Typological and Historical relations across sign languages
The view from articulatory features

Natasha Abner[1], Carlo Geraci[2,3,4], Jessica Lettieri[5], and Shi Yu[2,3,4,6]
[1] University of Michigan, [2] Institut Jean Nicod, [3] CNRS, [4] École Normal Supérieure, [5] University of Torino, [6] EHESS

Goals. The aim of this paper is twofold. First, we use articulatory (phonetic) features to describe the properties of sign language (SL) lexical items; second, we use these features to classify SLs into macro- and micro-families. The paper provides proof of concept that quantitative methods can be used to probe typological and historical classifications of SLs, along the lines of what has been done recently in spoken language phylogentics (Nichols 1992, Dunn et al. 2005) and the genetis of speech communities (Verdu et al. 2017).

Background. SLs are natural human languages that are perceived visually (vs. acoustically) and produced gesturally (vs. vocally). The perception-production systems of SL give rise to two macroscopic modality effects. One, SLs allow simultaneous production of a significant amount of contrastive phonemic material (Brentari 1998). Two, SLs display a degree of iconicity at the lexical level, understood here as a conceptually motivated mapping between sign form and sign meaning (Taub 2001 i.a.). Thus, SLs may exhibit a higher degree of cross-linguistically similarity (Wilbur 2010) and signers without a shared language may experience relative ease in converging on a shared communication system (Zeshan 2015). Nevertheless, SLs share many of the structural and functional phenomena of spoken languages. SLs may be classified into language families according to their historical relationships (Anderson & Peterson 1979, Wittman 1991), though additional reliable and verified documentation remains necessary. SLs may also be grouped typologically according to their linguistic properties (Brentari et al. 2015, Zeshan 2006). For example, pairwise comparisons of SLs based on global resemblance of handshape, movement, location and hand-orientation showed that it is possible to detect the degree of similarity/distance between SLs (Woodward 2000; McKee and Kennedy 2000). Here, we assess the efficacy of established statistical models in the classification of SLs based on linguistic features. Because typology and history exhibit patterns of convergence and divergence, we also evaluate the groupings statistically inferred from articulatory features relative to what is known about historical relatedness among SLs.

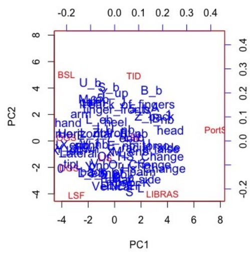
Methodology. To have the same baseline for cross-linguistic comparison, we used Woodward's SL adaptation of the Swadesh list. Following lexicostatistics practice (Rea 1990), this list identifies 100 items that represent some of the core concepts of human life/experience (e.g., *mother, live, fire,* etc.); the SL adaptation of this list removes items that are similar across sign languages (e.g., naming body parts by pointing) and likely to give rise to artificially inflated measures of similarity. Data from nine SLs (Austrian OGS, Brazilian LIBRAS, British BSL, French LSF, German DGS, Italian LIS, Portuguese PortSL, Spanish LSE, Turkish TİD) were sourced from an on-line dictionary (www.spreadthesign.com). Articulatory features were manually coded for items on the SL-adapted list for all nine languages (all items were not available for all languages). The set of articulatory features coded (55 handshapes, 36 locations, 11 movements, 8 hand-orientations) were modeled after Brentari (1998) but are common across SL phonological models (Sandler and Lillo-Martin 2006, Van der Kooij 2002). Coding was done using ad-hoc web-based software for annotation (Yu and Geraci 2018). Signs were also coded as one- or two-handed and as compound or non-compound signs. Figure (1) offers a partial overview of the dataset.

Analyses. Two analyses have been performed on the data: a divisive cluster analysis to identify the language groups in the data (Paradis 2006) and a principal component analysis (PCA) to identify the articulatory features driving these groupings (Baayan 1994). Figure (4) shows the results of the cluster analysis. The three main clusters group together LIBRAS, LIS, LSE and PortSL on one side, LSF, DGS and OGS on another side and BSL and TİD on the third cluster. The PCA identifies seven significant clusters (Figure 2). The PCA results indicate that BSL and TİD are clustered together due to bent base nodes handshapes (e.g., U_b =  and B_b = ). LIBRAS, LIS and LSE are clustered together due to distal movements (handshape and orientation change); while DGS and OGS are clustered together due to a high proportion of one-handed signs and very few signs produced on the horizontal and lateral planes. Figure (3) shows the plot of the first two PCs (which account for 38% of variance).

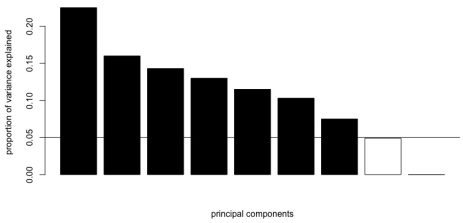
1. FIGURE: Overview of the annotated files

english	handshape	twohands	handpart	TwoH	PoA	PoA2	Mov	Mov2
all	5	true	palm	1	Lateral	Lateral	Y_down	Y_down
animal	5	true	palm	1	Horizontal	Horizontal	Y_down	Y_down
bad	I	false	ulnar_side	0	Vertical	Vertical	Z_front	Z_front
because	1	false	radial_side	0	Torso_top	torso	X_left	X_left
bird	L	false	heel	0	mouth	head	true false	Hs_Change
black	S_b	false	finger_fronts	0	cheek	head	Y_down	Y_down
blood	1	false	tip	0	chin	head	Z_back	Z_back
boring	E	false	finger_fronts	0	chin	head	Z_back	Z_back
brother	S_b	true	heel	1	Lateral	Lateral	Z_front	Z_front
bug	1	false	UNDEF	0	back_of_fingers	hand	Y_up	Y_up
cat	5	true	tip	1	mouth	head	X_right	X_right
child	E	false	palm	0	Horizontal	Horizontal	Y_down	Y_down

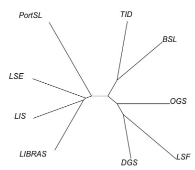
3. FIGURE: Plot of PC 1 and 2



2. FIGURE: Barplot of the PC analysis



4. FIGURE: Diagram of relatedness



Discussion. Documentation of historical relations among SLs (Ethnologue Languages of the World) would place LIS, LSE, LSF as part of the same linguistic group with LIBRAS also influenced by old LSF. DGS and OGS should be closely related, while no relation is documented between BSL and TİD. Finally, PortSL does not have historical relations with any of the languages in the sample. As history predicts, statistical analysis of articulatory features isolates PortSL while LIBRAS, LIS and LSE are clustered together; the clustering of DGS and OGS with LSF is also relatively expected given their historical relation. The clustering of BSL and TİD, however, is unexpected given historical documentation. We suggest that this is due largely to the fact that BSL and TİD are the only two languages in the sample with a two-hand manual alphabet, a property that may have heretofore underestimated influence on how the articulatory properties of the lexicon are structured. Thus, as in spoken languages, (quantitative) feature-based classification of SLs is possible and may reflect historical relations, but only partly.