sCIP: The s-Cluster Inventory Project

S-clusters present an interesting debate with regards to their structure (Boyd, 2006; Davis, 1990; Goad & Rose, 2004; Selkirk, 1982; Wiese, 1996) and markedness (Clements, 1990; Major, 1996). Although s-clusters have been studied in detail in L1 and L2 acquisition (Barlow, 2001; Major, 1996), a larger typological study of clusters has not been completed. The Sonority Sequencing Principle (SSP; Selkirk, 1984) and the Minimal Sonority Distance Principle (MSD; Clements, 1990) would both predict that s+glide clusters are the least marked type of s-cluster, followed by s+liquid and s+glide clusters, while s+obstruent clusters would be the most marked. This markedness hierarchy is based on each cluster type's adherence to the SSP (or violation in the case of s+obstruent clusters) and degree of sonority difference. Typologically, the predictions of the SSP and MSD would suggest that languages with the more marked s-cluster types would also have the less marked s-cluster types in their inventories. By developing a database of s-cluster typology cross-linguistically, we can look at the co-occurrence and implicational relationships of s-clusters and further investigate the structure and markedness of these clusters.

The sCIP database is a relational Microsoft Access database that will soon be freely available online (Figure 1). It provides the singleton consonant inventories, word-initial cluster inventories, and a breakdown word-initial cluster types by place and manner of articulation for both s-clusters and non-s-clusters in 231 languages selected from the UPSID database (Maddieson, 1984). The sCIP database is intended for the purpose of creating a typology of s-cluster inventories, but researchers interested in other types of initial clusters may also find it useful. The database is searchable, and users can, for example, compile a list of languages that contain a certain type of cluster, or examine how likely two cluster types are to co-occur in the same language.

			General Langu	age Information					
	Language Name Language Number Language Family		Achumawi	RefID	Ole	Olmsted64 🔍			
			744	Input Source		Phonology Lab			
			Hokan	Input by MW	Checked HG				
	Macro	area	North America 🖂	General Cluster Information					
	Notes	Cluster inve in text; glea dictionary; each of [km	ntory not discussed ined from only one example n kj sn xl]; while only	No Initial Clusters s-Clusters Non s-Clusters	 Initial Clusters sh sj sk sl sm sn sp sq st sw xl 				
NGLETONS									
Stops		Fricatives		Nasals		Approximants		5	
Labial Stop		Labial Fricative		Labial Nasal		Late	eral		
Dental/Alveolar	Stop	Coronal	Nonsibilant Fricative	Dental/Alveolar	Nasa	I Rho	rtic/Flap/	Тар	
Retroflex Stop		Retrofle	x Fricative	Retroflex Nasal		Glid	le		
Palatal Stop		Coronal	Sibilant Fricative	Palatal Nasal		Affric	ates		
Velar Stop		🗌 Velar Fri	cative	🗌 Velar Nasal		Affr	ricate		
Uvular Stop		Uvular F	ricative	Other Singletons					
		Glottal F	ricative						
Glottal Stop									
-Clusters				Non s-Cluster	s				
-Clusters S+obstruent	⊡ S+la	bial [2	Non s-Cluster Stop+Fricative	s e [Nasal+I	Liquid		
-Clusters S+obstruent S+nasal	⊙ S+la ⊙ S+ci	ibial [3	Non s-Cluster Stop+Fricative Stop+Nasal	s e [Nasal+I	Liquid Glide		
-Clusters S+obstruent S+nasal S+liquid	⊙ S+la ⊙ S+ci ⊙ S+d	ibial [pronal [prsal [2 2 2	Non s-Cluster Stop+Fricative Stop+Nasal Stop+Liquid	e [Nasal+I Nasal+I Liquid+	Liquid Glide Glide		
-Clusters S+obstruent S+nasal S+liquid S+glide	⊙ S+la ⊙ S+o ⊙ S+d ⊙ The	ibial [pronal [orsal [majority of s	2 2 2	Non s-Cluster Stop+Fricative Stop+Nasal Stop+Liquid Stop+Glide	e [Nasal+I Nasal+I Liquid+ Nasal II	Liquid Glide Glide nitial		
-Clusters S+obstruent S+nasal S+liquid S+glide	S S+la S S+or S S+d The clus	ibial [oronal [majority of s ters have soli	2 2 2	Non s-Cluster Stop+Fricative Stop+Nasal Stop+Liquid Stop+Glide Fricative+Nas	e [[al [Nasal+1 Nasal+1 Liquid+ Nasal II	Liquid Glide Glide nitial te Initial		
Glusters S+obstruent S+nasal S+liquid S+glide s-Cluster Notes	S S+la S S+o S S+d The clus evid	ibial [oronal [orsal [majority of s ters have soli ence.	2 2 2	Non s-Cluster Stop+Fricative Stop+Nasal Stop+Liquid Stop+Glide Fricative+Nas Fricative+Liqu	e [[[al [Nasal+1 Nasal+1 Liquid+ Nasal II Affricat Three E	Liquid Glide Glide nitial te Initial Element		

Figure 1. Sample screen from the sCIP database

Of the 231 languages in the database, 62 contain at least one type of initial s-cluster. The general trend predicted by MSD holds: s+glide clusters occur more frequently than s+liquid clusters, which in turn occur more commonly than s+nasal clusters (Figure 2). However, an even larger number of languages include s+obstruent clusters which violate SSP, and can violate language specific MSD restrictions. Focusing specifically on the types of clusters that co-occur within a

language, we find that 30 languages have s-cluster inventories that run contrary to the implicational relationships predicted by the SSP and MSD. These 30 languages include marked-leaning inventories that contain marked but not unmarked structures and gapped inventories (de Lacy, 2002) that contain very marked and very unmarked s-clusters, but not s-clusters of intermediary markedness. These initial results suggest that the marked status of s-clusters cannot be based on SSP and MSD alone.



References

- Barlow, J. (2001). The structure of /s/-sequences: Evidence from a disordered system. *JCL* 28: 291-324.
- Boyd, J. (2006). On the representational status of s-clusters *San Diego Linguistics Papers 2*: 39-84.
- Clements, G.N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology 1* (pp. 283-333). Cambridge, MA: Cambridge University Press.
- Davis, S. (1990). Italian onset structure and the distribution of *il* and *lo. Linguistics* 28: 43-55.
- Goad, H. & Rose, Y. (2004). Input elaboration, head faithfulness and evidence for representation in the acquisition of left-edge clusters in West Germanic. In R. Kager, J. Pater & W. Zonneveld (Eds.), *Constraints in Phonological Acquisition* (pp. 109-157). Cambridge, MA: Cambridge University Press.
- Maddieson, I. (1984). Patterns of Sounds. Cambridge, MA, Cambridge University Press.
- Major, R. (1996). Markedness in second language acquisition of consonant clusters. In R. Bayley & D.R. Preston, eds., *Variation and Second Language Acquisition* (pp. 75-96). Amsterdam: Benjamins.
- Selkirk, E.O. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *The Structure of Phonological Representations* (pp. 337-383). Dordrecht, Netherlands: Foris.
- Selkirk, E.O. (1984). On the major class features and syllable theory. In M. Aronoff & R. Oerhle (Eds.), *Language Sound Structure*, (pp. 107-136). Cambridge, MA: MIT Press.