# Decoding Sign Languages: The SL-FE Framework for Phonological Analysis and Automated Annotation

**Karahan Şahin, Kadir Gökgöz**
Bogazici University
Bogazici University. Department of Linguistics. 34342 Bebek/İstanbul. Turkey
{karahan.sahin, kadir.gokgoz}@boun.edu.tr

**Abstract**

SL-FE is a framework designed for the phonological representation of sign languages, bridging the gap between theoretical phonology and practical sign language annotation. SL-FE defines phonological information as a continuous signal from pose estimation information that enables not only the extraction of the comprehensive set of discrete phonological information but also provides a quantitative framework for theoretical analyses. By utilizing our framework, we conduct case studies to test empirical claims of feature dominance and symmetry on phonological complexity in Turkish Sign Language (TID). Only by defining a ranking function, we were able to classify these conditions with high lexical retrieval accuracy offering empirical evidence to support theoretical claims. The framework proves to be an essential tool for research in sign language linguistics.

**Keywords:** sign language phonology, automatic annotation, pose estimation

## 1. Introduction

The field of sign language research has seen considerable advancements in automatic annotation technologies, significantly enhancing the efficiency and accuracy of sign language recognition and translation. However, a gap persists in integrating theoretical phonological models into these frameworks. Traditional automatic annotation systems primarily focus on feature extraction, serving the immediate needs of recognition and translation without delving into the theoretical aspects of sign languages (Skobov and Lepage, 2020; Lucie Naert and Gibet, 2018; Gonzalez et al., 2012). While functional for specific applications, this approach overlooks the phonological information crucial for comprehensive linguistic analysis and understanding, with the .

In response to this need, our framework, **S**ign **L**anguage **F**eature **E**xtraction (SL-FE), emerges as a novel solution for the limitations of existing annotation systems. Unlike its predecessors, SL-FE is not merely an automatic annotation tool but a robust framework incorporating a continuous mathematical representation of phonological information specifically tailored for sign languages. Drawing upon prosodic models (Fenlon et al., 2017), SL-FE represents each phonological feature — finger selection, movement, and location information— through normalized feature-scoring methods. This method leverages pose-estimation technology to calculate the probability of feature occurrences, utilizing both orthogonal and angular distances between joints and normalizing these measurements according to the body proportions of the signer. Such an approach ensures that our scoring remains invariant to variations in signer and camera angles, providing a consistent and interpretable analysis of phonological features in continuous sign language videos as demonstrated in Figure 1.

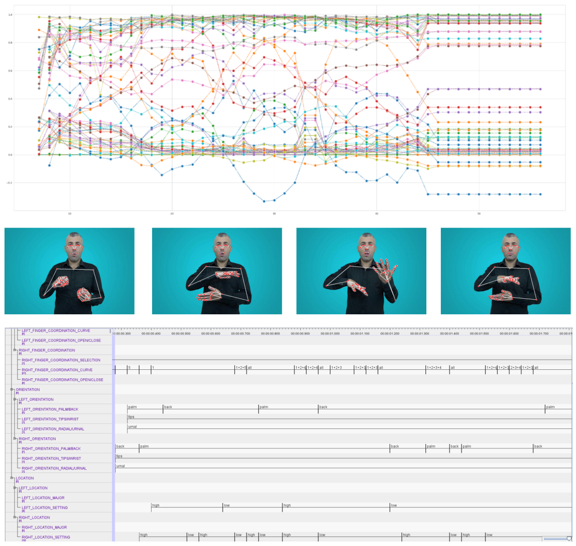

Figure 1: The pipeline of our framework for the lexical item "EVENT", in TID Sözlük. **The top side** is the cumulative plot of extracted continuous phonological information from the sign language video. **On the bottom side**, the annotations are exported to the ELAN interface after the classification pipeline is applied to the continuous feature set.

A significant achievement of our framework is its capacity to operationalize and validate typological claims within sign language research, such as Feature Dominance and Feature Symmetry (Bat-



*Proceedings of the 11th Workshop on the Representation and Processing of Sign Languages*, pages 335–342
25 May 2024. 

tison, 1978). By applying SL-FE to the TID (Turkish Sign Language) Sözlük Dictionary database, we have successfully computed the phonological complexity of isolated lexical items, offering empirical support for these theoretical constructs. This capability not only demonstrates the framework's analytical power but also contributes to the broader understanding of sign language phonology.

Furthermore, SL-FE is designed with accessibility in mind. The framework includes a user-friendly graphical user interface (GUI) that facilitates the viewing and exporting of annotations. This feature supports real-time and pre-recorded video analysis, making SL-FE a versatile tool for sign language research.

In summary, SL-FE yields a new line of methodology of sign language phonological research. Through its theory-driven approach to phonological feature representation and analysis, SL-FE addresses the limitations of previous annotation frameworks and paves the way for new directions in sign language research and applications.

## 2. Related Works and Theoretical Aspects

### 2.1. Related Works

Traditional automatic sign language annotation frameworks have largely been oriented with a focus on feature extraction utilized in recognition and translation models for classifying handshape (Mukushev et al., 2022; Lucie Naert and Gibet, 2018), detecting sign boundaries (Momeni et al., 2022) or the recognition of lexical items (Dreuw and Ney, 2008). In the automatic annotation process, these models either utilize RGB images from sign language videos or pose estimation information in the classification of the feature set. Although these methods introduce novel architectures for automation, they heavily rely on the prior annotations done for the training. Despite the practical utility of these systems, their contribution to theoretical linguistic inquiry is less pronounced. Theoretical research on sign language linguistics, focusing on systemic structure and function, requires a detailed interpretation of sign language as a linguistic system. Recent literature reflects an increasing interest in applying pose estimation techniques to provide quantitative insights into sign languages. These studies aim to bridge the gap between signs' physical articulation and linguistic implications (Chizhikova and Kimmelman, 2022; Ghaleb et al., 2024; Keleş et al., 2023; Stamp et al., 2022). This shift has been partly propelled by advancements in pose estimation technologies, enabling the articulatory components of sign languages to be quantitatively analyzed. In response to this growing interest, our framework, SL-FE, has been developed for both the automatic annotation of sign languages and the quantitative analysis of their phonological features concerning theoretical components of linguistic research.

### 2.2. Theoretical Aspects

Our framework's core innovation lies in its ability to provide a continuous representation of phonological features (i.e. Selected Fingers, Location, Orientation, and Movement) within a given sign language video. In the process of grounding our framework, we rely on the literature on theoretical aspects of sign language phonology where features are grouped into Inherent Features (IF) and Prosodic Features (PF) (Fenlon et al., 2017; Van der Hulst, 1993; Brentari, 1998). Namely, while Inherent Features provide a static snapshot within a single frame, the transition between position features (the thumb's interaction with the selected fingers, i.e. open to close or close to open), the transitions between settings in major locations (i.e. from proximal to distal, or from ipsilateral to contralateral), and changing orientation features (i.e. from palm to back of the hand, or from ulnar to radial parts of the hand) give rise to dynamic, Prosodic Features (PF). This treatment of phonological features and the appropriate mathematical modeling of these respective feature types are essential not only for extracting phonological information in a theoretically more informed manner from large corpora to be used in the different domains and tasks (i.e. sign segmentation and sign recognition in computer science), but they also provide a novel quantitative basis for theories of sign language phonology and typology.

## 3. Methodology

Our methodology focuses on four primary phonological feature types: Finger Selection, Orientation, Location, and Movement. Each feature type is extracted through a series of computational steps, leveraging pose-estimation technology and mathematical models to achieve a continuous and interpretable representation of sign language phonology regarding the variation and noise within sign language videos.

### 3.1. Pose Estimation

The preprocessing stage employs the Mediapipe hand and pose estimation models (Lugaresi et al., 2019), a tool for accurate human pose estimation. The model is critical to our framework, as it identifies and tracks various landmarks across the signer's body and hands in each frame, facilitating detailed phonological analysis. The landmarks include:

- **Hand Pose Landmarks**: Essential for analyzing movement, orientation, and finger selection, the model provides detailed information on the hand by identifying 21 joints per hand. Each joint is crucial for the in-depth examination of handshapes, movements, and orientations.
- **Pose Landmarks**: Primarily utilized for extracting location information, the model outputs 31 pose landmarks. These landmarks enable the framework to analyze how the signer's body interacts with space. These are either selected or generated according to the major and minor locations defined for sign languages.

Although we utilize the Mediapipe model in the current preprocessing due to its real-time processing and low CPU requirements, we are considering integrating the OpenPose framework (Cao et al., 2019). This prospective addition aims to broaden the framework's applicability and enhance its analytical depth to offer a more versatile and detailed tool for sign language research.

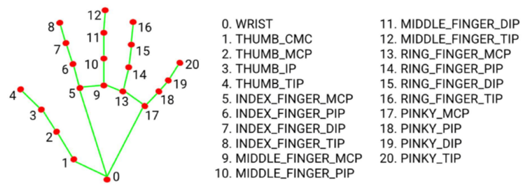


Figure 2: Hand Landmark list for left hand from Mediapipe

### 3.2. Finger Selection

Finger Selection is the first critical phonological feature our framework addresses. This process involves identifying key anchor points across each finger, focusing on four main inner joints for both hands (Joints 2-3, 6-7, 10-11, 14-15, 18-19 as designated in Figure 2). The angular distances between these joints are calculated to represent the fingers' selectional properties, such as curvature and contact points. The final feature values are obtained through min-max scaling of these angles across the video data, providing a continuous measure of finger selection within a normalized range of [0,1]. This normalization allows for a comparative analysis across different signers and sign languages, ensuring that the variations in individual signer's hand shapes do not skew the analysis.

$$FS(h,f) = \frac{1}{|J|}\sum_{p\in J}^{J}\frac{\angle(p_{j-1},p,p_{j+1})}{180} \quad (1)$$

In the finger selection feature extraction process defined in Eq. 1, $FS(h,f)$ serves as a quantifier for the selection state of a given finger $f$ on a given hand $h$. This mathematical representation is central to our framework, encapsulating the finger's posture in a numerical format. The set $J$ denotes the collection of joint indices, namely, Metacarpophalangeal (MCP), Proximal Interphalangeal (PIP), and Distal Interphalangeal (DIP) joints. These joints are pivot points that define the curvature and extension of each finger.

The formula calculates the normalized average angular difference between consecutive joints in the set $J$. For each joint $p$ in $J$, the angle $(p_{j-1}, p, p_{j+1})$ is computed, which measures the angle at joint $p$ formed by the line segments connecting it to its immediate neighboring joints $p_{j-1}$ and $p_{j+1}$. This angle is then normalized by dividing the angle by 180 degrees to scale the value between 0 and 1. Summing these normalized angles and dividing by the cardinality of the set $|J|$ gives us an average value, $FS(h,f)$, that represents the overall curvature of the finger.

The resulting feature score $FS$ is then categorized into one of three states based on its value: "unselected" if $FS(h,f)$ smaller than 0.2, "curved" if $FS(h,f)$ falls between 0.2 and 0.7, indicating a partially flexed finger posture, and "selected" if $FS(h,f)$ is greater than 0.7, signifying a finger that is actively selected by extending the finger in the formation of a sign shown in Eq. 2. This ternary categorization simplifies the interpretation of the finger's importance, distinguishing the overall handshape.

$$FS = \begin{cases} \text{unselected}, & \text{if } FS(h,f) \leq 0.2 \\ \text{curved}, & \text{if } 0.2 \leq FS(h,f) \leq 0.7 \\ \text{selected}, & \text{if } 0.7 \leq FS(h,f) \end{cases} \quad (2)$$

### 3.3. Orientation

The Orientation feature encompasses three main sub-features, each reflecting a distinct aspect of hand orientation in signing space during signing:

- **Palm-Back Score**: This score is derived from the relative orientation of the hand along the (x,y) axes, using the index knuckle and the pinky finger knuckle joints (Joints 5-17). It quantifies the extent to which the palm or back of the hand faces the interlocutor.
- **Radial-Ulnar Score**: Based on the hand's orientation along the (y,z) axes, this score also utilizes the index and pinky finger knuckle joints. It assesses the radial or ulnar deviation of the hand.
- **Tips-Wrist Score**: This score measures the orientation of the fingertips relative to the wrist along the z-axis, using the wrist and middle

fingertip joints (Joints 0-12). It captures the flexion or extension of the fingers relative to the wrist.

The granularity of phonological feature analysis is done by normalizing each orientation score to the absolute length of the signer's hand. This axis-specific normalization ensures that the resulting scores are relative to the signer's unique hand dimensions. Subsequently, these normalized scores are constrained within a [0,1] range for each feature tuple. We employ a softmax function to classify these orientation labels, which provides a probabilistic interpretation of each hand orientation.

$$\hat{O}(h) = \sigma(\frac{\sum_{ax}^{A} |p_{ax}^{1} - p_{ax}^{2}|}{||p^{1} - p^{2}||}) \quad (3)$$

The equation for deriving the orientation feature vector is formulated to capture the relative position of the hand in space. In this equation, $\hat{O}(h)$ represents the orientation feature vector for a hand $h$. The function $\sigma$ denotes the softmax function, which is applied to the sum of normalized differences across a set of axes $A$ for each feature used to define the orientation.

For each axis in $A$, the difference between the normalized joint positions $p_{ax}^{1}$ and $p_{ax}^{2}$ is calculated. These joint positions correspond to specific points on the hand, like knuckles or fingertips, relevant to the orientation being measured. The absolute value of this difference is then taken to ensure a non-negative measure of displacement. The normalization $\|p^{1} - p^{2}\|$ is the Euclidean distance between the two joints for each hand, serving as the denominator in the equation, which scales the orientation score relative to the size of the hand.

### 3.4. Location

Location analysis involves determining the relative positioning of each hand to major and minor locations (namely, Head, Nose, Ear, Mouth, Torso, Shoulder, and Chest). The technique measures the distance between the center of each hand and these landmarks, scaling these distances to the minimum and maximum values observed in each video frame. This scaling normalizes the data, accommodating variations in signer physique and positioning relative to the camera, thus ensuring the reliability of our phonological feature extraction across diverse datasets. We represent the overall relativized locations as the unit vector $L$ of the distance between the center point of selection of the hand and all selected locations.

### 3.5. Movement

The Movement feature extraction is the most complex because the model synthesizes continuous phonological information derived from each hand's Finger Selection, Orientation, and Location analyses. Our framework models primary movement types (i.e. path movement, aperture change, and orientation change) while we are still working on modeling secondary movement types, which cannot be derived from changes in IF features (i.e. path-shape and temporal alignment properties). In this regard, although we do not provide a comprehensive movement feature set, we provide a basis for the derivation of the movement in accordance with the theoretical aspects of movement features.

To demonstrate that our framework lays a basis for deriving complex features within sign language corpora, we empirically test and display the practical implications of our model with case studies within the TID Sözlük dataset. These studies focus on phonological information complexity to substantiate theoretical claims about feature Dominance and Symmetry which we define in the next section.

## 4. Case Studies

Our framework's application in these case studies is primarily motivated by the need to empirically test and validate phonological theories in Turkish Sign Language (TID). Utilizing the TID Sözlük dataset, we apply our framework to quantify phonological complexity to derive dominance and symmetricity conditions. We have selected these two conditions regarding the theoretical discussion on these conditions indicating that the definitions are derived by difference or the similarity between information complexity between hands in two-handed signs. Earlier claims only provide hand configuration limitation on these conditions, while Eccarius and Brentari (2007) argue that each condition can be defined as the maximization of the difference in phonological information (Dominance) or the minimization (Symmetry) which is the initial motivation for selecting as our case studies.

### 4.1. Constraints on Two-handed Signs

The constraints on two-handed signs, concerning Dominance and Symmetry (Battison, 1978) where the Dominance Condition articulates that in two-handed signs if handshapes differ, one hand (typically the non-dominant, passive hand, or weak hand) adopts an unmarked handshape. These unmarked handshapes are typically simpler in structure. Eccarius and Brentari (2007) extends this by discussing featural complexity, positing a limit to the featural complexity permissible in a sign.

The study also introduces the Featural Symmetry Condition, which posits that signs reduce their featural complexity by making the two hands mirror each other regarding selected fingers and orientation changes in the articulation of a sign. This suggests a balance or trade-off in complexity within the sign, resonating with the Dependency model, which views sign language structure in terms of interdependent features.

By applying these theoretical constructs to the TID dataset within our framework, we aim to provide empirical evidence for these phonological constraints. Our approach mathematically quantifies phonological complexity and symmetry, allowing us to test and validate the theoretical claims posited by phonological theory in sign language.

$$C(h) = \frac{1}{|Ch|} \sum_{ch \in Ch}^{Ch} abs(\Delta_h[f](ch)) \tag{4}$$

Equation 4 defines the phonological complexity $C(h)$ for a hand $h$ by averaging the absolute changes in phonological features across a set of channels $C$. In this context, $Ch$ is a collection of channels, each representing a different aspect of phonological information, namely finger selection, orientation, and location. The function $\Delta_h[f](ch)$ is the absolute forward finite difference function that calculates the change in a specific phonological feature $f$ within the channel $ch$ from adjacent frames. We obtain a measure of total phonological change by taking the absolute value of this change and summing it across all channels. This sum is then normalized by the number of channels $|Ch|$, resulting in an average measure of complexity for the hand across all considered phonological features. This computation allows for the quantification of complexity in a sign, providing a scalar value that can be used to analyze and compare the phonological structures within sign language corpora.

Additionally, in refining our understanding of two-handed sign constraints within the categorizations and lists the unmarked handshapes for TID Kubuş (2008). These definitions are particularly relevant in evaluating the performance of our framework when retrieving lexical items. The research outlines a set of unmarked handshapes specific to TID, which serve as a benchmark for assessing phonological complexity and dominance in two-handed signs.

### 4.2. Dataset

The TID Sözlük Dictionary (Makaroğlu and Dikyuva, 2017) is a comprehensive online corpus for Turkish Sign Language. It includes over 3000 isolated lexical items and within-sentence examples for each synonym. This dataset is not only a valuable educational resource but also a rich corpus for linguistic analysis, as it contains annotated handshape and location information for each lexical variant. In our study, the distribution of handshapes from this dataset serves as a basis for examining symmetry and dominance, allowing us to assign a scalar value representing the phonological complexity for each hand.

### 4.3. Case Study on Dominance Condition in TID

Feature Dominance in sign language phonology posits that in two-handed signs, the less active hand, designated as $h_2$, should exhibit lower phonological complexity compared to the more active hand. This principle reflects the asymmetry often observed in the phonological structure of sign languages, where the dominant hand carries more articulatory burden.

To quantify and utilize the phonological complexity between hands in demonstrating Dominance within data, we define a ranking function for retrieving the signs that maximize the difference in complexity score.

$$\operatorname*{argmax}_{H \in V} f(H) = \{\{h_1, h_2\} \in H \mid |C(h_1) - C(h_2)|\} \tag{5}$$

Equation 5 is the ranking function $f(H)$ designed to order signs based on the maximization of phonological complexity differences between the hands. In the given sign, $H$ represents the set containing pairs of hands, where $h1$ is typically the more active or dominant hand, and $h2$ is the less active or non-dominant hand. The function $C(h)$ computes the phonological complexity for a given hand $h$.

The ranking function operates by identifying the pair of hands $(h1, h2)$ within the set $H$ that has the largest absolute difference [1] in phonological complexity $|C(h1) - C(h2)|$. The argmax operator is applied to select the pair $(h1, h2)$ for which this absolute difference is maximized across all possible hand pairs in the dictionary $V$. This approach inherently ranks signs in a way that emphasizes the contrast in complexity between the two hands, reflecting the dominance condition where the less active hand is expected to demonstrate less phonological complexity compared to the more active hand. The function provides a quantitative basis for ordering signs by their adherence to this phonological principle.

Investigating the Top-100 retrieved signs that exhibit the highest difference in complexity scores, we examine the distribution of handshapes for the non-dominant hand. This analysis reveals a correlation with the unmarked handshapes for TID, suggesting that less active hands tend to favor sim-

[1] We should note that some of the signs have the higher complexity score in left hand given dominance hands are marked general handedness of signers which is mostly right hand. It should be noted for additional studies.

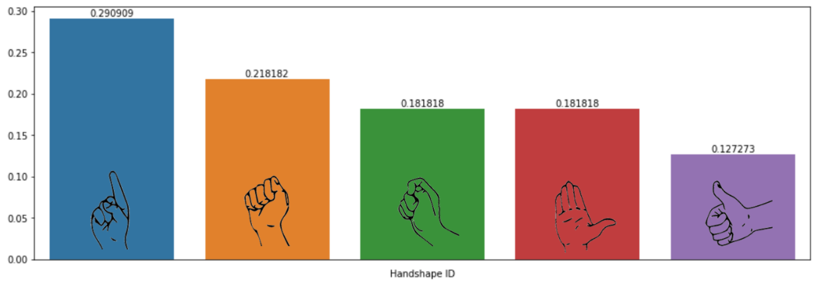


Figure 3: The handshape distribution of non-dominant hands (h2) for the Top-100 signs with the highest dominance ranking

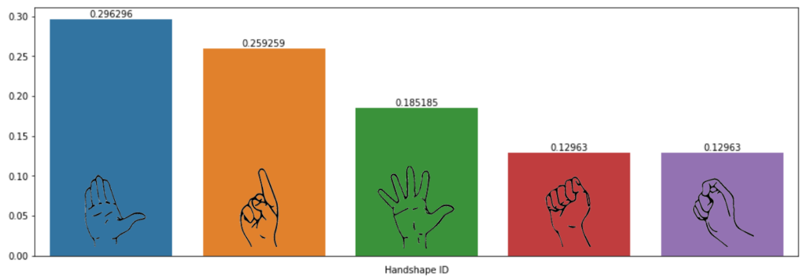


Figure 4: The handshape distribution for the Top-100 signs with the highest symmetricity ranking

pler, unmarked configurations, as shown in Figure 3.

A manual annotation process assesses the accuracy of the signs retrieved by our model, resulting in a 0.90 accuracy success rate in identifying dominant-hand signs shown in Table 1. This high degree of accuracy underlines the effectiveness of our phonological complexity formulation in predicting feature dominance within the signs.

### 4.4. Case Study on Symmetry Condition in TID

In contrast to feature dominance, the feature symmetry condition, suggests that two-handed signs should exhibit similar phonological features across both hands. This condition is motivated by featural symmetry, where both hands are expected to have similar finger selections, orientations, and movements, often resulting in unmarked handshapes.

To accommodate the symmetry condition, we revise our ranking function to focus on the minimization of phonological information complexity differences between hands as shown in Equation 6. This adjustment allows us to evaluate the degree of symmetry in the phonological structure of each sign by identifying and prioritizing those with the least complexity difference between the hands.

$$\underset{H \in L}{\operatorname{argmin}} f(H) = \{\{h_1, h_2\} \in H \mid |C(h_1) - C(h_2)|\} \tag{6}$$

Following the re-ranking of signs according to the updated function, we investigate the distribution of handshapes, particularly looking for the occurrence of unmarked shapes that would be indicative of symmetry. We then assess the accuracy of our model's ability to detect symmetric signs. A higher accuracy rate in this assessment would support our framework's capability to model phonological complexity effectively and validate the feature symmetry condition in sign language phonology. Similar to the Dominance Condition, we also observed the high distribution of unmarked handshapes in Top-100 retrieved sign as shown in Figure 4.

### 4.5. Results

In the dominance condition analysis, the model demonstrated high performance. This accuracy is attributed to the framework's capability to maximize the phonological information differences between the hands, which is a direct quantification of the dominance condition. The results were consistent with theoretical expectations, affirming the model's validity in discerning the more active hand's increased complexity. While still accurate, the analysis of the symmetricity condition revealed lower performance metrics compared to the dominance model. This outcome is due to the complexity of symmetricity, which is not solely about minimizing differences between hands but each hand should yield lower complexity separately. This dual requirement supports theoretical assertions of Eccarius and Brentari (2007) and highlights the additional constraints involved in modeling symmetricity within sign language phonology. Nevertheless, the baseline scores provided for the retrieval of symmetricity are still relatively high for further studies.

| Label | Acc. | Pre. | Rec. | F1 |
|---|---|---|---|---|
| Dominance | 0.90 | 1.00 | 0.90 | 0.95 |
| Symmetric | 0.84 | 0.71 | 0.84 | 0.77 |

Table 1: The results of the performance of retrieved Top-100 signs with highest Dominance and Symmetric ranking

## 5. Future Work

Further developments in our framework will address the integration of movement features, which are dynamic and complex components of sign language. We plan to utilize neural network models to effectively model these features, which can learn and generalize from large datasets. These models can potentially capture the temporal and spatial movement information across sign languages, translating them into meaningful phonological data that can be used for further linguistic analysis.

The ultimate goal of our research is to achieve a fully automated annotation process for sign language videos via advanced neural models. This automation will not only accelerate the annotation process but also enhance its accuracy, consistency, and scalability. As we integrate these advanced neural models, we will also re-evaluate and refine our annotation methodologies to ensure they remain robust and reliable for comprehensive sign language research.

## 6. Conclusion

In conclusion, SL-FE proves to be a transformative tool for sign language phonological analysis, adeptly bridging the gap between theoretical models and practical annotation. It offers a novel computational approach to quantify phonological complexity, providing empirical evidence for longstanding theoretical constructs. The case studies conducted with the TID dataset affirm the framework's capability to identify feature dominance and symmetry. Moreover, applying the two-handed sign criteria confirms the phonological constraints and others posited. As we continue to refine SL-FE, we anticipate its broader application in sign language research.

## 7. Acknowledgment

We would like to thank Lale Akarun and Umit Atlamaz for their contribution to reviewing the mathematical aspects of our framework and providing the necessary feedback and improvements.

## 8. Bibliographical References

Kairat Aitpayev, Shynggys Islam, and Alfarabi Imashev. 2016. Semi-automatic annotation tool for sign languages. In *2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)*, pages 1–4.

Robbin Battison. 1978. Lexical borrowing in american sign language.

Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign language recognition, generation, and translation: An interdisciplinary perspective. *CoRR*, abs/1908.08597.

Diane Brentari. 1998. *A prosodic model of sign language phonology*. Mit Press.

Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Anastasia Chizhikova and Vadim Kimmelman. 2022. Phonetics of negative headshake in Russian Sign Language: A small-scale corpus study. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 29–36, Marseille, France. European Language Resources Association.

Philippe Dreuw and Hermann Ney. 2008. Towards automatic sign language annotation for the ELAN tool. In *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 50–53, Marrakech, Morocco. European Language Resources Association (ELRA).

Petra Eccarius and Diane Brentari. 2007. Symmetry and dominance: A cross-linguistic study of signs and classifier constructions. *Lingua*, 117(7):1169–1201. The linguistics of sign language classifiers: phonology, morpho-syntax, semantics and discourse.

Jordan Fenlon, Kearsy Cormier, and Diane Brentari. 2017. The phonology of sign languages. In *The Routledge handbook of phonological theory*, pages 453–475. Routledge.

Esam Ghaleb, Ilya Burenko, Marlou Rasenberg, Wim Pouw, Peter Uhrig, Judith Holler, Ivan Toni, Aslı Özyürek, and Raquel Fernández. 2024. Co-speech gesture detection through multi-phase sequence labeling. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4007–4015.

Matilde Gonzalez, Michael Filhol, and Christophe Collet. 2012. Semi-automatic sign language corpora annotation using lexical representations of signs. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2430–2434, Istanbul, Turkey. European Language Resources Association (ELRA).

Marek Hrúz, Zdeněk Krňoul, Pavel Campr, and Luděk Müller. 2011. Towards automatic annotation of sign language dictionary corpora. In *International Conference on Text, Speech and Dialogue*, pages 331–339. Springer.

Onur Keleş, Ceren Oksal, and Emre Bilgili. 2023. Using pose estimation for reference tracking in turkish sign language. In *2023 31st Signal Processing and Communications Applications Conference (SIU)*, pages 1–4.

Okan Kubuş. 2008. An analysis of turkish sign language (tid) phonology and morphology. Master's thesis, Middle East Technical University.

Caroline Larboulette Lucie Naert, Clément Reverdy and Sylvie Gibet. 2018. Per channel automatic annotation of sign language motion capture data. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. Mediapipe: A framework for building perception pipelines.

Bahtiyar Makaroğlu and Hasan Dikyuva. 2017. *Güncel Türk İşaret Dili Sözlüğü*.

Liliane Momeni, Hannah Bull, KR Prajwal, Samuel Albanie, Gül Varol, and Andrew Zisserman. 2022. Automatic dense annotation of large-vocabulary sign language videos. In *European Conference on Computer Vision*, pages 671–690. Springer.

Medet Mukushev, Arman Sabyrov, Madina Sultanova, Vadim Kimmelman, and Anara Sandygulova. 2022. Towards semi-automatic sign language annotation tool: SLAN-tool. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 159–164, Marseille, France. European Language Resources Association.

Sylvie C. W. Ong and Surendra Ranganath. 2005. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(6):873–891.

Victor Skobov and Yves Lepage. 2020. Video-to-HamNoSys automated annotation system. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 209–216, Marseille, France. European Language Resources Association (ELRA).

Rose Stamp, Lilyana Khatib, and Hagit Hel-Or. 2022. Capturing distalization. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 187–191, Marseille, France. European Language Resources Association (ELRA).

Harry Van der Hulst. 1993. Units in the analysis of signs. *Phonology*, 10(2):209–241.