Evolutionary dynamics in the dispersal of sign languages

Justin M. Power¹*, Guido W. Grimm² and Johann-Mattis List³

¹ Department of Linguistics, University of Texas at Austin, USA
² Independent researcher, Orléans, France
³DLCE, Max Planck Institute for the Science of Human History, Jena, Germany

*corresponding author: justin(power@utexas.edu

June 2019

While the evolution of spoken languages is well understood and has been studied using traditional historical comparative methods as well as newer computational phylogenetic methods, evolutionary processes resulting in the diversity of contemporary sign languages are poorly understood, and scholars have been largely unsuccessful in grouping sign languages into monophyletic language families. To date, no published studies have attempted to use language data to infer relationships amongst sign languages on a large scale. Here, we report the results of a phylogenetic analysis of 40 contemporary and 36 historical sign language manual alphabets coded for morphological similarity. Our results support grouping sign languages in the sample into six main European lineages, with three larger groups of Austrian, British, and French origin, as well as three smaller groups centring around Russian, Spanish, and Swedish. The British and Swedish lineages support current knowledge of relationships amongst sign languages based on extra-linguistic historical sources. With respect to other lineages, our results diverge from current hypotheses by indicating (i) independent evolution of Austrian, French, and Spanish from Spanish sources; (ii) an internal Danish subgroup within the Austrian lineage; and (iii) evolution of Russian from Austrian sources.

Keywords: sign language, language phylogeny, language evolution, phylogenetic networks

1 Introduction

Linguistic analyses of the world’s sign languages (SLs) over the past 60 years have shown that the human capacity for language is not limited to the oral-aural modality. Instead, *homo symbolicus* has developed complex natural language in the gestural-visual modality as well, particularly in deaf signing communities throughout the world. The development of educational institutions for the deaf, which began during the Enlightenment in Europe, especially in the late 18th and early 19th centuries, contributed to the formation of these deaf signing communities and the emergence of widespread, conventional SLs [1, 2]. The languages that emerged in these newly formed communities were soon dispersed to other parts of Europe and beyond. The success of the first public school for the deaf, the *Institut National de Jeunes Sourds de Paris*, founded...
between 1759–71, attracted educators from across Europe and the Americas, who came to the Paris Institute to learn pedagogical methods, with the goal of establishing schools for the deaf in their countries of origin. Deaf students from other countries also came to Paris, graduated, and returned home to work as teachers or to found new schools. Thus, within a century of the founding of the Paris Institute, Langue des signes française (French SL) had reached the Netherlands in 1790, the US in 1817, Brazil in 1857, as well as other countries, creating linguistic connections between geographically distinct signing communities. Similarly, other European SLs spread along political and colonial pathways.

Although the history of European deaf educational institutions has been well-documented, much less is known about how SLs themselves relate to one another. Whereas the world’s spoken languages have been classified in families and subfamilies based on their evolutionary histories, few attempts have been made to form large-scale genetic classifications of the world’s SLs, which are typically missing from overviews of the world’s language families (e.g., see for overviews including SLs). This lacuna in our knowledge of the history of an entire class of human languages is due in part to challenges in understanding the evolutionary processes that have shaped the diversity of contemporary SLs. In particular, using traditional historical comparative methods, sign researchers have been unable to distinguish the results of tree-compatible evolutionary processes—that is, patterns of similarity reflecting inheritance in a vertical ancestor-descendant relationship—from tree-incompatible processes, such as borrowing and convergence. As a consequence of these methodological challenges, comparative studies of SLs at times conflate vertical and horizontal relationships in forming SL families or forgo historical interpretations of their results.

While relationships amongst spoken languages have been studied using both traditional methods and, more recently, computational phylogenetic methods, to date no published studies have attempted to use phylogenetic methods to infer relationships amongst SLs on a large scale. Here, we use network-based exploratory data analysis (EDA) with a sample of 76 SL manual alphabets (40 extant, 36 historical) as a first step in investigating the evolutionary histories of SLs and the processes that have shaped them. Our approach makes use of data-display networks, which represent both tree-compatible and tree-incompatible patterns within a data set and are therefore a useful starting point for understanding evolutionary processes and formulating phylogenetic hypotheses.

## 2 Materials and Methods

### 2.1 Data

We created a data set of 76 manual alphabets, comprising 2,124 total entries from contemporary (40, SI 2.1) and historical (36, SI 2.2) sources, both print and online, the geographical distribution of which is illustrated in Figure 1. We transcribed manual alphabet handshape forms using HamNoSys, a transcription system for SLs, and coded each for morphological similarity using the EDICTOR tool, which was originally developed for the curation and analysis of historical comparative spoken language data. The data can be accessed and inspected through the EDICTOR tool (SI 1). In addition, we share the data in the TSV format required for processing by the LingPy software package and in CLDF format, following the recommendations of the Cross-Linguistic Data Formats Initiative.

A manual alphabet (MA) is a set of forms representing an alphabet, in which one form corresponds to one letter. An individual form is invariably comprised of a handshape with a particular spatial orientation, and may include a characteristic location and/or movement. Historical comparative studies of languages typically use basic vocabulary as comparanda, but there...
are good reasons to begin with a comparison of MAs instead. First, because the creation and
transmission of many MAs are well-documented, the data provide a relatively well-understood
test case for studying evolutionary processes in SLs. There exist far fewer historical dictionaries
of the world’s SLs compared to historical examples of MAs, though some do exist [29–31]. Sec-
ond, while computer-readable transcription systems have been developed for SLs—HamNoSys
[25] and SignWriting [32] are the most commonly used—these are still not widely used in SL
lexicography (see [33] for a dictionary-based dictionary including transcriptions in HamNoSys; see
[34] for a dictionary with transcriptions in SignWriting), and transcriptions in the two systems
are not straightforwardly comparable in all respects. Open, computer-readable, cross-linguistic
comparative data sets of SL vocabulary do not yet exist, due in part to the lack of consensus on
transcription system, but also to the time-consuming nature of SL transcription. Transcribing
handshapes in MAs instead of lexical signs—which typically include specifications for orienta-
tion, location, and movement in addition to handshape—significantly reduces the time necessary
to create a large cross-linguistic comparative data set.

An MA should not be understood as equivalent to a language’s phonological inventory. Though features of an MA overlap with phonological features found in the lexicon, the two sets
of features are not co-extensive [35]. Further, the linguistic status of MAs and their relationships
to signed and written language have been the subject of debate. In some SLs fingerspelling—the
representation of a written word with a sequence of MA handshapes—is used frequently in every-
day discourse [36], and young children acquire MA handshapes and fingerspelling patterns before
they are able to read [37, 38]. However, usage of fingerspelling varies both cross-linguistically

Figure 1: Contemporary and historical sign languages in our sample, with locations being de-

and across signers within the same signing community, which may have consequences for evolutionary processes and mutation rates. Notwithstanding this variation in MA usage, we show here that many of the historically-attested extra-linguistic connections amongst SLs are represented clearly in the network analyses in Section 3. We take this as confirmation that coherent, historically-relevant information is recoverable from a data set consisting of MAs.

2.2 Character Coding

We considered only handshapes and movements for determining similarity of MA forms, as these are represented most consistently in both contemporary and historical sources. In coding similarity across MAs representing different types of alphabets—for example, Latin and Cyrillic alphabets—it is possible to consider both the form of the grapheme and the sound represented by the grapheme as bases for organising the comparison. For example, the graphemic forms of Latin B and Cyrillic Б are similar, but they represent different sounds: the voiced bilabial stop (IPA [b]) for Latin and the voiced labiodental fricative (IPA [v]) for Cyrillic. Thus, a decision must be made about whether to compare handshapes representing Cyrillic Б with handshapes for Latin B or V. In many such cross-alphabet comparisons, the documentary record is suggestive about how the comparisons should be organised. In the example just mentioned, we compared handshapes representing Cyrillic Б with Latin B and not V because historical records provide clues about how the Russian SL MA was adapted from other European MAs in the early 19th century (see Section 4). When graphemic forms across two alphabet types are similar, as in the case of Latin and Cyrillic just mentioned, the form of the grapheme typically becomes the basis of comparison. However, there are many graphemes that differ in form across Latin-based, Cyrillic-based, Greek, and Arabic-based alphabets, all of which are represented by MAs in our sample. In such cases, the sound represented by the grapheme is the only basis for comparison.

Character mapping on selected networks can help to indicate how the coding of individual concepts contributes to the overall differentiation pattern (see SI 4.2).

In Figure 2, we exemplify our coding approach and the resulting binary matrices for use in phylogenetic methods. For reasons of space we limit the number of taxa in the example to those necessary for detailing our methods, and provide a second, more complex example in the electronic supplementary material (SI 3.2). The left side of Figure 2 depicts handshape forms representing four different graphemes, three of which represent the voiced velar stop (IPA [ɡ]): Latin g, Cyrillic г, and Persian/Urdu ﮔ, and one for the voiced velar fricative (IPA [ɣ]): Greek γ. Starting at the top of the figure, Afghan SL represents the grapheme ﮔ by extending the index and middle fingers, a form that is unique in this comparison. To represent the grapheme g, historical Brazilian SL (Brazilian 1875) and French SL use forms with an extended index finger and the thumb orientated in a similar direction. Pakistan SL uses a similar handshape for the grapheme ﮔ. Finally, Greek SL, Russian SL, and historical Russian SL (Russian 1835) represent γ and r with extended index finger and thumb extended outward.

In the upper right side of Figure 2, we show how these handshapes were coded in the EDICTOR tool. The value in the “Concept” column enables comparisons across alphabets. Morphological similarity was coded by assigning the same arbitrary numerical value in the “Cogid” column for languages with similar forms. The column “Narrow concept” tracks the grapheme represented in each MA, and “Year” indicates when historical sources were published, an empty cell in the “Year” column reflecting that the source is contemporary. Afghan SL, not being similar with any other taxon, was assigned cogid 328. We coded the handshapes in Brazilian 1875, French SL, and Pakistan SL as similar and assigned them cogid 60. Handshapes in Greek SL, Russian SL, and Russian 1835 were coded as similar and assigned cogid 4. The bottom of Figure 2 shows how our coding translates to the binary matrix for use in phylogenetic methods. Taking the character ID from the “Cogid” column, Brazilian 1875, French SL, and Pakistan SL
were scored as 1 for ID 60, while the other four taxa were scored as 0. Greek SL, Russian 1835, and Russian SL were scored as 1 for ID 4, and the other taxa 0. Afghan SL was scored as 1 for ID 328, while the other taxa were scored as 0.

### 2.3 Phylogenetic analysis

The matrix includes characters supporting language cliques that are compatible (inherited patterns) and incompatible (horizontally-propagated patterns) with the splits in the unknown true tree. The true tree is, in this case, not necessarily a sequence of dichotomous splits because the tree can be anastomising: (i) one ancestor may have more than two direct descendants; and (ii) a descendant may have more than one direct ancestor in the case that, for example, an MA is the product of combining two (or more) different sources. Thus, any simple, dichotomising tree that we infer or select using tree-reconstruction methods and commonly-used optimality criteria will be incomprehensive, and any signal in the matrix reflecting aspects not covered by the tree’s selected or inferred topology will add to data incompatibility. Another source of tree-incompatible signal is the inclusion of putative “ancestors” in the form of historical MAs, as well as their direct or distant “descendants”. Spencer et al. [41] showed that the distance-based Neighbor-Nets (NNets; [42]), which were designed to counter the problem of signal incompatibility, outperform tree inferences when it comes to correctly depicting ancestor-descendant relationships.

With respect to the complex signal in the underlying matrices, we thus relied exclusively on network-based EDA [24, 43] using planar (2-dimensional), distance-based (NNets), and multidimensional, tree-sample-based splits graphs (Support Consensus Networks [44, 45], CNets). We
used PAUP* [46] to compute simple (Hamming) pairwise distances and establish non-parametric bootstrapping (BS) branch support under the Least-Squares (LS) and Maximum Parsimony (MP) optimality criteria. BS analysis used 10,000 pseudo-replicates; replicate trees were inferred using the BioNJ algorithm [47] for LS and quick-and-dirty BS for MP as outlined by Müller [48] (‘MulTrees’ option deactivated; only one tree saved per replicate). For Maximum Likelihood (ML) BS support, we used 10,000 replicate trees generated with RAxML v.8.0.20 [49] and the standard model for binary data allowing for site variation modeled via the Gamma function, and corrected for ascertainment bias (recommended setting for binary data without invariable sites; the effect has only been tested for phylogenomic binary data; hence, we also ran the same analysis without correcting for ascertainment bias). Splits graphs were inferred with SplitsTree v.4.13.1 [50]. All analysis files are included in the electronic supplementary material.

Post-analysis character mapping was done by hand-and-eye following the logical framework of median networks and guidelines provided by Bandelt et al. [51] for their manual reconstruction. In contrast to a dichotomised and/or anastomised tree, a median network considers taxa to be either tips or medians, representing ancestral variants connecting the tips. A full median network includes all possible most-parsimonious solutions for the mutation of a character, character complex, or data matrix. For this study, we establish the minimum amount of necessary changes in each set of binary sequences representing concepts found in the standard Latin alphabet (letters a to z) along time-filtered networks (SI 4.2).

3 Results

The NNet in Figure 3 allows defining eight main groups of differing coherence and uniqueness. Each group forms a neighbourhood in the graph defined by a single, more or less prominent, edge-bundle. Three of the groups collect SLs of (i) Austrian-, (ii) French-, and (iii) British-origin; the oldest SLs in the first two of these groups (Austrian 1786, French 1800) may reflect the older common bases from which the SLs in these groups are derived. The largely extinct Austrian-origin group includes a single surviving contemporary SL, Icelandic SL. Most other contemporary SLs (e.g., Austrian, Danish, and German SLs) of the Austrian-origin group are now found in the French-origin group, which includes the International Sign MA. In addition, we recognise (iv) an Afghan-Jordanian group, with lowest overall dissimilarity to the British-origin group; (v) a Polish group that is connected with (vi) the Russian group via Latvian SL; (vii) a distinct Spanish group including the oldest MAs in our data set (Yebra 1593, Bonet 1620); and (viii) the very unique Swedish group which includes Portuguese SL. The spiderweb structure of the centre of the NNet graph indicates the data cannot resolve the principal relationships between the distinguished eight main groups.

The robustness of discriminating signal in the underlying matrix for each main group, estimated using non-parametric BS support, is shown in Table 1. In general, highly coherent, distinct groups (Afgan-Jordanian, British-origin, Polish, Spanish, Swedish groups) received moderate to high support (BS > 48; usually > 90 for at least two optimality criteria) irrespective of the optimality criterion used. Less coherent groups (Austrian-origin, French-origin, Russian groups) received low support (BS < 42). This demonstrates that the distance matrix well reflects the overall diversity patterns. Inter-group relationships are essentially unresolved: best-supported alternatives have a BS ≤ 23. Ambiguous BS support (i.e., BS << 100) can result from a lack of discriminating signal or internal signal conflict, which can be explored using CNets (see SI 4.3; electronic supplementary material). In the case of the low-supported Austrian- and French-origin groups, no alternative finds a BS ≥ 15; these groups are poorly supported but lack alternatives. The same holds for the much higher BS support of the Spanish group. In the case of the Russian group, the low support relates to competing alternatives: the data prefer
Figure 3: Neighbour-net based on simple (Hamming) pairwise distances calculated from the standard-coded Cogid binary matrix. Colours highlight the main groups and the Danish subgroup (cf. Figure 4; SI 4.2) within the Austrian-origin lineage. Neighbourhood-defining edge-bundles are also highlighted.
Table 1: Non-parametric bootstrapping (BS) support for neighbourhood-defining splits of main groups (highlighted in Figure 3). ML = maximum likelihood; ASC = corrected for ascertainment bias; UNC = uncorrected for ascertainment bias; NJ = neighbour-joining; P = parsimony.

<table>
<thead>
<tr>
<th>SL group</th>
<th>MLBS ASC</th>
<th>NJBS UNC</th>
<th>PBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austrian-origin</td>
<td>&lt;15</td>
<td>&lt;15</td>
<td>24</td>
</tr>
<tr>
<td>British-origin</td>
<td>99</td>
<td>99.5</td>
<td>99.8</td>
</tr>
<tr>
<td>French-origin</td>
<td>&lt;15</td>
<td>&lt;15</td>
<td>42</td>
</tr>
<tr>
<td>Afghan-Jordanian</td>
<td>99.9</td>
<td>99.6</td>
<td>66</td>
</tr>
<tr>
<td>Russian</td>
<td>18</td>
<td>17</td>
<td>40</td>
</tr>
<tr>
<td>Polish</td>
<td>93</td>
<td>90</td>
<td>99</td>
</tr>
<tr>
<td>Spanish</td>
<td>84</td>
<td>85</td>
<td>71</td>
</tr>
<tr>
<td>Swedish</td>
<td>99</td>
<td>99</td>
<td>97</td>
</tr>
</tbody>
</table>

No alternative found with BS ≥ 15

* Not including Latvian SL (see SI 4.3)

and would support partly incongruent tree-topologies (SI 4.3). The two major sources of signal conflict are (i) Latvian SL, which is substantially less dissimilar to the Polish group than all SLs of the Russian group (BS_{NJ} = 37, but BS_{ML,MP} < 15), hence, its intermediate placement in the NNet (Figure 3); and (ii) the Russian 1835 MA. In this case, the BS support values can vary substantially between optimality criteria: the distance-based NJ vs. the character-based, mutation-probability naive MP vs. the character- and model-based ML. Within the Russian group, Russian 1835 is most-closely related to contemporary Cyrillic-representing MAs, while differing from Estonian SL, Estonian 1988, and Latvian SL. In the planar NNet, this conflict is resolved by placing Russian 1835 on the opposite side of Estonian and Latvian SLs, while the n-dimensional CNets show according 3-dimensional boxes (when using a cut-off of BS ≥ 15).

Figure 4 shows stacked NNets including SLs from specific time periods: the lowest NNet in the figure with SLs up to 1840; the middle from the mid-19th to the mid-20th century; and the uppermost NNet including contemporary SLs. The bottom graph demonstrates the substantial diversity amongst MAs by the early 19th century, with three main clusters: Austrian-origin, French-origin, and Spanish. The Austrian-origin group diversified further during the second half of the 19th century (middle graph), while the French SL MA was dispersed largely unmodified to the Americas. The Russian group is closest to the Austrian-origin group (the early Austrian SL MAs from 1786 and 1823, as well as their close relatives, the early German 1820 and Hungarian 1827 MAs) and most distant to the French-origin group. The third main distinct cluster in the bottom graph is the Spanish group. The Swedish group, appearing first in the middle graph, is already unique in the second half of the 19th century. In the topmost graph, the overall picture remains the same, with a few exceptions. First, Polish SL, which is found in the Austrian-origin group in the middle graph, is positioned between the Austrian-origin and Russian groups and forms a cluster with Lithuanian SL. Second, contemporary Norwegian SL separates from the middle of the graph and is no longer grouped closely with Danish or Icelandic SLs. Third, Austrian, Danish, and German SLs, earlier examples of which were found in the Austrian-origin group, are grouped closely with the International Sign MA and with American SL.

The results of the EDA indicate that each major group goes back to an independent founding event; hence, no to little support for the deepest splits, the spiderweb-like centre of the overall NNet in Figure 3 and the three distinct clusters earliest SLs in Figure 4. Contemporary Spanish
Figure 4: Time-/taxon-filtered stacked Neighbour-nets, based on the same distance matrix used for Figure 3. Bottom NNet: SLs up to ca.1840; middle NNet: 1815–late 20th century; top NNet: mid-/late 20th century–present. Abbr.: SG = potential subgroups within the French-origin group.
and Catalan SLs are direct derivatives of the oldest MAs in our dataset, Yebra 1593 and Bonet 1620, while the French- and Austrian-origin groups constitute the two main independent traditions in continental Europe. Slightly modified versions of the French SL MA were dispersed into the Americas, with the American SL MA later forming the basis for the International Sign MA, which had a homogenising effect on several European MAs. International Sign possibly affected Norwegian SL and fully replaced the Austrian-origin handshapes in Danish, German, and Austrian SLs. Standardisation also influenced internal relationships within the French-origin group: we observe a “taxonomic turnover” with the original SLs in Europe and the New World (closest to French 1800) being replaced by versions very similar (“International Sign subgroup”) or more similar (“contemporary French subgroup”) to American SL and International Sign than to the French original, with the exception of contemporary Dutch (unique development), as well as Greek and Irish SLs (still closest to original 18th/19th century French).

Contemporary Icelandic SL is a direct derivate from the Danish subgroup within the Austrian-origin group; the same holds for Norwegian SL, which started to strongly deviate from the closely related Icelandic SL in the second half of the 20th century. The Russian group can be linked historically to the Austrian-origin group (see the bottom-most NNet in Figure 3) and underwent substantial restructuring in the adaptation from representing a Latin-based alphabet to Cyrillic. The Estonian and Latvian SL MAs are, to a lesser degree, 20th century derivates, with more links to the contemporary Russian group than to the Russian MA from 1835 (see also SI 4.3). Swedish constitutes an isolated, mainly unique tradition and is the basis of contemporary Swedish and Portuguese SLs. Although we have not included any historical examples of either, the British-origin and Afghan-Jordanian groups constitute isolated traditions that evolved independently (British-origin) or largely independently (Afghan-Jordanian) of the European groups.

4 Discussion

Figure 5 shows five European lineages and their hypothesised dispersal from the late 16th to the late 19th century. While the earliest MAs in our sample, Yebra 1593 and Bonet 1620, are typically identified as the ancestors of most one-handed MAs in the world today [36], we argue that communities in Austria, France, and Spain independently formed MAs using the early sources. We suggest that this independent formation supports identifying three separate SL lineages. The Yebran and Bonetian MAs, together with other Spanish MAs in the sample (Spanish 1815, 1845, 1859, contemporary Spanish and Catalan SLs), constitute a Spanish lineage. The NNet in Figure 3 gathers these MAs in a highly coherent group with moderate to high BS support (Table 1) without alternatives (see CNet in the electronic supplementary material). Because there existed no large-scale deaf educational institutions in Spain until the early 19th century, it seems unlikely that the two early MAs were used widely prior to that period, particularly in signing communities. Preserved as examples in books, the early MAs would not have evolved due to processes implicated in their usage in a community, explaining the relative lack of change in this lineage.

While the French MA was clearly formed using Spanish sources, we argue here that it did not evolve directly from a shared Spanish-French origin. Independent formation of the French MA is suggested by the clear separation of the French and Spanish groups in Figure 3 and of the earliest French and Spanish MAs in Figure 4. The Paris Institute had been in existence for decades before the school in Madrid was founded, making it unlikely that there could have been a common Spanish-French basis from which the French MA evolved. Similarities between the earliest French and Spanish MAs are explained by the fact that educators throughout Europe, including the founder of deaf education in France, de l’Épée in Paris, had become aware of Bonet’s (1620) Reducion de las Letras and its MA by the late 18th century [3, 52]. While the Spanish MA from
Figure 5: Hypothesised dispersal of European sign languages from late 16th to late 19th century, based on results in section 3. Colour-coding reflects five hypothesised lineages: Spanish, French-origin, Austrian-origin, British-origin, Swedish. Timeline reflects approximate years of first transmission of SLs, typically coinciding with establishment of schools for the deaf or migrations of signers (except in the cases of 1 & 2, which track publication of earliest manual alphabets in sample).
1815 shows little innovation compared to Yebra and Bonet, many differences are observable between the Yebran and Bonetian MAs and the earliest French MA in our sample (French 1800), in which just eleven forms remained unmodified and forms were added representing new letters \( k \), \( v \), and \( w \). In adapting the early Spanish sources, the signing community in Paris changed them substantially, both consciously by innovating new forms and through usage in a community, which resulted in only minimal modifications to existing forms. It is clear that the French MA drew on the original Spanish sources directly: where the original and later Spanish forms differed, the French MA either kept the original form unmodified (e.g., in the handshapes representing \( m \), \( n \)) or modified the original and not the later Spanish form (e.g., in the handshape representing \( q \)).

The other main continental European lineage is the Austrian-origin lineage. Founders of the first school for the deaf in Vienna, Joseph May and Friedrich Storch, visited de l’Épée and the Paris Institute in 1777 to learn pedagogical methods and to subsequently establish deaf education in the Habsburg Empire \[10\]. Perhaps because of these historical connections, it has been thought that Austrian SL is related to French SL in an ancestor-descendant relationship \[15\]. A large number of uniquely shared innovations (synapomorphies in biology) would support such a conclusion, indicating a unified basis from which both MAs later diverged. Shared innovations would be, in this case, those forms that differed from Spanish sources but that were uniquely shared in French and Austrian MAs. In fact, we find relatively few such potentially derived forms between the earliest Austrian and French MAs, including those representing \( d \), \( e \), \( l \), \( r \), and \( w \). Forms for \( f \) and \( v \) were shared by Austrian 1786 and French 1800, but also by early unrelated SLs, such as Swedish 1866.

That these possible synapomorphies across Austrian and French MAs are too few is reflected in the lack of a French-Austrian neighborhood ‘trunk’ in the NNet of the earliest MAs in the sample (Figure 4). If Austrian had indeed evolved from a French basis, the network should show a prominent fan-like structure including both MAs, with the oldest MAs in the middle and the most derived within each group as wings towards either side of the fan. The potential shared innovations may be better characterised as early borrowings due to minimal contact between French and Austrian signing communities; or as convergent evolution, in which handshape forms independently evolved in similar ways either to iconically match the forms of similar graphemes, or because forms were selected for that conferred articulatory and perceptual advantages. Finally, compare the positions of the Austrian, French, and Spanish groups in Figure 5 to the positions of other languages thought to have evolved from French SL. For example, we know that deaf French educators helped establish deaf education in the US \[53\], Brazil \[54\], and Mexico \[55\]: and those MAs remain topologically close to contemporary French SL. In Figure 6, the earliest American SL MA in the sample (American 1821) has diverged minimally from the chronologically closest French MA from 1815. In contrast, early Austrian, French, and Spanish MAs are found in differing neighbourhoods.

The results support classifying Danish as a sublineage of the Austrian-origin lineage. Peter Atke Castberg, who founded the first school for the deaf in Copenhagen in 1807, visited deaf educational institutions in Germany, France, and Austria between 1802–5 \[56\]. Forms representing \( c \), \( g \), \( h \), \( o \), \( p \), and \( q \) in the earliest Danish MA from 1808 are shared with Austrian but not French. In contrast, none of the forms in the earliest Danish MA from 1808 are unambiguously French in origin. These patterns are reflected in the bottom graph in the NNet in Figure 4: Danish 1808 is topologically closest to the early Austrian-origin group languages. The early Danish MA shows several innovated handshapes, such as those representing \( d \), \( f \), \( k \), \( s \), \( u \), \( v \), \( w \), \( y \), and \( z \), as well as new forms for the Danish letters \( æ \) and \( ø \). Thus, there is some support for classifying Danish as a separate lineage based on similar argumentation to that used above. In contrast to the Austrian, French, and Spanish cases above, however, early Danish and early Austrian
MAs are consistently found in close topological proximity in the network analyses under various methods and optimality criteria.

While the results support classifying Russian as a separate lineage created using Austrian sources, any interpretation of the results with respect to Russian SL is complicated by substantial adaptation and restructuring of an existing MA to represent the Cyrillic alphabet. Historical connections between Austrian and Russian communities likely developed when an Austrian-trained educator, Father Sigmund, and a Frenchman, Jean-Baptiste Jauffret, helped establish the first school for the deaf in St. Petersburg during the period from 1806–10 [57, 58]. Similarities between Russian SL and early Austrian MAs are reflected in the NNet of early MAs in Figure 4, which places Russian 1835 closest to the Austrian-origin group languages compared to other groups. Because of mismatches between the two alphabets, however, the Austrian MA underwent substantial restructuring to represent graphemes in Cyrillic. New forms were invented for some Cyrillic letters not found in the Latin alphabet; other forms were dropped; and some forms were used to represent Cyrillic letters that appear similar to Latin letters, but which occupy different positions in the alphabet. One consequence of this mismatch between alphabets is that there are fewer cross-alphabet comparanda between Latin- and Cyrillic-representing MAs. In addition, as discussed in Section 2, there are serious methodological challenges in deciding which forms should be compared, with the resulting possibility that some connections between sources and adaptations cannot be recovered. Thus, differences between Latin and Cyrillic alphabets may cause phylogenetic methods to overestimate the distance between the Austrian-origin and Russian groups.

The British-origin group forms an independent lineage, with links to the early 2-handed MA in *Digit i lingua* published in 1698 [59]. The NNet in Figure 3 shows a clear split of this group from the centre of the graph. All of the MAs in this lineage are predominantly 2-handed, and some characters are shared throughout the group (e.g., handshapes representing *d* and *x*). While the links between British SL and SLs in Commonwealth countries are relatively well-documented [12], less is known about how the British-origin MA came to other SLs in the group, namely, Czech, Croatian, and Turkish SLs. Zeshan [60, 46] reports the presence of British educators in Turkey in the early 1950s. Both Kuhn et al. [61, 55] for Croatian SL and Hudáková [62, 30] for Czech SL report that 1- and 2-handed MAs are in use in Croatia and the Czech Republic: in Croatia, the 2-handed alphabet is thought to be older, while the opposite may be true in the Czech Republic. Future research could uncover historical examples of MAs from these SLs that can help to clarify their connections to the British-origin group. Similarly, lexical investigations are likely to shed more light.

Finally, the Swedish group, which consists of only historical and contemporary Swedish SL, as well as Portuguese SL, forms a separate lineage. Bergman and Engberg-Pedersen [56] suggest that Per Aron Borg, the founder of the first school for the deaf in Stockholm in 1809, though aware of de l’Épée’s work in the Paris Institute, may not have been familiar with the MA in use in France when he created the Swedish SL MA. There are some similarities between an early example of the Swedish SL MA from 1866 and MAs used in France and other parts of Europe, such as in handshapes representing *c*, *f*, *k*, *l*, *m*, *n*, *o*, *u*, and *v*. Thus, while the new Swedish SL MA was created mainly in isolation from other lineages, Borg may have borrowed handshapes known widely in Europe. The connection between Swedish SL and Portuguese SL can also be traced to Borg, who helped establish deaf education in Lisbon in 1823 [56].

5 Conclusion

Despite their relevance to our understanding of human linguistic diversity and to theories of language change, the evolutionary histories of the world’s SLs have not, until now, been studied...
using state-of-the-art methods. We have shown that computational phylogenetic methods can be applied to SL data to uncover new insights into the evolutionary histories of SLs, to generate new hypotheses about their relationships, and to better understand the evolutionary processes that have shaped the diversity of contemporary SLs. Our analysis supports some aspects of existing SL classifications, but adds complexity to the overall picture, in particular to our understanding of the evolution of SLs from early sources. Our discussion of the independent establishment of SL lineages points to a characterisation of similarities across lineages as primarily horizontal, and not due to descent from a common ancestor, while within-lineage diversification does appear to be characteristically vertical in many cases. We anticipate that future studies of lexical data may contradict our phylogeny based on MAs, in particular for SLs that adopted the International Sign MA, because this adoption did not likely affect a language’s lexicon to any great extent. Notwithstanding these limitations, we suggest that our analysis be taken in future research as the best available phylogenetic classification of these SLs.

Acknowledgement

JML was supported by the ERC Starting Grant 715618 Computer-Assisted Language Comparison (http://calc.digling.org). We thank Tiago Tresoldi for sharing initial ideas on this project, Harald Hammarström for providing information on proposed sign language classifications, and Russell D. Gray and Richard P. Meier for comments on an earlier version of this paper.

Author Contributions

JMP and JML initiated the study. JMP and JML designed the database and developed methods for data coding. JML programmed the database system. JMP coded the data. GWG carried out the phylogenetic analysis and designed graphics. JMP and GWG interpreted the results. JMP and GWG wrote the first draft. All authors read the last draft and agree on its contents.

Supplementary Material

The supplementary material accompanying this paper provides additional information in form of an appendix, as well as all data, code, and analyses needed to replicate or test our studies. It can be downloaded from https://github.com/lingpy/sign-language-evolution-paper.

References


[59] Author unknown. *Digití-lingua, or, The most compendious, copious, facil, and secret way of silent converse ever yet discovered: Shewing, how any two persons may be capable, in half an hours time, to discourse together by their fingers only, and as well in the dark as the light / by a person who has conversed no otherwise in above nine years*. P. Buck, London, 1698.


Supplementary Information for the Paper
“Evolutionary Dynamics in the Dispersal of Sign Languages”

Justin M. Power, Guido W. Grimm, and Johann-Mattis List

June 2019

Contents

1 Organisation of the supplementary material 2

2 Language Selection 2
   2.1 Extant Manual Alphabets ........................................... 2
   2.2 Historical Manual Alphabets ....................................... 3

3 Data Preparation and Curation 4
   3.1 Technical Aspects .................................................... 4
   3.2 Coding ................................................................. 4

4 Phylogenetic Analysis 5
   4.1 Reasons for exploratory data analysis and effects of incompatible signal .... 5
   4.2 Character Mapping .................................................. 6
      4.2.1 Comprehensive character map for sign languages up to 1840 ............ 8
      4.2.2 Character map for sign languages from mid-19th to mid-/late 20th century .... 9
      4.2.3 Character map for contemporary sign languages ............................ 10
   4.3 Split Support ......................................................... 10
1 Organisation of the supplementary material

The data was curated with help of the EDICTOR (List 2017). We used a server-based version to ease collaboration. A link to the database can be found at http://dighl.github.io/sign-languages. Since the database curation process was in flux for some time, and may change in the future, we provide a final stable dump of this database. The data itself is curated on GitHub (https://github.com/lexibank/powerma), while the versions underlying this draft along with the results and the code needed to convert the data into the formats required by the software packages we used can be found on GitHub (https://github.com/lingpy/sign-language-evolution-paper).

2 Language Selection

2.1 Extant Manual Alphabets

The table below shows the extant manual alphabets and sources in our sample, with IDs taken from Figures 1 and 5 in main text.

<table>
<thead>
<tr>
<th>ID</th>
<th>Manual Alphabet</th>
<th>Abbrev</th>
<th>Glottolog</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>Afghan Sign Language</td>
<td>ZEA</td>
<td>afg1239</td>
<td>Afghan Sign Language 2001</td>
</tr>
<tr>
<td>24</td>
<td>Albanian Sign Language</td>
<td>AlbSL</td>
<td>alba1271</td>
<td>gjishh.al/daktilim</td>
</tr>
<tr>
<td>27</td>
<td>American Sign Language</td>
<td>ASL</td>
<td>amer1248</td>
<td>Tennant and Brown 1998, Lydell 2018</td>
</tr>
<tr>
<td>18</td>
<td>Australian Sign Language</td>
<td>Auslan</td>
<td>aust1271</td>
<td>Johnston 2014</td>
</tr>
<tr>
<td>30</td>
<td>Austrian Sign Language</td>
<td>ÖGS</td>
<td>aust1252</td>
<td>Lydell 2018</td>
</tr>
<tr>
<td>31</td>
<td>Brazilian Sign Language</td>
<td>LSB</td>
<td>braz1236</td>
<td>Lydell 2018</td>
</tr>
<tr>
<td>4</td>
<td>British Sign Language</td>
<td>BSL</td>
<td>brit1235</td>
<td>Brien 1992, Lydell 2018</td>
</tr>
<tr>
<td>32</td>
<td>Bulgarian Sign Language</td>
<td>BZhE</td>
<td>bulg1240</td>
<td>Lydell 2018</td>
</tr>
<tr>
<td>33</td>
<td>Catalan Sign Language</td>
<td>LSC</td>
<td>cata1287</td>
<td>Perelló and Masclans 1998</td>
</tr>
<tr>
<td>34</td>
<td>Croatian Sign Language</td>
<td>HZJ</td>
<td>croa1242</td>
<td>Kuhn et al 2006, Lydell 2018</td>
</tr>
<tr>
<td>35</td>
<td>Czech Sign Language</td>
<td>CzSL</td>
<td>czec1253</td>
<td>Hudáková 2008</td>
</tr>
<tr>
<td>40</td>
<td>Danish Sign Language</td>
<td>DTS</td>
<td>dani1246</td>
<td>tegnsprog.dk</td>
</tr>
<tr>
<td>70</td>
<td>Dutch Sign Language</td>
<td>NGT</td>
<td>dutc1253</td>
<td>Zwitserlood 2010</td>
</tr>
<tr>
<td>42</td>
<td>Estonian Sign Language</td>
<td>EVK</td>
<td>esto1238</td>
<td>eki.ee/dict/viipekeel</td>
</tr>
<tr>
<td>43</td>
<td>Finnish Sign Language</td>
<td>FinSL</td>
<td>finn1310</td>
<td>Kuurojen Liitto 1998</td>
</tr>
<tr>
<td>44</td>
<td>Flemish Sign Language</td>
<td>VGT</td>
<td>vlaa1235</td>
<td>Vertaal 2012</td>
</tr>
<tr>
<td>48</td>
<td>French Sign Language</td>
<td>LSF</td>
<td>fren1243</td>
<td>Lydell 2018</td>
</tr>
<tr>
<td>49</td>
<td>French-Belgian Sign Language</td>
<td>LSFB</td>
<td>lang1248</td>
<td>dicto.lsf.b, sourdlang.be</td>
</tr>
<tr>
<td>52</td>
<td>German Sign Language</td>
<td>DGS</td>
<td>germ1281</td>
<td>Lydell 2018</td>
</tr>
<tr>
<td>53</td>
<td>Greek Sign Language</td>
<td>GSL</td>
<td>gree1271</td>
<td>Hatzopoulos 2008, Lydell 2018</td>
</tr>
<tr>
<td>54</td>
<td>Icelandic Sign Language</td>
<td>ÍTM</td>
<td>icel1236</td>
<td>Lydell 2018</td>
</tr>
<tr>
<td>22</td>
<td>Indian Sign Language</td>
<td>IPSL</td>
<td>ind1237</td>
<td>Lydell 2018</td>
</tr>
<tr>
<td>55</td>
<td>International Sign</td>
<td>IS</td>
<td>inte1259</td>
<td>Rubino et al 1975, Lydell 2018</td>
</tr>
<tr>
<td>56</td>
<td>Irish Sign Language</td>
<td>ISL</td>
<td>iris1235</td>
<td>Learn Irish Sign Language</td>
</tr>
<tr>
<td>57</td>
<td>Italian Sign Language</td>
<td>LIS</td>
<td>ital1275</td>
<td>Magarotto 1996, Lydell 2018</td>
</tr>
<tr>
<td>58</td>
<td>Jordanian Sign Language</td>
<td>LIU</td>
<td>jord1238</td>
<td>Hendriks 2008</td>
</tr>
<tr>
<td>59</td>
<td>Latvian Sign Language</td>
<td>LSL</td>
<td>latv1245</td>
<td><a href="http://zimjuvaloda.lv/lv/alphabet">http://zimjuvaloda.lv/lv/alphabet</a>, Lydell 2018</td>
</tr>
<tr>
<td>60</td>
<td>Lithuanian Sign Language</td>
<td>LGK</td>
<td>lith1236</td>
<td>gestai.ndt.lt/pirstu-abecele, Lydell 2018</td>
</tr>
<tr>
<td>61</td>
<td>Mexican Sign Language</td>
<td>LSM</td>
<td>mexi1237</td>
<td>Lydell 2018</td>
</tr>
<tr>
<td>20</td>
<td>New Zealand Sign Language</td>
<td>NZSL</td>
<td>newz1236</td>
<td>McKee et al 2011</td>
</tr>
</tbody>
</table>
2.2 Historical Manual Alphabets

The following table shows the historical manual alphabets and sources in our sample, with IDs taken from Figures 1 and 5 in main text, and abbreviations taken from the table in in Section 2.1.

<table>
<thead>
<tr>
<th>ID</th>
<th>Abbrev</th>
<th>Year</th>
<th>Location</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>ASL</td>
<td>1821</td>
<td>New York, USA</td>
<td>Akerly 1821</td>
</tr>
<tr>
<td>25</td>
<td>ASL</td>
<td>1886</td>
<td>Washington D.C., USA</td>
<td>Gordon 1886</td>
</tr>
<tr>
<td>26</td>
<td>ASL</td>
<td>1918</td>
<td>Iowa, USA</td>
<td>Long 1918</td>
</tr>
<tr>
<td>12</td>
<td>DTS</td>
<td>1808</td>
<td>Copenhagen, Denmark</td>
<td>Castberg 1818</td>
</tr>
<tr>
<td>36</td>
<td>DTS</td>
<td>1871</td>
<td>Copenhagen, Denmark</td>
<td>Nyegaard 1871</td>
</tr>
<tr>
<td>37</td>
<td>DTS</td>
<td>1907</td>
<td>Copenhagen, Denmark</td>
<td>Jorgensen 1907</td>
</tr>
<tr>
<td>38</td>
<td>DTS</td>
<td>1926</td>
<td>Copenhagen, Denmark</td>
<td>Døvstumme 1926</td>
</tr>
<tr>
<td>39</td>
<td>DTS</td>
<td>1967</td>
<td>Copenhagen, Denmark</td>
<td>Plum et al 1976</td>
</tr>
<tr>
<td>6</td>
<td>DGS</td>
<td>1821</td>
<td>Gmünd, Germany</td>
<td>Alle 1821</td>
</tr>
<tr>
<td>50</td>
<td>DGS</td>
<td>1909</td>
<td>Leipzig, Germany</td>
<td>Reuschert 1909</td>
</tr>
<tr>
<td>51</td>
<td>DGS</td>
<td>1916</td>
<td>Berlin, Germany</td>
<td>Riemann 1916</td>
</tr>
<tr>
<td>41</td>
<td>EVK</td>
<td>1988</td>
<td>Tallinn, Estonia</td>
<td>Toom 1988</td>
</tr>
<tr>
<td>19</td>
<td>ÍTM</td>
<td>1857</td>
<td>Akureyri, Iceland</td>
<td>Sigurðsson 1857</td>
</tr>
<tr>
<td>3</td>
<td>LSF</td>
<td>1799-1800</td>
<td>Paris, France</td>
<td>Unknown 1800</td>
</tr>
<tr>
<td>45</td>
<td>LSF</td>
<td>1803</td>
<td>Paris, France</td>
<td>Sicard 1803</td>
</tr>
<tr>
<td>46</td>
<td>LSF</td>
<td>1815</td>
<td>Paris, France?</td>
<td>de Ladebat 1815</td>
</tr>
<tr>
<td>47</td>
<td>LSF</td>
<td>1856</td>
<td>Paris, France</td>
<td>Pelissier 1856</td>
</tr>
<tr>
<td>10</td>
<td>LSE</td>
<td>1815</td>
<td>Madrid, Spain</td>
<td>Martí 1815</td>
</tr>
<tr>
<td>71</td>
<td>LSE</td>
<td>1845</td>
<td>Madrid, Spain</td>
<td>Ballesteros and Villabrille 1845</td>
</tr>
<tr>
<td>72</td>
<td>LSE</td>
<td>1859</td>
<td>Madrid, Spain</td>
<td>Carderera 1859</td>
</tr>
<tr>
<td>21</td>
<td>LSB</td>
<td>1875</td>
<td>Rio de Janeiro, Brazil</td>
<td>da Gama 1875</td>
</tr>
<tr>
<td>7</td>
<td>LIS</td>
<td>1897</td>
<td>Milan, Italy</td>
<td>Fornari 1897</td>
</tr>
<tr>
<td>9</td>
<td>MJ</td>
<td>1827</td>
<td>Vác, Hungary</td>
<td>Schwarzer 1827</td>
</tr>
<tr>
<td>8</td>
<td>NGT</td>
<td>1790</td>
<td>Groningen, Netherlands</td>
<td>Mörser and Guyot 1790</td>
</tr>
<tr>
<td>69</td>
<td>NGT</td>
<td>1820</td>
<td>Groningen, Netherlands</td>
<td>van Heijningen Bosch 1820</td>
</tr>
<tr>
<td>17</td>
<td>NTS</td>
<td>1893</td>
<td>Oslo, Norway</td>
<td>Svendsen 1893</td>
</tr>
<tr>
<td>62</td>
<td>NTS</td>
<td>ca. 1900</td>
<td>Trondheim, Norway</td>
<td>Bruun 1900</td>
</tr>
<tr>
<td>63</td>
<td>NTS</td>
<td>1955</td>
<td>Trondheim, Norway?</td>
<td>Norske Døves Landsforbund 1955</td>
</tr>
<tr>
<td>5</td>
<td>ÖGS</td>
<td>1786</td>
<td>Vienna, Austria</td>
<td>May 1789</td>
</tr>
</tbody>
</table>
3 Data Preparation and Curation

3.1 Technical Aspects

To edit the data in a machine- and human-readable way, we used the EDICTOR application (List 2017), which was originally designed for spoken languages and phonetic transcriptions. To annotate presumably cognate handshapes, we used the “full cognate” annotation schema which essentially assumes that cognacy is a transitive relation applying to a full form in a binary fashion (two forms are either cognate or not). To export the data to the Nexus format, we made use of LingPy (List et al. 2018). The accompanying repository with data-dump and source code shows how LingPy can be used for data conversion.

3.2 Coding

Simplified coding example for handshapes representing Latin h, n, and p, and Cyrillic н and п.

Coding is more complex when there is a mismatch across alphabet types of graphemic forms and sounds represented. The figure above exemplifies our coding for ten MAs of handshapes representing Latin h, n, and p, as well as Cyrillic н and п. Consider, first, the overlap in sounds represented across the two alphabets. Latin h represents the voiceless glottal fricative (IPA [h]), a sound for which Cyrillic has no corresponding letter; Latin n and Cyrillic н represent the voiced alveolar nasal (IPA [n]); and Latin p and Cyrillic п represent the voiceless bilabial stop (IPA [p]). However, when considering the forms of the graphemes, the two alphabets overlap in different ways: there are similarities between the forms of Latin h and Cyrillic н, as well as Latin n and Cyrillic п.

The left side of the figure shows our coding of handshapes representing the letters described above. We coded the handshapes representing Latin h in contemporary Austrian SL, German SL, and American SL, as
well as historical American SL (American 1821), as similar and assigned them character ID 145 (in the column “Cogid”). Handshapes for Latin h in historical Austrian SL (Austrian 1823) and German SL (German 1820) were coded as similar and assigned ID 118. We chose to compare handshapes for Cyrillic н in historical (Russian 1835) and contemporary Russian SL with the handshapes representing Latin h mentioned above. Due to historical connections between Austrian-trained educators and the establishment of deaf education in Russia in the early 19th century (Abramov 1993, Williams and Fyodorova 1993), we reasoned that the similarity in handshape and graphemic forms for Latin h and Cyrillic н in Austrian 1823 and Russian 1835 are due to the adaptation of the Austrian h handshape in the Russian SL MA to represent Cyrillic н. The alternative comparison, in which handshapes for Cyrillic н are compared with non-homologous handshapes representing Latin н, is a possible but, in our view, incorrect approach based on the historical record. Therefore, we coded the historical and contemporary Russian handshapes in the h-comparison, which can be seen by observing the “Concept” and “Narrow concept” columns. The handshapes representing Cyrillic н in Russian 1835 and contemporary Russian SL were coded as similar and assigned ID 299. Similarly, we compared the handshape forms representing Latin н in Estonian SL and h in Latvian SL to the forms in the h-comparison. That the handshapes in Estonian SL and Latvian SL are homologous to the Russian SL forms for Cyrillic н is likely due to the history of deaf education in the former Soviet Union and the use of Russian SL in Estonia. In addition, as we have described, the graphemic form of Latin h is similar to the form of Cyrillic н, and both Latin н and Cyrillic н represent the alveolar nasal (IPA [n]). We coded the forms for н in Estonian SL and h in Latvian SL as similar to the Russian SL form and assigned them the character ID 299.

Next, historical examples of American SL (American 1821), Austrian SL (Austrian 1823), and German SL (German 1820), as well as contemporary Latvian SL, use similar handshapes for Latin н and were assigned ID 255. Handshapes for Latin н in contemporary American SL, Austrian SL, and German SL were coded as similar and assigned ID 137. The cases of contemporary and historical Russian SL, as well as Estonian SL, were more complex. Using similar reasoning to that described above for comparing handshapes for Latin h and Cyrillic н, we compared the Austrian 1823 handshape representing Latin h with the Russian 1835 handshape for Cyrillic н, the latter grapheme representing the voiceless bilabial stop (IPA [p]). We included the Russian 1835 handshape for п in the n-comparison and coded the handshape with ID 255. The contemporary Russian SL handshape for н differs slightly from the Russian 1835 example, with greater bending of the index and third fingers, and was assigned ID 137. Estonian SL reversed the process just described, taking the Russian SL handshape for Cyrillic н to represent Latin п. Thus, we included the Estonian SL handshape for п in the n-comparison, assigning ID 255. In addition, because the Estonian SL handshape for н was coded in the h-comparison, and because Estonian SL also has a handshape representing Latin h, Estonian forms were coded twice in the set of character IDs for the h-comparisons. The Estonian SL handshape for Latin h is unlike any other form and was assigned ID 113.

4 Phylogenetic Analysis

4.1 Reasons for exploratory data analysis and effects of incompatible signal

We assume that, in the case of language, evolutionary history does not strictly follow a tree model, but includes reticulation as a result of borrowing and, in the case of MAs, partial or complete replacement by standardisation or due to complex socio-political constraints. In addition, we expect a substantial amount of positive selection due to articulatory and perceptual pressures within the framework of a limited number of possible morphologies (handshapes), resulting in a high level of homoplasy, as similar handshapes evolved or were conceived independently from each other. Homoplasy will provide support for topological alternatives, outcompeting those best reflecting the evolutionary history of SLs. Finally, our data includes historical and contemporary MAs (i.e., potential ancestors and their descendants), a situation poorly handled by trees and best by NNets (Spencer et al. 2004). Each aspect can add an additional evolutionary/historical dimension,
and a tree is, per se, a 1-dimensional graph.

For instance, if A, B, and C have a common origin only expressed by similarity of B to both A and C, it’s impossible to find a correct tree. The NNet can handle incompatible signal to some degree (B will be placed between A and C) but is restricted to distance matrices, which can be strongly biased by missing data artefacts (here, the number of Cogids applicable only for a subset of the SLs); and they are planar, 2-dimensional graphs. If a taxon or lineage shares traits with more than two distinct, unrelated lineages (e.g., if B borrowed from the not directly related D), one of these relationships will get lost in the graph. The limitation to two dimensions is the reason for the differences between the all-inclusive NNet in Figure 3 and the time-filtered N Nets in Figure 4, both in the main text (e.g., regarding the placement of Russian 1835). CNets are \( n \)-dimensional, that is, if the data reflect \( n \) different topologies, they will all be represented in the CNet. For the A-B-C-D example, we may find support for three partly incompatible splits: A + B, B + C, B + D. If the common origin of A, B, and C is well reflected in the character-matrix, we will get high support for a fourth split, A + B + C, which competes with B + D, and both will be seen in the CNet. The NNet will, in contrast, place B between A + C and D (aspect-wise correct), and the tree will place either (i) B as sister to D and both as sister to A + C, or (ii) D as sister to A + B + C, which is equally incorrect. However, CNets are based on a tree sample, hence, all limitations that apply for using tree-inference to reconstruct the history of (here) SLs, apply also for each inferred pseudoreplicate tree during BS (here, 10,000 BS pseudoreplicates).

In addition to general branching artefacts, we will have more or less random branching patterns (NJ- and ML-BS replicates are always fully-resolved trees, no branch has zero length, no polytomies). The use of different optimality criteria allows for testing the stability of potential relationships seen in the BS tree samples on which the BS-support CNets are based under different assumptions: NJ-BS CNets show the robustness of signal based on overall similarity (or dissimilarity), minimising the effect of single, potentially misleading characters, but also inflicting relationships based on dissimilarity to everything else (including long-branching artifacts, LBA); P-BS CNets, also affected by LBA, provide the most-conservative, but often also least-discriminating result under the assumption that all character changes (mutations) have exactly the same probability, an assumption that must be wrong for our data; and ML-BS CNets optimise a model that allows for between-character variation but which runs the risk of over-weighting certain character splits. If the CNets converge and support a neighbourhood seen in the NNet, it can be considered a data-unbiased result; if they deviate from or contrast with each other, the resulting competing alternatives may be biased by unrepresentative pairwise distance profiles due to missing data, or by inferred character mutations that strongly depend on the assumed model. Given the complexity of character similarity and historical pathways of SLs, it is impossible to judge, in such cases, which optimality criterion gives a better reflection of the true situation. However, it is safe to assume that the true situation is one of the preferred alternatives, if it can be explained by a single tree; or that the different preferred alternatives show different aspects of the true situation, if too complex for a single tree.

### 4.2 Character Mapping

The figures in this section show character maps based on the binary sequences encoding for the basic concepts shared among all Latin alphabets, i.e. the standard set of 26 letters from “a” to “z”. In principle, for each concept a (full) median network is reconstructed for the binaries encoding for a character and then mapped visually on the time-taxon-filtered neighbour-nets. The (inferred or deduced) mutation is indicated for each concept by arrows. When deduced, we consider not only the character cliques mapped on the networks but also the age and country.

For instance, for the concept “k” in the oldest MAs set, we have three binary sequences (addressed and abbreviated as “Dutch” = “Du.”, after the earliest MA showing this handshape, Dutch 1790 vs. “†Fr.” vs. “At.”; see annotation definitions below). The full median network for handshapes of the concept “k” is a triangle: to change from one binary sequence to another requires two steps (loss of the Cogid defining the one handshape, and gain of the Cogid defining the other handshape). “k” is not defined for the Spanish-origin group including the oldest MAs (Bonet, Yebra) that influenced the first MAs in all groups. Hence,
the Dutch version can be inferred as one primitive/commonly shared handshape evolved/changed from un-
known source (“→Du.”), possibly competing with the extinct French handshape “→ † Fr.” in the nearly
as old French 1800/1803 MAs. The original Austrian-origin group handshape (“At.”) is inferred to have
replaced the Dutch version (“At. → Du.”): the neighbourhood linked to the Austrian handshape is embed-
ded in a much larger neighbourhood including earliest Danish and Russian MAs as well as Austrian 1823
with a “Dutch”-type handshape. Note that this is an inference using the logical framework for median
networks and should not be viewed as conclusive evidence that the “Austrian”-type replaced the “Dutch”-type.
With respect to the age of the involved MAs and overall reconstructed history of SLs (see main text), it is
equally probable that younger MAs of the Austrian-origin group took over the more widespread “Dutch”-
type replacing the original “Austrian”-type typical for this lineage. Hence, there may be conflicts in the
reconstructions shown in all three figures, each one using a different taxon set and potentially a different set
of binaries. For example, in the mid-time network, we have an additional “Danish”-type and the “Dutch”-
type is inferred as the original handshape for “k” in all MAs including this concept. On the other hand, the
“Russian”-type handshape used in contemporary Norwegian SL can be deduced to represent a borrowing
or convergent development since older Norwegian MAs showed the “Danish”-type typical for the Danish
subgroup within the Austrian-origin group (“Da. → Ru.”).

All annotations in the graphs relate to the labels used for binary sequences in the file lists.xlsx in the
electronic supplementary material. We define the annotations below.

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Handshape that can be traced back to Bonet 1620, different from the handshape in Yebra’s MA.</td>
</tr>
<tr>
<td>Y</td>
<td>Handshape that can be traced back to Yebra 1593, different from the handshape in Bonet’s MA.</td>
</tr>
</tbody>
</table>
| O          | Original handshape: the handshape inferred as the original form of all covered European
            | lineages. When SLs of different groups share a handshape with Bonet and Yebra, this
            | handshape is labelled as “O”. Per definition, “B”, “Y” and “O” are mutually exclusive. |
| C          | Cosmopolitan: the label indicates that this handshape is not only shared among different
            | groups covered in the networks but also in the British-origin and/or Afghan-Lebanese
            | groups, for which our data set includes no historical MAs (hence, not part of any graph
            | in Figures 4.2.1–4.2.3). |
| D          | Derived handshape: a handshape shared by SLs of different lineages. Per definition, the
            | label “D” is mutually exclusive with all other labels, we only used it, when none of the
            | other labels apply. |
| “unique”   | Unique binaries: handshapes restricted to a single SL. |
| “†”        | Highlights that this particular handshape is only found in historical MAs. |

<table>
<thead>
<tr>
<th>Two letters</th>
<th>Handshapes (mostly) restricted to, exclusively found in a single main group.</th>
</tr>
</thead>
<tbody>
<tr>
<td>At.</td>
<td>Austrian/Austrian-origin group</td>
</tr>
<tr>
<td>Da.</td>
<td>Danish/Danish subgroup within the Austrian-origin group</td>
</tr>
</tbody>
</table>
| Du.         | Dutch: used for handshapes of the French-origin group not shared by French
            | SLs but found in the historical Dutch MAs (especially Dutch 1790, the oldest MA of the
            | French-origin group in our data set).                                   |
| Fr.         | French/French-origin group                                              |
| Ge.         | German: used for handshapes of the Austrian-origin group not diagnostic
            | for the Danish subgroup and not found in the oldest Austrian MAs but (historical) German
            | MAs                                                                      |
| Po.         | Polish/Polish group                                                     |
| Ru.         | Russian/Russian group                                                   |
Sp. Spanish/Spanish-origin group: this can be a handshape exclusively found in Bonet, Yebra and Spanish SLs as well as a handshape found only in younger Spanish MAs but not Bonet or Yebra.

Sw. Swedish/Swedish group

Coloured mutations refer to the accordingly coloured edge-bundle (taxon split) in the respective graph: mutations in light grey (and smaller) font unique mutations restricted to the respective OTU (“leaf”) of the network; dark grey indicates the (inferred) handshape set of a hypothetical common “ancestor” (median networks can place taxa/concepts in explicit ancestor-descendant relationships, with the inferred “medians” representing hypothetical ancestors).

### 4.2.1 Comprehensive character map for sign languages up to 1840

The list highlights the handshapes shared (labelled subsequently as original/“O”, or Spanish/“Sp.”) and differing between the two oldest MAs (labelled as “B” or “Y”) in our data set (Yebra 1693, Bonet 1620). The handshapes for the concept “d” are unique in both the Yebra and Bonet MA, all other MAs show different handshapes (most show a handshape characteristic of the oldest MAs of the Austrian- and French-origin groups).
4.2.2 Character map for sign languages from mid-19th to mid-/late 20th century

Similar to the character map above, each lineage can be characterised by group-specific handshapes (in coloured font). Note the striking difference between the diversification patterns in the French- vs. Austrian-origin group. The inflated fan-like structure of the neighbour-net for the Austrian cluster relates to a gradual accumulation of lineage-specific handshapes and their subsequent modification (and/or) replacement in subgroups. Potential older sources (Spanish 1815, Russian 1835, German 1820, Danish 1808, French 1815 and American 1821 MAs) are included to link this taxon set with the one shown in Figure 4.2.1.
4.2.3 Character map for contemporary sign languages

The taxon set also includes historical but post-war MAs (second half of the 19th century). In addition to lineage-specific handshapes, we find concepts directly linking the Polish and Austrian-origin group (coloured font) by shared handshapes. Note the high level of inferred single-SL (tip) mutations, especially in (phylogenetically) isolated SLs such as contemporary Norwegian or Albanian SL.

4.3 Split Support

Non-parametric bootstrapping (BS) support for competing relationships in the Polish and Russian groups with respect to Latvian SL. ML = maximum likelihood; ASC = corrected for ascertainment bias; UNC = uncorrected for ascertainment bias; NJ = neighbour-joining; P = parsimony.

<table>
<thead>
<tr>
<th>Alternative type</th>
<th>Group</th>
<th>MLBS</th>
<th>NJBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal</td>
<td>Contemp. Cyrillic: Bulgarian SL, Russian SL, Ukrainian SL</td>
<td>84</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>ASC = 71, UNC = 49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-level</td>
<td>CNet: Russian gr. incl. Russian 1835</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>ASC = 27, UNC = 54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNet: Contemp. Russian gr. + Latvian SL</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>ASC = 18, UNC = &lt;15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNet: Contemp. Russian gr. + Estonian SL</td>
<td>&lt;15</td>
<td>&lt;15</td>
</tr>
<tr>
<td></td>
<td>ASC = 46, UNC = 21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second-level</td>
<td>Preferred: Russian gr. incl. Russian 1835, excl. Latvian SL</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>ASC = 40, UNC = 39</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CNet: Russian gr. incl. Latvian SL excl. Russian 1835</td>
<td>18</td>
<td>&lt;15</td>
</tr>
<tr>
<td></td>
<td>ASC = 30, UNC = &lt;15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep split</td>
<td>NNet-alternative: Latvian SL part of Russian gr.</td>
<td>41</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>ASC = 33, UNC = 39</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NNet-alternative: Latvian SL part of Polish gr.</td>
<td>&lt;15</td>
<td>&lt;15</td>
</tr>
<tr>
<td></td>
<td>ASC = 37, UNC = 39</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Formari, Pasquale (1897). Il sordomuto e la sua istruzione: Manuale per gli allievi e le allieve delle R. scuole normali, maestri, genitori e filantropi. Milano: Ulrico Hoepli.


Yebra, Melchor de (1593). *Libro llamado Refugium Infirmonarum: Muy util y provechoso para todo genero de gente: En el qual se contienen muchos auisos espirituales para socorro de los afligidos enfermos, y para ayudar a bien morir a los que estan en lo ultimo de su vida*: con un Alfabeto de S. Buenauentura para hablar por la mano. Madrid: Luys Sa[n]chez.