

# Cross-Linguistic Phonological Similarity Analysis in Sign Languages Using HamNoSys

Abhishek Bharadwaj Varanasi, Manjira Sinha, Tirthankar Dasgupta

TCS Research

<sup>1</sup>{varanasi.abhishek,sinha.manjira,dasgupta.tirthankar}@tcs.com

## Abstract

This paper presents a cross-linguistic analysis of phonological similarity in sign languages using symbolic representations from the Hamburg Notation System (HamNoSys). We construct a dataset of 1000 signs each from British Sign Language (BSL), German Sign Language (DGS), French Sign Language (LSF), and Greek Sign Language (GSL), and compute pairwise phonological similarity using normalized edit distance over HamNoSys strings. Our analysis reveals both universal and language-specific patterns in handshape usage, movement dynamics, non-manual features, and spatial articulation. We explore intra and inter-language similarity distributions, phonological clustering, and co-occurrence structures across feature types. The findings offer insights into the structural organization of sign language phonology and highlight typological variation shaped by linguistic and cultural factors.

## 1 Introduction

Sign languages (SLs) are complex visual-gestural languages that convey meaning through a combination of hand configurations, movements, orientations, and spatial locations (Sinha, 2009). Unlike spoken languages, sign languages lack a standardized written form (Langer et al., 2014), making computational analysis and cross-linguistic comparison particularly challenging. One of the foundational aspects of sign language linguistics is phonology—the study of minimal visual units that distinguish signs. Phonological modeling in sign languages has been a growing area of interest in computational linguistics and sign language processing. Early work focused on rule-based systems and handcrafted features to capture phonological components such as handshape, location, and movement (Stokoe,

1960; Brentari, 1998). These approaches laid the foundation for formal linguistic analysis but lacked scalability and cross-linguistic generalization.

This paper addresses the problem of identifying signs that are phonologically similar within and across multiple sign languages. Specifically, we focus on four major sign languages: British Sign Language (BSL), German Sign Language (DGS), French Sign Language (LSF), and Greek Sign Language (GSL). For each language, we construct a dataset of 1000 signs, each annotated with its phonological structure using the Hamburg Notation System (HamNoSys) (Prillwitz and für Deutsche Gebärdensprache und Kommunikation Gehörloser, 1989). We compute pairwise phonological similarity between all sign pairs using a normalized edit distance over their HamNoSys representations, resulting in a  $1000 \times 1000$  similarity matrix per language.

Unlike prior work that focuses on building computational models for sign recognition or translation (Cihan Camgoz et al., 2017; Camgoz et al., 2018; Stoll et al., 2020; Saunders et al., 2020; Chen et al., 2022), our objective is to perform a detailed analytical study of phonological similarity patterns. We explore intra-language and inter-language similarity distributions, identify phonological clusters, and investigate the structural properties of the resulting similarity matrices. Our findings offer insights into the phonological organization of signs and provide a foundation for future work in multilingual sign language processing.

### 1.1 Overview of HamNoSys

The Hamburg Notation System (HamNoSys) (Prillwitz and für Deutsche Gebärdensprache und Kommunikation Gehörloser,



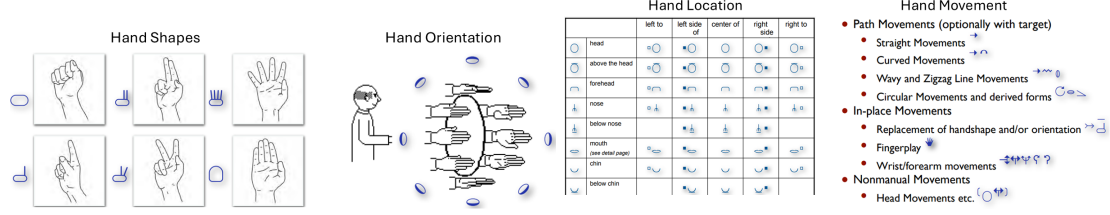


Figure 2: Examples of HamNoSys for hand shape, orientation, location and movement (Hanke, 2010)

ing phonological distance using feature-based representations, but their work was limited to small datasets and single-language settings. More recent work has explored neural models for sign similarity (Camgoz et al., 2020), though these approaches often rely primarily on visual features without explicit phonological grounding. To address this gap, Williams et al. (2017) operationalized phonological similarity by quantifying shared manual parameters, demonstrating psycholinguistic correlates of such similarity measures. Further advances have integrated phonological structure into neural models: Tavella et al. (2022) introduced the WLASL-LEX dataset annotated with phonological properties and showed that graph-based neural networks can recognize phonological features at scale; Rodriguez et al. (2023) proposed a phonological distance metric (“phdist”) over fourteen phonological specifications in NGT and used it to analyze deep sign embeddings; and Kezar et al. (2023) demonstrated that incorporating phonological representations improves isolated sign recognition performance on the Sem-Lex benchmark. These works highlight increased attention toward phonologically grounded representations in computational modelling of sign similarity.

Our work differs in that it focuses on symbolic phonological similarity across multiple sign languages using HamNoSys. By constructing large-scale similarity matrices and performing analytical studies, we aim to uncover structural patterns in sign language phonology that are both linguistically meaningful and computationally tractable.

### 3 Dataset Construction

The dataset used in this study is derived from the publicly available **Dicta-Sign Language Resources** (Efthimiou et al., 2012), a multilingual repository of sign language data developed as part of the Dicta-Sign project. The

resource provides a curated list of over 1000 concepts, each annotated with corresponding signs and phonological representations in four European sign languages: British Sign Language (BSL), German Sign Language (DGS), French Sign Language (LSF), and Greek Sign Language (GSL).

For each of the four languages, we selected 1000 signs corresponding to a shared set of concepts. Each sign is associated with a HamNoSys transcription that encodes its phonological structure, including handshape, location, orientation, and movement. These symbolic representations serve as the foundation for computing phonological similarity.

To quantify phonological similarity, we compute the normalized Levenshtein distance (Yujian and Bo, 2007) between HamNoSys strings. Given two signs  $i$  and  $j$  with HamNoSys representations  $H_i$  and  $H_j$ , the similarity score  $S_{ij}$  is defined as:

$$S_{ij} = 1 - \frac{d_{\text{lev}}(H_i, H_j)}{\max(|H_i|, |H_j|)} \quad (1)$$

where  $d_{\text{lev}}$  denotes the Levenshtein edit distance between the two strings, and  $|H_i|$  is the length of the string. This results in a similarity score in the range  $[0, 1]$ , where 1 indicates identical phonological structure.

For each language, we construct a  $1000 \times 1000$  similarity matrix  $\mathbf{S}^{(l)}$  capturing all pairwise phonological similarities. These matrices form the basis for the analytical tasks described in the next section.

## 4 Analysis and Results

We present a comprehensive analysis of phonological similarity patterns within and across four sign languages: British Sign Language (BSL), German Sign Language (DGS), French Sign Language (LSF), and Greek Sign Language (GSL). Each language’s dataset consists

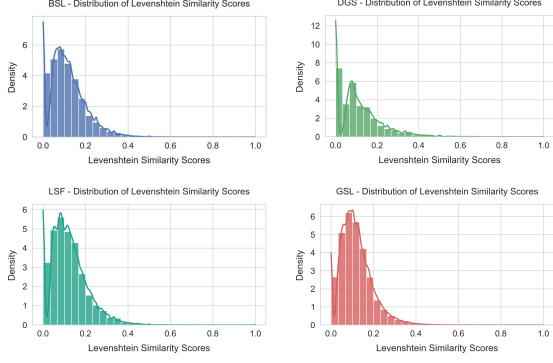


Figure 3: Intra-language phonological similarity distributions for BSL, DGS, LSF, and GSL.

of 1000 signs, and a  $1000 \times 1000$  similarity matrix was computed using normalized edit distance over HamNoSys representations.

#### 4.1 Intra-Language Similarity Distributions

Figure 3 shows the distribution of similarity scores within each language. All distributions are left-skewed, indicating that most sign pairs are moderately dissimilar, with a smaller proportion of highly similar signs. Notably, DGS and LSF exhibit slightly higher concentrations of high-similarity pairs, suggesting more phonologically compact lexicons.

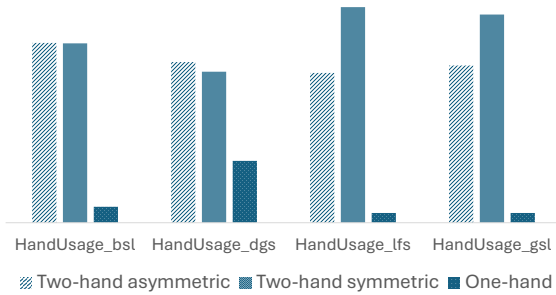


Figure 4: Intra-language hand-usage frequency for BSL, DGS, LSF, and GSL.

#### 4.2 Phonological Clustering

To explore the internal structure of each language’s phonological space, we applied hierarchical clustering on each similarity matrix. Figure 5 show the resulting clusters (only for the sake of clear visualization, we show the clustering results on a  $100 \times 100$  subset). Clear block structures emerge, indicating the presence of phonological families—groups of signs

that share similar handshapes, locations, or movements.

Language	Mean	Std Dev	Min	Max
BSL	0.115	0.084	0.000	1.000
DGS	0.112	0.106	0.000	1.000
LSF	0.122	0.085	0.000	1.000
GSL	0.118	0.075	0.000	1.000

Table 1: Summary statistics of phonological similarity scores

Table 1 present the mean, standard deviation, and range of similarity scores for each language. LSF and GSL show the highest average similarity, while DGS exhibits the widest spread, indicating greater phonological diversity.

#### 4.3 One-hand vs Two-hand Sign Analysis

The Figure 4 presents the distribution of signs based on hand usage—categorized into one-handed signs, two-handed symmetric signs, and two-handed asymmetric signs—across British Sign Language (BSL), German Sign Language (DGS), French Sign Language (LSF), and Greek Sign Language (GSL). A clear trend emerges: all four languages predominantly use two-handed signs, with symmetric and asymmetric configurations being nearly equally represented. For instance, BSL shows a near-even split between symmetric (461) and asymmetric (462) two-handed signs, while LSF and GSL lean slightly toward symmetric usage. In contrast, one-handed signs are significantly less frequent, especially in LSF and GSL (only 25 each), whereas DGS shows a relatively higher count (159), suggesting a greater preference or flexibility for one-handed articulation in German Sign Language. This distribution highlights both universal tendencies and language-specific variations in sign formation, which may reflect linguistic, cultural, or ergonomic factors influencing sign language structure.

#### 4.4 Phonological analysis across language

Table 2 showing the top 5 handshapes, movements, non-manual signs, and sign locations across British Sign Language (BSL), German Sign Language (DGS), French Sign Language



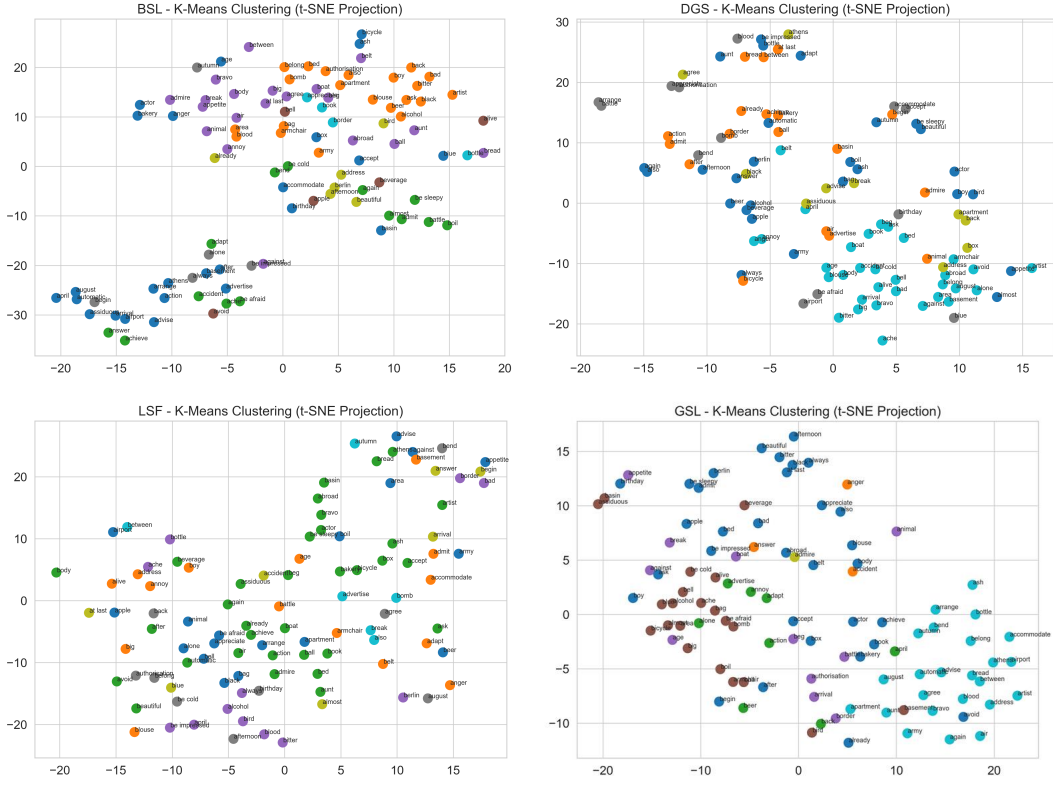


Figure 5: Phonological similarity using k-means clustering across languages. Same colored points belong to same cluster.

(LSF), and Greek Sign Language (GSL). The data reveals both shared and language-specific preferences in phonological features across the four sign languages.

**Handshapes (HS):** The handshapes hamthumboutmod, hamflathand, and hamfinger2 appear consistently across all languages with hamthumbacrossmod appearing in BSL, LSF, and GSL, indicating a core set of frequently used configurations. BSL and LSF favor hamthumboutmod most prominently. DGS and GSL show a high preference for hamflathand. DGS uniquely includes hamfist in its top 5, suggesting a more frequent use of closed hand configurations.

**Movements (MOV):** Universal dominance: hamrepeatfromstart is the most frequent movement across all four languages, highlighting repetition as a common linguistic strategy. hammoved and hammoveo are consistently present, but their ranks vary. GSL shows a higher frequency of hamfast, possibly reflecting a faster signing tempo or stylistic variation. LSF and GSL include hamrepeatfromstartseveral, suggesting more com-

plex repetition patterns.

**Non-Manual Features (NMA):** BSL, LSF, and GSL emphasize hamshoulders and hamchest, indicating upper torso involvement. Moreover, GSL shows the highest counts for hamchest and hamshoulders, suggesting strong reliance on torso-based non-manual cues. DGS has lower counts overall and includes hamchin and hamhead, pointing to more facial involvement.

**Sign Locations (LOC):** hambetween (likely referring to the space between hands or between signer and viewer) is dominant in BSL and GSL, suggesting spatial articulation is central. DGS and LSF favor hampalml and hampalmd, indicating signs are often articulated near the palm or lower body. hamsymmlr and hamextfingeru appear across multiple languages, reflecting symmetrical and extended finger placements.

These patterns suggest that while there is a shared phonological core across sign languages—especially in handshapes and movements suggesting inter-language phonological similarity—each language exhibits

unique tendencies in non-manual features and spatial articulation. This supports the idea that sign languages, though visually grounded, are shaped by distinct linguistic and cultural norms.

Category	Features	BSL	DGS	LSF	GSL
HS	hamthumboutmod	666	216	543	646
	hamthumbacrossmod	464	0	490	631
	hamflathand	448	343	350	656
	hamfinger2	345	219	326	445
	hamfinger2345	298	0	0	401
	hamfist	0	208	0	0
	hamfingerstraightmod	0	269	289	0
MOV	hamrepeatfromstart	232	298	260	232
	hammoved	210	211	187	216
	hammoveo	134	105	115	165
	hammover	134	103	115	0
	hammoveu	109	0	0	0
	hammoveor	0	81	0	0
	repeatfromstartseveral	0	0	125	160
NMA	hamshoulders	341	91	376	456
	hamchest	259	81	377	495
	hamshouldertop	99	0	115	135
	hamneck	75	0	0	0
	hamstomach	66	0	48	75
	hamchin	0	58	0	0
	hamhead	0	27	0	0
LOC	hamlips	0	25	56	49
	hambetween	903	0	318	937
	hampalml	543	401	494	543
	hampalmd	509	317	498	491
	hamextfingeru	462	0	448	413
	hamsymmmlr	375	330	320	552
	hamextfingerol	0	336	0	0

Table 2: Frequency distribution of the most frequent sign language phonological features across languages.

#### 4.5 Intra-Phonological Co-occurrences

Insights derived from the co-occurrence (pointwise mutual information, PMI) table across four sign languages—BSL, DGS, LSF, GSL—focusing on phonological feature interactions reveals that Across all languages, high PMI values are observed between compound or modified handshapes (e.g., hamthumboutmod, hamceeopen, hamfingerside), indicating that these handshapes frequently co-occur in signs with complex articulatory configurations. DGS shows strong co-occurrence between hamfingerpad and hamthumbball (PMI =  $5.59 \pm 0.035$ ), suggesting a preference for precision grip-like configurations. LSF and GSL both show high PMI between hamceeopen and hamfingerside, indicating a shared structural tendency toward open, lateral hand articulations. GSL also exhibits strong co-occurrence

between hamceeopen and hamfingernail, hinting at a visual emphasis on finger extension and orientation.

The highest PMI values in hand-location category are found in LSF (hamextfingerdi-hamextfingeri, PMI =  $6.91 \pm 0.043$ ) and DGS (hamarmextended-hamextfingerdi, PMI =  $5.91 \pm 0.037$ ), suggesting frequent use of extended arm and finger configurations in spatial articulation. GSL shows strong co-occurrence between hamarmextended and hamextfingerir (PMI =  $5.59 \pm 0.035$ ), indicating a preference for distal articulation zones. Across all languages, combinations involving hamhandback, hamwristback, and hamextfinger variants suggest a consistent use of backward or lateral orientations in sign production.

Movement features show the highest PMI values overall, with DGS (hamclockdr-hamclocku, PMI =  $9.91 \pm 0.061$ ) and GSL (hamcircleil-hamstirccw, PMI =  $8.91 \pm 0.055$ ) demonstrating highly structured temporal motion patterns. Circular and clock-like movements (hamcircle, hamclock, hamstir) dominate across all languages, indicating a shared visual rhythm in sign articulation. These patterns suggest that cyclic and directional movements are central to sign semantics and may serve as phonological markers for verb or action-related signs.

Non-manual features show lower PMI values overall, indicating more diffuse or context-dependent usage. BSL shows the strongest co-occurrence (hameyes-hamnose, PMI =  $4.30 \pm 0.027$ ), suggesting facial articulation plays a significant role in sign contrast. LSF and GSL show moderate co-occurrence between hamchin, hamhead, and hamneck, pointing to a layered use of facial and neck gestures. DGS shows relatively low PMI values, possibly reflecting a more manual-centric phonological system or less reliance on facial features.

#### 4.6 Inter-Phonological Co-occurrence

As depicted in Table 3 We also analyze how different category of phonological features interact among themselves. For example, how handshapes interact with movements, or locations in a particular language’s signing space. We found BSL and GSL show higher co-occurrence between hamthumboutmod and

	BSL		DGS		LSF		GSL	
Type	Pairs	Freq	Pairs	Freq	Pairs	Freq	Pairs	Freq
HS +	hamthumboutmod-hampalml	259	hamflathand-hamextfingerol	122	hamthumboutmod-hampalml	211	hamthumboutmod-hambetween	259
LOC	hamthumboutmod-hambetween	249	hamflathand-hampalml	107	hamthumbacrossmod-hampalmd	180	hamthumboutmod-hamsymmlr	248
	hamthumboutmod-hamsymmlr	227	hamflathand-hampalmd	102	hamthumboutmod-hampalmd	172	hamflathand-hambetween	229
	hamthumboutmod-hampalmd	208	hamfist-hampalml	97	hamthumbacrossmod-hamextfingeru	165	hamflathand-hamsymmlr	229
	hamthumbacrossmod-hamextfingeru	190	fingerstraightmod-hamsymmlr	97	hamthumbacrossmod-hampalml	143	hamthumboutmod-hampalml	225
HS+	hamthumboutmod-hamrepeatfromstart	111	hamfingerstraightmod-hamrepeatfromstart	89	hamthumboutmod-hamrepeatfromstart	104	hamthumboutmod-hamrepeatfromstart	90
NMA	hamthumboutmod-hammoved	99	hamfist-hamrepeatfromstart	75	hamthumbacrossmod-hamrepeatfromstart	91	hamthumbacrossmod-hamrepeatfromstart	87
	hamthumbacrossmod-hamrepeatfromstart	95	hamflathand-hamrepeatfromstart	65	hamfingerstraightmod-hamrepeatfromstart	79	hamflathand-hamrepeatfromstart	86
	hamflathand-hamrepeatfromstart	75	fingerstraightmod-hammoved	62	hamthumboutmod-hammoved	79	hamthumboutmod-hamfast	84
	hamfinger2-hamrepeatfromstart	67	hamfinger2-hamrepeatfromstart	62	hamthumbacrossmod-hammoved	75	hamthumboutmod-hammoved	79
HS+	hamthumboutmod-hamshoulders	166	hamthumboutmod-hamchest	29	hamthumboutmod-hamchest	163	hamthumbacrossmod-hamshoulders	181
NMA	hamthumboutmod-hamchest	139	hamfinger2345-hamchest	25	hamthumboutmod-hamshoulders	137	hamthumboutmod-hamchest	173
	hamthumbacrossmod-hamshoulders	101	fingerstraightmod-hamshoulders	22	hamthumbacrossmod-hamshoulders	133	hamflathand-hamchest	166
	hamflathand-hamshoulders	99	hamflathand-hamshoulders	22	hamflathand-hamchest	113	hamthumboutmod-hamshoulders	159
	hamflathand-hamchest	90	hamthumboutmod-hamshoulders	21	hamthumbacrossmod-hamchest	105	hamthumbacrossmod-hamchest	155
MOV+	hamrepeatfromstart-hampalml	119	hamrepeatfromstart-hampalml	124	hamrepeatfromstart-hampalml	108	hamrepeatfromstart-hambetween	135
LOC	hamrepeatfromstart-hambetween	113	hamrepeatfromstart-hamsymmlr	101	hamrepeatfromstart-hamextfingeru	100	hamrepeatfromstart-hamsymmlr	124
	hammoved-hamsymmlr	90	hammoved-hampalml	92	hamrepeatfromstart-hampalmd	92	hamrepeatfromstart-hampalml	113
	hamrepeatfromstart-hamextfingeru	88	hamrepeatfromstart-hamextfingerol	90	hammoved-hampalml	83	hamfast-hambetween	113
	hammoved-hampalml	88	hammoved-hamextfingerol	79	hammoved-hampalmd	78	hammoved-hambetween	112
MOV+	hamrepeatfromstart-hamshoulders	76	hammoved-hamshoulders	35	hammoved-hamshoulders	81	hammoved-hamchest	93
NMA	hammoved-hamshoulders	56	hamrepeatfromstart-hamshoulders	20	hamrepeatfromstart-hamshoulders	78	hamrepeatfromstart-hamchest	88
	hammoved-hamchest	53	hamrepeatfromstart-hamchin	20	hammoved-hamchest	73	hammoved-hamshoulders	81
	hamrepeatfromstart-hamchest	53	hamrepeatfromstart-hamchest	20	hamrepeatfromstart-hamchest	65	hamhalt-hamshoulders	80
	hammover-hamshoulders	34	hammoved-hamchest	13	repeatfromstartseveral-hamshoulders	53	hamrepeatfromstart-hamshoulders	77
NMA+	hamshoulders-hambetween	168	hamshoulders-hamsymmlr	46	hamshoulders-hampalml	148	hamchest-hambetween	260
LOC	hamshoulders-hampalml	151	hamchin-hampalml	42	hamshoulders-hampalmd	146	hamchest-hamsymmlr	236
	hamshoulders-hamsymmlr	142	hamchin-hamextfingerul	37	hamchest-hampalmd	140	hamshoulders-hamsymmlr	235
	hamchest-hambetween	137	hamchest-hampalml	35	hamchest-hampalml	129	hamshoulders-hambetween	229
	hamshoulders-hampalmd	121	hamshoulders-hamextfingeruo	34	hamchest-hamextfingero	122	hamchest-hampalml	206

Table 3: Frequency distributions of co-occurrences of phonological features across different sign languages.

spatial locations like `hampalml`, `hambetween`, and `hamsymmlr`, suggesting that this handshape is highly versatile and frequently used in central signing space. DGS favors combinations like `hamflathand-hamextfingerol` and `hamflathand-hampalml`, indicating a preference for flat hand configurations in extended or lateral orientations. LSF shows similar patterns to BSL, with `hamthumboutmod` and `hamthumbacrossmod` frequently paired with `hampalml` and `hampalmd`, reflecting a balanced use of thumb-based handshapes in mid-body locations.

Across all languages, `hamrepeatfromstart` is the most frequent movement paired with dominant handshapes (`hamthumboutmod`, `hamthumbacrossmod`, `hamflathand`), reinforcing its role as a core phonological motion. BSL and LSF show strong pairings of `hamthumboutmod` with both `hamrepeatfromstart` and `hammoved`, suggesting a dynamic use of thumb-based signs. GSL includes `hamfast` in its top co-occurrences, indicating a tendency toward rapid articulation in certain handshape-movement combinations.

BSL, LSF, and GSL show strong co-occurrence between `hamthumboutmod` and upper-torso cues, while GSL uniquely favors `hamthumbacrossmod-hamshoulders` and `hamflathand-hamchest`, reflecting rich manual-non-manual integration. DGS, with lower overall frequencies and modest pairings like `hamfinger2345-hamchest`, suggests a more manual-centric system.

`hamrepeatfromstart` usually co-occurs with `hampalml`, `hambetween`, and `hamsymmlr` across all languages, confirming its central role in spatially anchored sign articulation. GSL shows strong pairings of `hamfast` with `hambetween`, suggesting a preference for fast, centrally located signs. BSL, DGS and LSF include extended finger orientations (like `hamextfingeru`, `hamextfingerol`) in frequent pairings, indicating a nuanced use of directional movement.

Co-occurrence between `hamrepeatfromstart` and `hamshoulders` or `hamchest` is common in BSL, LSF, and GSL, reinforcing the idea that repetitive movements are often accompanied by expressive non-manual cues. GSL shows the highest integration, with `hammoved-hamchest` and `hamrepeatfromstart-hamchest`

appearing frequently, suggesting a significant coupling of motion and torso-based expression. DGS shows lower frequencies and more facial-centric pairings (e.g., `hamrepeatfromstart-hamchin`), indicating a different balance of articulatory features.

BSL and GSL show high co-occurrence between `hamshoulders` and `hambetween`, suggesting that upper-body non-manual features are often used in central signing space. LSF shows high frequencies for `hamchest-hampalml` and `hamshoulders-hampalmd`, indicating a preference for mid-body articulation zones. DGS includes more facial and lateral pairings (e.g., `hamchin-hampalml`, `hamchin-hamextfingerol`), reflecting a more distributed use of non-manual features.

In summary we observe that BSL and GSL exhibit strong centralization in signing space, with frequent use of `hambetween` and upper torso non-manuals. DGS shows a more distributed and facially oriented phonological structure, with lower integration of non-manuals and more lateral articulations. LSF balances manual and non-manual features with a preference for mid-body locations and thumb-based handshapes. These patterns reveal universal tendencies and regional variations in phonological feature co-occurrence, offering insights into the structural and cultural shaping of sign languages.

## 5 Conclusion

We analyzed phonological similarity across four sign languages using HamNoSys-based symbolic representations. By comparing 1000 signs per language, we identified consistent patterns in handshape, movement, and spatial usage, along with notable differences in non-manual features and articulation styles. Co-occurrence analysis revealed strong intra- and inter-feature dependencies, suggesting both universal phonological structures and geo-linguistic variation. LSF and DGS show higher internal consistency, and sign clustering reveals phonological families—laying the groundwork for multilingual sign language modeling and cross-linguistic phonological transfer. All these observations are based on raw frequency counts; formal statistical testing will be included in future work.



## References

- Suvarna Rajesh Bhagwat, RP Bhavsar, and BV Pawar. 2024. Marathi to indian sign language machine translation. *ACM Transactions on Asian and Low-Resource Language Information Processing*.
- Diane Brentari. 1998. *A prosodic model of sign language phonology*. Mit Press.
- Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7784–7793.
- Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020. Sign language transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10023–10033.
- Yutong Chen, Ronglai Zuo, Fangyun Wei, Yu Wu, Shujie Liu, and Brian Mak. 2022. Two-stream network for sign language recognition and translation. *Advances in Neural Information Processing Systems*, 35:17043–17056.
- Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, and Richard Bowden. 2017. Subunets: End-to-end hand shape and continuous sign language recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 3056–3065.
- OA Crasborn and IEP Zwitserlood. 2008. Annotation of video data in the corpus ngt.
- Eleni Efthimiou, Stavroula-Evita Fontinea, Thomas Hanke, John Glauert, Rihard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and François Goudenove. 2010. Dicta-sign—sign language recognition, generation and modelling: a research effort with applications in deaf communication. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 80–83.
- Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and François Lefebvre-Albaret. 2012. The dicta-sign wiki: Enabling web communication for the deaf. In *International conference on computers for handicapped persons*, pages 205–212. Springer.
- Eleni Efthimiou, Stavroula-Evita Fotinea, Christian Vogler, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and Jérémie Segouat. 2009. Sign language recognition, generation, and modelling: A research effort with applications in deaf communication. In *International Conference on Universal Access in Human-Computer Interaction*, pages 21–30. Springer.
- Thomas Hanke. 2004. Hamnosys—representing sign language data in language resources and language processing contexts. In *sign-lang@ LREC 2004*, pages 1–6. European Language Resources Association (ELRA).
- Thomas Hanke. 2010. Hamnosys—hamburg notation system for sign languages. *Institute of German Sign Language*, Accessed in, 7.
- Lee Kezar, Jesse Thomason, Naomi Caselli, Zed Sehyr, and Elana Pontecorvo. 2023. The semlex benchmark: Modeling asl signs and their phonemes. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–10.
- Gabriele Langer, Susanne König, and Silke Matthes. 2014. Compiling a basic vocabulary for german sign language (dgs)—lexicographic issues with a focus on word senses. In *Proceedings of the XVI EURALEX International Congress: The User in Focus*, pages 767–786.
- Sara Morrissey. 2008. *Data-driven machine translation for sign languages*. Ph.D. thesis, Dublin City University.
- Carolina Neves, Luísa Coheur, and Hugo Nicolau. 2020. Hamnosys2sigml: translating hamnosys into sigml. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6035–6039.
- Ellen Ormel, Daan Hermans, Harry Knoors, Angelique Hendriks, and Ludo Verhoeven. 2010. Phonological activation during visual word recognition in deaf and hearing children. *Journal of Speech, Language, and Hearing Research*, 53(4):801–820.
- Siegmund Prillwitz and Hamburg Zentrum für Deutsche Gebärdensprache und Kommunikation Gehörloser. 1989. *Hamnosys: Version 2.0; hamburg notation system for sign languages; an introductory guide*. Signum-Verlag.
- J Martinez Rodriguez, Martha Larson, and L ten Bosch. 2023. Exploring the importance of sign language phonology for a deep neural network.
- Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2020. Adversarial training for multi-channel sign language production. *arXiv preprint arXiv:2008.12405*.
- Samar Sinha. 2009. *A grammar of Indian sign language*. Ph.D. thesis, PhD dissertation, Jawaharlal Nehru University, New Delhi, India.
- William C Stokoe. 1960. Sign language structure (studies in linguistics. *Occasional paper*, 8.

- Stephanie Stoll, Necati Cihan Camgoz, Simon Hadfield, and Richard Bowden. 2020. Text2sign: towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*, 128(4):891–908.
- Sugandhi, Parteek Kumar, and Sanmeet Kaur. 2020. Sign language generation system based on indian sign language grammar. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 19(4):1–26.
- Federico Tavella, Viktor Schlegel, Marta Romeo, Aphrodite Galata, and Angelo Cangelosi. 2022. Wlasl-lex: a dataset for recognising phonological properties in american sign language. *arXiv preprint arXiv:2203.06096*.
- Harry Walsh, Ben Saunders, and Richard Bowden. 2022. Changing the representation: Examining language representation for neural sign language production. *arXiv preprint arXiv:2210.06312*.
- Joshua T Williams, Adam Stone, and Sharlene D Newman. 2017. Operationalization of sign language phonological similarity and its effects on lexical access. *The Journal of Deaf Studies and Deaf Education*, 22(3):303–315.
- Li Yujian and Liu Bo. 2007. A normalized levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):1091–1095.

## A HamNoSys Explanations

Table 4 shows the detailed explanation and category of what each HamNoSys symbol means along with its unicode (Neves et al., 2020).

HamNoSys	Unicode	Explanation	Category
hamspace	0020	Space separator (used to separate symbols or words).	Other
hamexclaim	0021	Punctuation marker (e.g., exclamation, comma, full stop, question) for transcriptions.	Other
hamcomma	002C	Punctuation marker (e.g., exclamation, comma, full stop, question) for transcriptions.	Other
hamfullstop	002E	Punctuation marker (e.g., exclamation, comma, full stop, question) for transcriptions.	Other
hamquery	003F	Punctuation marker (e.g., exclamation, comma, full stop, question) for transcriptions.	Other
hamaltbegin	007B	Alternative/parenthetical markers used to bracket alternate transcriptions or metadata.	Other
hammetaalt	007C	Alternative/parenthetical markers used to bracket alternate transcriptions or metadata.	Other
hamaltend	007D	Alternative/parenthetical markers used to bracket alternate transcriptions or metadata.	Other
hamfist	E000	Fist handshape (closed hand).	Hand Shapes
hamflathand	E001	Flat handshape (palm and fingers extended and close together, like a flat hand).	Hand Shapes
hamfinger2	E002	Two-finger configuration (usually index+middle extended).	Hand Shapes
hamfinger23	E003	Two adjacent fingers extended (index+middle) in non-spread configuration.	Hand Shapes
hamfinger23spread	E004	Two adjacent fingers extended and spread apart (index+middle spread).	Hand Shapes
hamfinger2345	E005	Fingers 2-5 extended (index through little finger), excluding thumb.	Hand Shapes
hampinch12	E006	Pinch-like handshape (thumb and one or more fingers pinching together).	Hand Shapes
hampinchall	E007	Pinch-like handshape (thumb and one or more fingers pinching together).	Hand Shapes
hampinch12open	E008	Pinch-like handshape (thumb and one or more fingers pinching together).	Hand Shapes
hamcee12	E009	C-shaped hand configuration (curved hand like letter 'C').	Hand Shapes
hamceeall	E00A	C-shaped hand configuration (curved hand like letter 'C').	Hand Shapes
hamceeopen	E00B	C-shaped hand configuration (curved hand like letter 'C').	Hand Shapes
hamthumboutmod	E00C	Thumb pointed outwards (thumb extended away from palm) a thumb position modifier.	Hand Shapes
hamthumbacrossmod	E00D	Thumb lying across the palm or fingers a thumb position modifier.	Hand Shapes
hamthumbopenmod	E00E	Thumb held open (not tucked in) modifier for thumb openness.	Hand Shapes
hamfingerstraightmod	E010	Handshape specifying particular fingers extended.	Hand Shapes
hamfingerbendmod	E011	Handshape specifying particular fingers extended.	Hand Shapes
hamfingerhookmod	E012	Handshape specifying particular fingers extended.	Hand Shapes
hamdoublebent	E013	Modifier for double-bent or double-hooked finger shapes (complex finger bend).	Hand Shapes
hamdoublehooked	E014	Modifier for double-bent or double-hooked finger shapes (complex finger bend).	Hand Shapes
hamextfingeru	E020	Finger direction marker extended finger points up (used to show finger orientation).	Location/Orientation
hamextfingerur	E021	Finger direction marker extended finger points up-right (used to show finger orientation).	Location/Orientation

HamNoSys	Unicode	Explanation	Category
hamextfingerr	E022	Finger direction marker extended finger points right (used to show finger orientation).	Location/Orientation
hamextfingerdr	E023	Finger direction marker extended finger points down-right (used to show finger orientation).	Location/Orientation
hamextfingerd	E024	Finger direction marker extended finger points down (used to show finger orientation).	Location/Orientation
hamextfingerdl	E025	Finger direction marker extended finger points down-left (used to show finger orientation).	Location/Orientation
hamextfingerl	E026	Finger direction marker extended finger points left (used to show finger orientation).	Location/Orientation
hamextfingerul	E027	Finger direction marker extended finger points up-left (used to show finger orientation).	Location/Orientation
hamextfingerol	E028	Finger direction marker extended finger points out-left (used to show finger orientation).	Location/Orientation
hamextfingero	E029	Finger direction marker extended finger points out/away (used to show finger orientation).	Location/Orientation
hamextfingeror	E02A	Finger direction marker extended finger points out-right (used to show finger orientation).	Location/Orientation
hamextfingeril	E02B	Finger direction marker extended finger points in-left (used to show finger orientation).	Location/Orientation
hamextfingeri	E02C	Finger direction marker extended finger points in/toward (used to show finger orientation).	Location/Orientation
hamextfingerir	E02D	Finger direction marker extended finger points in-right (used to show finger orientation).	Location/Orientation
hamextfingerui	E02E	Finger direction marker extended finger points up-in (used to show finger orientation).	Location/Orientation
hamextfingerdi	E02F	Finger direction marker extended finger points down-in (used to show finger orientation).	Location/Orientation
hamextfingerdo	E030	Finger direction marker extended finger points down-out (used to show finger orientation).	Location/Orientation
hamextfingeruo	E031	Finger direction marker extended finger points up-out (used to show finger orientation).	Location/Orientation
hampalmu	E038	Palm orientation indicator (which way the palm faces: up/down/left/right or variants).	Location/Orientation
hampalmur	E039	Palm orientation indicator (which way the palm faces: up/down/left/right or variants).	Location/Orientation
hampalmr	E03A	Palm orientation indicator (which way the palm faces: up/down/left/right or variants).	Location/Orientation
hampalmdr	E03B	Palm orientation indicator (which way the palm faces: up/down/left/right or variants).	Location/Orientation
hampalmd	E03C	Palm orientation indicator (which way the palm faces: up/down/left/right or variants).	Location/Orientation
hampalmdl	E03D	Palm orientation indicator (which way the palm faces: up/down/left/right or variants).	Location/Orientation
hampalmll	E03E	Palm orientation indicator (which way the palm faces: up/down/left/right or variants).	Location/Orientation
hampalmul	E03F	Palm orientation indicator (which way the palm faces: up/down/left/right or variants).	Location/Orientation
hamhead	E040	Head (general) indicates head as location or non-manual articulator.	Non-Manual Features
hamheadtop	E041	Top of the head (specific location).	Non-Manual Features
hamforehead	E042	Forehead (location; often for non-manuals like eyebrow movement).	Other
hameyebrows	E043	Eyebrows (non-manual feature raise/lower etc).	Non-Manual Features
hameyes	E044	Eyes (gaze direction or eye activity).	Non-Manual Features
hamnose	E045	Nose (facial location).	Non-Manual Features
hamnostrils	E046	Nostrils (specific part of nose).	Other
hamear	E047	Ear (location).	Other
hamearlobe	E048	Earlobe (location).	Other
hamcheek	E049	Cheek (facial location).	Other
hamlips	E04A	Lips / mouth area (non-manual/mouthings).	Non-Manual Features
hamtongue	E04B	Tongue (mouth articulation reference).	Other
hamteeth	E04C	Teeth (mouth reference).	Other

HamNoSys	Unicode	Explanation	Category
hamchin	E04D	Chin (location reference).	Non-Manual Features
hamunderchin	E04E	Under-chin (location).	Other
hamneck	E04F	Neck (location / non-manual).	Non-Manual Features
hamshouldertop	E050	Top of the shoulder (location).	Non-Manual Features
hamshoulders	E051	Shoulders (body reference).	Non-Manual Features
hamchest	E052	Chest (body location).	Non-Manual Features
hamstomach	E053	Stomach/abdomen area (location).	Non-Manual Features
hambelowstomach	E054	Lower stomach/abdomen (location).	Other
hamlrbeside	E058	Location: left/right beside (side position next to body).	Other
hamlrat	E059	Location: left/right at (side location marker) indicates side-relative placement.	Other
hamcoreftag	E05A	Coreference tag (used for referencing another element or anchor in notation).	Location/Orientation
hamcorefref	E05B	Coreference reference (points to a previously defined anchor or location).	Location/Orientation
hamneutralspace	E05F	Neutral signing space in front of the signer (space away from body).	Location/Orientation
hamupperarm	E060	Upper arm (location reference).	Other
hamelbow	E061	Elbow (location).	Other
hamelbowinside	E062	Inner side of the elbow (specific location).	Other
hamlowerarm	E063	Lower arm / forearm (location).	Other
hamwristback	E064	Back of the wrist (location).	Location/Orientation
hamwristpulse	E065	Wrist pulse area (location).	Location/Orientation
hamthumbball	E066	Bulbous part of thumb (thumb pad/ball) used as a location reference.	Hand Shapes
hampalm	E067	Palm orientation indicator (which way the palm faces: up/down/left/right or variants).	Location/Orientation
hamhandback	E068	Back of hand (dorsal side).	Location/Orientation
hamthumbside	E069	Thumb-related handshape or modifier.	Hand Shapes
hampinkyside	E06A	Pinky-side (ulnar side) of hand.	Location/Orientation
hamthumb	E070	Thumb-related handshape or modifier.	Hand Shapes
hamindexfinger	E071	Index finger (reference) used as location/orientation reference.	Other
hammiddlefinger	E072	Middle finger used as location/orientation reference.	Other
hamringfinger	E073	Ring finger used as location/orientation reference.	Other
hampinky	E074	Little finger / pinky used as location/orientation reference.	Location/Orientation
hamfingertip	E075	Handshape specifying particular fingers extended.	Hand Shapes
hamfingernail	E076	Handshape specifying particular fingers extended.	Hand Shapes
hamfingerpad	E077	Handshape specifying particular fingers extended.	Hand Shapes
hamfingermidjoint	E078	Handshape specifying particular fingers extended.	Hand Shapes
hamfingerbase	E079	Handshape specifying particular fingers extended.	Hand Shapes
hamfingerside	E07A	Handshape specifying particular fingers extended.	Hand Shapes
hamwristtopulse	E07C	Top/inner wrist near the pulse location reference.	Location/Orientation
hamwristtoback	E07D	From wrist top toward back of wrist orientation reference.	Location/Orientation
hamwristtothumb	E07E	Thumb-related handshape or modifier.	Location/Orientation
hamwristtopinky	E07F	Orientation/position from wrist toward pinky side.	Location/Orientation
hammoveu	E080	Hand movement direction: up (linear path in that direction).	Movements
hammoveur	E081	Hand movement direction: up-right (linear path in that direction).	Movements
hammovev	E082	Hand movement direction: right (linear path in that direction).	Movements
hammovedr	E083	Hand movement direction: down-right (linear path in that direction).	Movements
hammoved	E084	Hand movement direction: down (linear path in that direction).	Movements
hammovedl	E085	Hand movement direction: down-left (linear path in that direction).	Movements
hammovev	E086	Hand movement direction: left (linear path in that direction).	Movements
hammoveul	E087	Hand movement direction: up-left (linear path in that direction).	Movements
hammoveol	E088	Hand movement direction: out-left (linear path in that direction).	Movements
hammoveo	E089	Hand movement direction: out/away (linear path in that direction).	Movements
hammoveor	E08A	Hand movement direction: out-right (linear path in that direction).	Movements
hammoveil	E08B	Hand movement direction: in-left (linear path in that direction).	Movements
hammovei	E08C	Hand movement direction: in/toward (linear path in that direction).	Movements



HamNoSys	Unicode	Explanation	Category
hammoveir	E08D	Hand movement direction: in-right (linear path in that direction).	Movements
hammoveui	E08E	Hand movement direction: up-in (linear path in that direction).	Movements
hammovedi	E08F	Hand movement direction: down-in (linear path in that direction).	Movements
hammovedo	E090	Hand movement direction: down-out (linear path in that direction).	Movements
hammoveuo	E091	Hand movement direction: up-out (linear path in that direction).	Movements
hamcircleo	E092	Circular movement path around out/away (circle in that orientation).	Movements
hamcirclei	E093	Circular movement path around in/toward (circle in that orientation).	Movements
hamcircled	E094	Circular movement path around down (circle in that orientation).	Movements
hamcircleu	E095	Circular movement path around up (circle in that orientation).	Movements
hamcirclel	E096	Circular movement path around left (circle in that orientation).	Movements
hamcircler	E097	Circular movement path around right (circle in that orientation).	Movements
hamcircleul	E098	Circular movement path around up-left (circle in that orientation).	Movements
hamcircledr	E099	Circular movement path around down-right (circle in that orientation).	Movements
hamcircleur	E09A	Circular movement path around up-right (circle in that orientation).	Movements
hamcircledl	E09B	Circular movement path around down-left (circle in that orientation).	Movements
hamcircleol	E09C	Circular movement path around out-left (circle in that orientation).	Movements
hamcircleir	E09D	Circular movement path around in-right (circle in that orientation).	Movements
hamcircleor	E09E	Circular movement path around out-right (circle in that orientation).	Movements
hamcircleil	E09F	Circular movement path around in-left (circle in that orientation).	Movements
hamcircleui	E0A0	Circular movement path around up-in (circle in that orientation).	Movements
hamcircledo	E0A1	Circular movement path around down-out (circle in that orientation).	Movements
hamcircleuo	E0A2	Circular movement path around up-out (circle in that orientation).	Movements
hamcircledi	E0A3	Circular movement path around down-in (circle in that orientation).	Movements
hamfingerplay	E0A4	Handshape specifying particular fingers extended.	Hand Shapes
hamnodding	E0A5	General HamNoSys element (specific meaning depends on context).	Other
hamswinging	E0A6	General HamNoSys element (specific meaning depends on context).	Movements
hamtwisting	E0A7	General HamNoSys element (specific meaning depends on context).	Movements
hamstircw	E0A8	General HamNoSys element (specific meaning depends on context).	Movements
hamstirccw	E0A9	General HamNoSys element (specific meaning depends on context).	Movements
hamreplace	E0AA	General HamNoSys element (specific meaning depends on context).	Other

HamNoSys	Unicode	Explanation	Category
hammovecross	E0AD	Hand movement direction: directional movement (linear path in that direction).	Movements
hammoveX	E0AE	Hand movement direction: directional movement (linear path in that direction).	Movements
hamnomotion	E0AF	General HamNoSys element (specific meaning depends on context).	Other
hamclocku	E0B0	Clockwise/counterclockwise circular motion indicated by clock position 'u'.	Movements
hamclockul	E0B1	Clockwise/counterclockwise circular motion indicated by clock position 'ul'.	Movements
hamclockl	E0B2	Clockwise/counterclockwise circular motion indicated by clock position 'l'.	Movements
hamclockdl	E0B3	Clockwise/counterclockwise circular motion indicated by clock position 'dl'.	Movements
hamclockd	E0B4	Clockwise/counterclockwise circular motion indicated by clock position 'd'.	Movements
hamclockdr	E0B5	Clockwise/counterclockwise circular motion indicated by clock position 'dr'.	Movements
hamclockr	E0B6	Clockwise/counterclockwise circular motion indicated by clock position 'r'.	Movements
hamclockur	E0B7	Clockwise/counterclockwise circular motion indicated by clock position 'ur'.	Movements
hamclockfull	E0B8	Full circular clockwise motion (full rotation).	Movements
hamarcl	E0B9	Short arced movement (a small curved path).	Movements
hamarcu	E0BA	Short arced movement (a small curved path).	Movements
hamarcr	E0BB	Short arced movement (a small curved path).	Movements
hamarcd	E0BC	Short arced movement (a small curved path).	Movements
hamwavy	E0BD	Wavy oscillating movement (smooth wave-like motion).	Movements
hamzigzag	E0BE	Zig-zag oscillating movement (sharp alternating motion).	Other
hamellipseh	E0C0	Elliptical (oval) movement path, specifying orientation of ellipse.	Movements
hamellipseur	E0C1	Elliptical (oval) movement path, specifying orientation of ellipse.	Movements
hamellipsev	E0C2	Elliptical (oval) movement path, specifying orientation of ellipse.	Movements
hamellipseul	E0C3	Elliptical (oval) movement path, specifying orientation of ellipse.	Movements
hamincreasing	E0C4	Movement or parameter increasing (e.g., amplitude growing).	Other
hamdecreasing	E0C5	Movement or parameter decreasing (e.g., amplitude shrinking).	Other
hamsmallmod	E0C6	Modifier: small (subtle / small-amplitude) movement.	Other
hamlargemod	E0C7	Modifier: large (wide / large-amplitude) movement.	Other
hamfast	E0C8	Modifier: fast speed.	Movements
hamslow	E0C9	Modifier: slow speed.	Movements
hamtense	E0CA	Modifier: tense or stiff quality of movement/hand.	Movements
hamrest	E0CB	Rest position (hold without motion).	Movements
hamhalt	E0CC	Abrupt stop / halt in motion.	Movements
hamclose	E0D0	Hand closing or coming together (close action).	Other
hamtouch	E0D1	Touch/contact action (hand touches another part).	Other
haminterlock	E0D2	Hands interlocking (fingers interlaced) action.	Other
hamcross	E0D3	Crossing hands or crossing motion/placement.	Other
hamarmextended	E0D4	Arm is extended away from body (extended-arm posture).	Location/Orientation
hambehind	E0D5	Placed or moved behind body or another body-part.	Other
hambrushing	E0D6	Brushing motion (light stroke across surface).	Other
hamrepeatfromstart	E0D8	Repetition operator indicates repeating the movement or sequence.	Movements
hamrepeatfromstartseveral	E0D9	Repetition operator indicates repeating the movement or sequence.	Movements
hamrepeatcontinue	E0DA	Repetition operator indicates repeating the movement or sequence.	Movements
hamrepeatcontinueseveral	E0DB	Repetition operator indicates repeating the movement or sequence.	Movements

HamNoSys	Unicode	Explanation	Category
hamrepeatreverse	E0DC	Repetition operator indicates repeating the movement or sequence.	Movements
hamalternatingmotion	E0DD	Alternating motion (hands or fingers alternate in action).	Movements
hamseqbegin	E0E0	Sequence/grouping marker: begins/ends a sequence, parallel group, or fusion of actions.	Other
hamseqend	E0E1	Sequence/grouping marker: begins/ends a sequence, parallel group, or fusion of actions.	Other
hamparbegin	E0E2	Sequence/grouping marker: begins/ends a sequence, parallel group, or fusion of actions.	Other
hamparend	E0E3	Sequence/grouping marker: begins/ends a sequence, parallel group, or fusion of actions.	Other
hamfusionbegin	E0E4	Sequence/grouping marker: begins/ends a sequence, parallel group, or fusion of actions.	Other
hamfusionend	E0E5	Sequence/grouping marker: begins/ends a sequence, parallel group, or fusion of actions.	Other
hambetween	E0E6	Spatial relation: between (e.g., movement or placement between hands or body parts).	Location/Orientation
hamplus	E0E7	Plus symbol: combines or adds elements (used in composite descriptions).	Other
hamsymmpar	E0E8	Symmetry operator: indicates two-handed symmetry (how attributes mirror across hands).	Location/Orientation
hamsymmlr	E0E9	Symmetry operator: indicates two-handed symmetry (how attributes mirror across hands).	Location/Orientation
hamnondominant	E0EA	Marker referring to the non-dominant hand (used to describe NDH behaviour).	Location/Orientation
hamnonipsi	E0EB	Marker meaning non-ipsilateral / opposite-side reference (side-related indicator).	Location/Orientation
hametc	E0EC	Placeholder: 'etc.' or miscellaneous/other elements not explicitly listed.	Other
hamorirelative	E0ED	Orientation/relative reference marker (indicates orientation relative to something else).	Location/Orientation
hammime	E0F0	Mime or pantomime marker indicates mimed action rather than lexical sign.	Non-Manual Features

Table 4: Explanations of HamNoSys symbols