority hierarchy were considered and patterns within languages' cluster inventories and lexicons were examined. Parallel with the previously examined Obligatory Contour Principle for Place of Articulation (OCP-Place) as a functional constraint on similarity avoidance, quantitative patterns were found within languages that reflect typological generalizations in sonority modulation and sonority sequencing. In this preliminary analysis, the hypothesis is supported: Quantitative sonority constraints on consonant clusters are found reflecting potential functional motivations for sonority in both production and perception. The prediction is that functionally-motivated phonotactic patterns in sonority constraints should be found, at least statistically, everywhere we look.

Address of the Author

Stefan Frisch, University of South Florida, Dept. of Communication Sciences and Disorders, 4202 E. Fowler Ave, PCD1017, Tampa, FL 33620-8100 <sfrisch@usf.edu>

Bibliographical References

Alderete John, Tupper Paul & Stefan A. Frisch 2013. Phonological constraint induction in a connectionist network: Learning OCP-Place constraints from data. *Language Sciences* 37. 52-69.

Anttila Arto. 2007. Variation and optionality. In de Lacy Paul (ed.), *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press. 519-536.

Bailey Todd & Ulrike Hahn 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 4. 568-591.

Basbøll Hans 1994. How to derive the sonority syllable from the prototypical peak. *Acta Linguistica Hafniensia* 27. 51-65.

Berent Iris & Joseph Shimron 1997. The representation of Hebrew words: Evidence from the obligatory contour principle. *Cognition* 64. 39-72.

Berg Thomas & Hassam Abd-El-Jawad 1996. The unfolding of suprasegmental representations: A cross-linguistic perspective. *Journal of Linguistics* 32. 291-324.

Betancourt Kyna S. 2013. Nonword processing in bilingual five year olds: Do phonotactics count? Unpublished Ph.D. dissertation, University of South Florida.

Blevins Juliette 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins Juliette 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32. 117-166,

Brea-Spahn María R. 2009. Spanish-specific patterns, language proficiency, and nonword repetition performance in English language learners. Unpublished Ph.D. dissertation, University of South Florida.

Bybee Joan 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10. 425-455.

Bybee Joan 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14. 261-290.

Bybee Joan 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Clements George N. 1990. The role of the sonority cycle in core syllabification. In Kingston John & Mary Beckman (eds.), *Laboratory phonology I: Between the grammar and the physics of speech*. New York: Cambridge University Press. 283-333.

Coetzee Andries & Joe Pater 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26. 289-337.

Frisch Stefan A. 1996. Similarity and frequency in phonology. Unpublished Ph.D. dissertation, Northwestern University.

Frisch Stefan A. 2000. Temporally organized lexical representations as phonological units. In Broe Michael B. & Pierrehumbert Janet B. (eds.), *Acquisition and the lexicon: Papers in laboratory phonology V*. Cambridge: Cambridge University Press. 283-298.

Frisch Stefan A. 2004. Language processing and OCP effects. In Hayes Bruce, Kirchner Robert & Donca Steriade (eds.), *Phonetically-based phonology*. Cambridge: Cambridge University Press. 346-371.

Frisch Stefan A. & María R. Brea-Spahn 2010. Metalinguistic judgments of phonotactics by monolinguals and bilinguals. *Laboratory Phonology* 1. 345-360.

Frisch Stefan A., Pierrehumbert Janet B. & Michael B. Broe 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22. 179-228.

Frisch Stefan A. & Bushra Zawaydeh 2001. The psychological reality of OCP-Place in Arabic. *Language* 77. 91-106.

Greenberg Joseph H. 1950. The patterning of root morphemes in Semitic. *Word* 5, 162-181.

Greenberg Joseph H. 1978. In Greenberg Joseph H., Ferguson Charles A. & Edith A. Moravcsik (eds.), *Universals of human language: Phonology.* Stanford, Stanford University Press. 243-279.

Hay Jennifer, Pierrehumbert Janet B. & Mary Beckman 2003. Speech perception,wellformedness, and the statistics of the lexicon. In Local John, Ogden Richard & Rosalind Temple (eds.), *Papers in laboratory phonology VI: Phonetic interpretation*. Cambridge: Cambridge University Press. 58-74.

Hayes Bruce, Kirchner Robert & Donca Steriade 2004. *Phonetically-based phonology*. Cambridge: Cambridge University Press.

Hayes Bruce & Colin Wilson 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379-440.

Hayes Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Londe 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85. 822-863.

Lindblom Björn 1983. Economy of speech gestures. In MacNeilage Peter (ed.), *The production of speech*. New York: Springer-Verlag. 217-246.

Lindblom Björn 1990. Explaining phonetic variation: A sketch of the H & H theory. In Hardcastle William & Alain Marchal (eds.), *Speech production and speech modelling*. Dordrecht: Kluwer. 403-439.

Luce Paul. A. & David B. Pisoni 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19. 1-36.

MacEachern Margaret R. 1999. *Laryngeal cooccurrence restrictions*. New York: Garland.

McCarthy John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Keating Patricia (ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form*. Cambridge: Cambridge University Press. 191-283.

MacNeilage Peter F. & L. Davis Barabara 2000. On the origin of internal structure of word forms. *Science* 288. 527-531.

Messer Marielle H., Leseman Paul P.M., Boom Jan & Aziza Mayo. Y. 2010. Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology* 105. 306-323.

Munson Benjamin 2001. Phonological pattern frequency and speech production in children and adults. *Journal of Speech, Language, and Hearing Research* 44. 778-792.

Munson Benjamin, Edwards Jan and Mary E. Beckman 2011. Phonological representations in language acquisition: Climbing the ladder of abstraction. In Cohn Abigail C., Fougeron Cécile &Marie K. Huffman (eds.), *The Oxford handbook of laboratory phonology*. Oxford: Oxford University Press. 288-309.

Nusbaum Howard, Pisoni David & Christopher. K. Davis 1984. Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report No. 10* (Speech Research Laboratory, Bloomington IN). 357-376.

Ohala John J. & Haruko Kawasaki-Fukumori 1997. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In Eliasson Stih & Jahr Ernst Hakon (eds.), *Language and its ecology: Essays in memory of Einar Haugen*. Berlin: Mouton de Gruyter. 342-365.

Pierrehumbert Janet 1994. Syllable structure and word structure: A study of triconsonantal clusters in English. In Keating Patricia (ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form*. Cambridge: Cambridge University Press. 168-188.

Pierrehumbert Janet (2003) Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46. 115-154.

Prince Alan & Paul Smolensky 2004. *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell Publishers.

Redford Melissa A. 2000. *An articulatory basis for the syllable*. Unpublished Ph.D. dissertation, University of Texas at Austin.

Saffran Jenny R., Aslin Richard N. & Elissa L. Newport 1998. Statistical Learning by 8-month-old Infants. *Science* 274. 1926-1928.

Sebastián-Gallés Nuria, Echeverria Sagrario & Laura Bosch 2005. The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language* 52. 240-255.

Treiman Rebecca, Kessler Brett, Knewasser Stephanie, Tincoff Ruth & Margo Bowman 2000. English speakers' sensitivity to phonotactic patterns. In Broe, Michael & Pierrehumbert, Janet B. (eds.), *Papers in laboratory phonology V: Language acquisition and the lexicon*. Cambridge: Cambridge University Press. 269-282.

Vitevitch Michael S., Luce Paul. A., Pisoni David B. & Edward T. Auer 1999. Phonotactics, neighborhood activation and lexical access for spoken words, *Brain and Language* 68. 306-311.

Zuraw Kie 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28. 417-472.

Appendix: Sources

IJAL = *International Journal of American Linguistics*

HAIL = *Handbook of American Indian Languages*

Abramson A. (1962). *The vowels and tones of Standard Thai*. Bloomington, IN: Indiana University Press.

Aschmann H. P. (1946). Totonaco phonemes. *IJAL*, 12, 34-43.

Andrews H. (1949). Phonemes and morphophonemes of Temoayan Otomi. *IJAL*, 15, 213-222.

Berry K. & Berry C. (1999). *A description of Abun: A West Papuan language of Irian Jaya*. (Pacific Linguistics, Series B, 115.) Canberra: Australian National University.

Boas F. & Deloria E. (1976). *Dakota grammar*. New York: AMS press.

Bogoras W. (1922). Chukchee. In Boas F. (ed.), *HAIL 2*. Washington, DC: Government Printing Office. 631-903.

Callister B. (1993). Misiman phonology. In J.M. Clifton (ed.), *Phonologies of Austronesian Languages, No. 2*. SIL, Papua New Guinea.

Camara Jr J. M. & Mattoso J. (1979). *The Portuguese language*. Chicago: University of Chicago Press.

Chitoran I. (1994). Acoustic investigation of Georgian harmonic clusters. *Working papers of the Cornell phonetics laboratory*, 9, 27-65.

Deprez V. (1988). Georgian complex segments. *Proceedings of NELS*, 16, 109-123.

Fast P.W. (1953). Amuesha (Arawak) phonemes. *IJAL*, 19, 191-194.

Ferguson C.A. & Chowdhury M. (1960). The phonemes of Bengali. *Language*, 36, 22-59.

Fractenberg L.J. (1922). In Boas F. (ed.), *HAIL 2*. Washington, DC: Government Printing Office. 297-429.

Garvin P.L. (1950). Wichita I: Phonemics. *IJAL*, 16, 179-184.

Id. (1948). Kutenai I: Phonemics. *IJAL*, 14, 37-42.

Gibson L.F. (1956). Pame (Otomi) phonemics and morphophonemics. *IJAL*, 22, 242-265.

Harden M. (1946). Syllable structure of Terena. *IJAL*, 12, 60-63.

Jaeger J.J. & Van Valin R.D. (1982). Initial consonant clusters in Yatee Zapotec. *IJAL*, 48, 125-138.

Kohler K. (1999). German. In *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press. 86-89.

Kordić Snježana (1997). Serbo-Croatian (Languages of the world/materials, 148.) Munich: Lincom Europa.

Koutsoudas A. (1962). *Verb morphology of Modern Greek*. Bloomington, IN: Indiana University Press.

Lisker L. (1963). *Introduction to spoken Telugu*. New York: American Council of Learned Societies.

Marlett S.A. (1988). The syllable structure of Seri. *IJAL*, 245-278.

Matthews G.H. (1955). A phonemic analysis of a Dakota dialect. *IJAL*, 21, 56-59.

McArthur H. & McArthur L. (1956). Aguacatec (Mayan) phonemes in the stress group. *IJAL,* 22, 72-76.

McIntosh J.B. (1945). Huichol phonemes. *IJAL*, 11, 31-35.
McKaughan H.P. (1954). Chatino formulas and phonemes. *IJAL*, 20, 23-27.
Morelli F. (1999). The phonotactics and phonology of obstruent clusters in Optimality Theory. Unpublished Ph.D. dissertation, University of Maryland.
Morgan T.A. (1988). Positive and negative syllable structure conditions in Spanish. In T.A. Morgan, J.F. Lee & B. VanPatten (eds.), *Language and language use: Studies in Spanish*. New York: University Press of America. 37-54.
Nakaskul K. (1978). The syllabic and morphological structure of Cambodian words. In P. Jenner (ed.), *Mon-Khmer studies VII*. 183-200.
Nater H. F. (1984). The Bella Coola language (Canadian enthnology service paper no. 92). Ottawa: National Museums of Canada.
Needham D. & Davis M. (1946). Cuicateco phonology. *IJAL*, 20, 23-27.
Newmark L. (1957). *A structural grammar of Albanian*. Bloomington, IN: Indiana University Press.
Penzl H. (1955). *A grammar of Pashto*. Washington: American Council of Learned Societies.
Pike K.L. & Pike E.V. (1947). Immediate constituents of Mazateco syllables. *IJAL*, 12, 78-91.
Powell J.V. & Woodruff Sr F. (1976). *Quileute Dictionary*. Moscow, ID: University of Idaho Press.
Prick van Wely F.P.H. (1951). *Casssell's English-Dutch, Dutch-English dictionary*. London: Cassell.
Rabel L. (1961). *Khasi, a language of Assam*. Baton Rouge: Louisiana State University Press.
Rath D.D. (1993). Mengen phonology essentials. In J. M. Clifton (ed.), *Phonologies of Austronesian Languages, No. 2*. SIL, Papua New Guinea.
Ruhlen M. (1973). *Rumanian phonology*. Stanford: Stanford University Press.
Reichard G.A. (1933). Coeur d'Alene. In Boaz, F. (ed.), *HAIL 3*. Washington, DC: Smithsonian Institute. 515-707.
Robbins F.E. (1961). Quiotepec Chinantec syllable patterning. *IJAL*, 27, 237-250.
Salzmann Z. (1956). Arapaho I: Phonology. *IJAL*, 22, 49-56.
Sapir E. (1922). The Takelma language of Southwestern Oregon. In Boas, Franz (ed.), *HAIL* 2, 1-296. Washington, DC: Government Printing Office.
Siptar P. & Torkenczy M. (2000). *The phonology of Hungarian*. Oxford: Oxford University Press.
Skinner L.E. (1962). Usila Chinantec syllable structure. *IJAL*, 28, 251-255.
Spencer R.F. (1946). The phonemes of Keresan. *IJAL*, 12, 229-236.
Spotts H. (1953). Vowel harmony and consonant sequences in Mazahua (Otomi). *IJAL*, 19, 253-258.
Steriade D. (1982). Greek prosodies and the nature of syllabification. Unpublished Ph.D. dissertation, MIT.
Tarpent M.L. (1989). A grammar of the Nisgha language. Unpublished Ph.D. dissertation, University of Victoria.
Trager F.H. (1971). The phonology of Picuris. *IJAL*, 37, 29-33.

Vanvik A. (19720. A phonetic-phonemic analysis of Standard Eastern Norwegian. *Norwegian Journal of Linguistics*, 26, 119-164.
Waterhouse V. & Morrison M. (1950). Chontal phonemes. *IJAL*, 16, 35-39.
Whitman W. (1947). Descriptive grammar of Ioway-Oto. *IJAL*, 13, 233-248.
Wolff H. (1952). Osage I: Phonemes and historical phonology. *IJAL*, 18, 63-68.
Wonderly W.L. (1951). Zoque II: Phonemes and morphophonemes. *IJAL*, 17, 105-123.
Wright R. (1996). Consonant clusters and cue preservation in Tsou. Unpublished Ph.D. dissertation, UCLA.
Zhukova A.N. (1967). Russian-Koryak dictionary. Moscow: Izdat, Sovetskaja encilopedija.