

Introducing the Cross-Linguistic S-Cluster Inventory Database
Marissa Woods, Ashley Farris-Trimble, & Danica Reid
Simon Fraser University
corresponding: afarrist@sfu.ca

S-clusters have long inspired debate about their structure (Boyd 2006; Davis 1990; Goad & Rose 2004; Selkirk 1982; Wiese 1996) and markedness (Clements 1990; Major 1996). Though s-clusters have been studied a great deal in both L1 and L2 acquisition (e.g., Barlow 2001; Major 1996), a larger typology of s-clusters in fully developed languages has not been completed. Morelli's (1998) typological study of obstruent+obstruent clusters concluded that fricative+stop clusters are the least marked type but did not discuss different types of fricative+stop clusters, nor fricative+sonorant clusters. A larger study of the cross-linguistic typology of s-clusters could thus help answer questions about their structure and markedness. For instance, a typological study would allow us to characterize co-occurrences and implicational relationships among s-clusters and to describe the possible s-cluster inventories that exist among languages. The Sonority Sequencing Principle (SSP; Selkirk 1984) and the Minimal Sonority Distance Principle (MSD; Clements 1990) would predict that s+glide clusters are the least marked type of s-cluster (as they obey the SSP and have a large sonority difference), followed by s+liquid and s+nasal clusters, while s+obstruent clusters (which violate both principles) are the most marked. Typologically, we would predict that languages that have more marked s-clusters in their inventories also have less-marked s-clusters. Does this prediction pan out?

The current paper introduces the Cross-Linguistic S-Cluster Inventory Database (CLiSCID) with the goal of furthering exploration of typological relationships among s-clusters, their structure, and their markedness status. The CLiSCID is a Microsoft Access database that will soon be freely available online. It contains data from 231 languages selected from the UPSID database (Maddieson 1984), which documents singleton phonological segment inventories in a representative sample of 317 languages; we have compiled information on initial clusters for 231 of these languages. For each language, the CLiSCID contains information about the language's singleton inventory as described by Maddieson (1984), all clusters that appear word-initially, and a breakdown of the types of s-clusters and non-s-clusters that occur by place and/or manner of articulation (see Figure 1). Though s-clusters are the focus of the database, researchers interested in other types of clusters will find it useful as well. The database is searchable, and users can compile, for instance, lists of languages that contain a certain type of cluster, or can examine how likely two cluster types are to co-occur in the same language.

We present the database, our methods for populating it, and some possible uses. We also present preliminary results of an examination of s-cluster inventories by manner of articulation. 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 the SSP. Focusing specifically on the types of clusters that can 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 structures but not unmarked structures) and gapped inventories (de Lacy 2002) that contain very marked and very unmarked structures, but not structures of intermediate markedness. These results suggest that the marked status of s-clusters cannot be based on the SSP and MSD alone.

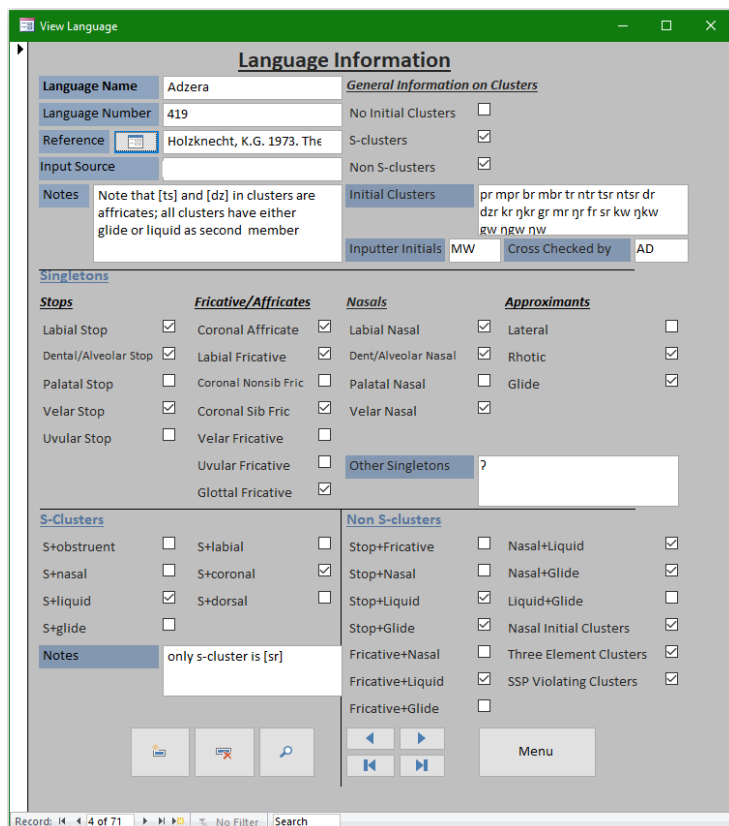


Figure 1: Sample screen from the CLiSCID

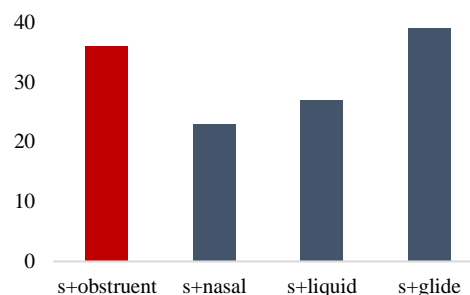


Figure 2: Occurrence in number of CLiSCID languages of different of s-cluster types

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 Linguistic 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*, 283-333. Cambridge, MA, CUP.
- Davis, S. 1990. Italian onset structure and the distribution of *il* and *lo*. *Linguistics* 28: 43-55.
- de Lacy, P. 2002. The formal expression of markedness, University of Massachusetts, Amherst.
- 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*, 109-57. Cambridge, MA, CUP.
- Maddieson, I. 1984. *Patterns of Sounds*. Cambridge, 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*, 75-96. Amsterdam: Benjamins.
- Morelli, F. 1998. Markedness relations and implicational universals in the typology of onset obstruent clusters. *Proceedings of NELS 28*, 107-20. UMass Amherst, GSLA.
- Selkirk, E.O. 1982. The syllable. In H. van der Hulst & N. Smith, eds., *The Structure of Phonological Representations*, 337-83. Dordrecht, Netherlands, Foris.
- Selkirk, E.O. 1984. On the major class features and syllable theory. In M. Aronoff & R. Oerhle, eds., *Language Sound Structure*, 107-136. Cambridge, MA: MIT Press.
- Wiese, R. 1996. *The Phonology of German*. Oxford, Clarendon Press.